UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE BIOCIÊNCIAS
DEPARTAMENTO DE GENÉTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

AVALIAÇÃO DE MODIFICADORES DO FENÓTIPO NA DOENÇA DE MACHADO-JOSEPH

Eduardo Preusser de Mattos

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular)

Orientadora: Profa. Dra. Laura Bannach Jardim
Co-orientadora: Profa. Dra. Maria Luiza Saraiva-Pereira

Porto Alegre, maio de 2018

O presente trabalho foi desenvolvido nas seguintes Instituições:

- Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre.
- Laboratório de Identificação Genética, Centro de Pesquisa Experimental, Hospital de Clínicas de Porto Alegre.
- Departamento de Biologia Celular, Centro Médico da Universidade de Groningen (UMCG/RuG), Holanda.

Agentes financiadores:

- Conselho Nacional de Desenvolvimento Científico e Tecnológico.
- Comissão de Aperfeiçoamento de Pessoal de Nível Superior.
- Programa Ciência Sem Fronteiras, modalidade doutorado sanduíche.
- Fundo de Incentivo a Pesquisas e Eventos, Hospital de Clínicas de Porto Alegre.

Variability is the law of life, and as no two faces are the same, so no two bodies are alike, and no two individuals react alike and behave alike under the abnormal conditions which we know as disease.

Sir William Osler (1849 – 1919)

Remember to look up at the stars and not down at your feet.
Try to make sense of what you see and wonder about what makes the universe exist.
Be curious. And however difficult life may seem,
there is always something you can do and succeed at.
It matters that you don't just give up.

Stephen Hawking (1942 – 2018)

Agradecimentos

A ciência é, por definição, uma atividade colaborativa que transcende barreiras geográficas, econômicas e linguísticas. Salvo raríssimas exceções, talvez mais bem ilustradas pela figura do exímio polímata Leonardo da Vinci, nenhum ser humano é capaz de dominar diversas áreas do conhecimento e, sozinho, promover avanços científicos e tecnológicos significativos. Como bem disse o físico americano John Bardeen na ocasião do recebimento do seu segundo prêmio Nobel em 1972:

> *"Science is a field which grows continuously with ever expanding frontiers. Further, it is truly international in scope. Any particular advance has been preceded by the contributions of those from many lands who have set firm foundations for further developments. […] The combined results of several people working together is often much more effective than could be that of an individual scientist working alone."*

Dessa maneira, muitas são as pessoas que, direta ou indiretamente, perto ou longe, tornaram a presente tese possível. A todos vocês o meu mais profundo agradecimento e satisfação por terem tornado esse trabalho uma realidade.

À Profª. Dra. Laura Bannach Jardim, orientadora da presente tese, meu eterno apreço e gratidão pela oportunidade de ter trabalhado e aprendido tanto, não somente com uma grande pesquisadora, mas também com uma pessoa extraordinária. Obrigado por todos os ensinamentos, paciência e inspiração ao longo dessa trajetória. Você é a prova máxima de que a ciência ética e de qualidade é uma realidade no Brasil.

À Profª. Dra. Maria Luiza Saraiva-Pereira, um agradecimento especial pela acolhida no Laboratório de Identificação Genética e por todos os aprendizados ao longo do doutorado.

À Profª. Dra. Vanessa Bielefeldt Leotti, super-estatística do Grupo da Neurogenética, agradeço imensamente toda a paciência e dedicação no desenvolvimento dos modelos estatísticos utilizados na presente tese. Muito

obrigado por ter me ensinado tudo sobre a utilização do software R e a interpretação dos resultados.

To Prof. Dr. Harm H. Kampinga, my eternal gratitude and appreciation for the unique opportunity of working at the Department of Cell Biology at the University Medical Center Groningen. I am so glad I decided to take your course in Brazil, when I was still a Master's student. You are an inspiration and a model of how science and scientific education should always be: of high quality, precise, and thorough, but also fun and humane.

To Prof. Dr. Steven Bergink, I am deeply indebted for the constant fruitful discussions and daily supervision in Groningen. You have always pushed me to be more critical and creative in the lab, and that has played a central role on my academic track and professional career. Thank you for all the feedback on my writing and for always being there when it was necessary.

Aos membros da banca avaliadora, Profª. Dra. Patrícia Espinheira Sá Maciel (Escola de Medicina, Universidade do Minho, Portugal), Profª. Dra. Ursula da Silveira Matte (PPGBM/UFRGS) e Prof. Dr. Artur Francisco Schumacher Schuh (Departamento de Farmacologia/UFRGS), meu agradecimento especial pela atenção dedicada à leitura e revisão técnica da presente tese.

Aos colegas e coautores dos manuscritos que compõem a presente tese, Gabriela Bolzan, Profª. Dra. Suzi Alves Camey, Helena Fussiger, Nathalia Kersting, Els F. E. Kuiper, Maiara Kolbe Musskopf, Camila Maria Oliveira, Anastácia Guimarães Rocha, Prof. Dr. Jonas Alex Morales Saute e Gabriele Nunes Souza, minha gratidão por todo o trabalho, esforço e atenção no desenvolvimento das atividades aqui descritas.

Aos grandes amigos e colegas de laboratório, Gabriel Vasata Furtado e Márcia Polese-Bonatto, muito obrigado por tudo. Muitas foram as vezes em que sentamos juntos para discutir experimentos, falar besteira ou simplesmente reclamar do doutorado e aconselhar outros a fugir do laboratório. Acho que, de

certo modo, fomos ajudando uns aos outros nos últimos quatro anos e, finalmente, chegamos lá.

Aos colaboradores portugueses do manuscrito 2 da presente tese, Profª. Dra. Maria M. M. Lima, Dr. João Vasconcelos e Dra. Mafalda Raposo, obrigado pela disponibilidade em participar do estudo e por todas as sugestões na redação final do artigo. A contribuição de vocês certamente enriqueceu imensamente o nosso trabalho.

A todos os professores, da pré-escola ao doutorado, que contribuíram para a minha formação, agradeço imensamente pela dedicação e empenho, muitas vezes mal remunerados, investidos em mim. Agradeço em especial a todos os professores da Universidade Federal do Rio Grande do Sul que por mais de uma década foram essenciais na minha educação. Estive sempre sobre os ombros de gigantes.

Finalmente, nunca haverá palavras suficientes para agradecer a toda a minha família por tudo, e por quem tudo vale a pena. Ao meu pai, Marcelo, minha mãe, Christine, meu irmão, Fernando, e minha esposa, Aline: vocês dividem comigo a primeira autoria deste trabalho. Vocês tornaram isso tudo possível, e eu amo vocês incondicionalmente. Obrigado, por tudo.

Sumário

Lista de abreviaturas

ADCA: ataxia cerebelar de herança autossômica dominante (*autosomal dominantly inherited cerebellar ataxia*)

AO: idade de início (*age at onset*)

ApoE: apolipoproteína E

ATP: adenosina trifosfato ou trifosfato de adenosina

CAGexp: alelo do gene *ATXN3* com um trato CAG expandido e patogênico

CCFS: escore funcional cerebelar composto (*composite cerebellar functional score)*

cDNA: DNA complementar

DMJ/SCA3: doença de Machado-Joseph/ ataxia espinocerebelar tipo 3

DUB: deubiquitilase

GBA: beta-glicosidase ou glucosilceramidase

HSP: proteína de choque térmico (*heat shock protein*)

ICARS: escala internacional cooperativa de escore de ataxia (*international cooperative ataxia rating scale*)

IMC: índice de massa corpórea

kB: quilobase

mRNA: RNA mensageiro

NESSCA: Escore do exame neurológico para avaliação da ataxia espinocerebelar (*neurological examination score for the assessment of spinocerebellar ataxia*)

Poli-Ub: poli-ubiquitina

PoliQ: poliglutamina

RE: retículo endoplasmático

SARA: escala para a avaliação e o escore da ataxia (*scale for the assessment and rating of ataxia*)

SCA: ataxia espinocerebelar (*spinocerebellar ataxia*)

SCAFI: índex funcional de ataxia espinocerebelar (*spinocerebellar ataxia functional index*)

sHSP: pequena proteína de choque térmico (*small heat shock protein*)

SNP: polimorfismo de nucleotídeo único (*single nucleotide polymorphism*)

Ub: ubiquitina

UIM: motivo de interação com ubiquitina (*ubiquitin-interacting motif*)

UPR: resposta a proteínas não enoveladas (*unfolded protein response*)

UPS: sistema ubiquitina-proteassomo (*ubiquitin-proteasome system*)

Resumo

A ataxia espinocerebelar tipo 3/ doença de Machado-Joseph (DMJ/SCA3) é uma condição neurodegenerativa devida à expansão de um trato CAG (CAGexp) no gene *ATXN3*, resultando em uma proteína anormal com uma sequência de poliglutaminas (poliQ) expandida. O CAGexp determina apenas parcialmente a idade de início dos sintomas (AO, *age at onset*) e evidências crescentes apontam para a modulação da AO por fatores adicionais. Diferentes estudos também sugerem que a velocidade de progressão da doença (VPD) é modificada pelo CAGexp e/ou fatores adicionais. O objetivo do presente trabalho foi identificar novos modificadores genéticos da DMJ/SCA3, com foco em dois fenótipos: AO e VPD. Inicialmente, uma revisão sistemática avaliou o atual "estado da arte" sobre modificadores da AO. Uma metanálise avaliou 140 estudos originais selecionados, incluindo dez coortes de pacientes com dados individuais (n=2.099) e duas coortes com dados agregados. Os dados gerados por essa metanálise direcionaram os estudos sobre os modificadores genéticos da AO. Em seguida, potenciais moduladores descritos anteriormente foram validados e um importante efeito população-específico do CAGexp sobre a AO foi descoberto, sugerindo a existência de modificadores da AO distintos em diferentes populações com DMJ/SCA3. Essa hipótese foi testada e validada e, como resultado, desenvolvemos fórmulas população-específicas para a predição da AO em indivíduos assintomáticos com DMJ/SCA3 de diferentes origens geográficas. A revisão sistemática permitiu-nos priorizar a análise de alguns candidatos a modificadores fenotípicos da DMJ/SCA3. O potencial efeito do gene para a apolipoproteína E (*apoE*), que é o maior fator de risco genético conhecido para o desenvolvimento da doença de Alzheimer, sobre a AO foi testado. Foi demonstrado que pacientes com DMJ/SCA3 homozigotos para *apoE* ε4/ε4 apresentaram AO significativamente mais precoce do que aqueles com outros genótipos de *apoE*, sugerindo que o controle dos níveis de apoE pode contribuir para o atraso do início da DMJ/SCA3. Além disso, o papel das chaperonas moleculares como modificadores fenotípicos da DMJ/SCA3 foi também investigado. Diversas chaperonas inibem a agregação e toxicidade de proteínas poliQ, mas a co-chaperona DNAJB6 tem sido demonstrada como um dos melhores agentes anti-amiloidogênicos. Esse tópico foi revisado como parte desse trabalho

e testamos a hipótese de que diversas chaperonas, e especialmente a DNAJB6, poderiam atuar como modificadores da AO e/ou VPD na DMJ/SCA3. Utilizando uma estratégia de amostragem de fenótipos extremos com foco na AO, níveis proteicos mais elevados de diferentes chaperonas, incluindo a DNAJB6, foram detectados em pacientes holandeses com DMJ/SCA3 de início significativamente mais tardia do que o esperado, em comparação a pacientes da mesma coorte com AO na média ou mais precoce do que o esperado. Além disso, níveis proteicos mais elevados da DNAJB6 foram fortemente associados a VPD mais lenta em uma coorte de pacientes brasileiros com DMJ/SCA3. No conjunto, esses resultados corroboram a noção de que chaperonas - especialmente a DNAJB6 - são potenciais alvos terapêuticos para se retardar o início e/ou atrasar a neurodegeneração na DMJ/SCA3, assim como em outras poliglutaminopatias. Portanto, foi demonstrado que fatores familiais e populacionais, o comprimento do trato CAG no gene *ATXN2*, o gene *apoE* e algumas chaperonas moleculares (especialmente a DNAJB6) são moduladores fenotípicos da DMJ/SCA3.

Abstract

Spinocerebellar ataxia type 3/Machado-Joseph disease (SCA3/MJD) is a neurodegenerative condition due to an expanded CAG tract (CAGexp) at *ATXN3*, resulting in an abnormally long polyglutamine (polyQ)-containing protein. CAGexp determines only partially age at onset (AO), and growing evidence supports a role for AO modulation in SCA3/MJD by additional factors. Different studies have also suggested that the velocity of disease progression (VDP) in SCA3/MJD is modulated by CAGexp and/or additional factors. The aim of the present work was to identify new genetic modifiers of SCA3/MJD, focusing on two phenotypes: AO and VDP. Initially, a systematic review assessed the current "state of the art" on genetic modifiers of AO. A meta-analysis was based on 140 selected original studies, including 10 individual participant cohorts (n=2,099) and 2 aggregated data cohorts. Data generated by this meta-analysis directed further analyses on genetic modifiers of AO. Then, candidate modifiers published previously were validated and an important population-specific effect of CAGexp on AO was uncovered, suggesting the occurrence of distinct AO modifiers in different SCA3/MJD populations. This hypothesis was tested and validated, and as an outcome we developed population-specific formulas for AO prediction in asymptomatic SCA3/MJD subjects from different geographical origins. The systematic review also indicated candidate modifiers of SCA3/MJD phenotypes to be prioritized. The potential effect of the *apolipoprotein E* gene (*apoE*), which is the strongest known genetic risk factor for developing Alzheimer's disease, on AO was tested. We then described that SCA3/MJD patients homozygous for *apoE* ε4/ε4 presented with significantly earlier AO than those with other *apoE* genotypes, suggesting that control of ApoE levels could prove beneficial for delaying disease initiation in SCA3/MJD. In addition, the role of molecular chaperones as phenotypic modifiers of SCA3/MJD was also investigated. Several chaperones suppress aggregation and toxicity of polyQ proteins, but DNAJB6 has been demonstrated as a superior anti-amyloidogenic agent. This topic was reviewed as part of this work and we tested the hypothesis that several chaperones, and especially DNAJB6, could act as modifiers of AO and VDP in SCA3/MJD. Using an extreme phenotype sampling approach focused on AO, increased protein levels of different chaperones, including DNAJB6,

were detected in Dutch SCA3/MJD patients with AO significantly later than expected when compared to individuals with early or average AO from the same cohort. Moreover, higher DNAJB6 protein levels were strongly associated to slower VDP in a cohort of Brazilian SCA3/MJD patients. Together, these data argue in favor of the therapeutic potential of chaperones – especially DNAJB6 – in delaying onset and/or slowing neurodegeneration in SCA3/MJD, as well as in other polyQ diseases. Therefore, we were able to show that familial and populational factors, CAG repeat length at *ATXN2*, *apoE* and some molecular chaperones (especially DNAJB6) are phenotypic modulators of SCA3/MJD.

Capítulo 1. Introdução e Objetivos

1.1 Aspectos clínicos da DMJ/SCA3

1.1.1 Descrição original e epidemiologia

A condição neurodegenerativa conhecida como doença de Machado-Joseph, ou ataxia espinocerebelar tipo 3 (DMJ/SCA3), foi inicialmente descrita na década de 1970 em duas grandes famílias dos Estados Unidos da América com ancestrais oriundos das ilhas Açorianas de São Miguel (família Machado) e Flores (família Joseph) (Nakano et al. 1972; Rosenberg et al. 1976). Observou-se à época que todos os indivíduos afetados de ambas as famílias desenvolviam a mesma manifestação neurológica de início na vida adulta: a ataxia de marcha. Esse distúrbio representa a falta de equilíbrio e dificuldade de coordenação motora decorrentes da disfunção do cerebelo e de suas vias (Rüb et al. 2013). Essas observações posteriormente levaram à classificação da DMJ/SCA3 como uma forma de ataxia cerebelar de herança autossômica dominante (ADCA) (Schöls et al. 1997). Atualmente, mais de 40 afecções genéticas distintas pertencem a esse grupo (Bird 2016). Apesar dessa grande diversidade, a maior parte das ADCAs é extremamente rara, e estima-se que apenas cinco doenças pertencentes a esse grupo (SCA1, SCA2, DMJ/SCA3, SCA6 e SCA7) sejam responsáveis por cerca de 50 a 60% de todos os casos de ADCA no mundo (Rüb et al. 2013). Dentre essas, a DMJ/SCA3 é a mais prevalente. Estudos de diferentes grupos de pesquisa demonstraram que a DMJ/SCA3 é a forma mais comum de SCA no Estado do Rio Grande do Sul (Jardim et al. 2001a; Trott et al. 2006; Souza et al. 2016), no Brasil (de Castilhos et al. 2014) e no mundo (Schöls et al. 2004; Sequeiros et al. 2012).

1.1.2 Achados clínicos

Não há sintomas específicos que auxiliem no diagnóstico clínico definitivo de determinada SCA dominante como a DMJ/SCA3. Uma exceção, talvez, é a SCA7, única SCA associada à degeneração retiniana (Rüb et al. 2013). De fato, a caracterização puramente clínica, baseada em exames neurológicos e de imagem,

dificilmente leva a um diagnóstico SCA-específico (Riess et al. 2008). Ainda que a ataxia de marcha seja o principal achado clínico na maioria dos pacientes com DMJ/SCA3, diversos outros sintomas frequentemente estão presentes, podendo afetar a coordenação e a velocidade dos movimentos, a deglutição e a fala, entre outros (Jardim et al. 2001b; Schöls et al. 2004; Riess et al. 2008; Jardim et al. 2010; Saute and Jardim 2015).

A DMJ/SCA3 foi historicamente dividida em três subtipos clínicos, de acordo com os principais sintomas neurológicos presentes e a idade de início (Coutinho et al. 1986; Sudarsky and Coutinho 1995). Posteriormente, outros subtipos foram propostos na literatura, com significado ainda não bem estabelecido (Riess et al. 2008; Saute and Jardim 2015). Os subtipos mais aceitos são os seguintes:

- tipo I: início precoce (entre 10 e 30 anos) e sintomas piramidais e extrapiramidais (principalmente distonia);
- tipo II: início intermediário (entre 20 e 50 anos) e sintomas piramidais e cerebelares;
- tipo III: início tardio (entre 40 e 75 anos) e sintomas cerebelares e neuropatia periférica;
- tipo IV: início em idade variável e sintomas parkinsonianos e neuropatia.

A classificação de pacientes nesses subtipos clínicos vem sendo menos utilizada desde a identificação da mutação causal da DMJ/SCA3 (Kawaguchi et al. 1994). Depois de terem sido associados ao tamanho da sequência repetitiva expandida e de sua ocorrência acontecer simultaneamente dentro das mesmas famílias, ficou claro que os subtipos clínicos da DMJ/SCA3 refletiam um espectro fenotípico de uma mesma doença, e não entidades clínicas distintas.

A variabilidade de sintomas na DMJ/SCA3 é objeto de muitos estudos que visam ao estabelecimento de correlações genótipo-fenótipo entre determinadas variações genéticas e manifestações clínicas específicas. Por exemplo, Siebert e colaboradores (2012) identificaram uma maior frequência de mutações em heterozigose no gene *GBA*, responsável pela doença de Gaucher, em pacientes

com DMJ/SCA3 com sintomas parkinsonianos (tipo IV), em relação a pacientes com os tipos I, II e III (Siebert et al. 2012).

Evidências também apontam para uma perturbação metabólica em indivíduos com DMJ/SCA3, que comumente possuem menor índice de massa corpórea (IMC) do que sujeitos saudáveis (Saute et al. 2012b). O IMC parece estar inversamente relacionado ao tamanho da expansão CAG em *ATXN3*, e ocorre independentemente da presença de disfagia (dificuldade de deglutição que afeta a ingesta de alimentos em estágios mais avançados da doença). A diminuição do IMC em pacientes com DMJ/SCA3 pode ser devida a alterações na rota bioquímica de sinalização da insulina, uma vez que indivíduos sintomáticos apresentam maior sensibilidade e menor resistência à insulina (Saute et al. 2011).

1.1.3 Avaliação da idade de início dos sintomas

Um dos fenótipos mais estudados na DMJ/SCA3 é, certamente, a idade de início dos sintomas (AO, do inglês *age at onset*). Historicamente, a maioria dos estudos na DMJ/SCA3 define a AO como a idade em que o paciente (ou um familiar próximo) notou os primeiros sintomas, sejam eles quais forem (Kawaguchi et al. 1994; Ranum et al. 1995; Jardim et al. 2001a; Infante et al. 2005; Saute and Jardim 2015). Apesar de o primeiro sintoma da DMJ/SCA3 ser geralmente a ataxia de marcha, diversas outras manifestações neurológicas podem já estar presentes antes das dificuldades de deambulação e equilíbrio (Globas et al. 2008; Luo et al. 2017). Por exemplo, uma parcela significativa dos pacientes apresenta alterações oculares, como visão dupla e nistagmo, anos antes de qualquer comprometimento de marcha (Globas et al. 2008). Por conseguinte, essa heterogeneidade do primeiro sintoma na DMJ/SCA3 pode ser um fator confundidor importante na busca de modificadores da AO, já que fatores distintos podem estar envolvidos na modulação de sintomas diferentes, como a ataxia de marcha e a visão dupla. Além disso, vieses importantes podem afetar a determinação acurada da AO na DMJ/SCA3 e em outras doenças do grupo das poliglutaminopatias. Por exemplo, pacientes com longa duração de doença são mais propensos a sub- ou superestimar sua AO

(Globas et al. 2008). Similarmente, há uma tendência de diagnóstico mais precoce nas gerações seguintes, comparado ao caso índice da família, uma vez que já se conhece a doença e seus sintomas.

Faz-se necessário notar também que a determinação da AO na DMJ/SCA3 como um evento específico no tempo representa uma simplificação, cientificamente conveniente, de um processo lento e contínuo ao longo da vida do indivíduo afetado. Por se tratar de uma doença geneticamente herdada (e, portanto, com exposição presente antes mesmo do nascimento), a AO pode ser melhor compreendida como um ponto arbitrário no tempo ao longo da história natural da DMJ/SCA3 em que os sintomas se tornam impactantes na vida do paciente. De fato, estudos na DMJ/SCA3 já identificaram diversas alterações funcionais em indivíduos classicamente reconhecidos como “assintomáticos”. Dentre essas manifestações pré-sintomáticas (em relação à ataxia de marcha, principalmente) encontram-se mudanças no perfil de citocinas (da Silva Carvalho et al. 2016) e alterações no metabolismo da glicose no sistema nervoso central (Soong and Liu 1998), no volume da substância branca (Wu et al. 2017b) e nos movimentos oculares (Wu et al. 2017a). Apesar de todas as complicações de definição e aferição da AO, estudos recentes têm proposto a uniformização do conceito de AO na DMJ/SCA3, focando na ataxia de marcha como o evento inicial do processo patológico (Globas et al. 2008; Tezenas du Montcel et al. 2014b; Tezenas du Montcel et al. 2014a; Raposo et al. 2015).

O principal fator determinante da AO na DMJ/SCA3 é o tamanho da sequência repetitiva CAG no gene *ATXN3* (CAGexp), que é a mutação causal da doença (ver seção 1.2 “Aspectos moleculares da DMJ/SCA3”). O CAGexp apresenta uma relação inversa com a AO: quanto maior a expansão, mais precoce a AO (Riess et al. 2008; Matilla-Dueñas et al. 2014). Entretanto, diversos estudos em coortes independentes ao longo dos últimos 25 anos estabeleceram claramente que a CAGexp não é o único determinante da AO. De fato, ao se comparar a distribuição da AO em função de CAGexp, é possível observar que a mutação causal determina, geralmente, de 40 a 60% da variabilidade da AO na DMJ/SCA3 (van de Warrenburg et al. 2005; de Castilhos et al. 2014; Tezenas du Montcel et al.

2014a; Raposo et al. 2015; Chen et al. 2016). Ainda que uma pequena parcela da variabilidade da AO seja devida às dificuldades de aferição discutidas acima, essa correlação parcial entre CAGexp e AO sugere que aproximadamente 50% da variação da AO é determinada por outros fatores genéticos e/ou ambientais. Por exemplo, para um CAGexp médio de 75 repetições CAG, a maioria dos indivíduos da população DMJ/SCA3 do estado do Rio Grande do Sul apresenta sintomas aos 34 anos de idade (Souza et al. 2016); contudo, há indivíduos com esse tamanho de CAGexp e que iniciam os sintomas aos 20, ou aos 45 anos (Saute and Jardim 2015). Indivíduos que apresentam AOs muito antes ou depois do esperado, de acordo com o comprimento de CAGexp, podem, desse modo, ser considerados *outliers* fenotípicos para AO.

Muitos estudos buscam desvendar novos moduladores fenotípicos da DMJ/SCA3, tendo como principal objetivo retardar o início dos sintomas ou diminuir a velocidade de progressão da doença. Contudo, a maioria das investigações até o presente momento detectou candidatos de pequeno efeito, que geralmente melhoram a explicação da variabilidade da AO em não muito mais do que 1 a 2% (de Castilhos et al. 2014; Tezenas du Montcel et al. 2014a; Raposo et al. 2015; Chen et al. 2016). Um dos poucos candidatos com efeito replicado em coortes distintas é o predito para o gene da apolipoproteína E (*apoE*) (Bettencourt et al. 2011; Zhou et al. 2014; Peng et al. 2014). O gene *apoE* está presente em três variantes principais em seres humanos, denominadas ε2, ε3 e ε4. A presença do alelo ε4 de *apoE* é o principal fator de risco genético conhecido para o desenvolvimento da doença de Alzheimer nas suas formas não monogênicas (Huynh et al. 2017a; Lautner et al. 2017), e evidências sugerem um cenário similar para a doença de Parkinson (Tsuang et al. 2013). Na DMJ/SCA3, dois estudos identificaram AO mais precoce – corrigindo-se para diferenças em CAGexp – entre portadores de pelo menos um alelo *apoE* ε2 (Bettencourt et al. 2011; Peng et al. 2014). Contudo, uma terceira investigação em uma coorte distinta não replicou esse achado e não observou nenhuma diferença significativa na AO de pacientes com ou sem o alelo ε2 (Zhou et al. 2014).

Alguns estudos, incluindo a presente tese, utilizaram o recrutamento de indivíduos *outliers* (ou seja, com fenótipos extremos) para AO, com o intuito de desvendar modificadores genéticos em sujeitos potencialmente portadores das suas variantes de maior efeito e que por isso teriam se desviado da norma (Zijlstra et al. 2010). A assim chamada "*extreme phenotype sampling*" propõe que sujeitos que se encontram nos extremos da uma distribuição fenotípica devam possuir uma ou mais variantes genéticas de grande impacto que são desfavoráveis, em um extremo, ou vantajosas, em outro (Li et al. 2011; Barnett et al. 2013). Esse tema é abordado com mais profundidade nos manuscritos que compõem a presente tese.

1.1.4 Avaliação da velocidade de progressão e gravidade da doença

Na DMJ/SCA3, assim como em outras condições neurodegenerativas, a gravidade e a velocidade progressão da doença são mensuradas por escalas clínicas de avaliação neurológica. Esses instrumentos são geralmente compostos por medidas padronizadas e objetivas que visam à determinação do grau de disfunção de um indivíduo e o quanto essa disfunção varia ao longo do tempo (Saute et al. 2012a). Escalas como a *International Cerebellar Ataxia Rating Score* (ICARS), *Spinocerebellar Ataxia Functional Index* (SCAFI) e muitas outras têm sido desenvolvidas e validadas para a DMJ/SCA3 (Saute et al. 2012a). Duas das escalas mais utilizadas na DMJ/SCA3 são a *Scale for Assessment and Rating of Ataxia* (SARA) (Schmitz-Hübsch et al. 2006) e a *Neurologic Examination Score for Spinocerebellar Ataxias* (NESSCA) (Kieling et al. 2008). A escala SARA avalia a performance de um indivíduo em quatro tarefas que medem a coordenação e gera um escore que varia de 0 (sem ataxia) a 40 (comprometimento extremo). A escala NESSCA quantifica alterações do exame neurológico como um todo, incluindo manifestações piramidais, extrapiramidais, de nervos cranianos e de sensibilidade, além de levar em conta o relato do indivíduo sobre a função dos esfíncteres e a presença de câimbras, vertigem e/ou disfagia. A NESSCA também varia entre 0 (exame neurológico completamente normal) a 40 (exame neurológico completamente anormal).

As escalas clínicas são extremamente úteis no estudo da história natural de condições neurodegenerativas como a DMJ/SCA3, possibilitando a determinação da velocidade de progressão da doença. Uma estimativa acurada desse parâmetro é essencial, por exemplo, para o desenvolvimento de ensaios clínicos que busquem parar ou retardar a progressão da doença. Entretanto, se a progressão de uma condição rara for muito lenta, a determinação da velocidade de progressão natural (sem intervenção medicamentosa) é uma tarefa bastante difícil, uma vez que exige o acompanhamento de muitos pacientes durante um longo intervalo de tempo. Por exemplo, Jardim e colaboradores (2010) avaliaram uma coorte de 105 pacientes da região Sul do Brasil durante 10 anos, a fim de estudar a história natural e a progressão da DMJ/SCA3 utilizando a escala NESSCA (Jardim et al. 2010). Foi possível determinar uma piora média de 1,26 pontos na escala NESSCA por ano de doença. A taxa de progressão da NESSCA foi dependente da idade de início do primeiro sintoma: quanto mais tardio o início da doença, mais devagar a sua progressão (Jardim et al. 2010). Uma taxa de progressão muito similar (1,61 pontos por ano) foi detectada em 139 pacientes europeus com DMJ/SCA3 que foram acompanhados ao longo de 2 anos com a escala SARA (Jacobi et al. 2011).

Ao contrário da AO, em que há uma clara associação com o tamanho da mutação causal, a influência do tamanho da expansão CAG no *ATXN3* sobre a progressão da DMJ/SCA3 ainda é controversa. Há evidências tanto a favor (Jardim et al. 2010; Donis et al. 2016) quanto contra essa associação (Jacobi et al. 2011; Tezenas du Montcel et al. 2012) em coortes com origens geográficas distintas. Certamente, estudos longitudinais mais longos e com maior número de pacientes serão necessários para esclarecer esse e outros aspectos a respeito da progressão da DMJ/SCA3.

1.2 Aspectos moleculares da DMJ/SCA3

1.2.1 Caracterização do gene *ATXN3*

A identificação original do *ATXN3* e sua ligação à DMJ/SCA3 remontam ao trabalho pioneiro de Kawaguchi e colaboradores (1994), que realizaram uma

varredura em uma biblioteca de cDNAs de cérebros humanos utilizando uma sonda de DNA composta por 13 repetições CTG, complementares a repetições CAG (Kawaguchi et al. 1994). Utilizando essa estratégia, os autores conseguiram isolar um novo transcrito de mRNA de 1.776 pares de bases e contendo uma repetição CAG. Essa molécula foi denominada de MJD1a, e demonstrou-se que sua repetição CAG estava expandida (CAGexp) em todos os indivíduos afetados de uma família com diagnóstico clínico de DMJ/SCA3.[1] O gene *ATXN3* compreende cerca de 48,2 kb de DNA genômico no braço longo do cromossomo 14 (14q32.1), e é dividido em 11 éxons que são submetidos a extenso *splicing* alternativo em diversos tecidos, incluindo cérebro, retina, testículos e sangue periférico (Ichikawa et al. 2001; Bettencourt et al. 2010b). Duas das principais isoformas do *ATXN3* diferem apenas na porção 3' do gene, com os éxons 10 ou 11 sendo utilizados como o éxon terminal (Ichikawa et al. 2001). O trato polimórfico CAG do *ATXN3* está localizado no éxon 10, e ele geralmente não é uma sequência CAG pura, sendo comumente interrompido na terceira, quarta e sexta repetições pelos trinucleotídeos CAA, AAG e CAA, respectivamente (Limprasert et al. 1996).

A proteína correspondente ao gene *ATXN3* é denominada ataxina-3, e sua expressão já foi detectada em todos os tecidos investigados, tanto no período embrionário como na vida adulta (Ichikawa et al. 2001; Riess et al. 2008). Ainda que existam algumas isoformas de ataxina-3 de significado incerto e que são detectadas exclusivamente em pacientes com DMJ/SCA3 (Bettencourt et al. 2010a), não há diferenças nos níveis de expressão de ataxina-3 entre pacientes e indivíduos saudáveis. Da mesma forma, não há correlação entre a gravidade da doença ou o tamanho da repetição CAG e níveis de expressão, tanto a nível do mRNA (Nishiyama et al. 1996), quanto da proteína ataxina-3 (Paulson et al. 1997a; Tait et al. 1998).[2] Além disso, a expressão da ataxina-3 não é restrita às áreas do

[1] Na verdade, a família descrita por Kawaguchi e colaboradores (1994) recebeu o diagnóstico de "DMJ". Naquela época, propôs-se que SCA3 e DMJ fossem enfermidades distintas. Estudos subsequentes determinaram que as duas descrições se referiam à mesma entidade clínica e molecular.

[2] Essas observações são extremamente importantes, já que refutam a hipótese da "perda de função" (*i.e.,* diminuição dos níveis proteicos normais de ataxina-3) como o mecanismo patológico primário da DMJ/SCA3. Pelo contrário, a maioria das evidências laboratoriais atuais apontam para um ganho de função tóxica por parte da expansão do trato CAG em pacientes com DMJ/SCA3, como será discutido adiante.

sistema nervoso central mais caracteristicamente afetadas pela DMJ/SCA3, ainda que seus níveis de expressão sejam maiores em neurônios do que em células gliais (Riess et al. 2008). Investigações na região promotora do *ATXN3* demonstraram características de um gene *housekeeping*, com um padrão de expressão global tanto em células neuronais, como em células não-neuronais. Sua região promotora não apresenta um elemento *TATA box* definido, mas contém sequências de DNA regulatórias de interação com os fatores de transcrição CBF e SP1, localizadas entre os nucleotídeos -291 a -104 em relação ao sítio de início de transcrição (Schmitt et al. 2003).

1.2.2 A natureza polimórfica e instável das repetições CAG em *ATXN3*

O tamanho do trato CAG no *ATXN3* é extremamente polimórfico na população em geral, usualmente contendo de 3 a 40 trinucleotídeos (Dürr et al. 1996; van de Warrenburg et al. 2005). Entretanto, alelos com 14 e 23 repetições CAG são os mais comuns e totalizam cerca de 50% de todos os alelos não-patogênicos do *ATXN3* (Limprasert et al. 1996). Nas duas últimas décadas, avançou-se muito no entendimento das origens e da instabilidade genômica das expansões CAG no *ATXN3*. De fato, a primeira observação de aumentos do trato CAG durante transmissões meióticas em famílias com DMJ/SCA3 remonta ao ano de 1995 (Maciel et al. 1995). Desde então, muitos autores propõem que a "instabilidade intergeracional", uma das características mais marcantes de todas as doenças relacionadas a expansões de poliglutaminas, aconteça mais na direção do aumento da expansão portada pelo genitor e que isso explique o fenômeno da antecipação. A questão é ainda controversa, pela dificuldade em se registrar transmissões CAG completas – de toda a prole – de um número substancial de sujeitos, o que evitaria vieses de observação e finalmente confirmaria se expansões seriam de fato mais comuns do que contrações do CAGexp instável (Maciel et al. 1995).

Observou-se que transmissões paternas do CAGexp eram muito mais suscetíveis a instabilidades intergeracionais do que transmissões maternas. De

fato, os primeiros relatos desse fenômeno verificaram um risco quase 8 vezes maior de instabilidade do trato CAG no *ATXN3* durante uma transmissão paterna, em relação a meioses maternas (Igarashi et al. 1996). Além das características particulares da gametogênese em homens e mulheres, esse efeito está ligado, pelo menos parcialmente, à própria configuração gênica do *ATXN3*. Transmissões paternas do CAGexp contendo a variante C987 (ver seção 1.2.3 “Origem ancestral da DMJ/SCA3” abaixo) apresentaram um risco 75 vezes maior de sofrer instabilidade intergeracional, quando comparadas a transmissões maternas sem essa variante genética (Igarashi et al. 1996). Além de fatores em *cis* (no mesmo alelo em que se encontra o CAGexp), também existem evidências de que fatores em *trans* (no alelo sem a expansão CAG patogênica) modulam a instabilidade do trato CAG do *ATXN3* (Martins et al. 2008; Takahashi et al. 2010). Por exemplo, a idade do genitor afetado mostrou uma forte correlação (r=0,965) com a magnitude da instabilidade (no caso expansões, e não contrações) entre 42 transmissões paternas da DMJ/SCA3, sugerindo que mecanismos de reparo e manutenção do DNA possam estar envolvidos na instabilidade do trato CAG no *ATXN3* (Martins et al. 2008; Martins et al. 2014).

Evidências cumulativas também apontam para a existência de outra característica interessante em meioses de indivíduos com CAGexp: a segregação preferencial do alelo mutante.[3] Por exemplo, Ikeuchi e colaboradores (1996) revisaram 80 transmissões em 7 famílias com DMJ/SCA3 e detectaram a segregação do CAGexp em 73% das meioses (Ikeuchi et al. 1996). A análise dos espermatozoides de seis pacientes com DMJ/SCA3 também mostrou uma proporção de alelos mutantes maior do que a esperada, com cerca de 65% dos gametas contendo o alelo expandido (407 de 629 células analisadas) (Takiyama et al. 1997). Em um número substancial de irmandades totalmente genotipadas também se encontrou uma proporção significativamente maior de portadores do CAGexp do que a esperada pela segregação mendeliana (Souza et al. 2016).

[3] Entende-se por segregação alélica preferencial, também chamada de não-Mendeliana, aquela que ocorre em frequências significativamente diferentes das esperadas pelos postulados da genética clássica. Para uma condição autossômica dominante, como a DMJ/SCA3, espera-se que, em média, 50% da prole de um indivíduo afetado herde o alelo mutante.

Essas evidências sugerem que a presença de um alelo do *ATXN3* com um CAGexp confere vantagem seletiva às células germinativas, mas essa hipótese ainda precisa ser formalmente testada.

Finalmente, existem evidências epidemiológicas de que a presença do CAGexp levaria a um aumento do *fitness* genético de um indivíduo.[4] Prestes e colaboradores (2008) revisaram os dados de nascimento de 415 pacientes com DMJ/SCA3 pertencentes a 82 famílias, e identificaram que mulheres afetadas tinham em média mais filhos do que suas irmãs não afetadas e do que a população feminina local não afetada (Prestes et al. 2008). *Fitness* aumentado – ou a capacidade de deixar maior número de descendentes férteis (portadores ou não do CAGexp) do que o grupo controle – foi também documentado entre sujeitos com SCA1, HD e SCA2 (Frontali et al. 1996; Sena et al. 2018, comunicação pessoal).

O maior número de descendentes de pacientes com DMJ/SCA3 (*fitness*) e a tendência de que mais do que 50% dos mesmos seja portadora do CAGexp (segregação distorcida) poderiam explicar a persistência de alelos *ATXN3* patogênicos na população, mesmo frente ao fenômeno de antecipação da doença em gerações subsequentes. Alguns pacientes com DMJ/SCA3 apresentam grandes expansões no *ATXN3*, o que acarreta em idades de início muito precoces e óbito antes da idade reprodutiva. Por outro lado, forças seletivas positivas, como o *fitness* aumentado e a segregação preferencial do alelo mutante, poderiam explicar porquê a DMJ/SCA3 se fixou em tantas populações humanas, apesar de aparentemente ter tido muito poucas origens ancestrais, como veremos a seguir.

1.2.3 Origem ancestral da DMJ/SCA3

O primeiro grande estudo sobre as origens ancestrais da DMJ/SCA3, baseado na determinação de haplótipos contendo o alelo mutante, foi realizado no

[4] O *fitness* genético pode ser definido como a capacidade de um indivíduo, dadas as suas características genéticas, de deixar descendentes. No nível celular, gametas com maior *fitness* genético são aqueles que, devido ao seu material genético, possuem vantagem evolutiva sobre outros gametas.

ano de 2001 a partir dos dados de 249 famílias afetadas de diversas regiões e etnias (Gaspar et al. 2001). Nessa investigação, cinco marcadores microssatélite (D14S1015, D14S995, D14S973, D14S1016 e D14S977) e três SNPs (rs1048755, rs12895357 e rs7158733) no *ATXN3* foram genotipados. Esses mesmos SNPs recebem denominações diferentes em estudos distintos, e correspondem às variantes $A^{669}TG/G^{669}TG$, $C^{987}GG/G^{987}GG$ e TAA^{1118}/TAC^{1118}, respectivamente, comumente descritas em estudos na DMJ/SCA3 (Matsumura et al. 1996; Goto et al. 1997; Maciel et al. 1999). De todas as configurações haplotípicas possíveis ligadas ao CAGexp, os haplótipos $A^{669}C^{987}A^{1118}$ (ACA) e $G^{669}G^{987}C^{1118}$ (GGC) são os mais antigos e frequentes (Martins et al. 2007). Eles são também denominados de haplótipos "Flores" (ACA) e "São Miguel" (GGC), devido à sua distribuição particular nas ilhas açorianas homônimas (Gaspar et al. 2001).

Gaspar e colaboradores (2001) descobriram que 72% de todas as famílias com DMJ/SCA3 estudadas no mundo possuíam o haplótipo ACA ligado ao CAGexp, enquanto apenas 2% de indivíduos sem a mutação causal da DMJ/SCA3 possuíam esse haplótipo em seus alelos *ATXN3* não expandidos (Gaspar et al. 2001). Esses resultados confirmaram observações anteriores em coortes menores (Igarashi et al. 1996; Matsumura et al. 1996) e sugeriram uma origem portuguesa da DMJ/SCA3, especialmente em regiões com histórico de comércio com Portugal nos séculos XVI e XVII, como a Índia (Mittal et al. 2005) e o Brasil (Jardim et al. 2001a; de Castilhos et al. 2014). Nessas regiões, praticamente todos os alelos do *ATXN3* com CAGexp possuem o haplótipo ACA. Entretanto, investigações subsequentes empregando um maior número de marcadores genéticos no *ATXN3* estabeleceram uma origem asiática para a DMJ/SCA3, via haplótipo ACA, há mais de 6.000 anos atrás (Martins et al. 2007). De fato, esse haplótipo foi identificado até em dois pacientes de origem aborígene australiana, uma população geneticamente isolada a milhares de anos (Martins et al. 2012). Ainda de acordo com esses dados, alelos *ATXN3* mutantes ligados ao haplótipo GGC foram decorrentes de um segundo evento mutacional mais recente, que ocorreu há cerca de 1.500 anos atrás (Martins et al. 2007).

Além de origens distintas, os haplótipos ACA e GGC possuem importantes diferenças funcionais. Por exemplo, o haplótipo ACA foi associado a uma maior tendência à expansão do trato CAG durante a meiose, quando comparado a alelos mutantes com o haplótipo GGC (Martins et al. 2008). Mesmo considerando-se apenas os alelos não patogênicos do *ATXN3*, observou-se que alelos com a variante C^{987} (A<u>C</u>A) possuíam quase o dobro de repetições CAG, em média, do que alelos com a variante G^{987} (G<u>G</u>C) (Matsumura et al. 1996). Além disso, estudos detectaram um risco 7,4 vezes maior de ocorrência de grande instabilidade intergeracional[5] entre alelos mutantes do *ATXN3* com a variante C^{987} (A<u>C</u>A), quando comparados a alelos com a variante G^{987} (G<u>G</u>C) (Igarashi et al. 1996). Curiosamente, esse risco aumentado de instabilidade do CAGexp só estava presente se o alelo não expandido continha a variante G^{987} (G<u>G</u>C), mas não em portadores da variante C^{987} (A<u>C</u>A) em homozigose. Esse achado sugere a existência de um efeito inter-alélico (ou seja, em *trans*) na instabilidade do trato CAG no *ATXN3*, não somente em células somáticas (Igarashi et al. 1996), mas também em células germinativas, já que esse fenômeno foi também detectado na análise de haplótipos em espermatozoides de pacientes com DMJ/SCA3 (Takiyama et al. 1997).

