

**MARTA DO ROSÁRIO CRISTINO SILVA VENTURA**

# **DOENÇA DE MACHADO-JOSEPH:**

**ANÁLISE DA REGIÃO 3'UTR DO GENE *ATXN3***



**Universidade dos Açores**

**Departamento Biologia**

**Ponta Delgada**

**2011**

**MARTA DO ROSÁRIO CRISTINO SILVA VENTURA**

# **DOENÇA DE MACHADO-JOSEPH:**

## **ANÁLISE DA REGIÃO 3'UTR DO GENE *ATXN3***

**Orientadoras:** Professora Doutora Maria Manuela de Medeiros Lima  
Doutora Maria da Conceição Félix Bettencourt

*Dissertação para a obtenção do grau de Mestre  
em Ciências Biomédicas, apresentada à  
Universidade dos Açores.*



**Universidade dos Açores**

**Departamento Biologia**

**Ponta Delgada**

**2011**

## AGRADECIMENTOS

A realização do presente trabalho, só foi possível com a colaboração de diversas pessoas, a todas as que contribuíram de forma directa ou indirecta para a sua realização, gostaria de expressar o meu mais profundo agradecimento.

- À Professora Doutora Maria Manuela de Medeiros Lima, orientadora deste trabalho, um agradecimento muito especial. Agradeço a constante disponibilidade, a sabedoria e os ensinamentos em todo o processo de orientação científica desta dissertação. Foi um enorme privilégio ter sido sua orientanda, apenas a sua orientação em todas as etapas, permitiu a concretização deste trabalho. Por todos os conhecimentos que me transmitiu, expresso desta forma o meu mais sincero agradecimento.
- À Doutora Maria da Conceição Félix Bettencourt, orientadora deste trabalho, um agradecimento muito especial. Ao meu lado, desde o primeiro dia de laboratório, em todos os momentos, mesmo nos mais atrozés, apenas a sua orientação em todas as fases, permitiu a realização do referido trabalho. Agradeço a disponibilidade constante, a transmissão de conhecimentos científicos durante todo o processo desta dissertação. Foi um enorme prazer ter sido sua orientanda, por tudo o que me ensinou, aqui expresso o meu mais sincero agradecimento.
- Ao Professor Doutor Nelson Simões, na qualidade de director do Centro de Investigação de Recursos Naturais (CIRN), em cujos laboratórios se desenvolveu todo o trabalho prático subjacente à presente tese.

- À Dra Mafalda, Bolseira de Investigação do CIRN, agradeço todo o apoio e disponibilidade constantes, que sempre me proporcionou, em todas as etapas do referido trabalho, por tudo, muito obrigada.
- À Dra Teresinha Raposo e ao Dr. José Aires, gostaria de agradecer toda a disponibilidade na logística do meu trabalho no laboratório Dr. José Aires Raposo Lda.. Pela compreensão em todos os momentos, o meu mais sincero agradecimento.
- Ao Marco e ao Tomás, gostaria de agradecer todo o apoio que me deram em todos os momentos, pelas ausências físicas, pela pouca atenção dispensada e pelos dias de mau humor, ao longo das etapas da elaboração da tese, muito obrigado.
- Aos meus pais e à minha irmã, por todo o apoio, o meu muito obrigado.
- Aos meus padrinhos, sempre presentes desde o primeiro dia, com o mesmo discurso de ânimo, pelo apoio constante, o meu mais profundo agradecimento.

Resumo .....	1
<i>Abstract</i> .....	2
<b>1. INTRODUÇÃO .....</b>	<b>3</b>
1.1. A doença de Machado-Joseph (DMJ) .....	3
1.1.1. Identificação e caracterização da DMJ .....	3
1.1.2. Caracterização genética e molecular da DMJ .....	6
1.2. Importância das regiões 3'UTR .....	11
<b>2. PERTINÊNCIA E OBJECTIVOS DO TRABALHO .....</b>	<b>18</b>
<b>3. MATERIAL E MÉTODOS .....</b>	<b>20</b>
3.1. População de estudo .....	20
3.2. Análise de regiões dos exões 10 e 11 do gene <i>ATXN3</i> .....	20
3.2.1. Amplificação .....	20
3.2.2. Sequenciação .....	21
3.2.3. Análise bioinformática .....	23
3.3. Análise estatística .....	23
<b>4. RESULTADOS E DISCUSSÃO .....</b>	<b>25</b>
<b>5. CONCLUSÕES GERAIS .....</b>	<b>34</b>
<b>6. BIBLIOGRAFIA .....</b>	<b>37</b>
<b>7. ANEXOS .....</b>	<b>43</b>

## Resumo

A doença de Machado-Joseph (DMJ), também denominada ataxia espinocerebelosa do tipo 3 (SCA3), é uma doença neurodegenerativa de início tardio (média 40,2 anos), transmitida de um modo autossômico dominante. A DMJ é causada por uma expansão do trinucleótido CAG na região codificante do gene *ATXN3* (14q32.1). O tamanho do tracto repetitivo explica apenas 50-75% da variação na idade de início desta doença, pelo que deverão existir outros factores responsáveis pela variabilidade clínica observada. Assim, apesar da mutação causal, no gene *ATXN3*, estar bem estabelecida, a informação acerca das regiões reguladoras deste gene, tais como a região 3'UTR, que podem influenciar os níveis de expressão do mesmo (e, em última instância, contribuir para a variabilidade clínica observada), é ainda muito reduzida. Com a realização deste trabalho pretendeu-se descrever e analisar, pela primeira vez, a variação genética na região 3'UTR (3' "untranslated region") do gene *ATXN3*, em amostras de doentes com DMJ e, analisar a eventual presença de variabilidade específica dos doentes. Pretendeu-se, para além disso verificar se a existência de polimorfismos, nesta região reguladora, podia interferir com os sinais de poliadenilação já descritos. Para tal, foram utilizadas 40 amostras de DNA genómico, das quais 30 eram de doentes DMJ (casos), pertencendo as restantes 10 amostras a indivíduos aparentemente saudáveis, sem história familiar de DMJ (controlo). Para cada amostra, foram amplificadas e sequenciadas porções do exão 10 e exão 11 do referido gene, que definem a maior parte da sua região 3'UTR. Por comparação das sequências obtidas com três sequências de cDNA, correspondendo a transcritos distintos, confirmou-se que as regiões analisadas correspondiam, efectivamente, à porção do DNA genómico que codifica para a região 3'UTR deste gene. A análise detalhada das 360 sequências obtidas, passíveis de serem analisadas, permitiu identificar, nas regiões estudadas, 10 posições polimórficas correspondentes a 10 SNPs, que já tinham sido previamente descritos. Constatou-se que os alelos mais frequentes, para cada um dos 10 SNPs detectados, eram coincidentes entre os dois grupos estudados (casos e controlos). Para além disso, verificou-se que não existiam diferenças significativas, relativamente às frequências genóticas, entre estes dois grupos. Os dados obtidos no presente trabalho apontam no sentido de que a região analisada do gene *ATXN3* não apresente variação exclusiva nos doentes. As regiões estudadas neste trabalho abrangem quatro sequências consenso para sinais de poliadenilação. A comparação das sequências obtidas com a sequência de referência permitiu constatar que os polimorfismos observados não implicam a eliminação dos sinais de poliadenilação. No entanto, também não originam sequências consenso para novos sinais de poliadenilação. Com base nos resultados obtidos, não será expectável que a poliadenilação na região 3'UTR, do gene *ATXN3*, que foi analisada neste trabalho, possa ter influência na regulação da expressão deste gene.

O presente estudo permitiu obter dados que permitiram conhecer o tipo de variação existente na região 3'UTR do gene *ATXN3* em amostras de doentes DMJ, visando contribuir para uma melhor compreensão das bases moleculares desta doença. Para além disso, os dados aqui gerados poderão servir de base para trabalhos futuros, nomeadamente os que visem analisar a variação genética em regiões reguladoras do gene *ATXN3* e a sua relação com a variabilidade fenotípica da DMJ.

## Abstract

Machado-Joseph disease (MJD), also called spinocerebellar ataxia, type 3 (SCA3), is a late onset neurodegenerative disorder (average age at onset of 40.2 years), transmitted as autosomal dominant. MJD is caused by the expansion of a CAG repeat in the coding region of the *ATXN3* gene (14q32.1). The size of the CAG tract explains only 50-75% of the variation of the age at onset for this disorder, and other factors should be responsible for the clinical variability observed in MJD. Although the causative mutation, at the *ATXN3* locus, is well established, information concerning the regulatory regions of this gene, such as the 3'UTR (3' untranslated), which can influence its levels of expression (and, ultimately, contribute to the clinical variability observed) is still very scarce. The aim of this work was to describe and analyze, for the first time, the variation in the 3'UTR region of the *ATXN3* locus, in MJD patients, and to infer the presence of variability specific of patients. The analysis of the impact of the polymorphisms identified, in terms of interference with the polyadenylation signs, already described for this gene, was, furthermore, an important objective of this work. Forty samples of genomic DNA were used, 30 belonging to MJD patients (cases), the remaining 10 corresponding to apparently healthy individuals without family history of MJD (controls). For each sample, portions of the exons 10 and 11 of the *ATXN3* gene, which define most of the 3'UTR region, were amplified and sequenced. The comparison of the sequences obtained with three DNA sequences, corresponding to different transcripts, confirmed that the analyzed regions corresponded, in fact, to the portion of the genomic DNA which encodes for the 3'UTR region of the *ATXN3* gene. The detailed analysis of the 360 analyzable sequences obtained allowed the identification of 10 polymorphic positions, corresponding to 10 SNPs, which had already been described. The most frequent alleles, for each of the SNPs identified, were coincident between the two groups analyzed (cases and controls). No significant differences on what concerns the genotypic frequencies, were found between the two groups. Data obtained in this work points to the lack of variation exclusive of the patients, in this region of the *ATXN3* gene. The regions analyzed in this work encompass four consensus sequences for polyadenylation signals. The comparison of the sequences obtained with the reference sequence showed that the observed polymorphisms are not implicated in the elimination of polyadenylation signals; however, they also fail to originate consensus sequences for new polyadenylation signals. Based on the results obtained, it does not seem expectable that polyadenylation in the 3'region of the *ATXN3* gene studied, can influence the expression of this gene.

The present study has generated data which will improve our knowledge concerning the variation of the 3'UTR region of the *ATXN3* gene in MJD patients, aiming to contribute to a better understanding of the molecular basis of this disease. Data here in generated, furthermore, can be the basis for future work, namely for those who will be targeted at analyzing the genetic variability in regulatory regions of the *ATXN3* gene and its relation with the phenotypic variability of MJD.

## 1. INTRODUÇÃO

### 1.1. A doença de Machado-Joseph (DMJ)

#### 1.1.1. Identificação e caracterização da DMJ

A doença de Machado-Joseph (DMJ), também denominada ataxia espinocerebelosa do tipo 3 (SCA3), é uma doença neurodegenerativa de início tardio (média de 40,2 anos), transmitida de modo autossômico dominante (Coutinho, 1992). Apesar de ser considerada uma doença de manifestação tardia, foram já reportados casos extremos, com idades de início aos 4 e aos 70 anos (Coutinho, 1992; Carvalho *et al.*, 2008). A sobrevida média estimada é de 21 anos (Coutinho, 1992).

