ACKNOWLEDGEMENTS

A elaboração deste trabalho não teria sido possivel sem a ajuda, companheirismo, generosidade e boavontade de muitos. Assim, gostaria de agradecer a todos aqueles que de alguma forma contribuíram para a sua concretização:

À Doutora Paula Jorge, minha orientadora, cujo espírito científico foi fundamental ao longo de todas as etapas que conduziram à realização deste trabalho. Desde o início disponibilizou o seu conhecimento, espírito crítico e amizade, incentivando-me a concretizar esta etapa da minha vida.

À Professora Doutora Maria Begoña Criado, minha co-orientadora, queria agradecer o apoio, orientação e amizade durante o meu trabalho e a minha formação académica.

A todos os elementos da Unidade de Genética Molecular do Centro de Genética Médica Jacinto de Magalhães, que me acolheram com grande amizade. À Doutora Rosário Santos, gostaria de agradecer a partilha de conhecimentos e oportunidade de poder pertencer ao seu grupo de trabalho, possibilitando assim a execução deste estudo; à Joana Loureiro, por ceder os seus resultados da construção de haplótipos em populações controlo e FXS; à Márcia pela generosidade e disponibilidade para me ajudar com o Western Blot; ao Pedro que, gentilmente, me permitiu "monopolizar" a sua bancada; à Elisa pela disponibilidade constante e pela preparação das células de cultura e, em especial à Isabel, cujo companheirismo e dedicação foram fundamentais em vários momentos.

À Sara Ricardo pela sua generosidade, sem a qual não teria sido possível optimizar a técnica de detecção imunocitoquímica da FMRP, bem como a divulgação dos resultados.

Ao Professor António Santos e à Luisa pela ajuda concedida na análise estatística dos meus resultados.

A todos os Professores do Mestrado, pela contribuição para a minha formação académica.

Aos meus Colegas de Mestrado, pela partilha de experiências, que me enriqueceram como profissional e como pessoa.

Aos novos e velhos amigos que fazem parte da minha história.

Aos meus Pais e Irmão, pelo carinho, incentivo e dedicação diários, sem os quais não seria possivel alcançar os meus sonhos.

RESUMO

O défice intelectual representa uma forma de deficiência cognitiva, responsável por limitações a nível mental e dos comportamentos adaptativos e sociais. Quando associado ao cromossoma X constitui uma causa frequente de atraso mental monogénico que afecta sobretudo indivíduos do sexo masculino. As mutações nos genes *FMR1, AFF2* e *ARX* surgem como as principais causas genéticas. Os genes *FMR1 e AFF2* contêm regiões repetitivas polimórficas sujeitas a mutações dinâmicas que podem originar expansões patogénicas. No gene *ARX* o segundo exão representa um *hotspot* mutacional, visto conter regiões repetitivas que codificam alaninas, sendo a mutação c.429_452dup24 a mais frequente.

Com o objectivo de caracterizar uma população de risco, foi efectuado um rastreio molecular por PCR multiplex dos *hotspot* mutacionais dos genes *FMR1, AFF2* e *ARX* em amostras de indivíduos com défice intelectual e numa população controlo. Para avaliar o número e o padrão de interrupções [AGG] da região repetitiva do gene *FMR1* foi optimizada a técnica de Triplet-primed PCR, permitindo estudar a sua estabilidade. As variações de tamanho detectadas no exão dois do gene *ARX* foram caracterizadas por sequenciação.

As frequências alélicas dos genes *FMR1* e *AFF2* na população portuguesa são semelhantes a outras populações Caucasianas. Este trabalho demonstra que a variação do número de tripletos no gene *FMR1* não está relacionado com variações no número de tripletos do gene *AFF2* (Pearson's R: ρ >0,05), apesar de, a combinação dos alelos com trinta tripletos (no gene *FMR1*) e com catorze (no gene *AFF2*) ser a mais frequente (13,4%) (χ^2: ρ<0,05).

Os alelos *FMR1* com a primeira interrupção [AGG] após nove ou dez tripletos [CGG] demonstraram ter origens diferentes (χ^2: ρ <0,05; Pearson's R: ρ <0,05). Foi observada uma forte correlação entre os alelos de vinte e nove tripletos e o haplótipo SNP CTC, bem como entre os alelos com trinta repetições e a linhagem TTT (χ^2: ρ <0,05).

Do rastreio do gene *ARX* resultou a identificação da mutação c.429_452dup24 em três indivíduos pertencentes a duas famílias distintas e da variante polimórfica c.441_464del24 em três amostras distintas. Também foram detectadas pequenas inserções, *in-frame*, que codificam para alaninas (c.304GCG[11] e c.304GCG[13]), com consequências patogénicas ainda por determinar.

É evidente a importância do estudo das interrupções [AGG] e do respectivo *background* genético dos alelos *FMR1* para um diagnóstico clínico exacto. As mutações no gene *ARX* de patogenicidade conhecida, apesar de pouco frequentes, revelam que este gene deve ser tido em conta no rastreio genético do défice intelectual.

PALAVRAS-CHAVE: Défice intelectual associado ao cromossoma X; *FMR1; AFF2; ARX*; Interrupções [AGG].

ABSTRACT

Intellectual disability (ID) is a common form of cognitive impairment, responsible by a significant number of limitations in intellectual function and adaptive behavior. X-linked intellectual disability represents a common cause of monogenic mental retardation affecting mostly males. Among the genetic causes involved, mutations in *FMR1, AFF2* and *ARX* genes emerge as the major causes. *FMR1* and *AFF2* genes contain polymorphic repetitive regions susceptible to suffer dynamic mutations, a process that may give rise to pathogenic expansions. In *ARX* gene, the second exon represents a mutational hot spot as holds repetitive regions that codify to alanines. Among the mutations described, variations in polyalanine tracts, namely the c.429_452dup24, are among the most frequent.

Aiming the characterization of a population at risk for ID, a multiplex molecular screening was performed regarding mutational hotspots in *FMR1*, *AFF2* and *ARX* genes in intellectually-disabled individuals and in a control population. A Triplet-primed PCR technique was optimized to evaluate [AGG] interspersion pattern in *FMR1* repetitive regions to assess allele stability. Furthermore, *ARX* exon 2 size variations were characterized by sequencing analysis.

Allelic frequencies of *FMR1* and *AFF2* genes in the Portuguese population are similar to other Caucasian populations. The present work shows that the increase of triplet content in *FMR1* gene is independent of variations in *AFF2* triplet content (Pearson's R: ρ >0,05), although, the occurrence of 30 repeat-sized alleles (for *FMR1*) and 14 repeat-sized alleles (for *AFF2)* is the most frequent combination (13,4%) ($\mathcal{X}^2$: ρ<0,05).

FMR1 alleles with the first interruption after nine or ten [CGG] triplets demonstrated to have different origins ($\mathcal{X}^2$: ρ <0,05; Pearson's R: ρ <0,05). A strong correlation among the 29 repeat-sized alleles and the CTC SNP haplotype and between the 30 repeat-sized alleles and the TTT lineage ($\mathcal{X}^2$: ρ <0,05) was observed.

In *ARX* analysis the c.429_452dup24 mutation was identified in three individuals from two different families and the polymorphic variant c.441_464del24 in other three patients. Small insertions codifying for alanines were also detected (c.304GCG[11] and c.304GCG[13]) with a pathogenic effect yet to be determined.

The importance of studying *FMR1* alleles in terms of [AGG] interruption and genetic background is evident for an accurate clinical diagnosis. *ARX* mutations of known pathogenicity, although few frequent, are a strong indication that this genes should be included in routine screening of ID.

KEY WORDS: X-linked intellectual disability; *FMR1; AFF2; ARX;* [AGG] interruptions.

LIST OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

***AFF2* –** AF4/FMR2 family, member 2

ARX – Aristaless Related Homeobox

***ARX* ex2p –** *ARX* exon 2 portion

bp – Base pair

CNS – Central Nervous System

ddNTP – Dideoxynucleotides

DMSO – Dimethyl Sulfoxide

DNA – Deoxyribonucleic Acid

EDTA – Ethylenediamine Tetraacetic Acid

FXS – Fragile X Syndrome

FRAXA – X chromosome fragile site, *locus* A

FRAXE – X chromosome fragile site, *locus* E

FMR1 – Fragile Mental Retardation 1 gene

FMRP – Fragile Mental Retardation 1 Protein

FXPOI – Fragile-X associated Premature Ovarian Insufficiency

FXTAS – Fragile-X associated Tremor/Ataxia Syndrome

ID – Intellectual Disability

KH – Nuclear ribonucleoprotein K Homology domain

mRNA – Messenger Ribonucleic Acid

NES – Nuclear Export Signal

NLS – Nuclear Localization Signal

NS-XLID – Nonsyndromic X-Linked Intellectual Disability

PBS – Phosphate Buffer Solution

PCR – Polymerase Chain Reaction

PolyA – Polyalanine tracts

POP – Performance Optimized Polymer

RNA – Ribonucleic Acid

STRs – Short Tandem Repeats

SNPs – Single Nucleotide Polymorphisms

TP-PCR – Triplet-Primed Polymerase Chain Reaction

XLID – X-Linked Intellectual Disability

5'-UTR – 5' Untranslated Region

$[CGG]_9$n – Alleles with the first [AGG] interruption after nine [CGG] repeats

$[CGG]_{10}$n – Alleles with the first [AGG] interruption after ten [CGG] repeats

1. Introduction

1.1 X chromosome and intellectual disability

Intellectual disability (ID) is a common form of cognitive impairment characterized by diminished functioning of the central nervous system (CNS) resulting in significant limitations, both in intellectual and adaptive behavior, as expressed in conceptual, social and practical adaptive skills, manifested before 18 years of age (Chiurazzi *et al.,* 2008). Is defined by the presence of incomplete or arrested mental development, mainly characterized by the deterioration of specific functions at each stage of development that contribute to the overall level of intelligence, such as cognitive, speech, motor and socialization tasks (World Health Organization, 2008).

The origin of ID can be related with genetic or environmental causes that affect the development and functioning of the CNS at prenatal, perinatal or postnatal period. Genetic causes include chromosome aneuploidies, chromosome structural abnormalities, genomic disorders and monogenic diseases, which can be related with either moderate or severe forms. On the other hand, environmental factors such malnutrition, premature birth and prenatal infections play a significant role in the pathogenesis of mild ID (Chiurazzi *et al.,* 2008).

The evolutionary origin of autosomes and sex chromosomes differs and may have consequences in the distinct contribution of the X chromosome to brain functioning. The X chromosome has played a crucial role in the development and segregation of sexual characteristics, accumulating an unbalanced number of genes related with mental functions (Skuse, 2005). This chromosome possesses an over-representation of genes expressed in the brain and frequently implicated in ID. The biased gene content of the human X chromosome explains the large amount of sex-linked mental disability syndromes (Delbridge *et al.*, 2008).

X-linked intellectual disability (XLID) represents a common cause of monogenic mental retardation affecting mostly males. The X-linked conditions are easily identified in affected males due to their hemizygozity, who inevitably manifest a phenotype when harboring a mutant allele (Chiurazzi *et al.*, 2008). Sex determination in humans follows the XX/XY system, with random inactivation of one of the two chromosomes in every female cell and this dosage compensation system in females increases the complexity of X-linked inheritance. Theoretically, heterozygous females will present a milder disease phenotype (compared to males with the same mutation) or be phenotypically normal, depending on the function of the gene involved (Shoubridge *et al.,* 2010). However it is not uncommon to have X-linked disorders with variable degrees of penetrance or a skewed X-inactivation, phenomenon that hurdle the prediction of the phenotype.

Based on their clinical presentation XLID can be categorized into three classes: *syndromes*, encompassing multiple congenital anomalies compromising other organs beyond the brain; *neuromuscular disorders* presenting neurological and/or muscular symptoms such as epilepsy, dystonia, spasticity and muscle weakness in the absence of malformations; *nonspecific conditions,* where ID represent the only consistent clinical sign and their discrimination depends entirely on the determination of their causative gene (Chiurazzi *et al.,* 2008).

1.2 Human chromosome fragility

The human genome has more than one hundred and twenty different fragile sites identified to date (Schwartz *et al.,* 2006). These heritable specific *loci* on human chromosomes exhibit an increased frequency of non-random gaps, contractions or disruptions and constitute areas where chromatin fails to compact during mitosis. Depending on their induction and frequency within the population, fragile sites are classified as rare or common. Common fragile sites are considered to be an intrinsic part of the normal chromosome structure, present in all individuals, whereas rare fragile sites are encountered in the chromosomes of merely a small fraction of the population (<5%) (Lukusa *et al.,* 2008).

Understand the mechanism and molecular basis of fragile sites is essential to unravel and manage the human pathologies caused by changes in these sensible *loci*. It has been shown that, after *in vitro* induction, fragile sites were frequently involved in deletions and translocations, in sister chromatid exchanges, in intrachromosomal gene amplifications and in plasmid integration. They were also related with the *in vivo* occurrence of duplications and amplifications or other chromosomal changes associated with human congenital diseases (Lukusa *et al.,* 2008). At the present time, thirty one rare fragile sites have been identified in different specific families and the majority of them are folate sensitive (Schwartz *et al.,* 2006). Until this moment, seven folate sensitive fragile sites have been characterized at the molecular level: FRA10A at 10q23.3; FRA11B at 11q23.3; FRA16A at 16p13.11; X chromosome fragile site, *locus* A (FRAXA) at Xq27.3; X chromosome fragile site, *locus* E (FRAXE) at Xq28; FRAXF at Xq28 and, more recently, FRA12A at 12q13.1. All these fragile sites represent *loci* with expanded [CCG] or [CGG] trinucleotide repeat sequences near to a CpG island (Lukusa *et al.,* 2008).

The progressive process that leads to a pathogenic expansion constitutes a dynamic mutation. First, these repetitive sequences undergo progressive increases in copy number, perhaps due to unequal crossing over, failure of mismatch repair

mechanism or to replication slippage events. Although these mechanisms can be the cause of some of these expansions, the initial event leading to the instability of a normal allele is thought to be slipped-strand mispairing during deoxyribonucleic acid (DNA) replication (Lukusa *et al.,* 2008). Okazaki fragments formed by longer uninterrupted repetitive sequences will not be anchored by unique sequences at their ends, constituting a risk of suffering slippage. If slippage occurs, DNA repairing enzymes could add nucleotides to the template strand to make it complementary to the newly formed strand, generating an expansion (Penagarikano *et al.,* 2007).

Along with these possible points of error, the length of the repeat tract is a fundamental element for the expansion to occur. Before they reach a critical copy number, these fragile sites undergo a premutation phase, where the [CCG] or [CGG] repeats become highly unstable when transmitted to the next generation but are not enough to cause fragility expression in the carrier individual (Sinden *et al.,* 1999). Once the copy number overtakes a critical limit it becomes a full mutation and as consequence the adjacent CpG island is hypermethylated (Eichler *et al.,* 1994).

Of the seven folate sensitive fragile sites characterized, only two have proven clinical expression, FRAXA and FRAXE, being FRAXA the most important of the fragile sites with a well-documented clinical impact (Lukusa *et al.,* 2008). The expansion of over than 200 [CGG] triplets in the Fragile Mental Retardation 1 (*FMR1*) gene is associated with the Fragile X Syndrome [FXS; MIM#300624], the most common form of familial severe ID. FRAXE is associated with X-linked non-specific intellectual disability, a moderated form of mental retardation where a pathogenic expansion of [CCG] triplets occurs in the AF4/FMR2 family, member 2 (*AFF2*) gene (Verkerk *et al.,* 1991).

1.2.1 X chromosome fragility at *locus* A and the Fragile X Syndrome

The presence of fragility in FRAXA *locus* is directly related with FXS, an X-linked dominant disorder, with reduced penetrance in women, associated with intellectual and emotional disabilities. Affected individuals may present a series of phenotypes, in addition to a reduced intellectual ability, which ranges from mild to severe. Many FXS patients also display hyperactivity, hypersensitivity to sensory stimuli, anxiety, impaired visual and spatial processing, development delay and autistic behaviors. Physical features have been described but are often nonspecific (Garber *et al.,* 2008).

The main cause of this monogenic disorder is the transcriptional silencing of the *FMR1* gene, due to an expansion of over 200 [CGG] repeats found in the 5'-untranslated region (5'-UTR) and consequent hypermethylation of the promoter region. This

hypermethylation constitutes a requirement for the fragility expression. The *FMR1* gene [Xq27.3; MIM*309550] belongs to a small gene family, that includes the Fragile X Mental Retardation, autosomal homolog 1 [3q28; MIM*600819] and Fragile X Mental Retardation, autosomal homolog 2 [17p13.1; MIM*605339] genes (D'Hulst *et al*., 2009). This gene consists of 17 exons spanning approximately 40 kilobases (kb) and prior to exon 1, in the 5'-UTR, possesses a repetitive region, composed by [CGG] variable number tandem repeats fragments interspersed by [AGG] triplets (Figure 1). Some other authors believe that repeats are more stable when interspersed by [AGG] triplets (Poon *et al.*, 2006). The "purity" of the repeats favors the instability, the [CGG] repeats of most stable alleles are interspersed with [AGG] triplets, while alleles with expanded triplets show a reduction or absence of these interruptions. The variability in [CGG] repeat length allows categorize the *FMR1* alleles in normal-sized alleles (6-39 [CGG]s), intermediate-sized or gray zone alleles (40-54 [CGG]s), premutated-sized alleles (55-200 [CGG]s) and full mutated alleles (>200) (Pfeiffer *et al.,* 2009).

The *FMR1* promoter is CG-rich and lacks a typical TATA element, but contains three initiator-like (Inr) sequences that correspond to sites I-III. It has been demonstrated that transcription of the *FMR1* gene is initiated at three different start sites (I-III), for both neuronal and non-neuronal cells and, the relative use of the three main start sites is significantly altered depending on the size of the [CGG] repeat, which displays a direct influence on transcription initiation (D'Hulst *et al*., 2009).

According to the size of its repetitive region *FMR1* messenger ribonucleic acid (mRNA) and FMRP content may suffer deviations. Individuals presenting normal and intermediate-sized alleles display normal ranges of *FMR1* mRNA and FMRP. Individuals with alleles in a premutation range present increased levels of *FMR1* mRNA and decreased FMRP levels. In a full mutation situation, hypermethylation of the repetitive region occurs, compromising the transcription of *FMR1* mRNA (Oostra *et al.,* 2003).