1.2.4 Funções fisiológicas da proteína ataxina-3

Dependendo da isoforma, a proteína ataxina-3 pode ter dois ou três domínios de interação com ubiquitina (UIMs), além de um domínio N-terminal denominado de Josefina (Burnett et al. 2003; Todi et al. 2010; Kuiper et al. 2017). O domínio Josefina compreende o sítio catalítico da ataxina-3, centrado em um aminoácido cisteína na posição 14 da proteína e em dois sítios de ligação à ubiquitina. A configuração desse domínio é típica de um grupo de enzimas da classe cisteína protease conhecidas como deubiquitilases (DUBs), que funcionam como enzimas editoras de cadeias de poli-ubiquitina (Burnett et al. 2003). As DUBs removem

[5] Grande instabilidade intergeracional foi definida no estudo de Igarashi e colaboradores (1996) como uma diferença de 2 ou mais repetições CAG, para menos ou para mais, entre os alelos *ATXN3* expandidos do genitor afetado e da prole.

ubiquitinas de proteínas alvo, regulando assim suas taxas de degradação via proteasomo. A ataxina-3 interage apenas com cadeias de 4 ou mais ubiquitinas, e esse é o tamanho mínimo que uma cadeia de poli-ubiquitina precisa ter para ser reconhecida pelo sistema de degradação do proteasomo (Burnett et al. 2003).

Duas moléculas de ubiquitina podem ser ligadas através de quaisquer um de sete resíduos de lisina distintos, gerando cadeias de poli-ubiquitina com diferentes configurações, dependendo de quais resíduos de lisina são utilizados. As configurações mais comuns compreendem as lisinas nas posições 48 (K48) e 63 (K63) (Todi et al. 2010). A ataxina-3 reconhece cadeias de poli-ubiquitina com ambas as configurações, mas ligações via K63 possuem interação preferencial (Orr 2012). Interessantemente, o CAGexp não abole a função DUB da ataxina-3. Ao contrário, estudos *in vitro* indicam que a ataxina-3 mutante possui uma atividade DUB maior do que a versão selvagem da proteína (Burnett et al. 2003; Todi et al. 2010). Além disso, modificações pós-traducionais também regulam a atividade da ataxina-3. Por exemplo, a ubiquitilação da própria ataxina-3 no resíduo de lisina 117 aumenta sua atividade DUB, e versões da ataxina-3 com sequências poliQ expandidas são mais suscetíveis a essa modificação (Todi et al. 2010).

Schmitt e colaboradores (2007) geraram camundongos nocaute para um ou os dois alelos de *ATXN3*, e não observaram nenhum efeito adverso no desenvolvimento embrionário ou diferenças morfológicas marcantes em relação a camundongos selvagens (Schmitt et al. 2007). Testes comportamentais e motores também não revelaram nenhuma consequência funcional da ausência da ataxina-3. Entretanto, os autores detectaram um aumento global nos níveis de proteínas ubiquitiladas, tanto em extratos proteicos de cérebro como de testículos, sem consequências neurológicas ou reprodutivas aparentes (Schmitt et al. 2007). Esse leve excesso de proteínas ubiquitiladas na ausência da ataxina-3 vai ao encontro da função DUB da ataxina-3 (Burnett et al. 2003; Todi et al. 2010).

1.2.5 Expansões de poliglutamina no contexto do sistema de controle de qualidade de proteínas

Na DMJ/SCA3, assim como nas demais poliglutaminopatias, a citotoxicidade e a neurodegeneração são desencadeadas por proteínas com forte tendência a agregação e desestabilização da rede de controle de qualidade de proteínas (Matilla-Dueñas et al. 2014; Kuiper et al. 2017). Diversos componentes moleculares fazem parte desse mecanismo de homeostase proteica (Williams and Paulson 2008; Balchin et al. 2016; Sontag et al. 2017) e os mais relevantes são brevemente discutidos abaixo.

1.2.5.1 Chaperonas moleculares

Chaperonas moleculares (referidas adiante apenas como chaperonas) são proteínas que auxiliam outros polipeptídeos a adotar suas conformações nativas, seja durante a tradução nos ribossomos, ou na formação de intermediários suscetíveis a agregação gerados em condições de estresse celular (Opal and Zoghbi 2002). As chaperonas também participam da destinação de proteínas irreversivelmente mal enoveladas para os sistemas celulares de degradação, como a autofagia e os proteasomos. Coletivamente, as chaperonas e os sistemas de síntese e degradação proteica fazem parte de um intrincado programa celular devotado à promoção do controle de qualidade de proteínas (Balchin et al. 2016). Esse conjunto de ações de homeostase proteica é também conhecido como proteostase (Saibil 2013).

As chaperonas são classicamente conhecidas como proteínas de choque térmico (HSPs). Contudo, essa não é uma definição totalmente correta, já que a expressão gênica da maioria das HSPs não é regulada por choque térmico (Hageman and Kampinga 2009). As chaperonas estão presentes em praticamente todos os compartimentos celulares, incluindo núcleo, citoplasma, mitocôndrias, retículo endoplasmático (RE), periplasma e cloroplastos (Doyle et al. 2013), onde possuem funções adaptativas centrais. Por exemplo, a localização de chaperonas

no RE é especialmente importante, dada a intensa produção de proteínas nessa organela. De fato, falhas no processo de enovelamento de proteínas no RE desencadeiam uma cascata apoptótica conhecida como “resposta a proteínas não enoveladas” (*unfolded protein response*, UPR) (Hetz et al. 2015).

De acordo com seu mecanismo de ação, as chaperonas podem ser divididas em dois grandes grupos: as chaperonas dependentes de ATP, compreendendo as famílias HSP100, HSP90, HSP70 e HSP60, e as chaperonas independentes de ATP, como as pequenas proteínas de choque térmico (sHSPs) (Doyle et al. 2013). Brevemente, chaperonas HSP100 são proteínas da superfamília AAA+ e possuem duas funções enzimáticas básicas: (i) “desenovelase” (*unfoldase*), direcionando clientes[6] para proteases especializadas; e (ii) “desagregase” (*disaggregase*), participando na ressolubilização de proteínas agregadas, como o fazem as proteínas Hsp104 de leveduras e ClpB de bactérias (Saibil 2013). Seres humanos não possuem desagregases da família HSP100, mas recentemente foram descobertos complexos HSP70 específicos que possuem atividade de desagregase *in vitro* (Nillegoda et al. 2015). Chaperonas da família HSP90 são constitutivamente expressas em praticamente todos os tecidos corporais, participando no direcionamento de clientes para rotas de degradação (Saibil 2013). Proteínas HSP70 constituem o maior e mais abundante grupo de chaperonas, com ortólogos na maioria das organelas. Muito do que se conhece sobre o mecanismo de ação das chaperonas dependentes de ATP é originário de estudos sobre a atividade de membros da família HSP70. Finalmente, chaperonas HSP60, também conhecidas como chaperoninas, localizam-se principalmente nas mitocôndrias e cloroplastos, e são homólogas à clássica chaperonina GroEL de bactérias (Saibil 2013).

O genoma humano também contém genes que codificam para pelo menos 10 sHSPs, que em mamíferos são chamadas de HSPBs (Carra et al. 2005; Garrido et al. 2012). Estruturalmente, todas as sHSPs possuem um domínio C-terminal

[6] Entende-se por “clientes” de chaperonas moleculares qualquer proteína que interage com uma chaperona molecular, seja durante seu enovelamento inicial, ou em situações de desenovelamento fisiológico ou causado por estresse celular e/ou mutações.

conservado de 80-90 aminoácidos chamado de alfa-cristalina. Fora desse domínio, a homologia entre sHSPs é extremamente baixa (Garrido et al. 2012; Kampinga and Garrido 2012). Diferentemente de outras famílias de chaperonas moleculares, as sHSPs possuem a habilidade de se associar em grandes complexos de homo- e/ou heteroligômeros com mais de 50 subunidades distintas. A oligomerização das sHSPs, mediada por modificações pós traducionais como a fosforilação, regula sua distribuição celular e afinidade por diversos substratos (Garrido et al. 2012).

Diversas chaperonas são extremamente eficientes no reconhecimento de proteínas mal enoveladas e propensas a agregação (Reis et al. 2016). A família HSP40/DNAJ de chaperonas moleculares é de especial interesse, já que vários membros desse grupo são capazes de suprimir a agregação de proteínas com tratos poliQ expandidos (Kampinga and Craig 2010; Kakkar et al. 2012). Em linha com essas observações, demonstrou-se correlação entre os níveis proteicos da chaperona DNAJB1 e a AO em uma coorte holandesa de pacientes com DMJ/SCA3 (Zijlstra et al. 2010). Esse estudo utilizou a estratégia de *outliers* fenotípicos para a AO e observou menor expressão de DNAJB1 em fibroblastos de pacientes com AO significativamente mais precoce do que o esperado, quando comparada à expressão em indivíduos com AO tardia. Desse modo, o trabalho de Zijlstra e colaboradores (2010) lançou a hipótese de que maiores níveis de expressão de certas chaperonas – principalmente DNAJB1 e, talvez, de outros membros da família HSP40/DNAJ – possam conferir um fator protetor a indivíduos portadores de doenças ocasionadas por proteínas amiloidogênicas, como é o caso da DMJ/SCA3.

Em um importante estudo subsequente, Hageman e colaboradores (2010) testaram a capacidade de todas as chaperonas humanas das famílias HSP110/HSPH, HSP70/HSPA e HSP40/DNAJ de inibir a agregação de proteínas com tratos poliQ expandidos, incluindo huntingtina, ataxina-3 e atrofina-1 (Hageman et al. 2010). Surpreendentemente, dois membros homólogos e pouco estudados da família HSP40/DNAJ, conhecidos como DNAJB6 e DNAJB8, foram as chaperonas que mais eficazmente inibiram a agregação de proteínas poliQ expandidas. O desempenho de DNAJB6 e DNAJB8 contra a agregação e

toxicidade de proteínas poliQ foi superior inclusive ao de DNAJB1, que era considerada até então um dos melhores candidatos a alvo terapêutico nas poliglutaminopatias (Zijlstra et al. 2010). Enquanto DNAJB8 é expressa apenas nos testículos, DNAJB6 está presente em todos os tecidos corporais, incluindo o cérebro (Hageman and Kampinga 2009). Desse modo, estudos subsequentes focaram na atividade de DNAJB6, demonstrando forte eficácia na supressão de agregação não somente de proteínas com tratos poliQ expandidos (Gillis et al. 2013; Kakkar et al. 2016), mas também das proteínas beta-amiloide (Månsson et al. 2013; Månsson et al. 2014) e alfa-sinucleína (Aprile et al. 2017). *In vivo*, a expressão neuronal de DNAJB6 retardou a neurodegeneração e aumentou significativamente a sobrevida de camundongos transgênicos para a doença de Huntington (Kakkar et al. 2016). Essas observações fazem da DNAJB6 um dos candidatos mais interessantes a modulador fenotípico da DMJ/SCA3 e outras doenças de origem amiloidogência.

1.2.5.2 O sistema ubiquitina-proteasomo (UPS)

Através de modificações pós-traducionais, proteínas podem ser conjugadas à ubiquitina (Ub), um polipeptídeo de 76 aminoácidos que é ligado ao grupamento ε-NH_2 de resíduos de lisina de proteínas alvo via o resíduo de glicina C-terminal da Ub. Esse processo é chamado de ubiquitilação, e exerce várias funções biológicas dependendo a quais resíduos específicos a Ub é conjugada, assim como do padrão de conjugação da Ub (Dantuma and Bott 2014). Mais da metade do montante celular total de Ub existe na forma de espécies monoubiquitiladas conjugadas principalmente a histonas (Clague et al. 2015).

Proteínas alvo podem ser marcadas com uma ou mais moléculas únicas de Ub (monoubiquitilação), ou conjugadas a várias unidades de Ub fisicamente ligadas na forma de cadeias de poli-Ub (Lilienbaum 2013; Braten et al. 2016). Cada tipo de ligação de poli-Ub possui uma consequência fisiológica específica, em termos de sinalização celular, para a proteína alvo. Por exemplo, cadeias de 4 ou mais Ub ligadas via K48 formam o sinal mais comum de degradação via proteasomos,

enquanto cadeias de Ub via K63 podem ter funções regulatórias ou direcionar proteínas para autofagia (Lilienbaum 2013; Collins and Goldberg 2017). A transferência e ligação da Ub a proteínas alvo é realizada por uma série de reações envolvendo enzimas ativadoras (E1), conjugadoras (E2) e ligases (E3) de Ub. O genoma humano codifica duas isoformas E1 (UBA1 e UBA6), pelo menos 37 proteínas E2 e mais de 1.000 E3 ligases (Lilienbaum 2013).

O sistema ubiquitina-proteasomo (UPS) é uma maquinaria de degradação proteica responsável pelo processamento de até cerca de 90% de todos os polipeptídeos celulares, incluindo, principalmente, proteínas de meia-vida curta e mal enoveladas (Lilienbaum 2013; Collins and Goldberg 2017). Em células de mamíferos, o ator principal do UPS é o proteasomo 26S, um complexo multiproteico dependente de ATP devotado à degradação de proteínas alvo. O proteasomo 26S é composto por duas unidades distintas: o componente central 20S (~700 kDa) e a partícula regulatória 19S (~900 kDa) (Lilienbaum 2013). O *pool* celular de proteasomos 26S maduros consiste de uma mistura de componentes 20S com um ou dois complexos 19S (Collins and Goldberg 2017). Vinte e oito polipeptídeos formam o componente 20S, com duas cópias cada de sete subunidades α e β distintas. Essas proteínas são dispostas em uma estrutura em forma de barril, com um poro central estreito que permite apenas a passagem de polipeptídeos não enovelados de modo dependente de ATP. As subunidades α possuem função estrutural, enquanto as subunidades β englobam as atividades proteolíticas do tipo caspase (β1, para resíduos com carga negativa), tripsina (β2, para resíduos com carga positiva) e quimiotripsina (β5, para resíduos hidrofóbicos), permitindo a clivagem de proteínas em sequências específicas. Os peptídeos gerados por esses processos de clivagem possuem de 2 a 10 aminoácidos em média (Collins and Goldberg 2017), que são posteriormente reduzidos a aminoácidos individuais por peptidases independentes do proteasomo (Dantuma and Bott 2014).

Especula-se que proteínas com tratos poliQ expandidos possam afetar os proteasomos, devido à toxicidade das espécies amiloidogênicas ou ao sequestro de subunidades proteasomais em agregados (Schmidt et al. 2002; Dantuma and Bott 2014). Enquanto polipeptídeos contendo cadeias de até cerca de 30

glutaminas são normalmente processados por proteasomos, as peptidases proteasomais não são capazes de clivar tratos poliQ mais longos (Venkatraman et al. 2004). A incapacidade de degradar tratos poliQ expandidos provavelmente se deve às especificidades particulares dos proteasomos eucarióticos, já que proteasomos de arqueobactérias são capazes de processar sequências poliQ expandidas (Venkatraman et al. 2004). Estudos *in vitro* e *in vivo* mostraram que a superexpressão de proteínas com tratos poliQ expandidos diminui a efetividade funcional dos proteasomos e aumenta o número total de proteínas celulares ubiquitiladas (Kraut et al. 2012). O aumento global de espécies ubiquitiladas é comumente considerado um indicador de disfunção da maquinaria UPS, provavelmente devido à toxicidade direta de oligômeros poliQ solúveis (Takahashi et al. 2008). Experimentos também sugerem um mecanismo indireto de inibição da via proteasomal por proteínas com tratos poliQ expandidos, provavelmente devido a perturbações globais na homeostase da Ub (Dantuma and Bott 2014). Contudo, é importante ressaltar que os sistemas de controle de qualidade proteica, em especial UPS e autofagia, são extremamente intrincados e possuem diversos mecanismos compensatórios frente a condições adversas. Por exemplo, pelo menos parte dos efeitos adversos decorrentes da inibição do UPS são aliviados por um aumento fisiológico nos níveis de autofagia (Rubinsztein 2007; Lim and Yue 2015). Desse modo, o estudo isolado de apenas um componente da maquinaria de controle de qualidade proteica pode levar a resultados inconsistentes.

1.2.5.3 O sistema de autofagia

A autofagia é um processo catabólico dependente de vesículas de membrana dupla conhecidas como autofagossomos, que sequestram componentes intracelulares e os direcionam para degradação via lisossomos (Lilienbaum 2013). A maioria das proteínas de meia-vida longa é degradada via autofagia, mas essa rota também está envolvida no processamento de agregados proteicos e até mesmo de organelas inteiras (Rose et al. 2011; Martin et al. 2014). A importância da autofagia para as funções celulares em geral e, especificamente

de neurônios, ficou clara com a demonstração de que a inibição da autofagia leva à acumulação de proteínas ubiquitiladas e de agregados e, subsequentemente, à neurodegeneração (Hara et al. 2006; Komatsu et al. 2006). Além disso, mutações em diferentes genes que codificam componentes da via da autofagia são patogênicas, como é o caso de mutações no gene *ATG5* (Kim et al. 2016a).

Os processos de autofagia são geralmente divididos em três tipos (Kaur and Debnath 2015). O primeiro e mais comum é a macroautofagia (geralmente denominada simplesmente de autofagia), um mecanismo de processamento inespecífico de proteínas e organelas por autofagossomos e lisossomos (Lilienbaum 2013). O segundo tipo de autofagia é a microautofagia, compreendendo a captura direta de substratos por invaginações das membranas lisossomais ou endossomais e posterior degradação em lisossomos (Kaur and Debnath 2015). O último tipo de autofagia recebe o nome de autofagia mediada por chaperonas. Esse processo é distinto da macroautofagia, pois utiliza a chaperona HSPA8/HSC70 e suas co-chaperonas no reconhecimento específico de proteínas que possuem o motivo de aminoácidos KFERQ. Polipeptídeos com esse sinal de degradação são transportados através da membrana lisossomal pelo receptor LAMP2 (Lilienbaum 2013). Durante esse processo, co-chaperonas como CHIP, BAG1 e BAG3, atuam como seletoras e direcionadoras, enviando proteínas alvo para a maquinaria UPS ou autofagia, dependendo de estímulos celulares e interações específicas. Por exemplo, interações via BAG1 favorecem a via UPS, enquanto BAG3 direciona proteínas alvo principalmente para degradação via autofagia (Lilienbaum 2013).

Recentemente, descobriu-se outros tipos de autofagia seletiva, mediados por receptores especializados que reconhecem sinais de degradação específicos, de modo similar ao que acontece na via UPS (Kaur and Debnath 2015). Em mamíferos, o sinal de degradação mais comumente utilizado na autofagia seletiva é a Ub (Khaminets et al. 2015), e os receptores de autofagia p62/SQSTM1 e HDAC6 são de especial interesse no estudo de proteínas amiloidogênicas. Por exemplo, ambos os receptores facilitam a degradação de agregados proteicos por

um tipo especial de autofagia denominado de agrefagia (Øverbye et al. 2007; Dikic 2017).

1.2.6 Agregação proteica: mecanismo tóxico ou protetor?

Desde as primeiras caracterizações morfológicas de agregados intraneuronais em pacientes com doenças neurodegenerativas se especula sobre a real função dessas inclusões proteicas (Paulson et al. 1997b; Chai et al. 1999b; Takahashi et al. 2010). Por exemplo, agregados de ataxina-3 e Ub são encontrados tanto em tecido cerebral de pacientes com DMJ/SCA3 como em indivíduos saudáveis, ainda que a frequência dessas alterações seja muito maior entre pacientes (Fujigasaki et al. 2000). Esse achado levou à hipótese de que a ataxina-3 talvez seja transportada para o núcleo de neurônios como uma resposta fisiológica normal a condições estressantes e durante o envelhecimento natural, mesmo na ausência de um trato poliQ expandido (Fujigasaki et al. 2000). Todavia, essas observações não são suficientes para o estabelecimento de uma relação de causalidade entre agregação proteica e neurodegeneração.

Alguns autores defendem a ideia de que o processo de agregação proteica faz parte de uma estratégia celular que visa ao isolamento espacial de proteínas altamente amiloidogênicas, como é o caso de proteínas com tratos poliQ expandidos (Zoghbi and Botas 2002; Takahashi et al. 2008). De acordo com essa linha de raciocínio, é possível que a formação de inclusões possa ser uma resposta neuronal fisiológica que tem como objetivo limitar a toxicidade de polipeptídeos propensos à agregação (Chai et al. 1999a; Sontag et al. 2017). De fato, um estudo pioneiro realizado por Arrasate e colaboradores (2004) na doença de Huntington demonstrou maior longevidade e menor taxa de morte celular entre neurônios que formavam inclusões de poliQ, em comparação a células em que inclusões não foram detectadas (Arrasate et al. 2004). Isso vai ao encontro de investigações anteriores que detectaram baixa correlação entre a quantidade de agregados e a extensão da neurodegeneração em diferentes áreas cerebrais de pacientes com

doenças ocasionadas por tratos poliQ expandidos (Paulson 1999; Gusella and MacDonald 2000; Opal and Zoghbi 2002).

Evidências recentes também sugerem que a agregação proteica não é um processo passivo, mas sim parte de um mecanismo celular ativo e especializado na formação de inclusões (Kaganovich et al. 2008; Sontag et al. 2017). Diversas proteínas dedicadas ao transporte e à localização de agregados em compartimentos celulares específicos já foram identificadas (Sontag et al. 2017), incluindo chaperonas moleculares (Malinovska et al. 2012; Escusa-Toret et al. 2013). Esses achados corroboram a hipótese de que a formação de agregados proteicos faz parte de uma estratégia celular que visa ao isolamento de espécies solúveis altamente amiloidogênicas e tóxicas. De fato, diversos estudos mostram que os intermediários solúveis (principalmente os chamados oligômeros), e não os agregados insolúveis, são as principais espécies tóxicas nas poliglutaminopatias (Nagai et al. 2007; Takahashi et al. 2008; Kim et al. 2016b). Como discutido antes, esse é precisamente o motivo pelo qual a utilização de chaperonas moleculares é tão atrativa como estratégia terapêutica na DMJ/SCA3 e em outras doenças amiloidogênicas, já que diversas chaperonas reconhecem polipeptídeos solúveis mal enovelados e propensos à agregação (Bukau et al. 2006; Brehme and Voisine 2016; Mannini and Chiti 2017).

1.3 Objetivos

1.3.1 Objetivo geral:

- Identificar novos modificadores genéticos de fenótipos na doença de Machado-Joseph/ ataxia espinocerebelar do tipo 3 (DMJ/SCA3), focando na idade de início dos sintomas e na velocidade de progressão neurológica da doença.

1.3.2 Objetivos específicos:

- Definir o atual “estado da arte” na investigação de modificadores genéticos da idade de início (AO) na DMJ/SCA3 através de uma revisão sistemática e metanálise.
- Viabilizar a metodologia de amostragem de fenótipos extremos para a AO para a população gaúcha com DMJ/SCA3.
- Com base nos resultados da metanálise proposta, selecionar potenciais candidatos a modificadores genéticos da AO na DMJ/SCA3 e testá-los em coortes selecionadas de pacientes utilizando a metodologia de amostragem de fenótipos extremos para AO, entre outras.
- Avaliar o papel das chaperonas moleculares, em especial a DNAJB6, na modulação da AO e da velocidade de progressão neurológica da DMJ/SCA3.

Capítulo 2. O estado da arte sobre modificadores da idade de início na DMJ/SCA3

2.1 Manuscrito 1: "*Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis*"

Artigo publicado no periódico
Journal of Neurology, Neurosurgery & Psychiatry
(DOI: *10.1136/jnnp-2018-319200*)

RESEARCH PAPER

Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis

Eduardo Preusser de Mattos,[1,2,3] Maiara Kolbe Musskopf,[2] Vanessa Leotti Torman,[4,5] Maria Luiza Saraiva-Pereira,[1,2,6,7] Laura Bannach Jardim[1,2,7,8]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jnnp-2018-319200).

For numbered affiliations see end of article.

Correspondence to
Dr Laura Bannach Jardim, Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rua Ramiro Barcelos 2350, Brazil; ljardim@hcpa.edu.br

Received 11 July 2018
Revised 24 August 2018
Accepted 11 September 2018

To cite: de Mattos EP, Kolbe Musskopf M, Leotti Torman V, *et al*. *J Neurol Neurosurg Psychiatry* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/jnnp-2018-319200

ABSTRACT

Objectives To perform a systematic review and meta-analysis of genetic risk factors for age at onset (AO) in spinocerebellar ataxia type 3/Machado-Joseph disease (SCA3/MJD).

Methods Two authors independently reviewed reports on the mathematical relationship between CAG length at the expanded *ATXN3* allele (CAGexp), and other genetic variants if available, and AO. Publications from January 1994 to September 2017 in English, Portuguese or Spanish and indexed in MEDLINE (PubMed), LILACS or EMBASE were considered. Inclusion criteria were reports with >20 SCA3/MJD carriers with molecular diagnosis performed by capillary electrophoresis. Non-overlapping cohorts were determined on contact with corresponding authors. A detailed analysis protocol was registered at the PROSPERO database prior to data extraction (CRD42017073071).

Results Eleven studies were eligible for meta-analysis, comprising 10 individual-participant (n=2099 subjects) and two aggregated data cohorts. On average, CAGexp explained 55.2% (95% CI 50.8 to 59.0; p<0.001) of AO variability. Population-specific factors accounted for 8.3% of AO variance. Cohorts clustered into distinct geographic groups, evidencing significantly earlier AO in non-Portuguese Europeans than in Portuguese/South Brazilians with similar CAGexp lengths. Presence of intermediate *ATXN2* alleles (27–33 CAG repeats) significantly correlated with earlier AO. Familial factors accounted for ~10% of AO variability. CAGexp, origin, family effects and CAG length at *ATXN2* together explained 73.5% of AO variance.

Conclusions Current evidence supports genetic modulation of AO in SCA3/MJD by CAGexp, *ATXN2* and family-specific and population-specific factors. Future studies should take these into account in the search for new genetic modifiers of AO, which could be of therapeutic relevance.

INTRODUCTION

Spinocerebellar ataxia type 3/Machado-Joseph disease (SCA3/MJD) is a neurological condition characterised by expansion of a polymorphic trinucleotide CAG tract (CAGexp) at *ATXN3*. SCA3/MJD is the most common dominantly inherited ataxia worldwide,[1 2] and *ATXN3* alleles with ≥45 repeats code for ataxin-3 proteins with abnormally long polyglutamine (polyQ) sequences. PolyQ-expanded ataxin-3 is prone to aggregation into neuronal inclusions and exerts a gain of toxic function, which leads to neuronal toxicity and degeneration,[3] similarly to what happens in Huntington's disease and other SCAs.[4]

The longer the CAGexp at *ATXN3*, the earlier the age at onset (AO) of disease. A large body of evidence has established that AO is not entirely explained by CAGexp, which explains 50% to 60% of the variability in AO,[5–8] and that AO should be modulated by additional genetic and/or environmental factors. Several candidates have been proposed, such as apolipoprotein E genotypic status,[9–11] CAG length at normal *ATXN3*[8 12 13] and *ATXN2*[6 8] alleles, and protein levels of the DNAJB1 chaperone.[14]

Most of the proposed modifiers in SCA3/MJD were not replicated or had small effects that usually improve the explanation of AO variance by not more than 1%. The greater part of the missing variability in AO remains unexplained, suggesting that the main CAGexp-independent modulators of AO are still unknown. Here, we performed a systematic review and meta-analysis of genetic risk factors associated to AO in SCA3/MJD. By analysing both aggregate and individual-participant data of more than 2000 patients from 16 countries across three continents, we were able to detect an important origin-specific effect of CAGexp on AO and to confirm the effect of some putative risk factors published previously.

METHODS

A detailed methodology protocol for this study was registered at the PROSPERO (International Prospective Register of Systematic Reviews) database prior to data extraction and is available at https://www.crd.york.ac.uk/PROSPERO/ under record CRD42017073071.

Literature search and data extraction

MEDLINE (PubMed), LILACS and EMBASE were searched from January 1994 to September 2017 for reports on genetic factors related to AO in SCA3/MJD. Search terms employed were 'sca3' OR 'mjd' OR 'spinocerebellar ataxia type 3' OR 'spinocerebellar ataxia type-3' OR 'machado-joseph disease' OR 'machado joseph disease' AND 'age of onset' OR 'age-of-onset' OR 'age at onset' OR 'age-at-onset'. Peer-reviewed articles and meeting abstracts were included, and references were checked to guarantee maximal coverage. Two reviewers

(EPDM and MKM) independently assessed and extracted data into evidence tables. Any disagreement regarding eligibility was discussed with a third reviewer (LBJ).

Population, exposure, comparators, outcomes, and inclusion and exclusion criteria

SCA3/MJD heterozygotes from diverse geographical origins comprised the population under study. The CAG length at CAGexp was the main exposure considered for meta-analysis; other genetic variants were included in the meta-analysis as risk factors (exposures) if reported at least twice in literature. The outcome was the quantitative variable AO defined as the age at the first symptom, usually gait ataxia. Included studies should report on both (1) molecularly confirmed SCA3/MJD symptomatic and/or asymptomatic heterozygotes and (2) the relationship between *ATXN3* CAGexp (main exposure) and AO (outcome). The term 'carrier' was used here as a synonym for heterozygotes, symptomatic or not, with one *ATXN3* allele with ≥45 CAG repeats. We excluded studies reporting on <20 carriers, in languages other than English, Portuguese or Spanish. If multiple publications reported the same data, the most up-to-date and complete data set was included. Corresponding authors were contacted to check for duplicated data and to grant access to their updated, pseudonymised, individual-participant databases (IPDs), whenever possible. Otherwise, we used summary statistics from aggregated databases (ADs). Besides AO and CAGexp, data on gender, family, length of normal *ATXN3* CAG tracts and at other CAG-containing *loci* and/or additional genetic variants were retrieved, if available.

Risk of bias assessment and quality control

The outcome AO was poorly defined in some studies. Since most patients with SCA3/MJD develop gait ataxia as the first symptom,[15] we combined in a single model carriers with known AO of gait ataxia (AOga) or of first symptom; when both criteria were available for the same individual, AOga was chosen. Only studies that measured CAG repeats by capillary electrophoresis were considered. Participation in molecular diagnosis quality control programmes was also questioned and informed here.

Analysis and data synthesis

Boxplots were used to describe the variability on both AO and CAGexp among studies. The meta-analysis was composed of three main models. First, the global relationship between *ATXN3* CAGexp length and AO was investigated using data from both IPDs and ADs,[16] aiming at comparing the degree of explanation of the variability in AO by CAGexp across studies, as reported by the linear R^2 measure. A second meta-analysis used IPDs only. Since complex models (quadratic and logarithmic) were only marginally better at explaining the data (see online supplementary file 1), AO was not mathematically transformed, and linear regression was used. A third analysis tested the effects of gender, family and CAG length at the non-expanded *ATXN3* allele and at other CAG-containing *loci*, focusing on the improvement of the R^2 measure. Geographical origin and interaction between origin and CAGexp were always included as independent variables. With the exception of *ATXN1*, which was considered a continuous variable, the effect of all CAG-containing *loci* was assessed as both continuous and discrete variables using CAG length groups as published previously[6 8] (online supplementary file 2). The percentage of AO variability explained by belonging to the same family was tested with a fixed-effects model. Analyses were performed using the software R V.3.4.1 with packages lsmeans and lmSupport, and SAS OnDemand for Academics V.3.1 (SAS Institute). Graphs were generated with ggplot2. Results were considered statistically significant when $p<0.05$.

RESULTS

Systematic review

The search yielded 641 unique abstracts (online supplementary file 3); 140 studies testing the relationship between AO and CAGexp at *ATXN3* were selected for the systematic review (online supplementary file 4). Thirty-one studies investigated additional modifying effects on AO, including CAG repeat length at the non-expanded *ATXN3* (n=19), *ATXN1* (n=5), *ATXN2* (n=6), *CACNA1A* (n=7), *ATXN7* (n=4), *HTT* (n=4), *TBP* (n=3) and *ATN1* (n=4) alleles. AO differences according to length of GGGGCC repeats at *C9ORF72*, and CAG repeats at *RAI1* and *KCNN3* were each reported once. Another report correlated ataxin-3 and selected chaperones protein levels with AO. Allelic and/or genotypic status of single-nucleotide polymorphisms at 15 genes were also correlated with AO, including variants at *ATXN3* (rs3814834, rs709930 and rs910369; n=1 each), *APOE* (rs429358 and rs7412; n=4) and *ATXN2* (rs7969300), *BDNF* (rs6265), *BECN1* (rs60221525 and rs116943570), *CHIP* (rs6597), *hCAD* (rs12738235), *IL1A* (rs1800587), *IL1B* (rs16944), *IL6* (rs1800795), *MT-ND3* (rs2853826), *OGG1* (rs1052133), *PPARGC1A* (rs7665116), *TNF* (rs1800629) and *UCHL1* (rs5030732; n=1 each). Differences in AO according to the degree of promoter methylation at *ATXN3* were evaluated by two studies, using distinct methodologies. Gender of the affected individual and transmitting parent were correlated with AO in 14 and 6 studies, respectively. Two reports considered the effect of population of origin on AO, and one evaluated the familial dependency of AO. Data extraction, references and detailed information of all AO modifiers reported in the literature, including those not selected for meta-analysis, are described in online supplementary file 5.

After contacting all corresponding authors of studies that met the inclusion and exclusion criteria (n=11), we retrieved updated information on 10 non-overlapping IPDs and 2 ADs of symptomatic individuals only (figure 1). CAGexp and geographical origin were available for all IPDs. Additional data included length of non-expanded CAG tracts at *ATXN3* (n=9 cohorts), and at *ATXN1*, *ATXN2*, *CACNA1A* and *ATXN7* (n=4 cohorts). Information on gender and family effects were available for six and seven cohorts, respectively. Geographical origin, sample sizes and retrieved data for each cohort included in the meta-analysis are summarised in table 1 and detailed in online supplementary file 6.

Effect of CAGexp and geographical origin

Exposure to diverse CAGexp repeat lengths at *ATXN3* was the most studied risk factor. IPD and AD retrieved from 11 studies comprised four cohorts from Europe,[6 7 17 18] three from Asia,[8 19 20] one from North America,[6] one from Central America[21] and three from Brazil.[5 22 23] Brazilian cohorts comprised the Rio Grande do Sul (Brazil-RS) cohort[23] and cohorts from other Brazilian regions (Brazil-non-RS cohorts): namely, subjects from São Paulo State[22] and those described by Neurogenetics Network, a consortium of Brazilian researchers.[5] Using both IPDs and ADs, the global linear correlation coefficient between CAGexp and $\log_{10}$(AO) was $r=-0.743$ (95% CI -0.768 to -0.713, $p<0.001$), meaning that, on average, the causative mutation determines about 55.2% (50.8%–59.0%) of the AO variability in SCA3/MJD worldwide.

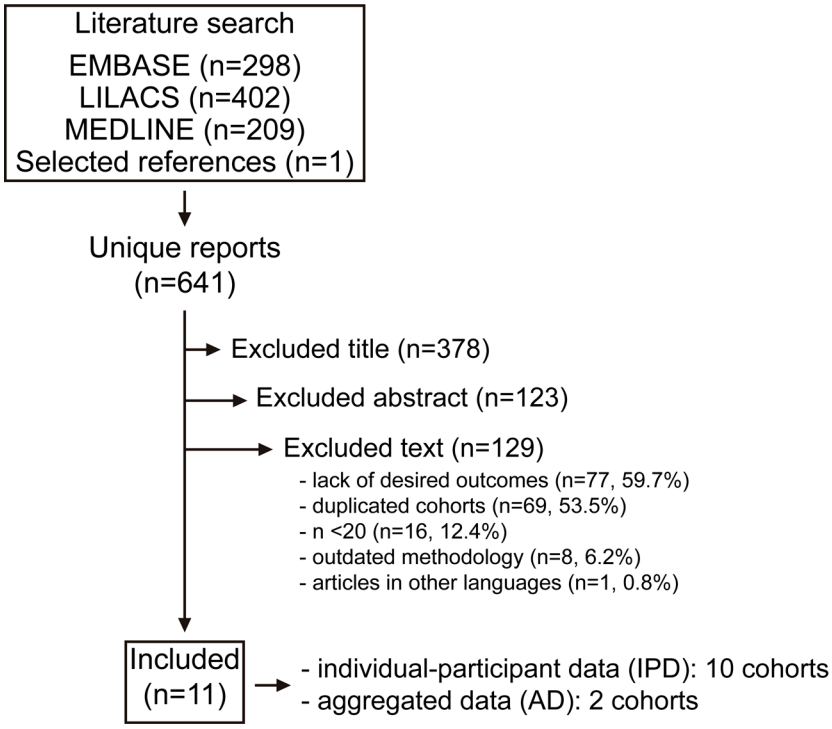


Figure 1 Workflow of the studies selected for the present meta-analysis.

Subsequent analyses used IPD cohorts only, totaling 2099 patients. Variability of CAGexp length among the 10 cohorts (figure 2A) was wider than variability in AO (figure 2B). Inclusion of geographical origin increased in 8.34% the explanation of AO variability (adjusted R^2=0.556; $F_{10,2091}$=263.8, p<0.001; online supplementary file 1). CAGexp significantly interacted with origin, which improved the model by an additional 1.02% (adjusted R^2=0.564; $F_{19,2082}$=144.1, p<0.001). The differential effect of CAGexp on AO among the 10 cohorts was evidenced by differences in slope and position of regression lines (figure 3A and B). Pairwise analysis of cohorts with similar slopes and/or intercepts allowed for data aggregation into three main

Table 1 Studies selected for meta-analysis of the modulation of age at onset in SCA3/MJD by CAG repeat length at *ATXN3* and additional factors

Cohort	N, published*	N, available†	Data type	QC programme‡	Available data (risk factors)	Reference
Rio Grande do Sul, Brazil	463	507	IPD	Yes	Family, gender, origin, *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A*, *ATXN7*	Souza *et al*[23] 2016
EUROSCA, Europe	403	403	IPD	Yes	Family, gender, origin, *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A*, *ATXN7*	Tezenas du Montcel *et al*[6] 2014
Taiwan	48	347	IPD	NS	Gender, origin, *ATXN3*	Wang *et al*[19] 2012
Portugal (Mainland)	48	226	IPD	Yes	Origin, *ATXN3*	Silveira *et al*[17] 1998
China	141	141	IPD	NS	Family, origin, *ATXN3* (expanded allele only)	Wang *et al*[20] 2017
Netherlands	342§	133	IPD	Yes	Family, origin, *ATXN3*	van de Warrenburg *et al*[18] 2005
São Paulo, Brazil	34	110	IPD	No	Origin, *ATXN3* (expanded allele only)	França Jr *et al*[22] 2009
Portugal (Azorean Islands)	93	106	IPD	Yes	Family, gender, origin, *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A*, *ATXN7*	Raposo *et al*[7] 2015
Neurogenetics network, Brazil	481¶	104	IPD	Yes	Family, gender, origin, *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A*, *ATXN7*	de Castilhos *et al*[5] 2014
Cuba	22	22	IPD	Yes	Origin, *ATXN3*	González-Zaldívar *et al*[21] 2015
China	802		AD	NS	Origin, *ATXN3* (expanded allele only)	Chen *et al*[8] 2016
USA	110		AD	NS	Origin, *ATXN3* (expanded allele only)	Tezenas du Montcel *et al*[6] 2014

*Number of patients reported on the original publication.
†Number of patients whose data were obtained after contacting the corresponding author of the selected publication.
‡Participation in molecular diagnosis quality control programme.
§This study reports on Dutch and French patients, but only the Dutch cohort was available for analysis.
¶This study includes patients from the Rio Grande do Sul cohort, but only patients from other Brazilian regions were retrieved for analysis.
AD, aggregated database; EUROSCA, European Consortium on Spinocerebellar Ataxias; IPD, individual-participant database; NS, not specified; SCA3/MJD, spinocerebellar ataxia type 3/Machado-Joseph disease.

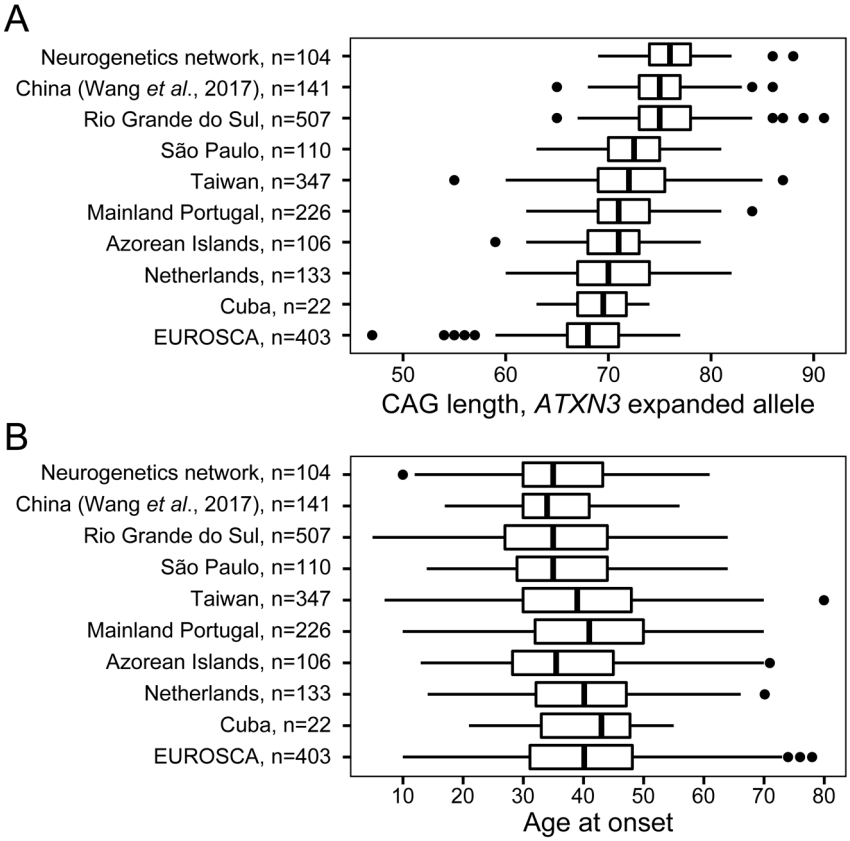


Figure 2 Variability of CAG length at the expanded ATXN3 allele and age at onset in SCA3/MJD. Data on CAG length at the expanded *ATXN3* allele and age at onset are shown for the 10 original, individual-participant data cohorts selected for meta-analysis. EUROSCA, European Consortium on Spinocerebellar Ataxias; SCA3/MJD, spinocerebellar ataxia type 3/Machado-Joseph disease.

geographical/ethnic groups with differential CAGexp modulation of AO: an average group with heterogeneous origins (China, Cuba, Brazil-non-RS and Taiwan cohorts), the group of non-Portuguese Europe (EUROSCA and Netherlands cohorts) and the group with clear Portuguese origin (Azorean Islands, mainland Portugal and Brazil-RS cohorts) (figure 3C and D; online supplementary file 7). Table 2 presents mean AO predictions for each geographically distinct group as a function of CAGexp tracts of three different lengths.

Effect of non-expanded CAG repeats at *ATXN3* and other CAG-containing *loci*

Data on CAG length of non-expanded *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A* and *ATXN7* alleles were available from 944 patients from four cohorts (EUROSCA, Azorean Islands, Brazil-RS and Brazil-non-RS) for meta-analysis. Inclusion of CAG length at the non-expanded *ATXN3* allele did not significantly improve the correlation between CAGexp and AO (p=0.327, continuous variable; p=0.388, discrete variable; online supplementary file 1). From the remaining candidate *loci*, only *ATXN2* significantly improved the explanation of AO variability, with longer *ATXN2* CAG tracts correlating with earlier AO (adjusted R^2=0.630; $F_{10,933}$=161.6, p<0.001; online supplementary file 1). There was a significant interaction between length of the longest CAG tract at *ATXN2* and CAGexp, which contributed an additional 0.39% to the explanation of AO variability (p=0.020; online supplementary file 1). Presence of at least one intermediate *ATXN2* allele (27–33 CAGs; 5% of alleles) significantly correlated with earlier AO (adjusted R^2=0.632; $F_{13,930}$=125.7, p<0.001; figure 4A and online supplementary file 1) in individuals with CAGexp tracts of up to 73 repeats (table 3).