Nos anos 70, mais precisamente entre 1972 e 1976, foram descritas “três novas” doenças neurológicas, em três famílias com ascendência açoriana, residindo nos Estados Unidos da América. A família Machado, identificada por Nakano e colaboradores, em 1972 (Nakano *et al.*, 1972 *fide* Coutinho, 1992), foi a primeira família a ser estudada, tendo sido a doença presente nessa família então denominada de “doença de Machado”. No mesmo ano, foi descrita, numa família que ficou conhecida como Thomas, o que se pensou ser uma segunda doença, que tomou a designação de “degenerescência nigro-espino-dentada com oftalmoplegia nuclear” (Woods e Schaumburg, 1972 *fide* Coutinho, 1992). A terceira descrição da doença, que mais tarde viria a ser conhecida como DMJ, foi feita em 1976 por Rosenberg e colaboradores, na família Joseph, sendo então designada de “forma autossômica dominante da degenerescência estriatonígrica” (Rosenberg *et al.*, 1976 *fide* Coutinho, 1992).

Ao longo dos anos esta doença teve várias designações, tais como “doença açoriana do sistema nevoso”, “doença de Joseph”, “doença de Machado” (já anteriormente mencionada), entre outras. Após a unificação da doença como uma única entidade clínica, a designação actualmente utilizada de “doença de Machado-Joseph” foi-lhe atribuída, em 1981, por Coutinho e Sequeiros. Essa designação foi-lhe atribuída pelo facto da família Machado ter sido a primeira a ser estudada, e da família Joseph incluir o maior grupo de indivíduos afectados, até então conhecido. Para além disso, estas duas famílias apresentavam os dois fenótipos extremos, em relação à idade de início da doença, bem como à severidade do quadro clínico (Coutinho e Sequeiros, 1981 *fide* Coutinho, 1992).

Clinicamente, a DMJ é caracterizada por um quadro complexo e heterogéneo, uma vez que envolve vários sistemas neurológicos, nomeadamente, o cerebeloso, o oculomotor, o piramidal, o extrapiramidal e o periférico (Coutinho, 1992). As principais manifestações desta doença são a ataxia cerebelosa (incoordenação dos movimentos, manifestação clínica mais frequente, presente em 98,6% dos doentes) e a oftalmoparésia externa progressiva - OEP (limitação dos movimentos oculares, segunda manifestação clínica mais frequente, presente em 93,1% dos doentes). A estas manifestações principais podem estar associados, em graus variáveis, sinais piramidais, extrapiramidais e periféricos (Coutinho, 1992).

Devido à elevada variabilidade fenotípica que caracteriza a DMJ, foram propostos três tipos clínicos principais (Coutinho e Andrade, 1978 *fide* Coutinho, 1992). O tipo 1 tem um início mais precoce, em média aos 24 anos, apresenta uma evolução mais grave, sendo caracterizado por ataxia

cerebelosa, OEP e sinais piramidais e extrapiramidais marcados. O tipo 2 tem uma idade de início intermédia (média aos 40 anos), é o tipo clínico mais frequente, no qual se manifestam a ataxia cerebelosa, a OEP, com ou sem sinais piramidais associados, e com sinais extrapiramidais e periféricos que, quando existem, são ligeiros. O tipo 3 tem um início mais tardio (média aos 47 anos), com desenvolvimento mais lento, caracterizado por ataxia cerebelosa, OEP, sinais piramidais e extrapiramidais ligeiros ou ausentes, e sinais periféricos muito evidentes. Um quarto tipo, que se caracteriza por sinais cerebelosos e parkinsonismo, foi descrito, mais tarde, por Suite e colaboradores (Suite *et al.*, 1986 *fide* Bettencourt, 2009). De um modo geral, todos os casos se manifestam, na fase mais inicial, por ataxia cerebelosa e OEP, diferenciando-se *a posteriori*, após um período de 5 a 10 anos, em tipo 1 ou 3, com o aparecimento de sinais piramidais e extrapiramidais ou sinais periféricos evidentes, respectivamente; podem, por outro lado, após o período anteriormente referido, estabelecerem-se como tipo clínico 2 (Coutinho, 1992).

Apesar de a DMJ ser, à semelhança de outras ataxias espinocerebelosas (SCAs), actualmente incurável, é possível a diminuição da intensidade de alguns dos sintomas, manifestados durante o curso da doença, mediante a administração de fármacos adequados, que podem, assim, melhorar a qualidade de vida destes pacientes (Paulson, 2007).

A DMJ é considerada a SCA mais comum a nível mundial, representando cerca de 15 a 45% da totalidade das SCAs em diversos países e etnias (Paulson, 2007). No arquipélago dos Açores, mais precisamente na ilha das Flores, a doença atinge os mais elevados valores de prevalência a nível mundial (1:239) (Bettencourt *et al.*, 2008a). Dada a elevada representação

desta doença, nas ilhas açorianas, a DMJ tem sido objecto de estudo por diversos investigadores desta área. Assim, em 1996 foi feito um extenso levantamento epidemiológico e genealógico das trinta e quatro famílias açorianas com DMJ (Lima, 1996). A investigação da origem da mutação nos Açores, com base na reconstituição das famílias afectadas, foi também alvo de análise (Lima, 1996; Lima *et al.*, 1997). Estudos dos mecanismos relacionados com a dinâmica mutacional do *locus* da DMJ foram também empreendidos (Lima *et al.*, 2005; Bettencourt *et al.*, 2008a; b). Do mesmo modo, trabalhos que incidem sobre aspectos moleculares relacionados com o gene causal da DMJ têm sido desenvolvidos, nomeadamente no que se refere ao estudo da variação existente a nível dos transcritos deste gene (Bettencourt *et al.*, 2010a; b). Será ainda pertinente referir os trabalhos sobre o programa açoriano de Aconselhamento Genético e Teste Preditivo para a DMJ (Lima *et al.*, 2001; Gonzalez *et al.*, 2004).

### **1.1.2. Caracterização genética e molecular da DMJ**

As doenças genéticas de transmissão dominante são, normalmente, raras. A DMJ, cuja transmissão é autossómica dominante, como foi referido anteriormente, está inserida nesse grupo de doenças raras, sendo os doentes, na sua grande maioria, heterozigóticos (Zlotogora, 2006).

Nos indivíduos homozigóticos, descritos em número muito reduzido, o fenótipo apresenta-se com uma maior severidade e com um início mais precoce, devido ao facto de possuírem o alelo mutado em ambos os cromossomas (Carvalho *et al.*, 2008). Isto indica que a dosagem génica é

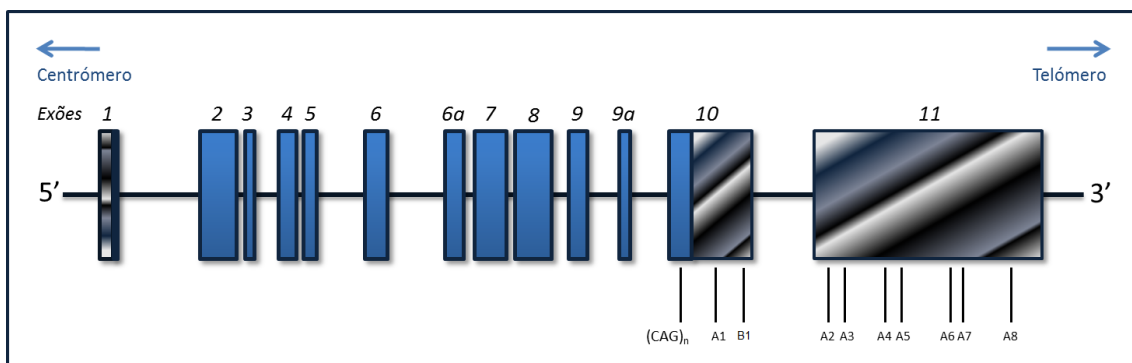
importante na determinação da idade de início desta doença, bem como da sua severidade (Lima, 2010).

O gene da DMJ apresenta uma penetrância praticamente completa, com um valor estimado de 98% (Sequeiros, 1996 *vide* Bettencourt *et al.*, 2008a). Os “saltos de gerações” são raros nas famílias com esta doença e os casos de portadores obrigatórios assintomáticos parecem ser explicados quer pela morte precoce dos indivíduos, em relação à idade média de início da doença, quer pela emigração, o que impede a confirmação do estatuto para a doença, nesses indivíduos (Coutinho, 1992). No entanto, o padrão penetrância é dependente da idade, o que significa que o risco residual depende da idade do indivíduo, ou seja, à medida que a idade avança, e o indivíduo em risco permanece assintomático, a probabilidade de vir a desenvolver a doença tornar-se-á cada vez mais reduzida (Bettencourt *et al.*, 2008a).

A DMJ é causada por uma expansão do trinucleótido CAG na região codificante do gene causal (Kawaguchi *et al.*, 1994), que conduz, ao nível da proteína, a uma expansão de um tracto de poliglutamina. A DMJ insere-se, desta forma, num grupo de doenças denominadas “doenças de poliglutamina” (doenças “poly-Q”).

O gene responsável pela DMJ (*ATXN3*) está localizado no braço longo do cromossoma 14 (14q32.1) (Takiyama *et al.*, 1993) e tem uma extensão de aproximadamente 48,2kb (Ichikawa *et al.*, 2001). A mutação da DMJ consiste, como mencionado anteriormente, na expansão de um tripleto CAG, que se localiza no exão 10 do referido gene, a 5' (Kawaguchi *et al.*, 1994).

Foram, inicialmente descritos 11 exões para o gene *ATXN3* (Ichikawa *et al.*, 2001); 2 exões adicionais foram, posteriormente, descritos (o exão 6a e o 9a) (Bettencourt *et al.*, 2010a) (Figura 1). Algumas porções do exão 10 e do exão 11, que correspondem às regiões estudadas no presente trabalho, são utilizadas por “splicing” alternativo e poliadenilação diferencial para formar a região 3' UTR (“untranslated region”, ou seja, região não traduzida) do RNA mensageiro (mRNA). Segundo a descrição de Ichikawa e colaboradores (2001), o exão 10 tem uma extensão de 1099pb e contém uma sequência consenso (AATAAA), para um sinal de poliadenilação (Figura 1); o exão 11, por sua vez, é composto por 6046pb e contém sete sequências consenso (AATAAA) para sete sinais de poliadenilação. Dependendo da porção dos exões 10 e 11 que é incluída no mRNA, a região 3'UTR poderá ter uma maior ou menor extensão.



**Figura 1.** Representação esquemática do gene *ATXN3*. Os exões estão numerados de 1 a 11 e estão representados como retângulos. Os retângulos preenchidos a azul indicam as regiões codificantes, o retângulo com linhas horizontais representa a região 5'UTR e os retângulos com linhas diagonais correspondem à região 3'UTR. A localização do tracto polimórfico (CAG)<sub>n</sub> está indicado. As sequências consenso para os 8 sinais de poliadenilação previamente descritos estão indicadas de A1 a A8 (Ichikawa *et al.*, 2001). B1 é um possível sinal de poliadenilação adicional localizado no final do exão 10 (adaptado de Bettencourt, 2009).