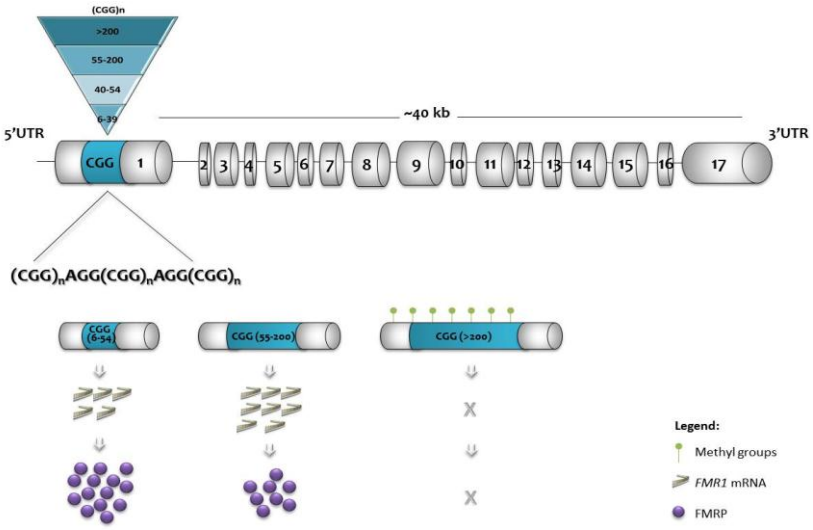


Figure 1. Fragile X Mental Retardation 1 gene structure. *FMR1* gene locates on the long arm of the X chromosome at position 27.3 and is composed by 17 exons spanning approximately 40 kb. In the 5'-UTR, is located the *FMR1* repetitive region, composed by [CGG] variable number tandem repeats interspersed by [AGG] triplets. (Adapted from: OOSTRA B. A. AND WILLEMSEN R. A fragile balance: *FMR1* expression levels, Hum Mol Genet, 2003, 12: 249 257 Review Issue 12)

Although the number of [CGG] repeats in the general population is highly variable, the majority of *FMR1* normal alleles have 29-30 repeats, interspersed by two stabilizing [AGG] triplets and are constant upon transmission (Oostra *et al.,* 2003). Interspersed [AGG] has been proposed to be the anchor which prevents DNA slippage during DNA replication. The loss of interspersed [AGG] results in a long pure [CGG] repeat sequence at the 3′ end which contributes to DNA instability. In general, the longer the 3′ pure [CGG] repeats, the more susceptible is the [CGG] to further expansion in later generations (Poon *et al.,* 2006).

In FXS families the premutated alleles usually have one or no interruptions. The smallest premutated allele (with 59 [CGG]) that expanded to a full mutation in one generation presented no interruptions, which support the general idea that the loss or lack of one of these interruptions is an important determinant for instability (Penagarikano *et al.,* 2007).

The genesis of this pathology obeys to a two-step expansion process as described by the Sherman paradox.[1] First, the common alleles have to expand to a premutation state of 55–200 unmethylated [CGG] repeats that are unstable during meiosis. Premutated alleles can be found in both males and females and may expand to a full

[1] The Sherman Paradox refers to an anomalous pattern of inheritance found in Fragile X syndrome. Sherman theorized that the *FMR1* gene becomes mutated through a two-step process. The first mutation is called a premutation and requires a second mutation to be converted into a full mutation leading to the clinical symptoms associated with FXS.

mutation only upon maternal transmission. The risk of transition is dependent on the size of the premutation. With a pathogenic expansion of over than 200 [CGG] repeats the alleles reach the condition of full mutation, causing the absence of the protein encoded (Oostra *et al.,* 2003; Oostra *et al.,* 2009).

Almost all fragile X patients carry an expanded full mutation and show methylation of the *FMR1* promoter region, with no other changes in the coding region. Rare male individuals with expanded but unmethylated repeats have been identified and are classified as *high functioning males*. Such individuals do not present the full spectrum of the FXS phenotype, demonstrating that methylation and not repeat elongation *per se* causes the typical features of this syndrome (Oostra *et al.,* 2009).

Silencing the expanded *FMR1* allele is a unique example of a local epigenetic change driven by variation in the gene sequence. Methylation of the *FMR1* gene acts inhibiting the binding of transcription factors and inducing chromatin condensation, which prevents the transcription machinery from binding (Penagarikano *et al.*, 2007). *FMR1* [CGG] repeat methylation is a dynamic process that occurs early in the embryonic development. The first methylation events occur in the fetal cells then, the extra-embryonic tissue is methylated to an increasing degree, as development progresses (Oostra *et al.,* 2009).

In nearly all cases, the FXS is caused by transcriptional silencing of *FMR1*. The lack of Fragile X Mental Retardation Protein (FMRP) is responsible for its phenotype. Other mutations, apart from the [CGG] expansion, could lead to FXS. Until now more than 15 deletions affecting part of or the entire gene have been described. Besides deletions, point mutations leading to the production of a nonfunctional protein have also been described (Penagarikano *et al*., 2007). An example of a small change described as pathogenic is the I304N missense mutation in *FMR1* exon 10, which compromises FMRP function. The I304N mutation maps to a position within the second KH domain of FMRP that is critical for stabilizing sequence-specific RNA-protein interactions and invalidates the association of the FMRP KH2 domain with its target, the kissing complex ribonucleic acid (RNA) (for more detail see Figure 2, page 17). Therefore, base pair changes ranging from one unit to the entire gene may disrupt the function of FMRP or prevent the protein from being formed (Weels *et al.*, 2009). It is important to consider that most diagnostic testing for FXS is limited to [CGG] repeat number assessment, ignoring other "conventional" mutations. Since the full mutation leads to a functional null allele, it is likely that much nonsense and missense mutations in *FMR1* gene are being missed.

The remarkable inheritance pattern of FXS results from the combination of the mutational mechanism with the location of the *FMR1* gene in the X chromosome. Indeed,

as observed in other dynamic mutations, the degree of instability depends on the gender of the transmitting parent, as the repeat expansion from the premutation to full mutation state predominantly occurs during maternal transmission. The maternal transmission bias could result from sex-specific differences in gametogenesis leading to the production of full expansion oocytes and premutation size repeats in sperm of carrier males. This is compatible with the observation that sperm from full mutation male patients present only premutated alleles (Oostra *et al*., 2009). Furthermore, full mutations are responsible for a replication delay of the *FMR1* gene, during the cell cycle, conferring a proliferative advantage of premutated germ cells that overgrow the full mutation germ cells (Oostra *et al*., 2009).

1.2.1.1 Phenotypic expression of premutated alleles

The clinical conditions associated with *FMR1* [CGG] premutation are Fragile-X Associated Premature Ovarian Insufficiency (FXPOI), previously described as Premature Ovarian Failure [POF; MIM#311360] and the Fragile X-associated Tremor/Ataxia Syndrome [FXTAS; MIM#300623].

Premutation carriers are at risk of developing a progressive neurodegenerative disorder, clinically and neuropathologically distinct from FXS, the FXTAS which affects mainly men aged more than fifty (Jacquemount *et al.,* 2003). Although both disorders are associated with expansions in *FMR1* gene, the clinical presentation and molecular mechanism underlying each disease are completely different. In adulthood, FXTAS carriers manifest a progressive action tremor and cerebellar ataxia, complemented by a progressive cognitive decline and behavioral difficulties, including memory loss, executive function deficits, anxiety, reclusive behavior and dementia in the most severe cases. The major neuropathological evidence and postmortem standard for a definitive diagnosis of FXTAS is the presence of intranuclear inclusions, eosinophilic and ubiquitin-positive, throughout the brain neurons and astrocytes (Oostra *et al.,* 2009).

FMR1 mRNA level is elevated five to ten times in premutation carriers, while the FMRP presents slightly reduced levels or no detectable change (Figure 1). To explain the FXTAS phenotype a RNA gain of function theory has emerged, based on the observation of elevated levels of expanded *FMR1* mRNA (Jin *et al.,* 2003). Indeed, these mRNAs have been shown to be toxic in human neural cell culture as in transgenic flies. In fly model of FXTAS, these altered noncoding mRNAs have been demonstrated to cause the formation of inclusion bodies and lead to neurodegeneration. In premutated cells the excess of [CGG] transcripts may attract and sequester other proteins with affinity to these

regions, affecting RNA metabolism, increasing cellular toxicity and leading to progressive cell death. This phenomenon is particularly important in the brain since it has the highest expression of *FMR1* gene. A combination of [CGG] repeat length and *FMR1* mRNA abundance may define a threshold for the clinical manifestations of FXTAS (Jin *et al.,* 2003).

This condition is more frequently associated with male premutation carriers as few women with FXTAS have been reported. The neurologic and cognitive involvement of women is smaller compared to affected males and this gender difference may be related to the random X inactivation pattern in women (Al-Hinti *et al.,* 2007). Although not associated in a syndrome, women with premutated alleles and psychiatric symptoms have been described (Roberts *et al*., 2009).

Roughly twenty per cent of women who carry premutation alleles experience inherited premature ovarian insufficiency, clinically defined as early depletion of ovarian follicles before the age of 40 (Rohr *et al.,* 2008). Other symptoms have been registered at higher frequencies among premutation carriers who are still cycling, including increased gonadotropins, altered menstrual cycle characteristics, infertility and increased levels of follicle stimulating hormone, an indicator of reduced ovarian function (Hunter *et al.,* 2008).

The risk of premutation carriers develop FXPOI is about twenty times higher than women bearing alleles within a normal range, being the alleles with eighty to one hundred repeats those who constitute the highest risk (Rohr *et al.,* 2008). Repeat size variation within the premutation range explains only a significant proportion of the variation in FXPOI as carriers of both smaller and larger premutation repeat sizes also suffer from FXPOI. The molecular mechanism behind FXPOI stills unraveled, although mRNA toxicity mediated by expanded *FMR1* transcripts is a possible candidate. However, further studies are necessary to elucidate the molecular and cellular pathways of FXPOI (Rohr *et al.,* 2008 and Hunter *et al.,* 2008).

Neither of these disorders (FXTAS and FXPOI) is seen in conjunction with FXS and are therefore thought to be distinct entities caused by a toxic gain-of-function effect of increased levels of the expanded *FMR1* gene transcripts (Hunter *et al.*, 2008).

1.2.1.2 Fragile Mental Retardation Protein

FMRP and its autosomal paralogs, fragile-X-related protein 1 and 2 (FXR1 and FXR2), consist of a small family of RNA-binding proteins, the fragile X-related gene family (D'Hulst *et al.,* 2009). As a result of intensive alternative splicing events, especialy in the 3' terminal half of the *FMR1* gene, twelve different protein isoforms, ranging from 67-80

kDa, can potentially be detected *in vivo*, with the most common having approximately 80 kDa (Schaeffer et al., 2003).

As a selective RNA-binding protein that associates with polyribosomes and negatively regulates local protein synthesis, FMRP plays a fundamental role in the development and plasticity of neurons (Garber *et al*., 2008). Due to FMRP's association with a variety of mRNAs, absence or malfunction of this protein affects many neuronal processes. FMRP is present in nearly all human cell types, being expressed primarily in the neurons of synapse-rich fetal and adult tissues, such as the hippocampus and cerebellum. It contains both nuclear localization and export signals, allowing FMRP to bind nuclear mRNA and further deliver to actively translating cytoplasmic polyribosomes (where 85% of cellular FMRP is located). FMRP also possesses the ability to repress translation of mRNAs at ribosomes (Pfeiffer *et al.,* 2009).

The best characterized motifs in FMRP are those that interact with RNA: two heterogeneous nuclear ribonucleoprotein K Homology (KH) domains and an RGG box containing a conserved Arg-Gly-Gly triplet. Both nuclear localization signal (NLS) and nuclear export signal (NES) are found in FMRP, suggesting that FMRP shuttles between the nucleocytoplasmic space. Two coiled coils (CC) involved in protein-protein interaction have also been identified (Figure 2) (Oostra *et al*., 2009 and D'Hulst *et al*., 2009).

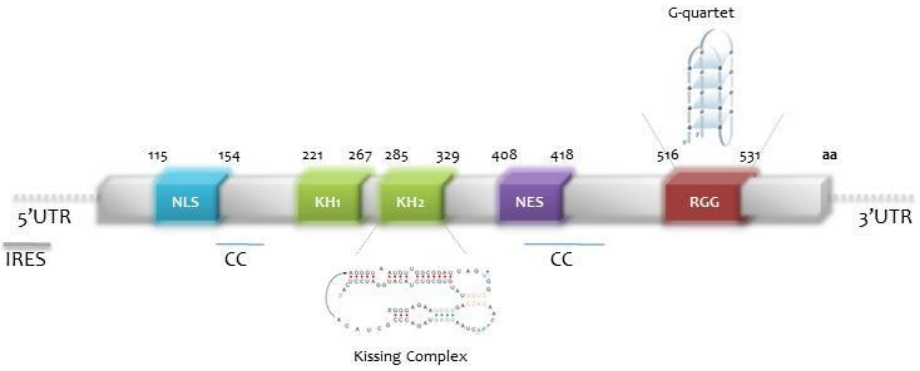


Figure 2. Fragile Mental Retardation Protein. FMRP is an RNA binding protein with a maximum of 632 aminoacids.

The specificity of interaction is dictated by the affinity of this protein for certain structures formed by mRNAs. FMRP binds a significant percentage of brain mRNAs and has a preference for three classes of mRNAs that contain a G-quartet structure, a kissing complex or a U-rich sequence. Therefore, two interactions have been described until now, KH2 domain links to a specific mRNA conformation denominated as Kissing Complex and RGG box to an RNA loop structure known as G-quartet. RNAs with these structures can

compete for the association of FMRP with polyribosomes, suggesting that this structure could be crucial for FMRP translational regulation (Penagarikano *et al*., 2007). An indirect or more complex interaction between FMRP and its target mRNAs could occur. The association of FMRP with the microRNA pathway represents a new way how FMRP modulates translation (D'Hulst *et al*., 2009).

1.2.2 X chromosome fragility at *locus* E and Non-Syndromic Intellectual Disability

ID associated with fragile *locus* FRAXE belongs to a class of non-syndromic X-linked intellectual disability (NS-XLID) that affects 1/50 000 newborn males (Chiurazzi *et al*., 2008). The disorder is the consequence of the silencing of *AFF2* gene [*AFF2*; MIM*300806] due to a [CCG] expansion in the region upstream to this gene (Bensaid *et al*., 2009). This NS-XLID condition is a form of mild to moderate intellectual disability associated to learning difficulties, communication deficits, attention problems, hyperactivity and autistic behavior. Since the FRAXE phenotype seems to be relatively mild and lacks dysmorphic features, this prevalence could be underestimated (Abrams *et al.,* 1997).

AFF2 gene is organized in 22 exons, showing several possibilities of alternative splicing for exons 2, 3, 5, 7 and 21, being translated into a 1,311 aminoacid nuclear protein (Figure 3). The longest of the *AFF2* isoform is composed of 1272 aminoacids and contains two nuclear localization signal sequences. Together with AF4/FMR2 family, member 1 [*AFF1*; MIM*159557] and AF4/FMR2 family, member 3 [*AFF3*; MIM*601464], forms a new family of nuclear proteins with DNA-binding capability and transcription transactivation potential. AFF2 is an RNA-binding protein that may be involved in alternative splicing regulation through an interaction with G-quartet RNA structure (Bensaid *et al*., 2009).

The repetitive region of *AFF2* is naturally polymorphic in the general population, presented as a pure repeat with no interspersed sequences. Normal-sized alleles (6-30 [CCG]s) are stable upon transmission, whereas intermediate-sized alleles (31-60 [CCG]s) may slightly vary in number between parents and their offspring. Alleles in a premutation range (61 to 200 [CCG]s) may expand to a full mutation state (>200 [CCG]s), which leads to methylation of the promoter region of *AFF2* gene and, consequently to mental impairment. The threshold between the normal and the premutated state is still poorly defined as the mechanism behind *AFF2* gene mutation is understudied and rare (Santos *et al.,* 2001).

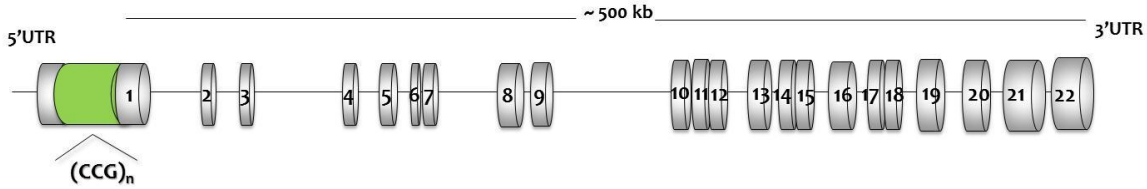


Figure 3. AF4/FMR2 Family member 2 gene structure. The *AFF2* gene locates on the long arm of the X chromosome at position 28.0 and is composed by 22 exons spanning approximately 500 kb. Prior to exon 1, in the 5'-UTR, the *AFF2* repetitive region is located, composed by uninterrupted [CCG] variable number tandem repeats.

The expansion mutation in *locus* E belongs to a category of dynamic mutations, as is the case of *locus* A. The mutational mechanism is represented by an inherited pathologic expansion of unstable [CCG] repeats, increased to more than two hundred hypermethylated triplets. The [CCG] content of *AFF2* repetitive region can either expand or contract, "apparently" being equally unstable when transmitted through the male or the female germ line (Bensaid *et al.*, 2009 and Santos *et al.,* 2001).

1.3 Aristaless Related Homeobox gene

The Aristaless Related Homeobox (*ARX*) gene [Xp22.13; MIM#300382] encodes the ARX protein, which belongs to a large family of homeodomain transcription factors. This relatively small gene locates at Xp22.13 presenting a genomic region of approximately 12,5 kb constituted by five coding exons (Figure 4). It is transcribed into a single mRNA of 2,8 kb and encodes a protein product with 562 amino acids. ARX is mainly expressed in the cerebral forebrain and floor plate, where exerts primary development functions. The concentration of ultra-conserved elements in the neighboring areas of *ARX* gene may suggest that they form a cluster of enhancers of this gene, which reinforces its importance in the CNS development (Gécz *et al*., 2006). Analysis of the distribution of these noncoding ultraconserved elements demonstrates that they tend to cluster in regions that are enriched for transcription factors and developmental genes (Bejerano *et al.,* 2004).

The ARX protein belongs to one of the largest classes of homeoproteins, the paired (*Prd*) class, thought to embrace important regulators of the fundamental events occurred during the vertebrate embryogenesis (Friocourt *et al.,* 2010). The homeodomain proteins act as nuclear transcriptional regulators via activation or repression of eukaryotic transcription. ARX contains four distinct polyalanine tracts (polyA), composed by recurrent alanine residues, three nuclear localization sequences, a central acidic domain and three highly conserved domains: two DNA binding domains, the homeodomain and the aristaless domain; and an octapeptide domain (Ohira *et al*., 2002; Gestarini-Duarte *et al*., 2006; Laperuta *et al.,* 2007).

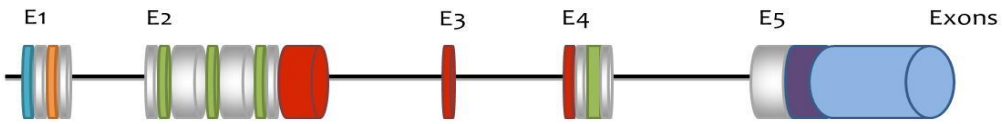


Figure 4. Aristaless Related Homeobox gene structure. Five exons of the *ARX* gene are boxed together with the open reading frame in grey. The known functional domains are highlighted: octapeptide domain is shown as an orange rectangle on exon 1, polyA tracts are shown as four green rectangles in exons 2 and 4, homeodomain sequences spanning exons 2, 3 and 4 are shown as red boxes and aristaless domain is shown in exon 5 as a purple box. (Adapted from: SHOUBRIDGE, *et al*., *ARX* Spectrum Disorders: Making Inroads into the Molecular Pathology, Hum Mutat, 2010, 31: 889-900.)

Since its identification, several mutations have been identified in *ARX* gene including insertions, deletions, duplications, missense, nonsense and splice mutations, leading to a wide spectrum of phenotypes that include syndromic as well as nonsyndromic

forms of intellectual disability (Shoubridge *et al.*, 2010). Even though there are mutations described across all extent of the gene, the majority of mutations occur in the largest coding exon of *ARX* gene, the exon two. The most frequent mutation in *ARX* gene involves a twenty-four base-pair duplication on exon 2 that results in an expansion of the second polyA in the ARX protein from twelve to twenty alanine residues, giving rise to a series of phenotypes (Laperuta *et al*., 2007; Suri *et al*., 2005).