Family and gender effects

Information on family effects was available for 1368 patients from 565 families (online supplementary file 8). Among these, CAGexp and origin alone explained ~60% of AO variance (adjusted R^2=0.599; $F_{11,1356}$=186.8; p<0.001). Inclusion of family data in a fixed-effects model increased the explanation by ~10% (adjusted R^2=0.702; $F_{888,479}$=4.6; p<0.001; online supplementary file 1). Data on gender were available for 1468 patients, and its inclusion contributed an additional 0.3% increase in the explanation of variability in AO (adjusted R^2=0.590; $F_{10,1457}$=211.9; p<0.001; online supplementary file 1). On average, male patients had younger ages at onset, especially among individuals with longer CAGexp tracts (figure 4B and online supplementary file 9). However, when considered together with CAG repeat length at *ATXN2* (n=942 individuals), the effect of gender was not significant (p=0.08; online supplementary file 1).

Combined effects

The final and best regression model considered CAGexp, origin, family effects and *ATXN2* genotypes, and explained 73.5% (95% CI 68.2 to 77.6) of the AO variance (adjusted R^2=0.735; $F_{682,245}$=4.8; p<0.001; online supplementary file 1).

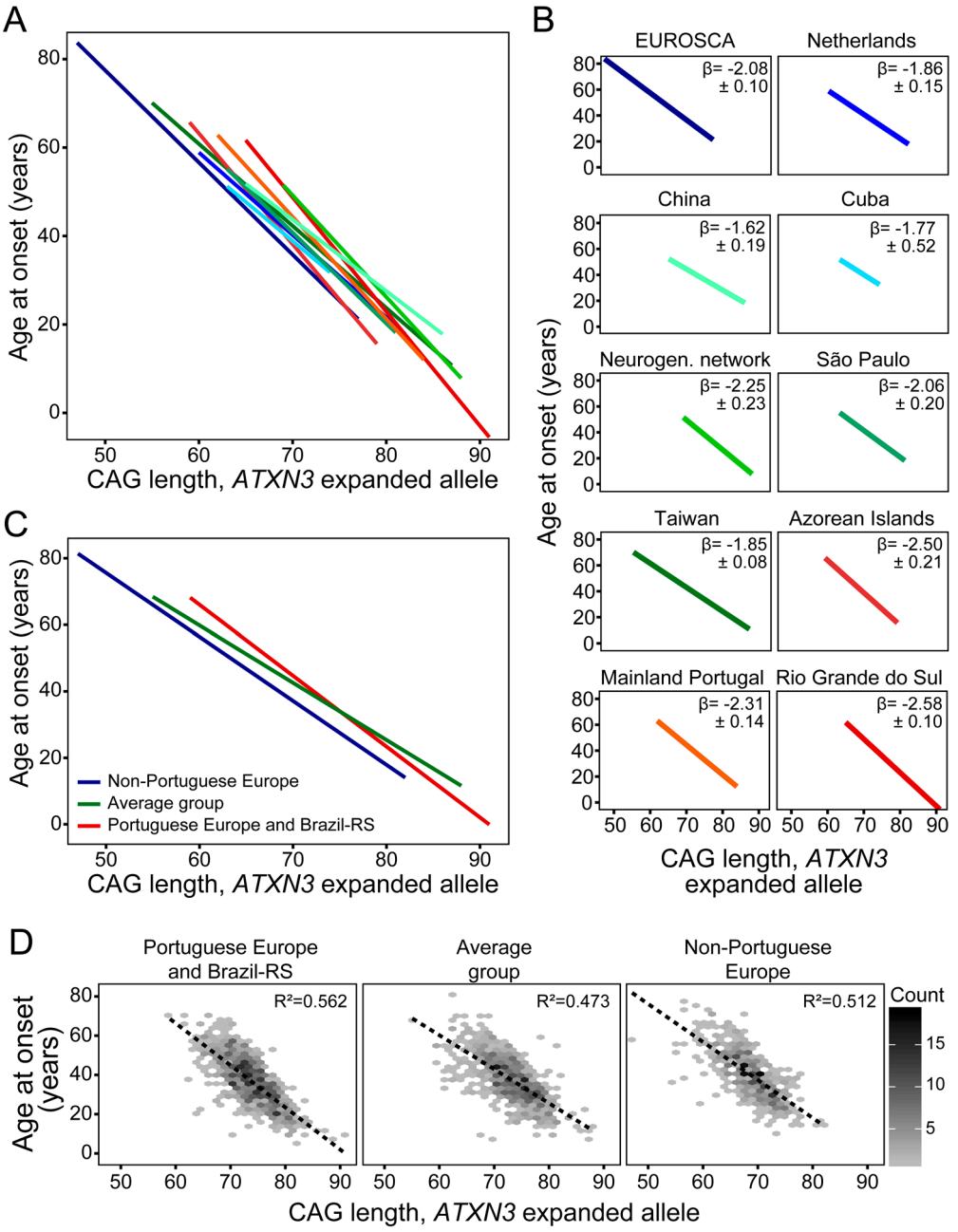


Figure 3 Population-specific modulation of the age at onset (AO) in spinocerebellar ataxia type 3/Machado-Joseph disease. (A) Linear regression of CAGexp and AO in the 10 distinct cohorts included in the present meta-analysis (cohorts are colour-coded as in panel B). (B) Same as in (A), but with individual plots for each cohort. The β coefficient of the linear regression, expressed as mean±SE, is also shown for each cohort. (C) Same as in (A), but with patients grouped in three main geographical origins, as a result of pairwise comparisons of intercepts and slopes of lines presented in (A). (D) Same as in (C), but with geographical origins depicted separately. Hexagons depict groups of patients, according to the colour scheme, and dashed lines represent linear regressions. For each plot, the R^2 correlation coefficient is also shown. EUROSCA, European Consortium on Spinocerebellar Ataxias.

DISCUSSION

This worldwide systematic investigation of risk factors for AO in SCA3/MJD detected that the CAGexp determines, on average, 55.2% of the phenotypic variability in AO. Additional modulation of AO by family factors, gender and CAG length at *ATXN2* were confirmed. Interestingly, currently unknown effectors related to geographical origin were also shown to modify AO. Although several more candidates have already been proposed, data were not robust enough to be meta-analysed, and further replication studies are necessary to assess their validity as phenotypic modulators in SCA3/MJD.

Effect of CAGexp and geographical/ethnic and family background

Clear geographical/ethnic differences on the effect of CAGexp on AO tell us that a universal correlation might not applicable to all carrier populations. Choice of statistical modelling might further evidence how populational differences in CAGexp can impact AO determination. Significant increase in explanation of AO variability was detected in Han Chinese,[8] European carriers from non-Portuguese populations and Americans[6] using quadratic models. Here, the quadratic modelling of CAGexp

Table 2 Effect of population of origin on the age at onset (AO) in SCA3/MJD

Parameter*	Portuguese Europe and Brazil-RS	Average group	Non-Portuguese Europe
Intercept	193.95 (184.36–203.55)[a]	163.12 (153.71–172.54)[b]	171.83 (160.78–182.88)[b]
Slope	−2.13 (−2.26 to −2.00)[a]	−1.72 (−1.85 to −1.59)[b,c]	−1.92 (−2.08 to −1.76)[a,c]
Predicted age at onset (years)			
For CAGexp=65	55.36 (54.10–56.62)[a]	51.20 (50.00–52.39)[b]	46.75 (45.84–47.66)[c]
For CAGexp=75	34.04 (33.45–34.62)[a]	33.98 (33.32–34.63)[a]	27.51 (26.27–28.75)[b]
For CAGexp=85	12.71 (11.14–14.29)[a]	16.76 (15.10–18.41)[b]	8.27 (5.55–10.99)[c]

Patients with SCA3/MJD from 10 cohorts with distinct geographical origins were grouped according to similarity of linear regression parameters. Mean AOs are presented for expanded CAG tracts at *ATXN3* (CAGexp) of different lengths.
*Data are presented as mean (95% CI). For each parameter, means sharing the same letter are not statistically different (Tukey-adjusted comparisons).
†
CAGexp, CAG length at the expanded ATXN3 allele; SCA3/MJD, spinocerebellar ataxia type 3/Machado-Joseph disease.

from IPDs yielded only a marginal improvement when compared with a simpler, linear regression modelling of AO variance. This is likely attributed to presence of individuals with larger CAGexp tracts (figure 2A), which correlate more strongly with AO,[7] compared with previous publications.[6 8]

Variation of CAGexp distribution was markedly larger than variation of AO among populations (figure 2). SCA3/MJD populations with larger CAGexp belonged to Brazilian and Asian cohorts. Inversely, subjects from Austria, Belgium, France, Germany, Hungary, Italy, Netherlands, Poland, Spain and UK (EUROSCA cohort)[6] had the shortest mean CAGexp. Reasons for such differences are still unknown. Although ascertainment bias usually operates in favour of recruiting more severe cases (ie, longer CAGexp tracts), this bias was unlikely in at least one cohort (Brazil-RS) with large CAGexp tracts since coverage in this population was shown to be very high.[23] Substantial differences in CAGexp determination are also unlikely since most included studies were performed in laboratories engaged in molecular diagnosis quality control programmes. Therefore, distinct CAGexp patterns likely represent true differences related to population of origin.

Pairwise comparisons allowed us to categorise carriers into three main geographical/ethnic groups, reflecting distinct relationships between CAGexp and AO (figure 3 and table 2), and suggesting that CAGexp does not have the same effect on AO of all SCA3/MJD carriers worldwide. Assuming that the 'average' group (figure 3C,D) represents the worldwide average relationship between CAGexp and AO in SCA3/MJD, our analysis suggests the existence of AO modifiers with opposing effects on non-Portuguese European carriers versus subjects with Portuguese ancestry (mainland Portuguese, Azorean and South Brazilians). There seems to be factors among non-Portuguese Europeans and carriers of Portuguese ancestry that effectively predispose to earlier and later AO, respectively, given a CAGexp of same length. It is also possible that the geographical/ethnic effect uncovered here reflects, at least partially, distinct *ATXN3* haplotypes and mutational origins, as different SCA3/MJD populations show distinct haplotypic frequencies.[24] Further research will be necessary to establish a causal link, if any, between CAGexp haplotypes and AO.

Familial effects might also be due to genetic AO modifiers, although the effect of shared environmental exposures within a family cannot be excluded. A significant decrease in residual AO variance within families, compared with that between families, was observed previously.[18 25] The ~10% improvement in R^2 observed here was smaller than the 25% observed in a French and Dutch cohort[18]; whether this was due to presence of several small families with one or two individuals in the meta-analysis remains to be established (online supplementary file 8).

Effect of the non-expanded *ATXN3* allele and of other non-expanded CAG-containing *loci*

In agreement with most original studies, there was no association between length of the non-expanded *ATXN3* allele and AO (online supplementary files 1 and 5) and that was also the case

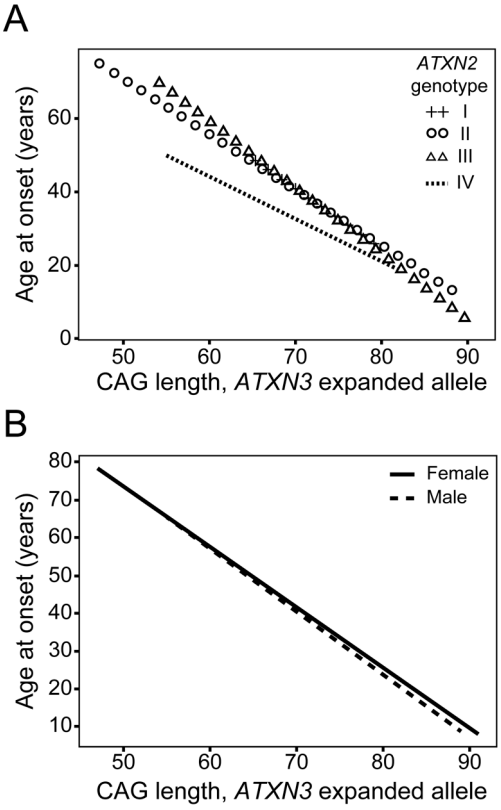


Figure 4 Effect of co-modifiers of the age at onset (AO) in SCA3/MJD. (A) Linear regression of CAG repeat length at the expanded *ATXN3* allele (CAGexp) and AO in patients with spinocerebellar ataxia type 3/Machado-Joseph disease, divided into four categories according to length of CAG tracts at both *ATXN2* alleles. I, at least one short (<22) allele; II, homozygous medium (22) alleles; III, at least one short intermediate allele (23–26) with or without a medium (=22) allele; IV, at least one intermediate (27–33) allele. (B) Same as in (A) but with individuals divided according to gender.

Table 3 Effect of *ATXN2* genotypes on the age at onset (AO) in SCA3/MJD

	ATXN2 genotype*			
Parameter	**I**	**II**	**III**	**IV**
Individuals, n (%)	32 (3.4)	634 (67.2)	47 (5.0)	231 (24.4)
Intercept†	228.03 (179.66–276.40)[a,b]	213.00 (199.41–226.60)[a]	213.85 (199.02–228.68)[a]	169.38 (142.50–196.27)[b]
Slope†	−2.64 (−3.31 to −1.97)[a,b]	−2.42 (−2.60 to −2.23)[a]	−2.43 (−2.64 to 2.23)[a]	−1.90 (−2.28 to −1.52)[b]
Predicted age at onset†				
For CAGexp=65	56.54 (51.21–61.87)[a]	55.69 (53.94–57.43)[a]	55.61 (53.80–57.43)[a]	45.64 (42.62–48.65)[b]
For CAGexp=70	43.35 (40.50–46.20)[a]	43.59 (42.54–44.64)[a]	43.44 (42.30–44.58)[a]	36.12 (33.90–38.33)[b]
For CAGexp=73	35.43 (32.83–38.03)[a]	36.33 (35.45–37.20)[a]	36.14 (35.08–37.20)[a]	30.41 (27.98–32.83)[b]
For CAGexp=74	32.80 (29.95–35.64)[a,b]	33.90 (33.02–34.79)[a]	33.70 (32.59–34.81)[a]	28.50 (25.90–31.11)[b]
For CAGexp=75	30.16 (26.94–33.37)[a,b]	31.48 (30.55–32.42)[a]	31.27 (30.08–32.46)[a]	26.60 (23.78–29.42)[b]
For CAGexp=85	3.77 (−5.30 to 12.84)[a]	7.28 (4.92–9.64)[a]	6.92 (4.11–9.74)[a]	7.56 (1.58–13.55)[a]

ATXN2 genotypes (*ATXN2*gen) were included in a linear regression model of AO as a function of CAG length at the expanded *ATXN3* allele (CAGexp) and geographical origin as follows: AO ~CAGexp+Origin+CAGexp*Origin+*ATXN2* gen+*ATXN2*gen*CAGexp. Mean AOs are presented for expanded CAG tracts at *ATXN3* of different lengths.
CAGexp, CAG length at the expanded *ATXN3* allele; SCA3/MJD, spinocerebellar ataxia type 3/Machado-Joseph disease.
**ATXN2* genotypes were defined as follows: (I) at least one short (<22) allele, (II) homozygous medium (=22) alleles, (III) at least one short intermediate (23–26) allele with or without a medium allele, and (IV) at least one intermediate (27–33) allele. †Data are presented as mean (95% CI). For each parameter, means sharing the same letter are not statistically different (Tukey-adjusted comparisons).
[†]Data are presented as mean (95% CI). For each parameter, means sharing the same letter are not statistically different (Tukey-adjusted comparisons).

for *ATXN1*, *ATXN7* and *CACNA1A*. However, since we did not have access to IPD from the large Chinese cohort that reported the associations between AO and *CACNA1A* and *ATXN7*,[8] population-specific differences in the range of CAG tracts at these loci—and in power to detect their potential effects—should not be overruled.

In contrast, we confirmed the association between non-expanded *ATXN2* alleles of intermediate CAG length (27–33 repeats) and earlier AO in SCA3/MJD (figure 4A, table 3 and online supplementary file 1), as reported previously.[6 8] Lack of confirmation in other cohorts is most likely attributed to small sample sizes[6 7] or inclusion of *ATXN2* in regression analysis as a continuous variable.[5] Whether the modulatory effect of *ATXN2* would be due to the CAG tract directly, or another genetically linked variant, is still unknown. Several observations support a biologically significant role for *ATXN2* and the normal ataxin-2 protein in neurodegenerative diseases. For instance, longer non-expanded *ATXN2* alleles have been related to increased risk of developing amyotrophic lateral sclerosis,[26] progressive supranuclear palsy,[27] frontotemporal dementia[28] and multiple systems atrophy.[29] Outside the CAG tract, a correlation between a missense polymorphism at *ATXN2* and earlier AO in Chinese patients with SCA3/MJD was recently shown.[30] Moreover, lower ataxin-2 levels were detected in brains of patients with SCA3/MJD and transgenic mice compared with healthy controls.[31] Importantly, restoration of ataxin-2 levels in affected mice led to significant morphological and behavioural improvements.[31] Therefore, the current evidence suggests that ataxin-2 is a strong candidate modifier of AO (and, maybe, disease progression) in SCA3/MJD.

Study limitations

Although great care was taken to control for potential biases and confounding factors, the present study is not without methodological limitations. Importantly, due to its retrospective assessment, it is possible that AO was not precisely defined for some of the individuals included. However, recalling biases were likely present in all patient cohorts, thus arguing in favour of true differences in AO among carriers from distinct populations/ethnicities. Moreover, different studies selected for meta-analysis had distinct definitions of AO, namely AO of the first symptom or AO of gait ataxia. Even though gait ataxia is usually the first symptomatic manifestation of SCA3/MJD, other symptoms might present before gait abnormalities.[15] While some of the largest patient cohorts included in this study had gait ataxia clearly stated as the parameter of choice for AO, which might have contributed to reduce AO heterogeneity, it is possible that distinct AO parameters are differentially modulated by CAGexp and/or other genetic factors.

Concluding remarks

The present analysis estimated that CAGexp is globally responsible for 55.2% of AO variance, on average. Gender and shared familial characteristics (most likely genetic) were confirmed as factors that influence AO in SCA3/MJD. Among candidate genes, CAG length at *ATXN2* was the only variant confirmed by the meta-analysis; future studies on the ataxin-3/ataxin-2 interactions might disclose promising discoveries.

Moreover, the IPD meta-analysis suggested protective factors in SCA3/MJD geographical groups of Portuguese origin, and probably a lack of some protective factors in non-Portuguese Europeans. Studying selected SCA3/MJD carrier groups, such as cohorts from specific geographical origins, or families with disease onset markedly different from the expected AO for their location, could significantly boost the search for genetic AO modulators.

The best model to assess the effect of the confirmed independent variables on AO determination in SCA3/MJD included CAGexp, geographical origin, family and CAG length at *ATXN2*: this model explained 73.5% of AO variability. That does not mean that factors responsible for the remaining variance should not have a genetic nature as well. In fact, several studies reviewed here assessed the effect of other genetic variants on AO, and some had promising results. Unfortunately, most were unique studies that were not qualified for meta-analysis. However, one of the advantages of meta-analyses is that updates can be performed in the future. Hopefully, further evidences on modifiers could increase the explanation of AO variability in SCA3/MJD, disclosing factors with potential therapeutic roles.

Author affiliations
[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[2]Laboratório de Identificação Genética, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
[3]Department of Biomedical Sciences of Cells & Systems, Section of Molecular Cell Biology, University Medical Center Groningen/Groningen University, Groningen, The Netherlands
[4]Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[5]Programa de Pós-Graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[6]Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[7]Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
[8]Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Acknowledgements This work would not be possible without the collaborative efforts of many people. We would like to thank all patients and corresponding authors of selected studies for sharing their data with us. This work was supported by the following Brazilian agencies: CAPES—Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Project 99999.01528/2015-2017; CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico, Project 402968/2012-3; FIPE-HCPA—Fundo de Incentivo à Pesquisa do Hospital de Clínicasde Porto Alegre, Projects GPPG HCPA 13-0303 and 14-0204. EPDM, MLS-P and LBJ were supported by CNPq.

Contributors EPDM, MLS-P and LBJ designed the study. EPDM, MKM and LBJ performed the systematic review. EPDM and VBL performed the statistical analysis. EPDM, MKM, VBL, MLS-P and LBJ wrote the manuscript.

Funding This study was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant no. 402968/2012-3), Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre (grant no. GPPG HCPA 13-0303GPPG HCPA 14-0204) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (grant no. 99999.01528/2015-2017).

Disclaimer Funding organisations had no role in study design; collection, analysis and interpretation of data; writing of the report; or decision to submit the article for publication.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Schöls L, Bauer P, Schmidt T, *et al*. Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. *Lancet Neurol* 2004;3:291–304.
2 Sequeiros J, Martins S, Silveira I. Epidemiology and population genetics of degenerative ataxias. *Handb Clin Neurol* 2012;103:227–51.
3 Matilla-Dueñas A, Ashizawa T, Brice A, *et al*. Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. *Cerebellum* 2014;13:269–302.
4 Kuiper EFE, de Mattos EP, Jardim LB, *et al*. Chaperones in polyglutamine aggregation: beyond the q-stretch. *Front Neurosci* 2017;11:1–11.
5 de Castilhos RM, Furtado GV, Gheno TC, *et al*. Spinocerebellar ataxias in Brazil—frequencies and modulating effects of related genes. *Cerebellum* 2014;13:17–28.
6 Tezenas du Montcel S, Durr A, Bauer P, *et al*. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain* 2014;137(Pt 9):2444–55.
7 Raposo M, Ramos A, Bettencourt C, *et al*. Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid? *Brain* 2015;138(Pt 12):e398.
8 Chen Z, Zheng C, Long Z, *et al*. (CAG)n loci as genetic modifiers of age-at-onset in patients with Machado-Joseph disease from mainland China. *Brain* 2016;139(Pt 8):e41.
9 Bettencourt C, Raposo M, Kazachkova N, *et al*. The APOE ε2 allele increases the risk of earlier age at onset in Machado-Joseph disease. *Arch Neurol* 2011;68:1580–3.
10 Peng H, Wang C, Chen Z, *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.e15–2179.e18.
11 Zhou Q, Ni W, Dong Y, *et al*. 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.
12 Dürr A, Stevanin G, Cancel G, *et al*. Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. *Ann Neurol* 1996;39:490–9.
13 França MC, Emmel VE, D'Abreu A, *et al*. Normal ATXN3 allele but not CHIP polymorphisms modulates age at onset in Machado-Joseph disease. *Front Neurol* 2012;3:164.
14 Zijlstra MP, Rujano MA, Van Waarde MA, *et al*. 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 Globas C, du Montcel ST, Baliko L, *et al*. Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. *Mov Disord* 2008;23:2232–8.
16 Pigott TD. *Advances in meta-analysis*. Boston, MA: Springer US, 2012.
17 Silveira I, Coutinho P, Maciel P, *et al*. Analysis of SCA1, DRPLA, MJD, SCA2, and SCA6 CAG repeats in 48 Portuguese ataxia families. *Am J Med Genet* 1998;81:134–8.
18 van de Warrenburg BP, Hendriks H, Dürr A, *et al*. Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort. *Ann Neurol* 2005;57:505–12.
19 Wang PS, Chen HC, Wu HM, *et al*. Association between proton magnetic resonance spectroscopy measurements and CAG repeat number in patients with spinocerebellar ataxias 2, 3, or 6. *PLoS One* 2012;7:e47479.
20 Wang C, Peng H, Li J, *et al*. Alteration of methylation status in the ATXN3 gene promoter region is linked to the SCA3/MJD. *Neurobiol Aging* 2017;53:192.e5–192.e10.
21 González-Zaldívar Y, Vázquez-Mojena Y, Laffita-Mesa JM, *et al*. Epidemiological, clinical, and molecular characterization of Cuban families with spinocerebellar ataxia type 3/Machado-Joseph disease. *Cerebellum Ataxias* 2015;2:1.
22 França MC, D'Abreu A, Nucci A, *et al*. Progression of ataxia in patients with Machado-Joseph disease. *Mov Disord* 2009;24:1387–90.
23 Souza GN, Kersting N, Krum-Santos AC, *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.
24 Martins S, Calafell F, Gaspar C, *et al*. Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. *Arch Neurol* 2007;64:1502–8.
25 DeStefano AL, Cupples LA, Maciel P, *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.
26 Elden AC, Kim HJ, Hart MP, *et al*. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 2010;466:1069–75.
27 Ross OA, Rutherford NJ, Baker M, *et al*. Ataxin-2 repeat-length variation and neurodegeneration. *Hum Mol Genet* 2011;20:3207–12.
28 Lattante S, Millecamps S, Stevanin G, *et al*. Contribution of ATXN2 intermediary polyQ expansions in a spectrum of neurodegenerative disorders. *Neurology* 2014;83:990–5.
29 Zhou X, Wang C, Ding D, *et al*. Analysis of $(CAG)_n$ expansion in ATXN1, ATXN2 and ATXN3 in Chinese patients with multiple system atrophy. *Sci Rep* 2018;8:3889.
30 Ding D, Li K, Wang C, *et al*. ATXN2 polymorphism modulates age at onset in Machado-Joseph disease. *Brain* 2016;139:aww176.
31 Nóbrega C, Carmo-Silva S, Albuquerque D, *et al*. Re-establishing ataxin-2 downregulates translation of mutant ataxin-3 and alleviates Machado-Joseph disease. *Brain* 2015;138(Pt 12):3537–54.

2.2 Manuscrito 2: "*Age at onset prediction in spinocerebellar ataxia type 3 changes according to population of origin*"

Artigo publicado no periódico
European Journal of Neurology
(DOI: *10.1111/ene.13779*)

ORIGINAL ARTICLE

Age at onset prediction in spinocerebellar ataxia type 3 changes according to population of origin

E. P. de Mattos[a,*], V. B. Leotti[b,*], B.-W. Soong[c,*], M. Raposo[d], M. Lima[d], J. Vasconcelos[e], H. Fussiger[f], G. N. Souza[g], N. Kersting[g], G. V. Furtado[a], J. A. M. Saute[g,h], S. A. Camey[b], M. L. Saraiva-Pereira[a,h,i] and L. B. Jardim[a,g,h,j]

[a]*Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul;* [b]*Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil;* [c]*Department of Neurology, Shuang Ho Hospital, Taipei Medical University School of Medicine, Taipei, Taiwan;* [d]*Faculdade de Ciências e Tecnologia, Universidade dos Açores, Ponta Delgada, Açores;* [e]*Serviço de Neurologia, Hospital do Divino Espirito Santo (HDES), Ponta Delgada, Açores, Portugal;* [f]*Programa de Pós-Graduação em Saúde da Criança e do Adolescente, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul;* [g]*Programa de Pós-Graduação em Ciências Médicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul;* [h]*Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul;* [i]*Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul; and* [j]*Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil*

Keywords:
age at onset, genetic modifier, Machado–Joseph disease, spinocerebellar ataxia type 3, survival models

Received 6 June 2018
Accepted 16 August 2018

European Journal of Neurology 2018, **0:** 1–8

doi:10.1111/ene.13779

Background and purpose: In spinocerebellar ataxia type 3/Machado–Joseph disease (SCA3/MJD), the length of CAG repeat expansions in *ATXN3* shows an inverse correlation with age at onset (AO). Recently, a formula for predicting AO based on CAG expansion was developed for European carriers. We tested this formula in SCA3/MJD carriers from distinct origins and developed population-specific models to predict AO.

Methods: This was a parametric survival modelling study.

Results: The European formula (EF) was tested in 739 independent SCA3/MJD carriers from South Brazil, Taiwan and the Portuguese Azorean islands, and it largely underestimated AO in South Brazilian and Taiwanese test cohorts. This finding challenged the universal use of the EF, leading us to develop and validate population-specific models for AO prediction. Using validation cohorts, we showed that Brazilian and Taiwanese formulas largely outperformed the EF in a population-specific manner. Inversely, the EF was more accurate at predicting AO among Portuguese Azorean patients. Hence, specific prediction models were required for each SCA3/MJD ethnic group.

Conclusions: Our data strongly support the existence of as yet unknown factors that modulate AO in SCA3/MJD in a population-dependent manner, independent of CAG expansion length. The generated models are made available to the scientific community as they can be useful for future studies on SCA3/MJD carriers from distinct geographical origins.

Introduction

Spinocerebellar ataxias (SCAs) are adult-onset neurodegenerative disorders with autosomal dominant inheritance. Several SCAs are caused by CAG repeat expansions (CAGexp) within coding regions of unrelated genes, being translated into neurotoxic polyglutamine-containing proteins [1].

In SCA type 3/Machado–Joseph disease (SCA3/MJD), CAGexp length determines ~50% of age at onset (AO) variability [2–5]. Although the average (range) CAGexp length in European and North American patients with SCA3/MJD was found to be

Correspondence: L. B. Jardim, Hospital de Clínicas de Porto Alegre, Rua Ramiro Barcelos 2350, 90035-003 Porto Alegre, Brazil
(tel.: + 55 51 3359 8011; fax: + 55 51 3359 8010;
e-mail: ljardim@hcpa.edu.br).
*These authors have contributed equally to the paper.

68.0 (54–77) [3,6,7], these values are skewed towards longer repeats (average, 75.1; range, 65–91) in Brazilian carriers [6,8]. A younger AO was consistently observed in Brazilian patients, when compared with European and North American patients [2,3,6,8–11].

Prediction of AO might impact genetic counselling and recruitment of pre-ataxic individuals for future clinical trials. To address this issue, maximum likelihood estimation of AO of first symptoms was previously used in SCA3/MJD [12], whereas others employed parametric survival models to estimate the probability of onset at a given age in Huntington's disease (e.g. [13]) and in Cuban SCA type 2 [14], European SCA type 1, SCA type 2, SCA3/MJD and SCA6 carriers [7]. These models were assumed to be applicable to independent patient cohorts. Recent evidence, however, suggests that additional CAGexp-independent factors might influence AO [4], arguing against a single AO prediction model and suggesting that the population background should be considered.

Here, we compared known AO of gait ataxia (AOga) in SCA3/MJD with predictions from the published European model [7], using independent cohorts from distinct geographical origins, i.e. South Brazil, Taiwan and the Portuguese Azorean Islands. We then developed and validated population-specific models for South Brazilian [Brazilian formula (BF)] and Taiwanese [Taiwanese formula (TF)] carriers.

Methods

A workflow illustrating the cohorts and analysis rationale used in this study is shown in File S1. Subjects with SCA3/MJD ($n = 739$) were stratified according to geographical origin. AOga was considered to be the age at the first walking disturbances, as reported by carriers and/or relatives (File S2 for details).

Study populations

South Brazilian cohort 1
A total of 100 symptomatic and 50 asymptomatic carriers were recruited from the Rio Grande do Sul SCA3/MJD population [8]. Their data were used to test the model built with data from European patients, referred here as to 'European formula' (EF) for AOga determination [7] (validation cohort for EF) and to generate BF (discovery cohort for BF).

South Brazilian cohort 2
An additional 107 patients from the South Brazilian SCA3/MJD population [8] were enrolled in a validation cohort to address predictions of BF, EF and TF.

Taiwanese cohort 3
A total of 40 asymptomatic and 227 symptomatic Taiwanese individuals, randomly assigned from an original cohort of 347 symptomatic subjects, were used to test EF (validation cohort for EF) and to generate TF (discovery cohort for TF).

Taiwanese cohort 4
The remaining 120 symptomatic Taiwanese individuals composed a validation cohort to address predictions of BF, EF and TF.

Portuguese Azorean cohort 5
A total of 95 symptomatic individuals from the Portuguese Azorean Islands [4] were used to test BF, EF and TF. SCA3/MJD in South Brazil is virtually entirely traced back to Azoreans who settled in this region between 1750 and 1770 [15]. Due to this genetic closeness, cohort 5 was chosen to validate prediction differences between BF and EF.

Clinical and molecular diagnosis

Individuals were genotyped at their local institutions where this study was conducted. Length of CAG repeats was determined by polymerase chain reaction using fluorescent primers for the *ATXN3* CAG repeat region and capillary electrophoresis.

Ethical aspects

This study was approved by the Ethics Committees of Hospital de Clínicas de Porto Alegre (14-0204), University of Azores (2/2016) and Shuang Ho Hospital, Taiwan. Confidentiality was guaranteed to all study participants, who gave written informed consent to participate in the study. Results from Brazilian ($n = 50$) and Taiwanese ($n = 40$) asymptomatic individuals were dealt with in a pseudonymized manner under an arrangement that ensured that results were not disclosed to anyone except the principal investigators in Brazil (L.B.J.) and Taiwan (B.W.-S.), respectively.

Statistical analysis

Prediction of median AOga for a given CAGexp length was calculated for all individuals from South Brazilian cohort 1, Taiwanese cohort 3 and Portuguese Azorean cohort 5, using EF for AOga determination [7]. Critical ranges (5th and 95th percentiles) were obtained for predicted AOga (File S3). Scatter plots were used to compare differences between observed and predicted AOga in symptomatic individuals from all cohorts or between the age at the last asymptomatic neurological evaluation

and the predicted AOga for asymptomatic individuals. Differences between observed and predicted values, expressed as mean prediction errors, were assessed by paired *t*-tests. Positive mean prediction errors indicated underestimations (observed AOga – predicted AOga > 0 years). Inaccurate AOga predictions for asymptomatic carriers were only detected when the predicted AO was earlier than the actual age of the individual at the time of her/his last clinical evaluation.

New prediction models were developed using data from the South Brazilian cohort 1 (BF) and Taiwanese cohort 3 (TF). For each cohort, data from symptomatic and asymptomatic carriers were fitted to four parametric survival models (log-normal, Gaussian, exponential and Weibull). The best-fitting models were chosen through residual analysis, Akaike information criterion and visual comparison with Kaplan–Meier curves (File S4). As data from both cohorts were better explained by a Gaussian parametric survival model, we adapted EF [7] to accommodate a Gaussian distribution (Files S5, S6 and S7).

Both BF and TF were fitted in the statistical software R [16], version 3.2.2, using the survreg function from the survival package. The print.psm function from the rms package was used to calculate R^2 [17]. Results were considered statistically significant when $P < 0.05$.

Results

Age at onset and CAG length of the expanded allele differences among distinct spinocerebellar ataxia type 3/Machado–Joseph disease populations

South Brazilian cohorts had mean AO of first symptoms, AOga and CAGexp values very similar to those reported previously for this population [8]. Mean CAGexp was statistically different among Portuguese Azorean (smaller mean CAGexp), Taiwanese (intermediate) and South Brazilian carriers (larger mean CAGexp) (two-tailed ANOVA with Tukey's *post hoc* test, $F_{2,736} = 69.51$, $P < 0.05$) (Table 1). However, mean AOga was only significantly different between Taiwanese and South Brazilian patients ($F_{2,646} = 5.08$, $P < 0.05$).

Accuracy of the European prediction model (European formula)

The EF underestimated known AOga of South Brazilian (Fig. 1a) and Taiwanese (Fig. 1b) patients. Mean prediction errors ranged from almost 6 years (Taiwanese group) to more than a decade (South Brazilian group) (File S8). EF predictions were more accurate for Portuguese Azorean patients (Fig. 1c and File S8).

The EF also underestimated the predicted AOga of South Brazilian and Taiwanese asymptomatic carriers (File S9). For 33/50 (66.0%) South Brazilian and 21/40 (52.5%) Taiwanese pre-clinical carriers, EF predicted younger AOga than the age of the individuals at the last asymptomatic neurological evaluation. Underestimations had a median of 6.93 (25th percentile, 5.31; 75th percentile, 13.71) and 6.12 (25th percentile, 3.44; 75th percentile, 13.42) years for South Brazilian and Taiwanese asymptomatic carriers, respectively.

Development and validation of population-specific models for age at onset of gait ataxia prediction

Based on parameters estimated from the South Brazilian cohort 1 and Taiwanese cohort 3 (Methods and

Table 1 Clinical and molecular data of spinocerebellar ataxia type 3/Machado–Joseph disease cohorts enrolled in the present study

	South Brazilian cohort 1		South Brazilian cohort 2	Taiwanese cohort 3		Taiwanese cohort 4	Portuguese Azorean cohort 5
	Patients	Carriers	Patients	Patients	Carriers	Patients	Patients
Total	100	50	107	227	40	120	95
Female	58 (58.0)	27 (54.0)	57 (53.3)	112 (49.3)	19 (47.5)	64 (53.3)	52 (49.1)
ATXN3 CAGnorm	22.5 (4.6; 14–31)	21.3 (5.7; 14–32)	20.9 (5.6; 13–36)	20.5 (6.7; 14–35)	18.5 (6.1; 14–32)	20.1 (6.7; 14–36)	21.5 (4.6; 14–29)
ATXN3 CAGexp	75.5 (3.2; 67–83)	73.7 (3.0; 68–84)	75.6 (3.4; 67–86)	72.0 (5.2; 55–87)	70.0 (4.2; 56–78)	70.5 (4.8; 60–87)	70.2 (3.8; 59–79)
AO gait ataxia	35.0 (10.1; 15–58)	NA	35.1 (10.6; 12–55)	38.6 (13.5; 7–80)	NA	39.8 (12.0; 8–69)	37.5 (12.1; 13–71)
Age at last asymptomatic evaluation	NA	31.8 (8.9; 19–54)	NA	NA	36.3 (7.5; 23–61)	NA	NA
SARA score	11.9 (4.7; 3.0–21.0)[a]	NA	13.0 (4.7; 4.0–29.0)[b]	ND	NA	ND	ND

AO, age at onset; CAGexp, CAG length of the expanded allele; CAGnorm, CAG length of the non-expanded allele; NA, not applicable; ND, not described; SARA, Scale for the Assessment and Rating of Ataxia. Data are given as *n* (%) and mean (SD; range). [a]Data available for 47 patients. [b]Data available for 57 patients.

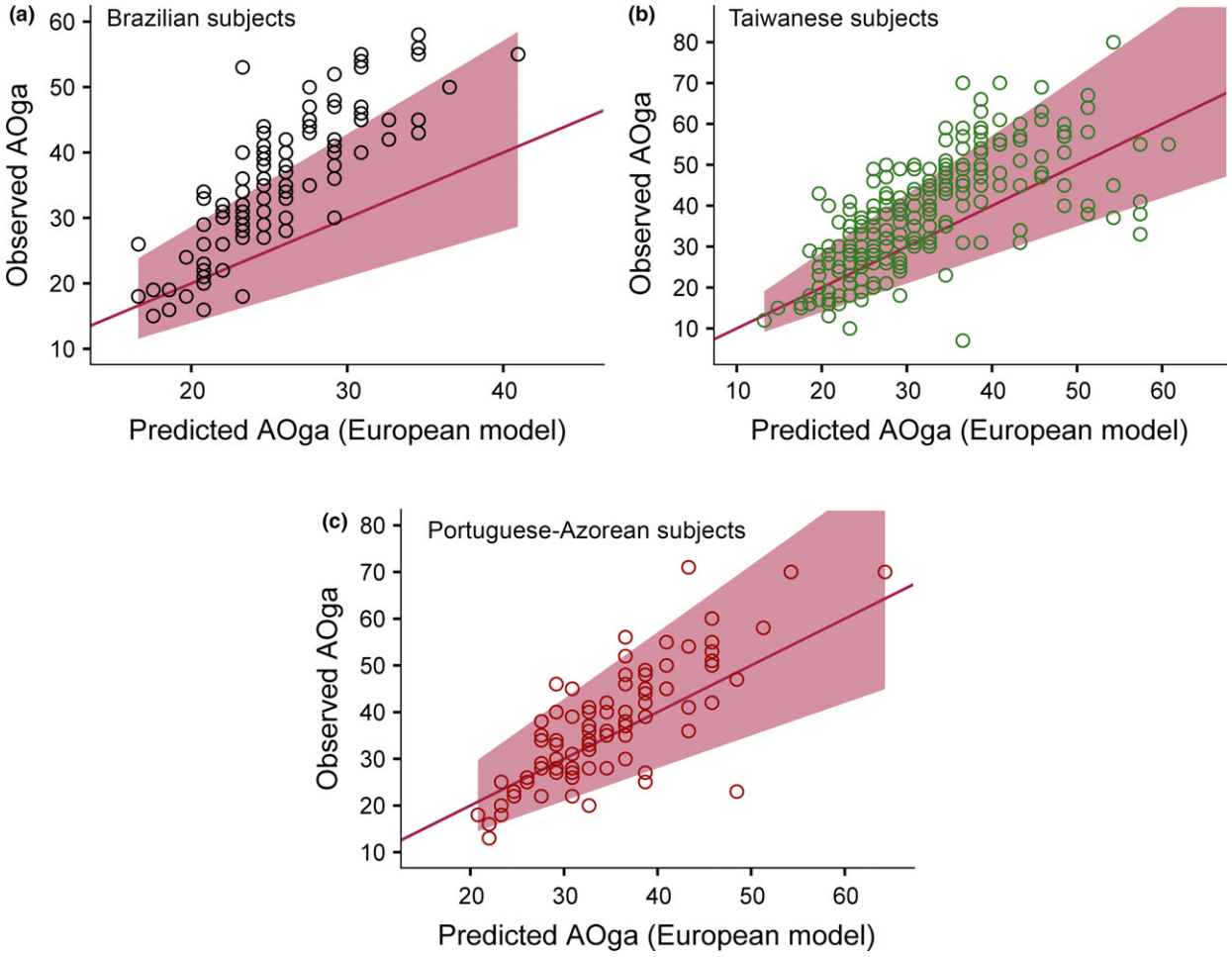


Figure 1 Prediction of age at onset of gait ataxia (AOga) in South Brazilian cohort 1 (a), Taiwanese cohort 3 (b) and Portuguese Azorean cohort 5 (c) symptomatic carriers using the European prediction model from Tezenas du Montcel *et al.* [7]. Regression line and confidence interval represent the expected AO, based on the European prediction model. [Color figure can be viewed at wileyonlinelibrary.com].

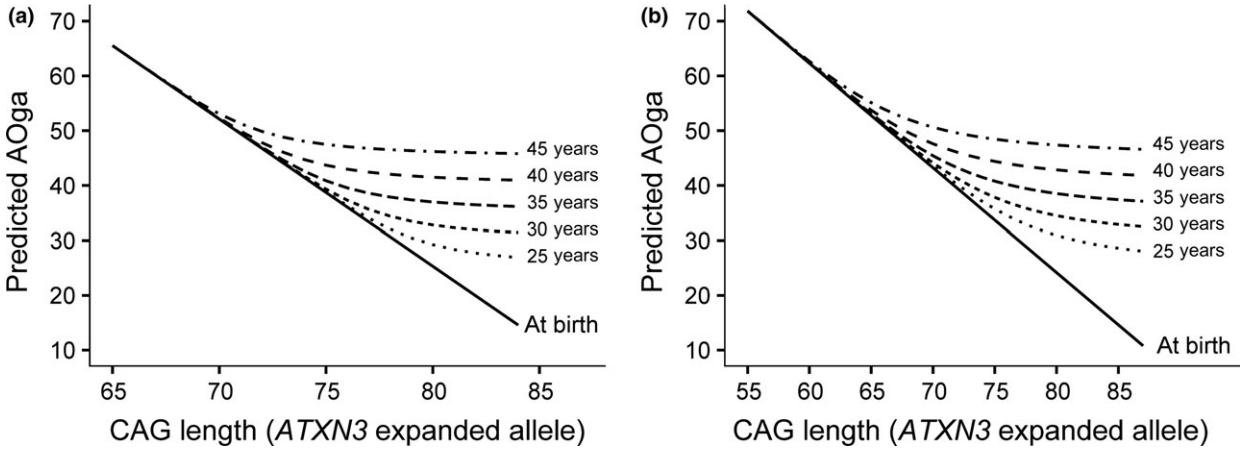


Figure 2 Population-specific predictions of age at onset of gait ataxia (AOga) in South Brazilian (a) and Taiwanese (b) individuals with spinocerebellar ataxia type 3/Machado–Joseph disease. Curves represent onset estimates at birth and at 25, 30, 35, 40 and 45 years of age. The x-axes of a and b are not drawn to the same scale due to distinct ranges of CAG repeat length in the two cohorts.