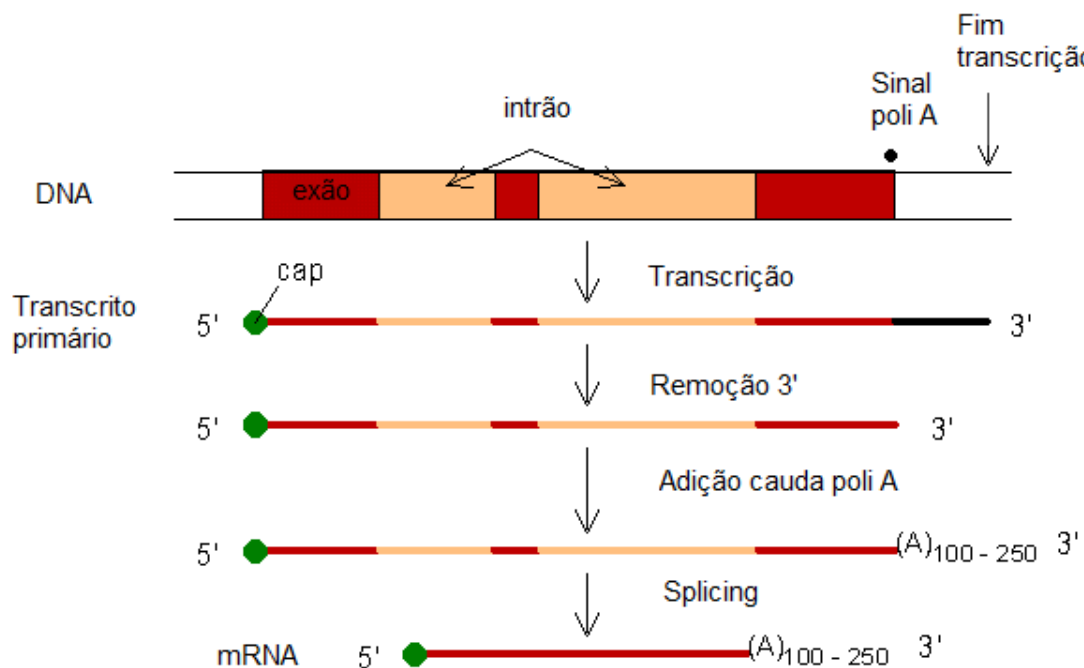
O gene *ATXN3* codifica uma proteína, denominada ataxina-3. A ataxina-3 pertence ao grupo das cisteína-proteases e tem um peso molecular de 42 kDa, na sua forma nativa. Estruturalmente é composta por 339 aminoácidos e por um número variável de glutaminas, que constituem o tracto poly-Q (Kawaguchi *et al*, 1994). A ataxina-3 é composta por um domínio Josefina (JD), na porção N-terminal, enquanto na extremidade C-terminal contém dois ou três motivos de interacção com a ubiquitina (UIMs) e o tracto poly-Q. Sabe-se que a forma nativa da proteína tem uma função de enzima desubiquitinante, na via proteossoma das ubiquitinas. Assim, e dadas as várias funções que têm sido propostas para a ataxina-3, pensa-se que a proteína normal participa em vias de controlo celular da qualidade das proteínas (Lima, 2010).

O gene *ATXN3* expressa-se ubiquamente, em tecidos neuronais e não neuronais. Nos neurónios, a proteína nativa encontra-se essencialmente no citoplasma; a proteína mutada, no entanto, é observada quer no núcleo quer no citoplasma das células (Paulson, 1997). Pensa-se que patologia poderá surgir, como consequência de um ganho de função neurotóxica pela proteína mutada, uma vez que, contém uma expansão do tracto poly-Q. Na presença da ataxina-3 mutada observa-se a formação de inclusões neuronais intranucleares (Paulson, 2007). Este tipo de inclusões intranucleares é frequentemente encontrado no cérebro de pacientes, portadores desta e outras doenças "poly-Q". Apesar destas inclusões existirem em doentes com DMJ, o padrão observado não se correlaciona totalmente com as regiões afectadas do cérebro e, portanto, o seu papel na neurodegeneração, ainda não está estabelecido (Paulson, 2007).

No locus da DMJ, os alelos normais apresentam entre 12 e 44 repetições do triplete CAG, enquanto os alelos expandidos apresentam consensualmente um número de repetições que varia entre 61 a 87 (Maciel *et al.*, 2001). O número de repetições CAG está inversamente relacionado com a idade de início da DMJ, ou seja, quanto maior o número de tripletos CAG, menor será a idade de início. No entanto, o tamanho do tracto repetitivo explica apenas cerca de 50 a 75% da variação na idade de início (Lima, 1996; Bettencourt *et al.*, 2010b). Para além de o tamanho do alelo expandido estar correlacionado de modo inverso com a idade de início, a extensão do tracto CAG correlaciona-se, ainda, com a frequência de sinais clínicos particulares, tais como sinais piramidais, que são mais frequentes em indivíduos com tractos repetitivos maiores (Takiyama *et al.*, 1995). O facto da correlação entre o tamanho do tracto CAG e a idade de início da doença ser incompleta implica que terão de existir outros factores que contribuam para a explicação da restante variação na idade de início da doença, nomeadamente genes modificadores, que poderão interagir com o gene da DMJ. Para além disso, poderão, ainda, ocorrer variações no próprio gene *ATXN3* (para além do próprio tracto CAG), quer na sua região codificante, quer em zonas reguladoras (nomeadamente a nível do promotor, da região 5'UTR e da região 3'UTR). Estas zonas têm, portanto, potencial para melhorar as correlações genótipo-fenótipo. Polimorfismos localizados, por exemplo, na região 3'UTR, que afectem o processamento e estabilidade do mRNA, poderão ter efeitos a nível da expressão génica, podendo dessa forma influenciar o fenótipo da doença (Cooper e Krawczak, 1993).

## 1.2. Importância das regiões 3'UTR

Nos seres eucariotas os precursores do mRNA, pré-mRNAs ou transcritos primários, são transcritos no núcleo, a partir do DNA genómico, pela enzima RNA polimerase II. Estes ainda não representam o mRNA maduro, ou seja, o mRNA funcional. Para que tal aconteça, esses transcritos têm que sofrer um conjunto de alterações bioquímicas, mecanismos denominados globalmente por “processamento do mRNA”; só então o mRNA pode ser transportado para o citoplasma, onde ocorre a sua tradução em proteínas. Os mecanismos que ocorrem a nível da maturação do mRNA são (Figura 2): o “capping”, o “splicing” e a poliadenilação que, seguidamente, serão descritos em maior detalhe (Mandel *et al.*, 2008).



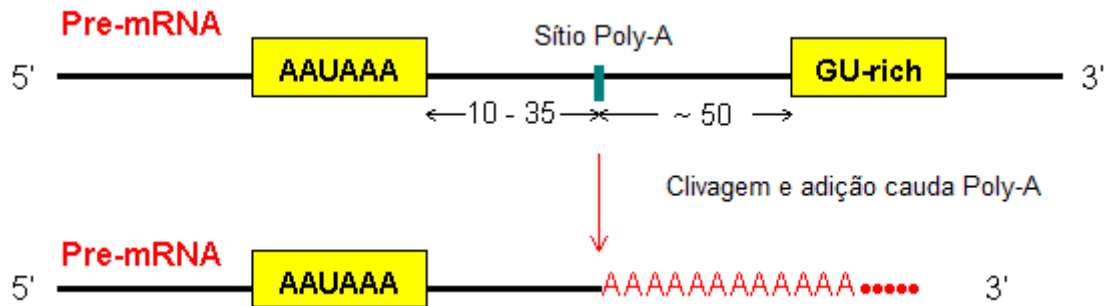
**Figura 2.** Mecanismos envolvidos na maturação do mRNA (adaptado de *Molecular Biology Web Book: Chapter 5*, disponível em: <http://www.web-books.com/MoBio/Free/Ch5A.htm>).

O “capping” ocorre logo após a transcrição e consiste na adição de um resíduo de guanina metilada (7-metilguanosina) ao primeiro nucleótido do pré-mRNA, por uma ligação fosfodiéster (5'-5'), formando uma estrutura denominada de “cap”. Todas essas reações são catalisadas por enzimas específicas. O “cap” tem algumas funções peculiares tais como: proteger a extremidade 5' do transcrito do ataque de exonucleases, facilitar o transporte do núcleo para o citoplasma e facilitar o “splicing” do pré-mRNA, tendo ainda um papel importante na síntese proteica (Strachan e Read, 2004).

O processo de poliadenilação, consiste na adição de uma cauda de poliadeninas (poly-A) à região 3'UTR e inclui duas etapas intimamente ligadas, a primeira das quais envolve a clivagem do pré-mRNA num local específico e, a segunda, a adição de múltiplos resíduos de adenina ao pré-mRNA. A poliadenilação, na região 3' UTR, desempenha um papel fundamental na expressão génica, porque as moléculas de mRNA processadas indevidamente não são transportadas do núcleo para o citoplasma para serem traduzidas. Para além disso, a cauda poly-A influencia a estabilidade do mRNA, bem como a sua tradução. Desta forma a poliadenilação é uma etapa fundamental na regulação da expressão génica (Lutz, 2008).

A sequência consenso do sinal de poliadenilação é altamente conservada. O hexanucleótido AAUAAA, é usado na maior parte dos genes (cerca de 80%) como sinal de poliadenilação, enquanto o hexanucleótido AUUAAA é usado com menor frequência (Edwalds-Gilbert *et al.*, 1997). O sinal de poliadenilação consiste numa sequência reconhecida pelos factores de mRNA, sendo essencial para o final da transcrição e para a eficácia da

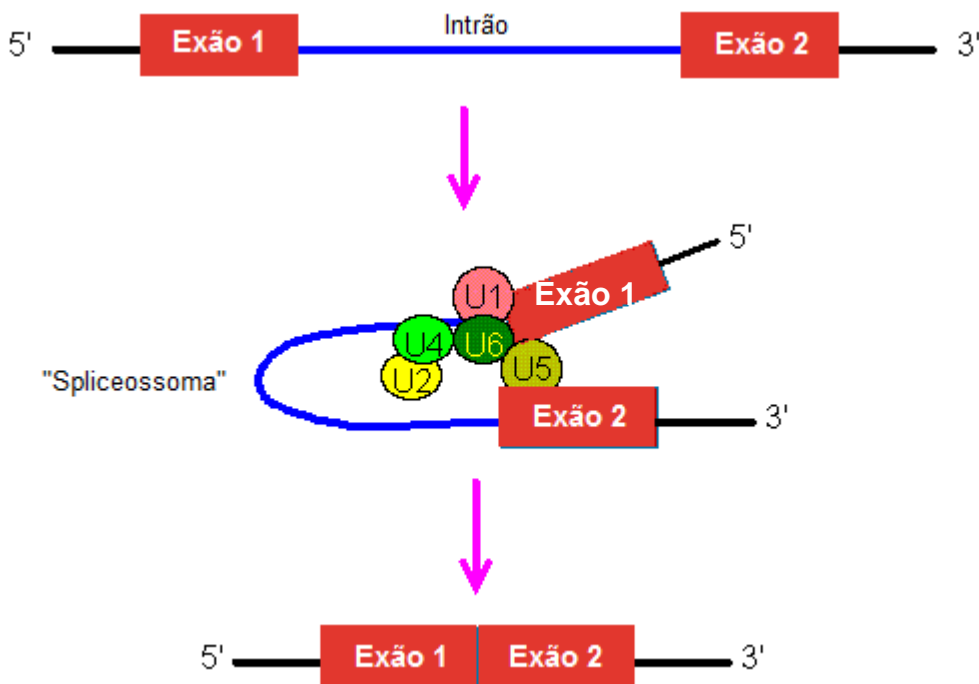
poliadenilação do mRNA (Chatterjee e Pal, 2009). Um esquema ilustrativo do mecanismo de poliadenilação pode ser observado na Figura 3.



**Figura 3.** Poliadenilação na região 3' do pré-mRNA. O sinal importante para a clivagem é a sequência consenso **AAUAAA**. A clivagem ocorre 10-35 nucleótidos a 3' desta sequência específica. Um segundo sinal, rico em GU, está localizado cerca de 50 nucleótidos a 3' do local de clivagem (adaptado de *Molecular Biology Web Book: Chapter 5*, disponível em: <http://www.web-books.com/MoBio/Free/Ch5A.htm>).