The dominant feature of all *ARX* mutations is ID, which constitutes a NS-XLID when presented as the sole consistent clinical feature. More frequently, ID is found in conjunction with other clinical signs. Phenotypes of the *ARX* mutations can be classified as non-malformation and malformation phenotypes (Shoubridge *et al.,* 2010). The additional clinical features include epilepsy, infantile spasms, autism, dystonia and dysarthria in non-malformation phenotypes or lissencephaly, hydrancephaly, abnormal genitalia and agenesis of the corpus callosum in malformation cases (Gestarini-Duarte *et al*., 2006; Gronskov *et al*., 2004).

1.3.1 *ARX* mutations leading to non-malformation phenotypes

Several mutations described in *ARX* gene are related with phenotypes that do not include brain malformations. As described in other alanine expansion disorders, the severity of clinical presentation increases with the length of the expansion (Shoubridge *et al*., 2010). ARX has four polyA tracts (polyA I, polyA II, polyA III, polyA IV) coded by several trinucleotide repeats (GCN). Although these repeats are stable during meiosis and mitosis, polymorphisms in length are frequent, involving more than 30% of tracts longer than seven alanines (Conti *et al.*, 2010).

Expansion of the polyA II from twelve to twenty alanines is a consequence of the most frequent *ARX* mutation, the c.429_452dup24 (previously described as 428_451dup24). The majority of patients display nonsyndromic intellectual disability, ranging from moderate to severe and some of them present dystonic movement of the hands and dysarthria, symptoms that characterize the Partington Syndrome [PRTS; MIM#309510]. The West syndrome or X-linked Infantile Spasms [ISSX; MIM#308350] has also been related with this *ARX* mutation. This mutation presents an unusual variation of the clinical manifestations and the environmental determinants involved on the phenotypic outcome of c.429_452dup24 are yet to be identified (Shoubridge *et al.,* 2010).

The c.304CGC[23] mutation, also reported as c.313_333dup21, occurs in polyA I and represents the second most frequent mutation in *ARX*. This recurrent mutation leads to the addition of seven alanines to a tract of sixteen, totalizing twenty three alanines. It

has been suggested that this particular mutation may cause protein aggregation, as has been demonstrated with *in vitro* transfection of c.304CGC[17] constructs (Shoubridge *et al.,* 2007). In fact, these mutant constructs stimulate protein aggregation, filamentous nuclear inclusions and increase cell death. In addition to those, a shift from nuclear to cytoplasmic localization of ARX protein has been detected. This new localization may contribute to a partial loss of function and the resulting pathogenesis. The clinical presentation of this mutation is variable and includes severe intellectual disability, generalized dystonia, infantile spasms and a hypsarrhythmic electroencephalogram pattern (Shoubridge *et al.,* 2007; Friocourt *et al*., 2010).

New mutations affecting polyA II have recently been identified. A *de novo* duplication of twenty seven base pairs (c. 430_456dup27) leading to an expansion from twelve to twenty-one alanines was described in a family with three affected males, presenting infantile spasms with a consistent early age of onset and death early in life (Reish *et al*., 2009). Other mutation described in the same region, c.423_455dup32, originate a less severe phenotype probably due to an interruption by a glycine after the tenth alanine residue. The mutations in *ARX* gene that cause non-malformation phenotypes, besides polyA mutations, illustrate the strong phenotypic heterogeneity of the *ARX* mutations and the associated pathologic consequences (for more detail see Appendix 1 – Table 15) (Shoubridge *et al*., 2010; Bienvenu *et al.*, 2002).

1.3.2 ARX mutations leading to malformation phenotypes

The absence of ARX protein is directly relate with alterations in brain development, such poorly developed olfactory bulbs and reduced volumes of the cerebral cortex and hippocampus (Laperuta *et al*., 2007). So far, the described *ARX* gene mutations that lead to brain malformation phenotypes are predominantly correlated with X-linked Lissencephaly with Ambiguous Genitalia [XLAG; MIM#300215]. These mutations include insertions, deletions, nonsense mutations, splice mutations, single nucleotide substitutions clustered in the homeodomain and the region prior to the aristaless domain (for more detail see Appendix 1 – Table 15) (Shoubridge *et al.,* 2010).

FXS emerges as the main cause of XLID. However, recent developments regarding *ARX* gene place it in a position that may rival FXS. Several studies suggest that *ARX* mutations as cause of ID may be underestimated, becoming *ARX* gene in a serious candidate for gene screening in routine DNA diagnostic practice in cases of ID.

2. OBJECTIVES

The present work represents a retrospective study of a Portuguese population at risk for intellectual disability (ID), referred for molecular diagnosis of Fragile X Syndrome that was excluded by Southern blot analysis. The first goal is the characterization of this population in a gene cluster based approach that comprises *FMR1*, *AFF2* and *ARX* genes. This approach shall allow the characterization of the Portuguese population regarding allele size variability of both the *FMR1* and *AFF2* gene repeat regions as well as identify variations in the second exon of *ARX* gene. The molecular strategy applied will putatively improve the underdiagnosed cases of conditions associated to these genes.

A second purpose is the assessment of the [AGG] interspersion pattern of the [CGG] region present in the 5' UTR of *FMR1* gene, namely the comparison of number and pattern variability between ID and control populations. The control population was detailed studied in terms of number, location of the [AGG] interspersions and its influence on allele stability. Using the previously established SNP haplotype of the control population those results shall be analyzed in terms of lineage background aiming to ascertain their ancestry. A subgroup of samples from the ID population with intermediate-sized alleles was also studied concerning [AGG] pattern. Altogether this information will contribute to understand the role of [AGG] interspersions and the effect of the number of [AGG] interruptions in allele stability and expansion risk assessment. Understanding the behavior of *FMR1* alleles will be of great importance in terms of genetic counseling for individuals at risk.

The third aim of this work consists on gathering clinical and molecular information in a spreadsheet for further epidemiological studies of the Portuguese ID population and database construction.

3. MATERIALS AND METHODS

3.1 Samples

Samples were carefully selected based on the family history of intellectual disability and clinical indications for Fragile X testing. All gDNA samples were previously extracted by the *salting out method* (Miller *et al*., 1988) from peripheral blood collected in ethylenediamine tetraacetic acid (EDTA), dissolved in EDTA buffer and stored at 4ºC, some of which for more than a decade.

A total of 1259 gDNA samples, with a normal Fragile X analysis by Southern Blot, were screened for variations in the repetitive regions of *FMR1* and *AFF2* genes and in a portion of the second exon of *ARX* gene (*ARX* ex2p) that includes two stretches of repetitive triplets codifying for alanines. The clinical information was assembled in an anonimized database that fulfill the requirements of the articles 4th, 17th and 18th of the law nº 12/2005 of 26 January published on the Portuguese Republic Diary, regarding the use of personal genetic and health information.

A set of 198 anonimized samples of a control population which do not present ID was used. Two other populations were established for [AGG] interspersion assessment: one composed by ID normal-size alleles and other by ID intermediate-size alleles.

3.2 Methodologies

3.2.1 Multiplex Polymerase Chain Reaction (mPCR)

The method established for molecular screening consisted in the combination of three primer pairs amplifying target regions of the X chromosome genes: *FMR1*, *AFF2* and *ARX*. Primers were labeled with different fluorescent molecules (6-FAM or HEX) depending on the amplicons size (Table 1).

Table 1. Primers used on mPCR.

Primer	Sequence	Labeling
g.FMR1_CGG_R*	5’ CCA TCT TCT CTT CAG CCC TGC 3’	(6-FAM)
g.FMR1_CGG_F	5’ TTC GGT TTC ACT TCC GGT G 3’	---
g.AFF2_CCG_F*	5’ TGT GAG TGT GTA AGT GTG TGA TGC TGC C 3’	(HEX)
g.AFF2_CCG_R	5’ CCG CGC GCA CCC AGC GAC 3’	---
g.ARX_2F	5’ CAG CAG CCC TGG CTG GGA CTC 3’	---
g.ARX_2R*	5’ CGG TAC GAC TTG CTG CGG CTG 3’	(HEX)

The PCR components and its concentrations were optimized for a multiplex PCR reaction (Table 2). The amplification reaction, performed in a 9800 Fast Thermal Cycler (Applied Biosystems), was performed as follows: 10' incubation at 98^{o}C; 42 cycles of denaturation at 98^{o}C for 45'', annealing at 58^{o}C for 45'' and extension at 68^{o}C for 2' 30''; final extension of 10' at 68^{o}C. After amplification, fragments were submitted to capillary electrophoresis.

The analysis of PCR amplified fragments was performed by capillary electrophoresis in *3130xl* Genetic Analyzer (Applied Biosystems). Following the PCR, the resultant products were mixed with Hi-Di^{TM} Formamide (Applied Biosystems) and the ladder, GeneScan^{TM} 500 ROX Size Std (Applied Biosystems). Once on the *3130xl* Genetic Analyzer (Applied Biosystems), each sample runs on a 36 cm capillary fulfilled with a running matrix, the Performance Optimized Polymer (POP 7 – Applied Biosystems) for 20' at 60^{o}C, with a voltage of 15 kV. Then samples are submitted to a D36 filter together with the Multi Capillary DS-30 (Dye set D) Matrix Standard Kit (Applied Biosystems). The resulting data were collected by the Foundation Data Collection v.3.0 Software (Applied Biosystems) and analyzed resorting to the GeneMapper Software version 4.0 (Applied Biosystems).

Table 2. Multiplex PCR components.

Reagents	
Distilled water (dH_2O)	Up to 25,0 µL
Betaine (5M)	1,0M
Accutaq buffer (10x) (Sigma)	1x
dATP (10mM) (Bioline)	0,12 mM
dCTP (10mM) (Bioline)	0,12 mM
dTTP (10mM) (Bioline)	0,12 mM
dGTP (10mM) (Bioline)	0,02 mM
7-deaza dGTP, Li-Salt (10 mM) (Roche)	0,4 mM
Dimethyl sulfoxide (DMSO) 100% (Sigma)	4,8 %
g.FMR1_CGG_F (10ρmol/µL)	10,0 ρmol
g.FMR1_CGG_R (10ρmol/µL)	10,0 ρmol
g.AFF2_CCG_F (10ρmol/µL)	6,0 ρmol
g.AFF2_CCG_R (10ρmol/µL)	6,0 ρmol
g.ARX_2F (5ρmol/µL)	5,0 ρmol
g.ARX _2R (5ρmol/µL)	5,0 ρmol
AccuTaq polymerase (5U/µL) (Sigma)	1,75 U
gDNA (300ηg/µL)	300,0 ηg

3.2.1.1 Allele size determination

The determination of both *FMR1* and *AFF2* allele sizes was accomplished resorting to sequenced controls with a known [CGG] and [CCG] content and by the application of the next formula:

$$\text{Triplet number} = \left[\frac{|A - B|}{3} \right] + C$$

"A" represents the sample distance in base pairs (bp) of migration during the capillary electrophoresis, which is directly proportional to the size of the fragment in question, "B" represents the control sample distance of migration and "C" the number of triplets of the control sample.

Due to inaccuracy in sizing when interpreting the results based on PCR and capillary electrophoresis, we decided to establish a maximum error of +/- 4 % of the total repeat size (for both *FMR1* and *AFF2* gene) and an error of 0,2% of fragment size, in base pairs, for *ARX* ex2p analysis.[2] Alleles were classified according to its size (Table 3).

Table 3. Alleles size classification.

Alleles	Normal	Intermediate	Premutation	Full mutation
***FMR1* [CGG]**	6-39	40-54	55-200	>200
***AFF2* [CCG]**	6-30	31-60	61-200	>200

3.2.2 gDNA Pool Polymerase Chain Reaction

A gDNA pool based methodology was designed to allow the screening of a larger number of samples for size variations in *ARX* ex2p (Figure 5). Thus, conditions were optimized for a pool of six different gDNAs, submitted to a standard PCR followed by capillary electrophoresis resolution. As control, one of the gDNA pools containing two samples, a deletion of 12 bp and an insertion/duplication of 24 bp was simultaneously amplified within each 96 well plate. All gDNA samples from a pool presenting abnormal sizes were further submitted to a new PCR, separately, in order to assess which of the six samples presented a size variation. Additionally, in order to confirm these variants sequencing analysis was performed.

[2] Size determination error is in agreement with The European Molecular Genetics Quality Network (EMQN) guidelines available at: www.emqn.org.

The PCR mix was designed having in mind the cost-benefit ratio (Table 4). The amplification reaction, performed in a 9800 Fast Thermal Cycler (Applied Biosystems) follows the same amplification and capillary electrophoresis conditions as previously described (for more detail see 3.2.1).

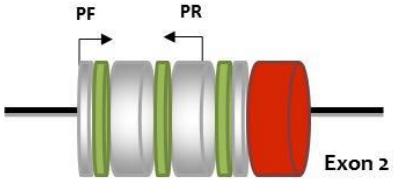


Figure 5. *ARX* ex2p amplification scheme. The primer pair used (g.ARX_2F (PF) and g.ARX_2R (PR)) allows amplification of a specific region of *ARX* exon two, that includes both polyalanine tracts I and II (first two green cylinders), and comprises the nucleotides 197-62 to 527.

Table 4. gDNA pool PCR components.

Reagents

dH_2O	Up to 38,0 µL
Betaine (5M)	0,7M
Multiplex PCR Master Mix (2x) (Applied Biosystems)	1x
DMSO 100% (Sigma)	6,6%
g.ARX2_F (10ρmol/µL)	13,0 ρmol
g.ARX2_R (10ρmol/µL)	13,0 ρmol
gDNA pool (300ηg/µL each sample)	~200,0 ηg each sample

3.2.3 Triplet-Primed Polymerase Chain Reaction

Traditionally, PCR amplification of unstable gDNA sequences involves the use of a pair of *locus* specific primers that flanks the variable trinucleotide sequence. In 1996, Warner *et al*. suggested a simple and elegant PCR modification, the TP-PCR for repeat expansions assessment. TP-PCR uses triplet priming to produce a set of amplicons, characteristic for the expanded allele, rather than amplifying across the full length of the gDNA fragments (Figure 6).

Based on the literature, it was possible to design a simple, rapid PCR-based method with capillary electrophoresis allowing the qualitative detection of [CGG] repeats in the *FMR1* gene repetitive region, as well as the presence of [AGG] triplets.

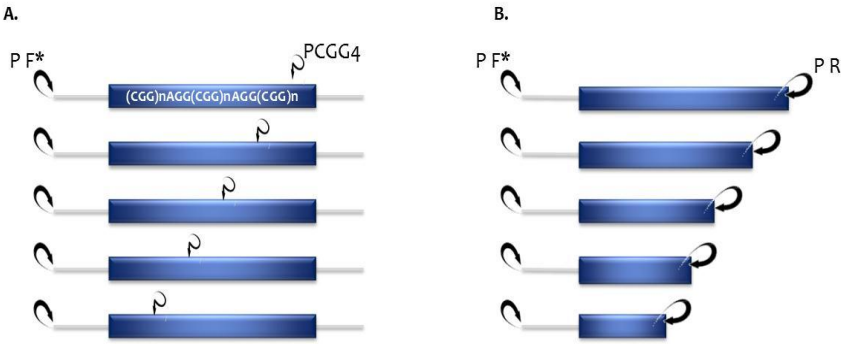


Figure 6. Triplet Primed Polymerase Chain Reaction method. A – The first phase consists of an amplification with a *locus* specific primer flanking the repeat together with a primer containing 4 [CGG] triplets, recognizing multiple priming sites within the repetitive region, and tagged with a unique sequence; B – The resulting multiple-sized fragments are then amplified, and reaction specificity is dictated by the same forward primer together with a reverse primer which is complementary to the tag in the $PCGG_4$. PF* - forward fluorescent labeled primer; $PCGG_4$ – reverse primer with a 4 [CGG] tract and a tail; PR – reverse primer.

TP-PCR components and concentration were optimized, including gDNA and additives concentration, primers combination and amplification program. (Tables 5 and 6). The amplification reaction, performed in a Verity 96 Well Thermal Cycler (Applied Biosystems), was performed as follow: 10' incubation at 98^{o}C; 48 cycles of denaturation at 98^{o}C for 1', annealing at 60^{o}C for 1' and extension at 68^{o}C for 6'; final extension of 10' at 68^{o}C. After amplification, fragments were submitted to capillary electrophoresis on a *3130xl* Genetic Analyzer (Applied Biosystems) as previously described (for more detail see 3.2.1).

Table 5. Primers used in TP-PCR.

Primer	Sequence	Labeling
g.FMR1_CGG_R*	5' CCA TCT TCT CTT CAG CCC TGC 3'	(6-FAM)
g.FMR1_CGG_F	5' TAC GCA TCC CAG TTT GAG ACG 3'	---
g.FMR1_P4_CGG	5' TAC GCA TCC CAG TTT GAG ACG CGG CGG CGG CGG CG 3´	---

Table 6. Triplet-primed PCR components.

Reagents	
Betaine (10M)	2,6M
Multiplex PCR Master Mix (2x) (Applied Biosystems)	1x
$MgCl_2$ (100mM) (Bioline)	5,1 mM
DMSO 100% (Sigma)	6,4%
g.FMR1_CGG_F (10ρmol/μL)	10,0 ρmol
g.FMR1_CGG_R (10ρmol/μL)	10,0 ρmol
g.FMR1_P4_CGG_F (5ρmol/μL)	7,5 ρmol
gDNA (300ηg/μL)	600,0 ηg
Final volume	39,0 μL

3.2.4 Sequencing procedure

Before the sequencing PCR, samples are submitted to an asymmetric amplification reaction. The PCR mix, according to the target gene regions is described on tables 7, 8 and 9. The amplification reaction for *FMR1* and *AFF2* genes, performed in a Verity 96 Well Thermal Cycler (Applied Biosystems) was as follows: 7' incubation at 96°C; 40 cycles of denaturation at 96°C for 45'', annealing at 60°C for 45'' and extension at 68°C for 3'; final extension of 15' at 68°C. For *ARX* sequencing, the reaction was performed with 10' incubation at 98°C; 45 cycles of denaturation at 96°C for 45'', annealing at 58°C for 45'' and extension at 68°C for 2' 30''; final extension of 15' at 68°C.

Table 7. Primers used for Sequencing PCR.

Primer	Sequence
g.FMR1_CGG_F	5' TGA CGT GGT TTC AGT GTT TAC 3'
g.FMR1_CGG_R	5' AGC CAA GTA CCT TGT AGA AAG CG 3'
g.AFF2_CCG_F	5' TGT GAG TGT GTA AGT GTG TGA TGC TGC C 3'
g.AFF2_CCG_R	5' TAG CCC GCG CAC CCA GCG AC 3'
g.ARX_2F	5' CAG CAC CGC TCC GTG GGA CTC 3'
g.ARX_2R	5' CGG TAC GAC GAC TTG CTG CGG TG 3'

Table 8. *FMR1* symmetric PCR components.

Reagents	
dH_2O	Up to 50,0 μL
Imomix Red (2x) (Bioline)	1x
Betaine (1,5M)	0,45 M
g.FMR1_CGG_R (10ρmol/μL)	10,0 ρmol
g.FMR1_CGG_F (10ρmol/μL)	10,0 ρmol
gDNA (300ηg/μL)	300,0 ηg

Table 9. *AFF2* and *ARX* symmetric PCR components.