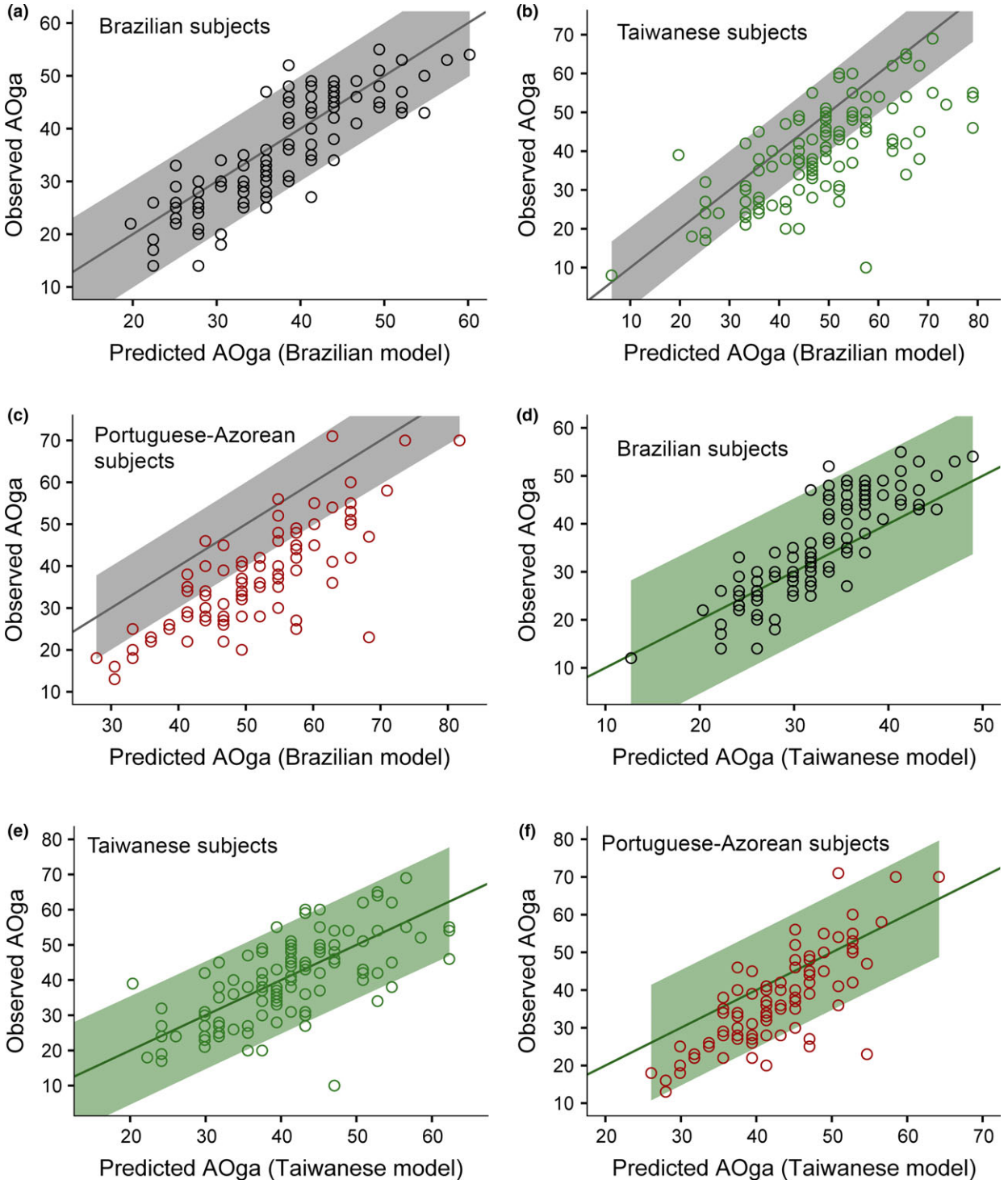


Figure 3 Validation of Brazilian and Taiwanese models for prediction of age at onset of gait ataxia in South Brazilian cohort 2 (a and d), Taiwanese cohort 4 (b and e) and Portuguese Azorean cohort 5 (c and f) patients with spinocerebellar ataxia type 3/Machado–Joseph disease. Shaded areas represent a 90% critical range. AOga, age at onset of gait ataxia. [Color figure can be viewed at wileyonlinelibrary.com].

File S7), new models BF and TF were respectively generated and used to estimate AOga in carriers within a wide range of CAGexp lengths (Fig. 2a and b, Files S5 and S6). CAGexp explained, on average, 58.9% and 50.0% of AOga variance in the South Brazilian cohort 1 and Taiwanese cohort 3, respectively. Inclusion of CAG length at the non-expanded *ATXN3* allele was not significant (BF, $P = 0.352$; TF, $P = 0.907$).

The BF and TF were then validated in the South Brazilian cohort 2, Taiwanese cohort 4 and Portuguese Azorean cohort 5. BF was accurate for South Brazilian individuals (Fig. 3a), but yielded large overestimations in Taiwanese (Fig. 3b) and Portuguese Azorean (Fig. 3c) patients (File S8). A similar population specificity was observed for TF, which was more accurate for Taiwanese carriers (Fig. 3e) when compared with predictions for South Brazilian (Fig. 3d) and Portuguese Azorean (Fig. 3f) individuals (Files S8 and S9). Excel files were made available to estimate AOga according to CAGexp for European (EF), Brazilian (BF) and Taiwanese (TF) SCA3/MJD carriers (Files S3, S5 and S6, respectively).

Discussion

The EF largely underestimated AOga in South Brazilian and Taiwanese SCA3/MJD carriers, even in subjects with CAGexp tracts in the range seen in European carriers. We then developed Brazilian and Taiwanese prediction models and validated them exclusively in a population-specific manner. These data suggest that (i) different AOga prediction formulas should be used for distinct ethnic groups and (ii) CAGexp at *ATXN3* has a differential contribution to AOga in distinct populations, possibly due to population-specific modifying factors.

Differences in AO and CAGexp between European and South Brazilian SCA3/MJD carriers have been acknowledged for a long time [8]. Although South Brazilian SCA3/MJD individuals tend to have longer CAGexp tracts than Europeans [2,3,5,6,8–11], we expected that EF should be more accurate, at least for individuals with expansions within the European range. However, this was not the case (Fig. 1a and b, and Files S8, S9 and S10). For instance, EF predicted AOga at 39.6 years for individuals with 68 CAG repeats, whereas BF and TF predicted much later AOga, i.e. 57.5 and 47.0 years, respectively (Files S3, S5 and S6). Populational specificity for the CAGexp–AO relationship was also suggested by the fact that the Gaussian model yielded the best fit for South Brazilian and Taiwanese cohorts, instead of the log-normal model used for EF [7]. We speculate that, if the *ATXN3* CAGexp range observed in South Brazilians occurred among Europeans, the effect on AOga would have disastrous consequences, producing very early-onset cases.

As BF largely overestimated AOga for Portuguese Azoreans, perhaps the European pattern of AOga dependency on CAGexp might have suffered a bottleneck/founder effect when Portuguese Azoreans settled in South Brazil. These settlers might have brought longer CAGexp tracts than the median expansion range from the original European population and simultaneously might have encoded protective factors, maybe in *cis* with CAGexp, that resulted in delayed disease onset and partially counterbalance the anticipatory effect of longer CAGexp in South Brazil. Conversely, there could be protective factors in the general population background, either genetic or environmental.

Although it is difficult to foresee the accuracy of predictions using EF, BF and TF in cohorts from other geographical origins, we anticipate that these models will require further local adjustments of parameters.

In conclusion, EF, BF and TF were validated for Portuguese Azorean, South Brazilian and Taiwanese SCA3/MJD carriers, respectively. We generated open access files that help to predict AOga according to a given CAGexp length, both at birth and at any given age for evaluation, for European, South Brazilian and Taiwanese carriers (Files S3, S5 and S6, respectively). If regional differences in AOga determination are confirmed in additional SCA3/MJD cohorts, this finding could have a deep impact not only on genetic counselling and on recruitment strategies of asymptomatic individuals for clinical trials, but also on the search for modifier factors that delay disease onset.

Acknowledgements

The authors thank the individuals who agreed to participate in this study. This work was supported by the following Brazilian agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; 402968/2012-3), Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre (14-0204) and Fundo de Apoio à Pesquisa do Rio Grande do Sul (1209-2551/13-4). Work in the Azores was supported by the project EXOS3 (PTDC/DTP-PIC/2638/2014) funded by FEDER through the COMPETE Program and by national funding through Fundação para a Ciência e a Tecnologia (FCT).

Disclosure of conflicts of interest

E.P.d.M., G.V.F., M.L.S.-P. and L.B.J. were supported by the National Council for Research and

Development (CNPq), Brazil. M.L. received grants from Fundação para a Ciência e Tecnologia, Portugal. J.A.M.S. received an unrestricted research grant from and is on the advisory board of PTC Therapeutics. L.B.J. received grants from Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre and Fundo de Apoio à Pesquisa do Rio Grande do Sul, Brazil. The other authors declare no financial or other conflicts of interest.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Supplementary File S1. Recruitment and analysis workflow of spinocerebellar ataxia type 3/Machado–Joseph disease cohorts included in the present study.

Supplementary File S2. Age at onset of first symptom versus age at onset of gait ataxia in patients with spinocerebellar ataxia type 3/Machado–Joseph disease from Brazilian cohort 1 and a discussion on the differences between these concepts.

Supplementary File S3. Predictions of the age at onset of gait ataxia for spinocerebellar ataxia type 3/Machado–Joseph disease carriers based on length of the CAG expansion in *ATXN3*, according to the European formula [7].

Supplementary File S4. Generation of population-specific models for prediction of the age at onset of gait ataxia in individuals with spinocerebellar ataxia type 3/Machado–Joseph disease.

Supplementary File S5. Predictions of the age at onset of gait ataxia for spinocerebellar ataxia type 3/Machado–Joseph disease carriers based on length of the CAG expansion in *ATXN3*, according to the Brazilian formula described here.

Supplementary File S6. Predictions of the age at onset of gait ataxia for spinocerebellar ataxia type 3/Machado–Joseph disease carriers based on length of the CAG expansion in *ATXN3*, according to the Taiwanese formula described here.

Supplementary File S7. Extended statistical methods.

Supplementary File S8. Comparison of performances of the Brazilian, European and Taiwanese prediction models of age at onset of gait ataxia in geographically distinct spinocerebellar ataxia type 3/Machado–Joseph disease cohorts. Mean prediction error refers to the difference between observed and predicted ages at onset of gait ataxia.

Supplementary File S9. Prediction of the age at onset of gait ataxia in South Brazilian and Taiwanese asymptomatic spinocerebellar ataxia type 3/Machado–Joseph disease carriers using the European prediction model (European formula).

Supplementary File S10. Additional comparisons of three population-specific models for prediction of the age at onset of gait ataxia in spinocerebellar ataxia type 3/Machado–Joseph disease carriers.

References

1. Bird TD. Hereditary ataxia overview. In: Pagon RA, Adan MP, Ardinger HH, *et al*., eds. *GeneReviews(®)*. Seattle, Washington: University of Washington, 1993–2017.
2. de Castilhos RM, Furtado GV, Gheno TC, *et al*. Spinocerebellar ataxias in Brazil – frequencies and modulating effects of related genes. *Cerebellum* 2014; **13:** 17–28.
3. Tezenas du Montcel S, Durr A, Bauer P, *et al*. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain* 2014; **137:** 2444–2455.
4. 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.
5. Chen Z, Zheng C, Long Z, *et al*. (CAG)n loci as genetic modifiers of age-at-onset in patients with Machado-Joseph disease from mainland China. *Brain* 2016; **139:** e41.
6. Jardim LB, Hauser L, Kieling C, *et al*. Progression rate of neurological deficits in a 10-year cohort of SCA3 patients. *Cerebellum* 2010; **9:** 419–428.
7. Tezenas du Montcel S, Durr A, Rakowicz M, *et al*. Prediction of the age at onset in spinocerebellar ataxia type 1, 2, 3 and 6. *J Med Genet* 2014; **51:** 479–486.
8. Souza GN, Kersting N, Krum-Santos AC, *et al*. Spinocerebellar ataxia type 3/Machado-Joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. *Clin Genet* 2016; **90:** 134–140.
9. van de Warrenburg BPC, Sinke RJ, Bemelmans CCV, Scheffer H. Spinocerebellar ataxias in the Netherlands. *Neurology* 2002; **58:** 702–708.
10. van de Warrenburg BPC, Hendriks H, Dürr A, *et al*. Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort. *Ann Neurol* 2005; **57:** 505–512.
11. Schmitz-Hübsch T, Coudert M, Bauer P, *et al*. Spinocerebellar ataxia types 1, 2, 3, and 6: disease severity and nonataxia symptoms. *Neurology* 2008; **71:** 982–989.
12. França MC, Emmel VE, D'Abreu A, *et al*. Normal ATXN3 allele but not CHIP polymorphisms modulates age at onset in Machado-Joseph disease. *Front Neurol* 2012; **3:** 164.
13. Langbehn DR, Hayden MR, Paulsen JS, PREDICT-HD Investigators of the Huntington Study Group. CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. *Am J Med Genet B Neuropsychiatr Genet* 2010; **153B:** 397–408.
14. Almaguer-Mederos LE, Falcón NS, Almira YR, *et al*. Estimation of the age at onset in spinocerebellar ataxia type 2 Cuban patients by survival analysis. *Clin Genet* 2010; **78:** 169–174.

15. Jardim LB, Pereira ML, Silveira I, Ferro A, Sequeiros J, Giugliani R. Machado-Joseph disease in South Brazil: clinical and molecular characterization of kindreds. *Acta Neurol Scand* 2001; **104:** 224–231.
16. R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018. https://www.R-project.org/. (accessed 30/04/2018).
17. Harrell FEJ. rms: Regression Modeling Strategies. R Package Version 5.1-0 2017; https://cran.r-project.org/web/packages/rms/index.html. (accessed 30/04/2018)

Capítulo 3. A apolipoproteína E (apoE) e a DMJ/SCA3

3.1 Manuscrito 3: "*ApoEε4 allele is associated with earlier age at onset in spinocerebellar ataxia type 3*"

Manuscrito em preparação

Title: ApoEε4 allele is associated with earlier age at onset in spinocerebellar ataxia type 3

Corresponding author information:

Laura B. Jardim, MD, PhD
Medical Genetics Service, Hospital de Clínicas de Porto Alegre
Rua Ramiro Barcelos 2350, 90035-003 Porto Alegre, Brazil
Phone: + 55 51 3359 8011 | Fax: + 55 51 3359 8010
ORCID: 0000-0001-6907-5068
e-mail: ljardim@hcpa.edu.br

Author list: Eduardo Preusser de Mattos[1,2,3], Gabriel Vasata Furtado[1,2], Vanessa Bielefeldt Leotti[4,5], Laura Bannach Jardim[1,2,8,9,10,] *, Maria Luiza Saraiva-Pereira[1,2,6,7,8]

Affiliations: [1] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 91501-970, Brazil. [2] Laboratório de Identificação Genética, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, 90035-903, Brazil. [3] Department of Cell Biology, University Medical Center Groningen/ Groningen University, Groningen, 9713 AV, The Netherlands. [4] Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 91509-900, Brazil. [5] Programa de Pós-Graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-003, Brazil. [6] Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-000, Brazil. [7] Programa de Pós-Graduação em Ciências Biológicas: Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre,

Rio Grande do Sul, 90035-003, Brazil. [8] Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, 90035-903 Brazil. [9] Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-003, Brazil. [10] Programa de Pós-Graduação em Ciências Médicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-003, Brazil.

Word count: 1853 words

Summary:

Presence of *apolipoprotein E* (*apoE*) ε2 allele has been implicated in age at onset (AO) modulation of spinocerebellar ataxia type 3/ Machado-Joseph disease (SCA3/MJD), a dominantly inherited neurodegenerative condition. However, conflicting results challenge this association, remaining unclear which (if any) *apoE* genotypes modulate AO in SCA3/MJD. Based on a cohort of 482 SCA3/MJD patients from South Brazil, we demonstrate here that individuals homozygous for the *apoE* ε4 allele had mean AO of first symptom almost 7 years earlier than carriers of other apoE genotypes, leading to a significant improvement on the rational for AO variability. Moreover, onset of gait ataxia was more than a decade earlier in ε4/ε4 individuals, on average, when compared to patients with one or no ε4 alleles. Differently from others, no modulatory effect of *apoE2* was seen in this present cohort. To the best of our knowledge, this is the largest evaluation of *apoE* status in SCA3/MJD patients to date. Moreover, we provide the first evidence of a putative, strong anticipatory effect of *apoE4* in SCA3/MJD. If replicated in additional cohorts, this finding could lead to interventions aiming at reducing apoE4 levels in polyglutamine disorders, similarly to a recent proposal for Alzheimer's disease.

Keywords: spinocerebellar ataxia type 3, Machado-Joseph disease, apolipoprotein E, age at onset, genetic modulator.

1. Introduction

Spinocerebellar ataxia type 3/ Machado-Joseph disease (SCA3/MJD) is a dominantly inherited neurodegenerative condition affecting mainly motor coordination and balance [1, 2]. The age at onset (AO) of disease shows an inverse correlation with length of a trinucleotide CAG expansion (CAGexp) at *ATXN3*, which causes SCA3/MJD. Multiple reports have shown that CAGexp explains just 40-60% of the variability in AO, suggesting that AO is influenced by additional factors [3, 4]. Several genetic modifiers of AO in SCA3/MJD have been proposed, although few were systematically evaluated in distinct populations. One exception is the analysis of the *apolipoprotein E* (*apoE*) gene, which has already been studied in a mixed cohort of Portuguese Azorean and Brazilian [5] and two Chinese cohorts [6, 7] of SCA3/MJD patients.

ApoE is a glycoprotein highly expressed in the brain (mainly in astrocytes) and liver [8]. In the central nervous system, apoE engages in the transport of cholesterol and lipids across cells. Human *apoE* is represented by three main alleles (ε2, ε3, and ε4), which encode for apoE2, apoE3 and apoE4 isoforms [9]. Two non-synonymous single nucleotide polymorphisms (SNPs, rs429358 and rs7412) are responsible for the differences between the three main apoE isoforms that differ at two amino acid residues only. Multiple studies evidenced a strong link between presence of *apoE* ε4 and higher risk of developing late-onset Alzheimer's disease (AD) [10]. On the other hand, studies on Parkinson's (PD) and Huntington's (HD) disease have associated the *apoE* ε2 allele with worse clinical outcomes. For instance, presence of *apoE2* was correlated with increased risk of developing PD [11], and earlier AO in HD [12]. However, the evidence for the contribution of *apoE2* to PD and HD remains inconsistent [13–16].

In SCA3/MJD, two of three studies detected significantly earlier AO among carriers of at least one *apoE* ε2 allele, compared to exclusive carriers of ε3 and ε4 alleles [5, 7]. On the other hand, analysis of a third independent cohort did not observe any anticipation effect of *apoE2* on AO [6]. Therefore, a precise role for apoE isoforms on the AO of SCA3/MJD individuals remains to be established. To

broaden on this knowledge, we evaluated the distribution of both *apoE* alleles and genotypes, and their potential differential contribution to AO, in a large, homogeneous cohort of SCA3/MJD patients from South Brazil.

2. Methods

Study subjects

Four hundred eighty-two molecularly-confirmed, symptomatic SCA3/MJD individuals (242 female, 50.2%) from the Rio Grande do Sul cohort were included in the study. General characteristics of this cohort, as well as a detailed description of data collection, were recently published [4]. Patients and their relatives were asked for AO of the first symptom (AOfs, n=482) and/or of gait ataxia (AOga, n=333). DNA was isolated from peripheral blood and length of CAG repeats at *ATXN3* was determined by capillary electrophoresis with fluorescently-labeled primers. One hundred (50% female) unrelated, apparently healthy individuals were used as controls to estimate allelic and genotypic frequencies in the normal population.

ApoE genotyping

Rs429358 and rs7412 were genotyped using TaqMan® SNP genotyping assays (Applied Biosystems, assay IDs C___3084793_20 and C____904973_10) according to the manufacturer's instructions. Polymerase chain reactions were performed on a 7500 Fast Real-Time PCR System (Applied Biosystems). The *apoE* alleles ε2 (T/T), ε3 (T/C), and ε4 (C/C) were inferred based on the combined allelic status of rs429358/rs7412. Genotype and allele frequencies were determined by direct counting. The distribution of rs429358, rs7412, and *apoE* genotypes was tested for Hardy-Weinberg equilibrium in both patients and controls.

Ethical aspects

This study was approved by the Institutional Review Board from Hospital de Clínicas de Porto Alegre (Plataforma Brasil CAAE record number 34463214.3.0000.5327).

Statistical analysis

Age at onset (both AOfs and AOga) was considered the dependent variable, while CAGexp, gender, and apoE genotypes were independent variables. The percentage of variability in AO explained by CAGexp was reported as the R^2 measure, based on linear regression models. Analysis of co-variance (ANCOVA) was employed to test the effects of gender and *apoE* genotype on AO, corrected for differences in CAGexp among groups, and expressed as means ± standard errors (S.E.) and adjusted R^2 values. Differences in allelic and genotypic frequencies were assessed with Pearson Chi-Square and Fisher's exact tests. Differences between AOfs and AOga were compared with a paired t-test and reported as mean and 95% confidence interval (C.I.). A parametric survival model for AOga prediction was used to calculate differences between observed and predicted AOga. Patients with a known AOga were categorized into early, intermediate, or late onset cases, depending whether predictions were below the 25^{th}, between the 25^{th} and 75^{th}, or above the 75^{th} percentiles. The statistical analysis programs SPSS v.19.0 and R v.3.2.2 were used and results were considered statistically significant when $p<0.05$.

3. Results

Allelic and genotypic frequencies of rs429358, rs7412, and *apoE* were in the equilibrium of Hardy-Weinberg and no differences were observed between SCA3/MJD patients and controls (**Supplementary Files 1** and **2**). As expected, most subjects were carriers of the ε3/ε3 genotype, but all possible *apoE* genotypic configurations were observed (**Table 1**). *ApoE* allelic and genotypic frequencies in the present cohort were similar to those observed in other cohorts (**Supplementary Files 3** and **4**).

Among SCA3/MJD patients, both AO measures (AOfs and AOga) showed a strong, inverse correlation with CAGexp (**Supplementary File 5A** and **5B**, respectively), explaining almost 70% of the variability in AO (R^2=0.699 and R^2=0.685, respectively, p<0.001). For 75/333 patients (22.5%) with information on both AOfs and AOga, the first symptom preceded gait ataxia by almost 1 year, on average (p<0.001, 95% C.I.=-1.25 – -0.69 years, **Supplementary File 5C**). Inclusion of *apoE* genotypic status in a linear regression model of AOfs and CAGexp was statistically significant ($F_{3,470}$=3.51, p=0.015) and raised in 0.5% the explanation of AO variability (adjusted R^2=0.704). Strikingly, individuals homozygous for the ε4 allele presented initial disease symptoms 6.6 years earlier, on average, than patients with other *apoE* genotypes, when corrected for differences in CAGexp (**Table 1** and **Supplementary File 6**). This was exclusively due to presence of the ε4 allele in homozygosis, since mean AOfs of ε2/ε4 and ε3/ε4 carriers was not significantly different from ε2/ε3 or ε3/ε3 carriers.

Similar to AOfs, *apoE* genotypes modulated significantly the relationship between CAGexp and AOga ($F_{3,323}$=6.61, p<0.001), increasing in 1.6% the explanation of AOga variability (adjusted R^2=0.701; **Supplementary File 7**). Homozygous *apoE* ε4/ε4 patients showed the first signs of gait ataxia 11.3 years earlier, on average, than subjects with other *apoE* genotypes. Interestingly, when patients with known AOga were classified into statistically early-, intermediate-, or late-onset cases (**see Methods section**), all five *apoE* ε4/ε4 carriers received the early-onset status (**Table 2**).

Inclusion of gender also improved the correlation between CAGexp and AOfs ($F_{1,472}$=8.18, p=0.004, adjusted R^2=0.704). Male patients had the first symptom 1.7 years earlier, on average, than female subjects (mean AOfs ± S.E. = 33.2 ± 0.4 versus 34.9 ± 0.4 years, respectively). When considered together, both gender and apoE genotype independently modulated AOfs ($F_{3,466}$=0.88, p=0.453, for the gender**apoE* interaction term). In this case, the explanation of variability in AO increased by 1.0% (adjusted R^2=0.709). When gait ataxia was used as the AO standard, inclusion of gender did not further improve the model ($F_{1,319}$=0.15, p=0.695).

4. Discussion

We report here the first evidence of association between the *apoE* ε4/ε4 genotype and earlier AO in SCA3/MJD. Inclusion of *apoE* status led to a significant improvement in the explanation of AO variability, corresponding to 0.5% for AOfs and 1.6% for AOga. The strong effect of CAGexp on AO associated to a small number of *apoE* ε4/ε4 carriers is likely to be responsible for these small mathematical improvements that can explain AO variability. Nevertheless, the *apoE* ε4/ε4 genotype correlated with marked phenotypic differences in AO, corresponding to an average anticipation of 6 years in AOfs, and more than 11 years in AOga. Due to sample size, we detected a higher absolute frequency of *apoE* ε4/ε4 carriers among SCA3/MJD patients than previous reports [5–7] (**see Supplementary File 4**). Conflicting reports on the association of earlier AO with either *apoE2* or *apoE4* in SCA3/MJD might be due to either (a) different mechanisms operating on distinct genetic backgrounds; (b) population stratification; or (c) different sample sizes. Smaller sample sizes might have been insufficient to detect a representative sample of *apoE* ε4/ε4 carriers. However, basic research on the contribution of distinct apoE isoforms to amyloid deposition have yielded contrasting results [8], and a real pathogenetic complexity related to apoE cannot be ruled out yet.

Several lines of evidence suggest worse amyloid-β42 (Aβ)-related outcomes for carriers of the *apoE* ε4/ε4 genotype when compared to individuals with other *apoE* genotypes. For instance, there are significantly lower levels of circulating Aβ in the cerebrospinal fluid of cognitively healthy carriers of the *apoE* ε4/ε4 genotype when compared to carriers of other genotypes, and this has been associated to increased deposition of amyloid plaques [17]. Similarly, AD mice engineered to co-express human *apoE4* had significantly higher levels of Aβ deposition in brain tissue when compared to animals expressing *apoE2* or *apoE3* [17, 18]. ApoE4 is less effective at transporting Aβ across bioengineered vessels than apoE2, which could account for the increased Aβ deposition in the context of apoE4 [19]. The fact that we observed earlier AO only in homozygous carriers of the *apoE* ε4 allele, but not

in ε2/ε4 or ε3/ε4 individuals, could reflect a dose-dependent effect of *apoE4*, as observed in the risk of developing AD [10].

Bettencourt and colleagues (2011) investigated a cohort of Brazilian and Portuguese patients and detected 20 individuals with the *apoE* ε2/ε3 genotype. Mean AO of those individuals was 5 years earlier than carriers of the ε3/ε3 genotype [5]. A similar anticipating effect of *apoE2* was subsequently observed in 14 ε2/ε3 Chinese SCA3/MJD patients [7], and authors speculated that apoE2 could contribute to earlier AO in SCA3/MJD due to defective cholesterol homeostasis in the brain (Bettencourt et al. 2011; Peng et al. 2014). On the other hand, a third large Chinese cohort [6], as well as the present SCA3/MJD cohort, failed to detect any association between *apoE2* and AO in 53 and 50 *apoE* ε2/ε3 carriers, respectively. Support for a negative role of *apoE2* comes mainly from observations of a slightly increased risk of developing PD among carriers of the *apoE* ε2 allele [11]. However, there is still little mechanistic evidence to support a clear role for apoE2 in PD or other synucleinopathies [8]. Moreover, a recent study that stratified patients with Lewy body disease according to presence or absence of concomitant AD neurologic changes, associated *apoE4* to increased risk of developing pure dementia with Lewy bodies [9], suggesting that apoE4 might also significantly affect α-synuclein. Given that Aβ, α-synuclein, and polyglutamine (polyQ) tracts of SCA3/MJD and other polyQ diseases have similar amyloidogenic properties, we speculate that apoE4 might have also a negative impact on polyQ biology, perhaps by increasing formation of aggregates and/or decreasing their degradation rates.

In summary, we described here the first evidence supporting the association between *apoE4* and earlier AO in SCA3/MJD. Although future studies in larger patient cohorts with distinct ethnic background will be crucial to confirm this finding, we suggest that the results obtained so far should be studied in a meta-analysis context.

Acknowledgements: The authors would like to thank all patients who agreed to participate in the present study. This study was supported by the following Brazilian

Agencies: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundo de Incentivo à Pesquisa e Eventos do HCPA (FIPE-HCPA). EMP, GVF, LBJ and MLSP were supported by CNPq.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of Interest: On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1 Riess O, Rüb U, Pastore A, Bauer P, Schöls L. SCA3: neurological features, pathogenesis and animal models. Cerebellum 2008;7:125–37.

2 Saute JAM, Jardim LB. Machado Joseph disease: clinical and genetic aspects, and current treatment. Expert Opin Orphan Drugs 2015;:1–19.

3 Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, Forlani S, Rakowicz M, Schöls L, Mariotti C, van de Warrenburg BPC, Orsi L, Giunti P, Filla A, Szymanski S, Klockgether T, Berciano J, Pandolfo M, Boesch S, Melegh B, Timmann D, Mandich P, Camuzat A, Goto J, Ashizawa T, Cazeneuve C, Tsuji S, Pulst S-M, Brusco A, Riess O, Brice A, Stevanin G. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain 2014;137:2444–55.

4 Souza GN, Kersting N, Krum-Santos AC, Santos ASP, Furtado G V, Pacheco D, Gonçalves TA, Saute JA, Schuler-Faccini L, Mattos EP, Saraiva-Pereira ML, Jardim LB. Spinocerebellar ataxia type 3/Machado-Joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. Clin Genet 2016;90:134–40.

5 Bettencourt C, Raposo M, Kazachkova N, Cymbron T, Santos C, Kay T, Vasconcelos J, Maciel P, Donis KC, Saraiva-Pereira ML, Jardim LB, Sequeiros J, Lima M. The APOE ε2 allele increases the risk of earlier age at onset in Machado-Joseph disease. Arch Neurol 2011;68:1580–3.

6 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.

7 Peng H, Wang C, Chen Z, Sun Z, Jiao B, Li K, Huang F, Hou X, Wang J, Shen L, Xia K, Tang B, Jiang H. The 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 Huynh T-P V., Davis AA, Ulrich JD, Holtzman DM. Apolipoprotein E and Alzheimer's disease: the influence of apolipoprotein E on amyloid-β and other amyloidogenic proteins. J Lipid Res 2017;58:824–36.

9 Tsuang D, Leverenz JB, Lopez OL, Hamilton RL, Bennett DA, Schneider JA, Buchman AS, Larson EB, Crane PK, Kaye JA, Kramer P, Woltjer R, Trojanowski JQ, Weintraub D, Chen-Plotkin AS, Irwin DJ, Rick J, Schellenberg GD, Watson GS, Kukull W, Nelson PT, Jicha GA, Neltner JH, Galasko D, Masliah E, Quinn JF, Chung KA, Yearout D, Mata IF, Wan JY, Edwards KL, Montine TJ, Zabetian CP. APOE ε4 increases risk for dementia in pure synucleinopathies. JAMA Neurol 2013;70:223–8.

10 Neu SC, Pa J, Kukull W, Beekly D, Kuzma A, Gangadharan P, Wang LS, Romero K, Arneric SP, Redolfi A, Orlandi D, Frisoni GB, Au R, Devine S, Auerbach S, Espinosa A, Boada M, Ruiz A, Johnson SC, Koscik R, Wang JJ, Hsu WC, Chen YL, Toga AW. Apolipoprotein E

genotype and sex risk factors for Alzheimer disease: A meta-analysis. JAMA Neurol 2017;74:1178–89.

11 Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide B-MMM, Schjeide LM, Meissner E, Zauft U, Allen NC, Liu T, Schilling M, Anderson KJ, Beecham G, Berg D, Biernacka JM, Brice A, DeStefano AL, Do CB, Eriksson N, Factor SA, Farrer MJ, Foroud T, Gasser T, Hamza T, Hardy JA, Heutink P, Hill-Burns EM, Klein C, Latourelle JC, Maraganore DM, Martin ER, Martinez M, Myers RH, Nalls MA, Pankratz N, Payami H, Satake W, Scott WK, Sharma M, Singleton AB, Stefansson K, Toda T, Tung JY, Vance J, Wood NW, Zabetian CP, 23andMe Genetic Epidemiology of Parkinson's Disease Consortium, International Parkinson's Disease Genomics Consortium, Parkinson's Disease GWAS Consortium, Wellcome Trust Case Control Consortium 2), Young P, Tanzi RE, Khoury MJ, Zipp F, Lehrach H, Ioannidis JPAA, Bertram L. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: The PDGene database. PLoS Genet 2012;8:e1002548.

12 Kehoe P, Krawczak M, Harper PS, Owen MJ, Jones AL. Age of onset in Huntington disease: sex specific influence of apolipoprotein E genotype and normal CAG repeat length. J Med Genet 1999;36:108–11.

13 Kálmán J, Juhász A, Majtényi K, Rimanóczy A, Jakab K, Gárdián G, Raskó I, Janka Z. Apolipoprotein E polymorphism in Pick's disease and in Huntington's disease. Neurobiol Aging 2000;21:555–8.

14 Li YJ, Hauser M a, Scott WK, Martin ER, Booze MW, Qin XJ, Walter JW, Nance M a, Hubble JP, Koller WC, Pahwa R, Stern MB, Hiner BC, Jankovic J, Goetz CG, Small GW, Mastaglia F, Haines JL, Pericak-Vance M a, Vance JM. Apolipoprotein E controls the risk and age at onset of Parkinson disease. Neurology 2004;62:2005–9.

15 Panegyres PK, Beilby J, Bulsara M, Toufexis K, Wong C. A study of potential interactive genetic factors in Huntington's disease. Eur Neurol 2006;55:189–92.

16 Andresen JM, Gayán J, Cherny SS, Brocklebank D, Alkorta-Aranburu G, Addis EA, US-Venezuela Collaborative Research Group, Cardon LR, Housman DE, Wexler NS. Replication of twelve association studies for Huntington's disease residual age of onset in large Venezuelan kindreds. J Med Genet 2007;44:44–50.

17 Castellano JM, Kim J, Stewart FR, Jiang H, DeMattos RB, Patterson BW, Fagan AM, Morris JC, Mawuenyega KG, Cruchaga C, Goate AM, Bales KR, Paul SM, Bateman RJ, Holtzman DM. Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. Sci Transl Med 2011;3. doi:10.1126/scitranslmed.3002156

18 Bales KR, Liu F, Wu S, Lin S, Koger D, DeLong C, Hansen JC, Sullivan PM, Paul SM. Human APOE Isoform-Dependent Effects on Brain β-Amyloid Levels in PDAPP Transgenic Mice. J Neurosci 2009;29:6771–9.

19 Robert J, Button EB, Yuen B, Gilmour M, Kang K, Bahrabadi A, Stukas S, Zhao W, Kulic I, Wellington CL. Clearance of beta-amyloid is facilitated by apolipoprotein E and circulating high-density lipoproteins in bioengineered human vessels. Elife 2017;6:1–24.

Table 1: Distribution of the age at onset of first symptom in South Brazilian SCA3/MJD patients (n=482), according to different *apolipoprotein E* (apo*E*) genotypes.

	***apoE* genotype**					
	ε2/ε2	**ε2/ε3**	**ε2/ε4**	**ε3/ε3**	**ε3/ε4**	**ε4/ε4**
Patients, n (%)	1 (0.2)	50 (10.4)	6 (1.2)	336 (69.7)	80 (16.6)	9 (1.9)
AO first symptom (AOfs)						
Mean (SE)	-	35.6 (1.6)	34.8 (6.3)	33.3 (0.6)	36.5 (1.1)	32.7 (4.2)
Adjusted, mean (S.E.) [a]	-	34.2 (0.9)	-	34.3 (0.3)	33.7 (0.7)	27.5 (2.1) [b]
***ATXN3* CAG length**						
CAGexp, mean (S.E.) [range]	-	74.9 (0.5) [65-84]	75.2 (1.4) [72-80]	75.8 (0.2) [67-91]	74.3 (0.4) [68-82]	73.4 (1.4) [69-81]
CAGnorm, mean (S.E.) [range]	-	21.3 (0.8) [14-38]	19.5 (3.7) [14-35]	22.1 (0.3) [13-39]	21.6 (0.6) [14-37]	22.7 (2.0) [14-34]

[a] Estimated age at onset for mean CAGexp = 75.40; [b] statistically different from other *apoE* genotypes (ANCOVA, $F_{3,470}$=3.51, p=0.015); AO: age at onset; CAGexp: CAG length at the expanded *ATXN3* allele; CAGnorm: CAG length at the non-expanded *ATXN3* allele; S.E.: standard error.

Table 2: Distribution of *apolipoprotein E* (apo*E*) genotypes among South Brazilian SCA3/MJD patients (n=333) divided into early-, intermediate-, and late-onset groups, according to the difference between observed and predicted age at onset of gait ataxia.

	***apoE* genotype, n (%)**					
AOga status [a]	**ε2/ ε3**	**ε2/ ε4**	**ε3/ ε3**	**ε3/ ε4**	**ε4/ ε4**	**Total**
Early	15 (11.1)	3 (2.2)	96 (71.1)	16 (11.8)	5 (3.8)	135 (40.5)
Intermediate	17 (10.6)	1 (0.7)	113 (70.6)	29 (18.1)	0 (0.0)	160 (48.0)
Late	4 (10.5)	1 (2.7)	27 (71.0)	6 (15.8)	0 (0.0)	38 (11.5)
Total	36 (10.8)	5 (1.5)	236 (70.9)	51 (15.3)	5 (1.5)	333 (100.0)

[a] Refers to the difference between the observed age at onset of gait ataxia (AOga) and the predicted AOga based on a parametric survival model considering the CAG length of the expanded *ATXN3* allele. Patients were classified as early (below the 25th percentile), intermediate (between the 25th and 75th percentiles), or late (above the 75th percentile) onset cases.

Supplementary File 1: *Apolipoprotein E* (*apoE*) allelic frequencies, as determined by the single nucleotide polymorphisms rs429358 and rs7412, among 482 SCA3/MJD patients and 100 unrelated, healthy controls.

	Allele	Control	SCA3/MJD	Total
apoE, n (%) [a]	ε2	18 (9.0)	58 (6.0)	76 (6.5)
	ε3	163 (81.5)	802 (83.2)	965 (82.9)
	ε4	19 (9.5)	104 (10.8)	123 (10.6)
	Total	200 (100.0)	964 (100.0)	1164 (100.0)
rs429358, n (%) [b]	T	181 (90.0)	858 (89.0)	1039 (89.3)
	C	19 (10.0)	106 (11.0)	125 (16.7)
	Total	200 (100.0)	964 (100.0)	1164 (100.0)
rs7412, n (%) [c]	T	18 (9.00)	58 (6.0)	76 (6.5)
	C	182 (91.0)	906 (94.0)	1088 (93.5)
	Total	200 (100.0)	964 (100.0)	1164 (100.0)

[a] p=0.620; [b] p=0.162; [c] p=0.276, Pearson Chi-Square test.

Supplementary File 2: *Apolipoprotein E* (*apoE*) genotypic frequencies, as determined by the single nucleotide polymorphisms rs429358 and rs7412, among 482 SCA3/MJD patients and 100 unrelated, healthy controls.

	Genotype	Control	SCA3/MJD	Total
***apoE*, n (%) [a]**	ε2/ε2	0 (0.0)	1 (0.2)	1 (2.0)
	ε2/ε3	18 (18.0)	50 (10.4)	68 (11.7)
	ε2/ε4	0 (0.0)	6 (1.2)	6 (1.0)
	ε3/ε3	63 (63.0)	336 (69.7)	399 (68.6)
	ε3/ε4	19 (19.0)	80 (16.6)	99 (17.0)
	ε3/ε4	0 (0.0)	9 (1.9)	9 (1.5)
	Total	100 (100.0)	482 (100.0)	582 (100.0)
rs429358, n (%) [b]	TT	81 (81.0)	385 (79.9)	466 (80.1)
	TC	19 (19.0)	88 (18.2)	107 (18.4)
	CC	0 (0.0)	9 (1.9)	9 (1.5)
	Total	100 (100.0)	482 (100.0)	582 (100.0)
rs7412, n (%) [c]	TT	0 (0.0)	1 (0.2)	1 (0.2)
	TC	18 (0.18)	56 (11.6)	74 (12.7)
	CC	82 (0.82)	425 (88.22)	507 (87.1)
	Total	100 (100.0)	482 (100.0)	582 (100.0)

[a] p=0.172; [b] p=0.531; [c] p=0.252, Fisher's exact test with 10,000 Monte Carlo simulations.

Supplementary File 3: *Apolipoprotein E* (*apoE*) allelic frequencies among studies reporting on SCA3/MJD patients.

	***apoE* allele, n (%) [a]**			
Cohort	**ε2**	**ε3**	**ε4**	**Total**
Bettencourt *et al*., 2011	23 (6.0)	321 (84.3)	37 (9.7)	381 (100.0)
Peng *et al*., 2014	21 (6.8)	254 (81.9)	35 (11.3)	310 (100.0)
Zhou *et al*., 2014	62 (7.7)	654 (81.1)	90 (11.2)	806 (100.0)
Present study	58 (6.0)	802 (83.2)	104 (10.8)	964 (100.0)
Total	164 (6.7)	2031 (82.5)	266 (10.8)	2461 (100.0)

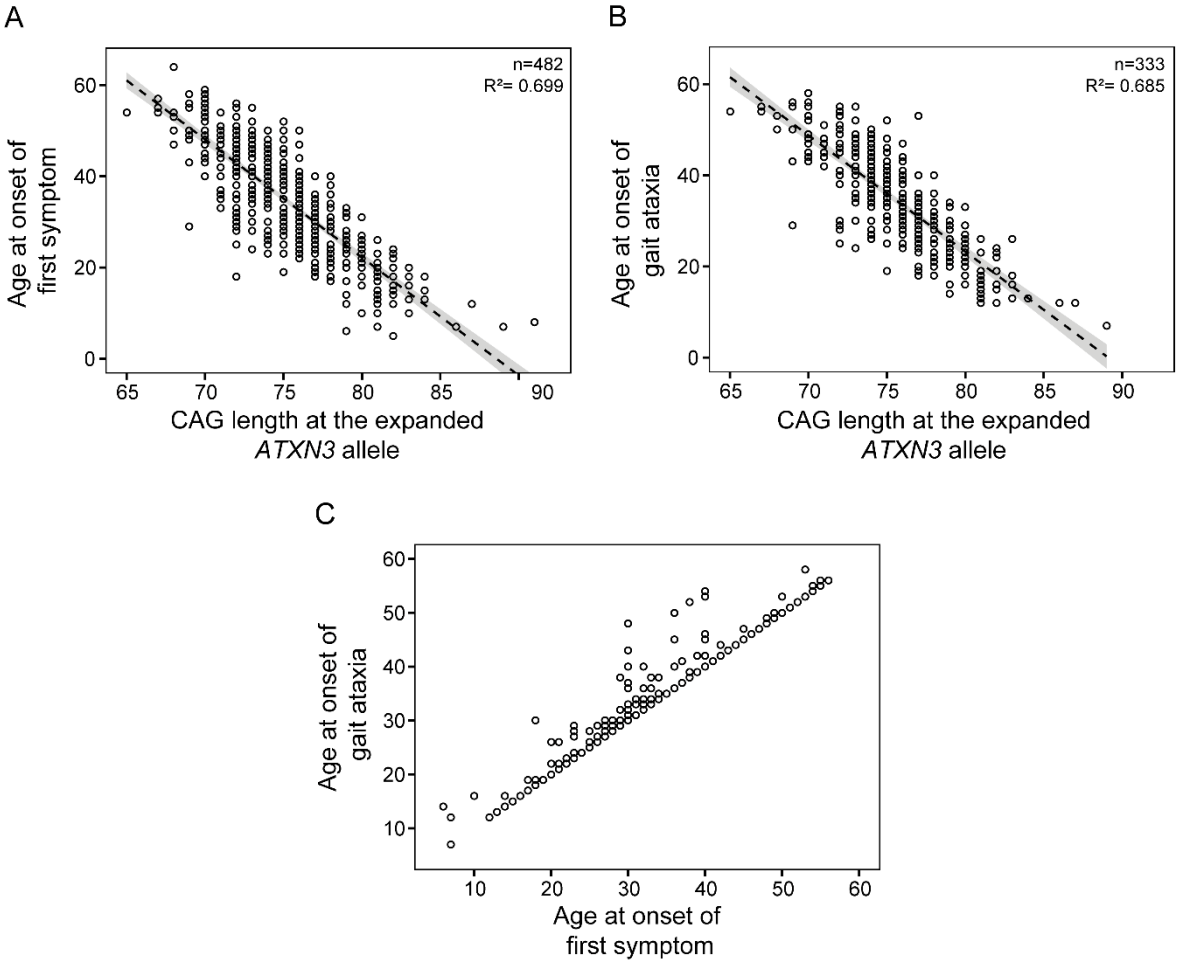
[a] p=0.797, Pearson Chi-Square test.