Nos eucariotas, os transcritos primários (pré-mRNA) apresentam dois tipos de sequências: os exões, porções codificantes do transcrito primário e que estão presentes no mRNA maduro; e os intrões, sequências não codificantes e que não estão presentes no mRNA maduro (Strachan e Read, 2004). Os intrões precisam ser retirados para dar origem a um mRNA funcional. Assim, após a adição do “cap” e da cauda poly-A, o pré-mRNA precisa sofrer o processo de excisão dos intrões e junção dos exões adjacentes. Esse mecanismo é conhecido como “splicing” do pré-mRNA. Para que o “splicing” ocorra correctamente existem sequências consenso (sequência dadora, sequência aceitadora e *branch point*), localizadas nas fronteiras

exões/intrões; as sequências dadoras e aceitadoras têm a finalidade de sinalizar o local onde irá ser efectuado o “splicing”. O “splicing” no pré-mRNA é mediado por um vasto complexo RNA-proteínas denominado “spliceossoma”, este complexo é constituído por 5 tipos de “small nuclear RNA” (snRNA) e por mais de 50 proteínas, as moléculas snRNA associam-se com proteínas para formar as pequenas ribonucleoproteínas nucleares (snRNP) (Strachan e Read, 2004). Um esquema ilustrativo do mecanismo de “splicing” pode ser observado na Figura 4.



**Figura 4.** Esquema representativo da actuação do “spliceossoma” durante o processo de “splicing” do pré-mRNA. U1, U2, U4, U5, U6 correspondem às pequenas ribonucleoproteínas nucleares (snRNPs) envolvidas (adaptado de *Molecular Biology Web Book: Chapter 5*, disponível em: <http://www.web-books.com/MoBio/Free/Ch5A4.htm>).

Até há relativamente pouco tempo existia uma percepção consensual de que as únicas secções do mRNA com impacto funcional eram as regiões codificantes. Durante as últimas décadas essa visão tem sido alterada, como consequência dos estudos que têm demonstrado que as regiões 5' e 3'UTR do mRNA contêm motivos capazes de regular muitos aspectos da função do mRNA (Hughes, 2006).

A expressão génica é um processo dinâmico e sujeito a vários mecanismos de regulação. No sistema nervoso, distúrbios a nível da regulação génica não só podem causar diversas formas de cancro, como podem levar ao desenvolvimento de várias doenças neurológicas, tais como: síndrome de Rett e síndrome de X frágil (Bolognani e Perrone-Bizzozero, 2008). O primeiro passo na expressão génica é a transcrição, um processo regulado por uma grande variedade de factores de transcrição e proteínas que afectam a metilação do DNA e a estrutura da cromatina. Para além da transcrição, importantes mecanismos de regulação génica estão presentes ao nível do processamento do pré-mRNA, ao nível da tradução e ao nível pós-tradução (Regateiro, 2003). As regiões UTRs do mRNA são ricas em elementos reguladores importantes, para a exportação nuclear do mRNA, para a sua localização citoplasmática e para a própria estabilidade do mRNA, bem como para a eficiência da sua tradução (Strachan e Read, 2004). Assim sendo, tanto a região 3'UTR como a região 5'UTR do mRNA desempenham um papel importante na regulação pós-transcricional da expressão génica (Day e Tuite, 1998). A região 5'UTR contém sequências responsáveis pelo início da tradução, de forma que tem que ser mais conservada que a região 3'UTR. A região 3'UTR, contudo, é decisiva para o destino de um mRNA específico, na

maioria dos casos (Grzybowska *et al.*, 2001). A região 3'UTR também é caracterizada por formar estruturas secundárias, que desempenham um papel importante na interacção dos mRNAs com as proteínas que se ligam ao RNA ("RNA binding proteins"). Uma perturbação nessas estruturas, devido a uma mudança na sequência, alterará a interacção do mRNA com essas proteínas. Existem várias doenças que são causadas por mutações nas sequências das regiões 3'UTR ou nas "RNA binding proteins", das quais se podem destacar alguns carcinomas, artrite crónica e algumas doenças inflamatórias (Grzybowska *et al.*, 2001). Um estudo realizado por Chen e colaboradores (2006b), em 83 variantes na região 3'UTR de vários mRNAs humanos de genes associados a doenças, revelou uma correlação entre a funcionalidade destas variantes e alterações na estrutura secundária prevista.

A investigação sobre a fisiopatologia das doenças e as mutações que afectam as funções da região 3'-UTR ainda é escassa. No entanto, os dados disponíveis sugerem que esta região do mRNA, que é muitas vezes negligenciada durante a selecção de variação genética em genes candidatos, desempenha um papel importante em várias doenças, nomeadamente na progressão da própria doença. Segundo o conhecimento actual, as doenças resultantes de mutações nos sinais de poliadenilação são raras (Chatterjee e Pal, 2009).

As doenças causadas por expansões de tripletos (grupo de cerca de 19 doenças genéticas humanas no qual também se inclui a DMJ), de acordo com a localização do motivo repetitivo, são classificadas em quatro classes: 5'UTR, exões, intrões e 3'UTR. Nesta última classe, destaca-se a distrofia miotónica do tipo 1, que é uma doença severa, causada por uma expansão do motivo CTG

na região 3'UTR do gene *DMPK*, cujo número de repetições CTG varia entre 5 e 37 nos alelos normais, podendo atingir até 3000 repetições nos alelos expandidos (Bettencourt *et al.*, 2007).

A investigação sobre as regiões 3'UTR pode ainda ser considerada incipiente, comparativamente aos dados disponíveis sobre a regulação da transcrição. Contudo, é cada vez mais claro que os mecanismos de regulação mediados pelas regiões 3'UTR poderão ter um grande impacto na patogénese. Algumas doenças multissistémicas são causadas por alterações em regiões 3'UTR, o que demonstra que um profundo conhecimento destas sequências reguladoras, bem como das proteínas específicas que se ligam a elas ("RNA binding proteins") poderá ser muito útil para a compreensão dos respectivos mecanismos da patogénese e para o desenvolvimento de uma terapia eficaz (Grzybowska *et al.*, 2001).

## 2. PERTINÊNCIA E OBJECTIVOS DO TRABALHO

Com a realização deste trabalho, pretendeu-se analisar e descrever, pela primeira vez em amostras de doentes DMJ, a variação genética na região 3'UTR do gene *ATXN3*, que corresponde ao gene causal da DMJ. Para tal, analisou-se, em amostras de DNA genómico (de 30 doentes DMJ e 10 indivíduos controlo), a sequência de porções do exão 10 e exão 11 do referido gene, que codificam para a região 3'UTR. O principal objectivo consistiu em identificar, nestes dois grupos de indivíduos, polimorfismos que possam, nomeadamente, interferir com os sinais de poliadenilação, eliminando os previamente descritos ou criando novos sinais. Para além disso, pretendeu-se, efectuar comparações entre o grupo de doentes DMJ e o grupo de indivíduos controlo, no que se refere aos polimorfismos encontrados, visando determinar se existe variabilidade que seja específica dos doentes.

A pertinência do estudo realizado está relacionada com o facto de, no presente momento, se conhecer muito pouco acerca da variabilidade genética presente na região que se analisou, principalmente em doentes. À semelhança de outras doenças, mencionadas na introdução do presente trabalho, também na DMJ as alterações na região 3'UTR poderão ter consequências na expressão do gene *ATXN3*. No caso da DMJ, apesar da mutação causal estar bem estabelecida, alterações nestas regiões reguladoras poderão, em última instância, contribuir para a heterogeneidade clínica observada nesta doença.

Assim, com o presente estudo pretendeu-se obter informações que possam contribuir para uma melhor compreensão das bases moleculares da DMJ e que possam servir de base para trabalhos futuros, nomeadamente

sobre variação genética em regiões reguladoras do gene *ATXN3* e a sua relação com a variabilidade fenotípica da DMJ.

### 3. MATERIAL E MÉTODOS

#### 3.1. População de estudo

Para o desenvolvimento deste trabalho foram utilizadas 40 amostras de DNA. Da totalidade das amostras, 30 eram de doentes DMJ (casos), não aparentados e confirmados molecularmente, apresentando a mutação no exão 10 do gene *ATXN3* (número de repetições CAG superior a 61). As restantes 10 amostras pertenciam a indivíduos controlo, aparentemente saudáveis, sem história familiar de DMJ, e previamente excluídos para a mutação (número de repetições CAG inferior a 44). As amostras de DNA tinham sido previamente obtidas por extracção a partir de sangue total, por métodos “standard”. Todas as amostras foram colhidas após consentimento informado. As amostras relativas aos doentes foram colhidas no âmbito de consultas de neurologia coordenadas pelo Dr. João Vasconcelos (Director de Serviço de Neurologia do Hospital Divino Espírito Santo em Ponta Delgada).

#### 3.2. Análise de regiões dos exões 10 e 11 do gene *ATXN3*

##### 3.2.1. Amplificação

Regiões dos exões 10 (fragmento com uma extensão de 1300 a 1500 pb) e 11 (fragmento com uma extensão de cerca de 2700pb) do gene *ATXN3*, descritas como sendo transcritas mas localizadas *downstream* ao codão «stop» (codificando parte da região 3'UTR), foram amplificadas, por PCR, usando primers específicos. Para a amplificação do exão 10 foram utilizados os primers MJD-10F (5'-AGATTTTCCTAAGATCAGCACTTCC-3') e MJD-10R (5'-

CTTCATTAAGCCATACCTACAGG-3'); para o exão 11 foram utilizados os primers MJD-11F (5'-AGGAAATGGGGATCGACCACAAAG-3') e MJD-11R (5'-CACAAACACATTCAAACGCATCCAG-3').

A reacção de amplificação foi efectuada num volume total de 25µl com 0,3µM de cada primer, 300µM de dNTPs, 2mM de MgCl<sub>2</sub>, 10xtampão de reacção de PCR [160mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 670mM Tris-HCl (pH 8,8 a 25°C), 0,1% Tween-20], 1U de Biotaq DNA polimerase (Bioline) e aproximadamente 100ng de DNA genómico. Para o exão 10 foram utilizadas as seguintes condições: uma primeira desnaturação durante 5' a 95°C; seguida de 37 ciclos de 30'' a 94°C, 30'' a 58°C e 1'20'' a 72°C; e uma extensão final de 7' a 72°C. Para o exão 11 as condições foram as seguintes: uma primeira desnaturação durante 5' a 95°C; seguida de 37 ciclos de 30'' a 94°C, 30'' a 59°C e 3' a 72°C; e uma extensão final de 10' a 72°C.

Os produtos de PCR foram purificados, utilizando o *Jetquick PCR Purification Spin Kit* (Genomed), de acordo com as instruções especificadas pelo fabricante.

### **3.2.2. Sequenciação**

Os produtos de PCR anteriormente obtidos foram sequenciados segundo método de Sanger, utilizando o *BigDye® Terminator v3.1 Cycle Sequencing Kit* (Applied Biosystems), de acordo com as instruções do fabricante. As amostras foram sequenciadas, em ambos os sentidos ("forward" e "reverse"), utilizando os "primers" descritos na Tabela 1, cobrindo-se, com este conjunto de sequências, uma extensão total de ~1100pb correspondentes ao exão 10 e de ~2000pb correspondentes ao exão 11. Os produtos das 10 reacções de

sequenciação efectuadas, para cada amostra, foram posteriormente purificados com EDTA/Etanol, de acordo com o seguinte procedimento: adicionou-se 5µl de EDTA (125mM) ao produto de sequenciação, de seguida adicionou-se 60µl de Etanol a 100%, agitou-se bem no “vortex”, deixou-se incubar à temperatura ambiente e ao abrigo de luz, durante 15’; seguidamente centrifugou-se 20’ a 13000rpm, desprezou-se o sobrenadante e, após secagem, adicionou-se 60µl de Etanol a 70%, centrifugou-se novamente 20’ a 13000 rpm, retirou-se o sobrenadante e secou-se o “pelet”. As amostras foram ressuspendidas em 20µl de formamida e as leituras foram efectuadas num sequenciador ABI PRISM® 310 (Applied Biosystems).