Reagents	*AFF2* PCR	*ARX* PCR
Distilled water	Up to 25,0 µL	Up to 25,0 µL
Betaine (5M)	1M	1M
Accutaq buffer (10x) (Sigma)	1x	1x
dATP (10mM) (Bioline)	0,12 mM	0,12 mM
dCTP (10mM) (Bioline)	0,12 mM	0,12 mM
dTTP (10mM) (Bioline)	0,12 mM	0,12 mM
dGTP (10mM) (Bioline)	0,02 mM	0,02 mM
7-deaza dGTP, Li-Salt (10 mM) (Roche)	0,4 mM	0,4 mM
DMSO 100% (Sigma)	4,8%	4,8%
g.AFF2_CCG_F (10ρmol/µL)	10,0 ρmol	---
g.AFF2_CCG_R (10ρmol/µL)	10,0 ρmol	---
g.ARX_2F (10ρmol/µL)	---	10,0 ρmol
g.ARX _2R (10ρmol/µL)	---	10,0 ρmol
AccuTaq polymerase (5U/µL) (Sigma)	1,75 U	1,75 U
gDNA (300ηg/µL)	12 ηg	300,0 ηg

To perform the sequencing PCR, components concentration were optimized. PCR mix, according to the gene submitted to sequencing reaction was composed as described on table 10. The amplification reaction, performed in 9800 Fast Thermal Cycler (Applied Biosystems) was started with 6' incubation at 95^{o}C followed by 27 cycles of denaturation at 95^{o}C for 10'', annealing at 50^{o}C for 5'' and extension at 60^{o}C for 4', the last step consisted in a final extension of 10' at 60^{o}C. When the sequencing reaction was completed, the products were purified, and then analyzed by capillary electrophoresis (for more detail see Appendix 2 – section 9.2.4).

Before being submitted to capillary electrophoresis the samples were eluted with Hi-Di^{TM} Formamide (Applied Biosystems). The conditions of capillary electrophoresis were similar to those applied to fragment analysis, with the exception of the run voltage applied of 13,2 kV and the running matrix replaced by the Big Dye Terminator v1.1 Matrix Standard Kit (Applied Biosystems). The resulting data were collected by the Foundation Data Collection v.3.0 Software (Applied Biosystems).

The variations in the sequences of *FMR1*, *AFF2* and *ARX* genes were established by comparing the sequences obtained with reference sequences. Using Sequence Scanner version 1.0 (Applied Biosystems) software that uses the reference sequences *FMR1* (NM_002024), *AFF2* (NM_002025) and *ARX* (NM_139058) genes it was possible to characterize all the variations found. The nomenclature of variations identified was

attributed according to the Human Genome Variation Society guidelines available at www.hgvs.org.

Table 10. Sequencing Reaction components.

Reagents	
dH_2O	Up to 15,0 µL
Big Dye Terminator v1.1 Cycle Sequencing Mix (Applied Biosystems)	2,0 µL
Betaine (5M)	1,0 M
Specific primer*	2,5/10/5 ρmol*
Purified symmetric PCR product	5 µL

*Primers should be used as follow: g.FMR1_CGG (2,5ρmol/µL); g.AFF2_CCG (10,0ρmol/µL); g.ARX_2 (5,0ρmol/µL), either for forward or reverse sequencing.

3.2.5 Database construction

After carefully analysis of clinical files, relevant information was gathered and, afterwards, complemented with results corresponding to molecular analysis of the *FMR1*, *AFF2* and *ARX* genes. The information assembled includes date, place of birth and sex, clinical manifestations and family history when referred. All the information obtained from molecular analysis was included in the database to characterize the population at a molecular level.

3.2.6 Statistical analysis

Statistical analysis was performed resorting to IBM® SPSS® Statistics version 19.0 software. A Chi-square Test of Independence was executed, with a significance level (ρ) of 0,05 and a confidence of 95%, to test relations between lineage background, positions of the first [AGG] interspersion and allele size and between the thirty-size *FMR1* alleles and the fourteen-size *AFF2* alleles.

Correlation measures were performed resorting to Pearson Correlation Coefficient, T-student Test or Interclass Correlation Coefficient. To test if an association between the variation in *FMR1* and *AFF2* genes size a Pearson Correlation Coefficient (ρ = 0,05) and the T-student test were applied. To assess how closely related the *FMR1* and *AFF2* allelic sizes are it was calculated the Cronbach's alpha (Interclass Correlation Coefficient), a measure of internal consistency. To test possible associations between the thirty-size *FMR1* alleles and the fourteen-size *AFF2* alleles a Pearson Correlation Coefficient was used.

4. Results and Discussion

4.1 Epidemiological evaluation

The analysis of the clinical signs of individuals referred for molecular diagnosis of Fragile X Syndrome (FXS) reveals that psychomotor delay emerges as the main symptom (23,5%) along with mental delay (22,5%). Other clinical manifestations and respective frequencies are represented on Figure 7. The majority of individuals included in this study were males (73,4%), probably due to the inheritance pattern of X-linked recessive intellectual disability, where males are the most affected as they inevitably manifest a phenotype when harboring a mutant allele on the X-chromosome (Chiurazzi *et al.*, 2008). The average of age, at the time of the molecular diagnosis request, was 11,6 years old. Furthermore, the age group of five to ten years old is significantly represented by 31,3% of the assessed individuals.

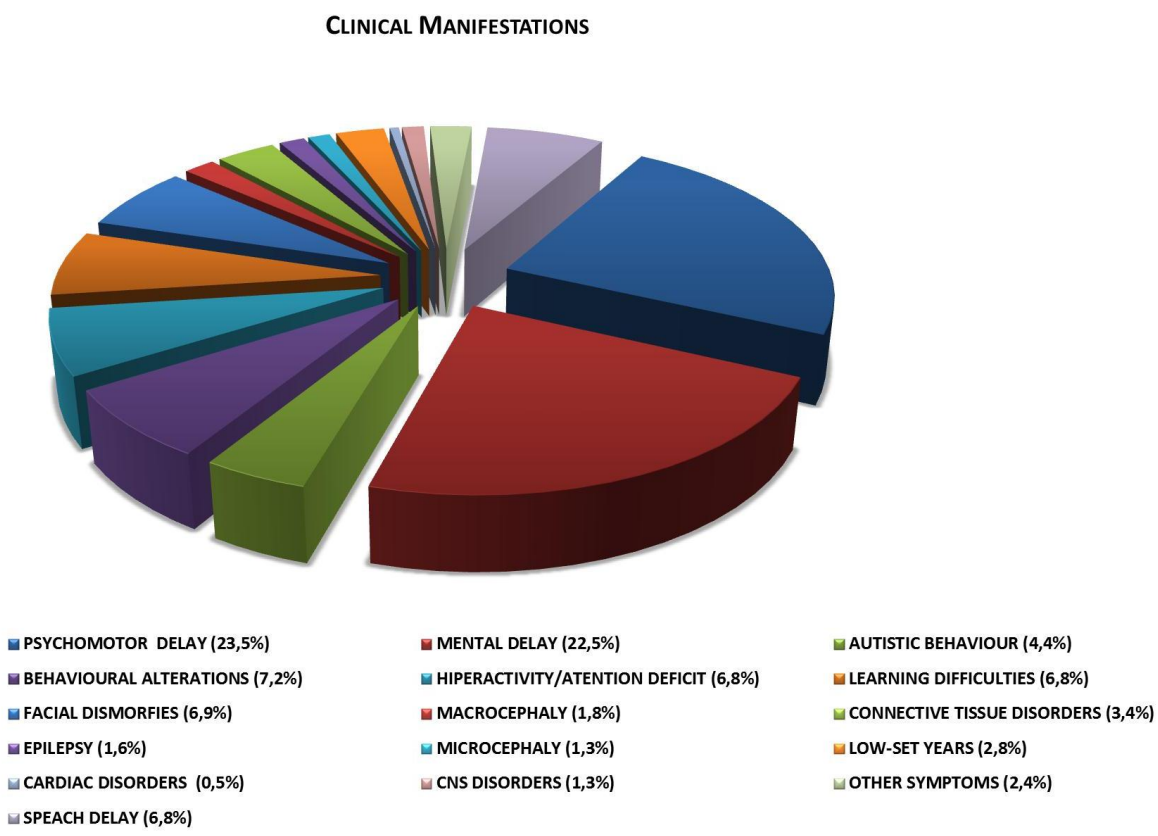


Figure 7. Frequencies of the different clinical manifestations of individuals with idiopathic intellectual disability (n=996).

4.2 X-linked intellectual disability assessment

Allele characterization of *FMR1* and *AFF2* repetitive regions, and detection of deletions/insertions in exon 2 of *ARX* gene was done using multiplex PCR (mPCR). This technique allows the identification of normal, intermediate and premutated-size alleles of *FMR1* and *AFF2* genes (Figures 8 A, B, and C) and distinction of size variations in *ARX* exon *2 (ARX* ex2p) (Figures 8 D and E). By amplifying these three genes at once it was possible to design a cost effective and a less time consuming technique, allowing to screen a considerable number of samples.

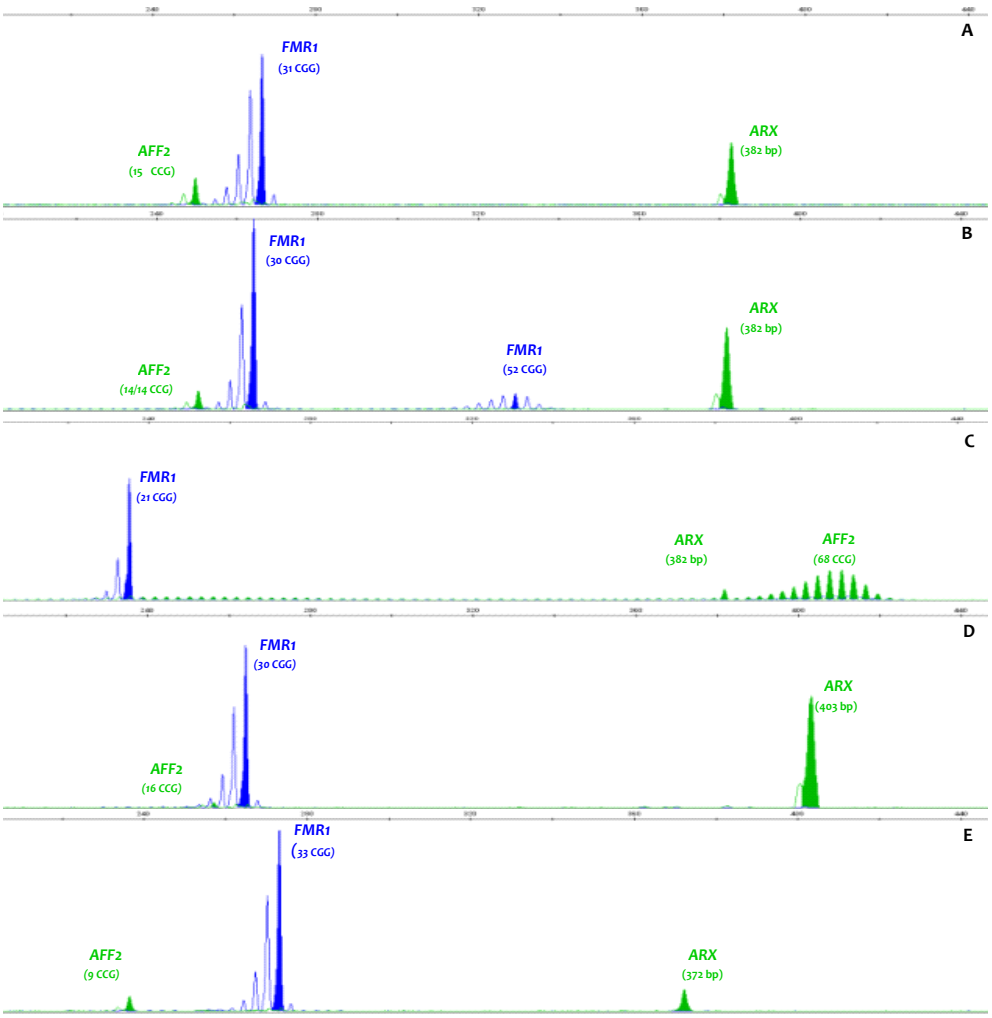
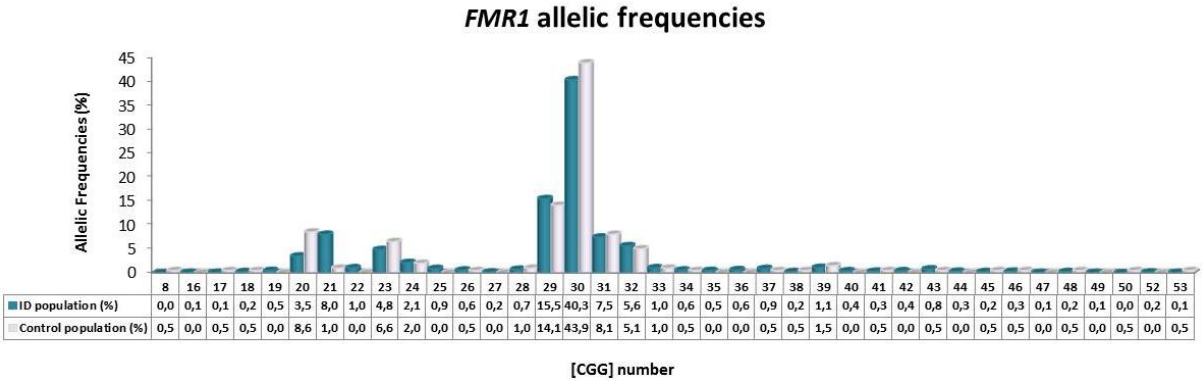


Figure 8. Electropherograms corresponding to multiplex PCR analysis of *FMR1, AFF2* and *ARX* genes. A – Profile of normal male individual. B – Female individual presenting one intermediate-size allele for *FMR1* gene, with approximately 52 [CGG] repeats, and a possible homoallelism for *AFF2* gene. C – Male individual presenting one *AFF2* allele in a premutation range, with approximately 68 [CCG] repeats. D and E – Samples from male individuals with *ARX* ex2p size variations, a 24 bp duplication and a 12 bp deletion, respectively.

FMR1 alleles presented a bimodal distribution in both the control and the ID populations. The allele with thirty repeats is the most frequent in both ID (40,3%) and control (43,9%) populations, followed by the twenty-nine triplet repeat alleles (15,5% and 14,1%, respectively) in ID and control populations.



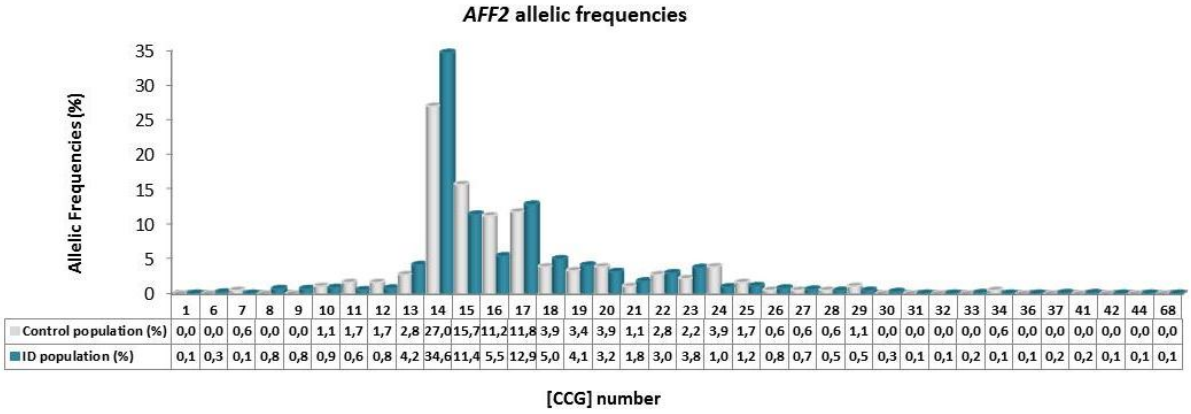
	8	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	52	53
ID population (%)	0,0	0,1	0,1	0,2	0,5	3,5	8,0	1,0	4,8	2,1	0,9	0,6	0,2	0,7	15,5	40,3	7,5	5,6	1,0	0,6	0,5	0,6	0,9	0,2	1,1	0,4	0,3	0,4	0,8	0,3	0,2	0,3	0,1	0,2	0,1	0,0	0,2	0,1
Control population (%)	0,5	0,0	0,5	0,5	0,0	8,6	1,0	0,0	6,6	2,0	0,0	0,5	0,0	1,0	14,1	43,9	8,1	5,1	1,0	0,5	0,0	0,0	0,5	0,5	1,5	0,0	0,5	0,0	0,5	0,0	0,5	0,5	0,0	0,5	0,0	0,5	0,0	0,5

Figure 9. Distribution of *FMR1* gene alleles. The distribution and frequency of different alleles are indicated for both control and intellectual disabled population. The control population was composed by a total of 198 alleles and the ID population comprised 1259 alleles.

The results obtained in the Portuguese population, is similar to other populations where the alleles with twenty-nine and thirty repeats are the most common. Caucasians (Kunst and Warren, 1994) as well as African Americans and Sub-Saharan West Africans (Peprah *et al.*, 2010), South-Americans comprising Brazilians (Mingroni-Netto *et al.*, 2002) Chilean (Jara *et al*., 1998) and several Asian populations including Chinese, Malays (Zhou *et al.,* 2006) and Taiwanese (Chiu *et al*., 2008) also present this alleles as the most frequent. There are some exceptions, like the Japanese (Otsuka *et al.,* 2009) and Mexicans (Rosales-Reynoso *et al.,* 2005), in which the twenty-seven and thirty-two repeats alleles are the most frequently encountered, respectively. Thus, it seems that either the twenty-nine or thirty repeat alleles are the ancestral alleles, founder effects due to geographic and cultural isolation or complex interactions between populations have changed the distribution of frequencies (Eichler and Nelson, 1996).

Apart from Japanese, both Asians and Caucasians share similar frequencies of twenty-nine and thirty triplet repeats alleles, but the mode is different in the two racial groups (Faradz *et al*., 2000). In a study of the Chinese population the most common repeat number was twenty-nine, with thirty being the second most common (Zhong et al., 1994). Other differences between Asians and Caucasians include a second modal peak at 36 repeats and the absence of modal peaks at 20 or 23 repeats observed in the Asian population (Zhong et al., 1994).

AFF2 alleles size distribution has a higher heterogeneity when compared to *FMR1* alleles. The mode of [CCG] repeat number observed in the Portuguese population is fourteen, in both control (27,0%) and ID (34,6%) populations (Figure 10). *AFF2* repeat distribution has been reported for several ethnic groups, with differences even among Caucasians (Murray *et al*., 2000). Comparison of absolute repeat number between studies is difficult due to intra-laboratory dissimilarity. However, a study where allele sizing was, presumably, standardized for the five populations tested, clearly demonstrated differences in modal repeat number: the predominant English mode is fifteen, whereas in the African, Chinese, Greek and Indian populations the mode is sixteen to eighteen (Ritchie *et al.,* 1997). Considering the error we have established we can clearly compare the Portuguese population with other European individuals where the mode of [CCG] triplets is similar to our population.