Supplementary File 4: *Apolipoprotein E* (*apoE*) genotypic frequencies among studies reporting on SCA3/MJD patients.

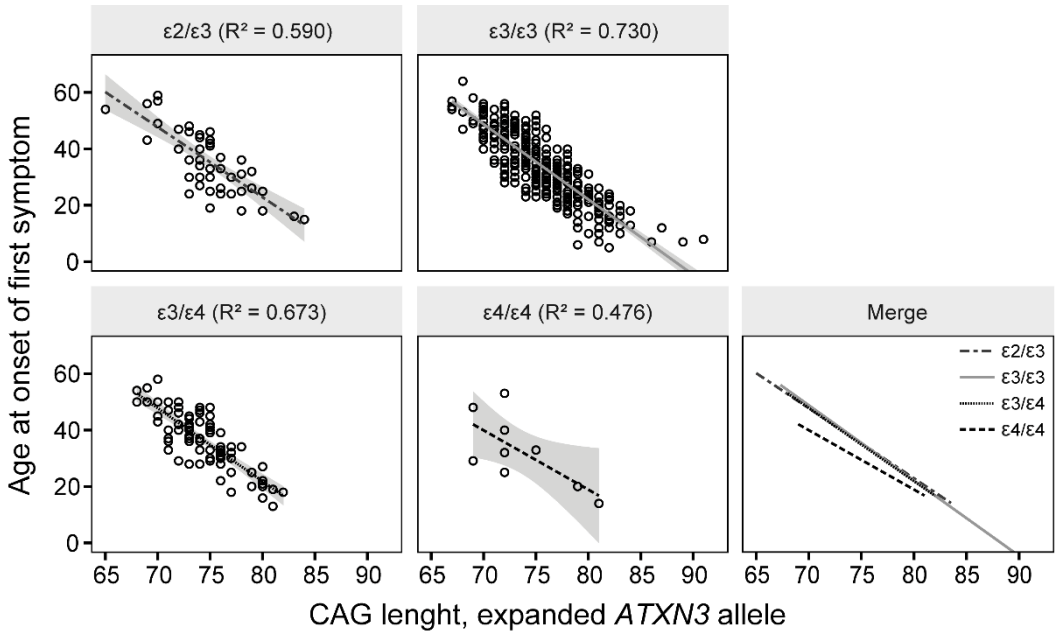
	***apoE* genotype, n (%) [a]**						
Cohort	**ε2/ ε2**	**ε2/ ε3**	**ε2/ ε4**	**ε3/ ε3**	**ε3/ ε4**	**ε4/ ε4**	**Total**
Bettencourt *et al*., 2011	0 (0.0)	20 (10.4)	3 (1.6)	134 (69.8)	33 (17.2)	2 (1.0)	192 (100.0)
Peng *et al*., 2014	3 (1.9)	14 (9.0)	1 (0.7)	103 (66.5)	34 (21.9)	0 (0.0)	155 (100.0)
Zhou *et al*., 2014	0 (0.0)	53 (13.2)	9 (2.2)	263 (65.3)	75 (18.6)	3 (0.7)	403 (100.0)
Present study	1 (0.2)	50 (10.4)	6 (1.2)	336 (69.7)	80 (16.6)	9 (1.9)	482 (100.0)
Total	4 (0.3)	137 (11.1)	19 (1.5)	836 (67.9)	222 (18.0)	14 (1.1)	1232 (100.0)

[a] p=0.163, Fisher's exact test with 10,000 Monte Carlo simulations.

Supplementary File 5: Relationship between age at onset and CAG length at the expanded *ATXN3* allele in Brazilian SCA3/MJD patients. A: Scatter plot showing the distribution of the age at onset of the first symptom according to the length of the CAG tract at the expanded *ATXN3* allele (CAGexp) in 482 SCA3/MJD patients. The dotted line represents the linear regression, and the gray area the 95% confidence interval. B: Same as in A, but for the relationship between age at onset of gait ataxia and CAGexp in 333 SCA3/MJD patients. C: Correlation between the ages at onset of the first symptom and gait ataxia among 333 SCA3/MJD patients.



Supplementary File 6: Modulation of the age at onset of the first symptom in SCA3/MJD by distinct *apolipoprotein E* (*apoE*) genotypes. Symptomatic individuals were divided in four groups according to *apoE* genotypes, and the age at onset of the first symptom was plotted against length of the CAG repeat at the expanded *ATXN3* allele. Lines represent the linear regression model, and gray areas the 95% confidence intervals. In the lower right panel, the four regression lines are shown together, and dots were omitted for clarity.



Supplementary File 7: Distribution of the age at onset of gait ataxia in South Brazilian SCA3/MJD patients (n=333), according to different *apolipoprotein E* (*apoE*) genotypes.

	***apoE* genotype**					
	ε2/ε2	**ε2/ε3**	**ε2/ε4**	**ε3/ε3**	**ε3/ε4**	**ε4/ε4**
Patients, n (%)	0 (0.0)	36 (10.8)	5 (1.5)	236 (70.9)	51 (15.3)	5 (1.5)
AO gait ataxia (AOga)						
Mean (SE)	-	35.4 (1.8)	34.2 (5.9)	34.3 (0.7)	37.3 (1.4)	25.6 (4.4)
Adjusted, mean (S.E.) [a]	-	33.7 (1.0)	-	35.1 (0.4)	35.1 (0.8)	23.3 (2.7) [b]
ATXN3 CAG length						
CAGexp, mean (S.E.) [range]	-	74.8 (0.6) [65-83]	75.0 (1.7) [72-80]	75.8 (0.2) [67-89]	74.6 (0.4) [68-82]	74.6 (2.3) [69-81]
CAGnorm, mean (S.E.) [range]	-	21.0 (0.8) [14-32]	16.4 (2.4) [14-26]	22.3 (0.3) [13-39]	22.0 (0.8) [14-37]	25.2 (2.8) [18-34]

[a] Estimated age at onset for mean CAGexp = 75.50; [b] statistically different from other *apoE* genotypes (ANCOVA, $F_{3,323}$=6.61, p<0.001); AO: age at onset; CAGexp: CAG length at the expanded *ATXN3* allele; CAGnorm: CAG length at the non-expanded *ATXN3* allele; S.E.: standard error.

Capítulo 4. A chaperona molecular DNAJB6 e a DMJ/SCA3

4.1 Manuscrito 4: “Chaperones in Polyglutamine Aggregation: Beyond the Q-Stretch”

Artigo publicado no periódico
Frontiers in Neuroscience
(DOI: *10.3389/fnins.2017.00145*)

frontiers in Neuroscience

MINI REVIEW
published: 23 March 2017
doi: 10.3389/fnins.2017.00145

Chaperones in Polyglutamine Aggregation: Beyond the Q-Stretch

E. F. E. Kuiper[1†], Eduardo P. de Mattos[1,2,3†], Laura B. Jardim[2,3,4], Harm H. Kampinga[1] and Steven Bergink[1]*

[1] Department of Cell Biology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands, [2] Programa de Pós-Graduação em Genética e Biologia Molecular, Department of Genetics, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil, [3] Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil, [4] Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

Expanded polyglutamine (polyQ) stretches in at least nine unrelated proteins lead to inherited neuronal dysfunction and degeneration. The expansion size in all diseases correlates with age at onset (AO) of disease and with polyQ protein aggregation, indicating that the expanded polyQ stretch is the main driving force for the disease onset. Interestingly, there is marked interpatient variability in expansion thresholds for a given disease. Between different polyQ diseases the repeat length vs. AO also indicates the existence of modulatory effects on aggregation of the upstream and downstream amino acid sequences flanking the Q expansion. This can be either due to intrinsic modulation of aggregation by the flanking regions, or due to differential interaction with other proteins, such as the components of the cellular protein quality control network. Indeed, several lines of evidence suggest that molecular chaperones have impact on the handling of different polyQ proteins. Here, we review factors differentially influencing polyQ aggregation: the Q-stretch itself, modulatory flanking sequences, interaction partners, cleavage of polyQ-containing proteins, and post-translational modifications, with a special focus on the role of molecular chaperones. By discussing typical examples of how these factors influence aggregation, we provide more insight on the variability of AO between different diseases as well as within the same polyQ disorder, on the molecular level.

Keywords: aggregation, Huntington's disease, Machado-Joseph disease, molecular chaperones, polyglutamine disease

OPEN ACCESS

Edited by:
Tiago Fleming Outeiro, University Medical Center Goettingen, Germany

Reviewed by:
Clevio Nobrega, University of the Algarve, Portugal
Pedro Domingos, Instituto de Tecnologia Quimica e Biológica—Universidade Nova de Lisboa, Portugal
Tatiana Rosado Rosenstock, Santa Casa de São Paulo School of Medical Sciences, Brazil

***Correspondence:**
Steven Bergink
s.bergink@umcg.nl

[†]These authors have contributed equally to this work.

Specialty section:
This article was submitted to Neurodegeneration, a section of the journal Frontiers in Neuroscience

Received: *20 January 2017*
Accepted: *08 March 2017*
Published: *23 March 2017*

Citation:
Kuiper EFE, de Mattos EP, Jardim LB, Kampinga HH and Bergink S (2017) Chaperones in Polyglutamine Aggregation: Beyond the Q-Stretch. Front. Neurosci. 11:145. doi: 10.3389/fnins.2017.00145

INTRODUCTION

Polyglutaminopathies are a family of diseases characterized by CAG trinucleotide expansions in the coding regions of at least nine unrelated genes, resulting in proteins with an abnormally long polyglutamine (polyQ) stretch, which have a high aggregation propensity. PolyQ aggregates can impede cellular protein homeostasis, loss of which is also observed in many other neurodegenerative diseases (Soto, 2003). These mutant proteins lead to one recessive inherited, X-linked spinal and bulbar muscular atrophy (SBMA), and eight dominantly inherited neuronal dysfunctions, Huntington's disease (HD), dentatorubral pallidoluysian atrophy (DRPLA), and the spinocerebellar ataxias (SCAs) type 1, 2, 3, 6, 7, and 17 (Margolis and Ross, 2001). All known polyglutaminopathies show a strong inverse correlation between expansion size and age at onset (AO) of the disease, with longer repeats significantly correlating with earlier onset of symptoms and

higher aggregation proneness of the affected protein, indicating that an expanded polyQ is tightly related to the diseases. There are two main features that are striking in the association between polyQ length and AO. First, there is marked variability between polyQ diseases in expansion thresholds that determines the pathogenicity, indicating that AO has only a partial dependence on the polyQ stretches and their absolute lengths (**Figure 1A**). Second, there is also CAG-length independent phenotypic variation within a given polyQ disease (**Figure 1B**). Both these findings imply that factors beyond the polyQ stretch are co-determining disease onset (Ranum et al., 1994; DeStefano et al., 1996; Hayes et al., 2000; Wexler et al., 2004; van de Warrenburg et al., 2005; Kaltenbach et al., 2007; Branco et al., 2008; Lessing and Bonini, 2008; Bettencourt et al., 2011; Tezenas du Montcel et al., 2014; Bečanović et al., 2015). It was hypothesized that the differential effects of distinct polyQ proteins with polyQ tracts of similar lengths could be, at least in part, due to the sequences flanking the polyQ expansion (Nozaki et al., 2001).

Here we discuss that, next to aggregation of the core polyQ stretch, which is common to all polyglutaminopathies (**Figure 2A**), the context around the cores can modulate aggregation in several ways and may be linked to differential handling of the protein quality control systems, including molecular chaperones, the ubiquitin proteasome system, and autophagy. These degradation processes, and their relationship with the chaperone system, are of importance and greatly influence the aggregation process (Rubinsztein, 2006). Certain chaperones act together with the protein degradation machineries to effectively clear aggregation-prone polypeptides, such as polyQ-containing proteins (Dekker et al., 2015). The molecular details of these downstream events are still unclear and will not be discussed here; instead we will focus on the impact of molecular chaperones on the aggregation process itself. Molecular chaperones are known to influence aggregation of polyQ proteins. This could either be directly by preventing the polyQ stretch from aggregating or via the flanking sequences. For only a few of the molecular chaperones the direct interaction with the polyQ proteins has been shown, although many chaperones are found to co-localize with polyQ inclusions (Cummings et al., 1998; Kazemi-Esfarjani and Benzer, 2000; Schmidt et al., 2002; Helmlinger et al., 2004; Bilen and Bonini, 2007; Hageman et al., 2010; Gao et al., 2011; Kakkar et al., 2014; Matilla-Dueñas et al., 2014; Reis et al., 2016; Zhao et al., 2016). However, co-localization of chaperones does not provide information on their mode of interaction and does not distinguish whether chaperones are truly interacting with the polyQ protein, or whether the presence of chaperones in the aggregates is a mere secondary effect due to a collapse of other cellular components with the inclusions. In this review, we will discuss: first, how polyQ tracts drive aggregation; second, how their flanking sequences could directly affect the aggregation proneness of the polyQ protein; and third, how polyQ proteins can be modified, changed in conformation, or fragmented, inducing aggregation (**Figure 2B**). We will not focus on the function, or loss of function, of the affected polyQ proteins, since this was so far not shown to be causative for disease, even though the native function of the protein might be important for normal cellular function. Furthermore, we will not go into the discussion on the toxicity of aggregation. For instance, it is still unclear whether the presence of aggregates contributes to SCA2 pathology (Huynh et al., 2000), even though aggregates are found in affected brain areas (Pang et al., 2002; Seidel et al., 2016). Finally, we will highlight the role of chaperones in the aggregation process and include only studies that provide insight in direct interaction of chaperones with the polyQ proteins. Rather than providing a complete overview, molecular mechanisms of typical examples will be discussed, aiming at providing general principles affecting polyQ aggregation on

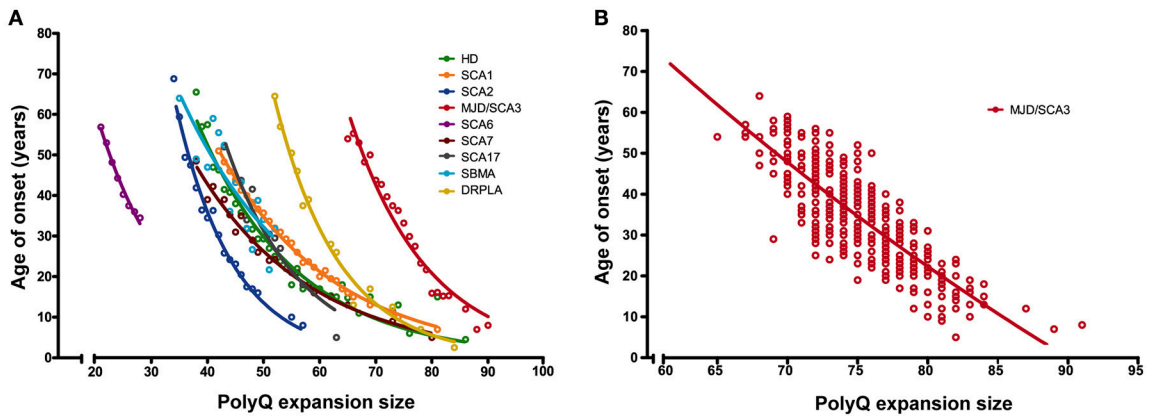


FIGURE 1 | Age of onset of disease inversely correlates with the size of the expanded polyQ tract in all known polyQ diseases. (A) Correlation between age of onset (AO) and CAG expansion size for all nine polyQ diseases identified so far. Circles depict mean AOs for a given expansion size based on multiple reported cohorts of patients. Lines represent the fitted data according to an exponential decay model. **(B)** Age of onset of disease is not completely determined by the expanded polyQ tract alone. Data on the variability of AO for a particular polyQ expansion size is shown as in **(A)** and was based on the large cohort of MJD/SCA3 patients reported by Saute and Jardim (2015). Circles represent single patients. Please refer to Supplementary File 1 for a complete list of references of the original cohort descriptions. Note that graph **(A,B)** are not drawn to the same scale.

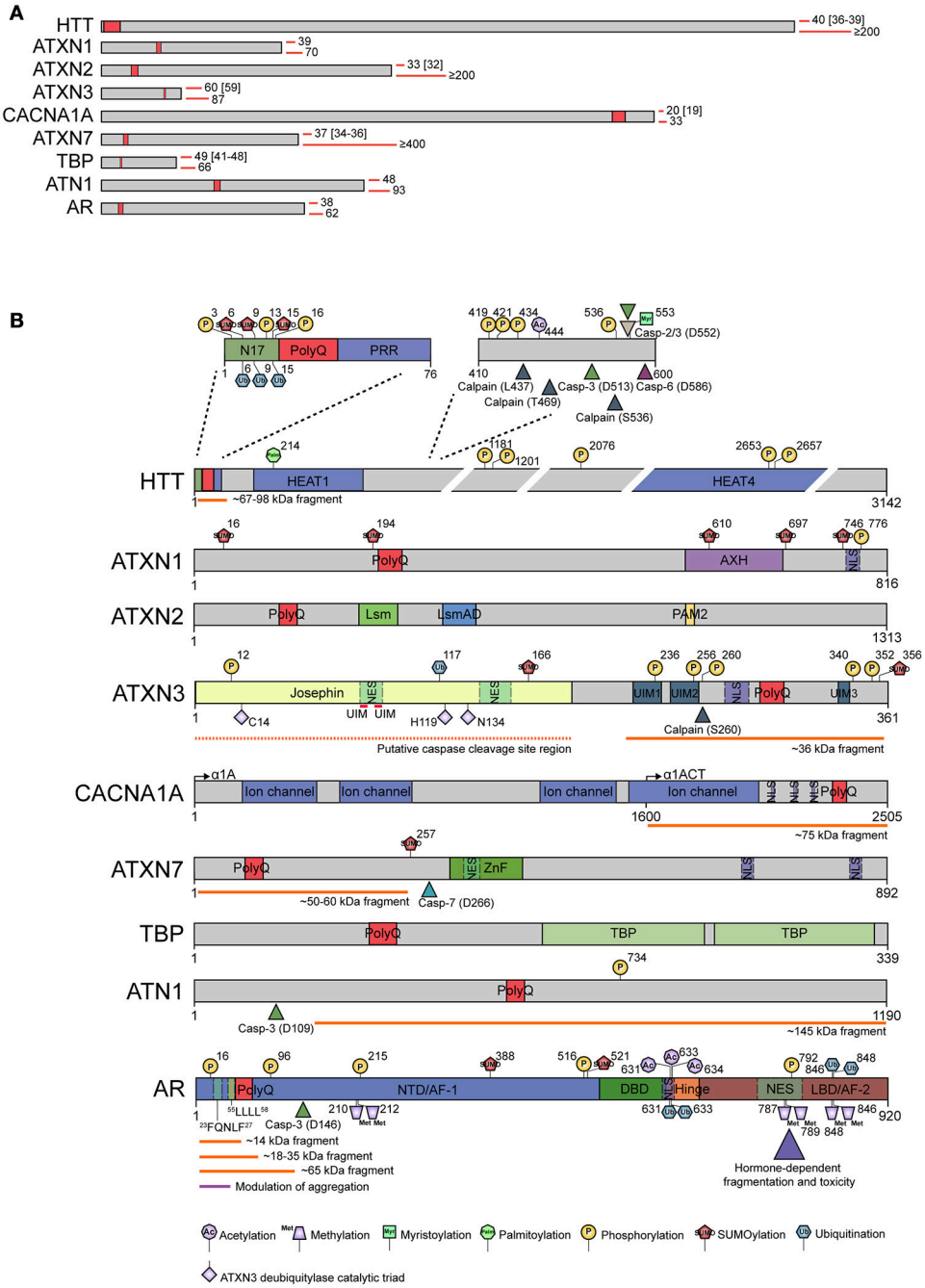


FIGURE 2 | Representation of pathogenic polyQ proteins and known modulating events associated with aggregation. (A) Schematic representation of the nine disease-related polyglutamine proteins drawn to scale. In each case, a polyQ stretch of fixed length is depicted at the approximate position (red boxes). Red bars on the right side of each protein show the smallest and largest number of glutamine repeats identified in patients of each polyQ disease to date. Numbers between brackets represent polyQ expansion sizes that have been reported to behave as incomplete penetrance alleles. **(B)** Detailed representation of all nine polyQ proteins. Domain organization is indicated. Known post-translational modifications associated with disease, caspase/calpain cleavage sites, and fragments identified are indicated. For ataxin-3, the long isoform with 3 ubiquitin-interacting motifs is shown. Residues C14, H119, and N134 depict the catalytic triad of the deubiquitylase activity of the Josephin domain. The CACNA1A locus encodes two proteins: α1A (full-length α1A) and α1ACT (C-terminal fragment of α1A) using a bicistronic mRNA with a cryptic internal ribosomal entry site. The polyQ is found in both. Many studies report a C-terminal fragment which probably represents α1ACT. For the androgen receptor, the only phosphorylation sites depicted are those with biochemical evidence of modulation of polyQ aggregation, cleavage and/or toxicity. Similarly, amino acid sequences 23FQNLF27 and 55LLLL58 highlight motifs shown to influence polyQ behavior. For simplicity, most huntingtin cleavage products are omitted and only the major N-terminal polyQ containing fragment is indicated. Amino acid numbering is based on Uniprot accession numbers P42858 (HTT), P54253 (ATXN1), Q99700 (ATXN2), P54252 (ATXN3), O00555 (CACNA1A), O15265 (ATXN7), P20226 (TBP), P54259 (ATN1), and P10275 (AR). However, for clarity, some residues

(Continued)

FIGURE 2 | Continued
are numbered according to their original publication, which might differ from the numbering according to the reference protein sequence (due to the expanding nature of polyQ proteins). AR, androgen receptor; ATN1, atrophin-1; ATXN1, ataxin-1; ATXN2, ataxin-2; ATXN3, ataxin-3; ATXN7, ataxin-7; AXH, ataxin-1/high-mobility group box containing protein-1; CACNA1, α1A subunit of the P/Q-type or CaV2.1 voltage-gated calcium channel; Casp, caspase; DBD, DNA binding domain; HTT, huntingtin; PolyQ, polyglutamine stretch; NTD/AF-1, amino-terminal domain/ activation function-1; LBD/AF-2, ligand-binding domain/ activation function-2; NES, nuclear export signal; NLS, nuclear localization signal; HEAT, huntingtin/elongation factor 3/PR65/A subunit of protein phosphatase 2A/ lipid kinase TOR domain; PRR, proline-rich region; N17, first 17 amino acids of huntingtin; TBP, TATA-binding protein (domain); UIM, ubiquitin-interacting motif; Ub-1/Ub2, ubiquitin-binding sites; Lsm, Like RNA splicing domain Sm and Sm2; LsmAD, Like-Sm-associated domain; PAM2, poly (A)-binding protein interacting motif 2; ZnF, SCA7-like zinc finger domain. For references to specific domains or post-translational modifications, please refer to Supplementary File 1.

the molecular level that may partially explain the individual differences between patients and steer future studies.

AGGREGATION PROPERTIES OF THE POLYQ STRETCH

Aggregates formed by polyQ stretches contain identical β-strand-based cores. Already in 1994, Perutz et al. described the ability of elongated polyQ stretches to form β-sheets (Perutz et al., 1994). Like many other amyloidogenic proteins (Sawaya et al., 2007), the polyQ chains can form β-sheets that are connected through interdigitating extended side chains and contain intramolecular β-hairpins (Hoop et al., 2016). Formation of β-hairpins allows for hydrogen bonding between the stacked side chains, providing a strong interaction (Hoop et al., 2016). The β-hairpins play an important role in the aggregation process. Q-stretches with a range up to 25Q are not able to form stable β-hairpins and therefore are not able to induce aggregation, except when mutations known to enhance β-hairpin formation are introduced (Kar et al., 2011, 2013). It is hypothesized that longer polyQ stretches can form more stable intramolecular β-hairpins, providing a critical monomeric nucleus necessary for inducing aggregation (Kar et al., 2011). The high affinity of the β-sheets affects interactions between molecules and might not only do so for the same pathogenic polyQ protein, but also as a secondary effect for other endogenous polyQ containing proteins (Nóbrega et al., 2015). For example, the endogenous, non-expanded TATA-box binding protein (TBP) was found to sequester into aggregates formed by other pathogenic polyQ proteins, such as huntingtin (HTT; Perez et al., 1998; Kim et al., 2002; Matsumoto et al., 2006). Similarly, inclusions containing ataxin-2 (ATXN2), ataxin-3 (ATXN3) and TBP are observed in SCA1, SCA2, SCA3, and DRPLA (Uchihara et al., 2001). Whether these secondary co-aggregating events contribute to disease is currently not clear (Kampinga and Bergink, 2016).

The crucial role for the formation of β-hairpins in the aggregation process is nicely illustrated by findings on missense CAG to CAT mutations. These mutations, coding for histidine, were found in the CAG-repeat in ATXN1, leading to insertion of one or more other amino acids and interrupting the Q-stretch (Sobczak and Krzyzosiak, 2004; Jayaraman et al., 2009; Menon et al., 2013). The AO is in these cases inversely correlated to the longer uninterrupted CAG stretch which, rather than a specific interruption pattern, dictates also the aggregation propensity *in vitro* (Menon et al., 2013). The structure of the polyQ-stretches is not changed because of the histidine-interruptions but the polyQ aggregation rates are decreased due to the Q-length dependent ability of the protein to form a critical nucleus to initiate aggregation (Jayaraman et al., 2009; Menon et al., 2013).

From all the different intracellular chaperones, so far the only ones described that could act on the β-sheets or β-hairpins formed by the Q-stretch are DNAJB6 and its closest homolog DNAJB8, two members of the DNAJ family of Hsp70 co-chaperones. In a screen for suppressors of aggregation of huntingtin (HTT-119Q) both DNAJB6 and DNAJB8 were superior suppressors of aggregation with a specificity for the polyQ tract, since they were similarly effective in the suppression of aggregation of HTT, ATXN3, the androgen receptor (AR), and polyQ alone (Hageman et al., 2010; Månsson et al., 2013). These DNAJ chaperones have a unique region containing 18 residues of the polar hydroxyl group amino acids serine and threonine, that is exposed on one face of the DNAJB6 monomer where it is predicted to interact with the hydrogen bonds in the polyQ β-hairpins (Månsson et al., 2013; Kakkar et al., 2016).

AGGREGATION INITIATION BY FLANKING DOMAINS IN POLYQ-CONTAINING PROTEINS

A longer Q-stretch not only has a higher aggregation propensity, but also affects the conformation of other parts of the protein. This can cause exposure of other regions in the proteins that have aggregation-prone properties by themselves (Ellisdon et al., 2006; Kelley et al., 2009; Tam et al., 2009). The intrinsic aggregation propensity leads to a two-stage aggregation mechanism (Ellisdon et al., 2006) in which the first aggregation step is actually thought to be a nucleation step of the non-polyQ-containing flanking domains. The formed nucleus can speed up the aggregation of the polyQ-stretch, which is then the second aggregation step. Aggregation of the flanking region and the polyQ stretch may enhance each other in a positive feedback loop accelerating aggregation and AO (Ellisdon et al., 2007; Saunders et al., 2011). The most striking examples of this process are known for HTT and ATXN3.

HTT is a relatively large protein with the polyQ stretch located in the first exon of the protein. The polyQ tract in HTT is flanked by a 17 amino acid long N-terminal (N17) domain and a polyproline domain on its C-terminus (Dehay and Bertolotti, 2006; Rockabrand et al., 2007; **Figure 2**). The N17 domain is highly soluble by itself and has an intrinsic tendency to collapse into an aggregation-resistant compact coil state (Thakur et al., 2009; Crick et al., 2013). When the Q-stretch is expanded,

the N17 domain undergoes a conformational change going into a more α-helical extended state (Tam et al., 2009; Thakur et al., 2009; Sivanandam et al., 2011), exposing a hydrophobic face through which self-association is induced (Kelley et al., 2009; Liebman and Meredith, 2010). Self-association provides an initial nucleus that increases the local concentration of the adjacent polyQ, promoting polyQ aggregation (Kelley et al., 2009; Liebman and Meredith, 2010; Sahoo et al., 2016). Aggregation of HTT can be prevented by modifying the hydrophobic face of the α-helix (Tam et al., 2009), confirming the important role of the N17 domain in initial aggregation. Moreover, synthetic polyQ peptides lacking the N17 domain show much slower aggregation kinetics (Månsson et al., 2013; Monsellier et al., 2015; Sahoo et al., 2016).

The exposed hydrophobic face on the N17 domain was identified as an interaction site for several chaperones amongst which the chaperonin TRiC, specifically the subunit CCT1 (Tam et al., 2006). CCT1 can suppress HTT aggregation by binding via its apical substrate-binding domains to the hydrophobic motifs in the N17, preventing the initial step of aggregation (Spiess et al., 2006; Tam et al., 2009; Shahmoradian et al., 2013; Sahl et al., 2015). The constitutively expressed Hsp70 (Hsc70/HSPA8) was found to co-localize, like many other Hsp70s including the prokaryotic DnaK and yeast Ssa1 (Jana et al., 2000; Muchowski et al., 2000; Novoselova et al., 2005; Tam et al., 2006), and interact with the N17 domain of HTT via its client protein binding domain (Monsellier et al., 2015). HSPA8 is not able to delay aggregation of a Q-stretch lacking flanking sequences (Månsson et al., 2013) and acts, similar to CCT1, by disrupting the interaction between N17 domains of HTT, slowing down aggregate formation (Monsellier et al., 2015).

Another example of a polyQ protein that undergoes a similar two-stage aggregation mechanism is ATXN3, causative for SCA3. ATXN3 is involved in proteostasis by editing specific ubiquitin sidechains that are targeting proteins to the proteasome (Kuhlbrodt et al., 2011). ATXN3 has an unstructured C-terminus containing the polyQ expansion and multiple ubiquitin interacting motifs (UIMs), and an N-terminus containing the Josephin domain (JD), which is a structured monomeric domain that folds into a globular conformation (Chow et al., 2004; Masino et al., 2004; **Figure 2**). The JD is the catalytic domain responsible for the deubiquitinating (DUB) properties of ATXN3 and has a high α-helical content forming a groove with two additional UIMs for recognition of the polyubiquitin chains of different linkages, and positioning them for cleavage (Masino et al., 2004; Nicastro et al., 2009, 2010). Sequence motifs on the helices in the groove are functionally important for binding conjugated ubiquitin but are predicted to be highly amyloidogenic and therefore responsible for the aggregation propensity of the JD itself (Masino et al., 2011; Lupton et al., 2015). Indeed, *in vitro* the isolated JD shows fibrillogenic behavior even under physiological conditions (Masino et al., 2004, 2011; Ellisdon et al., 2006), but when ubiquitin is added, the aggregation propensity of ATXN3 is lowered (Masino et al., 2011). Expansion of the polyQ stretch influences the conformation of the JD in such a way that the molecular mobility of two α-helices is increased and the amyloidogenic motif gets more exposed (Lupton et al., 2015; Scarff et al., 2015), providing a nucleus through which the first aggregation step of ATXN3 is initiated. This can in turn accelerate aggregation of the polyQ stretch (Gales et al., 2005; Ellisdon et al., 2007). In a dedicated screen, several modifiers of ATXN3 were identified that all fell into the canonical chaperone and ubiquitin pathways (Bilen and Bonini, 2007). Amongst the chaperones was alphaB-crystallin (HSPB5), which was found to interact with the JD in the distorted ubiquitin interacting groove, possibly masking the amyloidogenic motives, and having an effect on the initial nucleation step of ATXN3 (Robertson et al., 2010).

Flanking regions can also suppress aggregation of the polyQ stretch. For example, the proline-rich flanking domain (C38) in HTT has an opposite effect compared to the N17 domain. The C38 is also highly soluble, but actually lowers the rate of aggregation (Bhattacharyya et al., 2006; Dehay and Bertolotti, 2006; Duennwald et al., 2006; Crick et al., 2013). Other polyQ-containing proteins apart from HTT, also have a proline-rich region adjacent to the Q-stretch, like TBP, AR, and ATXN2 (Kim, 2014). It is tempting to speculate that these regions confer an evolutionary benefit and co-evolved with Q stretches to modulate their aggregation.

BINDING PARTNERS THAT CAN INFLUENCE AGGREGATION

As we have now seen, the opening up of physiologically needed hydrophobic, aggregation-prone, motifs in non-polyQ-containing parts of the protein, can lead to the unwanted formation of an initial nucleus for aggregation. These motifs are normally buried or in interaction with binding partners (or substrates), like ubiquitin in the case of ATXN3, which prevents exposure of the hydrophobic regions (Masino et al., 2011). Binding partners of polyQ-containing proteins can influence the aggregation to a great extent, also for ataxin-1 (ATXN1). ATXN1 is the protein that underlies SCA1, and has a Q-stretch in the N-terminal part of the protein and an AXH domain in the C-terminus (**Figure 2**). Just like the JD in ATXN3, the AXH domain in ATXN1 has aggregation-prone properties that are needed for its normal functioning, but therefore can be detrimental in the presence of an expanded polyQ stretch (De Chiara et al., 2013a). The AXH domain is responsible for transcriptional repression, RNA-binding activity, and is necessary for interacting with other proteins, mostly transcriptional regulators. For the domain to be able to bind all its different substrates, it has a remarkable conformational plasticity (Chen et al., 2004; De Chiara et al., 2013b; Deriu et al., 2016). Moreover, the AXH domain is responsible for ATXN1 self-association. Multimerization can bring polyQ stretches together, associated with aggregation and amyloid formation (De Chiara et al., 2005b, 2013a; Lasagna-Reeves et al., 2015). *In vivo* ATXN1 forms oligomers and interestingly the interaction partner transcriptional repressor Capicua (CIC) is found in these complexes (Lam et al., 2006; Lasagna-Reeves et al., 2015). The interaction of CIC with the AXH domain of ATXN1 stabilizes toxic soluble prefibrillar oligomers of ATXN1. When CIC levels are reduced, ATXN1

forms more fibrillar oligomers that are less toxic (Lasagna-Reeves et al., 2015). Also when the AXH domain is deleted, aggregate formation is reduced (De Chiara et al., 2005a,b). There are chaperones known to prevent ATXN1 aggregation and reduce toxicity, but the exact mechanism of action of the chaperones on ATXN1 is not known (Cummings et al., 1998; Zhai et al., 2008). A possible mechanism of action could be that chaperones bind to the AXH domain of ATXN1 to prevent complex formation or to prevent CIC from binding.

CLEAVAGE/FRAGMENTATION

Fragmented polyQ proteins have been found in patients and proteolytic processing of polyQ proteins into smaller, highly aggregation-prone fragments that are more toxic than the full-length protein has been described for most polyQ diseases, HD (Mangiarini et al., 1996; Martindale et al., 1998), DRPLA (Igarashi et al., 1998; Wellington et al., 1998), SBMA (Butler et al., 1998; Kobayashi et al., 1998; Wellington et al., 1998), and SCAs (Ikeda et al., 1996; Paulson et al., 1997; Zander et al., 2001; Goti et al., 2004; Helmlinger et al., 2004; Kordasiewicz et al., 2006; Matos et al., 2016a; **Figure 2B**). However, for SCA1, SCA2, and SCA17 the evidence for the presence of fragments is limited (Matos et al., 2016a). Proteases play a key role in the generation of these polyQ fragments, and inhibition of proteases or mutation of their cleavage sites can modulate the disease AO (Ona et al., 1999; Chen et al., 2000; Graham et al., 2006; Aharony et al., 2015). Importantly, expression of these fragments containing the polyQ stretch can already give rise to aggregation and the disease phenotype (Ikeda et al., 1996), although it is still not entirely clear why the polyQ fragments display enhanced toxicity when compared to their respective full-length proteins. Cleavage may lead to changes in aggregation propensity, conformation of the protein, localization, and molecular interactions (Matos et al., 2016a). For SBMA, it has been reported that a conformational change exposing the polyQ tract is already sufficient to drive aggregation (Heine et al., 2015) and cleavage might expose the polyQ stretch in a similar way as such a conformational change does. Protein domains that would otherwise prevent, or enhance, the aggregation may be removed, exposing the Q-stretch itself for aggregation. Finally, recognition sites and binding of molecular chaperones could be changed, exemplifying once more the importance of regions outside the polyQ tract in the modulation of aggregation.

For ATXN3, a cleavage product containing the C-terminal fragment from amino acid 221 with the 71Q expansion was found in mice showing the disease phenotype, but rarely in mice not showing the phenotype (Goti et al., 2004). This polypeptide was also found in SCA3 patients (Goti et al., 2004) indicating that fragmentation of the polyQ protein ATXN3 has a strong correlation with disease. Interestingly, while full-length ATXN3 with an expanded polyQ was mostly non-aggregating, co-expression with truncated ATXN3 makes the full-length protein co-localize with the truncated version in perinuclear aggregates (Paulson et al., 1997). More putative cleavage sites in ATXN3 were identified (Haacke et al., 2006; Colomer Gould et al., 2007) and it was shown that caspases are not the sole contributors to the fragmentation of ATXN3, but also the activity of calpains, such as calpain-2, is involved (Simões et al., 2012; Hübener et al., 2013). ATXN3 cleavage and translocation to the nucleus, and thus also aggregation, can be prevented by inhibiting calpains through overexpression of calpastatin in mice (Simões et al., 2012). Conversely, knocking down calpastatin worsened aggregation (Hübener et al., 2013). These data clearly show that under non-stressed conditions *in vivo*, fragmentation is both required and sufficient for aggregation of polyQ containing ATXN3. Similar data has been found for HTT. In almost all studies on HD, a fragment containing the first exon of HTT with the polyQ stretch is being used, since this fragment already gives rise to the HD phenotype. Toxic N-terminal fragments are found to be generated through cleavage by caspases, both in animal models and in patients (Wellington et al., 2002; Sawa et al., 2005; Graham et al., 2006; Maglione et al., 2006). Like in SCA3, fragmentation of HTT is crucial for disease progression, since the HD disease phenotype can be rescued by either mutating the cleavage site of caspase-6 in exon 13 (Graham et al., 2006), genetically ablating caspase-6 (Wong et al., 2015), or pharmacologically inhibiting caspases 1, 3, or 6 (Ona et al., 1999; Chen et al., 2000; Aharony et al., 2015). We have already discussed the ability of certain chaperones to bind to the N17 domain, which is present in the cleaved fragments.

POST-TRANSLATIONAL MODIFICATIONS

Post translational modifications (PTMs) like phosphorylation, ubiquitination, and SUMOylation, can affect the aggregation propensity of many polyQ proteins (Humbert et al., 2001; Steffan et al., 2004; Luo et al., 2005; Warby et al., 2005; Menon et al., 2012; Matos et al., 2016b; **Figure 2**). The transient nature of the PTMs usually indicates differential regulation of proteins and they can provide an interesting extra layer of modulation, possibly influencing all of the above-mentioned features of polyQ aggregation. PTMs can create alternative binding surfaces, affecting the affinity to binding partners like proteases and chaperones, and can lead to conformational changes to expose the Q-stretch. Therefore, either increased or decreased PTMs are associated with aggregation.

For most of the polyQ proteins there are several residues known to be modified (see **Figure 2B** for PTMs that impact aggregation). For ATXN3 six phosphorylation sites have been described, in the catalytic JD and in the UIMs (Fei et al., 2007; Mueller et al., 2009; Matos et al., 2016b; **Figure 2**). Phosphorylation of serine (S)340 and S352 in the third UIM did not change aggregation propensity, but shifted the localization of the aggregates from the cytoplasm to the nucleus (Mueller et al., 2009). Phosphorylation of S256 in the second UIM was shown to inhibit the formation of large insoluble polyQ complexes (Fei et al., 2007), and phosphorylation of S12 in the JD also reduces aggregation (Matos et al., 2016b). The protective effect of constitutive phosphorylation of S12 might be dependent on its close proximity to the catalytic sites in the JD, causing hindrance of the intramolecular aggregation. Phosphorylation of HTT on

S421 (Humbert et al., 2001) and S434 (Luo et al., 2005), leads to a decrease in polyQ aggregation due to a reduction in caspase-mediated cleavage thus preventing the formation of fragments (Luo et al., 2005; Warby et al., 2009). For ATXN1, S776 is the most studied phosphorylation site since it leads to reduced aggregate formation (Emamian et al., 2003; Orr, 2012). Another interesting PTM on ATXN1 is ubiquitination of K589 in the AXH domain. Mutating this residue leads to reduced degradation and, hence, more aggregation of ATXN1 (Kang et al., 2015), suggesting that PTMs may also affect the degradation of polyQ proteins resulting in a higher concentration of proteins at risk for aggregation.

Chaperone-dependent degradation of still soluble polyQ proteins could therefore be another important aspect in ameliorating disease. Interestingly, the co-chaperone CHIP (C-terminus of Hsp70-interacting protein), an E3 ligase that can interact with and modulate Hsp70 activity (Ballinger et al., 1999; Scheufler et al., 2000), has been implicated as a modulator in many polyQ diseases (Jana et al., 2005; Choi et al., 2007; Gao et al., 2011). CHIP interacts with ATXN1 via the phosphorylated S776 and the phospho-dead S776A mutation reduced this interaction. The CHIP-ATXN1 interaction is likely mediated via Hsp70, since the tetratricopeptide repeat (TPR) domain of CHIP, with which it interacts with Hsp70, is needed for the interaction and for promotion of ATXN1 degradation (Choi et al., 2007). A similar model of CHIP and Hsp70 interaction with HTT and ATXN3 was proposed, although no single modified residue was identified as a recognition site (Jana et al., 2005).

Members of DNAJ family of Hsp70 co-chaperones were also shown to play a role in the PTM dependent degradation of polyQ proteins, like in ATXN3 (Gao et al., 2011). DNAJB1 was identified to suppress aggregate formation of ATXN3 (Chai et al., 1999), but aggregation of the S256A mutant of ATXN3 could not be prevented by DNAJB1 (Fei et al., 2007), it is still unclear whether DNAJB1 has preferential affinity for phosphorylated ATXN3. Interestingly, Hsp70 can prevent S256A aggregation (Fei et al., 2007). Next to DNAJB1, DNAJB2 was found to suppress polyQ protein aggregation via two UIMs that were shown to be crucial for its interaction with K63-linked ubiquitination of HTT (Labbadia et al., 2012). Intriguingly, all the PTMs on HTT are less present in polyQ-expanded HTT, especially in the regions in the brain that are mostly affected, abolishing the possible protective effect of the modifications (Luo et al., 2005; Warby et al., 2005; Aiken et al., 2009). Currently it is unclear whether the drop in modification is causal or a consequence of aggregation.