**Tabela 1.** Sequências dos “primers” utilizados nas reacções de sequenciação.

Primers	Sequências
MJD 10aF	5'-GACCTATCAGGACAGAGTTCAC-3'
MJD 10bF	5'-TCCTGGGTTTCATGGCATTCTCC-3'
MJD 10R	5'-CTTCATTAAGCCATACCTACAGG-3'
MJD 10bR	5'-GTAAGTCTCCTTAATCCAGGG-3'
MJD 11F	5'-AGGAAATGGGGATCGACCACAAAG-3'
MJD 11aF	5'-GGATCTTTTGCAGACTAGC-3'
MJD 11bF	5'-CTAGGAGAAGCATAGTATGCAGG-3'
MJD 11aR	5'-CCTTTTGTGTAAGGGAACTTCAG-3'
MJD 11bR	5'-CACCTACATTCCATGACTCAACTG-3'
MJD 11cR	5'-CATGATGTGAGCCAACCTACC-3'

### **3.2.3. Análise bioinformática**

As sequências obtidas foram visualizadas e alinhadas com uma sequência de referência (GenBank código de acesso AB038653.1), utilizando o software BioEdit v 7.0.5.3. (Hall,1999). Para a confirmação de que as regiões analisadas correspondiam à porção do DNA genómico que codifica a região 3'UTR do gene *ATXN3*, foram utilizadas três sequências de cDNA, disponíveis na base de dados do “National Center for Biotechnology Information (NCBI)” ([www.ncbi.nlm.nih.gov/](http://www.ncbi.nlm.nih.gov/)), com os seguintes códigos de acesso: NM 004993.4, S75313.1, U64821.1.

### **3.3. Análise estatística**

As frequências genóticas e alélicas foram determinadas para os 10 “single nucleotide polymorphisms” (SNPs) observados. A conformidade com o equilíbrio de Hardy-Weinberg (HW), nos grupos de casos e de controlos, foi avaliada com um teste exacto, de acordo com Guo e Thompson (1992). O valor da diversidade genética, tal como descrita por Nei (1987), foi calculado para os mesmos SNPs.

Foram efectuados testes exactos de diferenciação, usando as frequências genóticas, para comparar o grupo de casos e o de controlos, considerando cada SNP separadamente.

A independência alélica entre SNPs foi testada através de comparações entre pares de SNPs, em casos e controlos, considerando apenas as amostras genotipadas para todos os SNPs (n=16 para o grupo dos casos e n=10 para o

grupo dos controlos). O teste efectuado (“Likelihood ratio test of linkage disequilibrium” - LD) está adaptado a dados genotípicos, nos quais a fase gamética é desconhecida. Aos valores de  $p$  obtidos, para os pares de SNPs, foi aplicada a correcção de Bonferroni para testes múltiplos.

Todos os testes acima referidos foram efectuados recorrendo ao programa Arlequin 3.5.1.2 (Excoffier, 1998-2010) e foi considerado um nível de significância de 0,05.

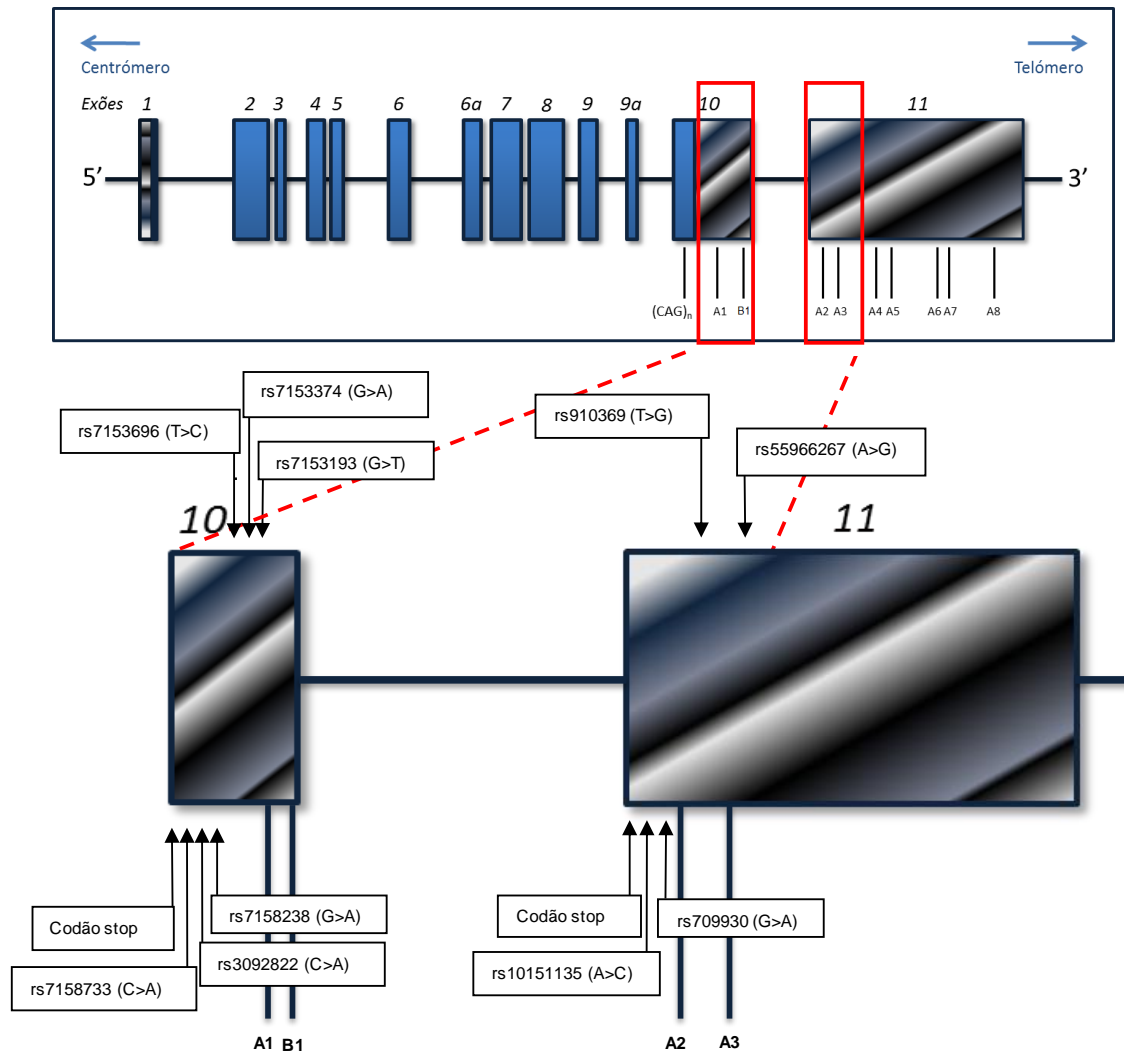
## 4. RESULTADOS E DISCUSSÃO

A sequenciação de uma região de cerca 1100pb do exão 10 e de uma região de cerca 2000pb do exão 11 do gene *ATXN3* permitiu obter um total de 360 sequências, passíveis de serem analisadas. A porção estudada destes dois exões abrange um total de quatro sequências consenso para sinais de poliadenilação, cobrindo os sinais de poliadenilação A1, A2 e A3, descritos por Ichikawa e colaboradores (2001). Para além desses três, abrange ainda um outro possível sinal de poliadenilação (uma vez que está presente o hexanucleótido AATAAA), existente no final do exão 10 (representado como B1 na Figura 1). Para toda a região sequenciada foi possível analisar a presença de polimorfismos, com excepção de uma região do exão 10 de cerca de 300pb para a qual, por motivos de ordem técnica (existência de motivos repetitivos nessa zona), não se obtiveram sequências de boa qualidade.

As sequências obtidas foram comparadas com três sequências de cDNA, correspondendo a transcritos distintos, para confirmar que as regiões analisadas correspondiam realmente à porção do DNA genómico que codifica para a região 3'UTR deste gene. Num dos transcritos utilizados para comparação (sequência com o código de acesso NM\_004993.4), devido à ocorrência de “splicing” alternativo é excluída a parte do exão 10 que está *downstream* ao tracto CAG e é utilizado um codão “stop” no início do exão 11, pelo que a região 3'UTR corresponderia aos aproximadamente 1500pb do exão 11 que se seguem, terminando cerca de 800pb após a sequência consenso para o segundo sinal de poliadenilação deste exão (A3, segundo Ichikawa *et al.*, 2001). Nos outros dois transcritos (sequências com os códigos de acesso

S75313.1 e U64821.1), é utilizado um codão “stop” que está imediatamente a 3' do tracto repetitivo CAG no exão 10, pelo que as suas regiões 3'UTR corresponderiam aos aproximadamente 400pb ou 200pb, respectivamente, que se seguem e que pertenciam ao exão 10. As sequências obtidas no presente trabalho englobam uma região mais vasta e abrangem as regiões 3'UTR dos transcritos anteriormente referidos na sua totalidade. Após esta comparação foi então possível confirmar que as regiões estudadas correspondem a porções do DNA genómico que codificam a região 3'UTR do gene *ATXN3*.

A análise detalhada das 360 sequências obtidas permitiu identificar, nas regiões estudadas, 10 posições polimórficas correspondentes a 10 SNPs (ver Figura 1A constante nos anexos), que já tinham sido previamente descritos e que se encontram depositados na base de dados de SNPs do NCBI (<http://www.ncbi.nlm.nih.gov/snp>). A localização relativa dos SNPs encontrados, identificados pelos respectivos rs, está evidenciada na Figura 5. Dos 6 SNPs identificados no exão 10, 3 correspondem a transições (rs7158238, rs7153696 e rs7153374) e 3 a transversões (rs7158733, rs3092822 e rs7153193). Dos 4 SNPs no exão 11, 2 correspondem a transições (rs709930 e rs55966267) e 2 a transversões (rs10151135 e rs910369).



**Figura 5.** Representação esquemática do gene *ATXN3*, com destaque para as regiões analisadas no presente trabalho dos exões 10 e 11. A localização relativa dos SNPs encontrados em ambos os exões, identificados pelos respectivos rs, bem como a indicação dos respectivos alelos (ancestral e derivado) também está em evidência na figura. Os possíveis codões stop, em ambos os exões, estão indicados. Três dos sinais de poliadenilação identificados por Ichikawa *et al.* (2001) estão representados com a numeração de A1 a A3. Com a designação de B1 está um possível sinal de poliadenilação adicional, localizado no final do exão 10 (adaptado de Bettencourt, 2009).

Os valores das frequências alélicas e genóticas para os 10 SNPs observados, para ambos os grupos de indivíduos estudados (casos e controlos), encontram-se descritos na Tabela 2. Na Tabela 1A constante dos anexos, encontram-se os genótipos de cada SNP para todas as amostras.

**Tabela 2.** Frequências genóticas e alélicas para os 10 SNPs encontrados nas regiões estudadas do gene *ATXN3*, nos grupos de casos e de controlos.

