MAFALDA SOFIA BASTOS RAPOSO

Predicting and tracking Machado-Joseph disease: biomarkers of diagnosis and prognosis

UNIVERSIDADE DOS AÇORES

FACULDADE DE CIÊNCIAS E TECNOLOGIA

PONTA DELGADA

2016

MAFALDA SOFIA BASTOS RAPOSO

Predicting and tracking Machado-Joseph disease: biomarkers of diagnosis and prognosis

Dissertation for PhD degree in Biology presented to the University of the Azores

Supervisor: Maria Manuela de Medeiros Lima

Co-supervisor: Maria da Conceição Félix Bettencourt

UNIVERSIDADE DOS AÇORES

FACULDADE DE CIÊNCIAS E TECNOLOGIA

PONTA DELGADA

2016


União Europeia
Fundo Social Europeu


PROEMPREGO


Governo dos Açores


AÇORES
2020
PROGRAMA OPERACIONAL
FEDER FSE


FRC
FUNDO REGIONAL PARA A CIÊNCIA

ACKNOWLEDGMENTS

The development of the present dissertation has been a great adventure, both personally and professionally, and I must express my eternal gratitude to all the people who shared with me this time:

• To my supervisors. To Professor Manuela Lima, which has been always present, for her unconditional support, for her ability of teaching me, for allowing me to grow up (after hitting often with my head on the walls, which, usually, were in the format of reviewers), by our scientific conversations (often mixed with talks less scientific) and her way of living science, being an example to follow (my example). I'm afraid that these words not thank you enough. To Dr. Conceição Bettencourt, who despite the distance has always been there when needed, by their exclamation points (usually above 5 was not a good sign), for your friendship and because we share the same birthday day (which itself says a lot of things).

• To the members of GAIN: Prof. Manuela, Dr. Carlos, Dr. Teresa and Dr. João. Our memorable trips to the islands, the team spirit and the capacity of joint efforts for a unique benefit (the MJD patients). The biggest laughs of this journey were undoubtedly shared with this group, which in turn made most of the shared moments, unforgettable. A special word to Dr. João, who allowed me to share with him, all these years, the clinical evaluation of the MJD patients, and that allowed me to live, closely, the reality of MJD. I must also mention Mariazinha, who received me with open arms and was always there when needed (a lot of times).

• To the MJD patients and their families, for their cooperation and availability. Often I shared their suffering, but, also, their hope.

• To my family. To my uncles, cousins, godmother, Andreia and Sofia that despite the distance we have always been together and support me unconditionally. My godson, who was born in the

same month and year of the beginning of this PhD project, and I hope that one day, he will be proud of me.

• To my friends, from Moncorvo and São Miguel. To my friends from Moncorvo for being my friends ever and forever (Dininha, Marisa, Tânia, Paula, Sandra, the Leonardo brothers, Julinho, Nuno, Zé Tó). To my friends from São Miguel, which are, currently, my support, who put me up all these years, and helping me to avoid missing home. To Zulmira, for her unconditional friendship, and to the other members of the Teves family (my Azorean family that received me with open arms), to Duarte and Ana (I will never be able to thank you enough for your friendship), to my housemates (especially to Claudia, for sharing with me many evenings), to Rita for her endless friendship, to my "lunch-box, trails and anything else UAc group" (Elizabete, Marta, Ana, Zita, Claudia, Ana Rosa).

• To the past and present members of the Human Genetics group. To Cristina (by skype and emails), Amanda (my favourite Catalan) and Nadiya for the helpful brainstormings during these years. To Pedro Mantua (the only man in the group, God help him), for given me laboratorial support and for listened my "thoughts" in the second floor.

• To Filipe, for all the moments we share, for being the other half. Of all the good things that sharing live brings love is the greatest of all.

• To my mammy for all the love and unconditional support, for being the best mother, for always being there, for being an example of life, for everything.

TABLE OF CONTENTS

MOLECULAR STATE BIOMARKERS

ABSTRACT

Machado-Joseph disease/spinocerebellar ataxia type 3 (MJD/SCA3) is a rare late-onset neurodegenerative disorder, which is currently untreatable. In this polyglutamine (polyQ) disease, a CAG repeat expansion located at the *ATXN3* gene, is the causative mutation. As a progressive disorder, MJD is characterized by different stages throughout its natural course, namely by a long preclinical period, during which molecular and even clinical alterations could be starting, many years before the ataxia onset. MJD onset is partially explained by the size of the CAG tract constituting its main trait biomarker. The incompleteness of this explanation, however, implies the existence of genetic modifiers, which should provide additional trait biomarkers. The identification of molecular trait biomarkers of MJD will allow to: (1) improve the prediction of age at onset and (2) empower interventional trials, by allowing genetic stratification of patients. The characterization of MJD disease stages has so far been performed using mainly clinical scales or imaging data, while few studies investigated molecular biomarkers. The identification of such molecular state biomarkers would allow: (1) the prediction of age at pathological onset (age at which cell damage is already present although a clinical phenotype is not yet evident), (2) monitoring different stages of disease progression, and (3) detecting subtle therapeutic benefits, in emergent clinical trials. The main goal of the present PhD project was to contribute to the development of molecular biomarkers for MJD. RNA and DNA, obtained from blood samples of preataxic subjects and MJD patients of Azorean background, including samples collected in two different moments, constituted the main resource used throughout the present PhD project. Samples from population controls were also used. The determination of the number of CAG repeats at the *ATXN3* was performed for all MJD subjects studied. The homogeneous MJD cluster comprising a total of 88 patients was characterized. The average of the neurological examination score for spinocerebellar ataxia (NESSCA) available for 69 patients was 12±5 (mean ± standard deviation); 40% of the variance explanation is provided by the number of the CAG repeats in the expanded allele and disease duration. Ocular alterations, namely nystagmus, are clinical features commonly described in MJD. The evaluation of nystagmus in the early stages of the disease was undertaken using asymptomatic carriers and non-carriers of *ATXN3* mutation. The frequency of nystagmus in asymptomatic carriers (17%) and its absence in non-carriers of the mutation suggests that nystagmus may appear before gait disturbance and can thus be considered an early sign of MJD. A candidate gene approach was used in this work to pursuit

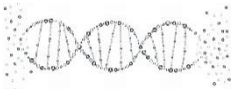
modifiers of the MJD onset. The CAG repeat number in loci associated with polyQ disorders, namely spinocerebellar ataxias (SCA) types 1, 2, 6, 7, 17, Huntington disease (HD), and dentato-rubro-pallydoluysian atrophy (DRPLA) was investigated as modulators of the age at onset in the Azorean MJD cohort; a negative correlation was found between the longer allele at *ATXN1* and disease onset, significantly increasing by 1.5% the explanation of onset variance. Furthermore, and because a role of inflammation in MJD has been postulated, promoter variants leading to alterations in cytokines expression were studied, and their impact in the disease onset was analysed. Patients carrying the IL6*C allele had a significant earlier onset. A complex net of mechanisms, which includes transcriptional regulation, ubiquitin-proteasome system, autophagy, apoptosis as well as mitochondrial function, has been implicated in MJD pathogenesis. Results from the transcriptome-wide gene expression analyses, followed by several steps of validation, showed that even in a peripheral tissue, transcription dysregulation was detectable and that *FCGR3B, P2RY13 and SELPLG* were dysregulated in patients. To further investigate dysregulated mRNA levels as a way to identify candidate biomarkers, the expression patterns of nine candidate genes - *HSPB1*, *DNAJB1*, *DNAJB12*, *DNAJB14*, *BAX*, *BCL2*, *SOD2*, *IL1B* and *IL6* genes were quantified. Decreased levels of *IL6* and *BCL2* mRNA were observed in preataxic subjects. Lower *HSPB1* and *BCL2* mRNA levels after the ataxia onset were, furthermore, evidenced. A significant decrease of *BCL2* mRNA levels during disease progression was observed in all patients used in the exploratory longitudinal study. Mitochondrial DNA (mtDNA) damage and depletion has been inconsistently reported in MJD. In this work the mtDNA common deletion was significantly more frequent in patients and preataxic carriers than in controls. In conclusion, although the clinical and physiological phenotype of MJD reflects mainly neuronal damage, novel peripheral molecular alterations were described in the present dissertation. The confirmation of the link between peripheral abnormalities and neurodegeneration offers a new window of opportunity for biomarker discovery. More steps of validation will be crucial to develop a battery of molecular biomarkers able to describe disease progression, as well as to be used as outcome measures in future preventive and/or disease-modifying clinical trials. Data from molecular biomarkers should also be further correlated with imaging and clinical measures. Efforts to test the promising molecular biomarkers in independent cohorts, as well as to investigate their potential in longitudinal studies are warranted.

RESUMO

A doença de Machado-Joseph/ataxia espinocerebelosa do tipo 3 (DMJ/SCA3) é uma doença neurodegenerativa rara e de início tardio, que permanece sem tratamento. À semelhança do que acontece nas restantes doenças de poliglutamina (poliQ), a mutação na base da DMJ corresponde à expansão de um motivo CAG na região codificante do respetivo gene causal, *ATXN3*. Sendo uma doença progressiva, a DMJ é caracterizada por várias fases, ao longo da sua progressão, nomeadamente por um longo período pré-clínico; durante esse período deverão ocorrer alterações moleculares e até mesmo clínicas, as quais podem surgir muitos anos antes da ataxia. A idade de inicio da ataxia, usualmente o primeiro sintoma da doença, é maioritariamente explicada pelo tamanho do tracto CAG que constitui assim o principal biomarcador de diagnóstico molecular da DMJ. Genes modificadores deverão, adicionalmente, contribuir para a explicação da idade de início da doença; tais modificadores genéticos, podem também ser considerados biomarcadores de diagnóstico da DMJ. A identificação deste tipo de biomarcadores permitirá: (1) melhorar a previsão da idade de início da doença e (2) aumentar o poder estatístico em ensaios clínicos, permitindo a estratificação dos doentes usando informação proveniente dos modificadores genéticos. As diferentes fases da DMJ são atualmente caracterizadas recorrendo à utilização de informação clínica (principalmente de escalas clínicas) ou de dados obtidos a partir de estudos de imagem, sendo o conhecimento acerca das alterações moleculares que deverão ocorrer ao longo da doença muito incipiente, o que se reflete na ausência de biomarcadores de monitorização da doença. A identificação deste tipo de biomarcadores permitirá: (1) a previsão da idade de início patológica (idade à qual alterações ao nível celular já estão presentes, mas o doente ainda não apresenta sintomas), (2) a monitorização das diferentes fases da progressão da doença e (3) a detecção de benefícios terapêuticos subtis, em futuros ensaios clínicos. O objetivo principal desta dissertação foi contribuir para o desenvolvimento de biomarcadores moleculares da DMJ. RNA e DNA, obtidos a partir de amostras de sangue de indivíduos pré-clínicos (portadores da mutação *ATXN3,* mas sem sintomatologia) e doentes DMJ de origem açoriana, incluindo amostras colhidas em dois momentos diferentes, constituíram o principal recurso para o desenvolvimento deste projeto de doutoramento. Foram também utilizadas amostras de indivíduos saudáveis da mesma população, que foram usados como controlo. O número de repetições CAG no gene *ATXN3* foi determinado para todos os indivíduos DMJ. A coorte DMJ açoriana, constituída por um total de 88 doentes, foi caracterizada. Para 69 dos doentes

avaliados, utilizando a escala clinica NESSCA (N*eurological Examination Score for Spinocerebellar Ataxia*), a média da pontuação foi de 12 ± 5 (média ± desvio padrão); 40% da variância desta pontuação é explicada pelo número de repetições CAG no alelo expandido e pela duração da doença. As alterações oculares, nomeadamente a presença de nistagmo, fazem parte do conjunto de características clínicas habitualmente descritas na DMJ. A avaliação do nistagmo em fases iniciais da doença, nomeadamente na fase pré-clínica, foi realizada em portadores assintomáticos e não-portadores da mutação *ATXN3*. A frequência do nistagmo em portadores assintomáticos (17%) e sua ausência em não-portadores da mutação sugere que o nistagmo pode preceder a ataxia e, portanto, ser considerado um sinal precoce da DMJ. Nesta dissertação, foi utilizada uma abordagem de genes candidatos para identificar modificadores da idade de início da DMJ. O número de repetições CAG em *loci* de doenças polyQ, nomeadamente as ataxias espinocerebelosas (SCA) dos tipos 1, 2, 6, 7, 17, a doença de Huntington (HD) e a atrofia dentato-rubro-pálido-luisiana (DRPLA) foi investigado como potencial modificador da idade de início em doentes açorianos; observou-se uma correlação negativa entre o alelo normal maior no gene *ATXN1* e a idade de início da doença, sendo que a presença deste alelo aumentou significativamente em 1,5% a explicação da variância da idade de início. Adicionalmente, e por ter sido descrito na DMJ um papel da inflamação, foram estudadas, como modificadores da idade de inicio, variantes alélicas na região promotora de genes de citocinas, uma vez que estas variações podem alterar os seus níveis de expressão. Os doentes que possuíam o alelo IL6*C apresentaram uma idade de início significativamente mais precoce. Um conjunto complexo de mecanismos, que inclui a regulação da transcrição, o sistema ubiquitina-proteassoma, a autofagia, a apoptose e a função mitocondrial, tem sido implicado na patogénese da DMJ. Os resultados de uma análise de expressão *whole-genome*, seguida de vários passos de validação, mostraram que mesmo num tecido periférico, a desregulação da transcrição estava alterada e que os genes *FCGR3B*, *P2RY13* e *SELPLG* estavam desregulados em doentes. Utilizando a desregulação dos níveis de mRNA como argumento para identificar novos candidatos a biomarcadores, foram analisados os padrões de expressão de nove genes - *HSPB1, DNAJB1, DNAJB12, DNAJB14, BAX, BCL2, SOD2, IL1B* e *IL6*. Observou-se que em indivíduos pré-clínicos os níveis de mRNA dos genes *IL6* e *BCL2* eram mais baixos relativamente aos controlos. Em indivíduos já com sintomatologia, foram observados níveis mais baixos de mRNA dos genes *HSPB1* e *BCL2,* comparativamente aos controlos. No estudo exploratório longitudinal foi observada uma diminuição significativa dos níveis de mRNA do gene *BCL2* durante a progressão da doença. Resultados inconsistentes sobre o dano e a deplecção

do DNA mitocondrial (mtDNA) em doentes DMJ têm sido reportados. Neste trabalho, verificou-se que os doentes e os portadores pré-clínicos apresentavam mais deleções do mtDNA (*common deletion*) do que os controlos. Embora as características clinicas e fisiológicas inerentes à DMJ reflitam principalmente alterações do processo neurodegenerativo, alterações moleculares quantificadas em sangue foram descritas nesta dissertação. A confirmação de uma relação entre as alterações periféricas e a neurodegeneração constitui uma nova oportunidade para o desenvolvimento de biomarcadores. Passos adicionais de validação serão cruciais para desenvolver um conjunto de biomarcadores moleculares capazes de monitorizar a progressão da doença, bem como de detectar alterações terapêuticas subtis em futuros ensaios clínicos, incluindo ensaios preventivos. Os biomarcadores moleculares devem também ser capazes de se correlacionar com marcadores clínicos e de imagiologia. Todos os esforços serão realizados no sentido de testar os biomarcadores moleculares aqui propostos em coortes independentes, bem como de investigar o seu potencial em estudos longitudinais.



INTRODUCTION

Worldwide, three in each 100,000 inhabitants are affected by an autosomal dominant cerebellar ataxia (ADCA; [1]), which collectively constitute a group of progressive neurodegenerative disorders. Within this group, forty subtypes of spinocerebellar ataxias (SCAs) are currently known, although 12 of the respective disease-causing genes remain to be identified (OMIM, https://www.ncbi.nlm.nih.gov/books/NBK1138/). Cerebellar ataxia, resulting mainly in unsteady gait, is often the most common and initial symptom of SCAs; other neurological manifestations, namely pyramidal or extrapyramidal signs, ocular alterations and cognitive impairment have also been reported (reviewed in [2]). The first signs of SCAs disease are usually reported during the third or fourth decade of life. The most frequently affected systems are the cerebellum and brainstem, although other structures are also frequently damaged (reviewed in [3]). SCAs can either be caused by unstable repeats expansions (in coding and non-coding regions of the respective genes) or by conventional mutations. The most frequent and studied type of ADCAs is caused by glutamine-encoding repeats, constituting the polyglutamine (polyQ) SCAs. Similarities between SCA 1, 2, 3, 6, 7 and 17 include, amongst others (1) the nature of the causative mutation - a CAG repeat within the coding region of the respective gene; (2) the negative correlation between age at onset and the number of CAG repeats in the expanded allele; and (3) the formation of intracellular aggregates and inclusions (reviewed in [4]). On the other hand, differences between polyQ SCAs are determined by (1) the vulnerability of specific populations of brain cells in each disease; and (2) the inherent function of the disease-causing protein; both facts result a substantial range of phenotypes.

••• Machado-Joseph disease (MJD)/Spinocerebellar ataxia type 3 (SCA3): an overview •••

Machado-Joseph disease (MJD), also known as spinocerebellar ataxia type 3 (SCA3; MIM#109150; ORPHA#98757) belongs to the group of polyQ disorders and its designation relates to the order by which this SCA locus was discovered. The term of Machado-Joseph disease is also widely used, representing the family names of the two Azorean ancestral families with extreme phenotypes,

which were firstly described in United States of America in 1970s (reviewed in [5]). MJD has also been designated by several other terms, namely Machado disease, Joseph disease, nigro-spinodentatal degeneration with nuclear ophthalmoplegia, autosomal dominant striatonigral degeneration and Azorean disease of the nervous system (reviewed in [5]), amongst others.

Epidemiology

Globally, MJD is a rare disease showing a prevalence of 1.5 per 100,000 individuals (http://www.orpha.net/orphacom/cahiers/docs/GB/Prevalence_of_rare_diseases_by_alphabetical_list.pdf). In Portugal, as well as in the majority of the countries for which the disease was described, MJD is the most frequent SCA subtype, representing 56% of all cases of ADCA [1]. In mainland Portugal, the prevalence reaches 3.1 per 100,000 individuals [1], being more than ten times higher in the Azores archipelago (39 per 100,000 individuals according to [6]). In the Azores , MJD families are originally from the islands of Flores, São Miguel, Terceira and Graciosa; the highest worldwide prevalence of MJD occurs in Flores island, where 1 in each 158 individuals is affected [6]. Currently, MJD patients are living in seven of the nine islands of the Azores archipelago (Figure 1). A tendency for the increase in prevalence during the last 35 years has been reported by Araujo and colleagues [6].

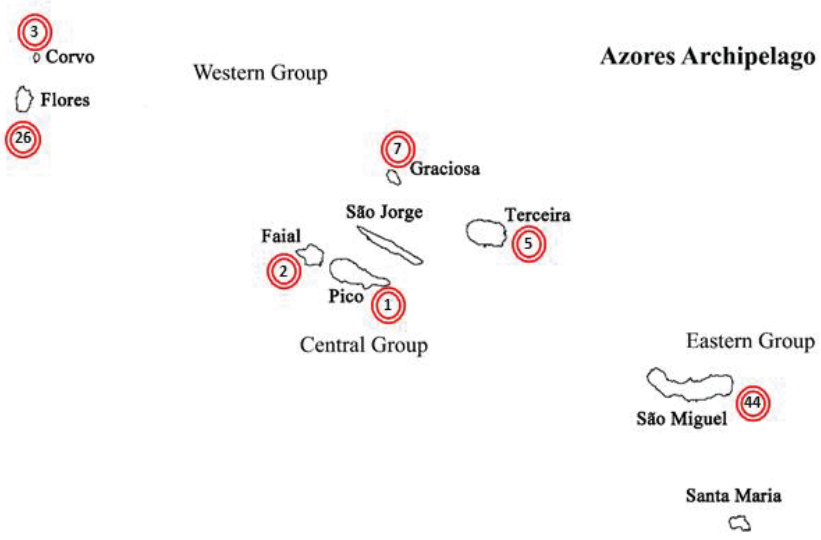


Figure 1. Geographical distribution of the 88 MJD patients alive and living in the Azores archipelago by the end of September 2016 (see Chapter I).

Clinical features

MJD is characterized by a high degree of pleomorphism, which is mainly explained by the degeneration of several central nervous systems (CNS); the cerebellar, pyramidal, extrapyramidal, motor neuron and oculomotor are the systems predominantly affected [7]. Gait ataxia, dysarthria, nystagmus, ophthalmoparesis are often observed; less frequent findings, namely dystonia and bulging eyes as well as occasionally blepharospasm are also described [7]. Some exploratory studies revealed correlations between neuroimaging data and cognitive as well as psychiatric dysfunction, although studies with a large number of patients are needed to validate these findings [8–11]. The clinical variability led to the classification of the disease into different clinical types, namely types I, II, III, IV and V, mainly distinguishable by differences in age at onset and clinical presentation or specific features such as parkinsonism and spastic paraplegia without cerebellar ataxia (reviewed in [5]). Usually the first symptom reported is gait ataxia, but diplopia is also described although less frequently [12]. Several studies have evidenced the presence of other clinical features preceding gait disturbances [13,14], which could compromise the definition of age at onset usually accepted (described as the age of appearance of gait ataxia). In European MJD cohorts, the mean age at onset, ranged between 38 [15] and 44 years of age [16], although extremes between 10 and 78 years have been reported [16]. The earliest reported onset belongs to a homozygous Brazilian girl, whose homozygosity enhanced the severity of the clinical phenotype [17]. Mean survival time after onset (ranging from 7 to 29 years) was estimated as 21 years in Portuguese as well as Brazilian patients [12,18].

The most frequent brain imaging-associated abnormality is the atrophy of the cerebellum and brainstem [19]. The degree of brain atrophy, assessed by magnetic resonance imaging (MRI), is correlated with clinical heterogeneity; more atrophy implies more symptomatology variability [20–22]. Evidences for the involvement of the sensory nerves as well as the motor neurons were found by nerve conduction velocity studies [23]. Degeneration is usually observed in the pons,

substantia nigra, thalamus, anterior horn cells and Clarke's column in the spinal cord, vestibular nucleus, many cranial motor nuclei, and other brainstem nuclei [24–26].

Molecular genetics and genotype-phenotype correlations

An unstable CAG repeat expansion localized at the *ATXN3* gene, which was mapped to chromosome 14 (14q32), is the mutation underlying MJD [27,28]. The number of CAG repeats in the normal alleles range from 13 to 41 CAGs repeats [29]. Consensually, alleles with more than 52 CAG repeats are related with MJD phenotype. In European cohorts, the range of mutated alleles varies from 47 to 79 CAG repeats [16]. An Indian subject with 45 CAG repeats was diagnosed with SCA3 [30]; for instance, an allele with 51 CAG repeats found in a member of a MJD Portuguese family seems to be non-pathogenic [31]. Alleles ranging from 45 to 51 CAG repeats are, therefore, intermediate alleles, which may or not be related with the MJD phenotype(s) and, currently, their behaviour is poorly understood. A tendency of the expanded alleles to further increase in size nearly in half of successive generations, particularly more evident upon male transmission, has been described [32–37]. Intergenerational instability provides, therefore, a molecular explanation to anticipation, a phenomenon reflecting the exhibition of earlier age at onset and/or more severe phenotypes in successive generations (reviewed in [38]). Instability of the CAG number in expanded alleles is also observed during mitosis, which implies changes in the number of CAG repeats in different cell types, a mechanism also known as somatic mosaicism [39,40].

The range of CAG repeats seems to be population-specific; in a Japanese cohort, described by Tezenas du Montcell and colleagues, a mutated allele with 86 CAG repeats was found [16]; on the other hand, in the Azorean cohort, the largest mutated allele described contained 79 CAG repeats [15]. Recently, in a Brazilian cohort, a mutated allele with 91 CAG repeats was described [37].

In MJD, a negative correlation between the size of the expanded allele and the age at onset has been widely described, the value for the Azorean cohort being of $r=-0.826$ ($p<0.001$; [15]). The explanation of the age at onset variance provided by the CAG length in expanded allele, can vary

from 50% (European cohort) to 68% (Azorean cohort) [15,41], depending of the series of patients studied. The lack of a complete explanation of the variability of the age at onset by the *ATXN3* mutation implies the existence of other non-CAG factors contributing to the phenotype, namely modifier genes (discussed with more detailed in chapter III and IV). Correlations between a higher number of CAG repeats in the expanded allele and some clinical features have been reported; pyramidal and dystonic signs as well as spasticity and hyperreflexia have been reported as more frequent in patients carrying a larger CAG tract [42,43].

The identification of the *ATXN3* mutation enabled the development of a molecular test to confirm the carrier/non-carrier status. Besides molecular diagnosis, predictive testing (PT) has been also available for at-risk MJD members, allowing the identification of mutation carriers even before the appearance of clinical manifestations. Prenatal Diagnosis (PND) as well as Preimplantation Genetic Diagnosis (PGD) are also currently available (reviewed in [5]).

Origins of MJD mutation

Originally, MJD was described as a Portuguese/Azorean entity; one of the several designations of the disease - Azorean disease of the nervous systems - was related with this hypothesis, which was further dismissed by more recent studies. The global geographic origin of MJD mutation was initially determined by the analysis of three intragenic single nucleotide polymorphisms (SNPs; $\underline{A}^{669}$TG/$\underline{G}^{669}$TG, $\underline{C}^{987}$GG/$\underline{G}^{987}$GG, and TA$\underline{A}^{1118}$/TA$\underline{C}^{1118}$) [44]; worldwide, two (ACA and GGC) of the four possible haplotypes were found in 94% of the MJD families (72% sharing the ACA haplotype). This finding suggested that two main founder mutations explained the global distribution of MJD [44]. In Azores islands, haplotype distribution was in accordance with the island of origin of the patient, namely the ACA haplotype was observed in families from Flores Island, while the GGC was found in families from São Miguel Island [44]. Distribution of these two haplotypes by island of origin confirmed the previous findings obtained by previous genealogical studies of MJD families [45]. Martins and colleagues, latter performed a more extensive haplotype analysis (6 intragenic

SNPs) using 264 MJD families from 20 different populations, revealing that the origin of the first mutational event occurred in Asia (Japan); an estimated age of 5.774 ± 1.116 years old was calculated for the appearance of mutation [46].

Ataxin-3 and its biological role

Ataxin-3, encoded by the *ATXN3* gene, is composed of 361 amino acid residues (the most common isoform found in brain - Ensembl transcript ID ENST00000393287) and a variable number of glutamine repeats, presenting an estimated molecular weight of 42 kDa in its native form [47]. A N-terminal domain, named Josephin domain (JD) followed by a carboxyl (C)-terminal tail constitutes ataxin-3 main structure [48]. The JD displays the catalytic triad of amino acids found in cysteine proteases [49]; several ubiquitin (Ub)-interacting motifs (UIMs), which binds to polyubiquitin chains plus the polyQ tract are located in the C-terminal region [50]. Despite the selective neuronal degeneration found in the brain of MJD patients, normal and mutated ataxin-3 are ubiquitously expressed, localizing in the cytoplasm and nucleus of several cell types [47,51–54]. Several ataxin-3 isoforms, resulting mostly from alternative splicing, have been described; in blood, 56 *ATXN3* splice variants have been identified, although in silico analysis revealed that only 20 could potentially be translated into a functional ataxin-3 protein [55]. The pathological hallmark of MJD is the formation of neuronal inclusions, both in the nucleus (neuronal nuclear inclusions – NNIs) as well as in the cytoplasm (neuronal citoplasmatic inclusions – NCIs) in affected and non-affected brain areas (see, amongst others, [56,57]). Currently, it is though that neuronal inclusions had a neuroprotective role since no correlations were found between the distribution of NNIs and neurodegeneration [58].

Ataxin-3 is a deubiquitinating enzyme (DUB) playing a major role in ubiquitination (reviewed in [59]). Several evidences suggest that ataxin-3 participates mostly in protein quality control mechanisms such us the ubiquitin proteasome system (UPS) and helps regulate the formation of aggresomes [60]. A role of ataxin-3 in the regulation of transcription, through different

mechanisms, has also been debated; ataxin-3 is able to interact with several transcriptional regulators (reviewed in [61]) and to target chromatin by directly binding DNA or histones [62].

The mutated polyQ repeat triggers conformational changes, which promote misfolding of the protein, affecting its stability, degradation, subcellular localization as well as molecular interactions with other proteins. Abnormal interactions of mutated ataxin-3 with its native partners could leads in some cases to a gain of function or to a partial loss of normal ataxin-3 function (reviewed in [59,61]). Alongside with full length mutated ataxin-3, the mutated protein can be also proteolytically cleaved, originating smaller fragments that are aggregation-prone; both entities can form soluble monomers, oligomers or large insoluble aggregates, which are located in the nucleus as well as in the cytoplasm (reviewed in [61]). Moreover, mutated ataxin-3 may also disturb several physiological processes, namely cellular protein homeostasis (the heat-shock response, the unfolded-protein response and autophagy), transcriptional regulation, mitochondrial function and oxidative stress, amongst others (reviewed in [59,61,63]). Expression levels of several molecular chaperones seems to be dysregulated in SCA3 models, showing, furthermore, a disease progression-associated pattern; up-regulation of several chaperones was observed in early stages of the disease, while during later stages expression levels of HSP70 and HSP40 were down-regulated (reviewed in [59]). Autophagy impairment has been widely described in mouse models [64–66] as well as MJD patient's fibroblasts [67]. Failure of protein clearance pathways correlate with the accumulation of mutant ataxin-3 during disease progression, pinpointing the chaperone system, the ubiquitin-proteasome system and autophagy potential therapeutic targets (reviewed in [68]). Dysregulation of transcription contributes to MJD pathogenesis by (1) abnormal interactions with transcription factors and coactivators and (2) the sequestration of transcription factors to polyQ aggregates (reviewed in [61]; for more details see Chapter V). Mitochondrial function and oxidative stress abnormalities has been widely reported in cell and animal models as well as MJD patients. MJD cells are more susceptibility to oxidative stress, namely reactive oxygen species, due to the reduced ability of mutated ataxin-3 to activate

FOXO4-mediated SOD2 transcription or the decreased levels of antioxidant molecules (reviewed in [59];[69]). Several studies have shown that mitochondrial function is also compromise in MJD, including a significant decrease in mitochondrial complex II activity in cell and mouse MJD models [70] and mitochondrial DNA (mtDNA) damage and depletion in cell and mouse MJD models as well as in blood of MJD patients (for more details see Chapter VI).

MJD clinical trials and its outcome measures

MJD is currently untreatable although some pharmacological and non-pharmacological therapies may be used to alleviate some clinical symptoms/signs. A limited number of interventional trials aiming to modify the disease progression or to reduce disease manifestations have been conducted (reviewed in [71]); results from such trials, however, have been either negative or non-conclusive. The small number of patients, the inadequate study design as well as the short periods of follow-up constitute some of the limitations pinpointed (reviewed in [72]). In 2012, results from two randomized, double blind, placebo controlled trials have been published; the drugs tested were varenicline [73] and lithium carbonate [74]. Both studies suggested that a larger number of patients was needed to conclude about drug efficacy [73,74]. During this year, an improvement of the SARA scores were observed in MJD patients treated with multi-doses of valproic acid [75]. Results are still pending for two interventional trials registered on clinicaltrials.gov: (1) a "single-center, randomized, double-blind, parallel-group, dose-controlled study, to assess safety, tolerability and efficacy of intravenous Cabaletta®" (NCT02147886); and (2) "the therapeutic effect of dalfampridine on gait incoordination in SCAs - a randomized, double-blinded, placebo-controlled, crossover clinical trial" (NCT01811706) is completed but no results have yet been published (www.clinicaltrials.gov).

The two measures selected as primary endpoints in the most recent clinical trials were the scale for the assessment and rating of ataxia (SARA) and the neurological examination score for the assessment of spinocerebellar ataxia (NESSCA), for which its behaviour during the natural history

of SCA3 was previously ascertained as expected [74,76–80]. SARA, an ataxia rating scale [80], is now considered standard in clinical ataxia studies. Functional measures, namely the Composite Cerebellar Functional Score (CCFS) and the SCA Functional Index (SCAFI) are thought to have advantages over ataxia rating scales, since both have been developed to reliably detect small clinical changes during short periods of time [81,82]. Notwithstanding, clinical rating scales if used alone can be subjective and insensitive to subtle changes over short periods of time, therefore requiring large sample sizes and longer and more expensive clinical trials. These limitations could be overcome by the development of molecular biomarkers, which are currently unavailable. More importantly, it is thought that future disease-targeted neuroprotective strategies will be more effective if administered before symptomatology appears; therefore, as SCA3 involves long preclinical stages, it will be crucial to detect molecular changes even before the clinical manifestations.

••• Biomarkers in SCA3: state of the art •••

Definitions and validation steps

A biomarker is defined as "a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacological responses to a therapeutic intervention" [83]. Levels of RNA and evaluation of mitochondrial DNA (mtDNA), presence/absence of mutations and/or polymorphisms, proteins, metabolites and brain structures are examples of the characteristics prone to be measure. The Biomarker Definitions Working Group (2001), specified that candidate markers should be easily and rapidly quantifiable within accessible biological fluids or tissues by using standardized techniques, which can be reliably reproduced at different study centers [83]. Additionally, markers must be independent from unrelated comorbidities and not be subjected to wide variations in the general population. Biomarkers should, furthermore, be associated with the pathogenic process and clinical manifestations of the disease, change linearly (either positively or negatively) with disease

progression, and should show a measurable response to treatments that modify the disease course (Figure 2) [83]. For polyQ diseases, two types of biomarkers can be considered - trait or state biomarkers. A trait biomarker informs about the presence or absence of a pathological condition; on the other hand, if some characteristic is changeable during the course of the disease and it is related with severity it would be considered as a state biomarker [84].