	1	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	36	37	41	42	44	68
Control population (%)	0,0	0,0	0,6	0,0	0,0	1,1	1,7	1,7	2,8	27,0	15,7	11,2	11,8	3,9	3,4	3,9	1,1	2,8	2,2	3,9	1,7	0,6	0,6	0,6	1,1	0,0	0,0	0,0	0,0	0,6	0,0	0,0	0,0	0,0	0,0	0,0
ID population (%)	0,1	0,3	0,1	0,8	0,8	0,9	0,6	0,8	4,2	34,6	11,4	5,5	12,9	5,0	4,1	3,2	1,8	3,0	3,8	1,0	1,2	0,8	0,7	0,5	0,5	0,3	0,1	0,1	0,2	0,1	0,1	0,2	0,2	0,1	0,1	0,1

Figure 10. Distribution of *AFF2* gene alleles. The distribution and frequency of different alleles are indicated for both control and ID population. The control population was composed by a total of 178 alleles and the population of intellectual disabled individuals comprised of 1190 alleles.

One allele in the premutation range was detected with sixty eight [CCG] repeats (Figure 10), which confirms the low rate of expansion thought to be associated to *AFF2* gene repetitive region. An intergenerational study by Murray *et al.* (2000), states that in 4124 transmissions analyzed (3824 female and 300 male), all alleles greater than sixty repeats were unstably transmitted. Furthermore, any expansion of a premutation to a full mutation was detected, suggesting that progression to a full mutation is a gradual process even from a relatively unstable premutation-size allele. There is evidence that male subjects with clinical evidences of action tremor, Parkinsonism, and cerebellar ataxia may have Fragile X premutations (Annesi *et al.*, 2004). In a study of 203 male subjects with Parkinson's disease (PD) and 370 healthy controls, *AFF2* intermediate alleles (31–60

repeats [CCG]) were found in 6.4% subjects with PD and in only 0.27% healthy controls ($P < 0.001$), thus indicating that "relatively" large *AFF2* alleles may be associated with PD (Annesi *et al.*, 2004).

Several *AFF2* alleles were sequenced and, contrary to *FMR1* alleles, no interruptions where found, only pure [CCG] sequences (Figure 11). Other studies in normal-sized alleles also support this observation (Zong *et al.*, 1996; Murray *et al.*, 2000). The absence of interspersions may suggest a simple relationship between size of the repeat and likelihood of expansion (Zong *et al.*, 1996).

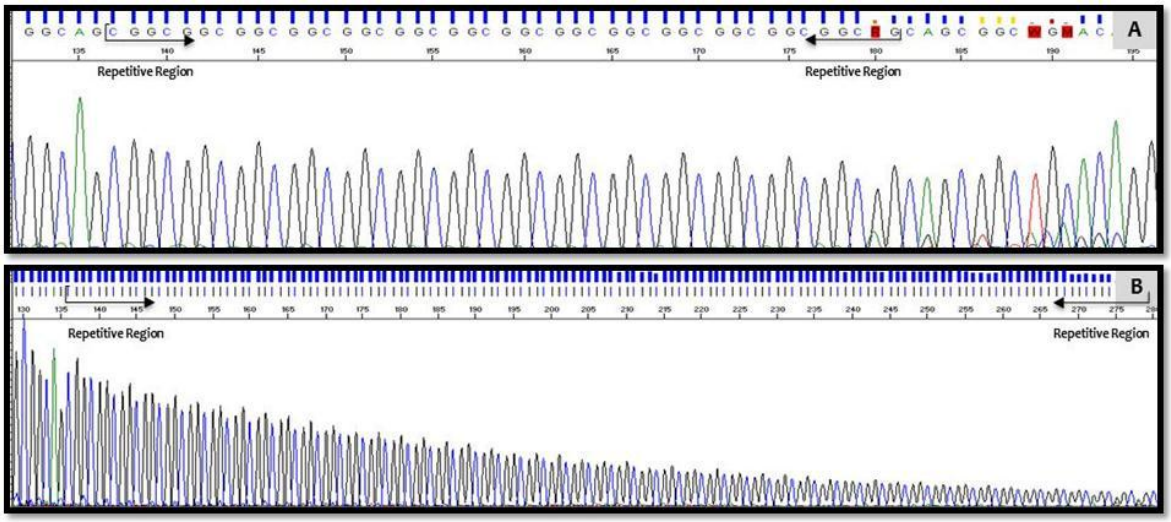


Figure 11. *AFF2* repetitive region sequencing analysis. The evaluation of alleles with different sizes revealed the absence of interruptions in all sequenced samples. A – *AFF2* repetitive region composed by 15 [CCG] repeats. B – *AFF2* repetitive region composed by 47 [CCG] repeats.

In the case of *FMR1* gene we observed that alleles with thirty [CGG] triplets were the most frequently found. The same situation occurred with *AFF2* alleles, being fourteen [CCG] triplets the most frequent content. To explain the triplet content variation of *FMR1* and *AFF2* at fragile *locus* A and E, respectively, a multistep mutation model has emerged. In 1992, Morton and Macpherson have suggested a model with four allelic states: N→S→P→F. The initial mutation occurs on a normal-sized allele (N) leading to susceptibility to instability. Next, the vulnerable allele (S) expands or contracts one or two repeats, being in an unstable state that could persist for as many as 90 generations (Chakravarti, 1992). If the allele repeatedly increases triplet content, it can reach the premutation state (P) and now is prone to expand to a full mutation range (F).

AFF2 alleles tend to be randomly dispersed, essentially over the fourteen [CCG] region (49,5%) (Figure 12). Indeed, the variation of these alleles seems to follow a model

of small triplet increase (one or two units at each time), which may explain the low progress to large expansions associated with this gene (Figure 12).

In the case of *FMR1* alleles there seems to be a bimodal distribution as one can find a second "group" of frequently observed alleles in the twenty (±2) region (Figure 9). It is also worth notice that 6,6% of the alleles studied were in the intermediate range (between 40 to 54 [CGG] triplets). According to the model described previously, this group of alleles represent the two possibilities of response to susceptibility to instability. The 20 [CGG] alleles represent a decrease in the triplet content due to contraction events, on the opposite the intermediate alleles correspond to a repeatedly increase in triplet content. In the latter case there is an increased risk of reaching the premutation state. Interestingly, the *FMR1* repeat is not [CGG] pure and there are several reports that associate the pattern of variation/"stability" with the presence of [AGG] interruptions. The number and the position of the [AGG] interspersion may be an important component of the *FMR1* repeat structure that might stabilize the triplet region during replication and slippage events (Sullivan et at., 2002). The content of 3' pure [CGG] may also account to instability when associated to specific haplotypes (Eichler et al., 1996).

We have tested if *FMR1* repeat content show any type of association with *AFF2* repeat number. The increase of triplets content in *FMR1* gene does not appear to have any association with variations in *AFF2* triplets content. The increase in *FMR1* [CGG] and *AFF2* [CCG] repeats is independent of each other (Pearson's R: ρ >0,05). Additionally, the T-student test proved that the increase in both genes is unrelated (ρ >0,05) and there is not a tendency for one gene to present the same behavior as the other (Cronbach's alpha value: <0,05), meaning that all kinds of contractions/expansions can be found in both *FMR1* and *AFF2* genes.

When we analyze the relationship between the most common alleles of *FMR1* (thirty repeat-sized) gene and *AFF2* gene (fourteen repeat-size) we found that these two alleles are the most frequent combination (13,4%) (χ^2: ρ<0,05). However, the occurrence of one is independent of the other (Pearson's R: ρ>0,05), meaning that, although they were found together in a significant number of cases, all combinations are possible, as the combination is random. In practical terms we can state that the fragile *loci* A and E are transmitted from parent to child in a separate way and in linkage desiquilibrium, allowing normal recombination events.

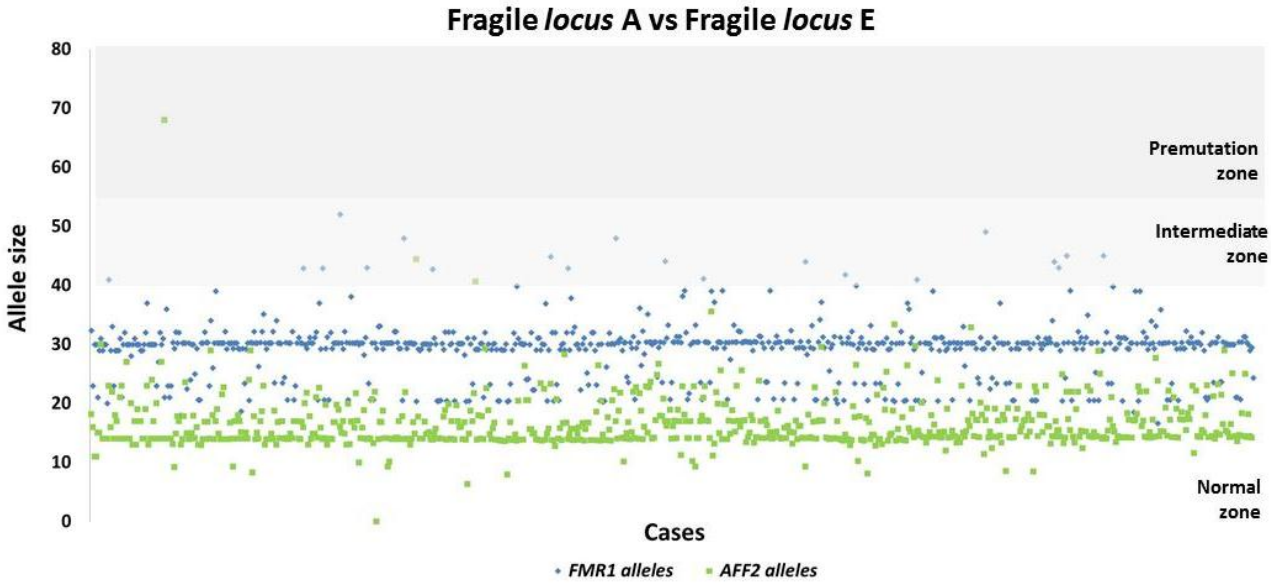


Figure 12. Triplet repeat number variation in Fragile *locus* A and E.

4.3 *FMR1* alleles characterization

The TP-PCR methodology implemented proved to be a suitable technique to study *FMR1* repetitive region. Similarly to sequencing analysis, this technique allowed an accurate evaluation of size and structure of *FMR1* alleles of both male and female samples (Figure 13). Additionally, TP-PCR represents a faster and economic technique for routine and high-throughput analysis.

Routine procedures, as Southern blot, allow size and methylation status determination of *FMR1* repetitive region. However, several recent publications refer that other features such as the presence or absence of [AGG] interspersions might influence the [CGG] repeat stability. We have decided to further characterize the *FMR1* alleles respecting to, not only, the number but also the pattern of interspersed [AGG] and the length of pure 3' [CGG] repeats. TP-PCR put in practice allows, in a single assay, the assessment of all those factors. Moreover, applied to the study of female samples allows identification of size homoallelism situations, in a cheaper and faster way than Southern blot. A careful analysis of *FMR1* alleles was performed and some examples of the diversity of patterns encountered are on Figure 14.

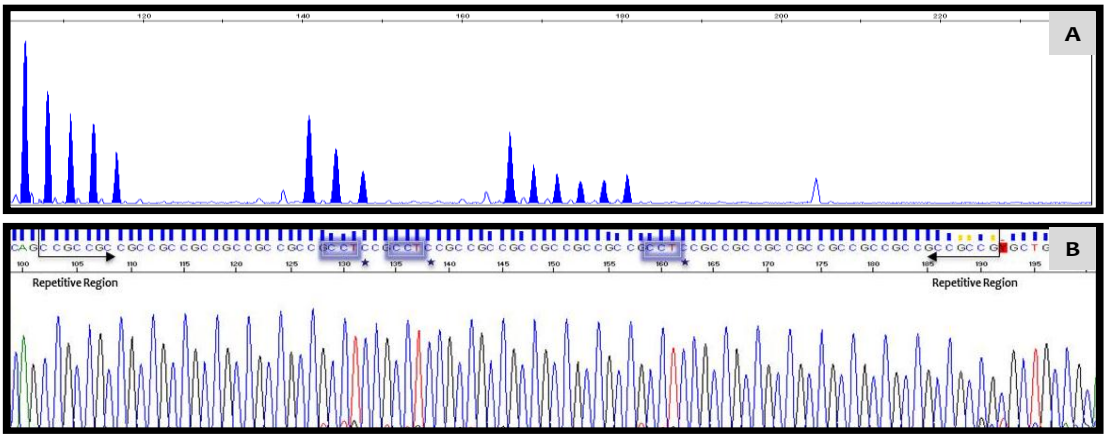


Figure 13. *FMR1* repetitive region assessment by TP-PCR. The analysis of *FMR1* alleles revealed the presence of [AGG] triplet interspersions across the repetitive region. This example illustrates a rare interspersion pattern: $[CGG]_{10}AGG[CGG]_7AGG[CGG]_1AGG[CGG]_9$ obtained by TP-PCR (A) and confirmed by sequencing analysis (B). (Note: both figures are in the reverse sense).

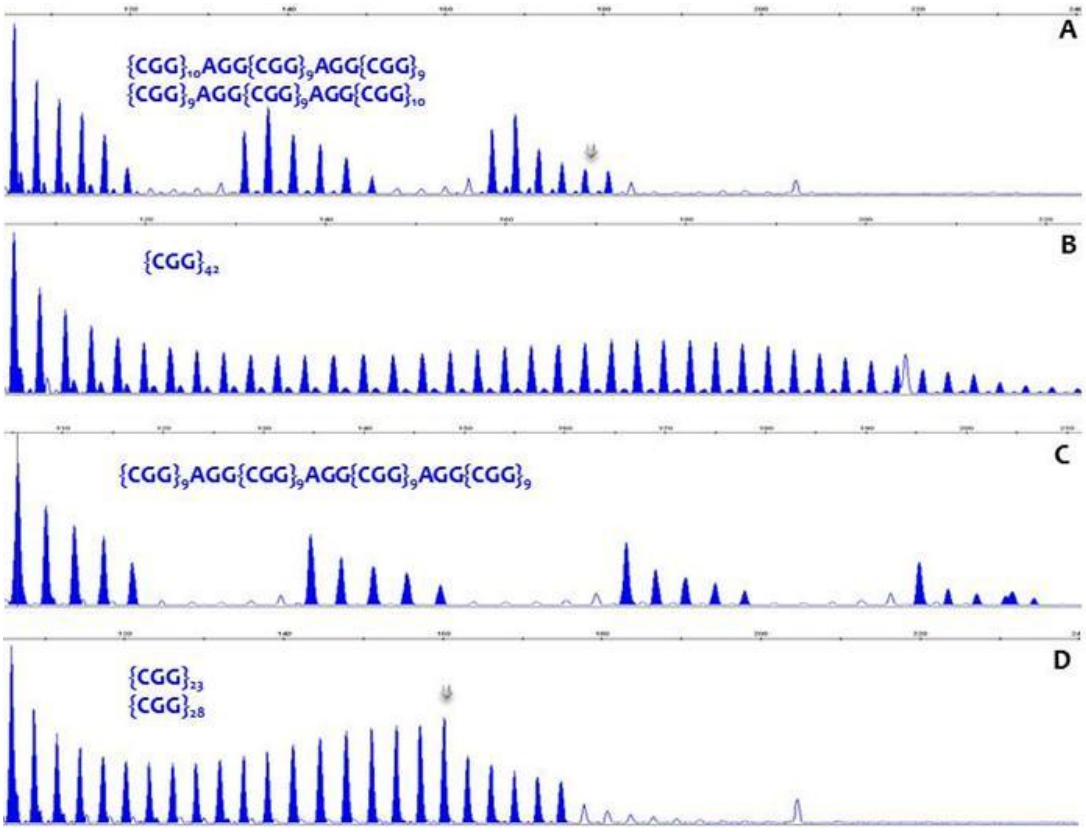


Figure 14. [AGG] interspersion patterns obtained by TP-PCR technique. Although the most frequent pattern of interspersion is in every ninth or tenth [CGG] repeat (A), several combinations were found. Bigger alleles tend to have less or no [AGG] interruptions (B) but some exceptional large alleles may preserve its [AGG] content (C). Less frequently, alleles may have pure uninterrupted stretches of [CGG] triplets (D).

Analysis of *FMR1* repetitive substructure has shown that this region is polymorphic respecting to size, presence of interspersions and their position (for more detail see Appendix 3 – Tables 16 and 17). A total of 138 alleles from a control population and 123 from a population presenting ID were studied for *FMR1* repetitive region substructure. All revealed less than 50 [CGG] repeats. The majority of those alleles had two [AGG] interruptions, being the pattern $[CGG]_{10}AGG[CGG]_9AGG[CGG]_9$ the most frequent in both control population (32,6%) and ID population (30,9%).

Among the control population thirty-five distinctive AGG interspersion patterns were identified, which represents a genetic diversity of 85,7%. Three substructures were identified as the most frequent: $[CGG]_{10}AGG[CGG]_9AGG[CGG]_9$ (32,6%) $[CGG]_9AGG[CGG]_9AGG[CGG]_9$ (12,3%) and $[CGG]_{10}AGG[CGG]_9$ (10,1%) (Figure 15).

The alleles belonging to ID population showed polymorphic internal substructures, in a total of thirty seven different patterns, presenting a genetic diversity of 87,3%. The same substructures seen in the control population were most common: $[CGG]_{10}AGG[CGG]_9AGG[CGG]_9$ (30,9%), $[CGG]_9AGG[CGG]_9AGG[CGG]_9$ (11,4%) and $[CGG]_{10}AGG[CGG]_9$ (8,9%) (Figure 17).

Three alleles with rare interspersion patterns were found, $[CGG]_9AGG[CGG]_9AGG[CGG]_2AGG[CGG]_6$ and $[CGG]_{10}AGG[CGG]_7AGG[CGG]_1AGG[CGG]_8$ (0,7% each) in the control population and $[CGG]_{10}AGG[CGG]_7AGG[CGG]_1AGG[CGG]_9$ found in the ID group (0,8%).

Previous studies have suggested that loss of [AGG] interruptions may be mediated by one of several mutational events such unequal sister chromatid exchange, [AGG] deletion, or A>G transversion (Eichler *et al.,* 1994; Kunst *et al.,* 1994). As the loss of [AGG] interruption, by non-reciprocal recombination events would be expected to change the total number of [CGG] repeats (Dover *et al*., 1982), it is unlikely that this represents the predominant or ”unique” mechanism. If changes in [AGG] content occur due to deletion or A→C and C→A transversions, the occurrence of rare interspersion patterns, as those observed in this study, may be a more plausible possibility.