SNPs	Frequências	Casos (n)	Controlos (n)	SNPs	Frequências	Casos (n)	Controlos (n)
rs7158733	Genóticas			rs7153193	Genóticas		
	AA	0,18 (4)	0,20 (2)		GG	0,37 (9)	0,60 (6)
	CA	0,41 (9)	0,20 (2)		GT	0,46 (11)	0,20 (2)
	CC	0,41 (9)	0,60 (6)		TT	0,17 (4)	0,20 (2)
	Alélicas				Alélicas		
	A	0,39	0,30		G	0,6	0,70
C	0,61	0,70	T	0,4	0,30		
rs3092822	Genóticas			rs10151135	Genóticas		
	AA	0,35 (8)	0,60 (6)		CC	0,50 (15)	0,90 (9)
	AC	0,43 (10)	0,20 (2)		CA	0,43 (13)	0,10 (1)
	CC	0,22 (5)	0,20 (2)		AA	0,07 (2)	0,00
	Alélicas				Alélicas		
	A	0,56	0,70		C	0,72	0,95
C	0,44	0,30	A	0,28	0,05		
rs7158238	Genóticas			rs709930	Genóticas		
	GG	0,37 (8)	0,60 (6)		GG	0,34 (10)	0,60 (6)
	GA	0,45 (10)	0,20 (2)		GA	0,53 (16)	0,20 (2)
	AA	0,18 (4)	0,20 (2)		AA	0,13 (4)	0,20 (2)
	Alélicas				Alélicas		
	A	0,41	0,30		G	0,77	0,70
G	0,59	0,70	A	0,33	0,30		
rs7153696	Genóticas			rs910369	Genóticas		
	CC	0,41 (7)	0,60 (6)		GG	0,35 (10)	0,60 (6)
	CT	0,35 (6)	0,20 (2)		GT	0,55 (16)	0,20 (2)
	TT	0,24 (4)	0,20 (2)		TT	0,10 (3)	0,20 (2)
	Alélicas				Alélicas		
	C	0,59	0,70		G	0,62	0,70
T	0,41	0,30	T	0,38	0,30		
rs7153374	Genóticas			rs55966267	Genóticas		
	GG	0,41 (9)	0,60 (6)		AA	0,93 (28)	1,00 (10)
	GA	0,41 (9)	0,20 (2)		AG	0,07 (2)	0,00
	AA	0,18 (4)	0,20 (2)		Alélicas		
	Alélicas				A	0,96	1,00
	A	0,61	0,70		G	0,04	0,00
G	0,39	0,30					

Em ambos os grupos estudados, foram observados os 3 genótipos possíveis para cada SNP, à exceção do rs55966267, cujo alelo A está fixado, nos controlos, observando-se, para este grupo apenas um genótipo (AA). Nos doentes, apenas foi possível encontrar 2 genótipos para o referido SNP, não tendo sido observado o genótipo GG.

Após a análise da tabela acima mencionada, constata-se que o alelo mais frequente em cada SNP, para o grupo dos casos, é também o mais frequente para o grupo dos controlos. No que diz respeito aos genótipos, verifica-se que os indivíduos do grupo dos casos tendem a ser, na sua maioria, heterozigóticos, para pelo menos 5 dos SNPs, enquanto os indivíduos do grupo dos controlos são, maioritariamente, homozigóticos para todos os SNPs identificados. Tendo em conta os alelos ancestrais descritos na base de dados do NCBI (<http://www.ncbi.nlm.nih.gov/snp>), para cada SNP, a maioria dos indivíduos deste último grupo estaria em homozigotia para o alelo ancestral em 6 dos SNPs (rs7158733, rs7158238, rs7153374, rs7153193, rs709930 e rs55966267) e para o alelo derivado nos restantes 4 SNPs (rs3092822, rs7153696, rs10151135 e rs910369).

Os testes efectuados para verificar a conformidade com o equilíbrio de HW no grupo dos casos, bem como no grupo dos controlos, produziram valores de  $p$  não significativos, constatando-se, desta forma, que as frequências genóticas para todos os SNPs estão de acordo com a situação de equilíbrio.

Os valores de diversidade genética obtidos para os 10 SNPs estão descritos na Tabela 3. Estes valores variaram entre zero e 0,50242, correspondendo o valor zero ao rs55966267, no grupo dos controlos, que, tal como previamente referido, apenas apresenta o alelo ancestral, neste mesmo grupo de indivíduos. O valor mais elevado de diversidade foi encontrado para o SNP rs3092822 que, no grupo de doentes, produziu valores de heterozigotia da ordem dos 50,24%. Após a comparação de ambos os grupos, para cada SNP, constatou-se que a diversidade genética é, em geral, ligeiramente mais elevada para o grupo de casos, o que está de acordo com o que se constatou

anteriormente, ao verificar-se que os indivíduos deste grupo eram mais frequentemente heterozigóticos. Por outro lado, o facto de o grupo dos controlos apresentar valores de diversidade mais baixos poderá estar relacionado com o reduzido número de indivíduos incluídos neste grupo.

**Tabela 3.** Valores de diversidade genética (DG) para os grupos de casos e de controlos. O número de indivíduos utilizado para o cálculo da DG, para cada um dos 10 SNPs, é idêntico ao apresentado na Tabela 2.

SNPs	DG	
	Casos	Controlos
rs7158733	0,48520	0,44211
rs3092822	0,50242	0,44211
rs7158238	0,49471	0,44211
rs7153696	0,49911	0,44211
rs7153374	0,48520	0,44211
rs7153193	0,48848	0,44211
rs10151135	0,41299	0,10000
rs709930	0,48814	0,44211
rs910369	0,47913	0,44211
rs55966267	0,06554	0,00000

Entre o grupo de casos e o grupo dos controlos, os resultados dos testes exactos de diferenciação, com base nas frequências genotípicas de cada um dos SNPs, separadamente, permitiram confirmar a ausência de diferenças estatisticamente significativas entre estes dois grupos.

A análise à independência alélica entre SNPs foi empreendida, separadamente, para casos e controlos. Os resultados, em termos de

significância estatística e após correcção de Bonferroni, estão evidenciados nas tabelas 4 (casos) e 5 (controlos).

**Tabela 4.** Significância dos resultados do teste de LD para todos os pares de SNPs, no grupo de casos (n =16).

SNPs	Exão 10						Exão 11			
	rs7158733	rs3092822	rs7158238	rs7153696	rs7153374	rs7153193	rs10151135	rs709930	rs910369	rs55966267
rs7158733		+	+	+	+	+	+	+	+	-
rs3092822	+		+	+	+	+	-	+	+	-
rs7158238	+	+		+	+	+	+	+	+	-
rs7153696	+	+	+		+	+	+	+	+	-
rs7153374	+	+	+	+		+	+	+	+	-
rs7153193	+	+	+	+	+		+	+	+	-
rs10151135	+	-	+	+	+	+		-	+	-
rs709930	+	+	+	+	+	+	-		+	-
rs910369	+	+	+	+	+	+	+	+		-
rs55966267	-	-	-	-	-	-	-	-	-	

(+ significativo, após Bonferroni)

**Tabela 5.** Significância dos resultados do teste ao LD para todos os pares de SNPs, no grupo de controlos (n=10).

SNPs	Exão 10						Exão 11			
	rs7158733	rs3092822	rs7158238	rs7153696	rs7153374	rs7153193	rs10151135	rs709930	rs910369	rs55966267
rs7158733		+	+	+	+	+	-	+	+	-
rs3092822	+		+	+	+	+	-	+	+	-
rs7158238	+	+		+	+	+	-	+	+	-
rs7153696	+	+	+		+	+	-	+	+	-
rs7153374	+	+	+	+		+	-	+	+	-
rs7153193	+	+	+	+	+		-	+	+	-
rs10151135	-	-	-	-	-	-		-	-	-
rs709930	+	+	+	+	+	+	-		+	-
rs910369	+	+	+	+	+	+	-	+		-
rs55966267	-	-	-	-	-	-	-	-	-	

(+ significativo, após Bonferroni)

Todos os SNPs localizados no exão 10, quer no grupo de casos, quer no grupo dos controlos, estão em LD, o que seria previsível considerando que estes SNPs têm uma grande proximidade física no gene *ATXN3*. Relativamente aos SNPs localizados no exão 11, verifica-se que em ambos os grupos o rs55966267 não se encontra em LD, o que poderá estar associado ao facto de, no grupo de casos, o alelo A apresentar uma frequência de 96%, sendo que nos controlos esta frequência é de 100%. Este SNP apresenta uma particularidade relativamente aos restantes encontrados, pois está fisicamente mais afastado dos restantes SNPs no exão 11; para além disso, localiza-se após a segunda sequência consenso para o sinal de poliadenilação no referido exão. No grupo dos controlos, para além do SNP anteriormente referido, o rs10151135 também não se encontra em LD; a frequência para o alelo C deste SNP é de 95%, o que estará associado ao resultado obtido para o LD neste grupo. No grupo dos casos o rs10151135 produziu valores de  $p$  consistentes com LD, na maior parte das comparações entre pares de SNPs.

Como referido no início desta secção, as regiões estudadas neste trabalho abrangem quatro sequências consenso para sinais de poliadenilação. A comparação das sequências obtidas com a sequência de referência permitiu constatar que os polimorfismos observados, não implicam a eliminação nem dos sinais de poliadenilação A1, A2 e A3, descritos anteriormente por Ichikawa *et al.* (2001), nem do B1. No entanto, também não originam sequências consenso para novos sinais de poliadenilação. Com base nos resultados obtidos, não será expectável que ocorram alterações a nível da poliadenilação devido aos SNPs presentes na porção 3'UTR do gene *ATXN3* que foi

analisada, não se prevendo portanto que, a este nível, exista uma influência na regulação da expressão deste gene.

## 5. CONCLUSÕES GERAIS

A importância da região 3'UTR é essencialmente atribuída ao papel que a mesma desempenha na regulação da expressão génica. Alterações na sequência destas regiões, tal como referido anteriormente neste trabalho, poderão ter consequências drásticas, nomeadamente no que se refere à ocorrência de doenças genéticas. No caso de muitos genes, como é o caso do gene *ATXN3*, que corresponde ao gene causal da DMJ, a informação existente sobre a variabilidade genética na região 3'UTR, especialmente em doentes, é muito reduzida ou inexistente. Apesar de, na DMJ, a mutação causal estar bem estabelecida, alterações nestas regiões reguladoras poderão, em última instância, contribuir para a heterogeneidade clínica observada nesta doença. Este trabalho pretendeu analisar e descrever pela primeira vez, a variação genética na região 3'UTR do gene *ATXN3*, usando amostras de doentes DMJ. Foi detectado um total de 10 SNPs que já tinham sido previamente descritos. Todos os SNPs anteriormente referidos foram identificados quer no grupo dos casos quer no grupo de controlos, não se tendo identificado variantes exclusivas dos doentes.

Pretendeu-se, também, efectuar comparações entre o grupo de doentes DMJ e o grupo de indivíduos controlo. Após a análise dos resultados obtidos, constatou-se que os alelos mais frequentes, para cada um dos 10 SNPs detectados, eram coincidentes entre os dois grupos estudados. Para além disso, verificou-se que não existiam diferenças significativas relativamente às frequências genotípicas entre estes dois grupos. Os dados obtidos no presente

trabalho apontam no sentido de que a região analisada do gene *ATXN3* não apresenta variação exclusiva nos doentes.