PERSPECTIVES

The expanded polyQ stretches in the different disease-associated proteins are the determining factor of disease onset and progression in all of the polyglutaminopathies. Above a certain threshold, Q-stretches are prone to aggregate. However, more often than not, the Q-stretch and its aggregation propensities are modulated by secondary events that we categorized here; flanking regions, which have modulating capacity due to intrinsic stability issues, binding of partners (including chaperones), modification by PTMs, and cleavage of the Q-stretch. The examples of molecular interactions described, clearly indicate that polyQ protein aggregation is a multifactorial and likely multistep process that not always has to go through the same sequence of events toward aggregate formation. For example, the intrinsic fibrillogenic behavior of the JD and cleavage of ATXN3 (leading to a fragment not containing the JD) can both trigger aggregation independently. It could very well be that initial aggregation can be triggered via different mechanisms leading to secondary events that stimulate aggregation further. Thus, *in vivo* aggregation of the JD might stimulate ATXN3 cleavage and, vice versa, cleavage might destabilize the JD domain resulting in a fast forward feedback loop of aggregation. Modulating events, together with the unique expression pattern and level of each polyQ protein, could explain the variation in AO between the nine diseases.

Moreover, the modulating events acting on the flanking regions might also explain the variation of AO among patients with a similar Q length within a given polyQ disease. By combining information on Q length (CAG repeat), expression levels of the chaperone DNAJB6, which modulates Q aggregation directly, and the expression levels of chaperones that act on the disease-specific flanking regions, with the PTM and fragmentation status, perhaps a better predication of AO could be made. A strategy targeting chaperones acting on the Q-stretch with those acting on the flanking regions might provide a synergistic approach for delaying AO, benefiting individuals diagnosed with an expanded polyQ tract. There is little information on the factors influencing progression of disease after onset and it would also be of interest to know whether progression of disease is influenced by the same factors that modulate aggregation propensity. If so, these could be used as a therapeutic modality as well.

AUTHOR CONTRIBUTIONS

EK and ED compiled all the data and contributed equally to this work. EK, ED, LJ, HK, and SB gave intellectual feedback and wrote the manuscript.

ACKNOWLEDGMENTS

EK was awarded a Topmaster fellowship from the Groningen University Institute for Drug Exploration (GUIDE). ED was awarded a Science without Borders fellowship from the Brazilian Ministry of Education. LJ and ED are supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (National Counsel of Technological and Scientific Development, CNPq). SB has received a grant from the Nederlandse Organisatie voor Wetenschappelijk Onderzoek Aard-en Levenswetenschappen.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fnins.2017.00145/full#supplementary-material

REFERENCES

Aharony, I., Ehrnhoefer, D. E., Shruster, A., Qiu, X., Franciosi, S., Hayden, M. R., et al. (2015). A Huntingtin-based peptide inhibitor of caspase-6 provides protection from mutant Huntingtin-induced motor and behavioral deficits. *Hum. Mol. Genet.* 24, 2604–2614. doi: 10.1093/hmg/ddv023

Aiken, C. T., Steffan, J. S., Guerrero, C. M., Khashwji, H., Lukacsovich, T., Simmons, D., et al. (2009). Phosphorylation of threonine 3: implications for huntingtin aggregation and neurotoxicity. *J. Biol. Chem.* 284, 29427–29436. doi: 10.1074/jbc.M109.013193

Ballinger, C. A., Connell, P., Wu, Y., Hu, Z., Thompson, L. J., Yin, L. Y., et al. (1999). Identification of CHIP, a novel tetratricopeptide repeat-containing protein that interacts with heat shock proteins and negatively regulates chaperone functions. *Mol. Cell. Biol.* 19, 4535–4545. doi: 10.1128/MCB.19.6.4535

Bečanović, K., Nørremølle, A., Neal, S. J., Kay, C., Collins, J. A., Arenillas, D., et al. (2015). A SNP in the HTT promoter alters NF-κB binding and is a bidirectional genetic modifier of Huntington disease. *Nat. Neurosci.* 18, 807–816. doi: 10.1038/nn.4014

Bettencourt, C., Raposo, M., Kazachkova, N., Cymbron, T., Santos, C., Kay, T., et al. (2011). The APOE ε2 allele increases the risk of earlier age at onset in Machado-Joseph disease. *Arch. Neurol.* 68, 1580–1583. doi: 10.1001/archneurol.2011.636

Bhattacharyya, A., Thakur, A. K., Chellgren, V. M., Thiagarajan, G., Williams, A. D., Chellgren, B. W., et al. (2006). Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* 355, 524–535. doi: 10.1016/j.jmb.2005.10.053

Bilen, J., and Bonini, N. M. (2007). Genome-wide screen for modifiers of ataxin-3 neurodegeneration in Drosophila. *PLoS Genet.* 3:e177. doi: 10.1371/journal.pgen.0030177

Branco, J., Al-Ramahi, I., Ukani, L., Pérez, A. M., Fernandez-Funez, P., Rincón-Limas, D., et al. (2008). Comparative analysis of genetic modifiers in Drosophila points to common and distinct mechanisms of pathogenesis among polyglutamine diseases. *Hum. Mol. Genet.* 17, 376–390. doi: 10.1093/hmg/ddm315

Butler, R., Leigh, P. N., McPhaul, M. J., and Gallo, J. M. (1998). Truncated forms of the androgen receptor are associated with polyglutamine expansion in X-linked spinal and bulbar muscular atrophy. *Hum. Mol. Genet.* 7, 121–127. doi: 10.1093/hmg/7.1.121

Chai, Y., Koppenhafer, S. L., Bonini, N. M., and Paulson, H. L. (1999). Analysis of the role of heat shock protein (Hsp) molecular chaperones in polyglutamine disease. *J. Neurosci.* 19, 10338–10347.

Chen, M., Ona, V. O., Li, M., Ferrante, R. J., Fink, K. B., Zhu, S., et al. (2000). Minocycline inhibits caspase-1 and caspase-3 expression and delays mortality in a transgenic mouse model of Huntington disease. *Nat. Med.* 6, 797–801. doi: 10.1038/80538

Chen, Y. W., Allen, M. D., Veprintsev, D. B., Löwe, J., and Bycroft, M. (2004). The structure of the AXH domain of spinocerebellar ataxin-1. *J. Biol. Chem.* 279, 3758–3765. doi: 10.1074/jbc.M309817200

Choi, J. Y., Ryu, J. H., Kim, H. S., Park, S. G., Bae, K. H., Kang, S., et al. (2007). Co-chaperone CHIP promotes aggregation of ataxin-1. *Mol. Cell. Neurosci.* 34, 69–79. doi: 10.1016/j.mcn.2006.10.002

Chow, M. K. M., MacKay, J. P., Whisstock, J. C., Scanlon, M. J., and Bottomley, S. P. (2004). Structural and functional analysis of the Josephin domain of the polyglutamine protein ataxin-3. *Biochem. Biophys. Res. Commun.* 322, 387–394. doi: 10.1016/j.bbrc.2004.07.131

Colomer Gould, V. F., Goti, D., Pearce, D., Gonzalez, G. A., Gao, H., Bermudez de Leon, M., et al. (2007). A mutant ataxin-3 fragment results from processing at a site N-terminal to amino acid 190 in brain of Machado-Joseph disease-like transgenic mice. *Neurobiol. Dis.* 27, 362–369. doi: 10.1016/j.nbd.2007.06.005

Crick, S. L., Ruff, K. M., Garai, K., Frieden, C., and Pappu, R., V (2013). Unmasking the roles of N- and C-terminal flanking sequences from exon 1 of huntingtin as modulators of polyglutamine aggregation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20075–20080. doi: 10.1073/pnas.1320626110

Cummings, C. J., Mancini, M. A., Antalffy, B., DeFranco, D. B., Orr, H. T., and Zoghbi, H. Y. (1998). Chaperone suppression of aggregation and altered subcellular proteasome localization imply protein misfolding in SCA1. *Nat. Genet.* 19, 148–154. doi: 10.1038/502

De Chiara, C., Menon, R. P., Adinolfi, S., De Boer, J., Ktistaki, E., Kelly, G., et al. (2005a). The AXH domain adopts alternative folds: the solution structure of HBP1 AXH. *Structure* 13, 743–753. doi: 10.1016/j.str.2005.02.016

De Chiara, C., Menon, R. P., Dal Piaz, F., Calder, L., and Pastore, A. (2005b). Polyglutamine is not all: the functional role of the AXH domain in the ataxin-1 protein. *J. Mol. Biol.* 354, 883–893. doi: 10.1016/j.jmb.2005.09.083

De Chiara, C., Menon, R. P., Kelly, G., and Pastore, A. (2013a). Protein-protein interactions as a strategy towards protein-specific drug design: the example of ataxin-1. *PLoS ONE* 8:e76456. doi: 10.1371/journal.pone.0076456

De Chiara, C., Rees, M., Menon, R. P., Pauwels, K., Lawrence, C., Konarev, P. V., et al. (2013b). Self-assembly and conformational heterogeneity of the AXH domain of ataxin-1: an unusual example of a chameleon fold. *Biophys. J.* 104, 1304–1313. doi: 10.1016/j.bpj.2013.01.048

Dehay, B., and Bertolotti, A. (2006). Critical role of the proline-rich region in Huntingtin for aggregation and cytotoxicity in yeast. *J. Biol. Chem.* 281, 35608–35615. doi: 10.1074/jbc.M605558200

Dekker, S. L., Kampinga, H. H., and Bergink, S. (2015). DNAJs: more than substrate delivery to HSPA. *Front. Mol. Biosci.* 2:35. doi: 10.3389/fmolb.2015.00035

Deriu, M. A., Grasso, G., Tuszynski, J. A., Massai, D., Gallo, D., Morbiducci, U., et al. (2016). Characterization of the AXH domain of Ataxin-1 using enhanced sampling and functional mode analysis. *Proteins* 84, 666–673. doi: 10.1002/prot.25017

DeStefano, A. L., Cupples, L. A., Maciel, P., Gaspar, C., Radvany, J., Dawson, D. M., et al. (1996). A familial factor independent of CAG repeat length influences age at onset of Machado-Joseph disease. *Am. J. Hum. Genet.* 59, 119–127.

Duennwald, M. L., Jagadish, S., Muchowski, P. J., and Lindquist, S. (2006). Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11045–11050. doi: 10.1073/pnas.0604547103

Ellisdon, A. M., Pearce, M. C., and Bottomley, S. P. (2007). Mechanisms of ataxin-3 misfolding and fibril formation: kinetic analysis of a disease-associated polyglutamine protein. *J. Mol. Biol.* 368, 595–605. doi: 10.1016/j.jmb.2007.02.058

Ellisdon, A. M., Thomas, B., and Bottomley, S. P. (2006). The two-stage pathway of ataxin-3 fibrillogenesis involves a polyglutamine-independent step. *J. Biol. Chem.* 281, 16888–16896. doi: 10.1074/jbc.M601470200

Emamian, E. S., Kaytor, M. D., Duvick, L. A., Zu, T., Tousey, S. K., Zoghbi, H. Y., et al. (2003). Serine 776 of ataxin-1 is critical for polyglutamine-induced disease in SCA1 transgenic mice. *Neuron* 38, 375–387. doi: 10.1016/S0896-6273(03)00258-7

Fei, E., Jia, N., Zhang, T., Ma, X., Wang, H., Liu, C., et al. (2007). Phosphorylation of ataxin-3 by glycogen synthase kinase 3β at serine 256 regulates the aggregation of ataxin-3. *Biochem. Biophys. Res. Commun.* 357, 487–492. doi: 10.1016/j.bbrc.2007.03.160

Gales, L., Cortes, L., Almeida, C., Melo, C. V., Costa, M. D. C., Maciel, P., et al. (2005). Towards a structural understanding of the fibrillization pathway in Machado-Joseph's disease: trapping early oligomers of non-expanded ataxin-3. *J. Mol. Biol.* 353, 642–654. doi: 10.1016/j.jmb.2005.08.061

Gao, X. C., Zhou, C. J., Zhou, Z. R., Zhang, Y. H., Zheng, X. M., Song, A. X., et al. (2011). Co-chaperone HSJ1a dually regulates the proteasomal degradation of ataxin-3. *PLoS ONE* 6:e19763. doi: 10.1371/journal.pone.0019763

Goti, D., Katzen, S. M., Mez, J., Kurtis, N., Kiluk, J., Ben-Haïem, L., et al. (2004). A mutant ataxin-3 putative-cleavage fragment in brains of Machado-Joseph disease patients and transgenic mice is cytotoxic above a critical concentration. *J. Neurosci.* 24, 10266–10279. doi: 10.1523/JNEUROSCI.2734-04.2004

Graham, R. K., Deng, Y., Slow, E. J., Haigh, B., Bissada, N., Lu, G., et al. (2006). Cleavage at the caspase-6 site is required for neuronal dysfunction and degeneration due to mutant huntingtin. *Cell* 125, 1179–1191. doi: 10.1016/j.cell.2006.04.026

Haacke, A., Broadley, S. A., Boteva, R., Tzvetkov, N., Hartl, F. U., and Breuer, P. (2006). Proteolytic cleavage of polyglutamine-expanded ataxin-3 is critical for aggregation and sequestration of non-expanded ataxin-3. *Hum. Mol. Genet.* 15, 555–568. doi: 10.1093/hmg/ddi472

Hageman, J., Rujano, M. A., van Waarde, M. A. W. H., Kakkar, V., Dirks, R. P., Govorukhina, N., et al. (2010). A DNAJB chaperone subfamily with HDAC-dependent activities suppresses toxic protein aggregation. *Mol. Cell.* 37, 355–369. doi: 10.1016/j.molcel.2010.01.001

Hayes, S., Turecki, G., Brisebois, K., Lopes-Cendes, I., Gaspar, C., Riess, O., et al. (2000). CAG repeat length in RAI1 is associated with age at onset variability

in spinocerebellar ataxia type 2 (SCA2). *Hum. Mol. Genet.* 9, 1753–1758. doi: 10.1093/hmg/9.12.1753

Heine, E. M., Berger, T. R., Pluciennik, A., Orr, C. R., Zboray, L., and Merry, D. E. (2015). Proteasome-mediated proteolysis of the polyglutamine-expanded androgen receptor is a late event in spinal and bulbar muscular atrophy (SBMA) pathogenesis. *J. Biol. Chem.* 290, 12572–12584. doi: 10.1074/jbc.M114.617894

Helmlinger, D., Bonnet, J., Mandel, J. L., Trottier, Y., and Devys, D. (2004). Hsp70 and Hsp40 chaperones do not modulate retinal phenotype in SCA7 mice. *J. Biol. Chem.* 279, 55969–55977. doi: 10.1074/jbc.M409062200

Hoop, C. L., Lin, H.-K., Kar, K., Magyarfalvi, G., Lamley, J. M., Boatz, J. C., et al. (2016). Huntingtin exon 1 fibrils feature an interdigitated β-hairpin–based polyglutamine core. *Proc. Natl. Acad. Sci. U.S.A.* 113, 1546–1551. doi: 10.1073/pnas.1521933113

Hübener, J., Weber, J. J., Richter, C., Honold, L., Weiss, A., Murad, F., et al. (2013). Calpain-mediated ataxin-3 cleavage in the molecular pathogenesis of spinocerebellar ataxia type 3 (SCA3). *Hum. Mol. Genet.* 22, 508–518. doi: 10.1093/hmg/dds449

Humbert, S., Bryson, E. A., Cordelie, F. P., Connors, N. C., Datta, S. R., Finkbeiner, S., et al. (2001). The IGF-1/Akt pathway is neuroprotective in Huntington's disease and involves huntingtin phosphorylation by Akt. *Dev. Cell.* 2, 831–837. doi: 10.1016/S1534-5807(02)00188-0

Huynh, D. P., Figueroa, K., Hoang, N., and Pulst, S. M. (2000). Nuclear localization or inclusion body formation of ataxin-2 are not necessary for SCA2 pathogenesis in mouse or human. *Nat. Genet.* 26, 44–50. doi: 10.1038/79162

Igarashi, S., Koide, R., Shimohata, T., Yamada, M., Hayashi, Y., Takano, H., et al. (1998). Suppression of aggregate formation and apoptosis by transglutaminase inhibitors in cells expressing truncated DRPLA protein with an expanded polyglutamine stretch. *Nat. Genet.* 18, 111–117. doi: 10.1038/ng0298-111

Ikeda, H., Yamaguchi, M., Sugai, S., Aze, Y., Narumiya, S., and Kakizuka, A. (1996). Expanded polyglutamine in the Machado-Joseph disease protein induces cell death *in vitro* and *in vivo*. *Nat. Genet.* 13, 196–202. doi: 10.1038/ng0696-196

Jana, N. R., Dikshit, P., Goswami, A., Kotliarova, S., Murata, S., Tanaka, K., et al. (2005). Co-chaperone CHIP associates with expanded polyglutamine protein and promotes their degradation by proteasomes. *J. Biol. Chem.* 280, 11635–11640. doi: 10.1074/jbc.M412042200

Jana, N. R., Tanaka, M., Wang, G. H., and Nukina, N. (2000). Polyglutamine length-dependent interaction of Hsp40 and Hsp70 family chaperones with truncated N-terminal huntingtin: their role in suppression of aggregation and cellular toxicity. *Hum. Mol. Genet.* 9, 2009–2018. doi: 10.1093/hmg/9.13.2009

Jayaraman, M., Kodali, R., and Wetzel, R. (2009). The impact of ataxin-1-like histidine insertions on polyglutamine aggregation. *Protein Eng. Des. Sel.* 22, 469–478. doi: 10.1093/protein/gzp023

Kakkar, V., Månsson, C., de Mattos, E. P., Bergink, S., van der Zwaag, M., van Waarde, M. A. W. H., et al. (2016). The S/T-rich motif in the DNAJB6 chaperone delays polyglutamine aggregation and the onset of disease in a mouse model. *Mol. Cell.* 62, 272–283. doi: 10.1016/j.molcel.2016.03.017

Kakkar, V., Meister-Broekema, M., Minoia, M., Carra, S., and Kampinga, H. H. (2014). Barcoding heat shock proteins to human diseases: looking beyond the heat shock response. *Dis. Model. Mech.* 7, 421–434. doi: 10.1242/dmm.014563

Kaltenbach, L. S., Romero, E., Becklin, R. R., Chettier, R., Bell, R., Phansalkar, A., et al. (2007). Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLoS Genet.* 3:e82. doi: 10.1371/journal.pgen.0030082

Kampinga, H. H., and Bergink, S. (2016). Heat shock proteins as potential targets for protective strategies in neurodegeneration. *Lancet Neurol.* 15, 748–759. doi: 10.1016/S1474-4422(16)00099-5

Kang, A. R., Park, S. H., Lee, S., Choi, D. Y., Kim, K. P., Song, H. K., et al. (2015). A key lysine residue in the AXH domain of ataxin-1 is essential for its ubiquitylation. *Biochim. Biophys. Acta* 1854, 356–364. doi: 10.1016/j.bbapap.2015.01.012

Kar, K., Hoop, C. L., Drombosky, K. W., Baker, M. A., Kodali, R., Arduini, I., et al. (2013). β-Hairpin-mediated nucleation of polyglutamine amyloid formation. *J. Mol. Biol.* 425, 1183–1197. doi: 10.1016/j.jmb.2013.01.016

Kar, K., Jayaraman, M., Sahoo, B., Kodali, R., and Wetzel, R. (2011). Critical nucleus size for disease-related polyglutamine aggregation is repeat-length dependent. *Nat. Struct. Mol. Biol.* 18, 328–336. doi: 10.1038/nsmb.1992

Kazemi-Esfarjani, P., and Benzer, S. (2000). Genetic suppression of polyglutamine toxicity in Drosophila. *Science* 287, 1837–1840. doi: 10.1126/science.287.5459.1837

Kelley, N. W., Huang, X., Tam, S., Spiess, C., Frydman, J., and Pande, V. S. (2009). The predicted structure of the headpiece of the huntingtin protein and its implications on huntingtin aggregation. *J. Mol. Biol.* 388, 919–927. doi: 10.1016/j.jmb.2009.01.032

Kim, M. (2014). Pathogenic polyglutamine expansion length correlates with polarity of the flanking sequences. *Mol. Neurodegener.* 9:45. doi: 10.1186/1750-1326-9-45

Kim, S., Nollen, E. A. A., Kitagawa, K., Bindokas, V. P., and Morimoto, R. I. (2002). Polyglutamine protein aggregates are dynamic. *Nat. Cell Biol.* 4, 826–831. doi: 10.1038/ncb863

Kobayashi, Y., Miwa, S., Merry, D. E., Kume, A., Mei, L., Doyu, M., et al. (1998). Caspase-3 cleaves the expanded androgen receptor protein of spinal and bulbar muscular atrophy in a polyglutamine repeat length-dependent manner. *Biochem. Biophys. Res. Commun.* 252, 145–150. doi: 10.1006/bbrc.1998.9624

Kordasiewicz, H. B., Thompson, R. M., Clark, H. B., and Gomez, C. M. (2006). C-termini of P/Q-type Ca2+ channel α1A subunits translocate to nuclei and promote polyglutamine-mediated toxicity. *Hum. Mol. Genet.* 15, 1587–1599. doi: 10.1093/hmg/ddl080

Kuhlbrodt, K., Janiesch, P. C., Kevei, É., Segref, A., Barikbin, R., and Hoppe, T. (2011). The Machado-Joseph disease deubiquitylase ATX-3 couples longevity and proteostasis. *Nat. Cell Biol.* 13, 273–281. doi: 10.1038/ncb2200

Labbadia, J., Novoselov, S. S., Bett, J. S., Weiss, A., Paganetti, P., Bates, G. P., et al. (2012). Suppression of protein aggregation by chaperone modification of high molecular weight complexes. *Brain* 135, 1180–1196. doi: 10.1093/brain/aws022

Lam, Y. C., Bowman, A. B., Jafar-Nejad, P., Lim, J., Richman, R., Fryer, J. D., et al. (2006). ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell* 127, 1335–1347. doi: 10.1016/j.cell.2006.11.038

Lasagna-Reeves, C. A., Rousseaux, M. W. C., Guerrero-Munoz, M. J., Park, J., Jafar-Nejad, P., Richman, R., et al. (2015). A native interactor scaffolds and stabilizes toxic Ataxin-1 oligomers in SCA1. *Elife* 4:e07558. doi: 10.7554/eLife.07558

Lessing, D., and Bonini, N. M. (2008). Polyglutamine genes interact to modulate the severity and progression of neurodegeneration in Drosophila. *PLoS Biol.* 6:e29. doi: 10.1371/journal.pbio.0060029

Liebman, S. W., and Meredith, S. C. (2010). Protein folding: sticky N17 speeds huntingtin pile-up. *Nat. Chem. Biol.* 6, 7–8. doi: 10.1038/nchembio.279

Luo, S., Vacher, C., Davies, J. E., and Rubinsztein, D. C. (2005). Cdk5 phosphorylation of huntingtin reduces its cleavage by caspases: implications for mutant huntingtin toxicity. *J. Cell Biol.* 169, 647–656. doi: 10.1083/jcb.200412071

Lupton, C. J., Steer, D. L., Bottomley, S. P., Hughes, V. A., Wintrode, P. L., and Ellisdon, A. M. (2015). Enhanced molecular mobility of ordinarily structured regions drives polyglutamine disease. *J. Biol. Chem.* 290, 24190–24200. doi: 10.1074/jbc.M115.659532

Maglione, V., Cannella, M., Gradini, R., Cislaghi, G., and Squitieri, F. (2006). Huntingtin fragmentation and increased caspase 3, 8 and 9 activities in lymphoblasts with heterozygous and homozygous Huntington's disease mutation. *Mech. Ageing Dev.* 127, 213–216. doi: 10.1016/j.mad.2005.09.011

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., et al. (1996). Exon I of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 87, 493–506. doi: 10.1016/S0092-8674(00)81369-0

Månsson, C., Kakkar, V., Monsellier, E., Sourigues, Y., Härmark, J., Kampinga, H. H., et al. (2013). DNAJB6 is a peptide-binding chaperone which can suppress amyloid fibrillation of polyglutamine peptides at substoichiometric molar ratios. *Cell Stress Chaperones* 19, 227–239. doi: 10.1007/s12192-013-0448-5

Margolis, R. L., and Ross, C. A. (2001). Expansion explosion: new clues to the pathogenesis of repeat expansion neurodegenerative diseases. *Trends Mol. Med.* 7, 479–482. doi: 10.1016/S1471-4914(01)02179-7

Martindale, D., Hackam, A., Wieczorek, A., Ellerby, L., Wellington, C., McCutcheon, K., et al. (1998). Length of huntingtin and its polyglutamine tract influences localization and frequency of intracellular aggregates. *Nat. Genet.* 18, 150–154. doi: 10.1038/ng0298-150

Masino, L., Nicastro, G., Calder, L., Vendruscolo, M., and Pastore, A. (2011). Functional interactions as a survival strategy against abnormal aggregation. *FASEB J.* 25, 45–54. doi: 10.1096/fj.10-161208

Masino, L., Nicastro, G., Menon, R. P., Dal Piaz, F., Calder, L., and Pastore, A. (2004). Characterization of the structure and the amyloidogenic properties of

the Josephin domain of the polyglutamine-containing protein ataxin-3. *J. Mol. Biol.* 344, 1021–1035. doi: 10.1016/j.jmb.2004.09.065

Matilla-Dueñas, A., Ashizawa, T., Brice, A., Magri, S., McFarland, K. N., Pandolfo, M., et al. (2014). Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. *Cerebellum* 13, 269–302. doi: 10.1007/s12311-013-0539-y

Matos, C. A., de Almeida, L. P., and Nóbrega, C. (2016a). Proteolytic cleavage of polyglutamine disease-causing proteins: revisiting the toxic fragment hypothesis. *Curr. Pharm. Des.* 22. doi: 10.2174/1381612822666161227121912. [Epub ahead of print].

Matos, C. A., Nóbrega, C., Louros, S. R., Almeida, B., Ferreiro, E., Valero, J., et al. (2016b). Ataxin-3 phosphorylation decreases neuronal defects in spinocerebellar ataxia type 3 models. *J. Cell Biol.* 212, 465–480. doi: 10.1083/jcb.201506025

Matsumoto, G., Kim, S., and Morimoto, R. I. (2006). Huntingtin and mutant SOD1 form aggregate structures with distinct molecular properties in human cells. *J. Biol. Chem.* 281, 4477–4485. doi: 10.1074/jbc.M509201200

Menon, R. P., Nethisinghe, S., Faggiano, S., Vannocci, T., Rezaei, H., Pemble, S., et al. (2013). The role of interruptions in polyQ in the pathology of SCA1. *PLoS Genet.* 9:e1003648. doi: 10.1371/journal.pgen.1003648

Menon, R. P., Soong, D., de Chiara, C., Holt, M. R., Anilkumar, N., and Pastore, A. (2012). The importance of serine 776 in Ataxin-1 partner selection: a FRET analysis. *Sci. Rep.* 2:919. doi: 10.1038/srep00919

Monsellier, E., Redeker, V., Ruiz-Arlandis, G., Bousset, L., and Melki, R. (2015). Molecular interaction between the chaperone Hsc70 and the N-terminal flank of huntingtin exon 1 modulates aggregation. *J. Biol. Chem.* 290, 2560–2576. doi: 10.1074/jbc.M114.603332

Muchowski, P. J., Schaffar, G., Sittler, A., Wanker, E. E., Hayer-Hartl, M. K., and Hartl, F. U. (2000). Hsp70 and hsp40 chaperones can inhibit self-assembly of polyglutamine proteins into amyloid-like fibrils. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7841–7846. doi: 10.1073/pnas.140202897

Mueller, T., Breuer, P., Schmitt, I., Walter, J., Evert, B. O., and Wüllner, U. (2009). CK2-dependent phosphorylation determines cellular localization and stability of ataxin-3. *Hum. Mol. Genet.* 18, 3334–3343. doi: 10.1093/hmg/ddp274

Nicastro, G., Masino, L., Esposito, V., Menon, R. P., De Simone, A., Fraternali, F., et al. (2009). Josephin domain of ataxin-3 contains two distinct ubiquitin-binding sites. *Biopolymers* 91, 1203–1214. doi: 10.1002/bip.21210

Nicastro, G., Todi, S. V., Karaca, E., Bonvin, A. M. J. J., Paulson, H. L., and Pastore, A. (2010). Understanding the role of the josephin domain in the polyub binding and cleavage properties of ataxin-3. *PLoS ONE* 5:e12430. doi: 10.1371/journal.pone.0012430

Nóbrega, C., Carmo-Silva, S., Albuquerque, D., Vasconcelos-Ferreira, A., Vijayakumar, U. G., Mendonça, L., et al. (2015). Re-establishing ataxin-2 downregulates translation of mutant ataxin-3 and alleviates Machado-Joseph disease. *Brain* 138, 3537–3554. doi: 10.1093/brain/awv298

Novoselova, T. V., Margulis, B. A., Novoselov, S. S., Sapozhnikov, A. M., Van Der Spuy, J., Cheetham, M. E., et al. (2005). Treatment with extracellular HSP70/HSC70 protein can reduce polyglutamine toxicity and aggregation. *J. Neurochem.* 94, 597–606. doi: 10.1111/j.1471-4159.2005.03119.x

Nozaki, K., Onodera, O., Takano, H., and Tsuji, S. (2001). Amino acid sequences flanking polyglutamine stretches influence their potential for aggregate formation. *Neuroreport* 12, 3357–3364. doi: 10.1097/00001756-200110290-00042

Ona, V. O., Li, M., Vonsattel, J. P., Andrews, L. J., Khan, S. Q., Chung, W. M., et al. (1999). Inhibition of caspase-1 slows disease progression in a mouse model of Huntington's disease. *Nature* 399, 263–267. doi: 10.1038/20446

Orr, H. T. (2012). SCA1-phosphorylation, a regulator of ataxin-1 function and pathogenesis. *Prog. Neurobiol.* 99, 179–185. doi: 10.1016/j.pneurobio.2012.04.003

Pang, J. T., Giunti, P., Chamberlain, S., An, S. F., Vitaliani, R., Scaravilli, T., et al. (2002). Neuronal intranuclear inclusions in SCA2: a genetic, morphological and immunohistochemical study of two cases. *Brain* 125, 656–663. doi: 10.1093/brain/awf060

Paulson, H. L., Perez, M. K., Trottier, Y., Trojanowski, J. Q., Subramony, S. H., Das, S. S., et al. (1997). Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron* 19, 333–344. doi: 10.1016/S0896-6273(00)80943-5

Perez, M. K., Paulson, H. L., Pendse, S. J., Saionz, S. J., Bonini, N. M., and Pittman, R. N. (1998). Recruitment and the role of nuclear localization in polyglutamine-mediated aggregation. *J. Cell Biol.* 143, 1457–1470. doi: 10.1083/jcb.143.6.1457

Perutz, M., Johnson, T., and Suzuki, M. (1994). Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* 91, 5355–5358. doi: 10.1073/pnas.91.12.5355

Ranum, L. P., Chung, M. Y., Banfi, S., Bryer, A., Schut, L. J., Ramesar, R., et al. (1994). Molecular and clinical correlations in spinocerebellar ataxia type I: evidence for familial effects on the age at onset. *Am. J. Hum. Genet.* 55, 244–252.

Reis, S. D., Pinho, B. R., and Oliveira, J. M. A. (2016). Modulation of molecular chaperones in Huntington's disease and other polyglutamine disorders. *Mol. Neurobiol.* doi: 10.1007/s12035-016-0120-z. [Epub ahead of print].

Robertson, A. L., Headey, S. J., Saunders, H. M., Ecroyd, H., Scanlon, M. J., Carver, J. A., et al. (2010). Small heat-shock proteins interact with a flanking domain to suppress polyglutamine aggregation. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10424–10429. doi: 10.1073/pnas.0914773107

Rockabrand, E., Slepko, N., Pantalone, A., Nukala, V. N., Kazantsev, A., Marsh, J. L., et al. (2007). The first 17 amino acids of Huntingtin modulate its sub-cellular localization, aggregation and effects on calcium homeostasis. *Hum. Mol. Genet.* 16, 61–77. doi: 10.1093/hmg/ddl440

Rubinsztein, D. C. (2006). The roles of intracellular protein-degradation pathways in neurodegeneration. *Nature* 443, 780–786. doi: 10.1038/nature05291

Sahl, S. J., Lau, L., Vonk, W. I. M., Weiss, L. E., Frydman, J., and Moerner, W. E. (2015). Delayed emergence of subdiffraction-sized mutant huntingtin fibrils following inclusion body formation. *Q. Rev. Biophys.* 49, e2. doi: 10.1017/s0033583515000219

Sahoo, B., Arduini, I., Drombosky, K. W., Kodali, R., Sanders, L. H., Greenamyre, J. T., et al. (2016). Folding landscape of mutant huntingtin Exon1: diffusible multimers, oligomers and fibrils, and no detectable monomer. *PLoS ONE* 11:e0155747. doi: 10.1371/journal.pone.0155747

Saunders, H. M., Gilis, D., Rooman, M., Dehouck, Y., Robertson, A. L., and Bottomley, S. P. (2011). Flanking domain stability modulates the aggregation kinetics of a polyglutamine disease protein. *Protein Sci.* 20, 1675–1681. doi: 10.1002/pro.698

Saute, J. A. M., and Jardim, L. B. (2015). Machado Joseph disease: clinical and genetic aspects, and current treatment. *Expert Opin. Orphan Drugs* 3, 517–535. doi: 10.1517/21678707.2015.1025747

Sawa, A., Nagata, E., Sutcliffe, S., Dulloor, P., Cascio, M. B., Ozeki, Y., et al. (2005). Huntingtin is cleaved by caspases in the cytoplasm and translocated to the nucleus via perinuclear sites in Huntington's disease patient lymphoblasts. *Neurobiol. Dis.* 20, 267–274. doi: 10.1016/j.nbd.2005.02.013

Sawaya, M. R., Sambashivan, S., Nelson, R., Ivanova, M. I., Sievers, S. A., Apostol, M. I., et al. (2007). Atomic structures of amyloid cross-beta spines reveal varied steric zippers. *Nature* 447, 453–457. doi: 10.1038/nature05695

Scarff, C. A., Almeida, B., Fraga, J., Macedo-Ribeiro, S., Radford, S. E., and Ashcroft, A. E. (2015). Examination of ataxin-3 aggregation by structural mass spectrometry techniques: a rationale for expedited aggregation upon polyglutamine expansion. *Mol. Cell. Proteomics* 14, 1241–1253. doi: 10.1074/mcp.M114.044610

Scheufler, C., Brinker, A., Bourenkov, G., Pegoraro, S., Moroder, L., Bartunik, H., et al. (2000). Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70-Hsp90 multichaperone machine. *Cell* 101, 199–210. doi: 10.1016/S0092-8674(00)80830-2

Schmidt, T., Lindenberg, K. S., Krebs, A., Schöls, L., Laccone, F., Herms, J., et al. (2002). Protein surveillance machinery in brains with spinocerebellar ataxia type 3: redistribution and differential recruitment of 26S proteasome subunits and chaperones to neuronal intranuclear inclusions. *Ann. Neurol.* 51, 302–310. doi: 10.1002/ana.10101

Seidel, K., Siswanto, S., Fredrich, M., Bouzrou, M., den Dunnen, W. F. A., Özerden, I., et al. (2016). On the distribution of intranuclear and cytoplasmic aggregates in the brainstem of patients with Spinocerebellar Ataxia Type 2 and 3. *Brain Pathol.* doi: 10.1111/bpa.12412. [Epub ahead of print].

Shahmoradian, S. H., Galaz-Montoya, J. G., Schmid, M. F., Cong, Y., Ma, B., Spiess, C., et al. (2013). TRiC's tricks inhibit huntingtin aggregation. *Elife* 2:e00710. doi: 10.7554/eLife.00710

Simões, A. T., Gonçalves, N., Koeppen, A., Déglon, N., Kügler, S., Duarte, C. B., et al. (2012). Calpastatin-mediated inhibition of calpains in the mouse brain prevents mutant ataxin 3 proteolysis, nuclear localization

and aggregation, relieving Machado-Joseph disease. *Brain* 135, 2428–2439. doi: 10.1093/brain/aws177

Sivanandam, V. N., Jayaraman, M., Hoop, C. L., Kodali, R., Wetzel, R., and Van Der Wel, P. C. A. (2011). The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils. *J. Am. Chem. Soc.* 133, 4558–4566. doi: 10.1021/ja110715f

Sobczak, K., and Krzyzosiak, W. J. (2004). Patterns of CAG repeat interruptions in SCA1 and SCA2 genes in relation to repeat instability. *Hum. Mutat.* 24, 236–247. doi: 10.1002/humu.20075

Soto, C. (2003). Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat. Rev. Neurosci.* 4, 49–60. doi: 10.1038/nrn1007

Spiess, C., Miller, E. J., McClellan, A. J., and Frydman, J. (2006). Identification of the TRiC/CCT substrate binding sites uncovers the function of subunit diversity in eukaryotic chaperonins. *Mol. Cell.* 24, 25–37. doi: 10.1016/j.molcel.2006.09.003

Steffan, J. S., Agrawal, N., Pallos, J., Rockabrand, E., Trotman, L. C., Slepko, N., et al. (2004). SUMO modification of Huntingtin and Huntington's disease pathology. *Science* 304, 100–104. doi: 10.1126/science.1092194

Tam, S., Geller, R., Spiess, C., and Frydman, J. (2006). The chaperonin TRiC controls polyglutamine aggregation and toxicity through subunit-specific interactions. *Nat. Cell Biol.* 8, 1155–1162. doi: 10.1038/ncb1477

Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M. A., et al. (2009). The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nat. Struct. Mol. Biol.* 16, 1279–1285. doi: 10.1038/nsmb.1700

Tezenas du Montcel, S., Durr, A., Bauer, P., Figueroa, K. P., Ichikawa, Y., Brussino, A., et al. (2014). Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain* 137, 2444–2455. doi: 10.1093/brain/awu174

Thakur, A. K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V. M., Byeon, I.-J. L., et al. (2009). Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nat. Struct. Mol. Biol.* 16, 380–389. doi: 10.1038/nsmb.1570

Uchihara, T., Fujigasaki, H., Koyano, S., Nakamura, A., Yagishita, S., and Iwabuchi, K. (2001). Non-expanded polyglutamine proteins in intranuclear inclusions of hereditary ataxias - triple-labeling immunofluorescence study. *Acta Neuropathol.* 102, 149–152. doi: 10.1007/s004010100364

van de Warrenburg, B. P. C., Hendriks, H., Dürr, A., van Zuijlen, M. C. A., Stevanin, G., Camuzat, A., et al. (2005). Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort. *Ann. Neurol.* 57, 505–512. doi: 10.1002/ana.20424

Warby, S. C., Chan, E. Y., Metzler, M., Gan, L., Singaraja, R. R., Crocker, S. F., et al. (2005). Huntingtin phosphorylation on serine 421 is significantly reduced in the striatum and by polyglutamine expansion *in vivo*. *Hum. Mol. Genet.* 14, 1569–1577. doi: 10.1093/hmg/ddi165

Warby, S. C., Doty, C. N., Graham, R. K., Shively, J., Singaraja, R. R., and Hayden, M. R. (2009). Phosphorylation of huntingtin reduces the accumulation of its nuclear fragments. *Mol. Cell. Neurosci.* 40, 121–127. doi: 10.1016/j.mcn.2008.09.007

Wellington, C. L., Ellerby, L. M., Gutekunst, C.-A., Rogers, D., Warby, S., Graham, R. K., et al. (2002). Caspase cleavage of mutant huntingtin precedes neurodegeneration in Huntington's disease. *J. Cell Biol.* 22, 749–759. doi: 10.1038/35096019

Wellington, C. L., Ellerby, L. M., Hackam, A. S., Margolis, R. L., Trifiro, M. A., Singaraja, R., et al. (1998). Caspase cleavage of gene products associated with triplet expansion disorders generates truncated fragments containing the polyglutamine tract. *J. Biol. Chem.* 273, 9158–9167. doi: 10.1074/jbc.273.15.9158

Wexler, N. S., Lorimer, J., Porter, J., Gomez, F., Moskowitz, C., Shackell, E., et al. (2004). Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc. Natl. Acad. Sci. U.S.A.* 101, 3498–3503. doi: 10.1073/pnas.0308679101

Wong, B. K. Y., Ehrnhoefer, D. E., Graham, R. K., Martin, D. D. O., Ladha, S., Uribe, V., et al. (2015). Partial rescue of some features of Huntington Disease in the genetic absence of caspase-6 in YAC128 mice. *Neurobiol. Dis.* 76, 24–36. doi: 10.1016/j.nbd.2014.12.030

Zander, C., Takahashi, J., El Hachimi, K. H., Fujigasaki, H., Albanese, V., Lebre, A. S., et al. (2001). Similarities between spinocerebellar ataxia type 7 (SCA7) cell models and human brain: proteins recruited in inclusions and activation of caspase-3. *Hum. Mol. Genet.* 10, 2569–2579. doi: 10.1093/hmg/10.22.2569

Zhai, R. G., Zhang, F., Hiesinger, P. R., Cao, Y., Haueter, C. M., and Bellen, H. J. (2008). NAD synthase NMNAT acts as a chaperone to protect against neurodegeneration. *Nature* 452, 887–891. doi: 10.1038/nature06721

Zhao, X., Chen, X.-Q., Han, E., Hu, Y., Paik, P., Ding, Z., et al. (2016). TRiC subunits enhance BDNF axonal transport and rescue striatal atrophy in Huntington's disease. *Proc. Natl. Acad. Sci. U.S.A.* 113, e5655–e5664. doi: 10.1073/pnas.1603020113

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

4.2 Manuscrito 5: “*The molecular chaperone DNAJB6 as a phenotypic modulator of spinocerebellar ataxia type 3/ Machado-Joseph disease*”

Manuscrito em preparação

Title:

The molecular chaperone DNAJB6 as a phenotypic modulator of spinocerebellar ataxia type 3/ Machado-Joseph disease.

Authors:

Eduardo Preusser de Mattos [1,2,3], Vanessa Bielefeldt Leotti [4,5], Gabriel Vasata Furtado [1,2], Márcia Polese-Bonatto [2,6], Gabriela Bolzan [1], Camila Maria Oliveira [1], Anastácia Guimarães Rocha [7], Jonas Alex Morales Saute [8,9], Harm H. Kampinga [3], Maria Luiza Saraiva-Pereira [1,2,8,10], Laura Bannach Jardim [1,8,9]*

Affiliations:

[1] Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 91501-970, Brazil.

[2] Laboratório de Identificação Genética, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, 90035-903, Brazil.

[3]Department of Biomedical Sciences of Cells & Systems, Section Molecular Cell Biology, University Medical Center Groningen/ Groningen University, Groningen, 9713 AV, The Netherlands.

[4] Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 91509-900, Brazil.

[5] Programa de Pós-Graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-003, Brazil.

[6] Programa de Pós-Graduação em Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-003, Brazil.

[7] Faculdade de Medicina, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-003, Brazil.

[8] Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, 90035-903 Brazil.

[9] Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-003, Brazil.

[10] Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, 90035-000, Brazil.

*** Corresponding author:**

Laura B. Jardim, MD, PhD

Medical Genetics Service, Hospital de Clínicas de Porto Alegre

Rua Ramiro Barcelos 2350, 90035-003 Porto Alegre, Brazil

Phone: + 55 51 3359 8011 | Fax: + 55 51 3359 8010

e-mail: ljardim@hcpa.edu.br

Abstract

Background: Spinocerebellar ataxia type 3/Machado-Joseph disease (SCA3/MJD) is caused by an expanded CAG tract (CAGexp) at *ATXN3*, coding for ataxin-3 protein with an abnormally long polyglutamine (polyQ) stretch. Although no disease-modifying therapies are available for SCA3/MJD, several molecular chaperones modulate polyQ-induced phenotypes *in vitro* and *in vivo*, with the HSP70 co-chaperone DNAJB6 being the most prominent candidate.