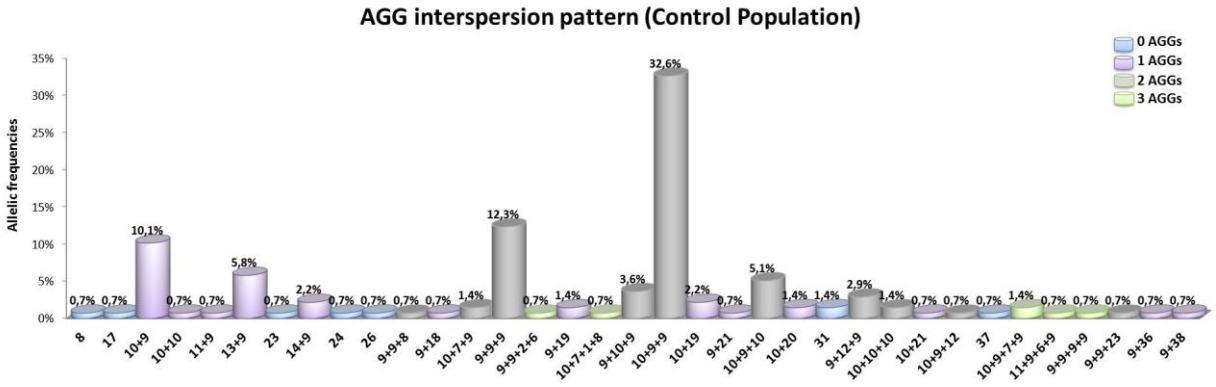


Figure 15. [AGG] interspersion patterns and corresponding frequencies observed in control population (138 alleles analyzed).[3]

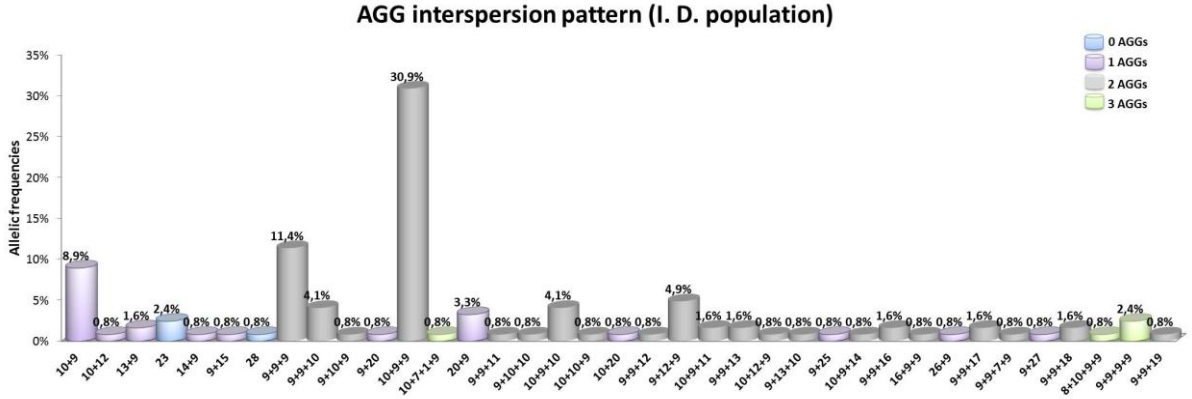


Figure 16. [AGG] interspersion patterns and corresponding frequencies found in ID population (123 alleles analyzed).

The comparison of the number of [AGG]s in a control and an ID population, with normal and intermediate-sized alleles, revealed that the majority of alleles presented two [AGG] interruptions. Although the majority of intermediate-sized alleles present two [AGG] interspersions, it is important to mention that the stretch of 3' [CGG] is increasing proportionally to the size of these alleles. This fact may compromise allele stability, as the length of pure 3' [CGG] repeats represent a structural factor known to influence *FMR1* alleles stability (Poon *et al.*, 2006). It has been suggested that *FMR1* alleles with more than twenty-four pure 3' [CGG] repeats are prone to instability, as the repeat slightly increases in length in each generation (Kunst *et al*., 1994). Moreover, alleles with more than thirty-four pure 3' [CGG] are unstable when transmitted (Eichler *et al*., 1994).

[3] Classification of CGG repeats is based on the position of AGG interruptions within the repeat. A ' + ' sign represents an AGG interruption and the numbers within the configuration refer to the length of uninterrupted CGG repeats.

The presence of four [AGG]s was found only in 1.3% of the ID intermediate-sized alleles group. This variant may represent a more stable structure, where the [AGG] interruptions are added along with the increase of [CGG] triplets in the 3' end or where the events leading to interruption loss have not yet occurred (Table 11).

A study of intermediate-sized alleles demonstrated that loss of [AGG] interspersion is not critical for expansion of normal alleles to an intermediate range (Dombrowski *et al*., 2002). Instead, they inferred that the loss of [AGG] interspersions is a late event that leads to greatly increased instability and may be related to haplotype background.

In this study, the analysis of *FMR1* repetitive region substructure demonstrated that all populations presented alleles lacking [AGG] interruptions and, interestingly, the higher frequency of pure [CGG] was found in the control population (5,8%). The total size of some of those alleles (21,1%) was between twenty-four and forty [CGG] triplets. According to Kunst *et al.* (1994), regardless of the size and classification, these alleles are prone to be unstable. Crawford *et al*. (2000a), reported an individual with twenty-eight pure repeats presenting a higher rate of sperm variants than did an individual with a $[CGG]_9AGG[CGG]_{23}$ repeat structure. Nolin *et al*. (1999) reported an individual with a sixty-eight pure [CGG] repeat allele that exhibited a higher rate of sperm variants than did individuals with $[CGG]_9AGG[CGG]_{65}$, $[CGG]_9AGG[CGG]_{68}$ or $[CGG]_9AGG[CGG]_9AGG[CGG]_{80}$ repeat substructures. Altogether, this data highlight the importance of assessing *FMR1* allele's internal substructure and suggest that, besides the size classification, is important to characterize them in terms of [AGG] number and position.

Table 11. Number of [AGG]s according to the allele size in control and ID populations

Population	[AGG]:1	[AGG]:2	[AGG]:3	[AGG]:4	Uninterrupted alleles	Total of alleles
Control normal sized alleles	**28,3%**	**61,6%**	4,3%	0,0%	**5,8%**	138
ID normal sized alleles	20,3%	**71,5%**	4,9%	0,0%	3,3%	123
ID intermediate-sized alleles	27,6%	**59,2%**	7,2%	**1,3%**	4,6%	152
Number of alleles	106	263	23	2	19	413

Single Nucleotide Polymorphisms (SNPs) possess a low mutation rate (10^{-9} for generation), meaning that are more resistant to recurrent mutations than do *Short Tandem Repeats* (STRs) (Satta *et al,* 1993). Information given by the stable SNPs provides evidence on ancestral lineages and their original background. A study of control and FXS Portuguese populations, resorting to SNPs analysis by High Resolution Melting

Curve Analysis (hrMCA), established that the TTT lineage was present in 58,0% of the control population alleles and the CTC lineage was the most frequently found in the FXS population with 58,0%, (Loureiro, 2010) (for more detail see Appendix 4 – Tables 19 and 20 and Figure 18). As the CTC lineage seems to be over-represented in FXS we decided to test if the presence and number of [AGG] triplets as well as the position of the first interruption is in linkage disequilibrium in the control group.

Some authors claim that the position of the first [AGG] interruption at 5' end may be an important factor to assess stability. Through association studies in a Caucasian population, the alleles presenting the first [AGG] after nine [CGG] triplets ($[CGG]_9n$) has been proposed to be more prone to expansion than those presenting the first [AGG] after ten [CGG] triplets ($[CGG]_{10}n$) (Gunter *et al*., 1998). However, no evidence of this positional effect was found in a large population of African-Americans (Crawford *et al*., 2000b).

In a control population, the lineages described previously have different frequencies according to the position of the first [AGG] interruption in the twenty-nine and thirty repeat alleles. 64,3% of alleles $[CGG]_9n$ showed the lineage CTC in contrast with 88,5% of alleles $[CGG]_{10}n$ have the TTT lineage (Table 12). This study suggest that the alleles $[CGG]_9n$ and $[CGG]_{10}n$ have different origins (χ^2: ρ <0,05; Pearson's R: ρ <0,05) and that the first are prone to expand towards pathogenic limits. It is important refer that the CTC lineage was simultaneously correlated with an FXS population and with the patterns $[CGG]_9n$, so together they may account to an increased risk of expansion. Another important fact is that, of all intermediate-sized alleles assessed, the majority of alleles (80,9%) presented the $[CGG]_9n$ pattern indicating that the position of the first interruption is a factor of vulnerability and that the $[CGG]_9n$ alleles are more prone to expand toward those limits.

Table 12. SNPs lineages and position of the first [AGG] interruption in the control population.

Lineage	1st [AGG]: 9	1st [AGG]: 10	1st [AGG]: >10	Uninterrupted alleles	Total of alleles
TTT	2,6%	**88,5%**	3,9%	5,1%	78
CTC	**64,3%**	10,7%	17,9%	7,1%	28
CCC	**62,5%**	12,5%	25,0%	0,0%	8
CTT	50,0%	25,5%	25,0%	0,0%	4
TTC	42,9%	57,1%	0,0%	0,0%	7
TCC	0,0%	100,0%	0,0%	0,0%	1
Total Number of alleles	30	79	11	6	126

Alleles with twenty-nine and thirty [CGG] repeats were the most frequently encountered in this study. Both demonstrated to have different [AGG] interspersion patterns and to be associated to different SNPs lineages in a control population of normal size alleles. A detailed analysis of their substructures revealed that twenty-nine repeat-size alleles, namely those with $[CGG]_9AGG[CGG]_9AGG[CGG]_9$ and $[CGG]_9AGG[CGG]_{19}$ internal substructures, have the CTC lineage (86,7 % and 100,0%, respectively). A rare allele with an unusual interspersion pattern ($[CGG]_{10}AGG[CGG]_7AGG[CGG]_1[CGG]_9$) was the only presenting the $[CGG]_{10}$n pattern and the TTT lineage (Table 13). On the contrary, thirty repeat sized-alleles presenting $[CGG]_{10}AGG[CGG]_9AGG[CGG]_9$ and $[CGG]_{10}AGG[CGG]_{19}$ substructures had the TTT lineage (88,9% and 100,0% respectively). Only two (3,7%) of the thirty repeat-sized alleles presenting the first interruption after nine [CGG] triplets were related to the TTT lineage. Association studies revealed a strong correlation among the twenty-nine repeat-sized alleles and the CTC lineage and between the thirty repeat-sized alleles and the TTT lineage (χ^2: ρ <0,05) (Table 13) indicating linkage disequilibrium.

This may suggest the existence of at least two different origins of the two more frequent allelic sizes (twenty-nine and thirty repeats) and suggests that one does not originates the other (by either expansions or contractions of one [CGG] repeat). Likewise, the existence of different genetic backgrounds may explain the existence of different susceptibility to instability.

Table 13. SNPs lineages in the most frequent *FMR1* allele sizes.

[AGG] pattern	TTT	CTC	CCC	CTT	TTC	TCC	Number of alleles	Allele size
9+9+9	0,0%	**86,7%**	0,0%	13,3%	0,0%	0,0%	15	**29**
9+19	0,0%	100,0%	0,0%	0,0%	0,0%	0,0%	2	
10+7+1+9	100,0%	0,0%	0,0%	0,0%	0,0%	0,0%	1	
9+9+2+7	100,0%	0,0%	0,0%	0,0%	0,0%	0,0%	1	**30**
9+9+10	100,0%	0,0%	0,0%	0,0%	0,0%	0,0%	1	
9+10+9	0,0%	**50,0%**	25,0%	0,0%	25,0%	0,0%	4	
10+9+9	**88,9%**	2,2%	0,0%	2,2%	4,4%	2,2%	45	
10+19	100,0%	0,0%	0,0%	0,0%	0,0%	0,0%	3	

4.4 *ARX* gene and intellectual disability

Mutations in *ARX* have been reported to give rise to a broad spectrum of phenotypes and suggested to be more frequent in XLID families than mutation in other known XLID genes, apart from *FMR1* expansions. We proposed to investigate whether mutations in polyalanine tracts (polyA) I and II of the *ARX* ex2p are a common cause of mental retardation in Portuguese patients with ID of unspecified cause, tested negative to FXS. The screening of 2728 gDNA samples resulted in detection of insertions and deletions in both polyA I and II of the *ARX* ex2p (Table 14).

Every pool samples presenting no size variations cannot be considered normal for *ARX* gene mutations, since only a portion of the second exon is being assessed and missense mutations are not detected by this methodology. Additionally, the amplification rate of all samples that compose each pool cannot be measured and is not presumably the same.

Table 14. Size variations detected on *ARX* exon 2, namely in polyalanine tracts I and II

DNA variation[1]	Predicted protein	Alanine content variation	Pathogenicity	Number of cases
c. 304GCG[11]	p.A111insA1	Ala(16)>Ala(17) PolyA I	Polymorphism[2]	1♀
c. 304GCG[13]	p.A111insA3	Ala(16)>Ala(19) PolyA I	Polymorphism[2]	1♂
c. 304GCG[6]	p.A111delA4	Ala(16)>Ala(12) PolyA I	To be determined	1♂
c. 441_464del24	p.A147delA8	Ala(12)>Ala(4) PolyA II	Polymorphism[3]	3♂
c. 429_452dup24	**p.A151insA8**	**Ala(12)>Ala(20) PolyA II**	**Mutation[4]**	**2♂ and 1♀**

1 – Designated according to the Human Genome Variation Society guidelines; 2-Gronskov *et al*., 2004; 3- Conti *et al*, 2010; 4 - Shoubridge *et al*., 2010.

Once sequenced, all the variants revealed to influence the protein alanine content, as all the altered codons translate exclusively to alanines. Both c.304GCG[11] and c.304GCG[13] result in [GCG] triplets insertion in polyA I, reflecting an addition of one and three alanine residues, respectively (Figure 17 A and B). Those variations were detected in one female presenting learning difficulties and speech delay and one male with psychomotor, mental and speech delay. Other studies have reported small insertions in polyA I, namely the c.304GCG[11] (Gronskov *et al.,* 2004), c.304GCG[12] and c.304GCG[13] (Bienvennu *et al.,* 2002). The pathogenic role of these nucleotide changes is uncertain, although it was suggested they may represent polymorphic variants (Gronskov *et al.,* 2004).The functional importance of these small polyalanine insertions is yet to be determined. Given that expansion of polyAs in transcription factors, to over than eighteen residues causes aggregation in the cytoplasm (Albrecht *et al.,* 2004), it is

feasible that expansion of the first polyA in *ARX* gene (composed by sixteen alanines) by two or three alanines might have functional consequences. However, such consequence is likely to be minor and thus have variable phenotypic expression (Gécz *et al.,* 2006).

Contractions in both polyA I and II of *ARX* ex2p were also detected. In the polyA I, the c. 304GCG[6], was detected in one male presenting mental delay and is responsible for a contraction of four alanine residues, due to a loss of four [GCG] triplets, with a pathological effect yet to be determined (Figure 17 C). The c.441_464del24 variant, found in three unrelated males: one presenting metal delay, one with a familial history of mental delay and dysmorphic features and the last with no clinical information available, results in a twenty-four bp deletion in the polyA II and, as consequence, a contraction of twelve to four alanine residues occurs (Figure 17 D). The *in frame* deletions in *ARX*ex2p that lead to contraction on the second polyA need to be interpreted cautiously, since some of these deletions are likely to represent rare polymorphisms rather than disease causing mutations. Contractions of variable length have been reported in the second polyA tract of *ARX,* of which contractions of twelve to either eight or four alanines have been considered to be rare, benign polymorphisms. Even though, those mutations have been reported in patients with intellectual disability, based on the current information they also have been identified on healthy unaffected males (Conti *et al.*, 2010). The observation that the c.429_452del24 was present in a healthy male from one family collected by the European XLID consortium (L45-MRX family) supported its classification as a length polymorphism (Conti *et al.*, 2010). Considering that many proteins with polyAs interact with each other by modifying their transcriptional activity (Amiel *et al.,* 2004), further studies will be necessary to elucidate if *ARX* polyA contractions may have a functional or pathogenic role.

The c.429_452dup24 mutation is a recurrent mutation identified in *ARX* gene, characterized by a twenty-four bp duplication in polyA II leading to an expansion of twelve to twenty alanine residues. This mutation was found in two male brothers presenting severe psychomotor delay and one female with learning and speech delay (Figure 17 E), being the only mutation with known pathogenic effect identified in our population. The c.429_452dup24 mutation has never been found in association with severe brain malformations. However, variable phenotypic expression is often observed within the same family with (Gronskov *et al.,* 2004), reinforcing the notion that *ARX* is a pleiotropic gene that, in a diverse genetic context controls different aspects of human brain morphogenesis and function. An unusual range of variation in clinical presentation associated with this mutation, both within and between families, has been reported (Turner et al., 2002). The likely genetic and environmental determinants modifying the

phenotypic outcome of the c.429_452dup24 are yet to be identified (Shoubridge *et al.,* 2010). The function of extended polyAs is unknown, but when they occur in transcription factor genes may result in either loss of function or gain of an abnormal function, suppressing the transcription (Nasrallah *et al*., 2004). When expanded, polyalanine tracts primarily promote protein misfolding that may modulate the normal protein activity and/or alter its cellular localization, which may affect the transcriptional regulation of cellular components essential for cell survival or differentiation (Messaed *et al.,* 2009). A study by Nasrallah *et al*., (2004) demonstrated that when expanded, the ARX protein remains in the nucleus and induces protein aggregation and cellular toxicity *in vitro.*

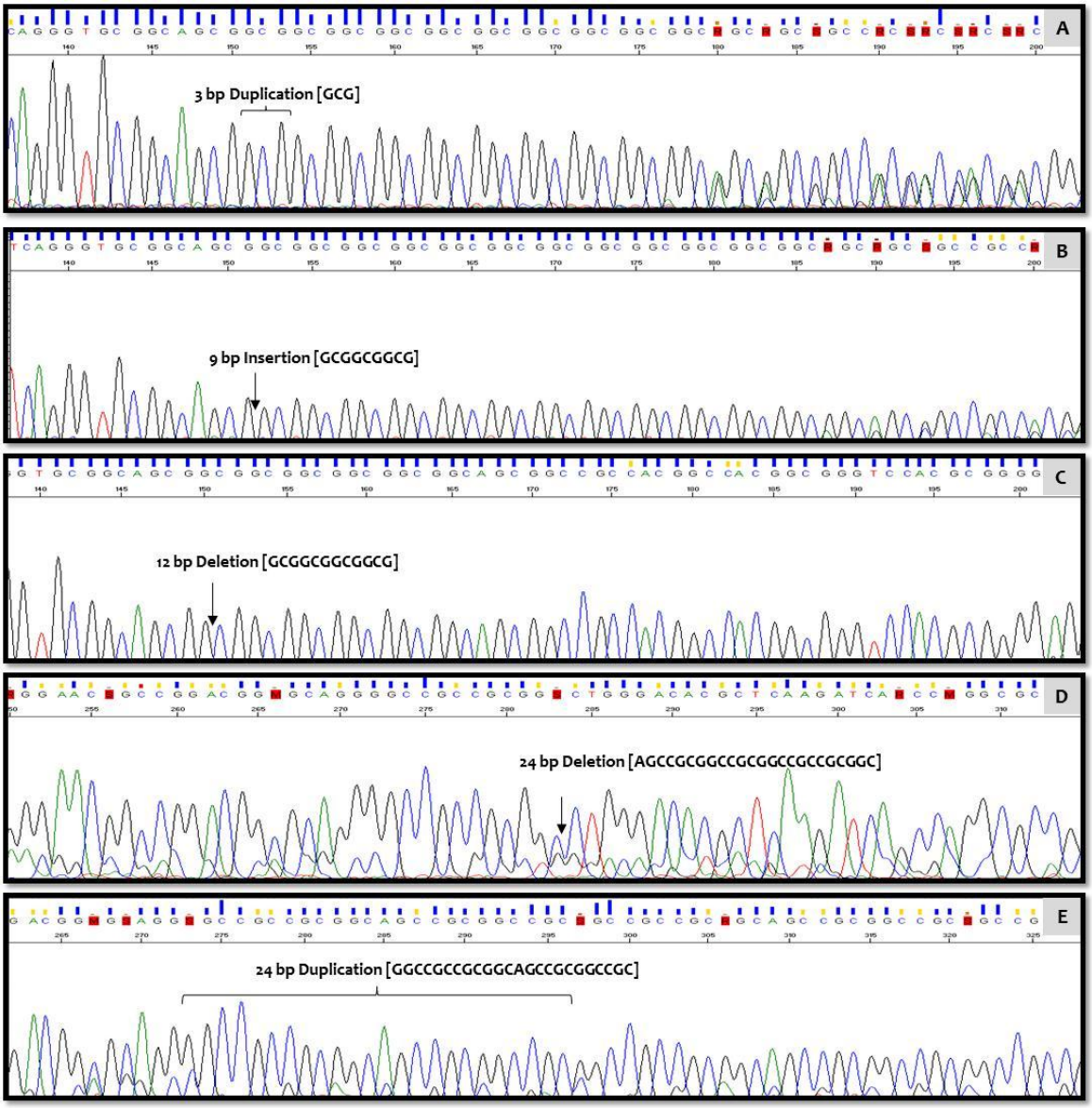


Figure 17. Variations detected in a portion of the second exon of the *ARX* gene. Several size variations were detected such as duplications (A and E), insertions (B) and deletions (C and D).