Outro dos objectivos deste trabalho era verificar se a existência de polimorfismos podia interferir com os sinais de poliadenilação. A análise da porção 3'UTR do gene *ATXN3*, estudada neste trabalho, demonstrou que os polimorfismos encontrados não implicam a eliminação dos sinais de poliadenilação anteriormente descritos (A1, B1, A2 e A3) nem a criação de sequências consenso para novos sinais de poliadenilação. Assim, poder-se-á concluir que não deverão existir alterações, causada pelos SNPs encontrados, no que respeita à poliadenilação desta porção da região 3'UTR, pelo que não se prevêem alterações nos níveis de expressão deste gene, que sejam mediadas por esse tipo de regulação. Contudo, é de salientar o facto de a região 3'UTR do gene *ATXN3*, não ter sido estudada na sua totalidade, uma vez que a restante porção do exão 11 que abrange as cinco últimas sequências consenso para sinais de poliadenilação (A4-A8), descritos por Ichikawa *et al.* (2001), não foi analisada, não permitindo desta forma extrapolar a mesma conclusão relativamente à influência dos sinais de poliadenilação para toda a região 3'UTR do gene em causa.

Esta e outras limitações do estudo conduzem a que haja uma série de perspectivas futuras interessantes que, se colocadas em execução, poderiam dar continuidade a este trabalho.

Em termos de perspectivas futuras, poder-se-ia aumentar o número de indivíduos estudados, quer no grupo dos controlos para poder ter uma maior percepção da variação normal, quer no grupo dos doentes, dividindo-os de

acordo com características fenotípicas, por exemplo pela idade de início da doença. No que se refere à região analisada, seria interessante estender a análise à restante região do exão 11 cobrindo toda a extensão do gene *ATXN3* capaz de formar a região 3'UTR. Poder-se-iam também fazer análises *in silico* dos factores que actuam em *trans*, de modo a tentar perceber efeitos adicionais de toda a variação encontrada.

Este trabalho, visando contribuir para uma melhor compreensão das bases moleculares da DMJ, permitiu obter informação sobre a variação existente na região 3'UTR do gene *ATXN3* em amostras de doentes DMJ, sobre a qual, até à data, o conhecimento era praticamente inexistente. Para além disso, os dados aqui gerados, poderão servir de base para trabalhos futuros, nomeadamente sobre variação genética em regiões reguladoras do gene *ATXN3* e a sua relação com a variabilidade fenotípica da DMJ.

## 6. BIBLIOGRAFIA

- BETTENCOURT C, SILVA-FERNANDES A, MONTIEL R, *ET AL.* (2007): Triplet repeats: Features, dynamics and evolutionary mechanisms. *In: Santos C, Lima M (Eds) Recent advances in molecular biology and evolution: applications to biological anthropology.* Research Signpost, Kerala; pp 83-114.
- BETTENCOURT C, SANTOS C, KAY T, VASCONCELOS J, LIMA M (2008a): Analysis of segregation patterns in Machado-Joseph disease pedigrees. *Journal of Human Genetics*; 53: 920-923.
- BETTENCOURT C, FIALHO R, SANTOS C *ET AL.* (2008b): Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: a study in normal families from Azores islands (Portugal). *Journal of Human Genetics*; 53: 333-339.
- BETTENCOURT C (2009): *Machado-Joseph disease: From genetic variability to clinical heterogeneity*, 204 pp. Dissertação de Doutoramento, Universidade dos Açores.
- BETTENCOURT C, SANTOS C, MONTIEL R *ET AL.* (2010a): Increased transcript diversity: novel splicing variants of Machado-Joseph Disease gene (*ATXN3*). *Neurogenetics*; 11: 193–202.
- BETTENCOURT C, SANTOS C, MONTIEL R *ET AL.* (2010b): The (CAG)<sub>n</sub> tract of Machado-Joseph disease gene (*ATXN3*): a comparison between DNA and mRNA in patients and controls. *European Journal of Human Genetics*; 18: 621-623.

- BOLOGNAMI F, PERRONE-BIZZOZERO N (2008): RNA-protein interactions and control of mRNA stability in neurons. *Journal of Neuroscience Research*; 86: 481-489.
- CARVALHO D R, LA ROCQUE-FERREIRA A, RIZZO I M ET AL. (2008): Homozygosity Enhances Severity in Spinocerebellar Ataxia Type 3. *Pediatric Neurology*; 38: 296-299.
- CHATTERJEE S, PAL J K (2009): Role of 5'- and 3'- untranslated regions of mRNAs in human diseases. *Biology of the Cell*; 101: 251-262.
- CHEN J M, FÉREC C, COOPER D (2006a): A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes I: general principles and overview. *Human Genetics*; 120: 1-21.
- CHEN J M, FÉREC C, COOPER D (2006b): A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding functionality of 3'UTR variants. *Human Genetics*; 120: 301-333.
- COOPER D N e KRAWCZAK M (1993): *Human Gene Mutation*, 402 pp. Bios Scientific Publishers Limited Oxford U.K.
- COUTINHO P (1992): *Doença de Machado-Joseph: Tentativa de definição*, 254 pp. Dissertação de Doutoramento, Instituto de Ciências Biomédicas Abel Salazar. Porto.
- COUTINHO P, ANDRADE C (1978): Autosomal dominant system degeneration in Portuguese families of the Azores Islands. A new genetic disorder involving cerebellar, pyramidal, extrapyramidal and spinal cord motor functions: *Neurology*; 28: 703-709. (não consultado)

- COUTINHO P, SEQUEIROS J (1980): Aspects cliniques, génétiques et pathologiques de la maladie de Machado-Joseph. *Journal de Génétique Humaine*; 29(3): 203-209. (não consultado)
- DAY D, TUTE M (1998): Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. *Journal of Endocrinology*; 157: 361-371.
- EDWALDS-GILBERT G, VERALDI K, MILCAREK C (1997): Alternative poly (A) site selection in complex transcription units: means to an end? *Nucleic Acids Research*; 25: 2547-2561.
- EXCOFFIER L G, LAVAL SCHNEIDER S, ARLEQUIN (1998-2010): An Integrated Software Package for Population Genetics Data Analysis. *Evolutionary Bioinformatics Online*, versão 3 5.5 1: 47-50.
- GONZALEZ C, LIMA M, KAY T, ET AL. (2004): Short-Term impact of predictive testing for the Machado-Joseph disease: Depression and Anxiety levels in individuals at risk from the Azores (Portugal). *Community Genetics*; 7(4): 196-201.
- GRZYBOWSKA E, WILCZYNSKA A, SIEDLECKI J A (2001): Breakthroughs and views Regulatory Functions of 3'UTRs. *Biochemical and Biophysical Research Communications*; 288: 291-295.
- Guo S e Thompson E (1992): Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*; 48: 361-372.
- HALL TA (1999): BIOEDIT: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98NT. *Nucleic Acids Symposium Series*: 41: 95-98.
- HUGHES T (2006): Regulation of gene expression by alternative untranslated regions. *Trends in Genetics*; 22: 119-122.

- ICHIKAWA Y, GOTO J, HATTORI M *ET AL.* (2001): The genomic structure and expression of MJD, the Machado- disease gene. *Journal of Human Genetics*; 46: 413-422.
- KAWAGUCHI Y, OKAMOTO T, TANIWAKI M *ET AL.* (1994): CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nature Genetics*; 8: 221-228.
- LIMA M (1996): *Doença de Machado-Joseph nos Açores: Estudo epidemiológico, biodemográfico e genético*, 299pp. Dissertação de Doutoramento Universidade dos Açores.
- LIMA M, MAYER M, COUTINHO P *ET AL.* (1997): Prevalence, geographical distribution and genealogical investigation of Machado-Joseph disease in the islands of Azores (Portugal). *Human Biology*, 69(3): 383-391.
- LIMA M, KAY T, VASCONCELOS J *ET AL.* (2001): Disease Knowledge and attitudes toward Predictive Testing and Prenatal Diagnosis in families with Machado-Joseph disease from the Azores islands (Portugal). *Community Genetics*; 4(1):36-42.
- LIMA M, COSTA M C, MONTIEL R *ET AL.* (2005): Population genetics of wild-type CAG repeats in the Machado-Joseph disease gene in Portugal. *Human Heredity*, 60: 156-163.
- LIMA M (2010): *Excepções ao Paradigma Mendeliano: A doença de Machado-Joseph como caso de estudo*, 39 pp. Provas de Agregação, Plano de Lição. Universidade dos Açores.
- LUTZ C (2008): Alternative polyadenylation: A twist on mRNA 3' end formation. *Acs Chemical Biology*; 3: 609-617.

- MACIEL P, COSTA M C, FERRO A *ET AL.* (2001): Improvement in the molecular diagnosis of Machado-Joseph disease. *Archives of Neurology*; 58: 1821-7.
- MANDEL C R, BAI Y, TONG L (2008): Protein factors in pre-mRNA 3'-end processing. *Cellular and Molecular Life Sciences*; 65: 1099-1122.
- NAKANO K, DAWSON D M, SPENCE A (1972): Machado disease-A hereditary ataxia in Portuguese emigrants to Massachusetts. *Neurology*; 22: 49-55. (não consultado)
- NEI M (1987): *Molecular Evolutionary Genetics*. Columbia University Press, New York, NY, USA.
- PAULSON H L, DAS S S, CRINO P B *ET AL.* (1997): Machado-Joseph disease gene product is a cytoplasmatic protein widely expressed in brain. *Annals of Neurology*; 41 (4): 453-62.
- PAULSON H L (2007): Dominantly inherited ataxias: lessons learned from Machado-Joseph disease/ spinocerebellar ataxia type 3. *Seminars in Neurology*; 27: 133-142.
- REGATEIRO F (2003): *Genética Médica*, 496 pp. Imprensa da Universidade de Coimbra. Coimbra.
- ROSENBERG R, NYHAN W L, BAY C, SHORE P (1976): Autosomal dominant striatonigral degeneration. A clinical, pathologic and biochemical study of a new genetic disorder. *Neurology*; 26: 703-714. (não consultado)
- SUITE ND, SEQUEIROS J, MCKHANN GM (1986): Machado-Joseph disease in a Sicilian-American family. *J Neurogenet*; 3: 177-182. (não consultado)
- STRACHAN T e READ A P (2004): *Human Molecular Genetics* 3ª Ed., 674 pp. Garland Publishing New York, USA.

TAKIYAMA Y, NISHIZAWA M, TANAKA H *ET AL.* (1993): The gene for Machado-Joseph disease maps to human chromosome 14q. *Nature Genetics*; 4: 300-304.

TAKIYAMA Y, Igarashi S, Rogaeva E A *ET AL.* (1995): Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes of flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. *Human Molecular Genetics*; 4 (7): 1137-46.