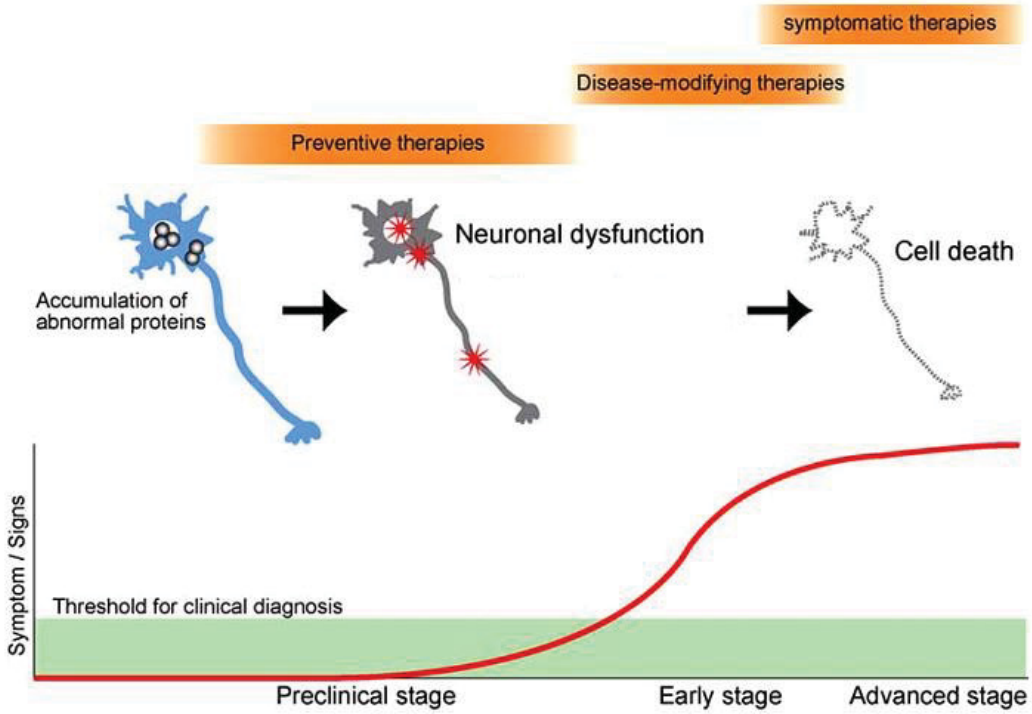


Figure 2. The several stages of MJD progression and the rationale for use of biomarkers. The production and accumulation of abnormal ataxin-3 triggers several downstream effects, which can be targets for biomarkers discovery. If neuronal dysfunction starts many years before clinical diagnosis, disease progression may be partially reversible in the preclinical stage (preventive therapies). Disease-modifying therapies can be applied at early stages of the disease and symptoms relief during advanced stages. Adapted from [85].

The pipeline for the development of biomarkers showing clinical utility involves several validation steps, therefore providing high levels of confidence for its use. First evidences of disease alterations should be described in cross-sectional studies. This phase (phase I) should allow the identification of changes which relate with the presence of the pathological condition. The markers previous identified in phase I should be further tested in longitudinal phase II studies; a phase III will study how markers obtained in phase II behave during an interventional trial. In phase

I (cross-sectional studies) if the discovery technique is a genome-wide technology, commonly chosen given that it is an easy and relatively inexpensive approach, a technical validation should also be needed [86].

In the last years, several cross-sectional studies aiming to identify candidate MJD biomarkers have been performed. Mainly for clinical assessments, a few longitudinal studies have been conducted. Imaging data during progression, for example, has started to be investigated but sample size remains limited. To our knowledge, only one prospective study concerning molecular alterations has been described for MJD [87], which will be addressed ahead. Although cross-sectional analyses allow an overview of the different disease stages, data is obtained at a specific time point, therefore constituting a static measure; moreover, due to the MJD-inherited clinical variability, the allocation of patients within different disease stages in the same study will imply difficulties in detecting highly variable changes, thus delaying the identification of reliable biomarkers. Therefore, longitudinal data should facilitate the development of realistic biomarkers to be used ideally as endpoints in future interventional trials of SCA3.

Clinical biomarkers

Clinical biomarkers are rating scales or functional measures that evaluate progression of different features of the MJD phenotype, namely motor and/or non-motor manifestations. These non-invasive measures can disclose the spectrum of phenotypic variability during the natural history of MJD, and are the basis for correlations with other measures, namely those derived from imaging and molecular data. Currently, clinical biomarkers are the most investigated measures in MJD, being the target of several large observational as well as longitudinal studies (see amongst others, [18,76–80,88]).

Quantitative/qualitative oculomotor data

The investigation of oculomotor abnormalities was described in several reports, although some of them are of a preliminary nature. Supranuclear and upward gaze paresis were reported to be

more evidenced with disease duration in a French and a Thai MJD cohort, respectively [89,90]. A negative correlation between retinal nerve fibre layer (RNFL) thickness (measured by optical coherence tomography - OCT), and SARA score was observed in MJD patients [91]. Vestibular dysfunction in MJD patients has also been described [92–95]. Some oculomotor features were also investigated in the preclinical stage; the frequency of gaze-evoked nystagmus was significantly higher in carriers of *ATXN3* mutation when compared to non-carriers [76].

Rating scales and functional measures

Three rating scales - SARA, NESSCA and International cooperative ataxia rating scale (ICARS) - have been used in MJD prospective studies. Both scales are semi-quantitative and only NESSCA, evaluates also non-ataxia features, SARA and ICARS being exclusively ataxia rating scales (reviewed in [96]). The usefulness of ICARS in MJD and in other SCAs was firstly investigated in 158 European patients [97]. Although results for MJD were not analysed independently, some issues compromising the utility of this scale in clinical trials were raised [97]. Thirty-four Brazilian patients were evaluated with ICARS during two years [98]. França and colleagues observed an increase in ICARS score of 5.1 points per year, which represents a change of 12% when compared to baseline data [98]. One year later, in another Brazilian SCA3 cohort (N=105), an annual increase of the NESSCA score of 1.26 points was obtained in a study performed by Jardim and colleagues [77]; these authors further reported that NESSCA score increased 0.15 points per year by each additional CAG repeat unit, and decreased 0.03 points by each additional year of age at onset [77]. Disease progression data from three larger SCA3 cohorts have been reported using SARA, which is currently the most used scale for SCA3. A progression score of 3±1.52, 1.14±0.04 and 1.56±0.08 points per year were calculated using 99 patients from Taiwan, 138 patients from USA and 122 patients from Europe, respectively [79,99,100]. In an USA cohort, other functional measures, namely the abbreviated SCA Functional Index (SCAFI-AB, which does not include the PATA rate) was also investigated; a score of -0.143 points per year was calculated [99]. SARA, INAS, SCAFI and

CCFS were also tested in a prospective study of individuals at risk for SCA1, SCA2, SCA3, SCA6, SCA7 (RISCA), which included preataxic carriers of SCA1, 2, 3 and 6. No significant alterations, however, were found in SCA3 preataxic subjects [76].

Imaging biomarkers

Structural and functional neuroimaging measures are appealing as biomarkers mainly because they can allow to detect changes many years before the clinical diagnosis [84]. These measures, however, could be invasive and expensive and, in addition, comparisons between different studies may be challenging. Moreover, this technology could not be readily accessible in most of the clinical centres. Both structural as well as functional imaging data are derived mainly from cross-sectional studies and, to our knowledge, only one study collected data in a longitudinal setup.

Structural data

Brain macrostructures and microstructures abnormalities were widely investigated in cross-sectional MJD studies; a significant atrophy of cerebellum and brainstem [22,101–104] as well as of subcortical structures [11,105] was described. The relationship between brain structural damage and genetic and clinical SCA3 features, namely the number of CAG repeats in expanded allele, age at onset, disease duration and rating scales (SARA or ICARS) was not consensual [11,22,102–106]. In a study by Reetz and colleagues [107], longitudinal atrophy patterns were obtained after the followed-up of 19 SCA3 patients during two years. A time-dependent reduction in the volume of brainstem, pons, putamen and caudate nucleus was found in SCA3 patients compared to controls [107]. These structural abnormalities were also recently investigated in preataxic carriers; volumetric analyses shown that the brainstem volumes of nine SCA3 preataxic carriers were lower than the 33 pooled-non carriers [76].

Functional data

The concentration of several neurometabolites have been widely investigated. Significant lower ratios of the N-acetylaspartate and N-acetylaspartylglutamate (NAA)/choline (Cho) as well as NAA/creatine (Cr) in the cerebellar hemispheres and vermis have been described; furthermore, both ratios correlated with early stages of the disease (patients presenting SARA scores lower than 10 points) [108]. Lopes and colleagues observed lower concentrations of NAA and glutamate + glutamine (Glx) in SCA3 patients [11]. These metabolites were further investigated in a recent study performed by Adanyegnh and colleagues; they found, in the cerebellar vermis and pons of 23 SCA3 patients, low levels of N-acetylaspartate and N-acetylaspartylglutamate (NAA) and glutamate (Glu) and high levels of myoinositol (*myo*-Ins) and creatine and phosphocreatine (tCr) [109]. Moreover, in pons only, tCr and *myo*-Ins concentrations were found to be positively correlated with SARA scores; also in pons, a negative correlation between SARA scores and NAA concentrations was observed. The concentration of the neurometabolites was not associated either with the number of CAG repeats in the expanded allele or with disease duration [109].

Abnormalities in the binding of decreased dopamine transporter (DAT) in SCA3 patients as well as preataxic carriers have been described. DAT binding in putamen and caudate nucleus was found in SCA3 patients, although this alteration was not correlated with the number of CAG repeats in the expanded allele, age at onset or disease duration [110]. These findings were further confirmed in an independent SCA3 cohort, including preataxic carriers; the reduction of DAT binding was more pronounced in SCA3 patients than in preataxic carriers [111].

Dysfunction of glucose metabolism was also described in preataxic carriers, showing alterations in the cerebellum and temporal and parietal cortices; these differences, moreover, were sufficient to distinguished carriers from non-carriers of the *ATXN3* mutation [112].

Maps of regional cerebral blood flow (rCBF), using arterial spin labelling – MRI were obtained for 22 SCA3 patients, 16 preataxic carriers and 27 controls. A significant decreased rCBF was observed

in pons, cerebellar dentate nucleus and cerebellar cortex from SCA3 patients; in preataxic carriers lower rCBF was found in cerebellar dentate nucleus and cerebellar cortex [113].

Duarte and colleagues recently reported the potential of functional MRI (fMRI) patterns to be biomarkers of SCA3 in early stages [114]. Findings were based on parametric performance-level-dependent signals and authors suggested that the functional reorganization of the motor network, provided by the fMRI, could be used to monitor disease progression in SCA3 [114].

Molecular biomarkers

In the search for biomarkers of SCA3, the quantification of molecular changes from easily collected biological fluids, namely blood, is of great appeal. Blood sampling could provide the identification of downstream effects promoted by abnormalities from CNS occurring in the periphery. Studies of serum or plasma are most frequent, but whole blood as well as other blood components such as leucocytes have also been investigated in SCA3. Some alterations obtained from CSF studies have also been described, although this procedure is more invasive, and therefore, samples are more difficult to obtain compared with blood.

Trait biomarkers

The presence of the *ATXN3* mutation is the primary trait biomarker of SCA3. However, as SCA3 phenotype is only partially explained by the number of CAG repeats in the expanded allele, other minor trait biomarkers remain to be identified, as previously referred. The discovery of future trait biomarkers will (1) improve the prediction of age at onset provided mostly by the size of the CAG tract; (2) allow the stratification of participants in future interventional trials, therefore empowering such trials; and (3) provide new insights into pathologically-associated SCA3 mechanisms.

Minor trait biomarkers can also be designated as genetic modifiers, which are one or more gene variants able to modulate the phenotypic expression of a monogenic disease [115]. In MJD, the

phenotypic variability which can be mostly modulated by other genes is the residual age at onset after accounting for the contribution of the CAG number repeats. The several genetic modifiers currently proposed for SCA3 are (1) the number of CAG repeats at several expansion loci [41,116]; (2) the allelic variants at the apolipoprotein E (*APOE*) [117,118] and at the glucosidase, beta, acid (*GBA*) genes [119]; (3) the variation in the 3'UTR at the *ATXN3* gene [120] and (4) the size of the normal *ATXN3* allele [121]. Recently, allelic variants of two DNA repair genes – the FANCD2/FANCI-associated nuclease 1 (*FAN1*) and the PMS1 homolog 2, mismatch repair system component (*PMS2*) were associated with age at onset if grouping all polyQ diseases in the analysis [122]. It is expectable that each genetic modifier will have a modest contribution, and therefore, the evaluation of the cumulative effect of a set of genes will be more useful to provide insights into shared mechanisms involved in SCA3 pathogenesis as well as to allow a better prediction of the appearance of clinical manifestations.

State biomarkers

Although realistic state biomarkers will probably only be provided by longitudinal studies, the large majority of candidate biomarkers were derived from cross-sectional studies; for some of them, correlations between molecular alterations and clinical and genetic features, namely disease duration and the number of CAG repeats in expanded allele were also available. One strategy to selected candidate molecular markers could be to investigate molecules (mRNA or protein levels) which played a role in pathologically-associated SCA3 mechanisms, namely transcriptional dysregulation, mitochondrial dysfunction and autophagy impairment. Moreover, some of these pathways can be also potential targets of drug therapies, suggesting that molecules previously identified as state biomarkers can be applied also as biomarkers of therapy response. Several molecules playing a role in mitochondria or autophagy pathways have been described in blood, CSF and fibroblasts of MJD patients.

Using samples of cerebrospinal fluid (CSF) from seven SCA3 patients, Matsuishi and colleagues found increased levels of lactate and lactate/pyruvate ratio as well as decreased pyruvate. These results were not in agreement with those found by the same authors in blood of SCA3 patients [123]. A case-control study, using seven SCA3 patients and seven paired healthy controls showed lower levels of thiols (indirect measure of glutathione peroxidase levels) and higher levels of the antioxidant enzyme catalase in SCA3 patients; these authors also found high levels of DNA damage in SCA3 samples [69].

MtDNA content and damage (m.8470_13446del4977 deletion, also known as common deletion, an age-related marker) previously investigated in several animal and cell SCA3 models (see amongst others [124,125]), have also been studied in blood of SCA3 subjects. Liu and colleagues analysed leucocytes from 61 SCA3 patients, reporting that mtDNA copy number/cell was lower in patients compared with controls. Furthermore, they reported that mtDNA depletion may occur in a CAG-dependent manner ($r=-0.301$; $p=0.048$). Low levels of mtDNA copy number/cell were also described in preataxic subjects with polyQ disease [126]. mtDNA damage and depletion were further confirmed in a limited number of SCA3 patients [127]. These findings were not further replicated by Zeng and colleagues [128].

Beclin 1 is an autophagy related protein whose levels were diminished in the presence of mutated ataxin-3. Overexpression of beclin-1 was found to ameliorate the phenotype of a SCA3 transgenic mice [65,66]. As impairment of autophagy is related with SCA3 pathogenesis, and beclin-1 seems to have an important role in this mechanism, Kazachkova and colleagues quantified the mRNA levels of *BECN1* (Kazachkova et al. submitted). *BECN1* mRNA levels were found to be 1.4 times higher in SCA3 patients compared to healthy controls (Kazachkova et al., submitted). SIRT1 protein, encoded by the *SIRT1* (sirtuin 1) gene, is a NAD^+-dependent deacetylase, which activity is required in caloric restriction regimen. Its neuroprotection effect in polyQ diseases, namely in SCA3, has been debated in a recent study by Cunha-Santos and colleagues [64]. They found that

levels of SIRT1 increased the alleviation of motor impairments of a SCA3 transgenic mice. Moreover, the authors reported that mRNA levels of *SIRT1*, as well as protein levels were severely decreased in human fibroblasts of SCA3 patients in comparison with controls [64].

The peripheral insulin/insulin-like growth factor 1 (IGF1) system (IIS) profile (insulin, IGF1, IGFBP1 and 3) and its correlation with clinical, molecular, and neuroimaging findings were evaluated in a case–control study, including 46 SCA3 patients and 42 healthy subjects [129]. Low levels of insulin and IGFBP3 and high levels of insulin sensitivity (HOMA2), free IGFI, and IGFBP1 where obtained in SCA3 patients. Furthermore, patients who had higher number of CAG repeats in expanded allele presented higher IGFBP1 levels; IGF1 levels were correlated with MRI data and an association between insulin levels and sensitivity with age at onset was described [129].

Cytokines profile in SCA3 patients and preataxic carriers of the *ATXN3* mutation has been recently described [87]. Alterations in the cytokine patterns were not found after comparing either patients with controls or patients with preataxic carriers or carriers with controls. A significantly higher eotaxin (member of the chemokine subfamily) concentration was observed in preataxic carriers after comparisons with patients and controls, although correlations of eotaxin with age, disease duration, CAG expansion, NESSCA score, and SARA score were missed. In a subset of SCA3 patients selected for a longitudinal analysis, eotaxin levels decreased after 360 days [87].

••• Pertinence and main goal•••

Although rare worldwide, MJD is highly prevalent in the Azores Archipelago [6]. Over 30 Azorean MJD families were identified and their genealogical data was compiled [45,130], allowing to start a comprehensive study of the disease. The Azorean MJD cohort, given its population context, displays several characteristics of a population isolate, facilitating the identification of molecular biomarkers, which can be further validate in independent cohorts. The development of trait biomarkers will allow: (1) to improve the prediction of age at onset, mainly provided by the CAG size in expanded allele, and (2) to empower interventional trials, by allowing genetic stratification of patients. State biomarkers will permit: (1) to track disease progression even before the ataxia onset, and (2) to detect subtle therapeutic benefits. The preclinical stage is of major importance, since has been hypothesised that cell damage starts many years before ataxia onset (pathological onset). The development of molecular biomarkers, complemented with imaging data will therefore constitute the only measures available to use in this non-clinical period.

The main goal of the present PhD thesis is to contribute for the development of objective, accessible and easily measured molecular biomarkers of MJD. Blood samples of preataxic subjects and MJD patients, including samples collected in two moments constitute the main resource to be used to fulfil the several objectives of the present PhD project.

A description of the present dissertation structure is shown in Figure 3. The Azorean MJD cohort can be considered a trial-ready cohort and its clinical and genetic characterization is described in Chapter I. The presence of a specific clinical sign, namely nystagmus, was evaluated in at-risk Azorean individuals - Chapter II. In total, five chapters were devoted to the investigation of molecular biomarkers, which were further divided as trait or state biomarkers. Chapter III and IV investigate the potential of candidate genetic modifiers to improve age at onset prediction. The remaining Chapters – V, VI and VII – describe the search for state biomarkers, based in the investigation of mRNA levels alterations and mitochondrial DNA (mtDNA) damage and depletion.

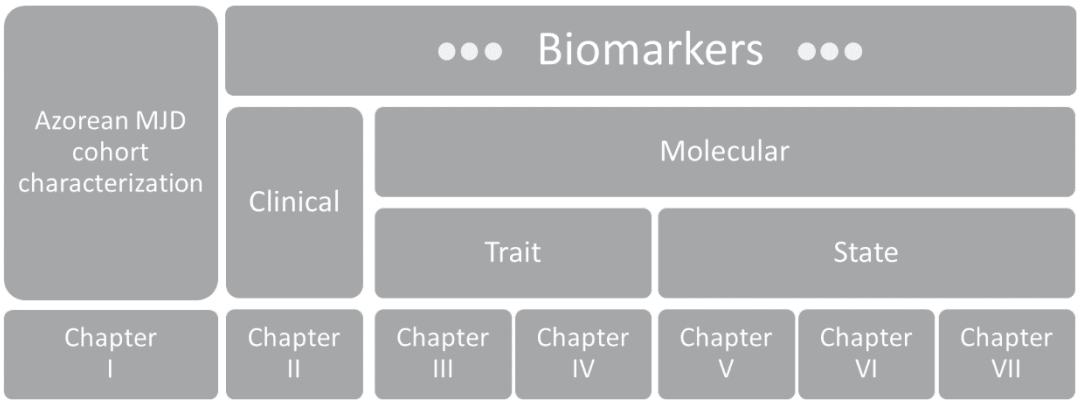


Figure 3. Structure of the present dissertation: workflow.

REFERENCES

1. Ruano L, Melo C, Silva MC, Coutinho P. The global epidemiology of hereditary ataxia and spastic paraplegia: a systematic review of prevalence studies. Neuroepidemiology. 2014;42:174–83.

2. Sun Y-M, Lu C, Wu Z-Y. Spinocerebellar ataxia: relationship between phenotype and genotype - a review. Clin. Genet. 2016;90:305–14.

3. Seidel K, Siswanto S, Brunt ERP, den Dunnen W, Korf H-W, Rüb U. Brain pathology of spinocerebellar ataxias. Acta Neuropathol. 2012;124:1–21.

4. Michalik A, Van Broeckhoven C. Pathogenesis of polyglutamine disorders: aggregation revisited. Hum. Mol. Genet. 2003;12:R173-86.

5. Bettencourt C, Lima M. Machado-Joseph Disease: from first descriptions to new perspectives. Orphanet J. Rare Dis. 2011;6:35.

6. de Araújo M, Raposo M, Kazachkova N, Vasconcelos J, Kay T, Lima M. Trends in the Epidemiology of Spinocerebellar Ataxia Type 3/Machado-Joseph Disease in the Azores Islands, Portugal. JSM Brain Sci. 2016;1:1001.

7. Lima L, Coutinho P. Clinical criteria for diagnosis of Machado-Joseph disease: report of a non-Azorena Portuguese family. Neurology. 1980;30:319–22.

8. Braga-Neto P, Dutra LA, Pedroso JL, Felício AC, Alessi H, Santos-Galduroz RF, et al. Cognitive deficits in Machado-Joseph disease correlate with hypoperfusion of visual system areas. Cerebellum. 2012;11:1037–44.

9. Braga-Neto P, Felicio AC, Hoexter MQ, Pedroso JL, Dutra LA, Alessi H, et al. Cognitive and olfactory deficits in Machado-Joseph disease: a dopamine transporter study. Parkinsonism Relat. Disord. 2012;18:854–8.

10. Zawacki TM, Grace J, Friedman JH, Sudarsky L. Executive and emotional dysfunction in Machado-Joseph disease. Mov. Disord. 2002;17:1004–10.

11. Lopes TM, D'Abreu A, França MC, Yasuda CL, Betting LE, Samara AB, et al. Widespread neuronal damage and cognitive dysfunction in spinocerebellar ataxia type 3. J. Neurol. 2013;260:2370–9.

12. Coutinho P. Doença de Machado-Joseph: Tentativa de definição. Universidade do Porto; 1992.

13. Globas C, du Montcel ST, Baliko L, Boesch S, Depondt C, DiDonato S, et al. Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. Mov. Disord. 2008;23:2232–8.

14. Raposo M, Vasconcelos J, Bettencourt C, Kay T, Coutinho P, Lima M. Nystagmus as an early ocular alteration in Machado-Joseph disease (MJD/SCA3). BMC Neurol. 2014;14:17.

15. Raposo M, Ramos A, Bettencourt C, Lima M. Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: Can homogeneous cohorts aid? Brain. 2015. p. e398.

16. Tezenas du Montcel S, Durr A, Bauer P, Figueroa K, Ichikawa Y, Brussino A, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain. 2014;137:2444–55.

17. Carvalho DR, La Rocque-Ferreira A, Rizzo IM, Imamura EU, Speck-Martins CE. Homozygosity Enhances Severity in Spinocerebellar Ataxia Type 3. Pediatr. Neurol. 2008;38:296–9.

18. Kieling C, Rieder CRM, Silva ACF, Saute JAM, Cecchin CR, Monte TL, et al. A neurological examination score for the assessment of spinocerebellar ataxia 3 (SCA3). Eur. J. Neurol. 2008;15:371–6.

19. Murata Y, Yamaguchi S, Kawakami H, Imon Y, Maruyama H, Sakai T, et al. Characteristic magnetic resonance imaging findings in Machado-Joseph disease. Arch. Neurol. 1998;55:33–7.

20. Abe Y, Tanaka F, Matsumoto M, Doyu M, Hirayama M, Kachi T, et al. CAG repeat number correlates with the rate of brainstem and cerebellar atrophy in Machado-Joseph disease. Neurology. 1998;51:882–4.

21. Onodera O, Idezuka J, Igarashi S, Takiyama Y, Endo K, Takano H, et al. Progressive atrophy of cerebellum and brainstem as a function of age and the size of the expanded CAG repeats in the MJD1 gene in Machado-Joseph disease. Ann. Neurol. 1998;43:288–96.

22. Eichler L, Bellenberg B, Hahn HK, Köster O, Schöls L, Lukas C. Quantitative assessment of brain stem and cerebellar atrophy in spinocerebellar ataxia types 3 and 6: impact on clinical status. AJNR. Am. J. Neuroradiol. 2011;32:890–7.

23. Lin K-P, Soong B-W. Peripheral neuropathy of Machado-Joseph disease in Taiwan: a morphometric and genetic study. Eur. Neurol. 2002;48:210–7.

24. Rüb U, de Vos RAI, Schultz C, Brunt ER, Paulson H, Braak H. Spinocerebellar ataxia type 3 (Machado-Joseph disease): severe destruction of the lateral reticular nucleus. Brain. 2002;125:2115–24.

25. Rüb U, Bürk K, Schöls L, Brunt ER, de Vos RAI, Diaz GO, et al. Damage to the reticulotegmental nucleus of the pons in spinocerebellar ataxia type 1, 2, and 3. Neurology. 2004;63:1258–63.

26. Rüb U, Brunt ER, Petrasch-Parwez E, Schöls L, Theegarten D, Auburger G, et al. Degeneration of ingestion-related brainstem nuclei in spinocerebellar ataxia type 2, 3, 6 and 7. Neuropathol. Appl. Neurobiol. 2006;32:635–49.

27. Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, et al. The gene for Machado-Joseph disease maps to human chromosome 14q. Nat. Genet. 1993;4:300–4.

28. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet. 1994;8:221–8.

29. Lima M, Costa MC, Montiel R, Ferro A, Santos C, Silva C, et al. Population genetics of wild-type CAG repeats in the Machado-Joseph Disease gene in Portugal. Hum. Hered. 2005;60:156–63.

30. Padiath QS, Srivastava AK, Roy S, Jain S, Brahmachari SK. Identification of a novel 45 repeat unstable allele associated with a disease phenotype at the MJD1/SCA3 locus. Am. J. Med. Genet. Part B, Neuropsychiatr. Genet. 2005;133B:124–6.

31. Maciel P, Costa MC, Ferro A, Rousseau M, Santos CS, Gaspar C, et al. Improvement in the molecular diagnosis of Machado-Joseph disease. Arch. Neurol. 2001;58:1821–7.

32. Igarashi S, Takiyama Y, Cancel G, Rogaeva EA, Sasaki H, Wakisaka A, et al. Intergenerational instability of the CAG repeat of the gene for Machado-Joseph disease (MJD1) is affected by the genotype of the normal chromosome: implications for the molecular mechanisms of the instability of the CAG repeat. Hum. Mol. Genet. 1996;5:923–32.

33. Takiyama Y, Sakoe K, Soutome M, Namekawa M, Ogawa T, Nakano I, et al. Single sperm analysis of the CAG repeats in the gene for Machado-Joseph disease (MJD1): evidence for non-Mendelian transmission of the MJD1 gene and for the effect of the intragenic CGG/GGG polymorphism on the intergenerational instability. Hum. Mol. Genet. 1997;6:1063–8.

34. Martins S, Coutinho P, Silveira I, Giunti P, Jardim LB, Calafell F, et al. Cis-acting factors promoting the CAG intergenerational instability in Machado-Joseph disease. Am. J. Med. Genet. Part B, Neuropsychiatr. Genet. 2008;147B:439–46.

35. Maciel P, Gaspar C, Guimarães L, Goto J, Lopes-Cendes I, Hayes S, et al. Study of three intragenic polymorphisms in the Machado-Joseph disease gene (MJD1) in relation to genetic instability of the (CAG)n tract. Eur. J. Hum. Genet. 1999;7:147–56.

36. Martins S, Pearson CE, Coutinho P, Provost S, Amorim A, Dubé M-P, et al. Modifiers of (CAG)(n) instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. Hum. Genet. 2014;133:1311–8.

37. Souza GN, Kersting N, Krum-Santos AC, Santos ASP, Furtado G V, Pacheco D, et al. Spinocerebellar ataxia type 3/Machado-Joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. Clin. Genet. 2016;90:134–40.

38. Paulson HL. Dominantly inherited ataxias: lessons learned from Machado-Joseph disease/spinocerebellar ataxia type 3. Semin. Neurol. 2007;27:133–42.

39. Tanaka F, Sobue G, Doyu M, Ito Y, Yamamoto M, Shimada N, et al. Differential pattern in tissue-specific somatic mosaicism of expanded CAG trinucleotide repeats in dentatorubral-pallidoluysian atrophy, Machado-Joseph disease, and X-linked recessive spinal and bulbar muscular atrophy. J. Neurol. Sci. 1996;135:43–50.

40. Cancel G, Gourfinkel-An I, Stevanin G, Didierjean O, Abbas N, Hirsch E, et al. Somatic mosaicism of the CAG repeat expansion in spinocerebellar ataxia type 3/Machado-Joseph disease. Hum. Mutat. 1998;11:23–7.

41. Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain. 2014;137:2444–55.

42. Matsumura R, Takayanagi T, Fujimoto Y, Murata K, Mano Y, Horikawa H, et al. The relationship between trinucleotide repeat length and phenotypic variation in Machado-Joseph disease. J. Neurol. Sci. 1996;139:52–7.

43. Jardim LB, Pereira ML, Silveira I, Ferro A, Sequeiros J, Giugliani R. Neurologic findings in Machado-Joseph disease: relation with disease duration, subtypes, and (CAG)n. Arch. Neurol. 2001;58:899–904.

44. Gaspar C, Lopes-Cendes I, Hayes S, Goto J, Arvidsson K, Dias A, et al. Ancestral origins of the Machado-Joseph disease mutation: a worldwide haplotype study. Am. J. Hum. Genet. 2001;68:523–8.

45. Lima M, Mayer FM, Coutinho P, Abade A. Origins of a mutation: population genetics of Machado-Joseph disease in the Azores (Portugal). Hum. Biol. an Int. Rec. Res. 1998;70:1011–23.

46. Martins S, Calafell F, Wong VCN, Sequeiros J, Amorim A. A multistep mutation mechanism drives the evolution of the CAG repeat at MJD/SCA3 locus. Eur. J. Hum. Genet. 2006;14:932–40.

47. Schmidt T, Landwehrmeyer GB, Schmitt I, Trottier Y, Auburger G, Laccone F, et al. An isoform of ataxin-3 accumulates in the nucleus of neuronal cells in affected brain regions of SCA3 patients. Brain Pathol. 1998;8:669–79.

48. Masino L, Musi V, Menon RP, Fusi P, Kelly G, Frenkiel TA, et al. Domain architecture of the polyglutamine protein ataxin-3: a globular domain followed by a flexible tail. FEBS Lett. 2003;549:21–5.

49. Nicastro G, Menon RP, Masino L, Knowles PP, McDonald NQ, Pastore A. The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. Proc. Natl. Acad. Sci. U. S. A. 2005;102:10493–8.

50. Tzvetkov N, Breuer P. Josephin domain-containing proteins from a variety of species are active de-ubiquitination enzymes. Biol. Chem. 2007;388:973–8.

51. Costa M do C, Gomes-da-Silva J, Miranda CJ, Sequeiros J, Santos MM, Maciel P. Genomic structure, promoter activity, and developmental expression of the mouse homologue of the Machado-Joseph disease (MJD) gene. Genomics. 2004;84:361–73.

52. Ichikawa Y, Goto J, Hattori M, Toyoda A, Ishii K, Jeong SY, et al. The genomic structure and expression of MJD, the Machado-Joseph disease gene. J. Hum. Genet. 2001;46:413–22.

53. Paulson HL, Das SS, Crino PB, Perez MK, Patel SC, Gotsdiner D, et al. Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. Ann. Neurol. 1997;41:453–62.

54. Trottier Y, Cancel G, An-Gourfinkel I, Lutz Y, Weber C, Brice A, et al. Heterogeneous intracellular localization and expression of ataxin-3. Neurobiol. Dis. 1998;5:335–47.

55. Bettencourt C, Santos C, Montiel R, Costa M do C, Cruz-Morales P, Santos LR, et al. Increased transcript diversity: novel splicing variants of Machado-Joseph Disease gene (ATXN3). Neurogenetics. 2009;

56. Paulson HL, Perez MK, Trottier Y, Trojanowski JQ, Subramony SH, Das SS, et al. Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. Neuron. 1997;19:333–44.

57. Hayashi M, Kobayashi K, Furuta H. Immunohistochemical study of neuronal intranuclear and cytoplasmic inclusions in Machado-Joseph disease. Psychiatry Clin. Neurosci. 2003;57:205–13.

58. Rüb U, de Vos RAI, Brunt ER, Sebestény T, Schöls L, Auburger G, et al. Spinocerebellar ataxia type 3 (SCA3): thalamic neurodegeneration occurs independently from thalamic ataxin-3 immunopositive neuronal intranuclear inclusions. Brain Pathol. 2006;16:218–27.

59. Costa M do C, Paulson HL. Toward understanding Machado-Joseph disease. Prog. Neurobiol. 2012. p. 239–57.

60. Burnett BG, Pittman RN. The polyglutamine neurodegenerative protein ataxin 3 regulates aggresome formation. Proc. Natl. Acad. Sci. U. S. A. 2005;102:4330–5.

61. Evers MM, Toonen LJA, van Roon-Mom WMC. Ataxin-3 protein and RNA toxicity in spinocerebellar ataxia type 3: current insights and emerging therapeutic strategies. Mol. Neurobiol. 2014;49:1513–31.

62. Evert BO, Araujo J, Vieira-Saecker AM, de Vos RAI, Harendza S, Klockgether T, et al. Ataxin-3 represses transcription via chromatin binding, interaction with histone deacetylase 3, and histone deacetylation. J. Neurosci. 2006;26:11474–86.

63. Matos C , de Macedo-Ribeiro S, Carvalho AL. Polyglutamine diseases: the special case of ataxin-3 and Machado-Joseph disease. Prog. Neurobiol. 2011;95:26–48.

64. Cunha-Santos J, Duarte-Neves J, Carmona V, Guarente L, Pereira de Almeida L, Cavadas C. Caloric restriction blocks neuropathology and motor deficits in Machado-Joseph disease mouse models through SIRT1 pathway. Nat. Commun. 2016;7:11445.

65. Nascimento-Ferreira I, Santos-Ferreira T, Sousa-Ferreira L, Auregan G, Onofre I, Alves S, et al. Overexpression of the autophagic beclin-1 protein clears mutant ataxin-3 and alleviates Machado-Joseph disease. Brain. 2011;134:1400–15.

66. Nascimento-Ferreira I, Nóbrega C, Vasconcelos-Ferreira A, Onofre I, Albuquerque D, Aveleira C, et al. Beclin 1 mitigates motor and neuropathological deficits in genetic mouse models of Machado-Joseph disease. Brain. 2013;136:2173–88.

67. Onofre I, Mendonça N, Lopes S, Nobre R, de Melo JB, Carreira IM, et al. Fibroblasts of Machado Joseph Disease patients reveal autophagy impairment. Sci. Rep. 2016;6:28220.

68. Li X, Liu H, Fischhaber PL, Tang T-S. Toward therapeutic targets for SCA3: Insight into the role of Machado-Joseph disease protein ataxin-3 in misfolded proteins clearance. Prog. Neurobiol. 2015;132:34–58.

69. Pacheco LS, da Silveira AF, Trott A, Houenou LJ, Algarve TD, Belló C, et al. Association between Machado-Joseph disease and oxidative stress biomarkers. Mutat. Res. 2013;757:99–103.

70. Laço MN, Oliveira CR, Paulson HL, Rego AC. Compromised mitochondrial complex II in models of Machado-Joseph disease. Biochim. Biophys. Acta. 2012;1822:139–49.