Aim: To test whether protein levels of selected chaperones, especially DNAJB6a and DNAJB6b isoforms, correlate with age at onset (AO) and velocity of disease progression (VDP) in SCA3/MJD.

Methods: Early-, average-, and late-onset patients, corrected for CAGexp, were recruited in Brazil (n=27) and the Netherlands (n=21). Chaperone protein levels were measured in fibroblasts derived from skin biopsies and compared among AO groups within each cohort. VDP was estimated in the Brazilian cohort by clinical scales at baseline and 15.0 ± 4.7 (mean ± S.D.) months follow-up. Mixed linear models assessed the relationship between distinct chaperones and VDP.

Results: Levels of all investigated chaperones were similar among Brazilian AO groups, but HSP40/DNAJB1, DNAJB6a, HSP60, and HSPA8 levels were significantly higher in late-onset Dutch patients. On longitudinal analysis, HSP60, DNAJB6a, and DNAJB6b protein levels were significant predictors of VDP. Remarkably, both DNAJB6a and DNAJB6b levels at baseline significantly correlated with slower Scale for Assessment and Rating of Ataxia (SARA; R^2=1.000 and R^2=0.879, p<0.001, respectively) and SCA Functional Index (SCAFI; R^2=0.447, p<0.001; and R^2=0.142, p=0.016, respectively) progression.

Discussion: Molecular chaperones, particularly DNAJB6, seem to modulate SCA3/MJD phenotype, at least in some groups of patients. Our findings argue in favor of therapeutic strategies aiming at manipulating chaperone levels to delay AO and/or slow down VDP in this disease.

Keywords: molecular chaperone, DNAJB6, spinocerebellar ataxia type 3, Machado-Joseph disease, age at onset, disease progression.

1. Introduction

Spinocerebellar ataxia type 3/Machado Joseph disease (SCA3/MJD) is a neurodegenerative disorder caused by a CAG repeat expansion (CAGexp) at *ATXN3*, which is translated into an abnormally long polyglutamine (polyQ) tract in the ataxin-3 protein.[1] Expanded polyQ sequences are prone to aggregation, and display a toxic gain of function in neurons, leading to cytotoxicity and neurodegeneration.[2,3] Although SCA3/MJD is an adult-onset disease, meaning that carriers are free from symptoms for several years, polyQ-induced toxicity probably starts early in life and at preclinical stages.[4–6]

Age at onset (AO) is usually defined as the age at which the patient or a close relative noticed the first symptom, usually gait ataxia.[7,8] However, recent data showed that subclinical ataxic manifestations are present well before the onset of gait ataxia in SCA3/MJD and in other polyQ diseases.[9,10] Although CAGexp length shows a strong, inverse correlation with AO, 50% of the interpatient variability in AO is not explained by CAGexp, suggesting the existence of genetic and/or environmental modifiers of AO.[11–15] Another way of measuring disease progression is to follow the rate of symptom changes in time in longitudinal or natural history studies. Velocity of disease progression (VDP) is also likely subject to genetic modulation, either by a common or distinct set of modifiers of AO. For instance, CAGexp was related to VDP in Brazilian SCA3/MJD patients,[16,17] whereas studies performed in European cohorts not always find this association.[18,19]

Given their well-established role on protein quality control,[20] certain molecular chaperones are among the best candidates for phenotypic modulation of polyQ disorders. Mounting evidence support a prominent role for molecular chaperones on prevention of polyQ aggregation and toxicity, including observations on cell and animal models of SCA3/MJD.[21] Moreover, higher protein levels of the HSP40 family member DNAJB1, measured in skin fibroblasts, were correlated to later AO in SCA3/MJD patients.[22] Subsequent data showed that the DNAJB6 and DNAJB8 chaperones are far superior suppressors of polyQ aggregation when compared to other HSP40, HSP70, and HSP110 members, both *in vitro* and *in vivo*.[23] While DNAJB8 is detected in the testes only, DNAJB6

is ubiquitously expressed, including the central nervous system.[24] DNAJB6 not only suppresses aggregation of polyQ, but also of different amyloidogenic proteins, including α-synuclein,[25] amyloid-β,[26,27] and some prions.[28] Targeted neuronal overexpression of DNAJB6 extended life-span of a mouse model of Huntington's disease (HD), resulting in fewer polyQ inclusions and improved performance, compared to control HD mice.[29] These observations make DNAJB6 a specially attractive candidate of phenotypic modifier in SCA3/MJD and other neurodegenerative diseases caused by amyloidogenic proteins.

Our aim was to study the potential effect of molecular chaperones, with a special focus on DNAJB6, as phenotypic modulators of SCA3/MJD by testing its association with the two standard hallmarks of disease state: AO and VDP. We used the extreme phenotype sampling, an approach intended to increase chances of detecting the effect of our candidate modifier factor: DNAJB6 levels. Subjects who started the disease much earlier or much later than expected by their CAGexp were compared. We hypothesized that higher levels of molecular chaperones, and DNAJB6 in particular, would be protective and then associated to later AO and/or slower disease progression in SCA3/MJD.

2. Materials and Methods

2.1 Patient cohorts and study designs

This study had two approaches: the cross-sectional arm, used to correlate chaperone levels (risk factors) with AO (outcome); and the longitudinal arm, used to investigate associations between chaperone levels (risk factors) and VDP (outcome; **Figure 1**). Two cohorts were available. The Dutch and Brazilian cohorts comprised the cross-sectional arm, whereas longitudinal data was available for the Brazilian cohort only.

The Brazilian cohort included molecularly confirmed SCA3/MJD patients from Rio Grande do Sul[30] recruited according to an extreme phenotype sampling approach focusing on AO. Between 2015 and 2016, and after informed consent, patients underwent a baseline interview to accurately establish AO based on

medical records and impression from the individuals and their relatives. AO was defined as the age at which the patient or a close relative noticed the first symptom. A skin biopsy was collected at baseline, and disease severity was assessed by the Scale for the Assessment and Rating of Ataxia (SARA; range: 0 to 40 points; increasing values indicating worsening),[31] the SCA Functional Index (SCAFI; results given in Z scores; decreasing values indicating worsening,[32] and the Neurological Examination Score for Spinocerebellar Ataxia (NESSCA; range: 0 to 40 points; increasing values indicating worsening)[33] at baseline and at a planned 1-year follow up visit. Expected AO was calculated by a parametric survival model for each CAGexp and validated specifically for Brazilian SCA3/MJD carriers published elsewhere.[34] Individuals with AO <25th, between the 25th and 75th, or >75th percentiles of expected AO comprised early-, average-, and late-onset cases, respectively, based on CAGexp predictions.

The Dutch cohort was previously described and included SCA3/MJD patients recruited between 2005 and 2010.[22] These subjects were also assigned to similar extreme AO groups, with AO predictions calculated using the parametric survival model validated specifically for European SCA3/MJD carriers (**Table S1**).[35]

2.2 Skin biopsies and establishment of fibroblast cultures

Skin biopsies were collected from the scapular or lower back area with a 4-mm punch and local anesthesia by a trained physician and placed on sterile vials containing Dulbecco's modified Eagle's medium (DMEM; Gibco) for immediate processing. Biopsies were rinsed with sterile phosphate buffered saline (PBS) and cut into 8-10 pieces, placed into T25 culture flasks with 3 ml DMEM supplemented with 10% fetal calf serum (FCS; Gibco), and 100 units/ml penicillin and 100 µg/ml streptomycin (P/S; Invitrogen, Carlsbad, CA, USA), and incubated at 37ºC with controlled humidity and 5% CO_2. After about 1-2 weeks, biopsies were trypsinized and fibroblast cultures were established by sub-culturing.

2.3 Cell extracts and sample preparation

Samples were prepared as described previously.[22] Fibroblasts were grown to 80-90% confluency in T75 culture flasks and harvested by trypsinization. Fibroblasts from the prospective Brazilian cohort were all collected at the 7th passage. Cells from the retrospective Dutch cohort were collected in different passages (mean: 12th passage; **Table S1**). Cells were pelleted by spinning at 1.000 rpm for 5 min at 4ºC and resuspended in 100 µl RIPA buffer [25 mm Tris-HCl pH 7.6, 150 mm NaCl, 1% NP-40, 1% sodium deoxycholate, 1% sodium dodecyl sulphate (SDS), and protease inhibitors]. Cells were sonicated for 5 s on ice, and protein content was determined with the DC protein assay (Bio-Rad, Hercules, CA, USA). Western blot samples were prepared at a final concentration of 3 µg/µl in SDS–polyacrylamide gel electrophoresis (PAGE) loading buffer (250 mm Tris-Cl, 20% glycerol, 4% SDS, 0.001% bromophenol blue and 10% β-mercaptoethanol), boiled for 5 min, and stored at -20ºC until use.

2.4 Western blot analysis

Thirty micrograms of each sample were loaded onto 12% SDS-PAGE gels (TGX Stain-Free FastCast system, Bio-Rad), and ran at 230 V. Proteins were transferred to nitrocellulose membranes (Schleicher and Schuell, PerkinElmer, Waltham, MA, USA), which were subsequently blocked in 10% skimmed milk diluted in PBS with 0.1% Tween-20 (PBS-T). Membranes were probed overnight at 4°C with: mouse anti-GAPDH (1:10,000; Fitzgerald), rabbit anti-HSP40 (1:1,000; SPA400, Stressgen Biotechnologies, Victoria, Canada), rabbit anti-HSP60 (1:500, ADI-SPA-805, Enzo Life Sciences, Farmingdale, NY, USA), anti-HSPA8 (1:5,000 SPA815, Stressgen Biotechnologies), or rabbit anti-DNAJB6 (1:500, home-made). Membranes were incubated with with the respective horseradish peroxidase-conjugated secondary antibody (1:5,000; Amersham, GE Healthcare, Uppsala, Sweden), and detection was performed using the ChemiDoc Touch Imaging System (Bio-Rad). Bands were quantified with the Image Lab software v. 6.0 (Bio-Rad). Mean protein levels in patient fibroblasts were determined from two to four technical replicates, depending on sample availability, and always normalized to the protein levels of a single control fibroblast line.

2.5 Statistical analysis

Mean protein levels of each chaperone were calculated considering the number of technical replicates per fibroblast line, which varied due to sample availability. Differences in mean protein levels between AO groups were assessed with the Kruskal-Wallis test. Correlations between protein levels and age at biopsy, passage number, disease duration, and ΔAO (observed - expected AO) were calculated with Spearman rank test. VDP was estimated according to the delta between scores obtained in two moments (baseline and follow up) and to the disease duration since the start of the first symptom.[36] Correlations between protein levels and VDP were estimated by linear mixed models adjusted for disease duration at baseline and follow up using the lme4 R package. Models considered protein levels as both continuous and categorical (high/low expression groups, based on median expression values) variables. Analyses were performed using the R software v. 3.4.1, and results were considered significant when $p<0.05$.

3. Results

3.1 Clinical description of patient cohorts

Twenty-seven symptomatic SCA3/MJD subjects from South Brazil were recruited, comprising 12 early-, 5 average-, and 10 late-onset cases (**Table S2**). There were no significant differences in CAGexp length or disease duration among groups, but early-onset cases had younger AO and ages at biopsy, on average, than late-onset individuals (**Table 1**). The Dutch cohort comprised 21 SCA3/MJD patients (8 early-, 7 average-, and 6 late-onset cases; **Table S1**), whose clinical characteristics were described previously.[22] AO, age at biopsy, disease duration, and CAGexp were not statistically different among Dutch onset groups (**Table 1**). Between the two cohorts, the number of patients in each onset group, as well as AO, age at biopsy, and disease duration were not significantly different. As expected,[15] mean CAGexp length was longer in Brazilian than in Dutch patients (**Table 2**).

3.2 Chaperone levels and the age at onset phenotype

Western blot analysis showed expression of DNAJB6 as both long (DNAJB6a, ~40 kDa) and short (DNAJB6b, ~26 kDa) isoforms (**Figures 2A** and **2F**). There was substantial interpatient variability in DNAJB6 levels, but neither DNAJB6a or DNAJB6b were significantly different among AO groups in the Brazilian cohort (**Figures 2B** and **2C**). Protein levels of both isoforms neither correlated with ΔAO (observed - expected AO; **Figures 2D** and **2E**), nor with age at biopsy or disease duration among Brazilian subjects (**Table S3**). Therefore, there was no differential expression of DNAJB6 in fibroblasts of Brazilian SCA3/MJD patients with extremely discordant AOs. Among Dutch patients, there was a non-significant trend towards higher DNAJB6a levels (**Figure 2G**), but not DNAJB6b (**Figure 2H**), in late-onset individuals when compared to early- and average-onset groups. When protein levels were correlated with ΔAO in the Dutch cohort, there was a moderate, statistically significant correlation between higher DNAJB6a protein levels and AO significantly later than expected based on CAGexp length (**Figure 2I** and **Table S3**), but not for DNAJB6b (**Figure 2J**).

We have also tested whether differential expression of HSP40/DNAJB1 could be observed in the Brazilian fibroblast cohort, as reported for the Dutch cohort previously,[22] and replicated here (**Figures 3H-J**). However, there were no significant differences in HSP40/DNAJB1 levels among Brazilian onset groups (**Figures 3B**), or correlation between HSP40/DNAJB1 and ΔAO (**Figure 3C**).

Potential differences in chaperone levels between AO groups were also investigated for members of the HSP60 and HSP70 families. Similarly to HSP40, a polyclonal anti-HSP60 antibody did not reveal any significant differences in chaperone levels in the Brazilian cohort (**Figures 3D** and **3E**). In the Dutch cohort, however, HSP60 protein levels were significantly higher in late-onset patients when compared to early- and average-onset individuals (**Figures 3K** and **3L**). There was also a significant correlation between HSP60 protein levels and age at biopsy (ρ=0.475, p=0.030, **Table S3**) in the Dutch cohort. Again, there were no significant differences in levels of the constitutively-expressed HSP70 member HSPA8 and AO groups, or other clinical variables, in fibroblasts from Brazilian patients (**Figure 3F, 3G** and **Table S3**). In the Dutch cohort we replicated the findings of Zijlstra *et al*. (2010), detecting a non-significant trend towards higher

HSPA8 protein levels in late-onset patients, compared to other AO groups (**Figure 3M**).[22] However, there was a moderate, significant correlation between HSPA8 protein levels and ΔAO in the Dutch cohort (**Figure 3N**), indicating an association between higher HSPA8 protein levels and AO significantly later than expected, based on CAGexp length, in the Dutch fibroblast cohort.

Since the most conspicuous difference between the Brazilian and Dutch cohorts was the CAGexp range, we looked for associations between the protein levels of the five chaperones and CAGexp in each cohort. As expected, however, no statistically significant correlations were found (**Table S3**).

3.3 Chaperone levels and the disease progression phenotype

Nineteen individuals (8 early-, 3 average-, and 8 late-onset cases; 74.1%) from the Brazilian cohort had a follow up visit at 15.0 ± 4.7 months (mean ± standard deviation) from baseline. SCAFI was not performed in two of the 19 included individuals. Total SARA (ρ=0.787, $p<0.001$), SCAFI (ρ=-0.509, p=0.031) and NESSCA (ρ=0.746, $p<0.001$) scores correlated with disease duration at baseline. Average SARA, SCAFI, and NESSCA progressions were 1.05 (95% C.I.: 0.61 - 1.49), -0.17 (95% C.I.: -0.24 - -0.10; Z-score), and 0.90 (95% C.I.: 0.48 - 1.32) points per year, respectively.

Initially, we tested whether extreme early- and late-onset AOs (SARA and NESSCA: n=8 each; SCAFI: n=8 early-, n=7 late-onset) could be considered risk factors for differences in the outcome velocity of disease progression, as measured by SARA, SCAFI, and NESSCA scores using the disease duration model. At baseline, early- and late-onset groups had similar average SARA (early: 14.5 vs late: 17.5; p=0.485), SCAFI (early: 0.10 vs late: 0.13; p=0.923), and NESSCA (early: 18.1 vs late: 18.6; p=0.885) scores. Upon follow up, progression velocities of early- and late-onset patients were not statistically different for SARA (p=0.983), SCAFI (p=0.790), or NESSCA (p=0.424; **Figures 4A-C**).

Then, DNAJB6 protein levels in fibroblasts were tested as risk factors for differences in the outcome velocity of SARA, SCAFI, and NESSCA progression.

The risk factor DNAJB6 was treated either as a dichotomous (similar to the analysis of AO effect) or as a continuous variable. Dichotomous analysis stratified groups between subjects with lower and higher DNAJB6 levels at baseline, based on median protein levels. A non-significant trend towards higher progression velocities among individuals with lower DNAJB6a and b levels was observed for both SARA (**Figures 4D** and **G**, respectively) and SCAFI (**Figures 4E** and **H**, respectively), but not for NESSCA (**Figures 4F** and **I**).

Continuous analysis showed that lower levels of DNAJB6a at baseline significantly correlated with faster SARA and SCAFI progressions, while lower DNAJB6b levels correlated with faster SARA progression (**Table 3**). For instance, every 0.1 increase in DNAJB6a and DNAJB6b protein levels correlated with a mean (95% C.I.) decrease of 0.16 (0.04 - 0.26) and 0.40 (0.01 - 0.74) points per year in SARA, respectively. No significant associations between DNAJB6a/b levels and NESSCA progression were observed. The correlations between DNAJB6a/b and disease progression, as measured by SARA or SCAFI, can be depicted as changes in scales slopes (i.e. difference in progression from baseline to follow up for each patient) as a function of chaperone levels (**Figure 5**). Larger changes in SARA strongly correlated with lower levels of both DNAJB6a and DNAJB6b protein levels (**Figures 5A** and **B**). Similarly, worse SCAFI scores at the follow up were correlated with DNAJB6a and DNAJB6b protein levels (**Figures 5C** and **D**, respectively). For the remaining molecular chaperones measured (HSP40, HSP60, and HSPA8), the only statistically significant association observed was between faster SARA progression and lower chaperone protein levels of HSP60 (**Table S4**).

4. Discussion

The present study detected associations between protein levels of selected chaperones (especially DNAJB6) and variation in two SCA3/MJD phenotypes: AO and velocity of disease progression. Higher levels of DNAJB6a, HSP40, HSP60, and HSPA8 were observed in patients from the Dutch cohort with AO significantly later than expected for CAGexp. The same associations were not found in a Brazilian patient cohort. And higher DNAJB6a, DNAJB6b,

and HSP60 levels in fibroblasts were associated with reduced velocities of disease progression, as measured by the clinical scales related to ataxia manifestations SARA and SCAFI, at least in the cohort where this study was feasible.

We initially hypothesized that DNAJB6 could be a candidate modifier of AO in SCA3/MJD, based on the powerful anti-amyloidogenic activity of DNAJB6[23,25–27,29,37] and on previous data regarding the role of HSP40/DNAJ proteins on polyQ diseases.[21] Given the association between higher DNAJB1 protein levels and later AO than expected by CAGexp in the Dutch SCA3/MJD cohort – first described by Zijlstra and colleagues (2010) and replicated here – we argued that this could also be the case for DNAJB6, with the advantage that DNAJB6 is a stronger suppressor of polyQ aggregation and toxicity than DNAJB1.[26] Such reasoning assumes that expression patterns and protein levels measured in skin fibroblasts are similar to those in neurons, which might not be the case. Any modulatory effect of DNAJB6 in SCA3/MJD would only be meaningful if present in neurons, and further cellular studies are needed to confirm our assumptions. Nevertheless, our results support the notion of phenotypic modulation of SCA3/MJD by molecular chaperones, especially DNAJB6.

4.1. The age at onset phenotype

We observed significant associations between protein levels of molecular chaperones (HSP40/DNAJB1, DNAJB6, HSP60, and HSPA8) and AO in the Dutch fibroblast cohort only, but not in the Brazilian cohort. Although the effects of population background and interpatient variability cannot be discarded, the lack of replication between the Brazilian and Dutch cohorts could be partially explained by differences in CAGexp length between these populations. Evidence suggests that CAGexp might not have a linear correlation with AO in SCA3/MJD, with shorter CAGexp tracts exerting a milder effect on AO when compared to longer expansions.[12–15] In this sense, patients with shorter CAGexp alleles might be more susceptible to disease modulation by additional modifiers when compared to carriers of longer CAGexp alleles. Chaperone levels could fail to modulate AO in the presence of longer CAGexp tracts, as those carried by the

Brazilian cohort (**Table 2**), which probably determine AO in a stronger, more dominant manner.[12] However, our analysis did not have enough statistical power to confirm this hypothesis, and further studies in larger cohorts will be required to address this issue. Another population-specific effect, independent from the CAGexp, could also explain these differences. A recent meta-analysis revealed the existence of substantial AO modification in geographically distinct SCA3/MJD populations by yet unknown factors.[15] In face of our present results, we speculate that a unique or a couple of factors is operating so that it would explain the data obtained in the Dutch cohort. We would call it a "stronger chaperone response", or a more effective activation of HSP40/DNAJB1, DNAJB6, HSP60, and HSPA8 in the presence of protein aggregation. Given our results, we hypothesize that in the presence of moderate CAG expansions at *ATXN3*, SCA3/MJD carriers equipped with a stronger chaperone response – caused by a yet unknown factor – can delay their disease onset.

4.2. The disease progression phenotype

Higher DNAJB6a/b and HSP60 protein levels, measured at baseline in skin fibroblasts, were associated with slower rate of ataxia progression in Brazilian SCA3/MJD patients, as determined by SARA. Similarly, slower SCAFI progression was correlated with higher DNAJB6a protein levels (see **Table 3** and **Figure 5**).

Unfortunately, longitudinal data on disease progression was not available for the Dutch cohort. Dutch and Brazilian cohorts were recruited at different periods of time (2005-2010 and 2015-2016, respectively). Although this difference has not impacted on laboratory analysis performed on established culture cells, the study of the Dutch subjects was not planned to perform longitudinal evaluations. Therefore, the velocity of disease progression was studied only in the Brazilian cohort.

The results obtained by two different instruments in the Brazilian group - SARA and SCAFI - were convergent. However, the same association was not obtained with NESSCA. Although SARA, SCAFI and NESSCA correlate quite well in SCA3/MJD,[31-33] they do not measure exactly the same neurologic

outcomes. SARA is a semi-quantitative scale and SCAFI is a quantitative, temporal task scale and both measure ataxic manifestations. In contrast, NESSCA quantifies the overall neurologic examination, addressing pyramidal, extrapyramidal, cerebellar, and sensory findings; extrapyramidal manifestations, for instance, have been shown to progress slower than the ataxic ones in SCA3/MJD.[16,31,40] The present results relating chaperone levels to a modulation of the ataxic progression can either be taken literally – i.e. one can propose that chaperone levels might change cerebellar but not the other neurological deteriorations - or can be taken as an effect of a "power threshold". The relatively low sensitivity of NESSCA to detect significant neurological changes in a short period of time, when compared to ataxic scales, might need larger sample sizes or longer study timeframes to detect the same phenomenon. The Rio Grande do Sul cohort, from which the Brazilian patients were recruited for the present study, has presented a substantial association between CAGexp and progression rate, measured by NESSCA and SARA.[16,17] The present subjects were recruited by the extreme phenotype sampling based on AO, which is a powerful approach to discover additional factors that modulate AO. Quite interestingly, both early- and late-onset groups from the Brazilian cohort showed similar progression rates for SARA, SCAFI, and NESSCA (see **Figure 4**). The unknown factors that determined those extreme AOs (independent from CAGexp) did not impact on progression rate - at least in the presence of the CAGexp range carried by this Brazilian cohort. We have seen that differences in chaperone levels, proposed to delay AO in carriers of short CAGexp, did not explain differences in AO independent from CAGexp among Brazilians, and we suggested that this could be due to their large CAGexp range. Therefore, it is quite impressive that DNAJB6a/b and HSP60 levels can impact on progression rate of this group. Again, further longitudinal studies on distinct SCA3/MJD cohorts are urgently needed to confirm our results. If we are correct, chaperone levels can be very important pharmaceutical candidates to modify disease progression in SCA3/MJD.

4.4 Concluding remarks

We suggest that molecular chaperones might contribute to the modulation of both AO and progression rate of disease in SCA3/MJD. The strong effect of larger CAGexp (such as those carried by Brazilian SCA3/MJD subjects) seems to neutralize the effect of molecular chaperones on the outcome AO. Given the promising *in vitro* and *in vivo* roles of DNAJB6 in suppressing aggregation and toxicity of polyQ [23,29,37] and other amyloidogenic proteins,[25–28] this chaperone is a strong candidate for further clinical investigations. If our observations are validated in other cohorts, the manipulation of protein levels of DNAJB6 and other chaperones could have deep therapeutic implications not only for affected, asymptomatic individuals (AO delay), but also for symptomatic patients (reduction of neurological deterioration rate).

Acknowledgements:

The authors would like to thank all patients who were willing to participate in the study. This work was supported by the following Brazilian agencies: CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Project 99999.01528/2015-2017; CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico - Project 402968/2012-3; FIPE-HCPA – Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre –Projects GPPG HCPA 13-0303 and 14-0204. EPM, GVF, MLSP and LBJ were supported by CNPq. MPB was supported by CAPES.

References:

1. Riess O, Rüb U, Pastore A, Bauer P, Schöls L. SCA3: neurological features, pathogenesis and animal models. *Cerebellum*. 2008;7(2):125-137.

2. Nóbrega C, Nascimento-Ferreira I, Onofre I, et al. Overexpression of mutant ataxin-3 in mouse cerebellum induces ataxia and cerebellar neuropathology. *Cerebellum*. 2013;12(4):441-455.

3. Matilla-Dueñas A, Ashizawa T, Brice A, et al. Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. *Cerebellum*. 2014;13(2):269-302.

4. DiFiglia M, Sapp E, Chase KO, et al. Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science*. 1997;277(5334):1990-1993.

5. Wellington CL, Ellerby LM, Gutekunst C-A, et al. Caspase cleavage of mutant huntingtin precedes neurodegeneration in Huntington's disease. *J Cell Biol*. 2002;22(4):749-759.

6. Seidel K, Meister M, Dugbartey GJ, et al. Cellular protein quality control and the evolution of aggregates in spinocerebellar ataxia type 3 (SCA3). *Neuropathol Appl Neurobiol*. 2012;38(6):548-558.

7. Globas C, du Montcel ST, Baliko L, et al. Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. *Mov Disord*. 2008;23(15):2232-2238.

8. Luo L, Wang J, Lo RY, et al. The Initial Symptom and Motor Progression in Spinocerebellar Ataxias. *Cerebellum*. 2017;16(3):615-622.

9. Ilg W, Fleszar Z, Schatton C, et al. Individual changes in preclinical spinocerebellar ataxia identified via increased motor complexity. *Mov Disord*. 2016;31(12):1891-1900.

10. Wu C, Chen D-B, Feng L, et al. Oculomotor deficits in spinocerebellar ataxia type 3: Potential biomarkers of preclinical detection and disease progression. *CNS Neurosci Ther*. 2017;23(4):321-328.

11. de Castilhos RM, Furtado GV, Gheno TC, et al. Spinocerebellar ataxias in Brazil--frequencies and modulating effects of related genes. *Cerebellum*. 2014;13(1):17-28.

12. Tezenas du Montcel S, Durr A, Bauer P, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain*. 2014;137(Pt 9):2444-2455.

13. 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(Pt 12):e398.

14. Chen Z, Zheng C, Long Z, et al. (CAG)n loci as genetic modifiers of age-at-onset in patients with Machado-Joseph disease from mainland China. *Brain*. 2016;139(Pt 8):e41.

15. de Mattos EP, Musskopf MK, Leotti VB, et al. Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis. *J Neurol Neurosurg Psychiatry*. 2018;pii: jnnp-2018-319200.

16. Jardim LB, Hauser L, Kieling C, et al. Progression rate of neurological deficits in a 10-year cohort of SCA3 patients. *Cerebellum*. 2010;9(3):419-428.

17. Donis KC, Saute JAM, Krum-Santos AC, et al. Spinocerebellar ataxia type 3/Machado-

Joseph disease starting before adolescence. *Neurogenetics*. 2016;17(2):107-113.

18. Jacobi H, Bauer P, Giunti P, et al. The natural history of spinocerebellar ataxia type 1, 2, 3, and 6: a 2-year follow-up study. *Neurology*. 2011;77(11):1035-1041.

19. Jacobi H, du Montcel ST, Bauer P, et al. Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: A longitudinal cohort study. *Lancet Neurol*. 2015;14(11):1101-1108.

20. Balchin D, Hayer-Hartl M, Hartl FU. In vivo aspects of protein folding and quality control. *Science (80-)*. 2016;353(6294):aac4354-aac4354.

21. Reis SD, Pinho BR, Oliveira JMA. Modulation of Molecular Chaperones in Huntington's Disease and Other Polyglutamine Disorders. *Mol Neurobiol*. September 2016.

22. 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(5):760-770.

23. Hageman J, Rujano MA, van Waarde MAWH, et al. A DNAJB chaperone subfamily with HDAC-dependent activities suppresses toxic protein aggregation. *Mol Cell*. 2010;37(3):355-369.

24. Hageman J, Kampinga HH. Computational analysis of the human HSPH/HSPA/DNAJ family and cloning of a human HSPH/HSPA/DNAJ expression library. *Cell Stress Chaperones*. 2009;14(1):1-21.

25. Aprile FA, Källstig E, Limorenko G, Vendruscolo M, Ron D, Hansen C. The molecular chaperones DNAJB6 and Hsp70 cooperate to suppress α-synuclein aggregation. *Sci Rep*. 2017;7(1):1-11.

26. Månsson C, Kakkar V, Monsellier E, et al. DNAJB6 is a peptide-binding chaperone which can suppress amyloid fibrillation of polyglutamine peptides at substoichiometric molar ratios. *Cell Stress Chaperones*. *Cell Stress Chaperones*. 2014;19(2):227-39.

27. Månsson C, Arosio P, Hussein R, et al. Interaction of the molecular chaperone DNAJB6 with growing amyloid-beta 42 (Aβ42) aggregates leads to sub-stoichiometric inhibition of amyloid formation. *J Biol Chem*. 2014;289(45):31066-31076.

28. Reidy M, Sharma R, Roberts B-L, Masison DC. Human J-protein DnaJB6b Cures a Subset of Saccharomyces cerevisiae Prions and Selectively Blocks Assembly of Structurally Related Amyloids. *J Biol Chem*. 2016;291(8):4035-4047.

29. Kakkar V, Månsson C, de Mattos EP, et al. The S/T-Rich Motif in the DNAJB6 Chaperone Delays Polyglutamine Aggregation and the Onset of Disease in a Mouse Model. *Mol Cell*. 2016;62(2):272-283.

30. Souza GN, Kersting N, Krum-Santos AC, et al. Spinocerebellar ataxia type 3/Machado-Joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. *Clin Genet*. 2016;90(2):134-140.

31. Schmitz-Hübsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology*. 2006;66(11):1717-1720.

32. Schmitz-Hübsch T, Giunti P, Stephenson DA, et al. SCA Functional Index: a useful compound performance measure for spinocerebellar ataxia. *Neurology*. 2008;71(7):486-492.

33. Kieling C, Rieder CRM, Silva ACF, et al. A neurological examination score for the assessment of spinocerebellar ataxia 3 (SCA3). *Eur J Neurol*. 2008;15(4):371-376.

34. de Mattos EP, Leotti VB, Soong B-w; et al. Age at onset prediction in spinocerebellar ataxia type 3 changes according to population of origin. *Eur J Neurol*. 2018. doi: 10.1111/ene.13779. [Epub ahead of print].

35. Tezenas du Montcel S, Durr A, Rakowicz M, et al. Prediction of the age at onset in spinocerebellar ataxia type 1, 2, 3 and 6. *J Med Genet*. 2014;51(7):479-486.

36. Monte TL, Reckziegel EDR, Augustin MC, et al. The progression rate of spinocerebellar ataxia type 2 changes with stage of disease. *Orphanet J Rare Dis*. 2018;13(1):20.

39. Gillis J, Schipper-Krom S, Juenemann K, et al. The DNAJB6 and DNAJB8 protein chaperones prevent intracellular aggregation of polyglutamine peptides. *J Biol Chem*. 2013;288(24):17225-17237.

40. Saute JA, Rieder CR, Castilhos RM, et al. Planning future clinical trials in Machado Joseph disease: Lessons from a phase 2 trial. *J Neurol Sci*. 2015;358(1-2):72-76.

Tables and Figure Legends:

Table 1: Comparison of clinical variables of patients from the Brazilian and Dutch SCA3/MJD fibroblast cohorts divided in early-, average-, and late-onset groups, according to deviation from the expected age at onset of symptoms. P-values refer to the comparisons within each cohort.

Table 2: Comparison of clinical variables of the Brazilian and Dutch SCA3/MJD fibroblast cohorts.

Table 3: Velocity of disease progression in SCA3/MJD patients from the Brazilian fibroblast cohort as a function of DNAJB6 protein levels at baseline. Correlations were performed for both DNAJB6a and b isoforms separately. For each scale, chaperone levels were considered a continuous variable.

Figure 1: Patient recruitment scheme and study designs. SCA3/MJD patients were recruited according to an extreme phenotype sampling approach focused on age at onset (AO). Early- (squares), average- (circles), and late-onset (triangles) individuals, corrected for the effect of CAG length at the expanded *ATXN3* allele (CAGexp), were recruited. Dashed regression lines divide patients below the 25th (early-onset), between the 25th and 75th (average-onset), and above the 75th (late-onset) percentiles of expected AO based on CAGexp. A cross-sectional design compared chaperone protein levels between AO groups in the Brazilian and Dutch cohorts. A longitudinal design evaluated SARA, SCAFI, and NESSCA progression scores (outcomes) as a function of AO groups and chaperone protein levels (risk factors).

Figure 2: DNAJB6 protein levels in cultured skin fibroblasts from Brazilian and Dutch SCA3/MJD patients. A and **F**: Representative Western blot images of DNAJB6 protein levels in selected individuals from the Brazilian (**A**) and Dutch (**F**) SCA3/MJD cohorts. DNAJB6 was expressed as both long (DNAJB6a; ~40 kDa) and short (DNAJB6b; ~26 kDa) isoforms. Not all individuals are represented. **B** and **C**: Quantification of DNAJB6a (**B**) and DNAJB6b (**C**) protein levels in Brazilian SCA3/MJD patients divided into groups according to deviation

from expected age at onset. Mean protein levels for each sample were estimated from three independent technical replicates. P-values refer to comparisons between AO groups. **D** and **E**: Relationship between DNAJB6a (**D**) and DNAJB6b (**E**) protein levels and age at onset deviation (observed - expected) in Brazilian SCA3/MJD patients. Spearman rank correlation coefficients (r_s) with p-values are also shown. **G** and **H**: Same as in **B** and **C**, respectively, but for patients from the Dutch cohort. **I** and **J**: Same as in **D** and **E**, respectively, but for patients from the Dutch cohort.

Figure 3: Quantification of HSP40, HSP60, and HSPA8 protein levels in cultured skin fibroblasts from Brazilian and Dutch SCA3/MJD patients. A and **H**: Representative Western blot images of HSP40, HSP60, and HSPA8 protein levels in selected individuals from the Brazilian (**A**) and Dutch (**H**) SCA3/MJD cohorts. Not all individuals are represented. **B**, **D**, and **F**: Quantification of total HSP40 (**B**), HSP60 (**D**), and HSPA8 (**F**) protein levels in Brazilian patients divided in groups according to deviation from expected age at onset. Mean protein levels for each sample were estimated from three independent technical replicates. P-values refer to comparisons between AO groups. **C**, **E**, and **G**: Relationship between HSP40 (**C**), HSP60 (**E**), and HSPA8 (**G**) protein levels and age at onset deviation (observed - expected) in Brazilian SCA3/MJD patients. Spearman rank correlation coefficients (r_s) with p-values are also shown. **I**, **K**, and **M**: Same as in **B**, **D**, and **F**, respectively, but for patients from the Dutch cohort. **J**, **L**, and **N**: Same as in **C**, **E**, and **G**, but for patients from the Dutch cohort.

Figure 4: Velocity of SARA and SCAFI progressions in Brazilian SCA3/MJD patients according to disease duration model. A and **B:** SARA (**A**) and SCAFI (**B**) progression scores obtained in groups built by the presence of determinants of AO independent from the CAGexp, studied by the extreme phenotype sampling strategy. Progressions from extremely early- and late-onset individuals were estimated and depicted. Mean slope values with 95% confidence intervals (C.I.) are also shown for each group. **C** and **D**: Mean progressions from patients stratified in groups with 50% lowest or highest DNAJB6a protein levels, as

measured in skin fibroblasts, for SARA (**C**) and SCAFI (**D**) clinical scales. **E** and **F**: Same as in **C** and **D**, respectively, but for DNAJB6b protein levels.

Figure 5: Relationship between SARA and SCAFI progression rates and DNAJB6 protein levels. A and **B**: Individual SARA slopes estimated from the differences in scores at baseline and follow up are shown as a function of DNAJB6a (**A**) and DNAJB6b (**B**) protein levels. R^2 correlation coefficients and p-values are also shown. **C** and **D**: Same as in **A** and **B**, respectively, but for SCAFI.

Supplementary material

Supplementary Table 1: Assignment of samples from Dutch SCA3/MJD patients to age at onset groups (early, intermediate, or late) based on deviation from the expected age at onset. Underlined cases denote differences between the present classification and the one used by Zijlstra and colleagues (2010). Additional clinical characteristics of these samples were described by Zijlstra *et al*., 2010.

Table S2: Clinical characteristics of Brazilian SCA3/MJD patients recruited for this study.

Supplementary Table 3: Spearman rank correlation coefficients between selected chaperone protein levels and clinical variables in Brazilian and Dutch SCA3/MJD fibroblast cohorts.

Supplementary Table 4: Velocity of disease progression in SCA3/MJD patients from the Brazilian fibroblast cohort as a function of protein levels of selected chaperones at baseline. Correlations were performed for HSP40, HSP60, and HSPA8 separately. For each scale, chaperone levels were considered either a continuous or dichotomous variable.

Table 1: Comparison of clinical variables of patients from the Brazilian and Dutch SCA3/MJD fibroblast cohorts divided in early-, average-, and late-onset groups, according to deviation from the expected age at onset of symptoms. P-values refer to the comparisons within each cohort.

	Brazilian cohort (onset groups)				Dutch cohort (onset groups)			
Variable	**Early**	**Average**	**Late**	**p-value**	**Early**	**Average**	**Late**	**p-value**
n (%)	12 (44.4)	5 (18.6)	10 (37.0)	0.236 [a]	8 (38.1)	7 (33.3)	6 (28.6)	0.867 [a]
Age at onset *	21.2 ± 9.5 (5 – 38) [§]	47.0 ± 7.7 (36 – 55) [¥]	48.3 ± 10.3 (31 – 65) [¥]	<0.001 [b]	30.0 ± 12.0 (13 – 49) [§]	35.3 ± 7.5 (29 – 49) [§]	41.2 ± 5.8 (35 – 47) [§]	0.151 [b]
Age at biopsy *	30.2 ± 12.0 (11 – 45) [§]	59.6 ± 5.7 (54 – 69) [¥]	58.2 ± 11.1 (38 – 70) [¥]	<0.001 [b]	45.9 ± 15.6 (24 – 66) [§]	50.8 ± 9.4 (32 – 59) [§]	58.3 ± 11.6 (42 – 75) [§]	0.302 [b]
Disease duration *	9.0 ± 5.4 (2 – 18) [§]	12.6 ± 5.3 (4 – 18) [§]	9.9 ± 4.9 (4 – 18) [§]	0.461 [b]	15.9 ± 14.1 (4 – 40) [§]	15.6 ± 9.2 (0 – 26) [§]	17.2 ± 11.3 (4 – 28) [§]	0.932 [b]
CAGexp *	76.6 ± 4.0 (69 – 82) [§]	73.0 ± 2.7 (70 – 77) [§]	73.9 ± 3.5 (68 – 80) [§]	0.128 [b]	69.4 ± 6.6 (61 – 80) [§]	68.8 ± 3.4 (64 – 72) [§]	72.5 ± 3.4 (68 – 78) [§]	0.303 [b]
CAGnorm *	21.1 ± 6.3 (14 – 33) [§]	22.2 ± 1.3 (20 – 23) [§]	21.4 ± 4.4 (14 – 28) [§]	0.980 [b]	NA	NA	NA	NA

* Mean ± standard deviation (range); [a] Pearson's Chi-squared test; [b] Kruskal-Wallis test. For each cohort, means sharing the same symbol are not statistically different (Tukey-adjusted comparisons); CAGexp/CAGnorm: CAG repeat length at the expanded/non-expanded *ATXN3* allele; NA: not available.

Table 2: Comparison of clinical variables of the Brazilian and Dutch SCA3/MJD fibroblast cohorts.

Variable	Brazilian cohort	Dutch cohort	p-value
Onset group, n (%)			
Early	12 (44.4)	8 (38.1)	
Average	5 (18.6)	7 (33.3)	0.495 [a]
Late	10 (37.0)	6 (28.6)	
Age at onset *	36.0 ± 16.3 (5 – 65)	35.0 ± 9.9 (13 – 49)	0.777 [b]
Age at biopsy *	46.0 ± 17.8 (11 – 70)	51.1 ± 13.1 (24 – 75)	0.262 [b]
Disease duration *	10.0 ± 5.2 (2 – 18)	16.1 ± 11.3 (0 – 40)	0.118 [c]
CAGexp *	74.9 ± 3.8 (68 – 82)	70.1 ± 4.9 (61 – 80)	<0.001 [b]

* Mean ± standard deviation (range); [a] Pearson's Chi-squared test; [b] Welch two sample t-test; [c] Wilcoxon rank sum test; CAGexp: CAG repeat length at the expanded/non-expanded *ATXN3* allele.

Table 3: Velocity of disease progression in SCA3/MJD patients from the Brazilian fibroblast cohort as a function of DNAJB6 protein levels at baseline. Correlations were performed for both DNAJB6a and b isoforms separately. For each scale, chaperone levels were considered a continuous variable.

Scale	Chaperone	Effect [a]	p-value
NESSCA	DNAJB6a	-0.06 (-0.18 – 0.07)	0.349
	DNAJB6b	-0.05 (-0.45 – -0.45)	0.818
SARA	DNAJB6a	-0.16 (-0.26 – -0.04)	**0.002**
	DNAJB6b	-0.40 (-0.74 – -0.01)	**0.025**
SCAFI	DNAJB6a	0.02 (0.00 – 0.04)	**0.046**
	DNAJB6b	0.03 (-0.02 – 0.09)	0.271

[a] Change in progression score (points/year) for every 0.1 increase in chaperone levels, mean (95% C.I.). NESSCA: Neurological Examination Score for Spinocerebellar Ataxia; SARA: Scale for the Assessment and Rating of Ataxia; SCAFI: Spinocerebellar Ataxia Functional Index.

Table S1: Assignment of samples from Dutch SCA3/MJD patients to age at onset groups (early, intermediate, or late) based on deviation from the expected age at onset. Additional clinical characteristics of these samples were described by Zijlstra *et al*., 2010.

ID	AO group (Zijlstra et al., 2010)	Expected AO [a, b]	ΔAO [a]	AO group (this study)	Harvesting passage number
1	Late	38.7	8.3	Late	10
2	Early	38.7	-8.7	Early	11
3	Late	22.0	16.0	Late	11
4	Average	34.6	0.4	Average	10
5[c]	Average	29.2	5.8	Late	13
6	Average	30.9	-0.9	Average	9
7[c]	Late	48.5	0.5	Average	11
8	Late	32.7	12.3	Late	11
9	Average	48.5	-6.5	Average	8
10[c]	Early	34.6	-4.6	Average	11
11	Late	32.7	14.3	Late	11
12	Average	32.7	-3.7	Average	8
13	Early	26.1	-4.1	Early	11
14[c]	Average	57.4	-8.4	Early	24
15[c]	Average	48.5	-9.5	Early	12
16	Late	27.6	7.4	Late	12
17[c]	Average	40.9	-5.9	Early	11
18	Early	45.8	-10.8	Early	14
19	Early	26.1	-9.1	Early	9
20	Early	19.7	-6.7	Early	15
21	Average	32.7	-0.7	Average	7

[a] In years; [b] calculated with the AO prediction model described by Tezenas du Montcel *et al*., 2014; [c] denotes differences between the present classification and the one used by Zijlstra and colleagues (2010). AO: age at onset.