ARX gene represents a hot spot for mutations in families with cognition disorders as its mutations account for 9.5% of XLID families (Gécz *et al.*, 2006). The c.429_452dup24 might account for 6.6% of all XLID and 41% of families with mutations in *ARX* gene (Nawara *et al.,* 2006).

5. FINAL REMARKS

X-linked intellectual disability (XLID) is an extremely heterogeneous group of disorders and many of the causative genes are still unknown. As a consequence, there are no major mutational hot spots for XLID, with the exception of [CGG] repeat expansions in *FMR1* gene, responsible for FXS.

Pathogenic expansion mutations have been reported in two X-chromosome genes, *FMR1* and *AFF2*. However, as allele size and triplet content variations occur in an independent way in these two genes and the mechanisms responsible for both dynamic mutations are probably different.

This retrospective study of individuals with intellectual disability of unspecified cause allowed identification of undiagnosed individuals with an *AFF2* premutation and *ARX* mutations (c.429_452dup24). In the genetic counseling point of view, it is extremely important to assess these individuals and their families in order to prevent future intergenerational transmissions. Particularly in the case of *AFF2* premutation, the possible relationship of these alleles with development of Parkinson disease in adulthood and the risk of expansion in future generations further demonstrate the importance of this retrospective study.

Molecular assessment of individuals presenting clinical signs should be as more complete as possible, especially in cases of *FMR1* gene related disorders. The current knowledge of how normal and intermediate-size alleles behave is limited, so a more accurate evaluation of individuals at risk for XLID should include analysis of [AGG] number, pattern, and position of the first interruption. This study provides evidences of how these variables, put together, may help to improve risk assessment accuracy and provide a more solid genetic counseling. Also, genetic background may prove to be helpful in search of factors involved in the mechanism of dynamic mutations. To better understand the mechanisms behind *FMR1* gene expansions, intergenerational studies should be carried to verify if alleles presenting the first [AGG] interruption after nine [CGG] repeats and associated with the CTC lineage, are more prone to instability and at risk to increase for an intermediate or premutation range. Also, information concerning to the genetic background of *FMR1* intermediate and premutation-sized alleles will help to clarify which alleles are more prone to increase it triplet content.

Our results show clear evidences that the majority of *FMR1* alleles of the control population and presenting twenty-nine or thirty [CGG] repeats, possess different genetic backgrounds. This observation is not compatible with the current accepted multistep mutation model of Morton and Macpherson (1992), where alleles in a relatively stable state expand or contract one or two repeats. Indeed, to fit in this model these alleles

should present a common lineage. A more extensive study of these alleles may help to redefine their behavior.

The twenty-four base pair duplication (c.429_452dup24) was the only mutation identified in our population, when assessing the *ARX* ex2p. Other variants, including small insertions (c.304GCG[11]; c.304GCG[13]) and deletions (c.441_464del24), were also identified. Even though, no pathogenic effect has been established as consequence of this insertions and deletions, they were found in a very small percentage in our sample and, in the case of small insertions, they were even rarer than the c.429_452dup24 mutation. If in fact these variants represent polymorphic changes, they must be classified as very rare. Other important question still to be answered is that if alone or in conjunction with other genetic or environmental factors, they may represent an increased risk of susceptibility to develop intellectual disability related disorders. Pathogenicity studies are needed to evaluate the molecular and functional consequences of small insertions and the role of *ARX* polymorphisms in ID susceptibility.

6. REFERENCES

ABRAMS, M.T.; DOHENY, K.F.; MAZZOCCO, M.M. *et al.* Cognitive, behavioral, and neuroanatomical assessment of two unrelated male children expressing FRAXE, Am J Med Genet, 1997, 74: 73–81.

ALBRECHT, A.N.; KORNAK, U.; BODDRICH, A. *et al.* A molecular Pathogenesis for transcription factor associated polyalanine tract expansions, Hum Mol Genet, 2004, 13: 2351-2359.

AL-HINTI, J.T.; NAGAN, N. AND HARIK, S. I. Fragile X premutation in a woman with cognitive impairment, tremor, and history of premature ovarian failure, Alzheimer Dis Assoc Disord, 2007, 21(3): 262-264.

AMIEL, J.; TROCHET, D.; CLÉMENT-ZIZA, M. *et al.* Polyalanine expansions in human, Hum Mol Genet, 2004, 13: 235-243.

ANNESI, G.; NICOLETTI, G.; TARANTINO, P. *et al.* FRAXE intermediate alleles are associated with Parkinson's disease, Neurosci Lett, 2004, 368(1): 21-24.

BEJERANO, G.; PHEASANT, M.; MAKUNIN, I. *et al.* Ultraconserved Elements in the Human Genome, Science, 2004, 304: 1321-1324.

BENSAID, M.; MELKO, M.; BECHARA, E. G. *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, 2009, 37: 1269-1279.

BIENVENU, T.; POIRIER, K.; FRIOCOURT, G. *et al. ARX*, a novel Prd-class-homeobox gene highly expressed in the telencephalon, is mutated in X-linked mental retardation, Hum Mol Genet, 2002, 11: 981-991.

CHIU, H.; TSENG, Y.; HSIAO, H. *et al.* The AGG interruption pattern within the CGG repeat of the *FMR1* gene amoung Taiwanese population, J Genet, 2008, 87: 275-277.

CHIURAZZI, P.; SCHWARTZ, C.; GECZ, J. *et al.* XLMR genes: update 2007, Eur J Hum Genet, 2008, 16: 422–434.

CONTI, V.; MARINI, C.; MEI, D. *et al.* Contractions in the second PolyA tract of *ARX* are rare, non-pathogenic polymorphisms, Am J Med Genet, 2010, Part A 9999: 1-4.

CRAWFORD, D.C.; WILSON B. AND SHERMAN S.L. Factors involved in the initial mutation of the fragile X CGG repeat as determined by sperm small pool PCR, Hum Mol Genet, 2000a, 9: 2909–2918.

CRAWFORD, D.C.; ZHANG, F.; WILSON, B. *et al.* Fragile X CGG repeat structures among African-Americans: identification of a novel factor responsible for repeat instability, Hum Mol Genet, 2000b, 9: 1759–1769.

DELBRIDGE, M.; MCMILLAN, D.; DOHERTY, R. *et al.* Origin and evolution of candidate mental retardation genes on the human X chromosome (MRX), BMC Genomics, 2008, 9: 65-70.

D'HULST, C. AND KOOY, R. F. Fragile X Syndrome: from molecular genetics to therapy, J Med Genet, 2009, 46: 577-584.

DOVER, G. Molecular drive: a cohesive mode of species evolution, Nature,1982, 299: 111–17.

DOMBROWSKI C.; LEVESQUE S.; MOREL M.L. *et al.* Premutation and intermediate-size *FMR1* alleles in 10 572 males from the general population: loss of an AGG interruption is a late event in the generation of fragile X syndrome alleles, Hum Mol Genet, 2002, 11: 371–378.

EICHLER, E.; HOLDEN, J. J. A.; POPOVICH, B.W. *et al.* Length of uninterrupted [CGG] repeats determines instability in the *FMR1* gene, Nature Genet, 1994, 8: 88–94.

EICHLER, E.; MACPHERSON, J.N.; MURRAY, A. *et al.* Haplotype and interspersion analysis of the *FMR1* CGG repeat identifies two different mutational pathways for the origin of the fragile X syndrome, Hum Mol Genet, 1996, 5: 319–330.

EICHLER, E. AND NELSON, L. Genetic variation and evolutionary stability of the *FMR1* CGG repeat in six closed human populations, Am J Hum Genet, 1996, 64: 220- 225.

FARADZ, S.M.H.; PATTIIHA, M.Z.; LEIGH, D.A. *et al.* Genetic diversity at the *FMR1 locus* in the Indonesian population, Ann Hum Genet, 2000, 64: 329-339.

FRIOCOURT, G. AND PARNAVELAS J.G. Mutations in *ARX* results in several defects involving GABAergic neurons, Front Cell Neurosci, 2010, 4: Art. 4.

GARBER, K.; VISOOTSAK, J. AND WARREN, S. T. Fragile X syndrome, Eur J Hum Genet, 2008, 16: 666–672.

GÉCZ, J.; CLOOSTERMAN, D. AND PARTINGTON, M. *ARX*: a gene for all seasons, Curr Opin Genet Dev, 2006, 16: 308-816.

GESTINARI-DUARTE, R.; SANTOS-REBOUÇAS, C.; BOY, R. *et al. ARX* mutation c.428-451dup (24bp) in a Brazilian family with X-linked mental retardation, Eur J Med Genet, 2006, 49: 269-275.

GRONSKOV, K.; HJALGRIM, H.; NIELSEN, I-M. *et al.* Screening of the *ARX* gene in 682 retarded males, Eur J Hum Genet, 2004, 12: 701-705.

GUNTER, C.; PARADEE, W.; CRAWFORD, D.C. *et al.* Re-examination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in *FMR1*, Hum Mol Genet, 1998, 7: 1935-1946.

HUNTER, J.; EPSTEIN, M.; TINKER, W. S. *et al.* Fragile X-associated Primary Ovarian Insufficiency: Evidence for Additional Genetic Contributions to Severity, Genet Epidemiol, 2008, 32(6): 553–559.

JACQUEMONT, S.; HAGERMAN R.J.; LEEHEY M. *et al.* Fragile X Premutation Tremor/Ataxia Syndrome: Molecular, Clinical, and Neuroimaging Correlates, Am J Hum Genet, 2003, 72: 869–878.

JARA, L.; ASPILLAGA, M.; AVENDANO, I. *et al.* Distribution of (CGG)n and *FMR1* associated microssatelite alleles in a normal Chilean population, Am J Med Genet, 1998, 75(3): 277-82.

JIN. P.; ZARNESCU, D. C. AND ZHANG F. RNA-mediated neurodegeneration caused by the fragile X premutation r[CGG] repeats in Drosophila, Neuron, 2003; 39: 739–747.

KUNST, C.B AND WARREN, S.T. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles, Cell, 1994, 77: 853–861.

LAPERUTA, C.; SPIZZICHINO, L.; D'ADAMO, P. *et al.* MRX87 family with Aristaless X dup24bp mutation and implication for polyalanine expansions, BMC Med Genet, 2007, 8: 25-29.

LAVOIE, H.; DEBANE, F.; TRINH, D. *et al.* Polymorphism, shared functions and convergent evolution of genes with sequences coding for polyalanine domain, Hum Mol Genet, 2003,12: 2967-2979.

LUKUSA, T. AND FRYNS, J. P. Human chromosome fragility, Biochim et Biophys Acta, 2008, 1779: 3–16.

MESSAED, C. AND ROULEAU, G.A. Molecular mechanisms underlying polyalanine diseases, Neurobiol Dis, 2009, 34(3): 397-405.

MINGRONI-NETTO, R.; ANGELI, C.; AURICCHIO, T. *et al.* Distribution of CGG repeats and FRAXAC1/DXS548 alleles in South American populations, Am J Med Genet, 111: 243-252.

MORTON, N.E. AND MACPHERSON, J.N. Population genetics of the fragile-X syndrome: multiallelic model for the *FMR1 locus*, Proc Natl Acad Sci USA, 1992, 89: 4215–4217.

MURRAY, A.; ENNIS, S.; YOUINGS, S.A. *et al.* Stability and haplotype analysis of the FRAXE region, Eur J Hum Genet, 2000, 8: 583-589.

NASRALLAH, I.M.; MINARCIK, J.C. AND GOLDEN, A. A polyalanine tract expansion in *ARX* forms intranuclear inclusion and results in increased cell death, J Cell Biol, 2004, 167: 411-416.

NAWARA, M.; SZCZALUBA, K.; POIRIER, K. *et al.* The *ARX* mutations: a frequent cause of X-linked mental retardation, Am J Med Genet, 2006, 140: 727-732.

NOLIN, S.L.; HOUCK, G.E. JR; GARGANO, A.D. *et al.*. *FMR1* CGG-repeat instability in single sperm and lymphocytes of fragile-X premutation males, Am J Hum Genet, 1999, 65: 680–688.

OHIRA, R.; ZHANG, Y.; GUO, W. *et al.* Human *ARX* gene: Characterization and expression, Mol Genet Metab, 2002, 77: 179-188.

OOSTRA B. A. AND WILLEMSEN R. A fragile balance: *FMR1* expression levels, Hum Mol Genet, 2003, 12: 249-257 (Review Issue 12).

OOSTRA B. A. AND WILLEMSEN R. *FMR1*: a gene with three faces, Biochim Biophys Acta. 2009, 1790(6): 467–477.

PENAGARIKANO, O.; MULLE, J. AND WARREN, S. The Pathophysiology of Fragile X Syndrome, Annu. Rev. Genomics Hum. Genet, 2007, 8: 109-129.

PEPRAH, E.K.; ALLEN, E.G.; WILLIAMS, S.M. *et al.* Genetic diversity of the Fragile X Syndrome gene *(FMR1*) in a large Sub-Saharan West African population, Ann of Hum Genet, 2010, 74: 316-325.

PFEIFFER, B. AND HUBER, K. M. The state of synapses in Fragile X Syndrome, Neuroscientist, 2009, 15: 549-567.

POON, P.; CHEN, Q.; ZHONG, N. *et al.* [AGG] interspersion analysis of the *FMR1* [CGG] repeats in mental retardation of unspecific cause, Clin Biochem, 2006, 39: 244-248.

REISH, O.; FULLSTON, T.; REGEV, M. *et al.* A novel de novo 27 bp duplication of the *ARX* gene, resulting from postzygotic mosaicism and leading to three severely affected males in two generations, Am J Hum Genet, 2009, 149A: 1655-1660.

RITCHIE, R.J.; CHAKRABARTI, L.; KNIGHT, S.J.L. *et al.* Population genetics of the FRAXE and FRAXF GCC repeats, and a novel CGG repeat, in Xq28, Am J Med Genet, 1997, 73: 463-469.

ROBERTS, J.E.; BAILEY, D.B.; MANKOWSKI, J. *et al.* Mood and anxiety disorders in females with the *FMR1* premutation, Am J Med Genet B Neuropsychiatr Genet, 2009, 150B(1): 130-139.

ROHR, J.; ALLEN, E.G.; CHAREN, K. *et al.* Anti-Mullerian hormone indicates early ovarian decline in fragile X mental retardation (*FMR1*) premutation carriers: a preliminary study, Hum Reprod, 2008, 23: 1220–1225.

ROSALES-REYNOSO, M.; MENDONZA-CARRERA, F.; TROYO-SAROMAN, R. *et al.* Genetics diversity at the *FMR1 locus* in Meexican population, Arc of Med Res, 2005, 36: 412-417.

SANTOS, C. B.; LIMA, M. A. AND PIMENTEL M.M.G. A New PCR assay useful for screening of FRAXE/FMR2 mental impartment among males, Hum Mutat, 2001, 18: 157-162.

SATTA, Y.; OHUIGIN, C.; TAKAHATA, N. *et al.* The synonymous substitution rate of the major histocompatibility complex *loci* in primates. Proc Natl Acad Sci USA, 1993, 90: 7480–7484.

SCHWARTZ, M.; ZLOTORYNSKI, E. AND KEREM, B. The molecular basis of common and rare fragile sites, Cancer Lett. 2006, 232: 13–26 (Review).

SCHAEFFER, C.; BEAULANDE, M.; EHRESMANN, C.; *et al.* The RNA binding protein FMRP: new connections and missing links, Biol Cell, 2003, 95: 221-228.

SHOUBRIDGE, C.; CLOOSTERMAN, D.; PARKINSON-LAWERENCE, E. *et al.*. Molecular pathology of expanded polyalanine tract mutations in the Aristaless-related Homeobox gene, Genomics, 2007, 90: 59-71.

SHOUBRIDGE, C.; FULLSTON, T. AND GÉCZ, J. *ARX* Spectrum Disorders: Making Inroads into the Molecular Pathology, Hum Mutat, 2010, 31: 889-900.

SINDEN, R. R. Biological implications of the DNA structures associated with disease-causing triplet repeats, Am J Hum Genet, 1999, 64: 346–353, (Review).

SKUSE, D. X-linked genes and mental functioning, Hum Mol Genet, 2005, 14: 27-32 (Review issue 1).

SULLIVAN, A.K.; CRAWFORD, D.C.; SCOTT, E.H. *et al.* Paternally transmitted *FMR1* alleles are less stable than maternally transmitted alleles in the common and intermediate size range, Am J Hum Genet, 2002, 70: 1532-1544.

SURI, M. The phenotypic spectrum of *ARX* mutations, Dev Med Child Neurol, 2005, 47: 133-137.

VERKERK, J; PIERETTI, M.J.; SUTCLIFFE, S. *et al.* Identification of a gene (FMR-1) containing a [CGG] repeat coincident with a break point cluster region exhibiting length variation in fragile X syndrome, Cell, 1991, 65: 905–914.

TURNER, G; PARTINGTON, M.; KERR, B. *et al.* Variable expression of mental retardation, autism, seizures and dystonic and movements in two families with an identical *ARX* mutation, Am J Med Genet, 2002, 112: 405-411.

WEELS, R. Mutation Spectra in Fragile X Syndrome Induced by deletions of [CGG].[CCG] Repeats, J Biol Chem, 2009, 284: 7407-7411.

ZHONG, N.; LIU, X.; GOU, S. *et al.* Distribution of *FMR1* and associated microsatellite alleles in a normal Chinese population, Am J Med Genet, 1994, 51(4): 417-422.

ZHONG, N.; JU, W.; AND CURLEY, D. A survey of FRAXE allele sizes in three populations, Am J Med Genet, 1996, 64: 415-419.

ZHOU, Y.; TANG, K.; LAW, H.Y. *et al. FMR1* CGG repeat patterns and flanking haplotypes in three Asian populations and their relationship with repeat instability, Ann Hum Genet, 2006, 70: 784-796.

ELECTRONIC REFERENCES:

BIANCALANA, V.; STEINBACH, P. AND STENHOUSE S. Draft Best Practice Guidelines for Molecular Analysis in Fragile X Syndrome, [Online] (Accessed on January 17, 2011). Available on: www.emqn.org

JOHAN DEN DUNNEN AND ANTONARAKIS SE. Human Genome Variation Society: Guidelines for mutation nomenclature, [Online] (Accessed on December 28, 2010). Available on: www.hgvs.org.

WORLD HEALTH ORGANIZATION: REGIONAL OFFICE FOR EUROPE. Mental Health: intellectual disability, [Online] (Accessed on September 12, 2010). Available on: www.euro.who.