71. Kazachkova N, Lima M. Towards A Therapeutic Intervention in Polyglutamine Ataxias: From Models to Clinical Trials. In: Atta-ur-Rahman, editor. Front. Clin. Drug Res. -CNS Neurol. Disord. 2015. p. 77–130.

72. Bezprozvanny I, Klockgether T. Therapeutic prospects for spinocerebellar ataxia type 2 and 3. Drugs Future. 2009;34:991.

73. Zesiewicz TA, Greenstein PE, Sullivan KL, Wecker L, Miller A, Jahan I, et al. A randomized trial of varenicline (Chantix) for the treatment of spinocerebellar ataxia type 3. Neurology. 2012;78:545–50.

74. Saute JAM, de Castilhos RM, Monte TL, Schumacher-Schuh AF, Donis KC, D'Ávila R, et al. A randomized, phase 2 clinical trial of lithium carbonate in Machado-Joseph disease. Mov. Disord. 2014;29:568–73.

75. Lei L-F, Yang G-P, Wang J-L, Chuang D-M, Song W-H, Tang B-S, et al. Safety and efficacy of valproic acid treatment in SCA3/MJD patients. Parkinsonism Relat. Disord. 2016;26:55–61.

76. Jacobi H, Reetz K, du Montcel ST, Bauer P, Mariotti C, Nanetti L, et al. Biological and clinical characteristics of individuals at risk for spinocerebellar ataxia types 1, 2, 3, and 6 in the longitudinal RISCA study: analysis of baseline data. Lancet Neurol. 2013;12:650–8.

77. Jardim LB, Hauser L, Kieling C, Saute JAM, Xavier R, Rieder CRM, et al. Progression rate of neurological deficits in a 10-year cohort of SCA3 patients. Cerebellum. 2010;9:419–28.

78. Jacobi H, Bauer P, Giunti P, Labrum R, Sweeney MG, Charles P, et al. The natural history of spinocerebellar ataxia type 1, 2, 3, and 6: A 2-year follow-up study. Neurology. 2011;77:1035–41.

79. Jacobi H, du Montcel ST, Bauer P, Giunti P, Cook A, Labrum R, et al. Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study. Lancet Neurol. 2015;14:1101–8.

80. Schmitz-Hübsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology. 2006;66:1717–20.

81. Schmitz-Hübsch T, Giunti P, Stephenson DA, Globas C, Baliko L, Saccà F, et al. SCA Functional Index: a useful compound performance measure for spinocerebellar ataxia. Neurology. 2008;71:486–92.

82. du Montcel ST, Charles P, Ribai P, Goizet C, Le Bayon A, Labauge P, et al. Composite cerebellar functional severity score: validation of a quantitative score of cerebellar impairment. Brain. 2008;131:1352–61.

83. Biomarkers Definitions Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. Clin. Pharmacol. Ther. 2001;69:89–95.

84. Weir DW, Sturrock A, Leavitt BR. Development of biomarkers for Huntington's disease. Lancet Neurol. 2011;10:573–90.

85. Katsuno M, Banno H, Suzuki K, Takeuchi Y, Kawashima M, Tanaka F, et al. Molecular genetics and biomarkers of polyglutamine diseases. Curr. Mol. Med. 2008;8:221–34.

86. Scherzer CR. Chipping away at diagnostics for neurodegenerative diseases. Neurobiol. Dis. 2009;35:148–56.

87. da Silva Carvalho G, Saute JAM, Haas CB, Torrez VR, Brochier AW, Souza GN, et al. Cytokines in Machado Joseph Disease/Spinocerebellar Ataxia 3. Cerebellum. 2015;

88. França MC, D'Abreu A, Nucci A, Cendes F, Lopes-Cendes I. Progression of ataxia in patients with Machado-Joseph disease. Mov. Disord. 2009;24:1387–90.

89. Dürr A, Stevanin G, Cancel G, Duyckaerts C, Abbas N, Didierjean O, et al. Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. Ann. Neurol. 1996;39:490–9.

90. Boonkongchuen P, Pongpakdee S, Jindahra P, Papsing C, Peerapatmongkol P, Wetchaphanphesat S, et al. Clinical analysis of adult-onset spinocerebellar ataxias in Thailand. BMC Neurol. 2014;14:75.

91. Pula JH, Towle VL, Staszak VM, Cao D, Bernard JT, Gomez CM. Retinal Nerve Fibre Layer and Macular Thinning in Spinocerebellar Ataxia and Cerebellar Multisystem Atrophy. Neuroophthalmology. 2011;35:108–14.

92. Yoshizawa T, Nakamagoe K, Ueno T, Furusho K, Shoji S. Early vestibular dysfunction in Machado-Joseph disease detected by caloric test. J. Neurol. Sci. 2004;221:109–11.

93. Gordon CR, Joffe V, Vainstein G, Gadoth N. Vestibulo-ocular arreflexia in families with spinocerebellar ataxia type 3 (Machado-Joseph disease). J. Neurol. Neurosurg. Psychiatry. 2003;74:1403–6.

94. Rüb U, Brunt ER, de Vos RAI, Del Turco D, Del Tredici K, Gierga K, et al. Degeneration of the central vestibular system in spinocerebellar ataxia type 3 (SCA3) patients and its possible clinical significance. Neuropathol. Appl. Neurobiol. 2004;30:402–14.

95. Zeigelboim BS, Teive HAG, Sampaio R, Jurkiewicz AL, Liberalesso PBN. Electronystagmography findings in spinocerebellar ataxia type 3 (SCA3) and type 2 (SCA2). Arq. Neuropsiquiatr. 2011;69:760–5.

96. Saute JAM, Donis KC, Serrano-Munuera C, Genis D, Ramirez LT, Mazzetti P, et al. Ataxia rating scales--

psychometric profiles, natural history and their application in clinical trials. Cerebellum. 2012;11:488–504.

97. Schmitz-Hübsch T, du Montcel ST, Baliko L, Boesch S, Bonato S, Fancellu R, et al. Reliability and validity of the International Cooperative Ataxia Rating Scale: A study in 156 spinocerebellar ataxia patients. Mov. Disord. 2006;21:699–704.

98. França MC, D'Abreu A, Nucci A, Cendes F, Lopes-Cendes I. Progression of ataxia in patients with Machado-Joseph disease. Mov. Disord. 2009;24:1387–90.

99. Ashizawa T, Figueroa KP, Perlman SL, Gomez CM, Wilmot GR, Schmahmann JD, et al. Clinical characteristics of patients with spinocerebellar ataxias 1, 2, 3 and 6 in the US; a prospective observational study. Orphanet J. Rare Dis. 2013;8:177.

100. Lee Y, Liao Y, Wang P, Lee I-H, Lin K, Soong B. Comparison of cerebellar ataxias: A three-year prospective longitudinal assessment. Mov. Disord. 2011;26:2081–7.

101. Schulz JB, Borkert J, Wolf S, Schmitz-Hübsch T, Rakowicz M, Mariotti C, et al. Visualization, quantification and correlation of brain atrophy with clinical symptoms in spinocerebellar ataxia types 1, 3 and 6. Neuroimage. 2010;49:158–68.

102. Stefanescu MR, Dohnalek M, Maderwald S, Thürling M, Minnerop M, Beck A, et al. Structural and functional MRI abnormalities of cerebellar cortex and nuclei in SCA3, SCA6 and Friedreich's ataxia. Brain. 2015;138:1182–97.

103. Camargos ST, Marques W, Santos AC dos. Brain stem and cerebellum volumetric analysis of Machado Joseph disease patients. Arq. Neuropsiquiatr. 2011;69:292–6.

104. Klockgether T, Skalej M, Wedekind D, Luft AR, Welte D, Schulz JB, et al. Autosomal dominant cerebellar ataxia type I. MRI-based volumetry of posterior fossa structures and basal ganglia in spinocerebellar ataxia types 1, 2 and 3. Brain. 1998;121 (Pt 9:1687–93.

105. de Rezende TJR, D'Abreu A, Guimarães RP, Lopes TM, Lopes-Cendes I, Cendes F, et al. Cerebral cortex involvement in Machado-Joseph disease. Eur. J. Neurol. 2015;22:277–83, e23-4.

106. Schulz JB, Borkert J, Wolf S, Schmitz-Hübsch T, Rakowicz M, Mariotti C, et al. Visualization, quantification and correlation of brain atrophy with clinical symptoms in spinocerebellar ataxia types 1, 3 and 6. Neuroimage. 2010;49:158–68.

107. Reetz K, Costa AS, Mirzazade S, Lehmann A, Juzek A, Rakowicz M, et al. Genotype-specific patterns of atrophy progression are more sensitive than clinical decline in SCA1, SCA3 and SCA6. Brain. 2013;136:905–17.

108. Lirng J-F, Wang P-S, Chen H-C, Soong B-W, Guo WY, Wu H-M, et al. Differences between spinocerebellar ataxias and multiple system atrophy-cerebellar type on proton magnetic resonance spectroscopy. PLoS One. 2012;7:e47925.

109. Adanyeguh IM, Henry PG, Nguyen TM, Rinaldi D, Jauffret C, Valabregue R, et al. In vivo neurometabolic profiling in patients with spinocerebellar ataxia types 1, 2, 3, and 7. Mov. Disord. 2015;30:662–70.

110. Yen TC, Lu CS, Tzen KY, Wey SP, Chou YH, Weng YH, et al. Decreased dopamine transporter binding in Machado-Joseph disease. J. Nucl. Med. 2000;41:994–8.

111. Yen T-C, Tzen K-Y, Chen M-C, Chou Y-HW, Chen R-S, Chen C-J, et al. Dopamine transporter concentration is reduced in asymptomatic Machado-Joseph disease gene carriers. J. Nucl. Med. 2002;43:153–9.

112. Soong BW, Liu RS. Regional decrease in brain glucose metabolism in asymptomatic gene carriers of Machado-Joseph disease: a preliminary report. Zhonghua yi xue za zhi = Chinese Med. J. 1998;61:121–6.

113. Xing W, Wang X, Liao X, Liao W, Shen L. Spin labeling artery method perfusion MRI study of SPG4 and SCA3/MJD. Magn. Reson. Imaging. 2014;32:1330–4.

114. Duarte JV, Faustino R, Lobo M, Cunha G, Nunes C, Ferreira C, et al. Parametric fMRI of paced motor

responses uncovers novel whole-brain imaging biomarkers in spinocerebellar ataxia type 3. Hum. Brain Mapp. 2016;37:3656–68.

115. Génin E, Feingold J, Clerget-Darpoux F. Identifying modifier genes of monogenic disease: Strategies and difficulties. Hum. Genet. 2008. p. 357–68.

116. Jardim L, Silveira I, Pereira ML, do Céu Moreira M, Mendonça P, Sequeiros J, et al. Searching for modulating effects of SCA2, SCA6 and DRPLA CAG tracts on the Machado-Joseph disease (SCA3) phenotype. Acta Neurol. Scand. 2003;107:211–4.

117. Bettencourt C, Raposo M, Kazachkova N, Cymbron T, Santos C, Kay T, et al. The APOE ε2 allele increases the risk of earlier age at onset in Machado-Joseph disease. Arch. Neurol. 2011;68:1580–3.

118. Peng H, Wang C, Chen Z, Sun Z, Jiao B, Li K, et al. APOE ε2 allele may decrease the age at onset in patients with spinocerebellar ataxia type 3 or Machado-Joseph disease from the Chinese Han population. Neurobiol. Aging. 2014;35:2179.e15-8.

119. Siebert M, Donis KC, Socal M, Rieder C, Emmel VE, Vairo F, et al. Glucocerebrosidase gene variants in parkinsonian patients with Machado Joseph/spinocerebellar ataxia 3. Park. Relat. Disord. 2012;18:185–90.

120. Long Z, Chen Z, Wang C, Huang F, Peng H, Hou X, et al. Two novel SNPs in ATXN3 3' UTR may decrease age at onset of SCA3/MJD in Chinese patients. PLoS One. 2015;10:e0117488.

121. França MC, Emmel VE, D'Abreu A, Maurer-Morelli C V, Secolin R, Bonadia LC, et al. Normal ATXN3 Allele but Not CHIP Polymorphisms Modulates Age at Onset in Machado-Joseph Disease. Front. Neurol. 2012;3:164.

122. Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann. Neurol. 2016;79:983–90.

123. Matsuishi T, Sakai T, Nagamitsu S, Shoji H, Ueda N, Kaneko S, et al. Decreased cerebrospinal fluid levels of substance P in Machado-Joseph disease. J. Neurol. Sci. 1996;142:107–10.

124. Ramos A, Kazachkova N, Silva F, Maciel P, Silva-Fernandes A, Duarte-Silva S, et al. Differential mtDNA damage patterns in a transgenic mouse model of Machado-Joseph disease (MJD/SCA3). J. Mol. Neurosci. 2015;55:449–53.

125. Kazachkova N, Raposo M, Montiel R, Cymbron T, Bettencourt C, Silva-Fernandes A, et al. Patterns of mitochondrial DNA damage in blood and brain tissues of a transgenic mouse model of Machado-Joseph disease. Neurodegener. Dis. 2013;11:206–14.

126. Liu C-S, Cheng W-L, Kuo S-J, Li J-Y, Soong B-W, Wei Y-H. Depletion of mitochondrial DNA in leukocytes of patients with poly-Q diseases. J. Neurol. Sci. 2008;264:18–21.

127. Yu Y-C, Kuo C-L, Cheng W-L, Liu C-S, Hsieh M. Decreased antioxidant enzyme activity and increased mitochondrial DNA damage in cellular models of Machado-Joseph disease. J. Neurosci. Res. 2009;87:1884–91.

128. Zeng A, Liu X, Shen L, Li W, Ding Z, Bai Y, et al. Analysis of mitochondrial DNA variations in a Chinese family with spinocerebellar ataxia. J. Clin. Neurosci. 2012;19:60–4.

129. Saute JAM, da Silva ACF, Muller AP, Hansel G, de Mello AS, Maeda F, et al. Serum insulin-like system alterations in patients with spinocerebellar ataxia type 3. Mov. Disord. 2011;26:731–5.

130. Lima M, Mayer F, Coutinho P, Abade A. Prevalence, geographic distribution, and genealogical investigation of Machado-Joseph disease in the Azores (Portugal). Hum. Biol. 1997;69:383–91.

CHAPTER I

The Azorean Machado-Joseph disease/Spinocerebellar Ataxia type 3 cohort: clinical and genetic characterization.

In prep.

Mafalda Raposo[1,2,3], João Vasconcelos[4], Conceição Bettencourt[5], Teresa Kay[6], Manuela Lima[1,2,3]

1. Faculdade de Ciências e Tecnologia, Departamento de Biologia, Universidade dos Açores, Ponta Delgada, Portugal
2. Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal
3. Instituto de Biologia Celular e Molecular (IBMC), Universidade do Porto, Porto, Portugal
4. Departamento de Neurologia, Hospital do Divino Espírito Santo, Ponta Delgada, Portugal
5. Department of Clinical and Experimental Epilepsy and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
6. Departamento de Genética Clinica, Hospital D. Estefânia, Lisboa, Portugal

ABSTRACT

Interventional trials in Machado-Joseph disease (MJD), an untreatable neurodegenerative disorder, are imminent. Genetic and clinical characterization of worldwide MJD clusters will favour several aspects of future interventional trials, such as allowing the stratification of patients by presence/absence of particular genetic variants. Moreover, tracking the behaviour of clinical features measured by a rating scale, during disease progression, is also valuable, since the trial-associated endpoints currently in use are clinical scales. In the Azores, where the MJD prevalence is the highest worldwide, a homogeneous MJD cluster comprising 88 patients has been genetically and clinically characterized. For a subset of 69 MJD subjects the score for the Neurological Examination Score for Spinocerebellar Ataxia (NESSCA) was available at baseline; one or two years after baseline, 23 MJD subjects were again evaluated. From the 88 SCA3 patients alive and living in Azores by the end of September 2016, 54 preserve gait autonomy. The explanation of age at onset variance by the CAG number in expanded allele is, to the best of our knowledge, the highest described in European cohorts (59%). Average NESSCA score at baseline was 12±5 (mean ± standard deviation), for which 40% of variance explanation is provided by the CAG number in expanded allele and disease duration. One year after baseline evaluation, the NESSCA score increased in average 2.4 ± 1.1 points (N=8, p=0.042); patients with 2 years of interval between evaluations showed an increase in average of 3.4 ± 0.8 points (N=15, p=0.002). The genetic and clinical characterization of the Azorean MJD patients should promote the understanding of several aspects of the disease and the establishment of a well-characterized trial-ready cohort.

INTRODUCTION

Machado-Joseph disease (MJD; MIM #109150; ORPHA #98757), also known as spinocerebellar ataxia type 3 (SCA3), is a late onset neurodegenerative disorder inherited as autosomal dominant. Similarly to clinical entities with a major genetic determinant, MJD is a rare disease; noteworthy, it is the most frequent within its own group of disorders, the autosomal dominant spinocerebellar ataxias (SCAs) [1]. Prevalence estimations for the Azores Archipelago (Portugal) have recently been reported as 39/100,000; the disease clusters in the island of Flores, where 1 in each 158 is affected [2].

Average onset of MJD has been reported as around 40 years of age, but initial disease manifestations can occur within a considerably wide age interval, spanning from 4 [3] to 78 years [4]; ataxia, progressive external ophtalmoplegia, pyramidal and extrapyramidal signs, dystonia with rigidity and distal muscular atrophies constitute the main features of MJD's clinical picture [5]. Cerebellar alterations have been considered as the most relevant in MJD, gait ataxia being the most frequent initial manifestation. Among the different neurological assessment tools that have been proposed to measure the severity of the cerebellar component of MJD is the Scale for the Assessment and Rating of Ataxia (SARA), currently considered as the worldwide standard, namely in ataxia trials (www.clinicaltrials.gov). SARA, however, needs to be complemented with other instruments that can cover non-ataxic features, such as the INAS (Inventory of Non-Ataxia Signs) [6]. In alternative, a global and comprehensive inventory for the assessment of SCAs in general and MJD in particular can be achieved with the use of the Neurological Examination Score for Spinocerebellar Ataxia (NESSCA), which has been described and validated in MJD by Kieling and colleagues [7]. This scale is based on a semiquantitative assessment of the standardized neurological examination, allowing the evaluation of ataxic and non-ataxic features, therefore providing a complete clinical profile of MJD patients. Longitudinal data obtained with NESSCA

scale, should, moreover, be highly informative to measure disease progression, as demonstrated by Jardim and colleagues [8].

The MJD gene, *ATXN3*, was mapped to 14q32.1 and the causative mutation identified as a coding unstable CAG expansion [9,10]. At the *ATXN3* locus normal alleles consensually range from 12 to 44 CAGs, whereas well-established limits for expanded alleles comprise from 61 to 87 repeats [11]. Although the size of the CAG tract is established as the main factor in the explanation of age at onset [12], the explanation of its variance is incomplete, indicating that other factors, namely genetic, should play a role in MJD's phenotype [13]. Genetic variants at several loci have been proposed as modulators of the onset of the disease; the ability to detect the effect of such genetic modifiers is dependent on the MJD cohort being analysed, since different population backgrounds display their own allelic profiles (reviewed in [14]).

Despite the considerable amount of research targeting MJD, several aspects of this disease remain challenging. Some of these interrogations are currently being addressed with the use of cell and/or animal models (reviewed in [15]). Animal models, namely rodents, present the obvious advantage of manipulation and their study remains the only facilitated way to access the affected tissues in MJD (see, amongst others [16,17]). Insights from models, however, present their own set of limitations (reviewed in [18]); therefore, for several important topics in MJD research the availability of clinical and genetic data from cohorts of patients is of major importance. Moreover, as potential treatments for spinocerebellar ataxias are being developed, the need for well-characterized trial-ready cohorts has been widely acknowledged. Neurological assessments could be of utility to create and improve instruments to be used in such interventional trials. On the other hand, the knowledge of the allelic profile at the *ATXN3* locus of each individual MJD cohort is relevant, both for the design of studies aiming to identify genetic modifiers as well as in the context of future interventional trials, being an important basis for patient's stratification.

The goal of this work was to provide a general clinical and genetic characterization of the Azorean MJD cohort. Clinical features, namely disease stage, age at onset, disease duration and NESSCA score are analysed. Longitudinal data for a subset of SCA3 patients is also provided. Moreover, *ATXN3* allelic profile and genotype-phenotype correlations were performed to characterize the Azorean MJD cohort.

SUBJECTS AND METHODS

Clinical and genetic data of MJD subjects

The Azorean MJD cohort comprised, by the 30 of September 2016, 88 clinically and molecularly confirmed patients, alive and living in the Azores archipelago. After informed consent, sixty-nine of these patients were clinically evaluated, by the same neurologist (JV), aiming to obtain the score for NESSCA at the baseline of the study. A subset of 23 SCA3 subjects was evaluated at a second visit, one or two years after baseline. After pseudonymization, clinical data was retrieved from the information available in the MJD subject's individual files at the Department of Neurology, Hospital do Divino Espirito Santo (HDES), Ponta Delgada, Azores.

Age at onset was defined as the age of the first complaints, usually gait ataxia and/or diplopia. Disease duration was the time elapsed between age at onset and the age at neurological evaluation. Using the median of disease duration, patients were further divided in two groups: patients with short (≤5 years) or long (6+ years) disease duration. Disease stage was adapted from Klockgether and colleagues [19] and defined in this study as: stage 1 - gait autonomy preserved; stage 2 – inability to walk without assistance; and stage 3 – total loss of independent gait (wheelchair).

Determination of the number of CAGs at the *ATXN3* locus was performed as described by Bettencourt and colleagues [20].

Statistical analysis

Normal distribution of continuous variables was tested using the one-sample Kolmogorov-Smirnov test. Nonparametric tests were employed for variables which were not normally distributed or whenever the sample size was small.

Linear relationship between age at evaluation, CAG number in expanded allele, age at onset, disease duration and NESSCA score was tested using Pearson correlation coefficient or Spearman rank correlation. If some covariate adjustment was needed, partial correlation test was used. Variables for which significant correlations were obtained were further used as independent variables to calculate the amount of explanation to the NESSCA score variance (linear regression). For the subset of subjects with longitudinal data available (N=23), the distribution of NESSCA score at baseline and at the second visit was compared, using the Wilcoxon signed-rank test. This test was performed either with all patients or using patients divided by one or two years of interval between evaluations. Baseline features of patients with one or two years of interval between evaluations were compared using the Mann-Whitney U test. Correlations between the size of the CAG size in the normal allele, the expanded allele, age at onset and disease duration were performed. The explanation of the age at onset variance provided by the number of CAG repeats in expanded allele was obtained using a linear regression.

Results were considered statistically significant if p-value was lower than 0.05. All statistical procedures were performed in IBM SPSS statistics 22.

RESULTS

General demographic, clinical and genetic characterization of the Azorean cohort is shown in Figure 1. The same number of females and males was evaluated and the majority of the patients are between 38 and 52 years of age. Fifty-one and 36 patients belong to families originally from São Miguel and Flores Islands, respectively. From the 88 SCA3 patients alive and living in Azores at the time of the study, 54 preserves gait autonomy – stage 1. The most frequent allele in the normal range presents 23 CAG repeats, whereas the most common allele in the pathogenic interval presents 71 repeats (Figure 1). Allelic size distribution of the normal allele as well as of the expanded allele is slightly skewed to the left (-0.580±0.258 (skewness ± standard error) and -0.288±0.257, respectively). This finding indicates that 74% and 62% of the repeats size in normal and expanded alleles, respectively, are equal or below the mode (which is of 23 repeats for the normal allele and 71 repeats for the expanded allele).

The negative correlation between CAG number in expanded allele and age at onset was calculated ($r=-0.768$, $p<0.0005$). In this cohort, 59% of the age at onset variance is explained by the CAG number in expanded allele (Figure 2; $r^2=0.590$, $p<0.0005$).

Baseline characterization of the 69 patients is shown in Table 1. The Azorean cohort data was compared with data previously reported for two Brazilian cohorts [7,8].

Average NESSCA score for the Azorean SCA3 cohort was 12±5 (mean ± standard deviation). This value was significantly lower when compared to the average of a Brazilian cohort I [7]. NESSCA score significantly correlated with CAG number in expanded allele (Pearson correlation coefficient, $r=0.244$, $p=0.043$), as well as with disease duration (Spearman rank correlation, $r=0.581$; $p<0.01$).

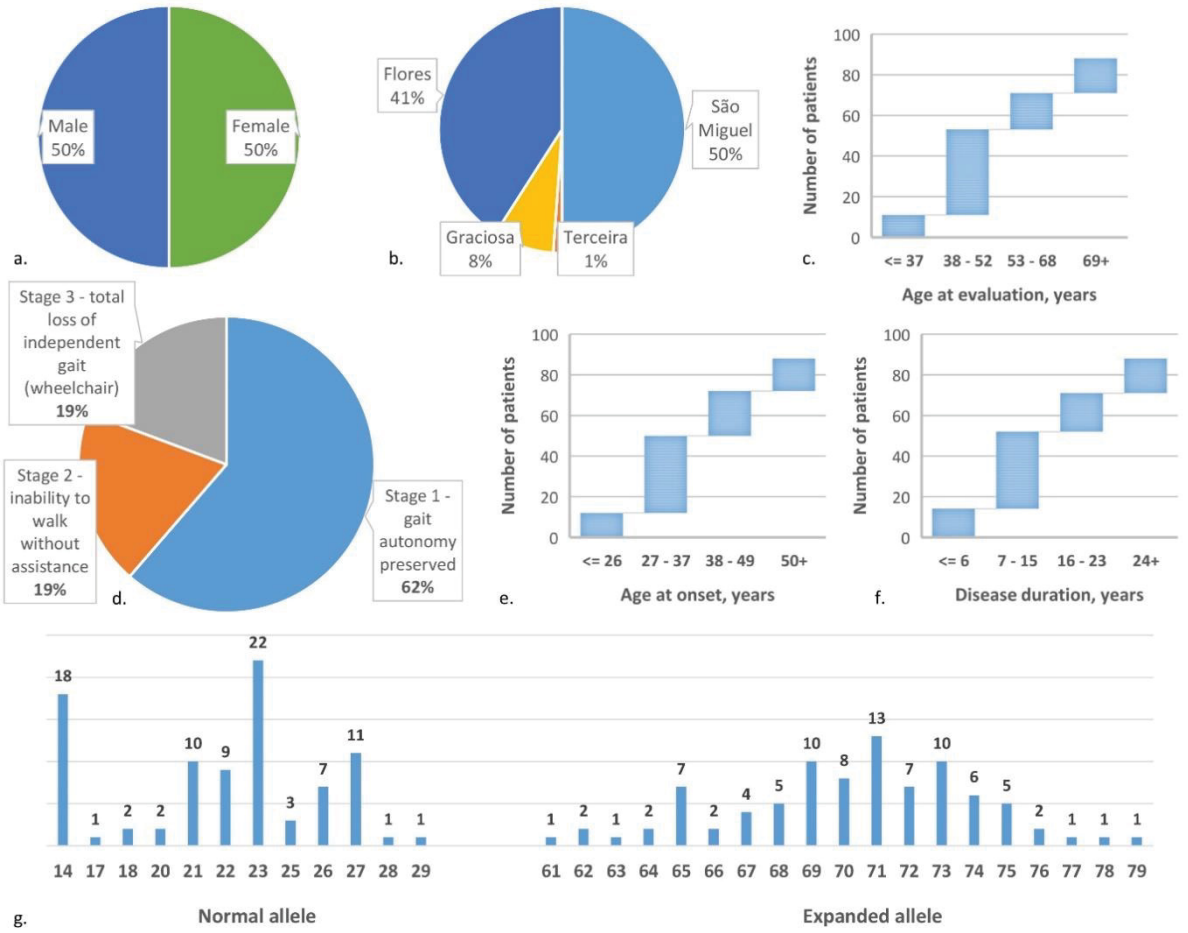


Figure 1. Characterization of Azorean MJD cohort (N=88), namely (a.) gender, (b.) island of birth, (c.) age at evaluation, (d.) disease stage, (e.) age at onset, (f.) disease duration and (g.) CAG number in normal and expanded alleles.

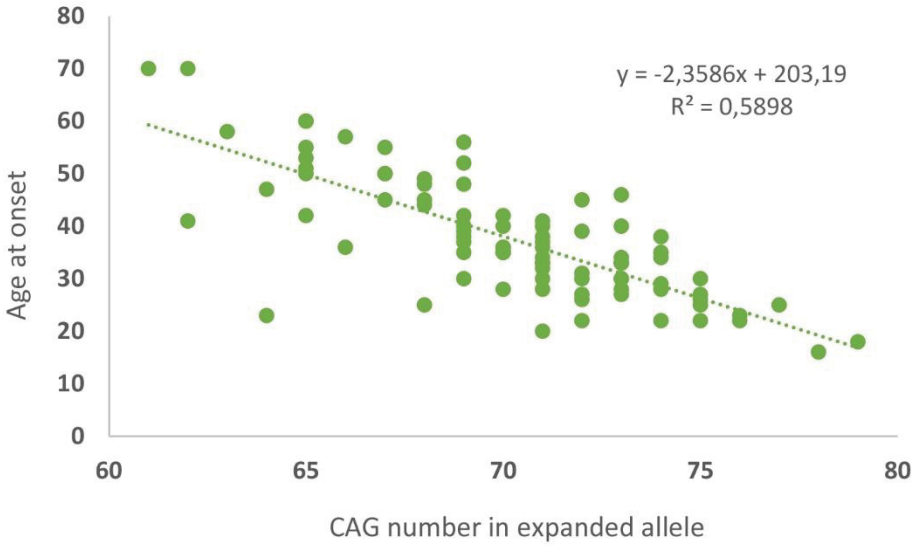


Figure 2. Linear relationship between CAG number in expanded allele and age at onset.

Table 1. Azorean SCA3 patient characteristics at the baseline of the study. Data from two Brazilian cohorts [7,8] were displayed for comparisons.

	Azorean cohort, N=69	**Brazilian cohort I [7], N=99**	**Brazilian cohort II [8], N=156**	
Gender (Female|Male)	36|33	50|49	85|71	ns
Age at evaluation, years	41 | 43 ± 12 [17-85]	41 ± 23 [9-75]	41 ± 13	p<0.05
CAG_n in expanded allele	71 | 71 ± 4 [61-78]	74 ± 2 [69-79]	74 | [67-85]	ns
Age at onset, years	35 | 36 ± 11 [16-70]	34 ± 10 [7-57]	33 ± 11	ns
Disease duration, years	5| 7 ± 6 [0-25]	8 | [0-26]	6 | [0-25]	
NESSCA score	12 | 12 ± 5 [0-24]	16 ± 6 [3-35]		p<0.05

Continuous variables, if available, are shown as median | mean ± standard deviation [minimum-maximum]; Statistical tests were performed using the OpenEpi software, version 3.03 (Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health. www.OpenEpi.com, updated in 2014/09/22). Comparisons between two variables were performed using the T-test; ns = not significant.

The relationship between the CAG number in the expanded allele and the NESSCA score increased after controlling for disease duration (partial correlation test; r=0.386, p=0.002). The CAG number in the expanded allele and disease duration significantly explained 40% of NESSCA score variance (r^2=0.403, p<0.0005; CAG repeats - beta = 0.338, p=0.001 and disease duration (log transformed) - beta= 0.593, p<0.0005). The CAG number in the expanded allele explains 11% of the variance of NESSCA score (part correlation coefficient=0.334), while disease duration contributes with 34%

(part correlation coefficient=0.586) for the explanation. SCA3 patients with short disease duration (9.55±0.68, mean ± standard error) had lower NESSCA score compared to patients with long disease duration (15.08±0.73), after controlling for CAG number in the expanded allele (70.6 repeats).

NESSCA scores for each of the 23 SCA3 patients evaluated in a second visit (one or two years after baseline) are shown in Figure 3.

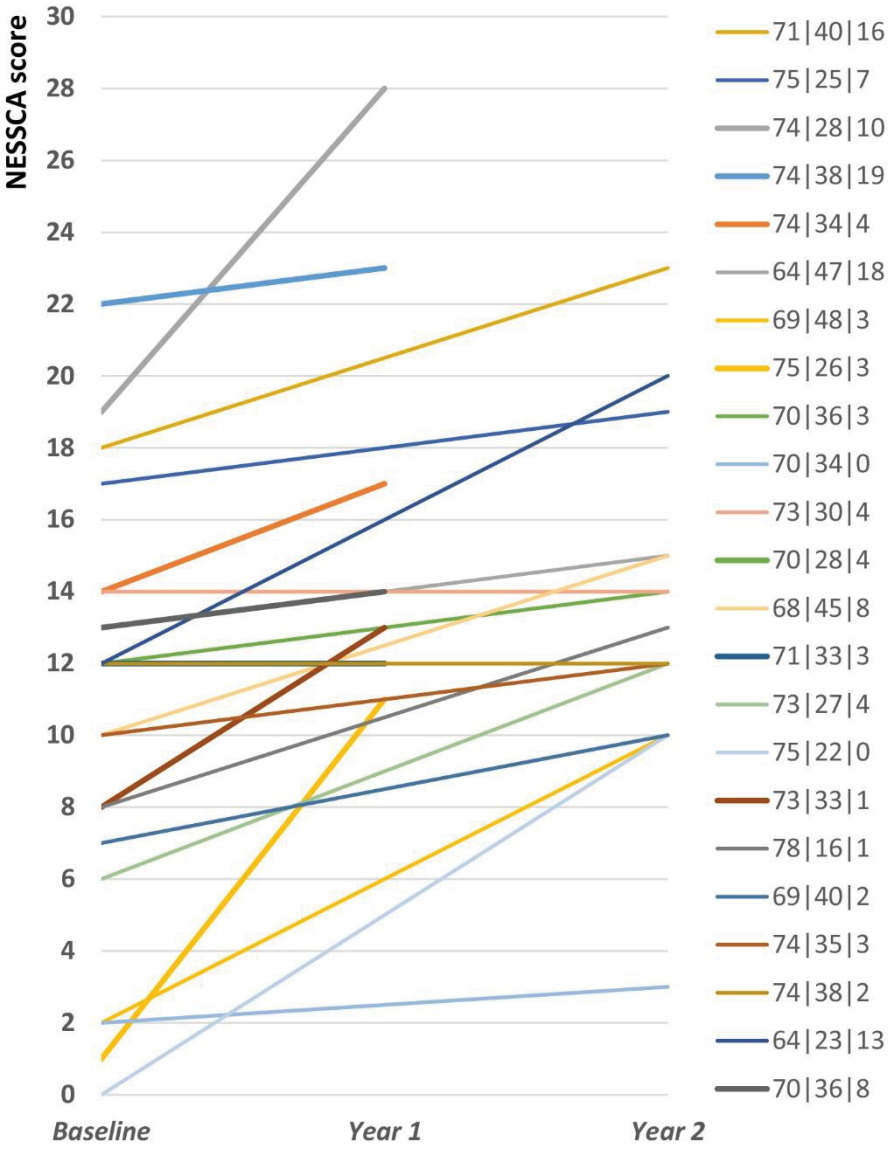


Figure 3. NESSCA score at baseline and at a second visit. CAG number of expanded allele | age at onset | disease duration is shown for each SCA3 subject under study (N=23).