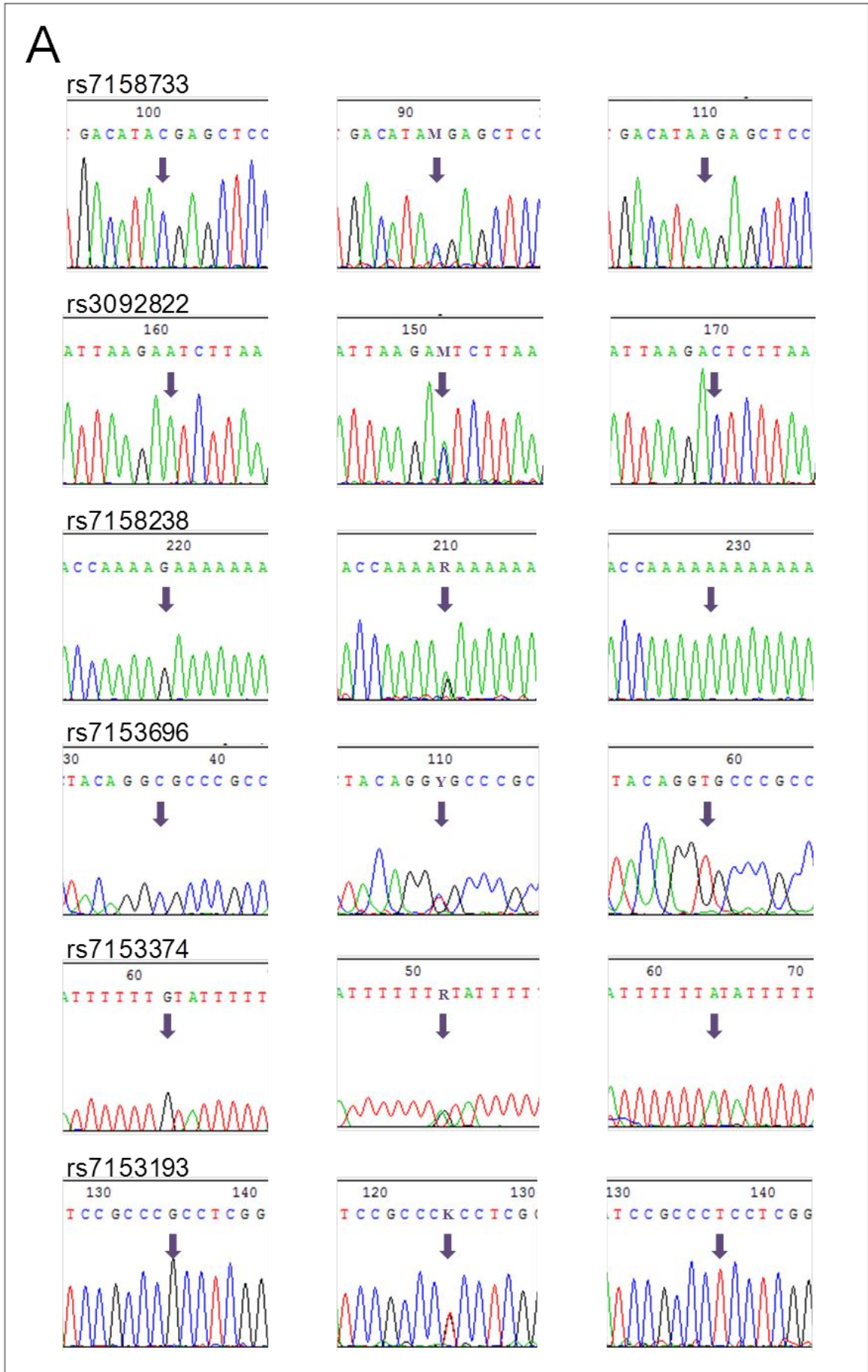
WOODS B T, SCHAUMBURG H H (1972): Nigro-spino-dentatal degeneration with nuclear ophthalmoplegia: A unique and partially treatable clinico-pathological entity. *Journal of the Neurological Sciences*; 17: 149-166. (não consultado)

ZLOTOGORA J 2006: Genetic diseases: Prevalence. *Encyclopedia of Life Sciences*. (DOI:101038/npg/els.0006004).

## **7. Anexos**

**Tabela 1A.** Representação dos genótipos para cada um dos SNP por indivíduo, para o grupo dos casos e dos controlos.

	Código Amostra	Exão 10						Exão 11			
		rs7158733	rs3092822	rs7158238	rs7153696	rs7153374	rs7153193	rs10151135	rs709930	rs910369	rs55966267
<b>DOENTES</b>	DMJ 1	CA	AC	GA	CT	GA	GT	CA	GA	GT	AA
	DMJ 2	CA	AC	GA	CT	GA	GT	CC	GA	GT	AA
	DMJ 6	CC	AA	GG	CC	GG	GG	AA	GG	GG	AA
	DMJ 10	CC	AA	GG	CC	GG	GG	CA	GG	GG	AA
	DMJ 14	AA	CC	AA	TT	AA	TT	CC	AA	TT	AA
	DMJ 17	CA	-	-	-	GA	GT	CA	GA	GT	AA
	DMJ 18	-	-	-	-	-	-	CC	GA	GT	AA
	DMJ 19	CA	AC	GA	-	GA	GT	CC	GA	GT	AA
	DMJ 20	-	-	-	-	-	-	CC	GA	GT	AA
	DMJ 21	AA	CC	AA	TT	AA	TT	CC	AA	TT	AA
	DMJ 22	CA	AC	GA	-	-	GT	CC	GA	GT	AG
	DMJ 30	CC	AA	GG	CC	GG	GG	CA	GG	GG	AA
	DMJ 32	CC	AA	GG	-	GG	GG	CA	GG	GG	AA
	DMJ 33	CC	AA	-	-	GG	GG	CA	GG	GG	AA
	DMJ 35	CC	AA	GG	CC	GG	GG	AA	GG	GG	AA
	DMJ 36	CC	AA	GG	CC	GG	GG	CA	GG	GG	AA
	DMJ 39	CC	AA	GG	CC	GG	GG	CA	GG	GG	AA
	DMJ 40	CC	CC	GG	CC	GG	GG	CA	GG	GG	AA
	DMJ 41	-	AC	GA	-	-	GT	CC	GA	GT	AA
	DMJ 42	CA	AC	GA	CT	GA	GT	CA	GA	GT	AA
	DMJ 43	CA	AC	GA	CT	GA	GT	CC	GA	GT	AG
	DMJ 45	AA	CC	AA	TT	AA	TT	CC	AA	TT	AA
	DMJ 46	AA	CC	AA	TT	AA	TT	CC	AA	-	AA
	DMJ 47	-	AC	GA	-	GA	GT	CC	GA	GT	AA
	DMJ 48	CA	AC	GA	CT	GA	GT	CC	GA	GT	AA
	DMJ 49	CA	AC	GA	CT	GA	GT	CC	GA	GT	AA
	DMJ 50	-	-	-	-	-	-	CC	GA	GT	AA
	DMJ 51	-	-	-	-	-	-	CA	GA	GT	AA
	DMJ 57	-	-	-	-	-	-	CA	GA	GT	AA
DMJ 59	-	-	-	-	-	-	CA	GG	GG	AA	
<b>CONTROLOS</b>	ContDMJ 9	CC	AA	GG	CC	GG	GG	CC	GG	GG	AA
	ContDMJ 10	CC	AA	GG	CC	GG	GG	CC	GG	GG	AA
	ContDMJ 20	CC	AA	GG	CC	GG	GG	CC	GG	GG	AA
	ContDMJ 21	CA	AC	GA	CT	GA	GT	CC	GA	GT	AA
	ContDMJ 22	CC	AA	GG	CC	GG	GG	CC	GG	GG	AA
	ContDMJ 23	AA	CC	AA	TT	AA	TT	CC	AA	TT	AA
	ContDMJ 30	CC	AA	GG	CC	GG	GG	CC	GG	GG	AA
	ContDMJ 31	CC	AA	GG	CC	GG	GG	CC	GG	GG	AA
	ContDMJ 32	CA	AC	GA	CT	GA	GT	CA	GA	GT	AA
	FLO12	AA	CC	AA	TT	AA	TT	CC	AA	TT	AA



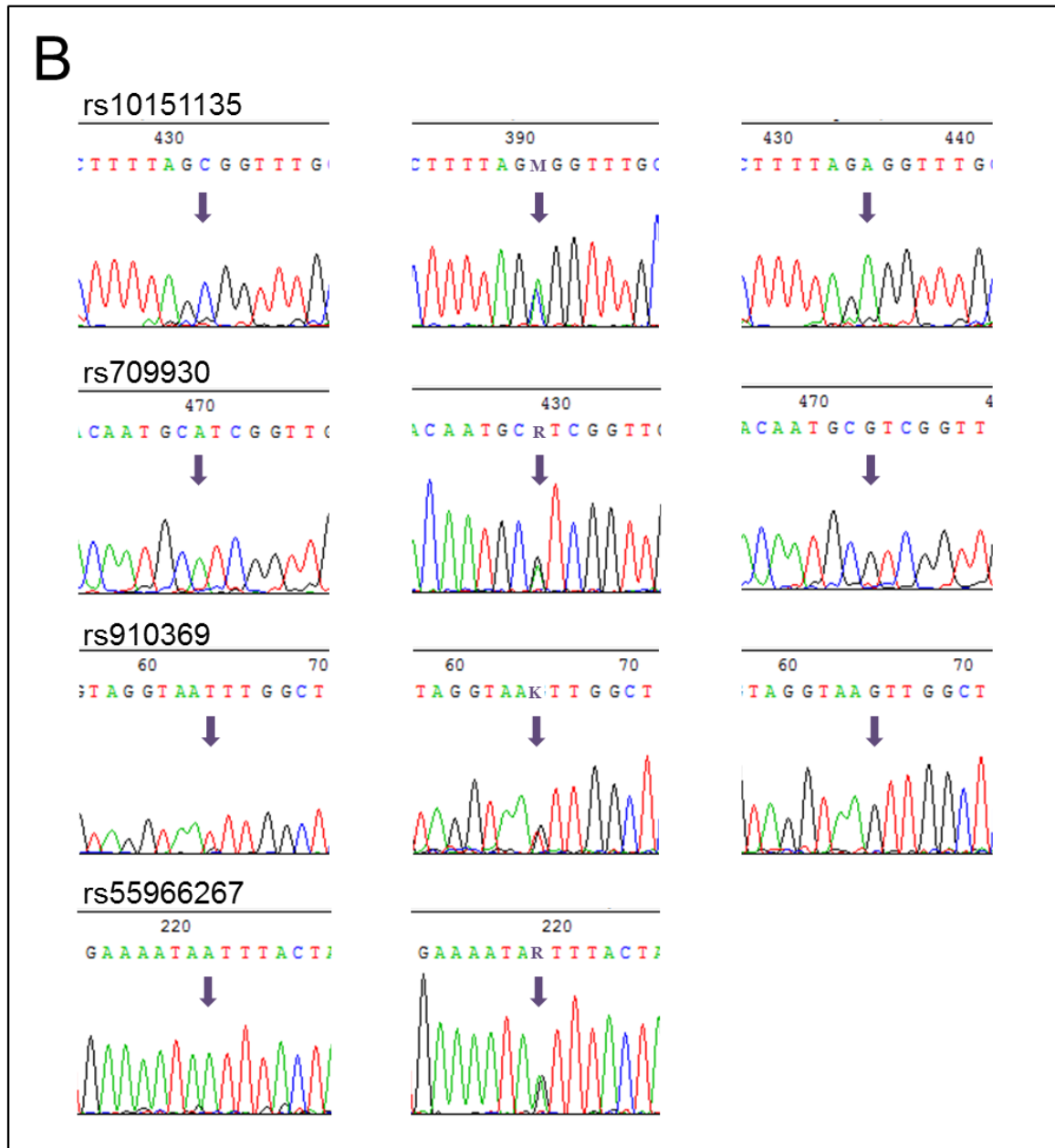


Figura 1A. Electroferogramas exemplificativos de cada um dos genótipos observados, para cada um dos 10 SNPs identificados nos exões 10 e 11 do gene *ATXN3*. A) SNPs encontrados no exão 10 (rs7158733, rs3092822, rs7158238, rs7153696, rs7153374, rs7153193); B) SNPs encontrados no exão 11 (rs10151135, rs709930, rs910369, rs55966267). A localização de cada SNP está destacada com uma seta. Nos electroferogramas correspondentes à situação de heterozigotia, o nucleótido correspondente ao SNP foi representado pelo código para bases degeneradas, assim: K=G/T, M=C/A, R=G/A, Y=C/T.