THESIS:

LOUREIRO, JOANA MARIA GERALDES DA ROCHA. Síndrome de X-Frágil em Portugal: caracterização molecular por SNPs e STRs, Lisbon: Lisbon University, 2010, Master Thesis.

7. APPENDIX

7.1 Appendix 1 – Mutations in *ARX* gene

Table 15. List of mutations in *ARX* leading to non-malformation and malformation phenotypes.

Mutation	Exon (Intron)	Domain affected	Mutation type	Phenotype	References
Exon1_2del	1, 2		Deletion	XLAG	Kitamura *et al*., 2002
c.196+2T>C	1	Oct	Splice	XLAG	Kato *et al*., 2004
c.81C>G	1	Oct	Nonsense	OS	Fullston *et al*., 2009
c.98T>C	1	Oct	Missense	NS-XLID	Bienvenu *et al*., 2002
c.112C>T	1	Oct	Missense	NS-XLID	Poirier *et al*., 2006
c.488A>G	2		Missense	NS-XLID	Bienvenu *et al*., 2002
c.856G>A	2		Missense	NS-XLID	Bienvenu *et al*., 2002
c.1058C>T	2	HD	Missense	XMESID	Stromme *et al*., 2002
Exon2_5del	2-5		Deletion	XLAG	Kato *et al*., 2004
c.232G>T	2		Nonsense	XLAG	Kato *et al*., 2004
c.335_368del34	2	PolyA I	Deletion	XLAG	Kato *et al*., 2004
c.392_452del61	2		Deletion	XLAG	Kato *et al*., 2004
c.420_451del32	2		Deletion	XLAG	Kitamura *et al*., 2002
c.617delG	2		Deletion	XLAG-HYD	Kato *et al*., 2004
c.619_647del28	2		Deletion	XLAG	Kato *et al*., 2004
c.790delC	2		Deletion	XLAG	Kitamura *et al*., 2002
c.980_983del4	2		Deletion	XLAG	Miyata *et al*., 2009
c.994C>T	2	HD	Missense	XLAG	Uyanik *et al*., 2003
c.995G>A	2	HD	Missense	XLAG	Kitamura *et al*., 2002
c.995G>C	2	HD	Missense	XLAG	Kato *et al*., 2004
c.996_1004del9	2	HD	Deletion	XLAG	Bhat *et al*., 2005
c.998C>A	2	HD	Missense	ACC/AG	Kato *et al*., 2004
c-1013A>CinsC	2	HD	Insertion	XLAG	Okazaki *et al*., 2008
c.1028T>A	2	HD	Missense	XLAG	Kitamura *et al*., 2002
c.1058C>G	2	HD	Missense	XLAG	Kato *et al*., 2004
c.1105G>T	3	HD	Nonsense	HYD/AG	Kato *et al*., 2004
c.1117C>T	3	HD	Nonsense	HYD/AG	Kitamura *et al*., 2002
c.1119+1G>C	(3)	HD	Splice	XLAG	Kato *et al*., 2004

c.1135C>A	4	HD	Missense	ISSX (WS)	Shoubridge *et al*., 2010
c.1449-816_*460del	(4)	Arist	Deletion	ISSX (WS)	Stromme *et al*., 2002
c.1136G>T	4	HD	Missense	XLAG	Shoubridge *et al*., 2010
c.1187_1188insC	4	HD	Insertion	XLAG	Kitamura *et al*., 2002
c.1372delG	4		Deletion	XLAG	Kitamura *et al*., 2002
c.1419_1420insAC	4		Insertion	XLAG/ACC	Hartmann *et al*., 2004
c.1561G>A	5	Arist	Missense	XLAG/LCH	Kato *et al*., 2004
c.1465delG	5	Arist	Deletion	ISSX (WS)	Wallerstein *et al*., 2008

Note: Mutations responsible for non-malformation phenotypes are highlighted.

Oct – Octapeptide domain; PolyA I – Polyalanine tract I; HD – Homeodomain; Arist – Aristaless domain; ACC/AG – Agenesis of the Corpus Callosum with Abnormal Genitalia (Proud Syndrome); HYS/AG – Hydrancephaly and Abnormal Genitalia; ISSX – Infantile Spasms X-linked (West Syndrome); LCH – Lissencephaly with Cerebral Hypoplasia; NS-XLID – Nonsyndromic X-linked Intellectual Disability; OS – Ohtahara Syndrome; XLAG – X-linked Lissencephaly with Ambiguous Genitalia; XMESID – X-linked Myoclonic Epilepsy with Spasticity and Intellectual Disability**.**

(Adapted from: SHOUBRIDGE *et al*. *ARX* Spectrum Disorders: Making Inroads into the Molecular Pathology. Human Mutation, 2010, 31: 889-900.)

7.2 Appendix 2 – Gene sequencing

7.2.1 Reagent preparation

- Tris - Acetic Acid - EDTA (TAE): 50X TAE (Tris 2M, acetato 1M EDTA 50 mM):
- Weight 242,2 g of Trizma base and dissolve in Distilled water. Add 57,2 mL of 100% glacial acetic acid and 100,0 mL of EDTA-NaOH 0,5 M, pH 8,0, mix and add distilled water until a 1000 mL final volume. Store this solution at 4ºC.
- Loading buffer (0,25% bromophenol blue, 0,25% xilenocianol; 15% (m/v) Ficoll 400):
  Dissolve 1,5 g of Ficoll 400 in 9,5 mL of Distilled water, add 250 µL of 10% (m/v) bromophenol blue and 250 µL of 10% xilenocianol. Mix by inversion.
- Hypper Ladder I:
  Dilute 2µL of loading buffer to 9 µL of Distilled water and add 1µL of Hypper ladder I (Bioline).

7.2.2 Agarose gel electrophoresis

1. Prepare a 0,8% agarose gel in Tris Acetate EDTA 1X (TAE) buffer. To 100 mL add 7,5 µL of Gel Red (Biotum) and mix by gentle agitation.
2. Mix 2µL of the symmetric PCR product with 2µL of Loading Buffer.
3. Place the mixture on the gel wells, cover with TAE buffer 1X and apply an electric field of 140V until separate PCR products. Place a molecular weight (Hypper Ladder I (Bioline)) on the first and last well.
4. Expose the gel to ultraviolet light to verify the amplification and quality of gDNA fragments.

7.2.3 Enzymatic PCR products purification

1. Mix 5 µL of the symmetric PCR product with 2 µL of ExoSAP-IT for a combined 7 µL reaction volume.
2. Incubate at 37^{o}C for 30 minutes to degrade remaining primers and nucleotides.
3. Incubate at 80oC for 15 minutes to inactivate the ExoSAP-IT.

7.2.4 Sequencing products purification: DyeEx 96 Kit

1. Remove the lower and upper film of the DyeEx 96 (EdgeBio) plate.
2. Place the DyeEx 96 (EdgeBio) plate above a collection plate and centrifuge at 1000x g for 3 minutes.
3. Reject the preservation fluid, transferred from the DyeEx plate to the collection plate by inversion.
4. Carefully place the DyeEx 96 plate above a new 96 well plate, where the sequencing products will be eluted.
5. Slowly apply 20 µL of the sequencing products on the gel columns of the DyeEx 96 plate.
6. Centrifuge the DyeEx 96 plate and the 96 well plate at 1000 g for 3 minutes.
7. Discard the DyeEx 96 plate and dry the eluted products placing the 96 well plate on a thermocycler at 70^{o}C for 30 to 40 minutes.
8. Resuspend the samples in 25 µL formamide and run on the *3130xl* Genetic Analyzer (Applied Biosystems).

7.3 Appendix 3 – [AGG] interspersion patterns

Table 16. [AGG] interspersion patterns and corresponding frequencies in control population.

Total repeat length	Substructure of repeat	Number of alleles	Allelic frequency (%)	Number of [AGG]s	3' pure [CGG] number
8	8	1	0,72	0	8
17	17	1	0,72	0	17
20	10+9	14	10,14	1	9
21	10+10	1	0,72	1	10
21	11+9	1	0,72	1	9
23	13+9	8	5,80	1	9
23	23	1	0,72	0	23
24	14+9	3	2,17	1	9
24	24	1	0,72	0	24
26	26	1	0,72	0	26
28	9+9+8	1	0,72	2	8
28	9+18	1	0,72	1	18
28	10+7+9	2	1,45	2	9
29	9+9+9	17	12,32	2	9
29	9+9+2+6	1	0,72	3	6
29	9+19	2	1,45	1	19
29	10+7+1+9	1	0,72	3	8
30	9+10+9	5	3,62	2	9
30	10+9+9	45	32,61	2	9
30	10+19	3	2,17	1	19
31	9+21	1	0,72	1	21
31	10+9+10	7	5,07	2	10
31	10+20	2	1,45	1	20
31	31	2	1,45	0	31
32	9+12+9	4	2,90	2	9
32	10+10+10	2	1,45	2	10
32	10+21	1	0,72	1	21
33	10+9+12	1	0,72	2	12
37	37	1	0,72	0	37
38	10+9+7+9	2	1,45	3	9
38	11+9+6+9	1	0,72	3	9
39	9+9+9+9	1	0,72	3	9
43	9+9+23	1	0,72	2	23
45	9+36	1	0,72	1	36
48	9+38	1	0,72	1	38

Table 17. [AGG] interspersion pattern and corresponding frequencies in intellectual disabled population.

Total repeat length	Substructure of repeat	Number of alleles	Allelic frequency (%)	Number of [AGG]s	3' pure [CGG] number
20	10+9	11	8,94	1	9
23	10+12	1	0,81	1	12
23	13+9	2	1,63	1	9
23	23	3	2,44	0	23
24	14+9	1	0,81	1	9
25	9+15	1	0,81	1	15
28	28	1	0,81	0	28
29	9+9+9	14	11,38	2	9
30	9+9+10	5	4,07	2	10
30	9+10+9	1	0,81	2	9
30	9+20	1	0,81	1	20
30	10+9+9	38	30,89	2	9
30	10+7+1+9	1	0,81	3	9
30	20+9	4	3,25	1	9
31	9+9+11	1	0,81	2	11
31	9+10+10	1	0,81	2	10
31	10+9+10	5	4,07	2	10
31	10+10+9	1	0,81	2	9
31	10+20	1	0,81	1	20
32	9+9+12	1	0,81	2	12
32	9+12+9	6	4,88	2	9
32	10+9+11	2	1,63	2	11
33	9+9+13	2	1,63	2	13
33	10+12+9	1	0,81	2	9
34	9+13+10	1	0,81	2	10
35	9+25	1	0,81	1	25
35	10+9+14	1	0,81	2	14
36	9+9+16	2	1,63	2	16
36	16+9+9	1	0,81	2	9
36	26+9	1	0,81	1	9
37	9+9+17	2	1,63	2	17
37	9+9+7+9	1	0,81	3	9
37	9+27	1	0,81	1	27
38	9+9+18	2	1,63	2	18
39	8+10+9+9	1	0,81	3	9
39	9+9+9+9	3	2,44	3	9
39	9+9+19	1	0,81	2	19

Table 18. [AGG] interspersion patterns of intermediate-sized alleles of an intellectual disabled population.

Total repeat length	Substructure of repeat	Number of alleles	Allelic frequency (%)	Number of [AGG]s	3' pure [CGG] number
40	9+9+9+10	2	1,3	3	10
40	9+9+20	3	2,0	2	20
40	10+9+19	1	0,7	2	19
41	9+11+9+9	1	0,7	3	9
41	9+9+21	6	4,0	2	21
41	9+31	3	2,0	1	31
41	10+9+9+10	2	1,3	3	10
41	10+30	1	0,7	1	30
41	11+29	1	0,7	1	29
41	41	1	0,7	0	41
42	9+9+9+12	1	0,7	1	12
42	9+9+22	12	7,9	2	22
42	9+12+19	1	0,7	2	19
42	9+22+9	1	0,7	2	9
42	9+32	2	1,3	1	32
42	10+9+21	2	1,3	2	21
42	42	1	0,7	0	42
43	9+10+12+9	4	2,6	3	9
43	9+13+9+9	1	0,7	3	9
43	9+9+23	10	6,6	2	23
43	9+33	6	4,0	1	33
43	43	2	1,3	0	43
44	9+9+24	6	4,0	2	24
44	9+24+9	1	0,7	2	9
44	9+34	5	3,3	1	34
44	10+9+23	1	0,7	2	23
44	10+33	1	0,7	1	33
45	9+9+7+7+9	1	0,7	4	9
45	9+9+25	5	3,3	2	25
45	9+15+7+1+9	1	0,7	4	9
45	9+35	5	3,3	1	35
45	10+9+24	1	0,7	2	24
45	10+34	1	0,7	1	34
46	9+9+26	6	4,0	2	26
46	9+36	1	0,7	1	36
46	10+9+25	1	0,7	2	25
46	10+35	2	1,3	1	35
46	18+27	1	0,7	1	27
46	46	1	0,7	0	46

47	9+8+9+18	1	0,7	3	18
47	9+9+27	7	4,6	2	27
47	9+37	5	3,3	1	37
47	10+9+26	1	0,7	2	26
47	10+36	1	0,7	1	36
48	9+9+28	3	2,0	2	28
48	9+38	1	0,7	1	38
49	9+9+29	3	2,0	2	29
49	10+9+28	2	1,3	2	28
50	9+9+30	4	2,6	2	30
50	9+40	1	0,7	1	40
50	10+9+29	1	0,7	2	29
51	9+9+31	2	1,3	2	31
51	9+41	1	0,7	1	41
51	10+9+30	1	0,7	2	30
52	9+9+32	1	0,7	2	32
52	9+42	1	0,7	1	42
52	52	2	1,3	0	52
53	9+9+33	2	1,3	2	33
53	9+43	1	0,7	1	43
54	9+9+34	2	1,3	2	34
54	9+44	1	0,7	1	44
55	9+9+35	1	0,7	2	35
56	9+9+36	2	1,3	2	36

7.4 Appendix 4 – SNPs analysis

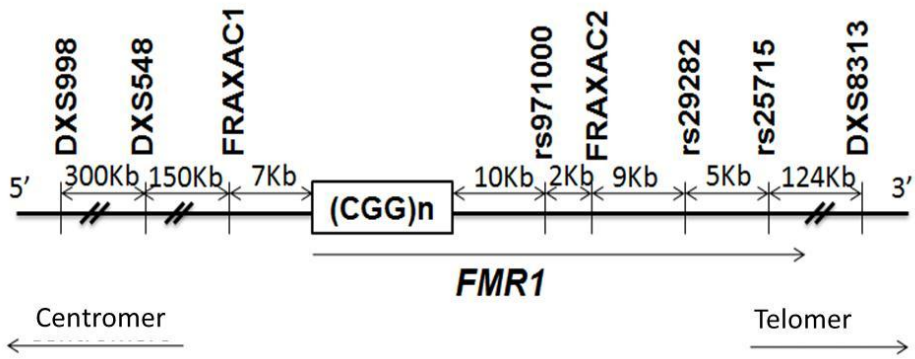


Figure 18. Localization of the molecular markers relatively to the polymorphic region [CGG]. The nucleotidic position of each STR is based on the AC_000066.1 reference sequence of *FMR1* gene (available on GeneBank Database) where these makers are distributed as follow: DXS998: 146959992; DXS548: 147155519; FRAXAC1: 147338298; FRAXAC2: 147357868 and DXS8313: 147480407. The nucleotidic position of each SNP, based on the NT_011681.16 reference sequence is: rs971000: NT_011681.16:g.3444469A>G; rs29282: NT_011681.16:g.3455280T>C and rs25715: NT_011681.16:g.3459968T>C. (Adapted from: LOUREIRO, JOANA MARIA GERALDES DA ROCHA. Síndrome de X-Frágil em Portugal: caracterização molecular por SNPs e STRs, Lisbon: Lisbon University, 2010, Master Thesis.)

Table 19. SNPs lineages: frequency in control population and FXS population.

Lineage	Control Population	FXS Population
TTT	**58,0%**	10,0%
CTC	26,0%	**58,0%**
CCC	8,0%	25,0%
CTT	3,0%	2,0%
TTC	5,0%	5,0%
TCC	0,1%	0,0%
Number of alleles	213	123

Table 20. SNPs analysis and [AGG] interspersion patterns.

[AGG] pattern	Allele Size	Position of the first [AGG] interspersion	rs971000	rs929282	rs25715
8	**8**	**--**	**T**	**T**	**T**
17	17	**--**	**C**	**C**	**C**
10+9	20	**10**	**T**	**T**	**C**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+9	20	**10**	**T**	**T**	**T**
10+10	21	**10**	**T**	**T**	**C**
11+9	21	**11**	**C**	**T**	**C**
13+9	23	**13**	**C**	**T**	**C**
13+9	23	**13**	**C**	**T**	**C**
13+9	23	**13**	**C**	**C**	**C**
13+9	23	**13**	**C**	**T**	**T**
13+9	23	**13**	**T**	**T**	**T**
23	23	**--**	**T**	**T**	**T**
13+9	23	**13**	**C**	**T**	**C**
14+9	24	**14**	**T**	**T**	**T**
14+9	24	**14**	**T**	**T**	**T**
14+9	24	**14**	**C**	**C**	**C**
24	24	**--**	**T**	**T**	**T**
26	26	**--**	**C**	**T**	**C**
9+18	28	**9**	**T**	**T**	**C**
9+9+8	28	**9**	**T**	**T**	**C**
10+7+9	28	**10**	**T**	**T**	**T**
10+7+1+9	28	**10**	**T**	**T**	**T**
10+7+9	28	**10**	**T**	**T**	**T**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**T**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**

9+19	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**T**
9+19	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+9+9	29	**9**	**C**	**T**	**C**
9+10+9	30	**9**	**C**	**C**	**C**
9+10+9	30	**9**	**C**	**T**	**C**
9+9+10	30	**9**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**C**	**C**
10+9+9	30	**10**	**T**	**T**	**C**
10+9+9	30	**10**	**T**	**T**	**T**
9+9+2+7	30	**9**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**C**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
9+10+9	30	**9**	**T**	**T**	**C**
9+10+9	30	**9**	**C**	**T**	**C**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+19	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**C**	**T**	**C**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**C**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**

10+19	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+19	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+9	30	**10**	**T**	**T**	**T**
10+9+10	31	**10**	**T**	**T**	**T**
10+9+10	31	**10**	**T**	**T**	**T**
10+9+10	31	**10**	**T**	**T**	**T**
10+9+10	31	**10**	**T**	**T**	**T**
10+9+10	31	**10**	**T**	**T**	**T**
10+9+10	31	**10**	**T**	**T**	**T**
10+9+10	31	**10**	**T**	**T**	**T**
10+20	31	**10**	**T**	**T**	**T**
31	31	**--**	**C**	**T**	**C**
31	31	**--**	**T**	**T**	**T**
9+12+9	32	**9**	**C**	**C**	**C**
9+12+9	32	**9**	**C**	**C**	**C**
9+12+9	32	**9**	**C**	**C**	**C**
9+12+9	32	**9**	**C**	**C**	**C**
10+10+10	32	**10**	**T**	**T**	**T**
10+21	32	**10**	**T**	**T**	**T**
13+9	32	**13**	**C**	**T**	**C**
10+9+12	33	**10**	**T**	**T**	**T**

10+9+7+9	38	**10**	**C**	**T**	**C**
10+9+7+9	38	**10**	**C**	**T**	**C**
10+9+6+9	38	**10**	**T**	**T**	**T**
9+9+9+9	39	**9**	**C**	**T**	**C**