One year after baseline evaluation, NESSCA score increased, on average, 2.375 ± 1.133 points (N=8; Wilcoxon signed rank: z=-2.032, p=0.042); patients with 2 years of interval between evaluations showed an increase, in average, of 3.400 ± 0.779 points (N=15; Wilcoxon signed rank: z=-3.076, p=0.002). The CAG number in the expanded allele, age at onset, disease duration and NESSCA score of MJD patients with one or two years of interval between evaluations were compared at baseline; no significant differences were obtained (Table 2). Data from the 23 Azorean patients was compared with results from the Brazilian longitudinal study (Brazilian cohort II) ([8]; Table 2); no significant differences were found.

Table 2. Description of demographic, genetic and clinical features, at baseline, of the subset of patients that underwent into the longitudinal study (N=23). Data from the total number of patients as well as data from patients which have been evaluated after 1 or 2 years of interval from baseline assessment were shown. Data from a Brazilian longitudinal study are also described.

	Azorean sub-cohort, N=23			**Brazilian cohort [8], N=105**
	Total	1 Year, N=8	2 Years, N=15	
Gender (Female|Male)	14|9	6|2	8|7	57|49
Age at evaluation, years	38 | 39 ± 11 [17-65]	37 | 39 ± 8 [32-57]	38 | 40 ± 13 [17-65]	41 ± 14
CAG_n in expanded allele	73 | 72 ± 4 [64-78]	74 | 73 ± 2 [70-75]	71 | 71 ± 4 [64-78]	74 | [67-82]
Age at onset, years	34 | 33 ± 8 [16-48]	33 | 32 ± 4 [26-38]	35 | 34 ± 10 [16-48]	34 ± 10
Disease duration, years	4 | 6 ± 6 [0-19]	6| 7 ± 6 [1-19]	4 | 6 ± 6 [0-18]	11 | [3-28]
NESSCA score	13 | 14 ± 5 [3-28]	12 | 14 ± 5 [8-22]	10 | 10 ± 5 [0-18]	*

Continuous variables, if available, are shown as median | mean ± standard deviation [minimum-maximum];
*information not available.

DISCUSSION

Globally, MJD is considered as a rare disorder, although regional clusters have been reported. Homogeneous groups of patients, such as the Azorean cohort [21], described in this paper, are expected to play an important role in the understanding of the disease, being particularly relevant

in the response to emerging interventional trials, following reports of promising preclinical data [22,23].

In the Azorean cohort, a clinical rating characterization with a validated scale was still lacking. The analysis of NESSCA data indicates that predictors with higher impact on the NESSCA score are the CAG size in the expanded allele and disease duration, which together explain 40% of the score variance. NESSCA score was significantly higher in the Brazilian cohort I [7] compared to the Azorean cohort. As CAG size in expanded alleles is one of the predictors of NESSCA score variance, we can speculate that because Brazilian patients, in average, present longer CAG repeats in the expanded allele (altough differences between CAG numbers from each cohort are not statistically significant), it is expected that the NESSCA score should be higher (Table 1). Azorean MJD patients presented an increase, in average, of two or three points after one or two years of interval between evaluations, respectively. Clinical scores obtained can be further used for correlations with data from molecular biomarkers studies, which are currently ongoing [24,25].

The age at onset variance explained by the number of CAG units in the expanded allele in the Azorean MJD cohort is the highest reported in Europe [4,26–28]. Differences in variance explanation could have implications in the identification of genetic modifiers. Incompleteness in age at onset explanation could be caused by the unconsistent report of age at onset by the patient and/or caregiver. In fact, this aspect was discussed previously by Gloubas and colleagues, which proposed instructions to better define age at onset [28]. In many cases, unreliable information concerning age at onset may be solved studying homogeneous cohorts, such as the Azorean, allowing prospective studies of disease onset.

As a rare disorder worldwide, the low number of MJD patients available could compromise studies concerning several aspects of the disease. This fact would be partially solved with the knowledge of clinical and genetic features of specific MJD clusters which could allow: (1) the acquisition of

longitudinal data, (2) a better understanding of disease pathogenesis, (3) the improvement of the design of interventional trials, and (4) the validation of molecular biomarkers.

REFERENCES

1. Durr A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. Lancet Neurol. 2010;9:885–94.

2. de Araújo M, Raposo M, Kazachkova N, Vasconcelos J, Kay T, Lima M. Trends in the Epidemiology of Spinocerebellar Ataxia Type 3/Machado-Joseph Disease in the Azores Islands, Portugal. JSM Brain Sci. 2016;1:1001.

3. Carvalho DR, La Rocque-Ferreira A, Rizzo IM, Imamura EU, Speck-Martins CE. Homozygosity Enhances Severity in Spinocerebellar Ataxia Type 3. Pediatr. Neurol. 2008;38:296–9.

4. Tezenas du Montcel S, Durr A, Bauer P, Figueroa K, Ichikawa Y, Brussino A, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain. 2014;137:2444–55.

5. Lima L, Coutinho P. Clinical criteria for diagnosis of Machado-Joseph disease: report of a non-Azorena Portuguese family. Neurology. 1980;30:319–22.

6. Jacobi H, Rakowicz M, Rola R, Fancellu R, Mariotti C, Charles P, et al. Inventory of Non-Ataxia Signs (INAS): validation of a new clinical assessment instrument. Cerebellum. 2013;12:418–28.

7. Kieling C, Rieder CRM, Silva ACF, Saute JAM, Cecchin CR, Monte TL, et al. A neurological examination score for the assessment of spinocerebellar ataxia 3 (SCA3). Eur. J. Neurol. 2008;15:371–6.

8. Jardim LB, Hauser L, Kieling C, Saute JAM, Xavier R, Rieder CRM, et al. Progression rate of neurological deficits in a 10-year cohort of SCA3 patients. Cerebellum. 2010;9:419–28.

9. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat. Genet. 1994;8:221–8.

10. Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, et al. The gene for Machado-Joseph disease maps to human chromosome 14q. Nat. Genet. 1993;4:300–4.

11. Maciel P, Costa MC, Ferro A, Rousseau M, Santos CS, Gaspar C, et al. Improvement in the molecular diagnosis of Machado-Joseph disease. Arch. Neurol. 2001;58:1821–7.

12. Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. Am. J. Hum. Genet. 1995;57:54–61.

13. DeStefano AL, Cupples LA, Maciel P, Gaspar C, Radvany J, Dawson DM, et al. A familial factor independent of CAG repeat length influences age at onset of Machado-Joseph disease. Am. J.

Hum. Genet. 1996;59:119–27.

14. Raposo M, Ramos A, Bettencourt C, Lima M. Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid? Brain. 2015;138:e398.

15. Colomer Gould VF. Mouse models of spinocerebellar ataxia type 3 (Machado-Joseph disease). Neurotherapeutics. 2012;9:285–96.

16. Kazachkova N, Raposo M, Montiel R, Cymbron T, Bettencourt C, Silva-Fernandes A, et al. Patterns of mitochondrial DNA damage in blood and brain tissues of a transgenic mouse model of Machado-Joseph disease. Neurodegener. Dis. 2013;11:206–14.

17. Ramos A, Kazachkova N, Silva F, Maciel P, Silva-Fernandes A, Duarte-Silva S, et al. Differential mtDNA damage patterns in a transgenic mouse model of Machado-Joseph disease (MJD/SCA3). J. Mol. Neurosci. 2015;55:449–53.

18. Cendelin J. From mice to men: lessons from mutant ataxic mice. Cerebellum & ataxias. 2014;1:4.

19. Klockgether T, Lüdtke R, Kramer B, Abele M, Bürk K, Schöls L, et al. The natural history of degenerative ataxia: a retrospective study in 466 patients. Brain. 1998;121 (Pt 4:589–600.

20. Bettencourt C, Fialho RN, Santos C, Montiel R, Bruges-Armas J, MacIel P, et al. Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: A study in normal families from the Azores islands (Portugal). J. Hum. Genet. 2008;53:333–9.

21. Raposo M, Ramos A, Bettencourt C, Lima M. Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid? Brain. 2015;awv206.

22. Cunha-Santos J, Duarte-Neves J, Carmona V, Guarente L, Pereira de Almeida L, Cavadas C. Caloric restriction blocks neuropathology and motor deficits in Machado-Joseph disease mouse models through SIRT1 pathway. Nat. Commun. 2016;7:11445.

23. Teixeira-Castro A, Jalles A, Esteves S, Kang S, da Silva Santos L, Silva-Fernandes A, et al. Serotonergic signalling suppresses ataxin 3 aggregation and neurotoxicity in animal models of Machado-Joseph disease. Brain. 2015;138:3221–37.

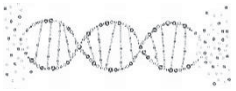
24. Raposo M, Bettencourt C, Maciel P, Gao F, Ramos A, Kazachkova N, et al. Novel candidate blood-based transcriptional biomarkers of Machado-Joseph disease. Mov. Disord. 2015;30:968–75.

25. da Silva Carvalho G, Saute J, Haas C, Torrez V, Brochier A, Souza G, et al. Cytokines in Machado Joseph Disease/Spinocerebellar Ataxia 3. Cerebellum. 2015;

26. Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. Am J Hum Genet. 1995;57:54–61.

27. van de Warrenburg BPC, Hendriks H, Dürr A, van Zuijlen MCA, Stevanin G, Camuzat A, et al. Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort. Ann. Neurol. 2005;57:505–12.

28. Globas C, du Montcel ST, Baliko L, Boesch S, Depondt C, DiDonato S, et al. Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. Mov. Disord. 2008;23:2232–8.



CLINICAL BIOMARKERS

CHAPTER II

Nystagmus as an early ocular alteration in Machado-Joseph disease (MJD/SCA3).


RESEARCH ARTICLE Open Access

Nystagmus as an early ocular alteration in Machado-Joseph disease (MJD/SCA3)

BMC Neurology

Journal subject category: Clinical Neurology

Journal rank and quartile: 113/192 – Q3

Impact factor 2014: 2.040

5-year impact factor 2014: 2.568

Mafalda Raposo[1,2], João Vasconcelos[3], Conceição Bettencourt[2,4], Teresa Kay[5], Paula Coutinho[6], Manuela Lima[1,2]

1. Centre of Research in Natural Resources (CIRN), University of the Azores, Ponta Delgada, Portugal
2. Institute for Molecular and Cellular Biology (IBMC), University of Porto, Porto, Portugal
3. Department of Neurology, Hospital do Divino Espírito Santo, Ponta Delgada, Portugal
4. Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
5. Department of Clinical Genetics, Hospital D. Estefania, Lisbon, Portugal
6. CGPP, Institute for Molecular and Cellular Biology (IBMC), University of Porto, Porto, Portugal

ABSTRACT

Background: Machado-Joseph disease (MJD), also named spinocerebellar ataxia type 3 (SCA3) is the most common autosomal dominant ataxia worldwide. Although nystagmus is one of the most frequently reported ocular alterations in MJD patients its behaviour during the course of the disease, namely in its early stages, has recently started to be investigated. The main goal of this work was to characterize the frequency of nystagmus in symptomatic and presymptomatic carriers of the MJD mutation, and investigate its usefulness as an early indicator of the disease.

Methods: We conducted an observational study of Azorean MJD family members, comprising a total of 158 subjects which underwent neurological evaluation. Sixty-eight were clinically and molecularly diagnosed with MJD, 48 were confirmed asymptomatic carriers and 42 were confirmed non-carriers of the MJD mutation. The frequency of nystagmus was calculated for the 3 groups.

Results: Nystagmus was present in 88% of the MJD patients. Seventeen percent of the at-risk subjects with a carrier result in the molecular test and none of the 42 individuals who received a non-carrier test result displayed nystagmus ($p<0.006$). Although not reaching statistical significance, symptomatic subjects showing nystagmus had a tendency for a higher length of the CAG tract in the expanded allele, when compared to individuals who did not have nystagmus.

Conclusions: The frequency of nystagmus in asymptomatic carriers and its absence in non-carriers of the mutation, suggests that nystagmus may appear before gait disturbance and can thus be considered an early sign of MJD.

BACKGROUND

Machado-Joseph disease (MJD; MIM #109150; ORPHA98757), also known as spinocerebellar ataxia type 3 (SCA3), is the most frequent autosomal dominant ataxia worldwide [1]. The MJD gene (*ATXN3*) was mapped to 14q32.1 and the causative mutation was identified as a coding unstable CAG expansion [2,3]; normal alleles consensually range from 12 to 44 CAGs, whereas well established limits for expanded alleles comprise from 61 to 87 repeats [4].

With an average onset in adulthood (around 40 years), MJD is characterized by a wide range of clinical manifestations, including ataxia, progressive external ophtalmoplegia, pyramidal and extrapyramidal signs, dystonia with rigidity and distal muscular atrophies [5,6]. Gait ataxia is reported as the initial complaint in the majority of MJD patients [7]; in some of the cases, other symptoms, such as diplopia have been reported as preceding gait instability [8]. Ocular alterations are clinical features commonly described in SCAs, occurring as a consequence of cerebellar and brainstem degeneration [9]. Amongst such alterations is nystagmus, a clinical sign associated with vestibular nuclei, found to be damaged in MJD post-mortem brains [10]. Nystagmus is one of the most frequently reported ocular alterations in MJD patients [7,11,12]; its behaviour during the course of the disease, namely in its early stages, has recently started to be investigated [13].

The main goal of this work was to characterize the frequency of nystagmus in symptomatic and presymptomatic carriers of the MJD mutation and investigate its usefulness as an early indicator of the disease. The availability of molecular testing to detect asymptomatic carriers of the MJD mutation, associated with biological and/or refined clinical markers of the disease process, would enable an early intervention with prophylactic treatment, which should lead to prevention, or at least delayed onset or slowed progression, by allowing intervention before the appearance of motor symptoms [14].

SUBJECTS AND METHODS

Subjects

From 1996 to 2011, 158 subjects belonging to 21 Azorean MJD families performed the molecular test for MJD, either in the context of molecular diagnosis (N=68) or predictive testing (PT) (N=90). Participants were asked for written informed consent to participate in research concerning the analysis of the clinical variability of MJD. During this period the molecular tests have been conducted by two distinct Portuguese genetic diagnosis laboratories.

Three main groups were defined for this study:

Group 1 - subjects with a clinical and molecular diagnosis of MJD (N=68). This group corresponds to individuals that were already symptomatic at the time of their first observation, and for which a molecular confirmation of MJD was subsequently obtained; Group 2 - subjects with a normal neurological evaluation, who were confirmed as carriers of the *ATXN3* mutation in the molecular test (N=48). Group 2 was therefore formed by subjects that were submitted to a neurological examination upon entering the PT program (before taking the molecular test) and were asymptomatic at that time. Some of the subjects from this group developed motor symptoms during the observational period of the study. For genotype-phenotype analysis these subjects were clustered with patients from group 1, constituting the pool of “symptomatic subjects” (N=87). Group 3 - subjects with a normal pre-test clinical evaluation and that subsequently received a negative result in the molecular test (non-carriers of the *ATXN3* mutation) (N=42).

This study is part of a larger project which was approved by the Ethics Committee of the Hospital do Divíno Espírito Santo (São Miguel Island, Azores – Portugal).

Clinical features

Neurological examination was performed by two experienced neurologists, following the protocol established by Coutinho [7]. The examination included the evaluation of nystagmus, which consists in small, horizontal, saccade-like movements that lead the eye away from the target trajectory and, after a delay, bring it back onto the target [9,15]. The evaluation of nystagmus followed the ICARS procedure [16]. The age at onset and the duration of disease were also assessed. Onset corresponded to the age of appearance of the first symptoms (such as gait disturbance) reported by the patient and/or a close relative. Disease duration was calculated as the time elapsed between the age at onset and the last neurological evaluation available. Time from onset was calculated as the difference between age at first neurological observation and reported age at ataxia onset.

Statistical analysis

Differences in the frequency of nystagmus among the three study groups were measured by the Fisher's Exact Test. Unrelated subjects (N=46) from the pool of symptomatic subjects were analysed for the following variables: CAG length of the normal and expanded allele, age at onset and disease duration. Pearson correlation test was performed to evaluate the relationship between the CAG repeat length of normal and expanded alleles and age at onset. Univariate and multivariate linear regression were used to test the effect of several predictors on age at onset, namely: the number of CAGs in expanded and normal alleles, gender, island of origin, molecular diagnosis laboratory, and presence/absence of nystagmus. The best model was used to predict the age at onset for all members of group 2. The results were accepted as statistically significant whenever $p<0.05$. All analyses were performed with PASW Statistics 18 [17].

RESULTS

Frequency of nystagmus in symptomatic and presymptomatic MJD carriers

For the 158 Azorean subjects included in the study, information on gender, age and size of the CAG tract of MJD alleles is displayed in Table 1. Mean age is higher in group 1 (48±15 years), than in the other 2 studied groups (30±9 and 36±12, respectively). The number of CAG repeats in the expanded allele is 71±4 (mean ± standard deviation) in both group 1 and group 2 (Table 1). Nystagmus was present in 88% of the MJD patients (group 1) and 17% of the at-risk subjects with a carrier result in the molecular test (group 2). However, nystagmus was not observed in any of the 42 individuals who received a non-carrier test result (group 3), which resulted in significant differences between groups 2 and 3 (p=0.006).

Table 1. Characterization of the studied subjects according to group assignment.

		N	Group 1	N	Group 2	N	Group 3	
	Gender	68	33♀|35♂	48	29♀|19♂	42	26♀|17♂	
	Age	68	48±15 [16-82]	48	30±9 [18-59]	42	36±12 [18-60]	
CAG repeats	**Normal allele**	60	20±5 [13-28]	44	22±5 [13-27]	42	17±4 [13-27]	23±4 [13-31]
	Expanded allele	64	71±4 [61-80]	46	71±4 [63-80]			

All variables were presented as mean ± standard deviation [range]. Gender was represented as ♀ for female and ♂ for male.

From the 68 patients in group 1, only eight (12%) failed to present nystagmus at the first neurological examination; five out of these eight, however, presented nystagmus at the last observation performed in the course of this study, increasing to 96% the number of patients who presented nystagmus. The remaining three patients were lost from follow-up, therefore preventing the verification of the presence of nystagmus at a later stage of the disease.

From the group of 48 asymptomatic carriers (group 2), 19 (40%) developed gait disturbance within the period of this study. From these 19, five already presented nystagmus at PT evaluation and developed gait ataxia during the follow-up period; the remaining three were not re-evaluated

(Figure 1). Fourteen individuals from group 2 also developed gait ataxia during this period (Figure 1); for such subjects although, in the respective PT evaluation nystagmus was absent, it was present at the latest observation performed. Individuals which presents nystagmus at PT evaluation were closest to reported age at onset (median of 2 years) than individuals without nystagmus (median of 3 years).

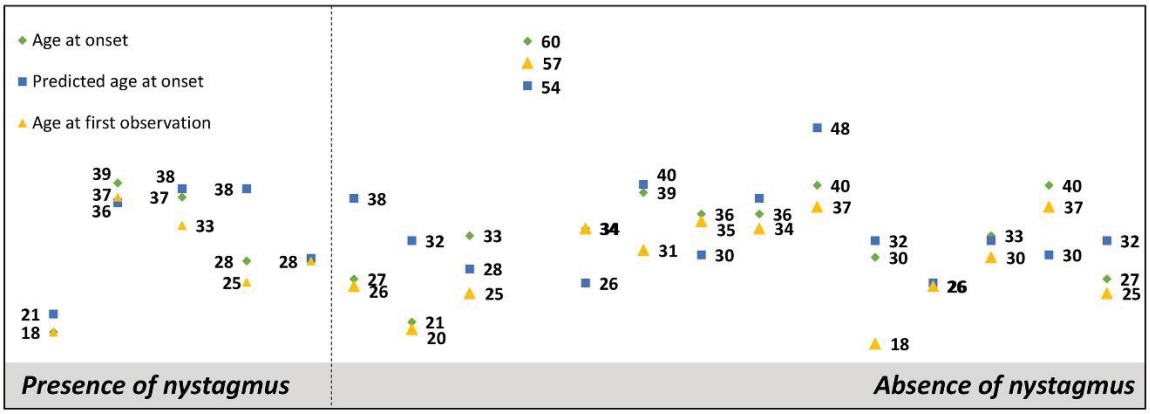


Figure 1. Age at onset, predicted age at onset and present age was displayed in individuals from group 2, which during this study developed the disease, divided by presence (N=5) and absence of nystagmus (N=14).

Genotype-phenotype correlations

The 87 symptomatic carriers were clinically characterized, showing a mean value for age at onset of 39±12 years (mean ± standard deviation) and disease duration of 15±9 years. All subjects displayed cerebellar signs; moreover, 65% presented mostly corticospinal dysfunction, followed by mostly peripheral (26%) and extrapyramidal (6%) manifestations (Figure 2).

Using information on unrelated symptomatic subjects with available CAG data, a negative correlation between the size of the expanded allele and the age at onset was observed (N=42, r=-0.667, p<0.0005). The explanation of the age at onset variance provided by the CAG length in expanded allele was 45% (t=-5.740, p<0.0005). An improvement of the model using additional predictors (CAG repeat number in normal allele, gender, island of origin and genotyping laboratory) raised the explanation of the age at onset to 48% (p<0.0005).

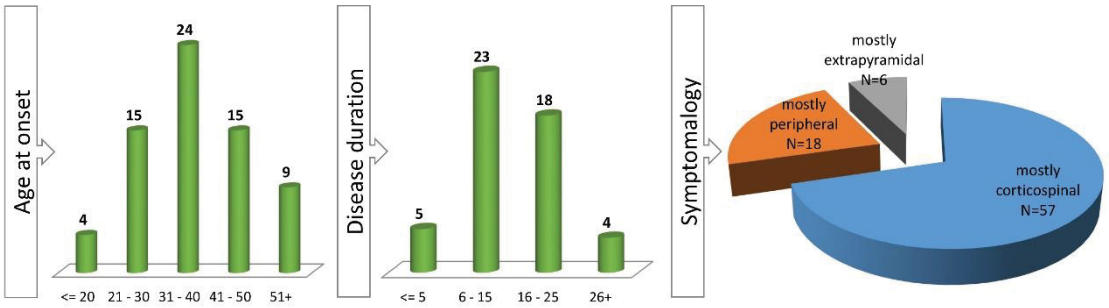


Figure 2. Neurological features of MJD patients, such as age at onset (N=86), disease duration (N=81) and symptomatology (N=68). Cerebellar dysfunction was observed in all patients and therefore they were distinguished also by dysfunction in other systems: mostly corticospinal, mostly extrapyramidal or mostly peripheral signs.

Individuals showing nystagmus had a tendency for a higher CAG length in expanded allele, when compared to individuals who did not have nystagmus [71.54±0.67 and 69.71±0.97 (mean ± standard error), respectively]; this tendency, however, failed to reach significance. The presence/absence of nystagmus did not improve significantly the prediction of the age at onset when adding this variable to the number of CAG repeats in the expanded allele as well as the other predictors previously mentioned.

DISCUSSION

The present study confirmed nystagmus as a very frequent sign in MJD patients (up to 96% of the MJD patients from the studied series), which is in agreement with previous studies where this sign has been reported with frequencies that vary between 55 and 92% [7,8,18,19]. Considering the genotype-phenotype results, a long term follow-up study would be needed to understand if there is a correlation between CAG repeat length in expanded allele and the age of appearance of nystagmus.

The present study further shows that nystagmus is also observed in carriers of the MJD mutation, before the manifestation of gait disturbance, indicating that it can be an early sign of the disease,

which should be monitored in mutation carriers even before motor signs appear. Recently a higher rate of horizontal gaze-evoked nystagmus in MJD carriers than non-carriers (ten [39%] of 26 vs one [5%] of 20; p=0.013) has been described [13]. From ataxia clinical scales currently being applied, the complementary use of one that includes the evaluation of nystagmus, such as ICARS [16] or NESSCA [20] is important, to avoid the under-estimation of this clinical sign and the failure to detect early alterations. In other SCAs, early detection of impaired eye movements in presymptomatic individuals has been a matter of study. Slowing of horizontal saccades was observed in SCA1, 2, 6 and 7 [19,21–24] whereas a decrease in pursuit gain was found in SCA6 [25]. The present study is based in the qualitative analysis of nystagmus, i.e., presence or absence of this clinical sign; although reasonable information about oculomotor movements might be obtained with such qualitative measure, the use of semi-quantitative techniques such as oculographic evaluations, should improve the characterization of nystagmus types as well as the study of other ocular features, which might help to further understand the importance of eye abnormalities in early stages of MJD.

CONCLUSIONS

In summary, the presence of nystagmus in asymptomatic MJD carriers suggests that the detection of nystagmus may be useful in the identification of early stages of MJD, potentially facilitating the enrolment of presymptomatic individuals in future preventive clinical trials.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

MR and ML participated in the design of the study and drafted the manuscript. JV and TK were involved in the collection of the clinical data. CB and PC participated in the discussion of the study and in the drafted of the manuscript. JV and ML participated in coordination of the study. All authors read and approved the final manuscript.

AKNOWLEDGMENTS

This study was supported by a grant PTDC/DTP-PIC/0370/2012 from Fundação para a Ciência e Tecnologia (FCT) and by a PhD fellowship M3.1.2/F/006/2011 (MR) from Fundo Regional para a Ciência (FRC).

REFERENCES

1. Bettencourt C, Lima M. Machado-Joseph Disease: from first descriptions to new perspectives. Orphanet J. Rare Dis. 2011;6:35.

2. Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, et al. The gene for Machado-Joseph disease maps to human chromosome 14q. Nat. Genet. 1993;4:300–4.

3. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32. 1. Nat. Genet. 1994;8:221–8.

4. Maciel P, Costa MC, Ferro A, Rousseau M, Santos CS, Gaspar C, et al. Improvement in the molecular diagnosis of Machado-Joseph disease. Arch. Neurol. 2001;58:1821–7.

5. Coutinho P, Andrade C. 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. 1978;28:703–9.

6. Lima L, Coutinho P. Clinical criteria for diagnosis of Machado-Joseph disease: report of a non-Azorena Portuguese family. Neurology. 1980;30:319–22.

7. Coutinho P. Doença de Machado-Joseph: Tentativa de definição. Universidade do Porto; 1992.

8. Globas C, du Montcel ST, Baliko L, Boesch S, Depondt C, DiDonato S, et al. Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. Mov. Disord. 2008;23:2232–8.

9. Rodríguez-Labrada R, Pérez-Velázquez L. Eye movement abnormalities in Spinocerebellar

ataxias. In: Gazulla J, editor. Spinocerebellar Ataxia. Intech; 2012. p. 1–198.

10. Sequeiros J, Coutinho P. Epidemiology and clinical aspects of Machado-Joseph disease. Adv. Neurol. 1993;61:139–53.

11. Vale J, Bugalho P, Silveira I, Sequeiros J, Guimarães J, Coutinho P. Autosomal dominant cerebellar ataxia: frequency analysis and clinical characterization of 45 families from Portugal. Eur. J. Neurol. 2010;17:124–8.

12. Jardim LB, Pereira ML, Silveira I, Ferro A, Sequeiros J, Giugliani R. Neurologic findings in Machado-Joseph disease: relation with disease duration, subtypes, and (CAG)n. Arch. Neurol. 2001;58:899–904.

13. Jacobi H, Reetz K, du Montcel ST, Bauer P, Mariotti C, Nanetti L, et al. Biological and clinical characteristics of individuals at risk for spinocerebellar ataxia types 1, 2, 3, and 6 in the longitudinal RISCA study: analysis of baseline data. Lancet Neurol. 2013;12:650–8.

14. Paulsen JS, Hayden M, Stout JC, Langbehn DR, Aylward E, Ross CA, et al. Preparing for preventive clinical trials: the Predict-HD study. Arch. Neurol. 2006;63:883–90.

15. Leigh R, Zee D. The neurology of eye movements. 4th ed. New York: Oxford University Press; 2006.

16. Trouillas P, Takayanagi T, Hallett M, Currier RD, Subramony SH, Wessel K, et al. International Cooperative Ataxia Rating Scale for pharmacological assessment of the cerebellar syndrome. The Ataxia Neuropharmacology Committee of the World Federation of Neurology. J. Neurol. Sci. 1997;145:205–11.

17. SPSS_Inc. PASW Statistics for Windows. Chicago: SPSS Inc.;

18. Rivaud-Pechoux S, Dürr A, Gaymard B, Cancel G, Ploner CJ, Agid Y, et al. Eye movement abnormalities correlate with genotype in autosomal dominant cerebellar ataxia type I. Ann. Neurol. 1998;43:297–302.

19. Bürk K, Fetter M, Abele M, Laccone F, Brice a, Dichgans J, et al. Autosomal dominant cerebellar ataxia type I: oculomotor abnormalities in families with SCA1, SCA2, and SCA3. J. Neurol. 1999;246:789–97.

20. Kieling C, Rieder CRM, Silva ACF, Saute JAM, Cecchin CR, Monte TL, et al. A neurological examination score for the assessment of spinocerebellar ataxia 3 (SCA3). Eur. J. Neurol. 2008;15:371–6.

21. Klostermann W, Zühlke C, Heide W, Kömpf D, Wessel K. Slow saccades and other eye movement disorders in spinocerebellar atrophy type 1. J. Neurol. 1997;244:105–11.

22. Seifried C, Velázquez-Pérez L, Santos-Falcón N, Abele M, Ziemann U, Almaguer LE, et al. Saccade velocity as a surrogate disease marker in spinocerebellar ataxia type 2. Ann. N. Y. Acad. Sci. 2005;1039:524–7.

23. Christova P, Anderson JH, Gomez CM. Impaired eye movements in presymptomatic spinocerebellar ataxia type 6. Arch. Neurol. 2008;65:530–6.

24. Oh AK, Jacobson KM, Jen JC, Baloh RW. Slowing of voluntary and involuntary saccades: an early sign in spinocerebellar ataxia type 7. Ann. Neurol. 2001;49:801–4.

25. Takeichi N, Fukushima K, Sasaki H, Yabe I, Tashiro K, Inuyama Y. Dissociation of smooth pursuit and vestibulo-ocular reflex cancellation in SCA-6. Neurology. 2000;54:860–6.

MOLECULAR TRAIT BIOMARKERS

CHAPTER III

Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid?

Brain

Journal subject categories: Clinical Neurology and Neuroscience

Journals rank and quartile: 5/192 – Q1; 12/256 – Q1

Impact factor 2015: 10.103

5-year impact factor 2015: 10.545

Mafalda Raposo[1, 2, 3], Amanda Ramos[1, 2, 3], Conceição Bettencourt[4] & Manuela Lima[1, 2, 3]

1. Centre of Research in Natural Resources (CIRN), University of the Azores, Ponta Delgada, Portugal
2. Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
3. Institute for Molecular and Cell Biology (IBMC), University of Porto, Porto, Portugal
4. Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK

Sir,

We read with interest the paper by Tezenas du Montcel and colleagues [1], which identified modifiers of the age at onset in spinocerebellar ataxias (SCAs). In polyglutamine (polyQ) SCAs, the size of the CAG expansion is incompletely correlated with the age at onset of the disease, suggesting the involvement of additional genetic factors. The identification of modifier genes for polyQ SCAs should not only contribute to more precise onset predictions, of relevance for genetic counseling, but also unveil aspects of the underlying pathogenesis, and pinpoint possible therapeutic targets [2]. Proposed modifiers for SCAs have often failed to be validated in independent studies, a fact that has been mostly attributed to the lack of statistical power (as discussed, for example, in [3]). The report by Tezenas du Montcel and colleagues [1] represents therefore an important contribution to the problematic of modifier genes in polyQ SCAs; this study includes the largest cohorts of patients for SCA1, 2, 3, 6, 7, 17 and dentato-rubro-pallydoluysian atrophy (DRPLA) analyzed so far, which were recruited under the scope of the European Consortium on Spinocerebellar Ataxias (EUROSCA). Moreover, main findings from the EUROSCA series were tested in four independent smaller cohorts (three Caucasian cohorts from the USA, France and Italy, as well as one Asian from Japan).

SCA3, also named Machado-Joseph disease, is the most frequent polyQ SCA, reaching in the Azores Islands (Portugal), the highest value of prevalence worldwide [4]. This particular epidemiological situation has allowed us to thoroughly follow a homogeneous SCA3 cohort, currently constituted by 93 patients (Table 1) belonging to extensively studied Azorean pedigrees. The Azorean population itself has been relatively isolated and is derived from a limited number of founders, factors which have shaped its genetic landscape [5–9]. Moreover, these islands have a centralized health system with a regional program specifically for SCA3, which has promoted the follow up of all patients by the same neurologist, minimizing inter observational error, and facilitating accurate annotation of clinical data, namely the age at onset.

Table 1. Characterization of the Azorean SCA3 cohort, and comparisons with EUROSCA and replication cohorts [1].

	AZORES	EUROSCA	USA	JAPAN	FRANCE
N	93	403	110	126	44
(CAG)n length in expanded allele	71±4 [62-79]	68±4 [47-77]	70±4 [61-79]	73±5 [60-86]	68±5 [53-75]
Statistical differences		*t=-6.520; **p<0.001***	*t=-1.775; p=0.08*	*t=3.170; **p=0.002***	*t=-3.775; **p<0.001***
(CAG)n length in normal allele	22±5 [14-29]	21±5 [12-35]	22±6 [14-36]	21±7 [14-34]	22±6 [14-37]
Statistical differences		*t=-1.738; p=0.08*	*t=0; p=0.999*	*t=-1.233; p=0.220*	*t=0; p=0.999*
Age at onset, years	38±12 [13-71]	40±12 [10-78]	40±11 [15-68]	41±15 [12-84]	44±12 [15-67]
Statistical differences		*t=1.822; p=0.07*	*t=1.574; p=0.12*	*t=1.874; p=0.05*	*t=3.029; **p=0.003***

	AZORES cohort	EUROSCA cohort
Model with expanded allele		
Linear	-0.032±0.002 (y=3.810 - 0.032x) r=-0.826, p<0.001 r^2=0.682, p<0.001 adjusted r^2=0.679	-0.024±0.001 r^2=0.504, p<0.001 adjusted r^2=0.503
+ Normal allele	r^2=0.687, p=0.272	p=0.437
Quadratic	$-1.0x10^{-3}\pm0.1x10^{-3}$ (y=-0.699 + 0.096x – 0.001x^2) r=-0.833, p<0.001 r^2=0.694, p<0.001 adjusted r^2=0.687	$-1.4x10^{-3}\pm0.1x10^{-3}$ r^2=0.594, p<0.001 adjusted r^2=0.592

All quantitative variables were represented as mean ± standard deviation [minimum-maximum]. An independent t-test (2-tailed) was conducted to compare CAG repeats size in normal and expanded alleles, as well as age at onset for the different cohorts. Statistical analysis was performed in OpenEpi software (Dean AG *et al.*, 2014) and IBM SPSS Statistics, version 22 (IBM Corp. Released 2013). A p-value lower than 0.05 was considered significant.

We tested the findings of Tezenas du Montcel and colleagues [1] relatively to SCA3. To do so, molecularly confirmed Azorean SCA3 patients were genotyped for SCA1, 2, 6, 7, DRPLA and Huntington disease (HD) loci (all patients were genotyped in the same laboratory; primers and conditions are available upon request). In the Azorean cohort, we observe a higher mean number

of CAGs in the expanded allele with an accordingly earlier age at onset, when compared to the EUROSCA as well as to the French cohorts (Table 1). The opposite significant result is observed when we compare Azorean and Japanese SCA3 patients, the last displaying a higher repeat length. Although average CAG length is higher in the Japanese patients, the onset is significantly delayed comparatively to ours (Table 1). In their study, Tezenas du Montcel and colleagues [1] found significant quadratic effects when estimating the variance in the explanation of onset provided by the expanded CAG tract size (improvement of 9% relatively to a linear model). Noteworthy, since Azorean SCA3 patients have a higher mean number of CAGs in the expanded allele, and larger alleles are more strongly correlated with onset than smaller expanded ones, we only obtain a modest improvement of 1.2% (Table 1). All the previous observations indicate the presence of differences on the profile and behavior of the expanded alleles in SCA3, which seem to be population dependent.

In the EUROSCA cohort of SCA3, Tezenas du Montcel and colleagues [1] failed to observe an association between onset and the size of the normal allele, which had previously been reported in other large studies [12,13]; this lack of association was also evidenced in their replication cohorts and is confirmed in ours. When testing additional effects of non-causative polymorphic (CAG)n-containing genes using the EUROSCA cohort, Tezenas du Montcel and colleagues [1] proposed associations between the size of the *ATXN2* (SCA2) and *ATN1* (DRPLA) alleles and the age at onset of SCA3. These effects were not replicated in any of the alternative cohorts, or in our Azorean SCA3 cohort. An association was also found between shorter *HTT* alleles and SCA3 onset; this finding was not investigated in the replication cohorts and our Azorean SCA3 cohort does not support this association. For *ATXN2*, whose intermediate alleles (frequency of 7%) had been associated with an earlier SCA3 onset in the EUROSCA series, the lack of replication was interpreted by Tezenas du Montcel and colleagues [1] as reflecting the rarity of *ATXN2* intermediate alleles in the replication populations, a justification especially applicable to the French and Japanese SCA3 patients, which lack this class of alleles. Notwithstanding in our cohort,

where intermediate alleles reach 5% (Figure 1), patients with an intermediate *ATXN2* allele also show a tendency for earlier estimated onset (34±3 years, mean ± standard error) when compared with patients without an intermediate *ATXN2* allele (38±1 years). This result shows how a population-specific genetic profile at a candidate locus impacts the ability to confirm/dismiss a modifier effect. In this sense, replication studies should be performed in populations where the allelic profile for the candidate modifier locus/loci is similar; otherwise the "lack" of replication becomes hard to interpret.

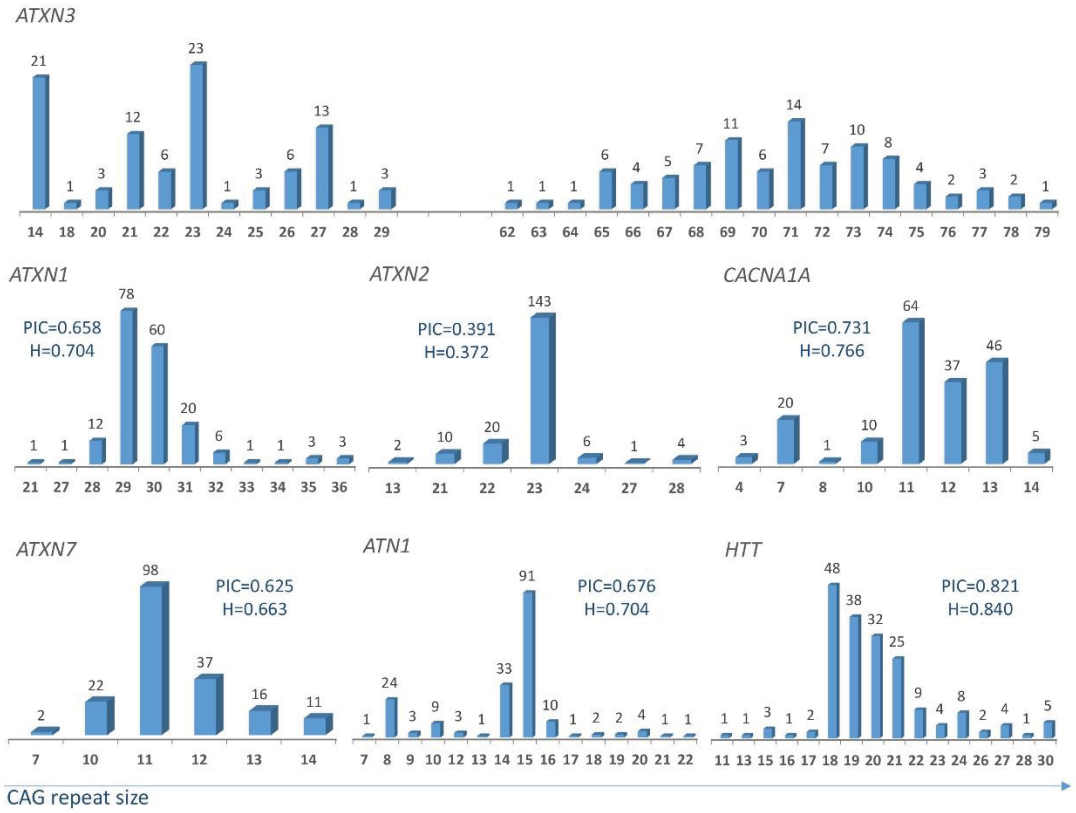


Figure 1. Allelic distribution of CAG repeats in *ATXN3, ATXN1, ATXN2, CACNA1A, ATXN7, ATN1* and *HTT* loci in SCA3 Azorean patients. Polymorphic information content (PIC) and expected heterozygosity (H) are shown for each candidate loci.

In the work of Tezenas du Montcel and colleagues [1], when the *ATXN3* locus was analyzed as a potential modifier of other SCAs, it was reported as displaying a low level of polymorphism. A study of nearly 2000 Portuguese chromosomes determined the *ATXN3* locus heterozygosity to be 85% [14]. The high level of polymorphism in the *ATXN3* locus was also described in the Japanese

population (polymorphic information content (PIC) = 0.78; [15]). Noteworthy, in our cohort, with the exception of *ATXN2*, all analyzed candidate loci are highly polymorphic (Figure 1), with values of PIC [16] ranging from 63% (*ATXN7*) to 82% (*HTT*). As a positive finding, not highlighted in the EUROSCA series, we can further report is that in our series, the longer allele at the *ATXN1* is negatively correlated with onset, significantly increasing by 1.5% the explanation of SCA3 age at onset variability (part correlation coefficient; r^2=0.701, p=0.038). Noteworthy, when compared to EUROSCA, we observe a higher frequency of longer normal *ATXN1* alleles in our cohort.

In summary, we believe that large multicenter collaborative studies, such as the EUROSCA initiative are crucial in the analysis of genetic modifiers for SCAs. The ability to replicate findings from such large cohorts is, notwithstanding, dependent on how comparable the replications cohorts are relatively to the discovery studies, both in the allelic profile of the causal mutation (CAG tracts size in polyQ SCAs) as well as in the allelic frequencies of the candidate modifier locus/loci.

REFERENCES

1. Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain. 2014;137:2444–55.

2. Génin E, Feingold J, Clerget-Darpoux F. Identifying modifier genes of monogenic disease: Strategies and difficulties. Hum. Genet. 2008. p. 357–68.

3. Pulst SM, Santos N, Wang D, Yang H, Huynh D, Velazquez L, et al. Spinocerebellar ataxia type 2: PolyQ repeat variation in the CACNAIA calcium channel modifies age of onset. Brain. 2005;128:2297–303.

4. Araújo M. Geographic profiling of Machado-Joseph disease: a health geographic pilot-study in the Azores. University of the Azores, Ponta Delgada.; 2012.

5. Pato CN, Azevedo MH, Pato MT, Kennedy JL, Coelho I, Dourado A, et al. Selection of homogeneous populations for genetic study: The Portugal genetics of psychosis project. Am. J. Med. Genet. - Neuropsychiatr. Genet. 1997;74:286–8.

6. Santos C, Abade A, Lima M. Testing hierarchical levels of population sub-structuring: the Azores islands (Portugal) as a case study. J. Biosoc. Sci. 2008;40:607–21.

7. Santos C, Lima M, Montiel R, Angles N, Pires L, Abade A, et al. Genetic structure and origin of peopling in the Azores Islands (Portugal): The view from mtDNA. Ann. Hum. Genet. 2003;67:433–56.

8. Montiel R, Bettencourt C, Silva C, Santos C, Prata MJ, Lima M. Analysis of Y-chromosome variability and its comparison with mtDNA variability reveals different demographic histories between islands in the Azores Archipelago (Portugal). Ann. Hum. Genet. 2005;69:135–44.

9. Kristiansson K, Naukkarinen J, Peltonen L. Isolated populations and complex disease gene identification. Genome Biol. 2008;9:109.

10. Dean AG, Sullivan KM SM. OpenEpi: Open Source Epidemiologic Statistics for Public Health, Version 2014/09/22. 2014;

11. IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. NY: IBM Corp.;

12. França MC, Emmel VE, D'Abreu A, Maurer-Morelli C V., Secolin R, Bonadia LC, et al. Normal ATXN3 allele but not CHIP polymorphisms modulates age at onset in Machado-Joseph disease. Front. Neurol. 2012;3.

13. Zhou Q, Ni W, Dong Y, Wang N, Gan S-R, Wu Z-Y. The Role of Apolipoprotein E as a Risk Factor for an Earlier Age at Onset for Machado-Joseph Disease Is Doubtful. PLoS One. 2014;9:e111356.

14. Lima M, Costa MC, Montiel R, Ferro A, Santos C, Silva C, et al. Population genetics of wild-type CAG repeats in the Machado-Joseph Disease gene in Portugal. Hum. Hered. 2005;60:156–63.

15. Sasaki H, Wakisaka A, Fukazawa T, Iwabuchi K, Hamada T, Takada A, et al. CAG repeat expansion of Machado-Joseph disease in the Japanese: Analysis of the repeat instability for parental transmission, and correlation with disease phenotype. J. Neurol. Sci. 1995;133:128–33.

16. Nagy S, Poczai P, Cernák I, Gorji AM, Hegedus G, Taller J. PICcalc: An online program to calculate polymorphic information content for molecular genetic studies. Biochem. Genet. 2012;50:670–2.

CHAPTER IV

Promoter variation and expression levels of inflammatory genes IL1A, IL1B, IL6 and TNF in blood of spinocerebellar ataxia type 3 (SCA3) patients.

Neuromolecular medicine

Journal subject categories: Neuroscience

Journals rank and quartile: 75/256 – Q2

Impact factor 2015: 3.692

5-year impact factor 2015: 4.120

Mafalda Raposo[1,2,3], Conceição Bettencourt[4], Amanda Ramos[1,2,3], Nadiya Kazachkova[1,2,3], João Vasconcelos[5], Teresa Kay[6], Jácome Bruges-Armas[2,3,7], Manuela Lima[1,2,3]

1. Department of Biology, University of the Azores, Ponta Delgada, Portugal
2. Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
3. Institute for Molecular and Cell Biology (IBMC), University of Porto, Porto, Portugal
4. Department of Clinical and Experimental Epilepsy and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
5. Department of Neurology, Hospital do Divino Espírito Santo, Ponta Delgada, Portugal
6. Department of Clinical Genetics, Hospital of D. Estefania, Lisbon, Portugal
7. SEEBMO, Hospital do Santo Espírito da Ilha Terceira, Angra do Heroísmo, Portugal

ABSTRACT

Age at onset in spinocerebellar ataxia type 3 (SCA3/MJD) is incompletely explained by the size of the CAG tract at the *ATXN3* gene, implying the existence of genetic modifiers. A role of inflammation in SCA3 has been postulated, involving altered cytokines levels; promoter variants leading to alterations in cytokines expression could influence onset. Using blood from 86 SCA3 patients and 106 controls this work aimed to analyse promoter variation of four cytokines (*IL1A, IL1B*, *IL6* and *TNF*) and to investigate the association between variants detected and their transcript levels, evaluated by quantitative PCR. Moreover, the effect of APOE isoforms, known to modulate cytokines, was investigated. Correlations between cytokine variants and onset were tested; the cumulative modifier effects of cytokines and APOE were analysed. Patients carrying the IL6*C allele had a significant earlier onset (4 years in average) than patients carrying the G allele, in agreement with lower mRNA levels produced by IL6*C carriers. The presence of APOE*ε2 allele seems to anticipate onset in average 10 years in patients carrying the IL6*C allele; a larger number of patients will be needed to confirm this result. These results highlight the pertinence of conducting further research on the role of cytokines as SCA3 modulators, pointing to the presence of shared mechanisms involving *IL6* and *APOE.*

INTRODUCTION

Spinocerebellar ataxia type 3 (SCA3/MJD; MIM#109150; ORPHA98757) is the most common spinocerebellar ataxia worldwide. The number of coding CAG repeats at the causative locus, *ATXN3*, explains from 50% to 75% of the age at onset variance (reviewed in [1]) therefore implying the existence of additional familial factors, namely genetic. Several genetic modifiers have been proposed: the number of CAG repeats at several expansion loci [2,4,5]; allelic variants at the apolipoprotein E (*APOE)* [6,7] and glucosidase, beta, acid (*GBA)* genes [8]; variation in the 3'UTR at the *ATXN3* gene [9] as well as the size of the normal SCA3 allele [10]. In the *ATXN3* gene a repeat expansion above 50 triplets encodes an abnormally long polyglutamine (polyQ) stretch in the ataxin-3 protein [11]; mutant ataxin-3 is prone to misfolding and aggregation, triggering a cascade of pathological events [12]. The putative role of inflammation, namely the behaviour of interleukine 1 alpha (IL1A), interleukine 1 beta (IL1B), interleukine 6 (IL6) and tumor necrosis factor (TNF), has been investigated in polyQ diseases [13]. In SCA3 brain tissue, IL1B and IL6 staining was found to be enhanced, as compared to controls; activated microglia and reactive astrocytes have also been observed [14–16]. Recently, eotaxin was found to be higher in serum of SCA3 asymptomatic carriers and in patients [17]. *IL1A* c.-889C>T, *IL1B* c.-511C>T, *IL6* c.-174G>C and *TNF* c.-308G>A localized at the promoter of respective cytokine genes have been related *in vitro*, *ex* and *in vivo* studies with differences in mRNA and/or protein levels of these cytokines [18–21]. Moreover, a link between APOE and cytokines has been investigated, since APOE modulates inflammatory and immune responses in an isoform-dependent manner [22]. We have previously shown that the APOE*ε2 allele was significantly associated with an earlier age at onset in a cohort of SCA3 Azorean patients [6].

Given a possible role of inflammation in SCA3, we hypothesised that promoter variants leading to alterations of expression levels of cytokines could influence disease manifestation, namely onset. Using peripheral blood from a homogenous Azorean cohort of SCA3 patients the present work

aimed to analyse variants in the promoter regions of four main cytokines: *IL1A* c.-889C>T (rs1800587), *IL1B* c.-511C>T (rs16944), *IL6* c.-174G>C (rs1800795) and *TNF* c.-308G>A (rs1800629), and to investigate the association between these variants and the respective transcript levels. Genotype-phenotype correlations were performed to test the loci previously reported as potential modifiers of SCA3 onset. Moreover, the cumulative modifier effects of cytokines loci and APOE were also tested.

SUBJECTS AND METHODS

Subjects

Eighty-six Azorean SCA3 patients, confirmed as carriers of the *ATXN3* mutation and 106 apparently healthy controls, were included in this study. Controls were selected taking into account the ancestry, age and gender distribution of cases. The size of the (CAG)n tract was determined as previously reported [23]. Age at onset, defined as the age of appearance of gait disturbance and/or diplopia reported by the patient and/or a close relative, was recorded during clinical assessments performed at the Department of Neurology (Hospital do Divino Espírito Santo - HDES, Ponta Delgada). *APOE* genotypes from SCA3 patients were considered for statistical analysis; genotyping was performed as in Bettencourt and colleagues [6]. All samples were collected after informed consent. This study is a part of a project approved by the Ethics Committee of HDES.

DNA isolation and multiplex PCR-RFLP

DNA was extracted from all samples using standard procedures. A multiplex PCR– restriction fragment length polymorphism (RFLP) was developed to analyse variants in the promoter of four cytokines: *IL1A* c.-889C>T (rs1800587), *IL1B* c.-511C>T (rs16944), *IL6* c.-174G>C (rs1800795) and *TNF* c.-308G>A (rs1800629). The set of primers (0.2µM of each one per reaction) for each cytokine

variant as well as multiplex PCR-RFLP reactions mixture and conditions are described in supplementary Table 1.

RNA isolation and qPCR

A subset of 54 SCA3 patients were selected to measure cytokines mRNA levels. Signs of inflammatory or infective conditions were annotated by accessing the clinical records of patients; patients presented any of the abovementioned conditions were not included. mRNA cytokine levels were also determined in 33 controls. Four pre-validated TaqMan Gene Expression Assays (Hs00174092_m1, Hs01555410_m1, Hs00985639_m1 and Hs99999043_m1 from Applied Biosystems) were used to measure cytokines mRNA levels. RNA isolation and quantification, cDNA synthesis, quantitative Real-Time PCR (qPCR) conditions, as well as calculation of relative expression values have been performed as described elsewhere [24].

Statistical analysis

Allele and genotype frequencies were estimated for all analysed loci and Hardy-Weinberg equilibrium (HWE) was tested. Allelic/genotypic frequencies for controls (N=106) were compared with available data for other European and non-European populations. An ANCOVA, using age at sampling as covariate, was run to compare transcript levels between cytokine genotypes. The effects of the CAG length in expanded allele on age at onset, as well as the presence/absence of each cytokine allelic variant were assessed using a linear fitting model. Equality of variances between groups was verified by the Levene's test. An ANCOVA, using the CAG length in expanded allele as covariate, was conducted to compare estimated age at onset between: (1) cytokine genotypes; or (2) cytokine alleles; or (3) interaction of APOE*ε2 allele and cytokine genotypes; or (4) interaction of APOE*ε2 allele and allelic variants. For two or more pairwise comparisons (comparisons between cytokine genotypes), p-values were adjusted using the Bonferroni procedure. Significant effects resulting from the ANCOVA comparisons, obtained only for IL6 allelic variants, were further tested. These correlations were confirmed: a) using a generalized

estimating equation test, where kinship was used as repetitive measure within-subjects; and b) using the subset of 38 unrelated SCA3 patients (patients who shared grand-parents were considered related). All statistical analyses were performed in IBM SPSS Statistics 22 (IBM Corp. Released 2013). A statistically significant result lower than 0.05 was considered for all tests performed.

RESULTS

Gender and age at collection for the studied subjects are shown in Figure 1. Genotypes for the *ATXN3* locus (N=86) in patients, as well as relevant clinical data are displayed in Figure 1.

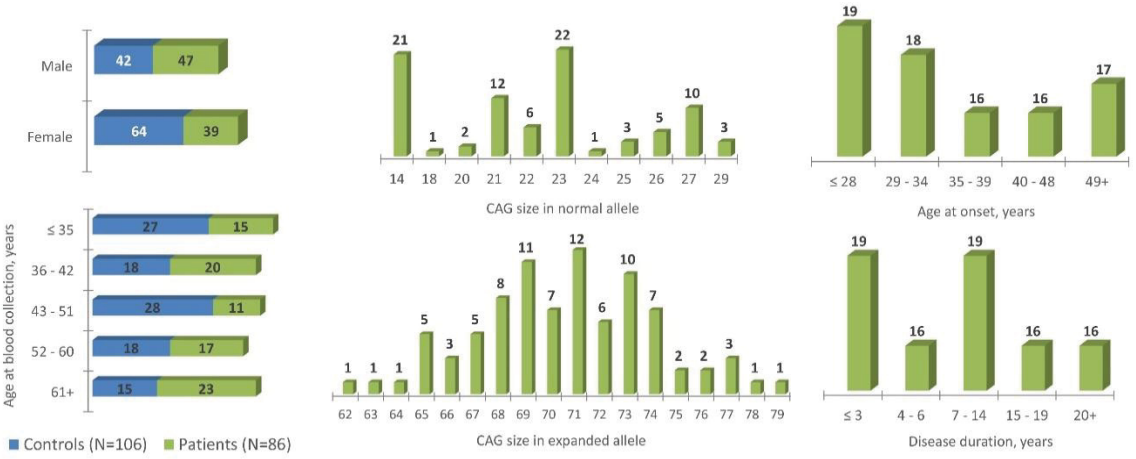


Figure 1. Demographic data and clinical features for studied individuals.

Loci were in conformity with Hardy-Weinberg equilibrium expectations, with the exception of IL1A and IL6 loci in SCA3 patients (supplementary Table 2). Pairwise differentiation exact test failed to detect significant differences in allelic or genotypic *IL1A, IL1B* and *TNF* frequencies between SCA3 patients and population-matched controls. At the IL6 locus, a statistically significant difference was obtained when comparing all patients with controls, which should reflect the excess of the C allele in the patients group; no differences, however, were detected when considering only unrelated patients (supplementary Table 2).

In SCA3 patients, no significant differences were obtained when comparing mRNA levels by genotypes (Figure 2). The effects of promoter allelic variants on mRNA levels were confirmed: the IL1A*T allele, the IL1B*T allele, the IL6*G allele and the TNF*A allele were associated with higher mRNA levels (Figure 2), in accordance with previous studies. The mRNA levels by cytokine genotype varied similarly in controls (data not shown).

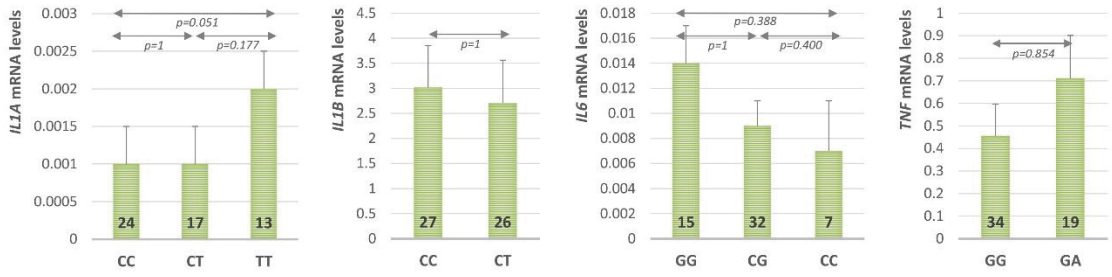


Figure 2. Cytokines mRNA levels (shown as $2^{-\Delta Ct}$) by genotypes in 54 SCA3 patients. Expression values were adjusted for age at blood collection (50 years). In the comparisons performed by *IL1A* or *IL6* genotype, Bonferroni adjusted p-values were obtained by an ANCOVA procedure. *IL1B* and *TNF* mRNA levels were not successfully quantified for one patient.

A negative correlation between the size of *ATXN3* expanded allele of SCA3 patients and age at onset was observed (N=86, r=−0.804, p<0.0005). The explanation of the age at onset variance provided by the CAG length in expanded allele was 65% (*F*=154.1, p<0.0005). An improvement of the previously model was observed only when the presence/absence of the IL6*C variant was added, which significantly contributed to the variance of the age at onset, by additionally explaining 1.9% (N=86, Part correlation coefficient=0.138, p<0.05).

Patients carrying the IL1A*T allele or the IL1B*T allele or the IL6*C allele all showed a tendency for an earlier age at onset (adjusted for mean CAG length) compared to patients homozygous for IL1A*C allele or the IL1B*C allele or the IL6*G allele (supplementary Table 3). Age at onset was anticipated by 4 years in average in patients carrying the IL6*C (N=66) (*F(1,83)*=4.7, p=0.03). The use of a generalized estimating equation test accounting for relatedness, and the earlier onset in patients carrying the IL6*C allele confirmed that this result was not due to patient's relatedness

(Wald X^2 = 3.8; p=0.05); the same tendency was observed when analysing only unrelated patients (N=38). In the present cohort (N=86), the presence of the APOE*ε2 allele explained 3.4% (p=0.003) of variance in age at onset. The presence of *APOE**ε2 allele significantly anticipated onset in average 10 years in patients carrying one or two copies of the IL6*C allele (p=0.005, Figure 3).

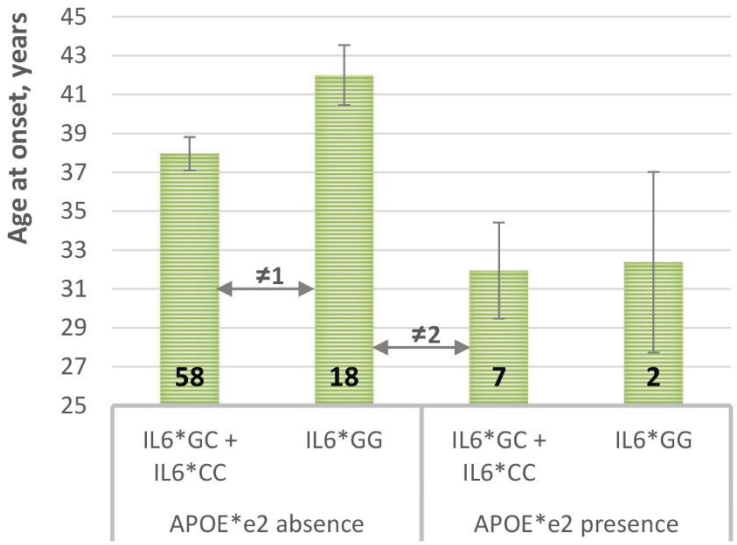


Figure 3. Estimated age at onset was calculated taking in consideration the cumulative effect of APOE*e2 allele and IL6 variation. ≠1The difference was in average of 4 years (ANCOVA, p=0.03). ≠2The difference was in average of 10 years (t-test, p=0.005). T-test was calculated using the OpenEpi, version 3.03 (Dean AG, Sullivan KM, Soe MM. OpenEpi: Open Source Epidemiologic Statistics for Public Health. www.OpenEpi.com, updated in 2014/09/22). APOE genotype was not successfully obtained in one patient.

Fitting a linear model, estimated age at onset (*F(3,81)*=63.001, p<0.0005) using the *APOE* and *IL6* loci alongside with the CAG size in expanded allele could be calculated applying the formula: age at onset = 230.117 – 2.686 x (CAG_n in expanded allele) – 3.272 x (presence/absence of IL6*C allele) – 6.911 x (presence/absence of APOE*ε2 allele).

DISCUSSION

A significant association between *IL6* c.-174G>C variation and age at onset was, nevertheless, identified; patients carrying the IL6*C allele presented, in average, an onset four years earlier than

the one displayed by patients homozygous for the G allele. We further observed a tendency for lower mRNA levels in patients carrying the IL6*C allele compared to GG homozygous, a result which is in agreement with previous findings [18]. Fishman and colleagues had suggested that *IL6* c.-174G>C variation is near to a glucocorticoid receptor (GR) binding site as well as G>C position could potentiate the creation for a binding site for the transcription factor nuclear factor 1 (NF1), implying, in both cases, the repression of transcription [18]. In our cohort, the low mRNA levels of IL6 produced by the IL6*C carriers were associated with a premature SCA3 onset, suggesting that in SCA3 patients' cells, such low levels could negatively contribute to cellular dysfunction, leading to the premature appearance of the first symptoms. Nishimura and colleagues [26] previously reported an association between IL1B*C allele and SCA6 onset; however, this association was not confirmed in our cohort of patients. In this study, an anticipation of onset (average of 10 years) was observed in patients carrying the APOE*ε2 allele and one or two copies of the IL6*C allele. Although there is no published data for IL6, APOE is known to suppress the secretion of TNF and IL1B, the APOE*ε2 isoform being associated with the lowest levels of secretion [22].

Even considering the homogeneity features of our patient's cohort, since sample size in this study is limited, the genotype-phenotype associations described should be replicated in a larger Azorean sample, when available, as well as in independent cohorts. Globally, results highlight the pertinence of further research on the role of cytokines as modulators of SCA3 onset, pointing to the presence of shared mechanisms involving *IL6* and *APOE*.

COMPETING INTERESTS

The authors declare no competing interests.

ACKNOWLEDGMENTS

This work was funded by FEDER funds through the Operational Competitiveness Programme – COMPETE and by National Funds through FCT – Fundação para a Ciência e a Tecnologia under the project FCOMP-01-0124-FEDER-028753 (PTDC/DTP/PIC/0370/2012). A PhD fellowship M3.1.2/F/006/2011 (MR) and postdoctoral fellowships M3.1.7/F/031/2011 (AR) and M3.1.7/F/002/2008 (NK) were supported by Fundo Regional para a Ciência (FRC), Governo dos Açores. CB is supported by the Wellcome Trust (UK).

REFERENCES

1. Bettencourt C, Lima M. Machado-Joseph Disease: from first descriptions to new perspectives. Orphanet J. Rare Dis. 2011;6:35.

2. Raposo M, Ramos A, Bettencourt C, Lima M. Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid? Brain. 2015;awv206.

3. Tezenas du Montcel S, Durr A, Bauer P, Figueroa K, Ichikawa Y, Brussino A, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain. 2014;137:2444–55.

4. Jardim L, Silveira I, Pereira ML, do Céu Moreira M, Mendonça P, Sequeiros J, et al. Searching for modulating effects of SCA2, SCA6 and DRPLA CAG tracts on the Machado-Joseph disease (SCA3) phenotype. Acta Neurol. Scand. 2003;107:211–4.

5. Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain. 2014;137:2444–55.

6. Bettencourt C, Raposo M, Kazachkova N, Cymbron T, Santos C, Kay T, et al. The APOE ε2 allele increases the risk of earlier age at onset in Machado-Joseph disease. Arch. Neurol. 2011;68:1580–3.

7. Peng H, Wang C, Chen Z, Sun Z, Jiao B, Li K, et al. APOE ε2 allele may decrease the age at onset in patients with spinocerebellar ataxia type 3 or Machado-Joseph disease from the Chinese Han population. Neurobiol. Aging. 2014;35:2179.e15-8.

8. Siebert M, Donis KC, Socal M, Rieder C, Emmel VE, Vairo F, et al. Glucocerebrosidase gene variants in parkinsonian patients with Machado Joseph/spinocerebellar ataxia 3. Park. Relat. Disord. 2012;18:185–90.

9. Long Z, Chen Z, Wang C, Huang F, Peng H, Hou X, et al. Two novel SNPs in ATXN3 3’ UTR may

decrease age at onset of SCA3/MJD in Chinese patients. PLoS One. 2015;10:e0117488.

10. França MC, Emmel VE, D'Abreu A, Maurer-Morelli C V, Secolin R, Bonadia LC, et al. Normal ATXN3 Allele but Not CHIP Polymorphisms Modulates Age at Onset in Machado-Joseph Disease. Front. Neurol. 2012;3:164.

11. Maciel P, Costa MC, Ferro A, Rousseau M, Santos CS, Gaspar C, et al. Improvement in the molecular diagnosis of Machado-Joseph disease. Arch. Neurol. 2001;58:1821–7.

12. Evers MM, Toonen LJA, van Roon-Mom WMC. Ataxin-3 protein and RNA toxicity in spinocerebellar ataxia type 3: current insights and emerging therapeutic strategies. Mol. Neurobiol. 2014;49:1513–31.

13. Olejniczak M, Urbanek M, Krzyzosiak W. The Role of the Immune System in Triplet Repeat Expansion Diseases. Mediators Inflamm. 2015;2015:1–11.

14. Evert B, Schelhaas J, Fleischer H, de Vos R, Brunt E, Stenzel W, et al. Neuronal intranuclear inclusions, dysregulation of cytokine expression and cell death in spinocerebellar ataxia type 3. Clin. Neuropathol. 2006;25:272–81.

15. Evert B, Vogt I, Kindermann C, Ozimek L, de Vos R, Brunt E, et al. Inflammatory genes are upregulated in expanded ataxin-3-expressing cell lines and spinocerebellar ataxia type 3 brains. J. Neurosci. 2001;21:5389–96.

16. Evert BO, Vogt IR, Vieira-Saecker AM, Ozimek L, de Vos RAI, Brunt ERP, et al. Gene expression profiling in ataxin-3 expressing cell lines reveals distinct effects of normal and mutant ataxin-3. J. Neuropathol. Exp. Neurol. 2003;62:1006–18.

17. da Silva Carvalho G, Saute J, Haas C, Torrez V, Brochier A, Souza G, et al. Cytokines in Machado Joseph Disease/Spinocerebellar Ataxia 3. Cerebellum. 2015;

18. Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin J, Humphries S, et al. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. J. Clin. Invest. 1998;102:1369–76.

19. Dominici R, Cattaneo M, Malferrari G, Archi D, Mariani C, Grimaldi L, et al. Cloning and functional analysis of the allelic polymorphism in the transcription regulatory region of interleukin-1 alpha. Immunogenetics. 2002;54:82–6.

20. Hall S, Perregaux D, Gabel C, Woodworth T, Durham L, Huizinga T, et al. Correlation of polymorphic variation in the promoter region of the interleukin-1 beta gene with secretion of interleukin-1 beta protein. Arthritis Rheum. 2004;50:1976–83.

21. Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW. Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. Proc. Natl. Acad. Sci. U. S. A. 1997;94:3195–9.

22. Zhang H, Wu L-M, Wu J. Cross-Talk between Apolipoprotein E and Cytokines. Mediators Inflamm. 2011;2011:1–10.

23. Bettencourt C, Fialho RN, Santos C, Montiel R, Bruges-Armas J, MacIel P, et al. Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: A study in normal families

from the Azores islands (Portugal). J. Hum. Genet. 2008;53:333–9.

24. Raposo M, Bettencourt C, Maciel P, Gao F, Ramos A, Kazachkova N, et al. Novel candidate blood-based transcriptional biomarkers of Machado-Joseph disease. Mov. Disord. 2015;30:968–75.

25. IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0. NY: IBM Corp.;

26. Nishimura M, Kawakami H, Maruyama H, Izumi Y, Kuno S, Kaji R, et al. Influence of interleukin-1beta gene polymorphism on age-at-onset of spinocerebellar ataxia 6 (SCA6) in Japanese patients. Neurosci. Lett. 2001;307:128–30.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Multiplex PCR primers, size of amplified fragments, restriction enzymes, size of restriction fragments as well as PCR-RFLP reactions mixture and conditions.

	Multiplex PCR		**RFLP**	
	Primers sequence 5' – 3'	**Product size (bp)**		
IL1A c.-889C>T				
IL1A-F[1]	TGTTCTACCACCTGAACTAGGC	99	*NcoI*	
IL1A-R[1]	TTACATATGAGCCTTCCATG		C → 79 + 20	T → 99
IL1B c.-511C>T				
IL1B-F[2]	TGGCATTGATCTGGTTCATCCA	244	*AvaI*	
IL1B-R	CCTGTCTGTATTGAGGGTG		C → 190 + 54	T → 244
IL6 c.-174G>C				
IL6-F	CAGAAGAACTCAGATGACTGGT	377	*NcoI*	
IL6-R	TGCAATGTGACGTCCTTTAC		G → 351 + 26	C → 377
TNF c.-308G>A				
TNF-F[3]	GAGGCAATAGGTTTTGAGGGCCAT	147	*NcoI*	
TNF-R[3]	GGGACACACAAGCATCAAG		G →126 + 21	A → 147
Reaction mixture				
	0.4mM each dNTP 1x NH4 Buffer 3mM MgCl2 solution 1x HiSpec solution 2U BIOTAQ DNA polymerase (Bioline) 200ng genomic DNA		1X reaction buffer Tango (10X) 5U NcoI 5U AvaI (Thermo Fisher Scientific) 3µl PCR product	
	Total volume	*25µl*		*10µl*
Thermocycler/Thermoblock conditions				
	initial denaturation: 95°C, 5min *34 cycles* *denaturation: 95°C, 30s* *annealing: 58°C, 90s* *extension: 72°C, 45s* final extension: 72°C , 10min		Incubation: overnight, 37°C (the digested product run on a 14% PAGE gel and was revealed using a silver nitrate standard protocol)	

_ mismatch primers (point mutation)

[1]Primers pair: De Freitas NM, Imbronito A V., Neves AC, Nunes FD, Pustiglioni FE, Lotufo RFM. Analysis of IL-1A(-889) and TNFA(-308) gene polymorphism in Brazilian patients with generalized aggressive periodontitis. Eur Cytokine Netw. 2007;18(3):142–7; [2]Primer forward: Brett PM, Zygogianni P, Griffiths GS, Tomaz M, Parkar M, D'Aiuto F, et al. Functional Gene Polymorphisms in Aggressive and Chronic Periodontitis. J Dent Res. 2005 1;84(12):1149–53; [3]Primers pair: Moorchung N, Srivastava AN, Gupta NK, Ghoshal UC, Achyut BR, Mittal B. Cytokine gene polymorphisms and the pathology of chronic gastritis. Singapore Med J. 2007;48(5):447–54.

Supplementary Table 2. Genotypic, allelic frequencies and p-values for the Hardy-Weinberg equilibrium test for each cytokine loci studied in all and unrelated SCA3 patients and population-matched controls. Differentiation exact test p-values are also shown.

			SCA3 patients		Controls (C), N=106	*Differentiation test p-value*[#]	
			All, N=86	Unrelated (U), N=38		C. *versus* all	C. *versus* U.
IL1A c.-889C>T (rs1800587)							
Frequencies	genotype	CC	0.424	0.474	0.491	0.139	0.200
		CT	0.365	0.263	0.396		
		TT	0.212	0.263	0.113		
	allele	C	0.606	0.605	0.689	0.099	0.153
		T	0.394	0.395	0.311		
			Hardy-Weinberg equilibrium				
			0.038	0.006	0.495		
IL1B c.-511C>T (rs16944)							
Frequencies	genotype	CC	0.558	0.500	0.462	0.136	0.523
		CT	0.419	0.474	0.481		
		TT	0.023	0.026	0.061		
	allele	C	0.767	0.737	0.703	0.181	0.555
		T	0.233	0.263	0.297		
			Hardy-Weinberg equilibrium				
			0.141	0.403	0.163		
IL6 c.-174G>C (rs1800795)							
Frequencies	genotype	GG	0.232	0.395	0.452	0.002	0.253
		GC	0.628	0.447	0.462		
		CC	0.140	0.158	0.086		
	allele	G	0.546	0.618	0.684	0.006	0.261
		C	0.454	0.382	0.316		
			Hardy-Weinberg equilibrium				
			0.019	0.739	0.316		
TNF c.-308G>A (rs1800629)							
Frequencies	genotype	GG	0.640	0.579	0.755	0.121	0.122
		GA	0.337	0.421	0.226		
		AA	0.023	0	0.019		
	allele	G	0.808	0.790	0.868	0.126	0.141
		A	0.192	0.210	0.132		
			Hardy-Weinberg equilibrium				
			0.726	0.171	1.000		

[#]p-value was calculated by exact G test in Genepop software (Raymond M. & Rousset F, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J. Heredity, 86:248-249
Rousset, F., 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol. Ecol. Resources 8: 103-106).
Pairwise differentiation exact test (Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online. 2005;1:47–50) for genotypic frequencies between apparently healthy Azorean individuals versus individuals from:
(1) mainland Portugal – no significant differences for *IL1B, IL6* and *TNF* loci, no data available for *IL1A* locus;
(2) Europe – differences significant for *IL6* and *TNF* loci, not significant for *IL1A* and *IL1B* loci;
(3) Asia – differences significant for *IL1A, IL1B* and *TNF* loci, no data available for *IL6* locus;
(4) Africa – differences significant for *IL1A* and *IL1B* loci, not significant for *TNF* locus, no data available for *IL6* locus.
Genotypes from European, Asiatic and African samples were obtained in dbSNP, NCBI (http://www.ncbi.nlm.nih.gov/snp/).

Supplementary Table 3. Genetic and clinical features of the 86 SCA3 patients divided by cytokines alleles, as well as presence/absence of APOE*ε2 allele.

	CAG length					
	Alleles	N	Normal allele	Expanded allele	Age at onset*	Disease duration
IL1A c.-889C>T	C	37	22±4	71±3	39±1	12±8
	T	48	21±5	70±4	37±1	11±9
IL1B c.-511C>T	C	49	22±5	71±4	38±1	11±8
	T	37	22±5	70±4	38±1	11±9
IL6 c.-174G>C	G	20	20±5	70±4	41±2≠	11±7
	C	66	22±5	71±3	37±1≠	12±9
TNF c.-308G>A	G	55	22±4	71±3	37±1	12±9
	A	31	20±5	70±4	38±1	10±7
APOE ε2 allele	Absent	76	21±5	70±4	39±1≠	11±8
	Present	9	22±6	70±3	32±2≠	13±12

CAG length in normal and expanded allele, as well as disease duration is represented as mean ± standard deviation; *Age at onset was adjusted for mean CAG length and are represented as mean ± standard error; ≠ $p<0.05$ was considered statistically significant.

MOLECULAR STATE BIOMARKERS

CHAPTER V

Novel candidate blood-based transcriptional biomarkers of Machado-Joseph Disease.


RESEARCH ARTICLE

Novel Candidate Blood-Based Transcriptional Biomarkers of Machado-Joseph Disease

Movement disorders

Journal subject categories: Clinical neurology

Journals rank and quartile: 14/192 – Q1

Impact factor 2015: 6.010

5-year impact factor 2015: 5.366

Mafalda Raposo, B.Sc.[1,2], Conceição Bettencourt, Ph.D.[3], Patrícia Maciel, Ph.D.[4,5], Fuying Gao, Ph.D.[6], Amanda Ramos, Ph.D.[1,2], Nadiya Kazachkova, Ph.D.[1,2], João Vasconcelos, M.D.[7], Teresa Kay, M.D.[8], Ana João Rodrigues, Ph.D.[4,5], Bruno Bettencourt, Ph.D.[2,9], Jácome Bruges-Armas, M.D., Ph.D.[2,9], Daniel Geschwind, M.D., Ph.D.[6], Giovanni Coppola, M.D., Ph.D.[6], Manuela Lima, Ph.D.[1,2]

1. Centre of Research in Natural Resources (CIRN), University of the Azores, Ponta Delgada, Portugal
2. Institute for Molecular and Cell Biology (IBMC), University of Porto, Porto, Portugal
3. Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
4. Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal
5. ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal
6. Departments of Psychiatry and Neurology, Semel Institute for Neuroscience and Human Behavior, David Geffen School of Medicine, University of California Los Angeles, Los Angeles CA, USA.
7. Department of Neurology, Hospital do Divino Espírito Santo, Ponta Delgada, Portugal
8. Department of Clinical Genetics, Hospital of D. Estefania, Lisbon, Portugal
9. Hospital de Santo Espírito da Ilha Terceira, SEEBMO, Angra do Heroísmo, Portugal

ABSTRACT

Background: Machado-Joseph disease (or spinocerebellar ataxia type 3) is a late-onset polyglutamine neurodegenerative disorder, caused by a mutation in the *ATXN3* gene, which encodes for the ubiquitously expressed protein ataxin-3. Previous studies on cell and animal models have suggested that mutated ataxin-3 is involved in transcriptional dysregulation. Starting with a whole-transcriptome profiling of peripheral blood samples from patients and controls, we aimed to confirm abnormal expression profiles in Machado-Joseph disease and to identify promising up-regulated genes as potential candidate biomarkers of disease status.

Methods: The Illumina Human V4-HT12 array was used to measure transcriptome-wide gene expression in peripheral blood samples from 12 patients and 12 controls. Technical validation and validation in an independent set of samples were performed by quantitative real-time PCR.

Results: Based on the results from the microarray, twenty-six genes, found to be up-regulated in patients, were selected for technical validation by quantitative real-time PCR (validation rate of 81% for the up-regulation trend). Fourteen of these were further tested in an independent set of 42 patients and 35 controls; ten genes maintained the up-regulation trend (*FCGR3B*, *CSR2RA*, *CLC*, *TNFSF14, SLA, P2RY13, FPR2, SELPLG, YIPF6* and *GPR96*); *FCGR3B, P2RY13 and SELPLG* were significantly up-regulated in patients when compared to controls.

Conclusions: Our findings support the hypothesis that mutated ataxin-3 is associated with transcription dysregulation, detectable in peripheral blood cells. Furthermore, this is the first report suggesting a pool of up-regulated genes in Machado-Joseph disease, which may have the potential to be used for fine phenotyping of this disease.

INTRODUCTION

Machado-Joseph disease (MJD; MIM #109150; ORPHA98757), also named spinocerebellar ataxia type 3 (SCA3), is the most common autosomal dominant ataxia worldwide [1], and the second most common polyglutamine (polyQ) disorder, following Huntington disease (HD) [2]. MJD is a clinically heterogeneous neurodegenerative disorder with an average onset around the fourth decade of life [3,4]. It is characterized by a wide range of manifestations, including ataxia, progressive external ophthalmoplegia, pyramidal and extrapyramidal signs, dystonia with rigidity and distal muscular atrophy [3,4]. A widespread neuronal loss in the cerebellum, pons, medulla oblongata, basal ganglia, thalamus, midbrain and cerebral cortex is also common in this disease [5]. Effective treatment for MJD is still lacking, and only symptomatic therapeutics is available [6]. The MJD's gene - *ATXN3* - mapped to 14q32.1 contains a (CAG)n tract at exon 10, which in mutant allele is expanded, typically beyond 51 repeat units [7,8]. A negative correlation between the number of CAG repeats in the expanded allele and the age at onset has been widely described, with the size of the CAG tract accounting for nearly 50% to 75% of the variation in the age at onset [6]. The *ATXN3* gene encodes for ataxin-3, a polyQ protein ubiquitously expressed in neuronal and non-neuronal tissues. Expansions of the polyQ tract above the pathological threshold initiates a cascade of pathogenic events that are being extensively studied (e.g. [9,10]). Similarly to MJD, it is known that in other polyQ diseases the expanded repeat triggers conformational changes in the corresponding proteins, leading to the formation of intracellular inclusions, considered the hallmark of this group of disorders [11]. In MJD, however, the characteristic neuronal intranuclear inclusions do not seem to be the main toxic entity (reviewed in [9]). Mutated ataxin-3 seems to be involved in transcriptional regulation via two processes: 1) recruitment of transcription factors to polyQ-rich inclusions [12–14]; and 2) altered interactions with transcription factors and co-activators [15–19]. So far, patterns of transcriptional dysregulation due to mutated ataxin-3 have been studied only in cellular and animal models. In such models, transcriptional alterations of genes involved in inflammatory processes, cell signaling and cell-surface associated proteins has

been described [17,20,21]. Studies of disease-modifying compounds, in the context of clinical trials, have been recently initiated for MJD [22,23], but gross clinical evaluation alone might not be sensitive enough to monitor disease progression and detect subtle therapeutic benefits; this is particularly true in diseases of slow progression, such as MJD. The development of disease-related biomarkers is therefore urgently needed. Gene expression profiling arrays, by identifying disease-specific transcriptional changes in blood, an easily accessible tissue, can boost the identification of potential biomarkers (reviewed in [24]). Thus far, these transcriptional changes have not been investigated in MJD patients, and their potential as biomarkers of disease remains to be evaluated. In this study we used microarrays to generate gene expression profiles in peripheral blood samples of MJD patients, aiming to confirm the presence of altered gene expression patterns in MJD and to identify up-regulated genes as potential candidate biomarkers of disease status.

METHODS

Subjects and sampling

Peripheral blood was collected from MJD Azorean patients and control individuals. A set of 12 patients and 12 controls was used for the gene expression microarray analysis and technical validation (Table 1). To ensure a homogeneous group of subjects, patients with an age at onset around the average (~40 years) in our series of cases, and presenting mainly cerebellar alterations, were selected for this phase of the study. An independent set of 42 patients and 35 controls was used for further validation (Table 1). All patients had a clinical and a molecular diagnosis of MJD (confirmed carriers of the *ATXN3* mutation). Patients were recruited through the Department of Neurology - Hospital Divino Espírito Santo (HDES, Ponta Delgada, Azores, Portugal). This study was approved by the Ethics Committee of the HDES, and all participants provided written informed consent.

Table 1. Characteristics of the MJD patients and controls according to the phases of the study.

	STUDY DESIGN	
MJD PATIENTS		
	Microarray & Technical validation	**Independent set validation**
N	12	42
Female|Male	10|2	26|16
Age (years)	42±6 [34-52]	47±15 [22-82]
Age at onset (years)	34±6 [28-46]	37±13 [12-70]
CAG repeats		
Normal allele	20±6 [14-29]	21±5 [14-29]
Expanded allele	72±2 [69-75]	71±4 [62-79]
Disease duration (years)	8±3 [6-17]	10±8 [0-30]
CONTROLS		
N	12	35
Female|Male	6|6	23|12
Age (years)	37±13 [19-61]	46±12 [24-77]

All quantitative variables were presented as mean ± standard deviation [range].

RNA isolation and gene expression microarrays

Whole-blood samples were collected in Tempus™ Blood RNA tubes (Applied Biosystems) and total RNA was isolated using the Tempus™ Spin RNA Isolation Kit (Applied Biosystems), following the manufacturer's protocol. The integrity of total RNA samples was assessed using the RNA 6000 Nano Kit on a 2100 Bioanalyzer (Agilent Technologies).

RNA was amplified, biotin-labeled and hybridized on a transcriptome-wide expression Illumina Human V4-HT12 array. Slides were scanned using Illumina BeadStation, and the signal was extracted using the Illumina BeadStudio software. All arrays were performed in the same core facility.

Technical validation of candidate blood-based biomarkers

As up-regulation is more reliably detected than down-regulation, since it is not dependent on the lower detection limit of the quantification methods, for the next phases of the study, we ranked and selected only those genes that in the gene expression microarray were up-regulated in patients. Therefore, up-regulated genes with a Benjamini-Hochberg false discovery rate (FDR)

adjusted p-value <0.05 and a log ratio > 1.2 were ranked. Consistency of the log ratio between pairs and maximal values of expression levels were further evaluated as criteria for the ranking procedure. Twenty-six genes fulfilled the previously described criteria were selected for quantitative real-time PCR (qPCR) technical validation, using the same set of samples previously used in the microarrays.

One microgram of total RNA was used to synthesize cDNA using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), according to the manufacturer's protocol. The TaqMan Gene Expression Master Mix and pre-validated TaqMan Gene Expression Assays (TaqMan IDs are described in Table 2) were used for qPCR (Applied Biosystems). TaqMan gene expression assays were designed, tested and optimized to address all MIQE guidelines [25]. More information on these assays is given in the white paper by Life Technologies (available at: https://tools.lifetechnologies.com/content/sfs/brochures/cms_088754.pdf. The *PPIB* (peptidylpropyl isomerase B; Hs00168719_m1) gene was used as the reference gene for qPCR validation. This reference gene is one of the most stably expressed genes in human blood, suitable for normalization in qPCR studies [26]. Moreover, *PPIB* gene was identified as an appropriated gene for expression normalization in blood of Huntington disease patients and R6/2 mice [27]. qPCR was performed in a 7900HT Fast Real-Time PCR System (Applied Biosystems). Each sample was run in triplicate alongside with the reference gene.

Validation in an independent set of samples

From the list of 26 genes subject to the technical validation, the 14 genes showing a consistent up-regulation trend in the technical validation phase, with the highest fold change and/or the lowest p-values, were further investigated in an independent set of samples (42 patients and 35 controls). The qPCR methodology was as described above (*technical validation of candidate blood-based biomarkers* subsection). Gene ontology (GO) annotations were performed in Qiagen's

interactive pathways analysis (IPA) software (QIAGEN Redwood City, www.qiagen.com/ingenuity).

Statistical analysis

Raw data obtained from the Illumina Human V4-HT12 arrays were analyzed with Bioconductor packages, as previously described [28]. Briefly, quality assessment included inter-array Pearson correlation and clustering based on top variant genes. Raw expression values were log2 transformed and normalized using quantiles. Batch effects were corrected using the ComBat algorithm [29]. Differential expression analysis (patients vs. controls) was performed using the limma package [28]. Up-regulated genes were considered for further analysis when FDR adjusted p-value was lower than 0.05, and log ratio was higher than 1.2-fold. Since controls were hybridized in separate batches an increased rate of false positives was expected. Therefore, to minimize this issue we applied a batch-correction algorithm (see above), adopted a very conservative statistical threshold (FDR <0.05), and verified all selected candidates by qPCR.

In the qPCR validation phases of this study (technical validation and validation in an independent set of samples), relative expression values were normalized to the reference gene (*PPIB*), and fold change values were calculated using the $2^{-\Delta\Delta Ct}$ method [30]. A two-sample, two-tailed Student's *t-test* was used to compare the ΔC_q (C_q, also named cycle threshold - Ct, is quantification cycle – Cq, as suggested by MIQE guidelines [25]) values between the two biological groups (patients vs. controls), and a p-value was calculated. All statistical procedures were performed using the DataAssist v3.0 software (Applied Biosystems).

RESULTS

Microarray-based gene expression profiling was used to analyze global gene expression in peripheral blood samples of MJD patients and controls. We identified a total of 5,523 up-regulated

probes, whereas 6,232 were down-regulated (FDR adjusted p-value < 0.05) when comparing patients to controls.

From the 26 top candidate biomarkers (Table 2) selected for technical validation by qPCR, we confirmed an up-regulation trend in 21/26 genes (validation rate of 81%), although only ten were statistically significant. From these 21 genes, fourteen (Table 2) were selected for validation by qPCR in an independent set of MJD patients and controls. Given the fact that the gender ratio in our first set of MJD samples was skewed towards females, we choose a 1:1 male to female ratio in our independent set of samples (both cases and controls) to ensure that the observed up-regulation was not due to a gender effect. In this phase, a validation rate of 71% was obtained, with 10/14 genes showing an up-regulation trend. In the new set of samples, the expression levels of *FCGR3B, CSF2RA, CLC, FPR2, SLA, GPR97, P2RY13, TNFSF14, SELPLG* and *YIPF6* were 1.11 to 2.60-fold higher in patients when compared to controls (Figure 1; Table 2). Noteworthy, *FCGR3B, P2RY13* and *SELPLG* genes were significantly up-regulated (p-value<0.05). According to gene ontology functional annotations, these genes are mostly related to the immune system response (*FCGR3B, CSF2RA, FPR2, TNFSF14,* and *SELPLG*) and G-protein coupled receptor signaling (*FPR2, GPR97* and *P2RY13*).

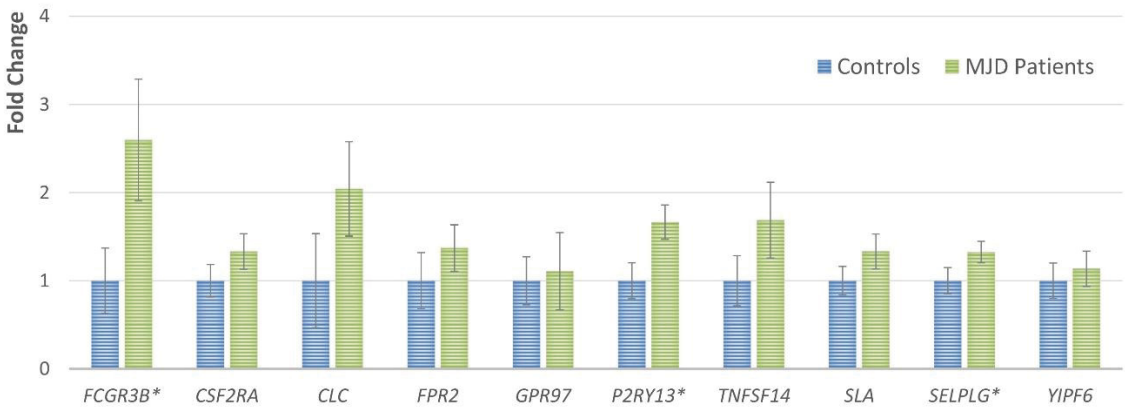


Figure 1. Fold change (FC) values for the 10 genes maintaining an up-regulation trend in an independent set of MJD patients versus controls. Error bars were obtained by the standard error of mean (SEM) difference of the ΔCq values, and presented as FC x (2^{SEM}-1) as described elsewhere [37]. FC=1 implies the absence of expression change. P-value was considered significant when lower than 0.05.

Within the patients group (patients from technical and independent set validation), our data shows a trend for higher expression levels in patients with shorter disease duration compared with patients with longer disease duration, especially for *FCGR3B* and *CLC*, which were statistically significant (Figure 2). The comparison between fold change values in patients grouped by CAG size in expanded allele failed to produce significant results (data not shown).

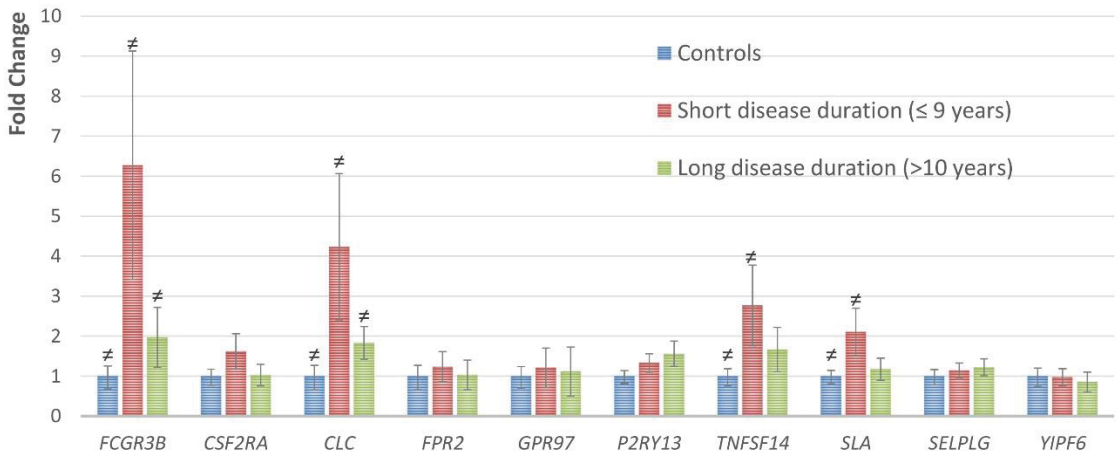


Figure 2. Fold change (FC) values of the 10 genes, previously validated as biomarkers of MJD status, in controls, patients with 9 or less years elapsed since onset (short disease duration) and with more than 10 years elapsed since onset (long disease duration). Disease duration was defined as the number of years elapsed from reported age at onset to the collection of blood samples for each patient. Error bars were obtained by the standard error of mean (SEM) difference of the ΔCq values, and presented as (FC) x (2^{SEM-1}) as described elsewhere [30]. FC=1 implies the absence of expression change. ≠ the statistical difference is significant (p-value used was <0.05).

Table 2. Candidate blood-based transcriptional biomarkers selected for the qPCR technical validation and validation in an independent set of samples.

				Microarray data		Technical Validation		Independent Validation	
TaqMan ID	**Gene ID**	**Symbol[1]**	**Full name[1]**	*Log Ratio[2]*	*p-value[3]*	*FC[4]*	*p-value*	*FC[4]*	*p-value*
Transmembrane receptors[5]									
Hs00275547_m1	2215	***FCGR3B***	Fc fragment of IgG, low affinity IIIb, receptor (CD16b)	1.943	1.18E-02	4.0582	6.80E-03	2.597	9,20E-03
Hs04191069_gH	3802	***KIR2DL1***	Killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 1	1.767	1.30E-03	1.8267	1.94E-01	Not selected	
Hs00427106_m1	3805	***KIR2DL4***	Killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 4	1.396	1.19E-03	1.2968	4.38E-01	Not selected	
Hs00538900_m1	1438	***CSF2RA***	Colony stimulating factor 2 receptor, alpha, low-affinity (granulocyte-macrophage)	1.294	5.80E-05	1.5071	1.44E-01	1.331	1.37E-01
Enzymes									
Hs01055743_m1	1178	***CLC***	Charcot-Leyden crystal protein	1.763	4.91E-03	3.4077	1.70E-02	2.041	1.13E-01
Hs01587865_g1	4051	***CYP4F3***	Cytochrome P450, family 4, subfamily F, polypeptide 3	1.512	2.14E-03	4.0785	4.77E-02	0.911	8.01E-01
Hs00193422_m1	3029	***HAGH***	Hydroxyacylglutathione hydrolase	1.368	4.84E-03	0.7296	3.10E-01	Not selected	
G-protein coupled receptors									
Hs01891184_s1	3579	***CXCR2***	Chemokine (C-X-C motif) receptor 2	2.040	4.99E-03	1.9372	2.13E-01	Not selected	
Hs02759175_s1	2358	***FPR2***	Formyl peptide receptor 2	1.541	4.56E-03	1.7242	9.79E-02	1.370	2.77E-01
Hs00416889_m1	222487	***GPR97***	G protein-coupled receptor 97	1.315	8.40E-03	3.3142	7.61E-02	1.107	7.77E-01
Hs03043902_s1	53829	***P2RY13***	Purinergic receptor P2Y, G-protein coupled, 13	1.313	1.37E-03	2.1724	1.07E-02	1.665	6.20E-03
Peptidases									
Hs00153519_m1	4311	***MME***	Membrane metallo-endopeptidase	1.476	4.07E-02	2.1544	1.84E-01	0.994	9.90E-01
Hs01073631_m1	64167	***ERAP2***	Endoplasmic reticulum aminopeptidase 2	1.410	2.42E-03	0.5328	5.63E-02	Not selected	

Table 2. (cont.)

TaqMan ID	Gene ID	Symbol[1]	Full name[1]	Microarray data		Technical Validation		Independent Validation	
				Log Ratio[2]	*p-value*[3]	*FC*[4]	*p-value*	*FC*[4]	*p-value*
Cytokines									
Hs00542477_m1	8740	***TNFSF14***	Tumor necrosis factor (ligand) superfamily, member 14	1.562	4.33E-04	5.6394	3.00E-04	1.687	9.02E-02
Others									
Hs00331399_m1	57535	***KIAA1324***	KIAA1324	2.101	1.77E-04	1.3916	2.65E-01	Not selected	
Hs00846590_s1	11026	***LILRA3***	Leukocyte immunoglobulin-like receptor, subfamily A (without TM domain), member 3	2.037	2.98E-05	0.5844	5.40E-02	Not selected	
Hs00747812_m1	3813	***KIR3DS1***	Killer cell immunoglobulin-like receptor, three domains, short cytoplasmic tail, 1	1.993	1.49E-03	1.1759	6.18E-01	Not selected	
Hs00962914_m1	7057	***THBS1***	Thrombospondin 1	1.684	4.68E-05	0.7767	5.77E-01	Not selected	
Hs00160066_m1	5266	***PI3***	Peptidase inhibitor 3, skin-derived	1.628	3.79E-02	3.1875	4.40E-02	0.525	1.74E-01
Hs01066294_m1	23762	***OSBP2***	Oxysterol binding protein 2	1.471	2.66E-03	0.9111	8.40E-01	Not selected	
Hs00190581_m1	8875	***VNN2***	Vanin 2	1.465	4.46E-03	1.4282	2.87E-01	Not selected	
Hs00275682_s1	3310	***HSPA6***	Heat shock 70kDa protein 6 (HSP70B')	1.344	1.22E-04	1.2347	3.82E-01	Not selected	
Hs00218346_m1	55363	***HEMGN***	Hemogen	1.328	8.09E-03	3.7584	4.84E-02	0.534	8.01E-02
Hs00277129_m1	6503	***SLA***	Src-like-adaptor	1.308	6.37E-05	2.5137	1.60E-03	1.333	1.15E-01
Hs00356602_m1	6404	***SELPLG***	Selectin P ligand	1.251	4.06E-06	1.9473	1.60E-03	1.324	4.04E-02
Hs00397742_m1	286451	***YIPF6***	Yip1 domain family, member 6	1.248	1.98E-04	2.5564	1.06E-02	1.136	5.17E-01

[1] Official gene symbol and full name provided by HUGO Gene Nomenclature Committee; [2] Log ratio = $\log_2$ (Cy3'/Cy5'); [3] p-value calculated by Student's t-test and adjusted for multiple comparisons by FDR (false discovery rate) correction; [4] Fold Change (FC) = geometric mean $2^{(-\Delta Cqpatients)}$ / geometric mean $2^{(\Delta Cqcontrols)}$; [5]Gene ontology annotations for technical validated genes were performed in Qiagen's interactive pathways analysis (IPA) software (QIAGEN, Redwood City, www.qiagen.com/ingenuity).

DISCUSSION

In this study we confirmed that transcriptional dysregulation can be detected in peripheral blood samples of MJD patients. We report a set of 10 genes - *FCGR3B, CSF2RA, CLC, FPR2, SLA, GPR97, P2RY13, TNFSF14, SELPLG* and *YIPF6* - that consistently show an up-regulation trend in MJD patients when compared to controls. Three of these genes - *FCGR3B*, *P2RY13* and *SELPLG* - were significantly up-regulated, and therefore should be considered in future studies as biomarkers of disease status. Moreover, the significantly increased levels of *FCGR3B and CLC* mRNAs observed in early MJD stages suggests that, even in a non-affected tissue (such as blood), a cellular response closer to the disease onset is activated in the presence of mutated ataxin-3; with disease progression (specifically more than 10 years after onset) expression levels of these genes fall to levels closer to those found in healthy individuals. These findings suggest a potential of the reported genes to monitor disease stages; further investigation is nevertheless, required. The potential of molecular alterations in the fine phenotyping of a polyQ disease was shown by Bjorkqvist and collaborators, which observed higher levels of interleukin (IL) 6 in plasma of Huntington disease (HD) gene carriers, in average 16 years before the predicted onset of clinical symptoms [31]. This study, also reported, that IL12 and granulocyte-macrophage colony-stimulating factor (GM-CSF) levels significantly decreased from early to moderate stages of the disease [31].

To our knowledge, from the abovementioned genes, only *TNFSF14, CSF2RA* and *FPR2* have been previously associated with neurodegenerative disorders. *TNFSF14* has been associated with amyotrophic lateral sclerosis (ALS) [32], and the latter two genes with Alzheimer's disease (AD) [33,34]. *TNFS14*, which is known to be expressed in immature dendrocytes, activated lymphocytes, monocytes and natural killer cells, motor neurons and astrocytes, has been shown to be up-regulated in the spinal cord of ALS patients [35]. This cytokine is known to be important for both innate and adaptive immune processes, and also functions with interferon-γ to induce a

singular slow apoptotic death in tumour cells. Soluble TNFSF14 produced by astrocytes in ALS acts as a death-inducing ligand in motor neurons [35]. The genetic ablation of the *tnfsf14* gene (orthologue to human *TNFSF14*) in an ALS mouse model has been related to a slow disease progression as well as an extended life expectancy, suggesting a possible modifier role for this gene [32].

CSF2RA encodes the alpha subunit of the heterodimeric receptor for colony stimulating factor 2 (CSF2), a hematopoietic factor. In recent years, *CSF2* has been shown to be an important neurotrophic factor in the central nervous system via binding to its receptor. Both *CSF2* and *CSF2RA* are expressed in neurons throughout the CNS, astrocytes, ependymal cells, and choroid plexus cells. In AD, expression of the protein encoded by *CSF2RA* has been found to be dysregulated in the hippocampus of patients [34]. No further links to neurodegenerative diseases have been established for this gene.

The formyl peptide receptor 2 (encoded by *FPR2*) is a G-protein-coupled receptor (GPCR) of bacterial chemotactic peptides expressed in B cells, mononuclear phagocytes and microglia. In AD models, the interaction between Aβ42 and *FPR2* is clearly associated with microglial cell activation. Moreover, a persistent internalization of Aβ42/*FPR2* complexes, which culminates in intracellular fibrillar formation and apoptotic death of the cells, suggests that *FPR2* can be a possible therapeutic target for AD [33]. Other members of the GPCR superfamily, which in our study, includes two dysregulated genes (*P2RY13* and *GPR97*), have been previously reported in polyQ-associated neurodegeneration. In MJD models, the involvement of the adenosine A2a receptor (A2aR), a GPCR subtype, has been reported [36]. It has been suggested that striatal pathology associated with mutant *ATXN3* overexpression in a lentiviral-based model can be abolished by inactivation of A2aR, providing the first evidence that manipulation of a neuromodulation system operated by A2aRs is effective in controlling the initial cascade of events triggered by the pathogenic ataxin-3 protein (synaptotoxicity and gliosis) [36]. The *P2Y5* gene, a

member of the same purinergic receptor family as *P2RY13*, has been validated as a candidate biomarker in HD blood cells [37].

FCGR3B and *SELPLG* genes were significantly up-regulated in MJD patients. Fc fragment of IgG, low affinity IIIb, receptor (FCGR3B) is an Fc gamma receptor (FcγR), which belong to a family of immunoglobulin-like receptors that bind to the Fc portion of IgG, and mediate the response of effector cells to immune complexes [38]. Activation of these receptors can result in a pro-inflammatory response including the release of cytokines and other mediators. Evidences concerning the ligation of specific FcγRs in the CNS by IgG and alternate ligands, which promote neuroinflammation and/or enhance neurodegeneration, were previously reported [38]. Selectin P ligand (*SELPLG*) gene encodes a glycoprotein that functions as a high affinity counter-receptor for the cell adhesion molecules P-, E- and L- selectin expressed on myeloid cells and stimulated T lymphocytes [39]. The selectin family mediates the tethering and rolling of leukocytes on the vascular wall during the process of leukocyte migration into the tissues under physiological and pathological conditions [40]. The role of selectins in leukocyte-endothelial interactions in the pathogenesis of neurological diseases had been debated [40]. Rodrigues and collaborators reported that depletion of *ATXN3,* using small-interference RNA in human and mouse cells, causes a decrease in expression of important cell adhesion molecules, indicating that the extracellular matrix–cell or cell–cell interconnection was compromised [41].

Five of the ten candidate biomarkers identified in the present study play a role in immune system pathways. This is not surprising given the cellular source of the mRNA used in our analysis. Noteworthy, this finding is in line with previous studies, which have shown that mutant ataxin-3 mediates up-regulation of several cytokines and cytokine-inducible transcription factors in cell models and in brain tissue of MJD patients [17,20,42], although none of the previously reported genes were up-regulated in our study. Immune system alterations have also been reported in peripheral blood of HD patients, inclusively before the manifestation of motor symptoms

[31,43,44]. Globally, it seems that genes involved in immune system and GPCRs may constitute good candidates for biomarkers in polyQ disorders, and should be considered in future studies of MJD progression.

In summary, our results support the hypothesis that mutated ataxin-3 is associated with transcription dysregulation in peripheral blood cells. Based on the transcriptional profile of MJD patients, this study identified for the first time a pool of up-regulated genes in this disease. The understanding of how the up-regulation of these genes relates to clinical-associated progressive features during the natural history of MJD will be crucial in clinical trials measuring the effectiveness of new drugs.

AUTHOR ROLES

1. Research project: A. Conception: Raposo, Lima, Bettencourt C. and Maciel; B. Organization: Raposo, Lima, Bettencourt, Maciel, Geschwind, Coppola, Lima, Bruges-Armas, Vasconcelos, Kay and Bettencourt B.; C. Execution: Raposo, Lima, Bettencourt C. and Maciel.

2. Statistical Analysis: A. Design, B. Execution: Raposo and Gao; C. Review and Critique: Raposo, Gao, Coppola, Lima, Bettencourt C, Rodrigues, Ramos and Kazachkova.

3. Manuscript Preparation: A. Writing of the first draft: Raposo, Lima, Bettencourt C, Maciel and Coppola; B. Review and Critique Critical revision of the manuscript for important intellectual content: Ramos, Kazachkova, Rodrigues, Bettencourt B, Bruges-Armas and Geschwind. Statistical analysis: Raposo and Gao. Obtained funding: Lima. Administrative, technical, and material support: Study supervision: Lima and Bettencourt C.

FINANCIAL DISCLOSURE

Nothing to report.

FUNDING SOURCES

This work was funded by FEDER funds through the Operational Competitiveness Programme – COMPETE and by National Funds through FCT – Fundação para a Ciência e a Tecnologia under the project FCOMP-01-0124-FEDER-028753 (PTDC/DTP/PIC/0370/2012). A PhD fellowship M3.1.2/F/006/2011 (M.R.) and postdoctoral fellowships M3.1.7/F/031/2011 (A.R.) and M3.1.3/F/004/2009 (N.K.) were supported by Fundo Regional para a Ciência (FRC), Governo dos Açores, and SFRH/BPD/33611/2009 (A.J.R.) was funded by FCT – Fundação para a Ciência e a Tecnologia. C.B. is supported by the UK Medical Research Council (MRC).

REFERENCES

1. Schöls L, Bauer P, Schmidt T, Schulte T, Riess O. Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. Lancet Neurol. 2004;3:291–304.

2. Pringsheim T, Wiltshire K, Day L, Dykeman J, Steeves T, Jette N. The incidence and prevalence of Huntington's disease: a systematic review and meta-analysis. Mov. Disord. 2012;27:1083–91.

3. Coutinho P, Andrade C. 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. 1978;28:703–9.

4. Lima L, Coutinho P. Clinical criteria for diagnosis of Machado-Joseph disease: report of a non-Azorena Portuguese family. Neurology. 1980;30:319–22.

5. Seidel K, Siswanto S, Brunt ERP, den Dunnen W, Korf H-W, Rüb U. Brain pathology of spinocerebellar ataxias. Acta Neuropathol. 2012;124:1–21.

6. Bettencourt C, Lima M. Machado-Joseph Disease: from first descriptions to new perspectives. Orphanet J. Rare Dis. 2011;6:35.

7. Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, et al. The gene for Machado-Joseph disease maps to human chromosome 14q. Nat. Genet. 1993;4:300–4.

8. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32. 1. Nat. Genet. 1994;8:221–8.

9. Evers MM, Toonen LJA, van Roon-Mom WMC. Ataxin-3 protein and RNA toxicity in spinocerebellar ataxia type 3: current insights and emerging therapeutic strategies. Mol. Neurobiol. 2014;49:1513–31.

10. Silva SD da, Jalles A, Maciel P. Therapeutic strategies for polyQ diseases : from cellular and animal models to the clinic. In: Almeirao E, Honrado T, editors. Neuropathol. New Res. New York: Nova Science Publishers Inc; 2013.

11. Koeppen AH. The pathogenesis of spinocerebellar ataxia. Cerebellum. 2005;4:62–73.

12. Perez MK, Paulson HL, Pendse SJ, Saionz SJ, Bonini NM, Pittman RN. Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. J. Cell Biol. 1998;143:1457–70.

13. Takahashi J, Tanaka J, Arai K, Funata N, Hattori T, Fukuda T, et al. Recruitment of nonexpanded polyglutamine proteins to intranuclear aggregates in neuronal intranuclear hyaline inclusion disease. J. Neuropathol. Exp. Neurol. 2001;60:369–76.

14. Chai Y, Wu L, Griffin JD, Paulson HL. The role of protein composition in specifying nuclear inclusion formation in polyglutamine disease. J. Biol. Chem. 2001;276:44889–97.

15. Li F, Macfarlan T, Pittman RN, Chakravarti D. Ataxin-3 is a histone-binding protein with two independent transcriptional corepressor activities. J. Biol. Chem. 2002;277:45004–12.

16. Nicastro G, Menon RP, Masino L, Knowles PP, McDonald NQ, Pastore A. The solution structure of the Josephin domain of ataxin-3: structural determinants for molecular recognition. Proc. Natl. Acad. Sci. U. S. A. 2005;102:10493–8.

17. Evert BO, Vogt IR, Vieira-Saecker AM, Ozimek L, de Vos RAI, Brunt ERP, et al. Gene expression profiling in ataxin-3 expressing cell lines reveals distinct effects of normal and mutant ataxin-3. J. Neuropathol. Exp. Neurol. 2003;62:1006–18.

18. Evert BO, Araujo J, Vieira-Saecker AM, de Vos RAI, Harendza S, Klockgether T, et al. Ataxin-3 represses transcription via chromatin binding, interaction with histone deacetylase 3, and histone deacetylation. J. Neurosci. 2006;26:11474–86.

19. Rodrigues A-J, Coppola G, Santos C, Costa M do C, Ailion M, Sequeiros J, et al. Functional genomics and biochemical characterization of the C. elegans orthologue of the Machado-Joseph disease protein ataxin-3. FASEB J. 2007;21:1126–36.

20. Evert BO, Vogt IR, Kindermann C, Ozimek L, de Vos RA, Brunt ER, et al. Inflammatory genes are upregulated in expanded ataxin-3-expressing cell lines and spinocerebellar ataxia type 3 brains. J. Neurosci. 2001;21:5389–96.

21. Chou A-H, Yeh T-H, Ouyang P, Chen Y-L, Chen S-Y, Wang H-L. Polyglutamine-expanded ataxin-3 causes cerebellar dysfunction of SCA3 transgenic mice by inducing transcriptional dysregulation. Neurobiol. Dis. 2008;31:89–101.

22. Zesiewicz TA, Greenstein PE, Sullivan KL, Wecker L, Miller A, Jahan I, et al. A randomized trial of varenicline (Chantix) for the treatment of spinocerebellar ataxia type 3. Neurology. 2012;78:545–50.

23. Saute JAM, de Castilhos RM, Monte TL, Schumacher-Schuh AF, Donis KC, D'Ávila R, et al. A randomized, phase 2 clinical trial of lithium carbonate in Machado-Joseph disease. Mov. Disord. 2014;29:568–73.

24. Coppola G, Geschwind DH. Technology Insight: querying the genome with microarrays-progress and hope for neurological disease. Nat. Clin. Pract. Neurol. 2006;2:147–58.

25. Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, et al. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clin. Chem. 2009;55:611–22.

26. Stamova BS, Apperson M, Walker WL, Tian Y, Xu H, Adamczy P, et al. Identification and validation of suitable endogenous reference genes for gene expression studies in human peripheral blood. BMC Med. Genomics. 2009;2:49.

27. Diamanti D, Lahiri N, Tarditi A, Magnoni L, Fondelli C, Morena E, et al. Reference genes selection for transcriptional profiling in blood of HD patients and R6/2 mice. J. Huntingtons. Dis. 2013;2:185–200.

28. Coppola G, Burnett R, Perlman S, Versano R, Gao F, Plasterer H, et al. A gene expression phenotype in lymphocytes from Friedreich ataxia patients. Ann. Neurol. 2011;70:790–804.

29. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. 2007;8:118–27.

30. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. 2008;3:1101–8.

31. Björkqvist M, Wild EJ, Thiele J, Silvestroni A, Andre R, Lahiri N, et al. A novel pathogenic pathway of immune activation detectable before clinical onset in Huntington's disease. J. Exp. Med. 2008;205:1869–77.

32. Aebischer J, Cassina P, Otsmane B, Moumen A, Seilhean D, Meininger V, et al. IFNγ triggers a LIGHT-dependent selective death of motoneurons contributing to the non-cell-autonomous effects of mutant SOD1. Cell Death Differ. 2011;18:754–68.

33. Cui Y, Le Y, Yazawa H, Gong W, Wang JM. Potential role of the formyl peptide receptor-like 1 (FPRL1) in inflammatory aspects of Alzheimer's disease. J. Leukoc. Biol. 2002;72:628–35.

34. Ridwan S, Bauer H, Frauenknecht K, von Pein H, Sommer CJ. Distribution of granulocyte-monocyte colony-stimulating factor and its receptor α-subunit in the adult human brain with specific reference to Alzheimer's disease. J. Neural Transm. 2012;119:1389–406.

35. Aebischer J, Moumen A, Sazdovitch V, Seilhean D, Meininger V, Raoul C. Elevated levels of IFNγ and LIGHT in the spinal cord of patients with sporadic amyotrophic lateral sclerosis. Eur. J. Neurol. 2012;19:752–9, e45-6.

36. Gonçalves N, Simões AT, Cunha RA, de Almeida LP. Caffeine and adenosine A(2A) receptor inactivation decrease striatal neuropathology in a lentiviral-based model of Machado-Joseph disease. Ann. Neurol. 2013;73:655–66.

37. Borovecki F, Lovrecic L, Zhou J, Jeong H, Then F, Rosas HD, et al. Genome-wide expression profiling of human blood reveals biomarkers for Huntington's disease. Proc. Natl. Acad. Sci. U. S. A. 2005;102:11023–8.

38. Fuller JP, Stavenhagen JB, Teeling JL. New roles for Fc receptors in neurodegeneration-the impact on Immunotherapy for Alzheimer's Disease. Front. Neurosci. 2014;8:235.

39. Carlow DA, Gossens K, Naus S, Veerman KM, Seo W, Ziltener HJ. PSGL-1 function in immunity and steady state homeostasis. Immunol. Rev. 2009;230:75–96.

40. Angiari S, Constantin G. Selectins and their ligands as potential immunotherapeutic targets in neurological diseases. Immunotherapy. 2013;5:1207–20.

41. Rodrigues A-J, do Carmo Costa M, Silva T-L, Ferreira D, Bajanca F, Logarinho E, et al. Absence of ataxin-3 leads to cytoskeletal disorganization and increased cell death. Biochim. Biophys. Acta. Elsevier B.V.; 2010;1803:1154–63.

42. Evert BO, Schelhaas J, Fleischer H, de Vos RAI, Brunt ER, Stenzel W, et al. Neuronal intranuclear inclusions, dysregulation of cytokine expression and cell death in spinocerebellar ataxia type 3. Clin. Neuropathol. 2006;25:272–81.

43. Ellrichmann G, Reick C, Saft C, Linker RA. The role of the immune system in Huntington's disease. Clin. Dev. Immunol. 2013;2013:541259.

44. Tai YF, Pavese N, Gerhard A, Tabrizi SJ, Barker RA, Brooks DJ, et al. Microglial activation in presymptomatic Huntington's disease gene carriers. Brain. 2007;130:1759–66.

CHAPTER VI

HSPB1 and BCL2 mRNA levels in blood of spinocerebellar ataxia type 3 subjects: results from cross-sectional and longitudinal analyses.

In prep.

Mafalda Raposo[1,2,3], Conceição Bettencourt[4], João Vasconcelos[5], Teresa Kay[6], Manuela Lima[1,2,3]

1. Faculdade de Ciências e Tecnologia, Universidade dos Açores, Ponta Delgada, Portugal
2. Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
3. Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Porto, Portugal
4. Department of Clinical and Experimental Epilepsy and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
5. Departamento de Neurologia, Hospital do Divino Espírito Santo, Ponta Delgada, Portugal
6. Departamento de Genética Clinica, Hospital D. Estefânia, Lisboa, Portugal

ABSTRACT

The development of molecular biomarkers for spinocerebellar ataxia type 3 (SCA3), a currently untreatable late-onset neurodegenerative disorder, is challenging. Transcriptional dysregulation, recently confirmed in blood samples of SCA3 patients, prompted the investigation of dysregulated mRNA levels as a way to identify candidate biomarkers. To pursuit this hypothesis, the expression patterns of nine genes - *HSPB1*, *DNAJB1*, *DNAJB12*, *DNAJB14*, *BAX*, *BCL2*, *SOD2*, *IL1B* and *IL6* genes, previously reported as dysregulated in SCA3, was investigated by quantitative real-time PCR. After a cross-sectional analysis enrolling 16 preataxic subjects, 74 SCA3 patients and 36 community controls, we retrieved the most relevant dysregulated genes and further evaluated their transcript levels in a longitudinal analysis, using a subset of 14 SCA3 patients for which samples from two visits were available. Decreased levels of *IL6* and *BCL2* mRNA were observed in preataxic subjects. Lower *HSPB1* and *BCL2* mRNA levels after the ataxia onset were, furthermore, evidenced. *HSPB1* mRNA levels evaluated in a longitudinal design tended to increase, although a larger cohort will be need to confirm this finding. All 14 patients shown a significant decrease of *BCL2* mRNA levels during disease progression. Although several steps of validation, namely testing in other SCA3 cohorts and correlations with clinical rating scores, *BCL2* can be considered a promising biomarker of SCA3 progression. Such molecular markers will be mandatory to complement clinical and imaging measures, already at use in interventional trials.

INTRODUCTION

Despite the fact that polyglutamine (polyQ) disorders remain untreatable, interventional studies have become a reality for a number of them. The success of clinical trials, however, is dependent on the improvement of sensitivity and reliability of current clinical and imaging outcome measures. In fact, limitations of these measures are widely acknowledged: imaging remains an expensive and time-consuming approach, whereas clinical measures can be subjective, not being suitable, moreover, for studies including preclinical subjects [1]. The identification of molecular biomarkers is therefore of particular importance, since these should allow the fine tracking of disease progression, starting from the preclinical stage, as well as enable the detection of subtle therapeutic benefits. The spinocerebellar ataxias (SCAs) are a subgroup of polyQ diseases, from which the most prevalent is the type 3 (SCA3; Machado-Joseph disease - MJD). SCA3 is a late-onset degenerative disorder affecting mainly the cerebellum and brainstem [2]; this disorder is caused by an instable CAG repeat motif in the *ATXN3* gene [3,4], encoding for the ubiquitously expressed protein ataxin-3 [5]. An expanded polyQ tract, consensually corresponding to more than 50 glutamines [6] results in a toxic gain of ataxin-3 function, leading to selective neuronal death [5]. A complex net of mechanisms, which includes transcriptional dysregulation, is implicated in SCA3 pathogenesis (reviewed in [7]). Alterations in transcription have been recently reported in blood samples of SCA3 patients [8], further confirming that dysregulation in the presence of mutated ataxin-3 is not limited to the central nervous system. In a microarray based cross-sectional study we have previously identified three genes, *FCGR3B* (Fc fragment of IgG, low-affinity IIIb, receptor (CD16b)), *P2RY13* (Purinergic receptor P2Y, G-protein coupled, 13) and *SELPLG* (Selectin P ligand), whose expression was significantly dysregulated in blood samples of SCA3 patients; moreover, *FCGR3B* and *CLC* (Charcot-Leyden crystal protein) mRNA levels significantly correlated with disease duration [8].

To further improve our knowledge on gene expression alterations in SCA3 we searched in the literature for molecules reported as being dysregulated in the presence of ataxin-3. In this study a candidate-approach was used to investigate the expression patterns of nine genes - *HSPB1* (heat shock 27kDa protein 1), *DNAJB1* (DnaJ (Hsp40) homolog, subfamily B, member 1), *DNAJB12* (DnaJ (Hsp40) homolog, subfamily B, member 12), *DNAJB14* (DnaJ (Hsp40) homolog, subfamily B, member 14), *BAX* (BCL2-associated X protein), *BCL2* (B-cell CLL/lymphoma 2), *SOD2* (superoxide dismutase 2, mitochondrial), *IL1B* (interleukin 1, beta) and *IL6* (interleukin 6) genes. After a cross-sectional analysis enrolling preataxic subjects, SCA3 patients and community controls, we retrieved the most relevant dysregulated genes and further evaluated their transcript levels in a longitudinal analysis, using a subset of SCA3 patients for which samples from two different visits were available.

The present study should contribute to the identification of molecular markers of SCA3, whose behavior can be associated with different stages of disease progression, including the preclinical stage. Such molecular markers will be mandatory to complement clinical and imaging measures, already at use in interventional trials.

SUBJECTS AND METHODS

Subjects

Cross-sectional analysis

RNA samples from 74 patients and 16 preataxic subjects as well as from 36 apparently healthy individuals (community controls) were used in this study. Preataxic subjects were defined in accordance with Maas and colleagues [9]. Age at onset was defined as the age of appearance of the first symptoms (gait disturbance and/or diplopia) reported by the patient and/or a close relative. The size of CAG repeats in the normal and expanded alleles was determined in a single

laboratory, using the protocol described by Bettencourt and colleagues [10]. A summary of genetic and clinical features of the SCA3 subjects analysed in this study is shown in Table 1. All participants completed the informed consent process and signed a consent form.

Table 1. Characterization of the SCA3 subjects included in the present study.

		Patients N=74	**Preataxic subjects** N=16	**Community controls** N=36
Female|Male		33|41	10|6	26|10
Age at collection		50±14 | [22-82] | 50	30±8 | [21-44] | 29	42±13 | [22-77] | 41
CAG repeat size	Normal allele	21±5 | [14-29] | 23	20±5 | [14-28] | 21	
	Expanded allele	71±4 | [62-79] | 71	69±4 | [62-75] | 68	
Age at onset		38±12 | [13-71] | 36		
Disease duration		12±9 | [0-36] | 12		

[1]mean ± standard deviation | [minimum - maximum] | median. Age at collection, age at onset and disease duration are shown in years.

Longitudinal analysis

For 14 of the 74 patients used in the cross-sectional analysis, samples collected at two distinct visits were available. Intervals between the baseline of the study and at the second visit ranged from one to seven years.

Gene selection

A literature search in Pubmed [11] was conducted to identify studies reporting dysregulation of mRNA and/or protein levels in the presence of mutated ataxin-3. Transcripts/proteins whose expression levels were reported as being dysregulated were listed (Table 2); those whose expression pattern (up- or down-regulation) was further supported with data from our previous microarray experiment [8], were selected: *HSPB1* [12–15]*, DNAJB1* [13,14]*, DNAJB12, DNAJB14, BAX* [16–18]*, BCL2* [16,17,19]*, SOD2* [20]*, IL1B* [21] and *IL6* [21,22] (Table 2).

Table 2. Genes selected for qPCR analysis. mRNA levels of *HSPB1, DNAJB1, DNAJB12, DNAJB14, BAX, BCL2, SOD2, IL1B* and *IL6* were quantified in 16 preataxic subjects, 74 SCA3 patients as well as 36 community controls.

GENE			PREVIOUS STUDIES					TaqMan assay ID
ID	**Symbol[1]**	**Name[1]**	**mRNA/protein levels**	**Dysregulation pattern**	**Organism**	**Tissue**	**Reference**	
3315	*HSPB1*	Heat shock 27kDa protein 1	Protein	Up-regulation	*Homo sapiens*	Fibroblasts	[14]	Hs03044127_g1
						Brain	[12]	
					Mus musculus	Brain		
					Homo sapiens	SK-N-SH cells		
				Down-regulation	*Homo sapiens*	SK-N-SH cells	[15]	
						Lymphoblastoid cells		
						HEK293 cells	[13]	
3337	*DNAJB1*	DnaJ (Hsp40) homolog, subfamily B, member 1	Protein	Down-regulation (NS)*	*Homo sapiens*	Fibroblasts	[14]	Hs00428680_m1
			mRNA Protein	Down-regulation	*Mus musculus*	Brain	[13]	
54788	*DNAJB12*	DnaJ (Hsp40) homolog, subfamily B, member 12	Selected from microarray study					Hs00430067_m1
79982	*DNAJB14*	DnaJ (Hsp40) homolog, subfamily B, member 14	Selected from microarray study					Hs01016831_m1
581	*BAX*	BCL2-associated X protein	Protein	Up-regulation	*Rattus norvegicus*	Cultured neurons	[16]	Hs00180269_m1
			mRNA Protein	Up-regulation	*Mus musculus*	Brain	[17]	
				Down-regulation (NS)	*Homo sapiens*	SK-N-SH cells	[18]	

Table 2. Cont.

GENE			PREVIOUS STUDIES					TaqMan assay ID
ID	**Symbol[1]**	**Name[1]**	**mRNA/protein levels**	**Dysregulation pattern**	**Organism**	**Tissue**	**Reference**	
596	*BCL2*	B-cell CLL/lymphoma 2	Protein	Down-regulation (NS)	*Rattus norvegicus*	Cultured neurons	[16]	Hs00608023_m1
			mRNA	Down-regulation	*Cercopithecus aethiops*	COS7 cells	[19]	
			Protein					
			mRNA		*Homo sapiens*	SK-N-SH cells	[18]	
			Protein			Fibroblasts		
6648	*SOD2*	Superoxide dismutase 2, mitochondrial	mRNA	Down-regulation	*Homo sapiens*	Brain (pons)	[20]	Hs00167309_m1
						Lymphoblastoid cells		
			Protein	Down-regulation (NS)		Brain (pons)		
						Lymphoblastoid cells		
3553	*IL1B*	Interleukin 1, beta	Protein	Up-regulation	*Homo sapiens*	Brain (pons)	[21]	Hs01555410_m1
3569	*IL6*	Interleukin 6	mRNA	Up-regulation	*Rattus norvegicus*	CSM14.1 cells	[21,22]	Hs00985639_m1
			Protein		*Homo sapiens*	Brain (pons)		

[1]Gene symbol and name provided according to HUGO Gene Nomenclature Committee (HGNC); NS = not statistically significant; *The antibody used was not specific for DNAJB1 immunostaining.

Quantitative real-time PCR (qPCR)

mRNA levels were quantified as described elsewhere [8]. The pre-validated TaqMan Gene Expression Assays used in this study are listed in Table 2. Relative expression values, normalized to the reference gene (*PPIB*), were calculated using the $2^{-\Delta Ct}$ method [23].

Statistical procedure

Normal distribution of continuous variables was tested using the Kolmogorov-Smirnov test; when violation of the assumption of normality was verified, non-parametric tests were performed. If no alternative for the non-parametric tests was available (e.g. ANCOVA), log-transformed variables were used.

Cross-sectional analysis

mRNA levels were compared between biological groups by the ANCOVA procedure; p-value was corrected using the Bonferroni method. Correlations between mRNA levels, CAG size in expanded allele, age at onset and disease duration were obtained by the partial correlation test. Age at blood collection was used as covariate in ANCOVA as well as in the partial correlation test.

Longitudinal analyses

In order to account for irregular intervals between patients visits, unstandardized residuals from a linear regression model were obtained. Briefly, a regression model was created using the age at blood collection and the disease duration as predictors and mRNA level for each gene as the dependent variable; the difference between the quantified mRNA level and the value predicted by the model is the unstandardized residual (supplementary Table 1). For each patient the estimated mRNA level was obtained by the difference between the quantified mRNA expression value and the unstandardized residual value. Unstandardized residuals were calculated at the baseline, as well as at the second visit. Estimated mRNA values were further tested using statistical procedures for related samples (Wilcoxon Matched-Pair Signed Rank).

All statistical tests were run in IBM SPSS Statistics 22; p-value lower than 0.05 was considered statistically significant.

RESULTS

HSPB1, BCL2 and IL6 mRNA levels are altered in non-ataxic and ataxic stages

Results from the cross-sectional analyses are shown in Figure 1. For *HSPB1* contradictory results have been reported (Table 2); this gene was down-regulated in our set of preataxic subjects as well as patients, reaching significance in these last. Although not significant, a tendency for decreased *HSPB1* levels during progression was evidenced. *BCL2* maintained the dysregulation trend reported in the literature (Table 2) being significantly down-regulated in preataxic subjects as well as in SCA3 patients. *BCL2* mRNA levels tended to increase during progression of the disease, however, differences were not significant. *IL6* mRNA levels, reported in the literature as increased in the presence of ataxin-3 (Table 2), were decreased in carriers of the *ATXN3* mutation, reaching significance in the comparison between preataxic subjects and controls; *IL6* mRNA levels, however, shown a high inter-individual variability for all groups analyzed.

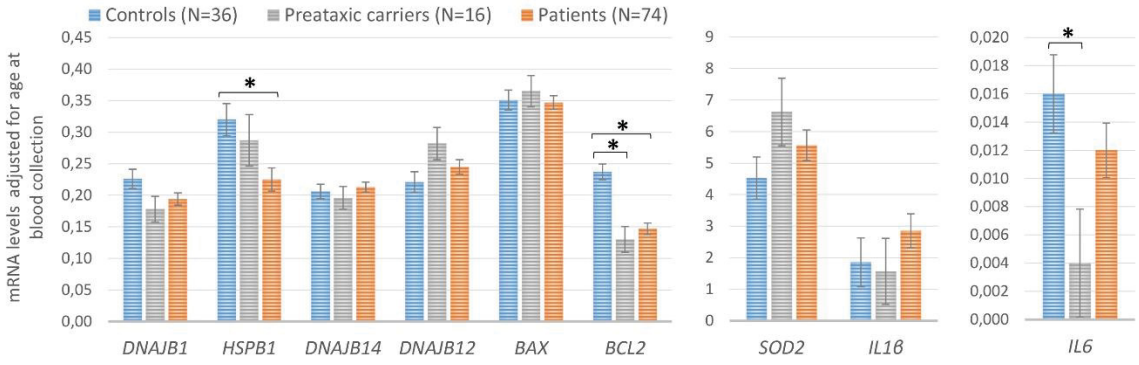


Figure 1. mRNA levels for the nine genes under study in preataxic subjects, patients and controls. mRNA levels were adjusted for age at blood collection (45 years). *p<0.05 was corrected for multiple comparisons using the Bonferroni procedure.

BCL2 mRNA levels are altered during disease progression

Based on the cross-sectional analysis, the most relevant genes differentially expressed in patients as compared with controls - *HSPB1* and *BCL2* - were selected for a longitudinal study, aiming to analyse the behaviour of these genes in follow-up samples. In the set of 14 patients analysed, *HSPB1* mRNA adjusted levels decreased in 9 patients between the baseline and the second visit, although differences were not significant. *BCL2* mRNA adjusted levels were found to be significantly increased between the baseline and the second visit (z=-3.296, p=0.001; Figure 2); noteworthy, the same expression behaviour (increase in expression) was observed for all 14 patients.

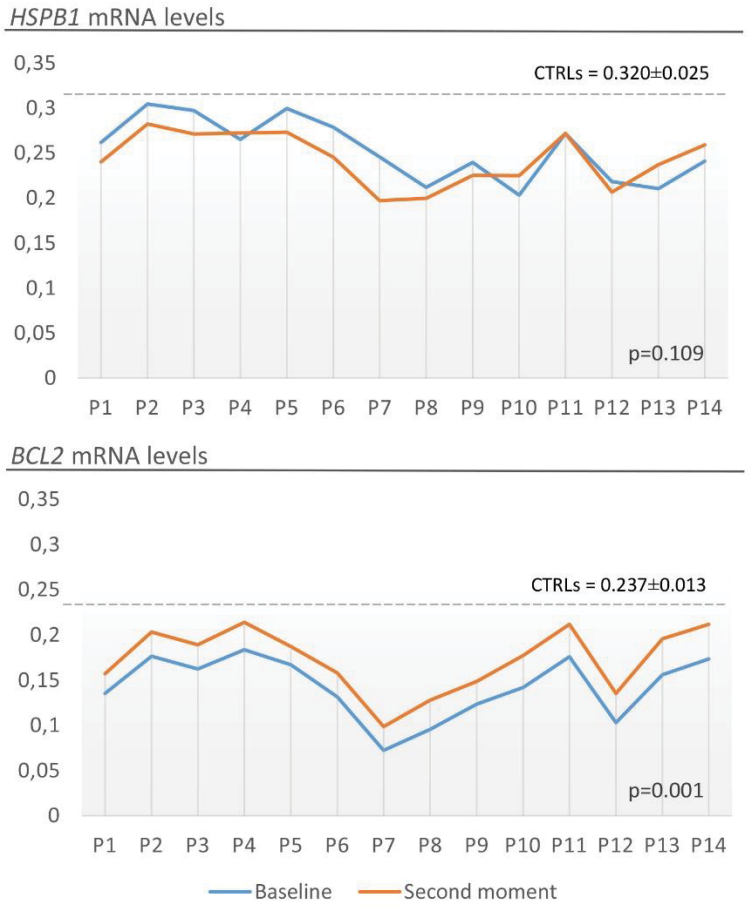


Figure 2. *HSPB1* and *BCL2* mRNA levels for the 14 patients used in the longitudinal analysis. Estimated mRNA levels (after removing the effect of age at blood collection and disease duration in each visit) were compared by the Wilcoxon Matched-Pair Signed Rank test. The mean ± standard error of mRNA levels quantified in controls is also shown.

DISCUSSION

In the present work nine candidate genes, whose expression had been previously reported as altered in the presence of mutant ataxin-3 were analysed in SCA3 preataxic carriers, patients and controls. When compared to controls *HSPB1, BCL2* and *IL6* mRNA levels were altered in SCA3 subjects. For *HSPB1,* a decrease tendency during disease progression was observed, although differences between mRNA levels of preataxic carriers and patients were not significant. HSPB1 was further tested in the longitudinal setup for which results confirm a trend for decrease during progression, but failed to reaching significance. The consistent tendency for decrease in SCA3, starting from the preclinical phase, however, cannot be ruled out since longitudinal study could be underpowered. HSPB1 belongs to the family of small heat-shock proteins, whose role in neurodegenerative diseases has been widely investigated (reviewed in [24]). Results from previous studies concerning HSPB1 levels are not consensual (Table 2); our results was in agreement with findings observed in a SCA3 mutant neuronal cell line and in lymphoblastoid cell lines [15]. The dual role of HSPB1, acting as chaperone or having anti-apoptotic properties (reviewed in [25]), could be on the basis of differences in expression levels between studies. The up-regulation of HSPB1 levels in SCA3 brain tissues [12] could be a compensatory mechanism to revert cell death in late stages of the disease.

BCL2 mRNA levels were significantly down-regulated both in preataxic subjects as well as in patients compared to controls. The cross-sectional analysis also revealed that, in SCA3 subjects, *BCL2* levels show a non-significant tendency for increasing from the preataxic to the disease stage. This tendency is in agreement with the significant increase in *BCL2* levels observed in the longitudinal study. *BCL2* encodes an anti-apoptotic protein named B-cell CLL/lymphoma 2 (reviewed in [26]), for which low levels in the presence of mutant ataxin-3 were described (Table 2). Tien and colleagues have described that decreased levels of BCL2 may be caused by defects in mRNA stability; *BCL2* mRNA decay was ~3.5 fold faster in a mutant SCA3 cell line (COS7-MJD78-

GFP) compared to parental cells (COS7-MJD26-GFP) [19]. In SK-N-SH-MJD78 cells, a cell line which mimics early stages of SCA3 disease, a higher number of apoptotic cells were found when compared to SK-N-SH-MJD26 control cells [18]; moreover, overexpression of *BCL2* protected mutant cells against polyQ-induced cell death [18]. We can postulate that an increase in *BCL2* expression levels during disease progression represents an attempt to modulate cell death.

In the cross-sectional study, significantly decreased levels of *IL6* mRNA levels were observed in preataxic subjects; the high inter-individual variability for the expression values one this gene was, however, acknowledged. *IL6*, a member of the cytokine family, is a modulator of the immune system whose association with other polyQ disorders, namely Huntington disease, has been reported (reviewed in [27]). A previous study by Carvalho and collaborators [28], using blood samples from patients and from a similar number of preataxic SCA3 carriers also failed to detect significant differences in the amounts of interleukin-6. Furthermore, no dysregulation was detected in blood from a genetrap SCA3 mouse model without disease phenotype [29]. On the other hand, higher levels of interleukin-6 staining have been described in brain tissues of SCA3 patients [21,22]; the same up-regulation trend has been reported in a rat model of SCA3 [22]. Non-consensual results available for *IL6* in SCA3 could be related with several factors, namely: a) studies from Evert and colleagues [21,22] used a limited number of SCA3 brain samples, which represented the final stage of the disease; therefore, the reported inflammatory process could be reflecting the general neuronal death and not the presence of the mutant ataxin-3 itself; and b) other studies [29] for *IL6* also report an important degree of inter-individual variability for the expression levels, similarly to what was seen in this study, corroborating the hypothesis that such levels are affected by several non-specific factors.

Although further work will be necessary to confirm the value of *BCL2 and HSPB1* as transcriptional biomarkers of SCA3, both can be considered promising candidates. To our knowledge, this is the first study investigating molecular alterations in longitudinal blood samples of SCA3 patients; due

to the limited group size, however, longitudinal analyses with more samples would be further required, as well as validation in independent cohorts. Moreover, correlations of mRNA levels with clinical measures and imaging data will also be mandatory. Because these genes are implicated in pathways related to mechanisms involved in disease pathogenesis, namely apoptosis and protein quality control, evaluating their expression levels during a therapeutic intervention may also be potentially relevant.

ACKNOWLEDGEMENTS

This work was funded by FEDER funds through the Operational Competitiveness Programme – COMPETE and by National Funds through FCT – Fundação para a Ciência e a Tecnologia under the project FCOMP-01-0124-FEDER-028753 (PTDC/DTP/PIC/0370/2012). MR (PhD fellowship, M3.1.2/F/006/2011) is supported by Fundo Regional para a Ciência (Governo dos Açores) and CB by the Wellcome Trust (UK).

REFERENCES

1. Andre R, Scahill RI, Haider S, Tabrizi SJ. Biomarker development for Huntington's disease. Drug Discov. Today. 2014;19:972–9.

2. Bürk K, Abele M, Fetter M, Dichgans J, Skalej M, Laccone F, et al. Autosomal dominant cerebellar ataxia type I clinical features and MRI in families with SCA1, SCA2 and SCA3. Brain. 1996;119 (Pt 5):1497–505.

3. Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, et al. CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat. Genet. 1994;8:221–8.

4. Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, et al. The gene for Machado-Joseph disease maps to human chromosome 14q. Nat Genet. 1993;4:300–4.

5. Paulson HL, Das SS, Crino PB, Perez MK, Patel SC, Gotsdiner D, et al. Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. Ann. Neurol. 1997;41:453–62.

6. Gan S-R, Ni W, Dong Y, Wang N, Wu Z-Y. Population genetics and new insight into range of CAG repeats of spinocerebellar ataxia type 3 in the Han Chinese population. PLoS One. 2015;10:e0134405.

7. Evers MM, Toonen LJA, Van Roon-Mom WMC. Ataxin-3 protein and RNA toxicity in spinocerebellar ataxia type 3: Current insights and emerging therapeutic strategies. Mol. Neurobiol. 2014. p. 1513–31.

8. Raposo M, Bettencourt C, Maciel P, Gao F, Ramos A, Kazachkova N, et al. Novel candidate blood-based transcriptional biomarkers of Machado-Joseph disease. Mov. Disord. 2015;30:968–75.

9. Maas RPPWM, van Gaalen J, Klockgether T, van de Warrenburg BPC. The preclinical stage of spinocerebellar ataxias. Neurology. 2015;85:96–103.

10. Bettencourt C, Fialho RN, Santos C, Montiel R, Bruges-Armas J, MacIel P, et al. Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: A study in normal families from the Azores islands (Portugal). J. Hum. Genet. 2008;53:333–9.

11. National Center for Biotechnology Information. Pubmed [Internet]. Available from: https://www.ncbi.nlm.nih.gov/pubmed

12. Chang WH, Cemal CK, Hsu YH, Kuo CL, Nukina N, Chang MH, et al. Dynamic expression of Hsp27 in the presence of mutant ataxin-3. Biochem. Biophys. Res. Commun. 2005;336:258–67.

13. Vos MJ, Zijlstra MP, Kanon B, van Waarde-Verhagen MAWH, Brunt ERP, Oosterveld-Hut HMJ, et al. HSPB7 is the most potent polyQ aggregation suppressor within the HSPB family of molecular chaperones. Hum. Mol. Genet. 2010;19:4677–93.

14. Zijlstra MP, Rujano MA, Van Waarde MA, Vis E, Brunt ER, Kampinga HH. Levels of DNAJB family members (HSP40) correlate with disease onset in patients with spinocerebellar ataxia type 3. Eur. J. Neurosci. 2010;32:760–70.

15. Wen F-C, Li Y-H, Tsai H-F, Lin C-H, Li C, Liu C-S, et al. Down-regulation of heat shock protein 27 in neuronal cells and non-neuronal cells expressing mutant ataxin-3. FEBS Lett. 2003;546:307–14.

16. Chou AH, Yeh TH, Kuo YL, Kao YC, Jou MJ, Hsu CY, et al. Polyglutamine-expanded ataxin-3 activates mitochondrial apoptotic pathway by upregulating Bax and downregulating Bcl-xL. Neurobiol. Dis. 2006;21:333–45.

17. Chou A-H, Yeh T-H, Ouyang P, Chen Y-L, Chen S-Y, Wang H-L. Polyglutamine-expanded ataxin-3 causes cerebellar dysfunction of SCA3 transgenic mice by inducing transcriptional dysregulation. Neurobiol. Dis. 2008;31:89–101.

18. Tsai HF, Tsai HJ, Hsieh M. Full-length expanded ataxin-3 enhances mitochondrial-mediated cell death and decreases Bcl-2 expression in human neuroblastoma cells. Biochem. Biophys. Res. Commun. 2004;324:1274–82.

19. Tien CL, Wen FC, Hsieh M. The polyglutamine-expanded protein ataxin-3 decreases bcl-2 mRNA stability. Biochem. Biophys. Res. Commun. 2008;365:232–8.

20. Araujo J, Breuer P, Dieringer S, Krauss S, Dorn S, Zimmermann K, et al. FOXO4-dependent upregulation of superoxide dismutase-2 in response to oxidative stress is impaired in spinocerebellar ataxia type 3. Hum. Mol. Genet. 2011;20:2928–41.

21. Evert B, Vogt I, Kindermann C, Ozimek L, de Vos R, Brunt E, et al. Inflammatory genes are upregulated in expanded ataxin-3-expressing cell lines and spinocerebellar ataxia type 3 brains. J. Neurosci. 2001;21:5389–96.

22. Evert BO, Vogt IR, Vieira-Saecker AM, Ozimek L, de Vos RAI, Brunt ERP, et al. Gene expression profiling in ataxin-3 expressing cell lines reveals distinct effects of normal and mutant ataxin-3. J. Neuropathol. Exp. Neurol. 2003;62:1006–18.

23. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C(T) method. Nat. Protoc. 2008;3:1101–8.

24. Muchowski PJ, Wacker JL. Modulation of neurodegeneration by molecular chaperones. Nat. Rev. Neurosci. 2005;6:11–22.

25. Hsieh M, Tsai H-F, Chang W-H. HSP27 and cell death in spinocerebellar ataxia type 3. Cerebellum. 2005;4:31–6.

26. Siddiqui WA, Ahad A, Ahsan H. The mystery of BCL2 family: Bcl-2 proteins and apoptosis: an update. Arch. Toxicol. 2015;89:289–317.

27. Ellrichmann G, Reick C, Saft C, Linker RA. The role of the immune system in Huntington's disease. Clin. Dev. Immunol. 2013;2013:541259.

28. da Silva Carvalho G, Saute JAM, Haas CB, Torrez VR, Brochier AW, Souza GN, et al. Cytokines in Machado Joseph Disease/Spinocerebellar Ataxia 3. Cerebellum. 2015;

29. Hübener J, Casadei N, Teismann P, Seeliger MW, Björkqvist M, von Hörsten S, et al. Automated behavioral phenotyping reveals presymptomatic alterations in a SCA3 genetrap mouse model. J. Genet. Genomics. 2012;39:287–99.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Characterization of the 14 patients enrolled in the longitudinal analyses as well as the predicted values of *BCL2*, *DNAJB14* and *HSPB1* mRNA levels after adjusting for age at blood collection and disease duration at the baseline of the study and at a second visit.

	Age at onset	CAG size	
		N. allele	Exp. allele
P1	38	14	74
P2	50	14	65
P3	47	20	64
P4	45	27	72
P5	48	21	68
P6	40	14	70
P7	27	23	68
P8	25	21	68
P9	33	22	71

	mRNA levels at baseline							
			BCL2			*HSPB1*		
	Age at blood collection	Disease duration	obs.[1]	res.[2]	pred.[3]	obs.	res.	pred.
P1	57	19	0.129	-0.007	0.135	0.150	-0.112	0.262
P2	65	15	0.188	0.012	0.176	0.274	-0.030	0.304
P3	65	18	0.153	-0.010	0.162	0.342	0.045	0.297
P4	49	4	0.305	0.122	0.184	0.231	-0.034	0.265
P5	65	17	0.117	-0.050	0.167	0.309	0.009	0.300
P6	64	24	0.100	-0.031	0.131	0.455	0.176	0.279
P7	63	36	0.100	0.027	0.072	0.324	0.078	0.246
P8	46	21	0.132	0.037	0.096	0.162	-0.050	0.212
P9	51	18	0.119	-0.004	0.123	0.078	-0.161	0.240
P10	34	4	0.153	0.011	0.142	0.190	-0.013	0.203
P11	53	8	0.170	-0.006	0.176	0.267	-0.005	0.272
P12	47	20	0.085	-0.018	0.103	0.154	-0.064	0.218
P13	34	1	0.084	-0.072	0.156	0.488	0.278	0.210
P14	42	2	0.162	-0.011	0.174	0.124	-0.117	0.241

Supplementary Table 1. Cont.

	Age at onset	CAG size		*mRNA levels at second visit*								
							BCL2			*HSPB1*		
		N. allele	Exp. allele		Age at blood collection	Disease duration	obs.[1]	res.[2]	pred.[3]	obs.	res.	pred.
P10	30	23	73	P1	64	26	0.115	-0.042	0.157	0.113	-0.127	0.240
P11	45	14	68	P2	69	19	0.175	-0.028	0.203	0.262	-0.021	0.282
P12	27	27	72	P3	69	22	0.258	0.069	0.189	0.267	-0.004	0.271
P13	33	23	73	P4	53	8	0.270	0.056	0.214	0.235	-0.037	0.272
P14	40	27	69	P5	72	24	0.198	0.011	0.187	0.279	0.006	0.273
				P6	68	28	0.133	-0.025	0.158	0.549	0.304	0.245
				P7	67	40	0.094	-0.005	0.099	0.161	-0.037	0.197
				P8	49	24	0.178	0.050	0.128	0.144	-0.056	0.200
				P9	57	24	0.153	0.004	0.149	0.183	-0.042	0.225
				P10	37	7	0.190	0.013	0.177	0.270	0.045	0.225
				P11	54	9	0.175	-0.037	0.212	0.239	-0.033	0.272
				P12	50	23	0.107	-0.029	0.135	0.177	-0.030	0.206
				P13	35	2	0.195	-0.001	0.196	0.416	0.179	0.237
				P14	43	3	0.175	-0.037	0.212	0.111	-0.148	0.259

N.= normal; Exp. = expanded; obs.[1] = observed mRNA value (quantification performed by qPCR); res.[2] = unstandardized residual calculated by the difference between an observed value and the value predicted by the regression model; pred.[3] = the difference between the observed mRNA value and the unstandardized residual.

CHAPTER VII

Mitochondrial DNA content and common deletion quantification in blood of spinocerebellar ataxia type 3 preataxic carriers and patients.

In prep.

Mafalda Raposo[1,2,3], Amanda Ramos[1,2,3], Nadiya Kazachkova[1,2,3], Balbina Teixeira[1], Conceição Bettencourt[4], Manuela Lima[1,2,3]

1. Faculdade de Ciências e Tecnologia, Universidade dos Açores, Ponta Delgada, Portugal
2. Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Portugal
3. Instituto de Biologia Molecular e Celular (IBMC), Universidade do Porto, Porto, Portugal
4. Department of Clinical and Experimental Epilepsy and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK

ABSTRACT

Mitochondrial DNA (mtDNA) damage and depletion has been inconsistently reported in spinocerebellar ataxia type 3 (SCA3), a late-onset proteinopathy. Correlations between mtDNA content or common deletion frequency and clinical features, as well as their behavior during the natural history of SCA3, including the preclinical stage, have never been investigated. This study aims to contribute to the clarification of the potential of these mtDNA dysfunction-associated alterations as markers of SCA3 progression. Sixteen preataxic SCA3 carriers, 93 patients and 103 population controls were enrolled in the study. The comparison of mtDNA content between Azorean SCA3 patients, preataxic carriers and controls failed to reveal significant differences. The common deletion was significantly more frequent in patients (1.8 times higher) and preataxic carriers (4.2 times higher) than in controls. This study suggests that the frequency of the common deletion could be a potential marker of the SCA3 preclinical stage. The investigation of its potential in larger and independent cohorts as well as in a longitudinal setup is worth pursuing.

INTRODUCTION

Mitochondrial DNA (mtDNA) depletion as well as increased number of deletions (namely the m.8470_13446del4977 deletion, also known as common deletion, an age-related marker), previously associated with neurodegeneration, have been described in spinocerebellar ataxia type 3 (SCA3) mutant cell and animal models [1–3], as well as in blood samples from SCA3 patients [1,4,5]. SCA3, also known as Machado-Joseph disease (MJD; MIM#109150; ORPHA98757), is a late-onset proteinopathy, inherited as an autosomal dominant trait and caused by an abnormal number of CAG repeats in the gene encoding for the ataxin-3 protein, *ATXN3* [reviewed in 6]. First clinical signs of SCA3, usually gait ataxia, occur at around 40 years of age [reviewed in 6], although pathophysiologic changes are thought to occur many years or even decades earlier (preclinical stage) [7]. Altered conformation of mutated ataxin-3 promotes a toxic gain of function, compromising several cellular mechanisms, namely mitochondrial function [reviewed in 8]. Although mtDNA damage and depletion have been reported in SCA3 patients' blood, previous reports were not conclusive, either because they lacked power, enrolling from 14 to 61 patients [1,4,5] or because age at sampling was not considered as covariate, a limitation that could be distorting the results [4,5]. Correlations between mtDNA content or common deletion and clinical features, namely age at onset and disease duration, as well as the behavior of both mtDNA alterations during the natural history of SCA3, including the preclinical stage, have not yet been reported.

To our knowledge, mtDNA depletion has been one of the most studied measures of mtDNA dysfunction in blood cells of SCA3 patients, which justifies its investigation in a larger and independent cohort, including SCA3 mutation carriers within the preclinical stage, whose results are reported in this paper. Furthermore, correlations between mtDNA content or common deletion frequency and the CAG size of expanded allele, age at onset as well as disease duration were tested. The investigation of the behavior of mtDNA depletion and damage during SCA3

progression could generate knowledge to be of use in interventional trials, namely those with compounds targeting mitochondrial-related pathways. Furthermore, early detection of mtDNA dysfunction, in a preclinical stage, is potentially of value in future preventive trials.

SUBJECTS AND METHODS

Sixteen preataxic carriers of the SCA3 mutation, 93 patients and 103 community controls were enrolled in this study. The preataxic carriers were classified as in Maas and colleagues [7]. Age at disease onset was defined as the age of the first appearance of gait disturbances, reported by the patient and/or a close relative. Disease duration was defined as the number of years elapsed from the reported onset to the time of blood collection. The number of CAG repeats at the *ATXN3* locus was determined using the protocol described by Bettencourt and colleagues [9]. The characterization of the SCA3 subjects and controls used in this study is displayed in Table 1. All participants provided written informed consent; this study is part of a project approved by the Ethics Committee of Hospital do Divino Espírito Santo (Ponta Delgada).

Table 1. Gender and age at blood sampling of SCA3 patients, preataxic subjects and controls analysed in the present study. Clinical and genetic features were described for patients and preataxic subjects.

	Patients	***Preataxic subjects***	***Controls***
	N=93	N=16	N=103
Gender (male|female)	46|57	5|11	48|45
Age at sampling (years)[#]	48±14 (17-82)	30±6 (22-43)	43±14 (18-77)
CAG repeat size			
Normal allele	22±5 (14-29)	20±4 (14-28)	
Expanded allele*	71±4 (62-79)	68±2 (65-74)	
Age at onset (years)	37±12 (13-71)		
Disease Duration (years)	11±9 (0-36)		

Values are presented as mean ± standard deviation (minimum-maximum); [#]p<0.05 (mean differences between the three biological groups were tested using an ANOVA test); *p<0.05 (means were compared by the independent-samples T-Test).

Quantitative real-time PCR (qPCR) was performed to obtain the relative expression values ($2^{-\Delta Ct}$) for the *RPPH1* (RNase P H1 RNA), the *MT-ND1* (NADH dehydrogenase 1) and the *MT-ND4* (NADH dehydrogenase 4) genes, following the manufacturer instructions (TaqMan® Gene Expression Assays Protocol by Applied Biosystems). Each multiplex reaction (*MT-ND1* and *RPPH1 or MT-ND4* and *MT-ND1*) was run in triplicate and raw data were collected in the Applied Biosystems 7900HT Fast Real-Time PCR system. MtDNA content was calculated as the cycle threshold (Ct) difference between the *MT-ND1* and the *RPPH1*; mtDNA common deletion frequency was calculated as the Ct difference between *MT-ND4* and the *MT-ND1*.

An ANCOVA test, followed by a Bonferroni correction, was performed to compare mtDNA content and common deletion frequency between the three biological groups. Correlations between mtDNA content, common deletion frequency, the number of CAG repeats in the expanded allele, age at onset and disease duration were tested using the partial correlation test. Age at blood collection was used as covariate. All statistical tests were performed in IBM SPSS Statistics 22; a p-value lower than 0.05 was considered as statistically significant.

RESULTS AND DISCUSSION

The comparison of mtDNA content between Azorean SCA3 patients and controls failed to reveal significant differences between these two groups; therefore, similarly to Zeng and colleagues [4], we were not able to replicate results of mtDNA depletion, previously described in two different SCA3 cohorts [1,5]. Several factors can contribute to differences in the outcome of the evaluation of mtDNA content in SCA3 studies. The small number of patients analysed in some studies, as well as the existence of confounding factors unaccounted for, such as age at sampling, are two factors of particular relevance. Because a decrease in mtDNA content, as well as an increase in common deletion frequency have been reported during the natural aging process [10], age is a factor that consistently needs to be taken in consideration in studies investigating mtDNA. In the present

work, although we were not able to select controls matched by age, its effect was adjusted in all the analyses. A third factor could be the inclusion of patients in different stages of disease progression, putatively reflecting different levels of cell dysfunction, blurring milder pathogenic effects.

Preataxic subjects displayed a slightly higher (although non-significant) mtDNA content compared to patients and controls (Figure 1, A). Higher mtDNA copy number has been reported in asymptomatic carriers of other polyglutamine (polyQ) disorders, including SCA3, compared to controls [5]. A longitudinal study in Huntington disease (HD), the most common polyQ disorder, described a biphasic pattern of changes in the mtDNA/nDNA copy number during disease progression [11]; mtDNA content increased before onset of motor symptoms and decreased after symptoms appearance [11]; this pattern is in accordance with the tendency observed in the present study. Noteworthy, because preataxic subjects were younger than SCA3 patients and controls (Table 1) in all comparisons of mtDNA content and the frequency of common deletion, the effect of age was accounted for in the ANCOVA model; the importance of this procedure has previously been emphasised.

The common deletion was significantly more frequent in patients (1.8 times) than in controls (Figure 1, B), a result which is in accordance with the report of Yu and collaborators [1]. The common deletion frequency was also significantly higher and more pronounced in preataxic subjects (4.3 times) than in controls (Figure 1, B). When comparing preataxic subjects and patients, however, differences in the frequency of the common deletion were not significant.

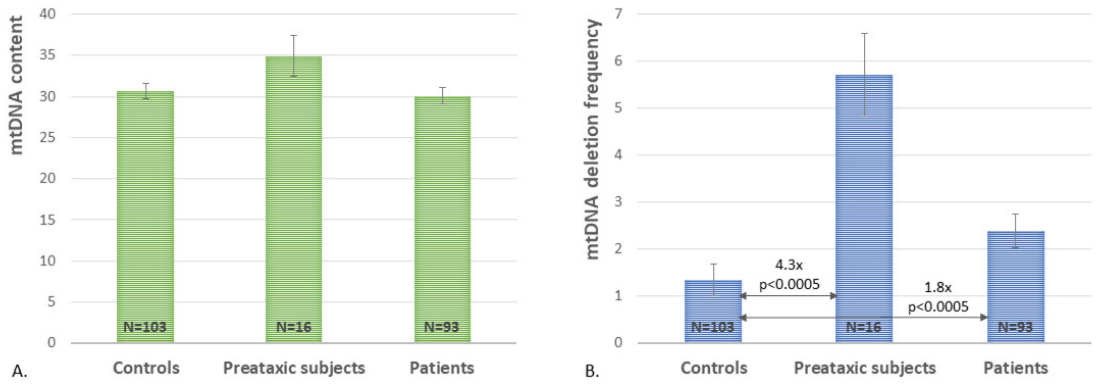


Figure 1. qPCR relative quantification of mtDNA content (A) and mtDNA deletion frequency (B) in blood from patients, preataxic subjects and controls. Mean differences between the three biological groups were tested using an ANCOVA, where age at blood collection was set as covariate. Relative expression levels of mtDNA content and mtDNA deletion frequency were log-transformed to improve normality assumptions, although non-transformed values are used in the graphics.

In SCA3 patients, no significant correlations between mtDNA content or common deletion with genetic (number of CAG repeats in expanded allele) or clinical variables (age at onset and disease duration), were detected (data not shown).

In preataxic subjects, a correlation between the frequency of mtDNA common deletion and CAG number in expanded allele was found (p=0.003, Figure 2); preataxic subjects carrying a higher number of CAG repeats in the expanded allele present a lower frequency of the mtDNA deletion.

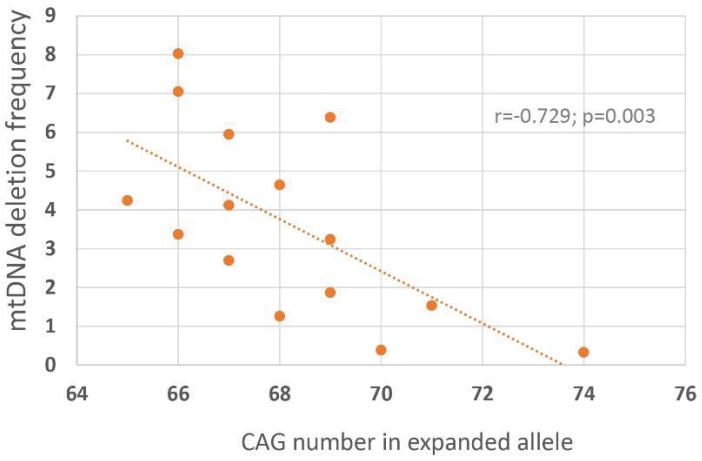


Figure 2. Partial correlation between mtDNA deletion frequency and CAG size in expanded allele in preataxic subjects (N=16). Age at blood collection was used as covariate.

Results concerning mtDNA content and frequency of the common deletion in the preclinical stage are promising; however, replication in larger and independent cohorts of preataxic subjects, as well as a more detailed investigation of the association between the frequency of the common deletion and clinical data (namely using scores from clinical rating scales) will also be mandatory. The availability of longitudinal data will also be of major importance to clarify the potential of mtDNA dysfunction-related biomarkers at the preclinical stage.

ACKNOWLEDGEMENTS

This work was funded by FEDER funds through the Operational Competitiveness Programme – COMPETE and by National Funds through FCT – Fundação para a Ciência e a Tecnologia under the project FCOMP-01-0124-FEDER-028753 (PTDC/DTP/PIC/0370/2012). A PhD fellowship M3.1.2/F/006/2011 (MR) and postdoctoral fellowships M3.1.7/F/031/2011 (AR) and M3.1.3/F/004/2009 (NK) were supported by Fundo Regional para a Ciência (FRC), Governo dos Açores. CB is supported by the Wellcome Trust (UK).

REFERENCES

1. Yu Y-C, Kuo C-L, Cheng W-L, Liu C-S, Hsieh M. Decreased antioxidant enzyme activity and increased mitochondrial DNA damage in cellular models of Machado-Joseph disease. J. Neurosci. Res. 2009;87:1884–91.

2. Ramos A, Kazachkova N, Silva F, Maciel P, Silva-Fernandes A, Duarte-Silva S, et al. Differential mtDNA damage patterns in a transgenic mouse model of Machado-Joseph disease (MJD/SCA3). J. Mol. Neurosci. 2015;55:449–53.

3. Kazachkova N, Raposo M, Montiel R, Cymbron T, Bettencourt C, Silva-Fernandes A, et al. Patterns of mitochondrial DNA damage in blood and brain tissues of a transgenic mouse model of Machado-Joseph disease. Neurodegener. Dis. 2013;11:206–14.

4. Zeng A, Liu X, Shen L, Li W, Ding Z, Bai Y, et al. Analysis of mitochondrial DNA variations in a Chinese family with spinocerebellar ataxia. J. Clin. Neurosci. 2012;19:60–4.

5. Liu C-S, Cheng W-L, Kuo S-J, Li J-Y, Soong B-W, Wei Y-H. Depletion of mitochondrial DNA in

leukocytes of patients with poly-Q diseases. J. Neurol. Sci. 2008;264:18–21.

6. Bettencourt C, Lima M. Machado-Joseph Disease: from first descriptions to new perspectives. Orphanet J. Rare Dis. 2011;6:35.

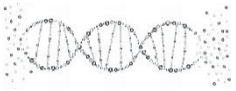
7. Maas RPPWM, van Gaalen J, Klockgether T, van de Warrenburg BPC. The preclinical stage of spinocerebellar ataxias. Neurology. 2015;85:96–103.

8. Evers MM, Toonen LJA, van Roon-Mom WMC. Ataxin-3 protein and RNA toxicity in spinocerebellar ataxia type 3: current insights and emerging therapeutic strategies. Mol. Neurobiol. 2014;49:1513–31.

9. Bettencourt C, Fialho RN, Santos C, Montiel R, Bruges-Armas J, MacIel P, et al. Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: A study in normal families from the Azores islands (Portugal). J. Hum. Genet. 2008;53:333–9.

10. Kazachkova N, Ramos A, Santos C, Lima M. Mitochondrial DNA damage patterns and aging: revising the evidences for humans and mice. Aging Dis. 2013;4:337–50.

11. Petersen MH, Budtz-Jørgensen E, Sørensen SA, Nielsen JE, Hjermind LE, Vinther-Jensen T, et al. Reduction in mitochondrial DNA copy number in peripheral leukocytes after onset of Huntington's disease. Mitochondrion. 2014;17:14–21.



CONCLUSIONS

The Azorean MJD cohort, given its particular features, should provide the adequate scenario to conduct research targeting this late-onset neurodegenerative disorder. In fact, the genetic and clinical background noise usually inherent to investigations based on multi-centric samples is reduced, therefore empowering studies. Exploratory studies to identify candidate biomarkers, both clinical and molecular, have been carried out in the present dissertation. These led to relevant positive findings, which establish a base for future studies. Notwithstanding, such findings should be further explored and validated in independent replication cohorts. The ongoing work should lead to the identification of molecular biomarkers, which will: (1) provide in depth knowledge about the natural history of MJD, by allowing better predictions of the age at onset and facilitating prognosis, (2) monitor different stages of disease progression, including the preclinical stage; and (3) empower future interventional trials. Noteworthy, studies aiming to identify trait and/or state biomarkers are favoured if the genetic and clinical features of the cohort under investigation are well known, as argued in Chapter I. Like other late-onset disorders, the preclinical stage in MJD remains poorly understood; the Azorean cohort, which includes preataxic carriers of the mutation, further potentiates research on this hardly accessible stage. The investigation of the presence of nystagmus in the preclinical stage is an example of this type of research (Chapter II). In this work, two genetic modifiers of MJD onset – the number of CAG repeats at the SCA1 locus (Chapter III) and the *IL6* c.-174G>C polymorphism (Chapter IV) have been proposed. These modifiers will now be added to those previously reported, globally constituting candidate trait biomarkers; their universal validity remains under investigation. In the quest for molecular state biomarkers, mRNA levels of *FCGR3B, P2RY13, SELPLG* measured in blood samples from MJD patients have been proposed as candidates, based on several experiments which demonstrated that their levels were dysregulated in patients, compared to controls (Chapter V). Although cross-sectional designs are widely used for biomarker discovery, the availability of samples from the same MJD subject collected at two or more different moments, therefore allowing longitudinal investigations, provides golden standard opportunities for state

biomarkers identification. The exploratory study investigating the pattern of expression of *HSPB1* and *BCL2* (Chapter VI) along the natural history of the disease, from the preataxic to the ataxic stage and then in progression, is a clear example of the potential of longitudinal studies. This type of approach could also constitute the only way to enlighten certain aspects which remain controversial in MJD, such as the presence and implications of mtDNA damage and depletion (Chapter VII).

Strengths, limitations and unsolved aspects of the PhD project

In the four years during which the work presented in this thesis was undertaken, two different aspects have been particularly discussed. Firstly, MJD is a rare disorder worldwide, bringing difficulties to the design of adequately powered studies. Secondly, to reach a sufficient number of samples so that power can be increased, the use of different cohorts is required. Nevertheless, some caution is necessary, since the analyses of larger but heterogeneous cohorts present its own set of limitations (see Chapter III). Small populations, such as the Azores Islands, which started from a limited group of founders present several of the advantages of population isolates: more uniform genetic background, genealogical records that can trace back in time (e.g., [1,2]), facilitated harmonization of phenotypic evaluation, as well as a more uniform environment [3]. The Azorean MJD cohort, being derived from the Azorean population, shares these advantages and, therefore, has a strong potential as a resource to identify new trait and state biomarkers. To ensure that results are not population-specific, findings from the Azorean MJD cohort should be further replicated in additional independent cohorts, which are currently difficult to access. Biomarkers showing a similar behaviour in the exploratory and replication cohorts will hopefully be robust enough to complement clinical measures, allow disease monitoring and improve the power of future clinical trials.

In this work, the identification of molecular biomarkers was grounded in the study of a non-affected tissue, namely blood; the use of an accessible tissue is in accordance with the definition

of biomarker. Furthermore, mutated ataxin-3 is widely expressed outside of the central nervous system (CNS; [4–8]) and it is expected that non-neuronal damage, even if representing only slight changes, could be quantified. Although the clinical and physiological phenotypes of neurodegenerative diseases reflect mainly neuronal damage, peripheral clinical signs in MJD, namely weight loss [9] as well as molecular alterations have been described (see Introduction and Chapters V, VI and VII for more details). The link between peripheral alterations and neurodegeneration has been also described in Huntington disease (reviewed in [10]) suggesting, once more, shared mechanisms between polyQ disorders.

The relationship between the disease and molecular biomarkers, namely gene expression data measured in a peripheral tissue has been described in the literature using mainly three theoretical models: (1) causal model – alterations of mRNA levels can be caused directly by the presence of mutated ataxin-3; (2) reactive model – gene expression changes may be a downstream effect of the causal mutation; and (3) sentinel model - transcriptional information can be shared, mediated by mononuclear white blood cells, between blood and brain compartments, as a response to brain damage [11,12]. Biomarkers identified through model 1 are ideal markers for measuring a therapeutic response; models 2 and 3 can provide markers of progression. Ideally, some of these biomarkers could also act as surrogate markers in trials testing disease modifying-compounds [11].

Future perspectives

The availability of a molecular test to confirm the presence of the *ATXN3* mutation in at-risk MJD individuals allowed to clearly define a new stage of the disease – the preclinical stage, which starts many years before the manifestation of ataxia (currently considered the hallmark of the clinical diagnosis). Abnormalities in the preclinical stage have been reported from structural and functional brain imaging studies, neurophysiologic studies as well as oculomotor system studies (reviewed in [13]); some of these abnormalities are detectable many years before the ataxia onset

[14]. As previously referred, the determination of the pathological onset is currently challenging as the brain can only be accessed post-mortem. The identification of MJD carriers before a considerable number of neurons has been injured will be of major importance to promote preventive trials and ideally prevent the disease from ever manifesting. To better understand the preclinical stage and early stages of the disease, a European project, ESMI (European Spinocerebellar Ataxia Type 3/Machado-Joseph Disease Initiative - JPCOFUND/0002/2015), which brings together several MJD cohorts, namely the Azorean cohort, is ongoing. Longitudinal data provided by this large and well-characterized multi-centre trial-ready cohort is expected to provide a major contribution to the investigation of several aspects of MJD, namely in the field of molecular biomarkers.

Findings from the present dissertation open new opportunities in the molecular biomarkers field, a pioneer field in MJD. More steps of validation will be crucial to developed a battery of molecular biomarkers able to better describe disease progression, as well as to be used as outcome measures in either preventive or disease-modifying future clinical trials. Data from molecular biomarkers should also be further correlated with imaging and clinical measures. Efforts to test the promising molecular biomarkers in independent cohorts, as well as to investigate their potential in longitudinal studies will be made.

REFERENCES

1. Lima M, Mayer F, Coutinho P, Abade A. Prevalence, geographic distribution, and genealogical investigation of Machado-Joseph disease in the Azores (Portugal). Hum. Biol. 1997;69:383–91.

2. Lima M, Mayer FM, Coutinho P, Abade A. Origins of a mutation: population genetics of Machado-Joseph disease in the Azores (Portugal). Hum. Biol. 1998;70:1011–23.

3. Peltonen L. Positional cloning of disease genes: advantages of genetic isolates. Hum. Hered. 50:66–75.

4. Schmidt T, Landwehrmeyer GB, Schmitt I, Trottier Y, Auburger G, Laccone F, et al. An isoform of ataxin-3 accumulates in the nucleus of neuronal cells in affected brain regions of SCA3 patients. Brain Pathol. 1998;8:669–79.

5. Costa M do C, Gomes-da-Silva J, Miranda CJ, Sequeiros J, Santos MM, Maciel P. Genomic structure, promoter activity, and developmental expression of the mouse homologue of the Machado-Joseph disease (MJD) gene. Genomics. 2004;84:361–73.

6. Ichikawa Y, Goto J, Hattori M, Toyoda A, Ishii K, Jeong SY, et al. The genomic structure and expression of MJD, the Machado-Joseph disease gene. J. Hum. Genet. 2001;46:413–22.

7. Paulson HL, Das SS, Crino PB, Perez MK, Patel SC, Gotsdiner D, et al. Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. Ann. Neurol. 1997;41:453–62.

8. Trottier Y, Cancel G, An-Gourfinkel I, Lutz Y, Weber C, Brice A, et al. Heterogeneous intracellular localization and expression of ataxin-3. Neurobiol. Dis. 1998;5:335–47.

9. Saute JAM, Silva ACF da, Souza GN, Russo AD, Donis KC, Vedolin L, et al. Body mass index is inversely correlated with the expanded CAG repeat length in SCA3/MJD patients. Cerebellum. 2012;11:771–4.

10. Carroll JB, Bates GP, Steffan J, Saft C, Tabrizi SJ. Treating the whole body in Huntington's disease. Lancet Neurol. 2015;14:1135–42.

11. Scherzer CR. Chipping away at diagnostics for neurodegenerative diseases. Neurobiol. Dis. 2009;35:148–56.

12. Schadt EE, Lamb J, Yang X, Zhu J, Edwards S, Guhathakurta D, et al. An integrative genomics approach to infer causal associations between gene expression and disease. Nat. Genet. 2005;37:710–7.

13. Maas RPPWM, van Gaalen J, Klockgether T, van de Warrenburg BPC. The preclinical stage of spinocerebellar ataxias. Neurology. 2015;85:96–103.

14. Jacobi H, Reetz K, du Montcel ST, Bauer P, Mariotti C, Nanetti L, et al. Biological and clinical characteristics of individuals at risk for spinocerebellar ataxia types 1, 2, 3, and 6 in the longitudinal RISCA study: analysis of baseline data. Lancet Neurol. 2013;12:650–8.