Table S2: Clinical characteristics of Brazilian SCA3/MJD patients recruited for this study.

ID	Sex	CAG norm	CAG exp	Age at biopsy [a]	Disease duration [a]	AO [a]	Expected AO [a]	ΔAO [a]	AO group
1	M	22	69	45	17	28	54.8	-26.8	Early
2	F	14	82	11	6	5	20.0	-15.0	Early
3	M	14	76	33	12	21	36.0	-15.0	Early
4	M	24	73	36	6	30	44.1	-14.1	Early
5	F	18	79	36	18	18	28.0	-10.0	Early
6	F	27	76	32	6	26	36.0	-10.0	Early
7	M	24	74	40	12	28	41.4	-13.4	Early
8	F	33	81	11	2	9	22.6	-13.6	Early
9	F	14	81	14	2	12	22.6	-10.6	Early
10	F	14	78	28	9	19	30.7	-11.7	Early
11	F	22	78	34	13	21	30.7	-9.7	Early
12	F	27	72	43	5	38	46.8	-8.8	Early
13	M	23	74	58	15	43	41.4	1.6	Average
14	M	23	77	54	18	36	33.4	2.6	Average
15	M	23	70	69	14	55	52.1	2.9	Average
16	M	22	71	57	4	53	49.4	3.6	Average
17	M	20	73	60	12	48	44.1	3.9	Average
18	M	23	78	43	8	35	30.7	4.3	Late
19	F	24	74	57	10	47	41.4	5.6	Late
20	F	23	74	65	18	47	41.4	5.6	Late
21	F	22	80	38	7	31	25.3	5.7	Late
22	M	23	73	63	13	50	44.1	5.9	Late
23	M	14	75	50	5	45	38.7	6.3	Late
24	F	28	68	70	5	65	57.5	7.5	Late
25	M	20	70	64	4	60	52.1	7.9	Late
26	F	23	72	70	15	55	46.8	8.2	Late
27	F	14	75	62	14	48	38.7	9.3	Late

[a] In years. AO: age at onset; CAGnorm/exp: length of CAG repeats at the non-expanded/expanded *ATXN3* allele; ΔAO: difference between observed and expected AO; F: female; M: male.

Table S3: Spearman rank correlation coefficients between selected chaperone protein levels and clinical variables in Brazilian and Dutch SCA3/MJD fibroblast cohorts.

		Age at biopsy		Disease duration		ΔAO [a]		Passage number		CAGexp [b]	
Cohort	**Chaperone**	**Spearman ρ**	**p-value**	**Spearman ρ**	**p-value**	**Spearman ρ**	**p-value**	**Spearman ρ**	**p-value**	**Spearman ρ**	**p-value**
Brazilian	DNAJB6a	-0.085	0.675	0.219	0.273	-0.132	0.512	NA	NA	0.146	0.469
	DNAJB6b	-0.112	0.579	-0.013	0.947	-0.288	0.145	NA	NA	-0.076	0.705
	HSP40	-0.250	0.208	-0.364	0.062	-0.106	0.598	NA	NA	0.266	0.180
	HSP60	-0.093	0.645	-0.255	0.199	-0.115	0.569	NA	NA	-0.054	0.788
	HSPA8	-0.003	0.988	-0.154	0.444	-0.215	0.281	NA	NA	-0.137	0.495
Dutch	DNAJB6a	0.216	0.346	0.183	0.428	0.488	**0.025**	0.146	0.528	0.366	0.103
	DNAJB6b	0.207	0.369	0.034	0.884	0.022	0.924	0.206	0.370	-0.200	0.385
	HSP40	0.288	0.206	-0.023	0.920	0.653	**0.001**	-0.302	0.183	-0.082	0.725
	HSP60	0.475	**0.030**	0.150	0.516	0.548	**0.010**	0.142	0.540	0.080	0.729
	HSPA8	0.175	0.449	-0.044	0.849	0.491	**0.024**	0.088	0.704	0.128	0.580

[a] Observed - expected age at onset; [b] expanded *ATXN3* allele; NA: not applicable.

Table S4: Velocity of disease progression in SCA3/MJD patients from the Brazilian fibroblast cohort as a function of protein levels of selected chaperones at baseline. Correlations were performed for HSP40, HSP60, and HSPA8 separately. For each scale, chaperone levels were considered either a continuous or dichotomous variable.

Model	Chaperone	Effect [a]	p-value
NESSCA; continuous chaperone levels	HSP40	0.01 (-0.08 – 0.11)	0.814
	HSP60	0.04 (-0.12 – 0.20)	0.636
	HSPA8	0.04 (-0.09 – 0.18)	0.565
NESSCA; dichotomous chaperone levels [b]	HSP40	0.04 (-0.06 – 0.15)	0.421
	HSP60	0.01 (-0.08 – 0.10)	0.858
	HSPA8	0.03 (-0.05 – 0.11)	0.454
SARA; continuous chaperone levels	HSP40	-0.06 (-0.16 – 0.04)	0.184
	HSP60	-0.05 (-0.20 – 0.10)	0.468
	HSPA8	0.01 (-0.12 – 0.15)	0.856
SARA; dichotomous chaperone levels [b]	HSP40	-0.04 (-0.14 – 0.07)	0.512
	HSP60	-0.09 (-0.18 – -0.01)	**0.033**
	HSPA8	-0.02 (-0.10 – 0.08)	0.698
SCAFI; continuous chaperone levels	HSP40	0.00 (-0.01 – 0.02)	0.539
	HSP60	0.01 (-0.01 – 0.02)	0.285
	HSPA8	0.00 (-0.02 – 0.02)	0.683
SCAFI; dichotomous chaperone levels [b]	HSP40	-0.01 (-0.03 – 0.01)	0.222
	HSP60	0.00 (-0.02 – 0.01)	0.582
	HSPA8	0.01 (-0.01 – 0.02)	0.313

[a] Change in progression score (points/year) for every 0.1 increase in chaperone levels, mean (95% C.I.). For dichotomous models, effects for the "high" expression group are shown; [b] protein levels of each fibroblast line were assigned to a "high" or "low" expression group based on median expression values; NESSCA: Neurological Examination Score for Spinocerebellar Ataxia; SARA: Scale for the Assessment and Rating of Ataxia; SCAFI: Spinocerebellar Ataxia Functional Index.

Figure 1: Patient recruitment scheme and study designs. SCA3/MJD patients were recruited according to an extreme phenotype sampling approach focused on age at onset (AO). Early- (squares), average- (circles), and late-onset (triangles) individuals, corrected for the effect of CAG length at the expanded *ATXN3* allele (CAGexp), were recruited. Dashed regression lines divide patients below the 25th (early-onset), between the 25th and 75th (average-onset), and above the 75th (late-onset) percentiles of expected AO based on CAGexp. A cross-sectional design compared chaperone protein levels between AO groups in the Brazilian and Dutch cohorts. A longitudinal design evaluated SARA, SCAFI, and NESSCA progression scores (outcomes) as a function of AO groups and chaperone protein levels (risk factors).

A

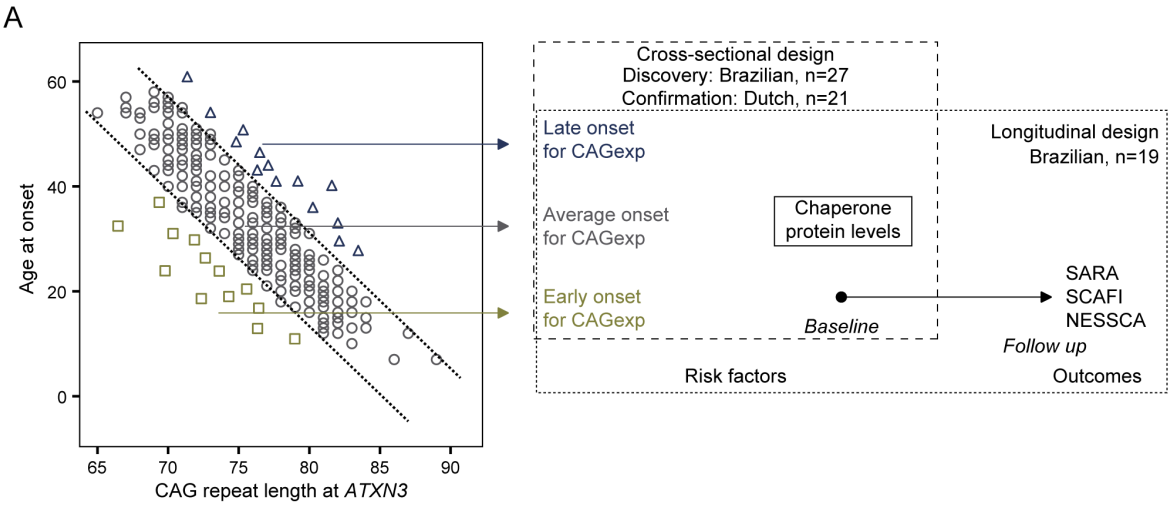


Figure 2 (next page): DNAJB6 protein levels in cultured skin fibroblasts from Brazilian and Dutch SCA3/MJD patients. A and **F**: Representative Western blot images of DNAJB6 protein levels in selected individuals from the Brazilian (**A**) and Dutch (**F**) SCA3/MJD cohorts. DNAJB6 was expressed as both long (DNAJB6a; ~40 kDa) and short (DNAJB6b; ~26 kDa) isoforms. Not all individuals are represented. **B** and **C**: Quantification of DNAJB6a (**B**) and DNAJB6b (**C**) protein levels in Brazilian SCA3/MJD patients divided into groups according to deviation from expected age at onset. Mean protein levels for each sample were estimated from three independent technical replicates. P-values refer to comparisons between AO groups. **D** and **E**: Relationship between DNAJB6a (**D**) and DNAJB6b (**E**) protein levels and age at onset deviation (observed - expected) in Brazilian SCA3/MJD patients. Spearman rank correlation coefficients (r_s) with p-values are also shown. **G** and **H**: Same as in **B** and **C**, respectively, but for patients from the Dutch cohort. **I** and **J**: Same as in **D** and **E**, respectively, but for patients from the Dutch cohort.

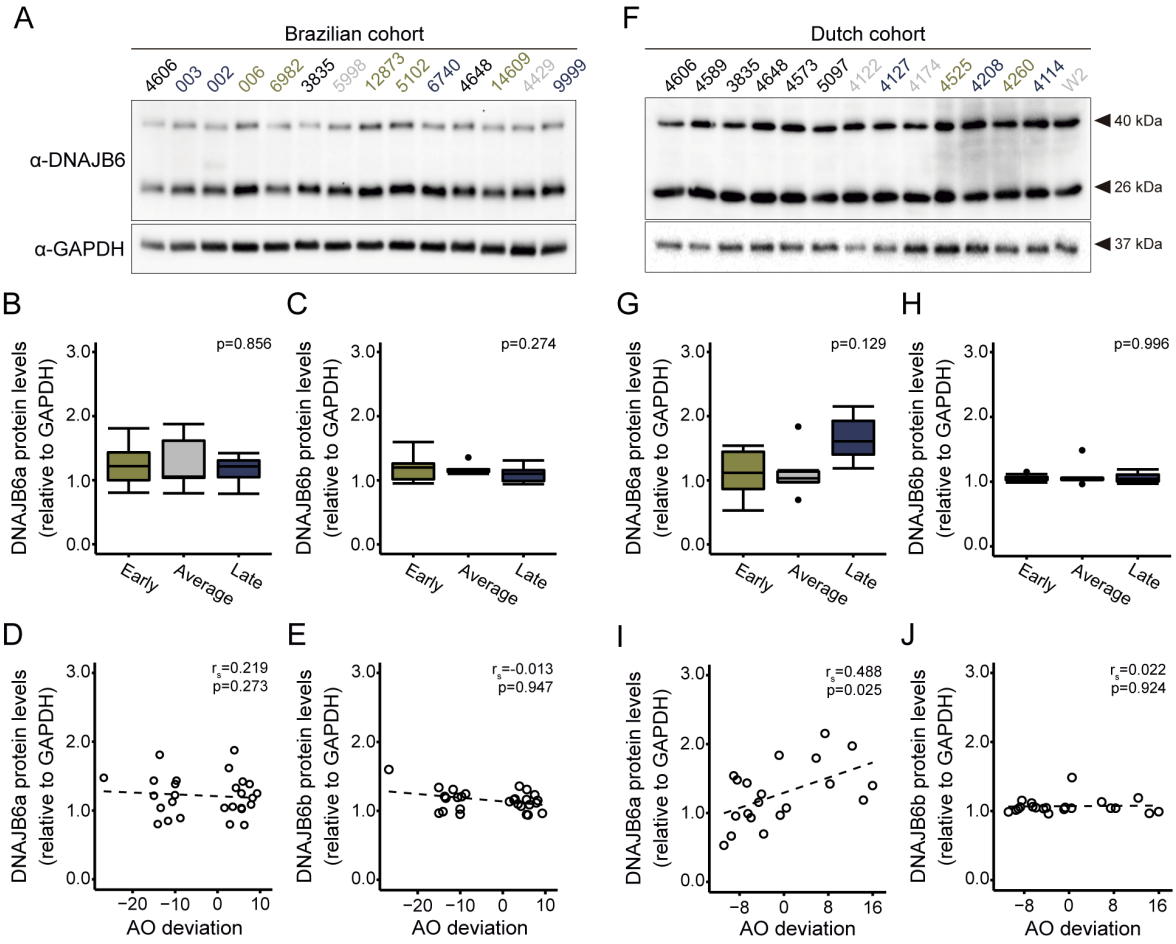

A
Brazilian cohort
4606
003
002
006
6982
3835
5998
12873
5102
6740
4648
14609
4429
9999
α-DNAJB6
α-GAPDH
F
Dutch cohort
4606
4589
3835
4648
4573
5097
4122
4127
4174
4525
4208
4260
4114
W2
40 kDa
26 kDa
37 kDa
B
C
G
H
DNAJB6a protein levels (relative to GAPDH)
DNAJB6b protein levels (relative to GAPDH)
p=0.856
p=0.274
p=0.129
p=0.996
Early
Average
Late
D
E
I
J
rs=0.219
p=0.273
rs=-0.013
p=0.947
rs=0.488
p=0.025
rs=0.022
p=0.924
AO deviation

Figure 3 (next page): Quantification of HSP40, HSP60, and HSPA8 protein levels in cultured skin fibroblasts from Brazilian and Dutch SCA3/MJD patients. A and **H**: Representative Western blot images of HSP40, HSP60, and HSPA8 protein levels in selected individuals from the Brazilian (**A**) and Dutch (**H**) SCA3/MJD cohorts. Not all individuals are represented. **B**, **D**, and **F**: Quantification of total HSP40 (**B**), HSP60 (**D**), and HSPA8 (**F**) protein levels in Brazilian patients divided in groups according to deviation from expected age at onset. Mean protein levels for each sample were estimated from three independent technical replicates. P-values refer to comparisons between AO groups. **C**, **E**, and **G**: Relationship between HSP40 (**C**), HSP60 (**E**), and HSPA8 (**G**) protein levels and age at onset deviation (observed - expected) in Brazilian SCA3/MJD patients. Spearman rank correlation coefficients (r_s) with p-values are also shown. **I**, **K**, and **M**: Same as in **B**, **D**, and **F**, respectively, but for patients from the Dutch cohort. **J**, **L**, and **N**: Same as in **C**, **E**, and **G**, but for patients from the Dutch cohort.

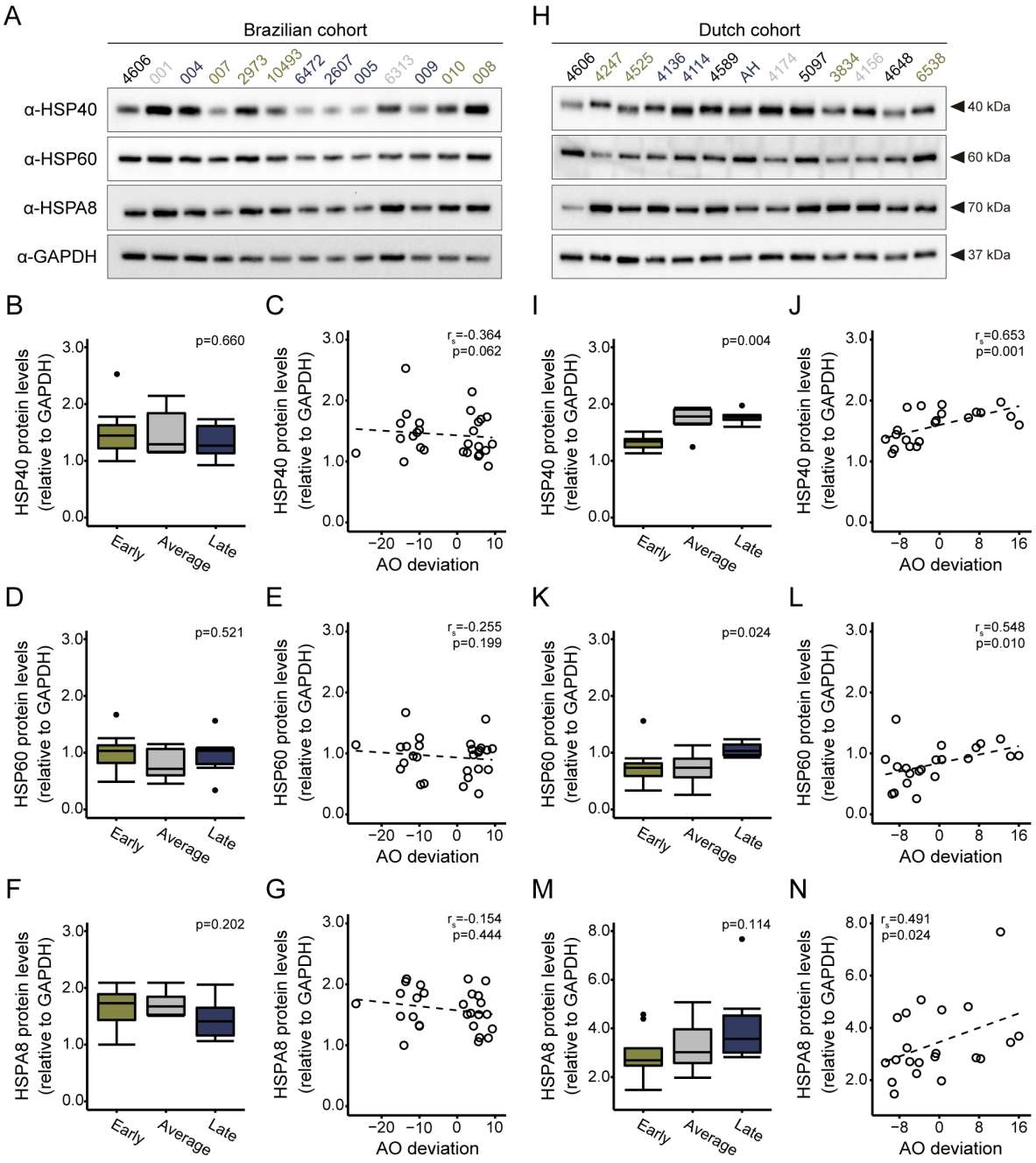

A
Brazilian cohort
4606
001
004
007
2973
10493
6472
2607
005
6313
009
010
008
α-HSP40
α-HSP60
α-HSPA8
α-GAPDH
H
Dutch cohort
4606
4247
4525
4136
4114
4589
AH
4174
5097
3834
4156
4648
6538
40 kDa
60 kDa
70 kDa
37 kDa
B
HSP40 protein levels (relative to GAPDH)
p=0.660
Early
Average
Late
C
HSP40 protein levels (relative to GAPDH)
rs=-0.364
p=0.062
AO deviation
I
HSP40 protein levels (relative to GAPDH)
p=0.004
J
HSP40 protein levels (relative to GAPDH)
rs=0.653
p=0.001
AO deviation
D
HSP60 protein levels (relative to GAPDH)
p=0.521
E
HSP60 protein levels (relative to GAPDH)
rs=-0.255
p=0.199
AO deviation
K
HSP60 protein levels (relative to GAPDH)
p=0.024
L
HSP60 protein levels (relative to GAPDH)
rs=0.548
p=0.010
AO deviation
F
HSPA8 protein levels (relative to GAPDH)
p=0.202
G
HSPA8 protein levels (relative to GAPDH)
rs=-0.154
p=0.444
AO deviation
M
HSPA8 protein levels (relative to GAPDH)
p=0.114
N
HSPA8 protein levels (relative to GAPDH)
rs=0.491
p=0.024
AO deviation

Figure 4 (next page): Velocity of SARA and SCAFI progressions in Brazilian SCA3/MJD patients according to disease duration model. A and **B:** SARA (**A**) and SCAFI (**B**) progression scores obtained in groups built by the presence of determinants of AO independent from the CAGexp, studied by the extreme phenotype sampling strategy. Progressions from extremely early- and late-onset individuals were estimated and depicted. Mean slope values with 95% confidence intervals (C.I.) are also shown for each group. **C** and **D**: Mean progressions from patients stratified in groups with 50% lowest or highest DNAJB6a protein levels, as measured in skin fibroblasts, for SARA (**C**) and SCAFI (**D**) clinical scales. **E** and **F**: Same as in **C** and **D**, respectively, but for DNAJB6b protein levels.

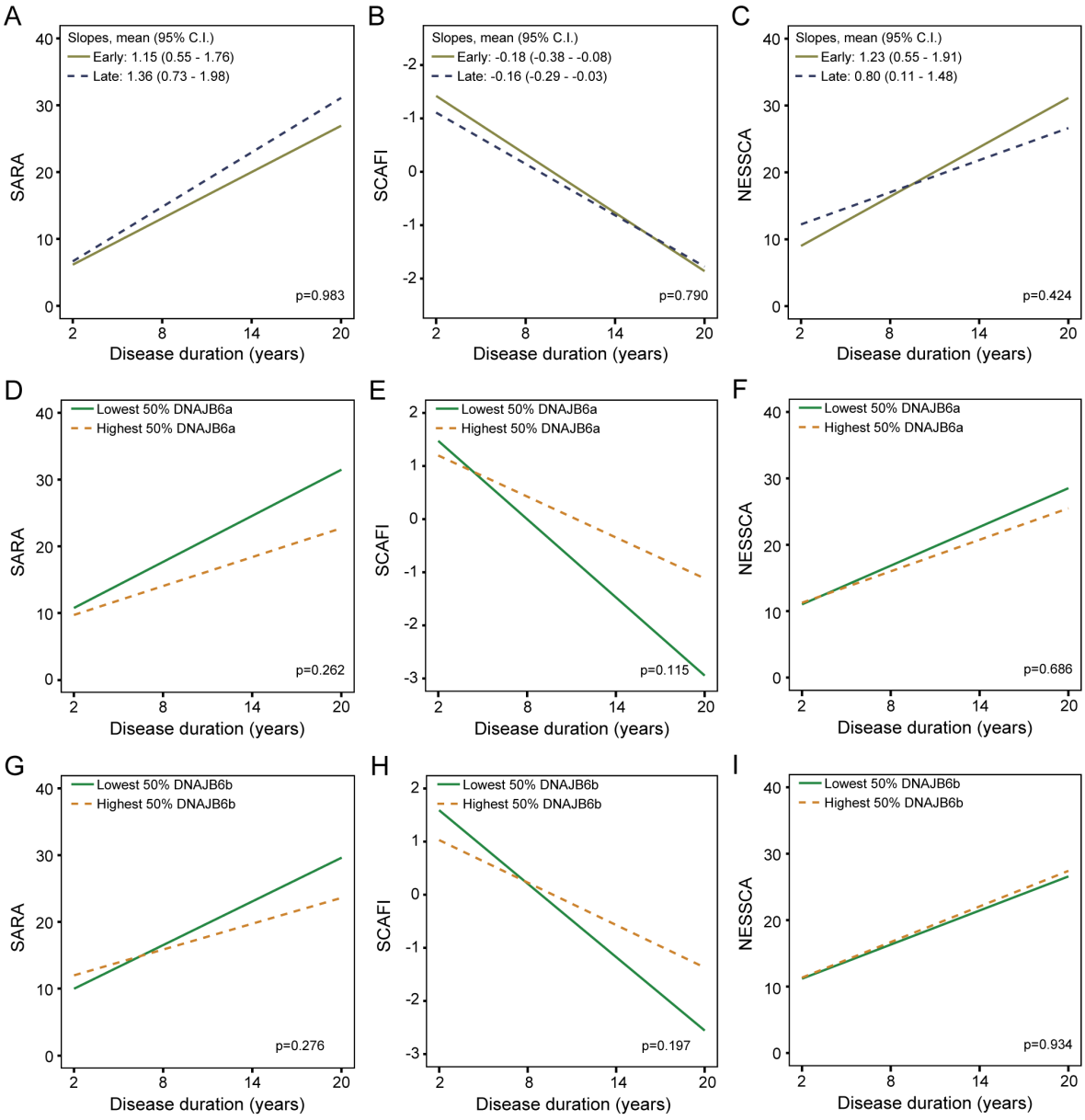
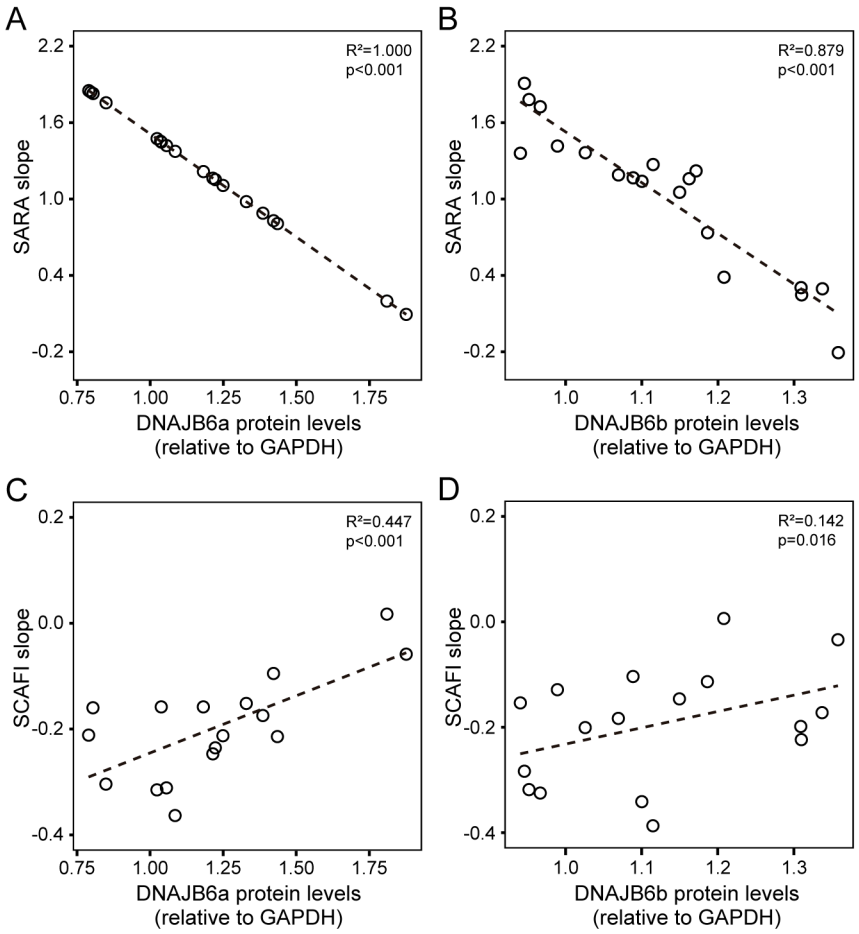

A
Slopes, mean (95% C.I.)
Early: 1.15 (0.55 - 1.76)
Late: 1.36 (0.73 - 1.98)
SARA
p=0.983
Disease duration (years)
B
Slopes, mean (95% C.I.)
Early: -0.18 (-0.38 - -0.08)
Late: -0.16 (-0.29 - -0.03)
SCAFI
p=0.790
Disease duration (years)
C
Slopes, mean (95% C.I.)
Early: 1.23 (0.55 - 1.91)
Late: 0.80 (0.11 - 1.48)
NESSCA
p=0.424
Disease duration (years)
D
Lowest 50% DNAJB6a
Highest 50% DNAJB6a
SARA
p=0.262
Disease duration (years)
E
Lowest 50% DNAJB6a
Highest 50% DNAJB6a
SCAFI
p=0.115
Disease duration (years)
F
Lowest 50% DNAJB6a
Highest 50% DNAJB6a
NESSCA
p=0.686
Disease duration (years)
G
Lowest 50% DNAJB6b
Highest 50% DNAJB6b
SARA
p=0.276
Disease duration (years)
H
Lowest 50% DNAJB6b
Highest 50% DNAJB6b
SCAFI
p=0.197
Disease duration (years)
I
Lowest 50% DNAJB6b
Highest 50% DNAJB6b
NESSCA
p=0.934
Disease duration (years)

Figure 5: Relationship between SARA and SCAFI progression rates and DNAJB6 protein levels. A and **B**: Individual SARA slopes estimated from the differences in scores at baseline and follow up are shown as a function of DNAJB6a (**A**) and DNAJB6b (**B**) protein levels. R^2 correlation coefficients and p-values are also shown. **C** and **D**: Same as in **A** and **B**, respectively, but for SCAFI.



Capítulo 5. Discussão

A presente tese teve como objetivo geral identificar novos modificadores genéticos da AO e da velocidade de progressão neurológica da DMJ/SCA3.

Para definirmos o estado da arte e partirmos em busca de bons candidatos ainda não completamente estudados, realizamos uma revisão sistemática e metanálise sobre fatores modificadores da AO na DMJ/SCA3. Disto resultou o primeiro artigo da presente tese, intitulado "*Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis*". Utilizando dados primários de 10 coortes de pacientes e dados agregados de 2 outras coortes, determinamos que o CAGexp no *ATXN3* determina 55,2% da variabilidade fenotípica da AO, em média. Desse modo, cerca de metade da variância da AO deve ser explicada por outros fatores (genéticos e/ou ambientais) que influenciam a idade em que um indivíduo com DMJ/SCA3 inicia com os primeiros sintomas. Esse trabalho também validou a influência de outros modificadores da AO que já haviam sido propostos na literatura: fatores familiais e o comprimento do trato CAG no *ATXN2* foram os mais relevantes, seguidos em menor medida pelo gênero do indivíduo afetado. Desse modo, estes fatores deverão ser levados em conta em futuros estudos de modificadores da AO na DMJ/SCA3.

No entanto, o principal e mais original achado da metanálise foi a observação de diferenças muito claras do efeito do CAGexp sobre a AO entre populações distintas de pacientes com DMJ/SCA3. Isso demonstra que um modelo universal para estimar o efeito da mutação causal da DMJ/SCA3 sobre sua AO pode não ser a melhor estratégia para comparar o impacto de moduladores em populações distintas. Isso fica claro quando observamos as dispersões do CAGexp e da AO entre coortes de pacientes com DMJ/SCA3 de origens distintas (de Castilhos et al. 2014; Tezenas du Montcel et al. 2014a; Raposo et al. 2015; Chen et al. 2016). A variabilidade do CAGexp entre populações de pacientes é muito maior do que a variabilidade da AO (ver a figura 2 do artigo 1). Se o efeito do CAGexp fosse o mesmo em todas as populações, coortes com tratos CAGexp mais longos – como

é o caso dos pacientes com DMJ/SCA3 no Estado do Rio Grande do Sul (Souza et al. 2016) – deveriam ter casos predominantemente de origem na infância, o que não ocorre na maioria dos indivíduos (Jardim et al. 2001a; Jardim et al. 2010; de Castilhos et al. 2014; Souza et al. 2016).

Propusemos a existência de três grandes grupos de pacientes com DMJ/SCA3 no mundo, com efeitos distintos do CAGexp sobre a AO, com base no efeito do CAGexp sobre a AO e nas suas relações de proximidade étnica: o grupo com ancestralidade europeia não-portuguesa, o grupo geral (ou intermediário) e o grupo com ancestralidade portuguesa. A figura 3 do artigo 1 dá a sustentação empírica para essa proposição. Do primeiro grupo, fazem parte as coortes recrutadas pelo consórcio europeu EUROSCA – e que praticamente apenas não incluiu portugueses – e pelos holandeses. Do último grupo, fazem parte as coortes recrutadas em Portugal continental, nos Açores e no Rio Grande do Sul. As demais coortes compreenderam o grupo geral e provieram de regiões muito distantes umas das outras: China continental, Taiwan, Brasil (à exceção do Rio Grande do Sul) e Cuba. Apesar da heterogeneidade geográfica, esse grupo geral apresenta uma relação CAGexp *versus* AO relativamente semelhante; por isso, propusemos que esta seja a relação “média”, ou geral, da causalidade da AO atribuível ao CAGexp na DMJ/SCA3.

O leitor poderá questionar em que se distinguem os sujeitos DMJ/SCA3 brasileiros do Rio Grande do Sul (pertencentes ao grupo tardio) dos demais brasileiros (pertencentes ao grupo geral). Ainda não há uma resposta objetiva para essa questão. No entanto, a colonização do Rio Grande do Sul por europeus foi bastante distinta daquela ocorrida no resto do território brasileiro. Açorianos (e não portugueses continentais) colonizaram o atual Rio Grande do Sul a partir de 1755, e estima-se que em um intervalo de não mais do que vinte anos mais de 2.000 pessoas tenham vindo como colonos e ocupado um até então vazio geográfico no sul do Brasil. A informação da ancestralidade genética dos indivíduos com DMJ/SCA3 do Rio Grande do Sul já é conhecida, e todas as famílias (salvo uma emigrada recentemente) apresentam o mesmo haplótipo ancestral ACA (Saraiva-Pereira, comunicação pessoal). Em contraste, o restante do território brasileiro

recebeu o aporte de portugueses tanto continentais como das ilhas açorianas, além de diversas outras etnias. Por isso, é provável que os sujeitos com DMJ/SCA3 que vivem em outros estados brasileiros tenham origens étnicas também mais heterogêneas.

Em relação ao grupo que denominamos de geral no artigo 1, os pacientes de origem europeia não-portuguesa parecem estar em desvantagem. Para um trato CAGexp de mesmo tamanho, esses pacientes manifestam AO mais precoce do que aqueles no grupo intermediário. Ao contrário, pacientes de Portugal, ou com ancestralidade mais definidamente portuguesa, como é o caso da coorte Rio Grande do Sul, iniciam os sintomas significativamente mais tarde do que indivíduos de outros grupos, para um trato CAGexp de mesmo tamanho. Essas observações sugerem que europeus não-portugueses (grupo “precoce”) e portugueses e gaúchos (grupo “tardio”) possuem fatores que predispõem a AO mais precoce e tardia, respectivamente. Em relação especificamente à coorte gaúcha, chama a atenção que esse efeito protetor se dê mesmo na presença de tamanhos de CAGexp relativamente maiores do que os tamanhos médios encontrados entre seus distantes “primos” açorianos - separados por cerca de 20 a 25 gerações.

Não é possível ainda sequer propor quais seriam esses fatores modificadores. Nosso “viés cognitivo” tende a supor que se tratem de genes modificadores, cujos alelos mais ou menos impactantes tenham distribuições diferentes entre as populações. Fenômenos epigenéticos também seriam plausíveis. Embora se possa pensar em fatores ambientais, será bem mais difícil encontrar hipóteses que unifiquem de forma aceitável, por exemplo, gaúchos e portugueses continentais em relação a seus hábitos ou exposições – além de distingui-los dos demais. É provável que o estudo comparativo de coortes desses dois grupos extremos de pacientes venha a ser de grande ajuda na busca dos reais modificadores da AO responsáveis por essas diferenças.

Os resultados da revisão sistemática e metanálise serviram como referencial teórico para os demais manuscritos desta tese. Primeiro, o efeito diferencial do CAGexp na AO de pacientes com DMJ/SCA3 de populações distintas foi abordado

no segundo artigo, intitulado "*Age at onset prediction in spinocerebellar ataxia type 3 changes according to population of origin*". Esse trabalho teve como ponto de início as equações publicadas por Tezenas du Montcel e colaboradores (2014b) para a predição da AO em indivíduos europeus portadores de tratos CAGexp, mas ainda assintomáticos para a DMJ/SCA3 (Tezenas du Montcel et al. 2014b). Há diversas vantagens em se desenvolver métodos estatísticos para a predição acurada da AO em sujeitos assintomáticos. Por exemplo, tanto o aconselhamento genético como o delineamento de novos ensaios clínicos se beneficiariam enormemente de estimativas mais precisas de quando determinado paciente com DMJ/SCA3 irá iniciar os sintomas.

De forma marcante, a fórmula europeia para a predição da AO foi muito imprecisa para pacientes com DMJ/SCA3 de populações não-europeias. Demonstramos nesse artigo que a fórmula europeia subestimou a AO, tanto de pacientes da coorte Rio Grande do Sul, como de Taiwan. Por outro lado, a predição europeia foi mais acurada para pacientes portugueses açorianos, ainda que erros também tenho sido produzidos. Esses resultados corroboraram os dados do artigo 1, que indicaram que pacientes europeus com DMJ/SCA3 parecem ter fatores que predispõem à AO mais precoce, em relação a pacientes de outras populações. Uma vez que a fórmula europeia de predição da AO foi baseada em dados de CAGexp e AO de pacientes europeus, esse efeito população-específico foi incorporado nas predições. Isso ficou evidente quando a fórmula europeia foi testada em pacientes do Rio Grande do Sul e de Taiwan.

Desse modo, optou-se por desenvolver fórmulas população-específicas para a predição da AO em pacientes gaúchos e taiwaneses. Isso só foi possível porque ambas as coortes possuíam portadores do CAGexp assintomáticos, o que permitiu o desenvolvimento de modelos paramétricos de sobrevida similares aos propostos para a doença de Huntington (Langbehn et al. 2004; Langbehn et al. 2010). Os modelos brasileiro e taiwanês para predição da AO na DMJ/SCA3 foram validados em grupos distintos de pacientes oriundos das mesmas populações para as quais as fórmulas foram geradas, demonstrando sua aplicabilidade clínica. Além disso, optou-se por deixar as equações de predição disponíveis a todos os pesquisadores,

de modo que esses dados possam contribuir para o desenvolvimento de novas equações população-específicas e para a pesquisa de novos moduladores da AO na DMJ/SCA3.

Depois desse estágio de levantamento do estado da arte, passamos a testar o efeito de dois candidatos a modificadores fenotípicos na DMJ/SCA3. O manuscrito 3, intitulado "*ApoEε4 allele is associated with earlier age at onset in spinocerebellar ataxia type 3*" abordou um candidato específico a modificador genético da AO na DMJ/SCA3: o gene da *apolipoproteína E* (*apoE*). A revisão sistemática realizada no artigo 1 apontou o *apoE* como um dos mais promissores candidatos a modulador da AO, e que já havia sido estudado em três coortes distintas de pacientes com DMJ/SCA3 (Bettencourt et al. 2011; Zhou et al. 2014; Peng et al. 2014). A presença de alelos *apoE* ε4 é o principal fator genético de risco conhecido para o desenvolvimento da doença de Alzheimer, em suas formas não monogênicas (Corder et al. 1994; Huynh et al. 2017a; Lautner et al. 2017; Liu et al. 2017). Na DMJ/SCA3, entretanto, dois dos três estudos existentes apontaram para um efeito deletério do alelo ε2 do *apoE*, ao invés de ε4, que estaria contribuindo para o desenvolvimento da AO mais precoce, em relação a pacientes sem alelos *apoE* ε2 (Bettencourt et al. 2011; Peng et al. 2014). Por outro lado, o terceiro estudo do *apoE* na DMJ/SCA3, realizado em uma grande coorte chinesa, não identificou nenhuma diferença significativa nas AOs de pacientes com diferentes genótipos do *apoE* (Zhou et al. 2014). Assim, não estava completamente claro se *apoE* modula de fato a AO na DMJ/SCA3.

O manuscrito 3 compreendeu a maior avaliação do *apoE* na DMJ/SCA3 já reportada na literatura, incluindo 482 pacientes da coorte Rio Grande do Sul. Esse tamanho amostral permitiu a detecção de uma frequência de indivíduos homozigotos para *apoE* ε4 significativamente maior do que nos estudos anteriores (Bettencourt et al. 2010; Zhou et al. 2014; Peng et al. 2014). Mais do que isso, em relação a outros genótipos, a presença de dois alelos *apoE* ε4 foi correlacionada com uma antecipação média de 6 anos da AO do primeiro sintoma e de mais de 11 anos da AO da ataxia de marcha. Nenhum efeito significativo do *apoE* ε2 sobre a AO foi detectado, contrariando os achados dos estudos anteriores (Bettencourt et

al. 2010; Peng et al. 2014). Contudo, a associação do *apoE* ε4 com AO mais precoce na DMJ/SCA3 vai ao encontro de todas as evidências *in vitro* e *in vivo* que demonstraram piores desfechos clínicos entre pacientes com as doenças de Alzheimer e Parkinson que eram portadores de alelos ε4, em relação a portadores do *apoE* ε2 ou ε3 (Tsuang et al. 2013; Emamzadeh et al. 2016; Lautner et al. 2017; Liu et al. 2017; Robert et al. 2017; Nuriel et al. 2017). Uma vez que tanto as doenças de Alzheimer e Parkinson, quanto a DMJ/SCA3 e demais poliglutaminopatias possuem mecanismos moleculares similares de patogenicidade, a modulação dos níveis neuronais da ApoE pode ser uma estratégia terapêutica interessante para doenças com tratos poliQ expandidos, assim como se tem observado para a doença de Alzheimer (Huynh et al. 2017b).

Enquanto temos a convicção do efeito do alelo *apoE* ε4 sobre a AO da DMJ/SCA3 a partir dos nossos achados – e basicamente permitidos pelo poder da nossa amostra –, o mesmo não se pode dizer sobre o efeito do alelo *apoE* ε2. Quando resultados controversos são obtidos em populações distintas, a primeira hipótese explicativa que vem à mente é a da presença de estratificação populacional. A distribuição dos alelos pode divergir muito entre diferentes grupamentos humanos. Nós sugerimos que esta controvérsia sobre se os alelos *apoE* ε2 afinal tenham ou não um efeito sobre a AO na DMJ/SCA3 seja dirimida, no futuro, por uma metanálise com os dados dos participantes individuais (IPDs) à maneira do que realizamos no artigo 1 com os dados disponíveis da literatura à época.

Por fim, os dois últimos manuscritos da presente tese foram dedicados às chaperonas moleculares, consideradas por muitos autores como um dos grupos mais promissores de potenciais modificadores do fenótipo nas poliglutaminopatias e outras doenças com mecanismo semelhante (Bukau et al. 2006; Balchin et al. 2016; Ciechanover and Kwon 2017; Mannini and Chiti 2017). Deve-se frisar que boa parte do trabalho realizado em ambos os manuscritos foi realizado durante estágio sanduíche no Departamento de Biologia Celular do *Universitair Medisch Centrum Groningen*, Universidade de Groningen, Holanda. Esse estágio aconteceu

dentro de um projeto de pesquisa financiado pelo Programa Ciência sem Fronteiras do Conselho Nacional de Desenvolvimento Científico e Tecnológico.

O artigo 4, intitulado "*Chaperones in Polyglutamine Aggregation: Beyond the Q-Stretch*" consistiu em uma revisão da literatura sobre chaperonas moleculares e doenças desencadeadas por tratos poliQ expandidos, como é o caso da DMJ/SCA3. Nesse manuscrito, abordou-se o papel das chaperonas na modulação diferencial de poliglutaminopatias distintas devido às composições diferentes de cada uma das nove proteínas com tratos poliQ ligadas a doenças neurodegenerativas. Realizamos uma extensa revisão da literatura sobre os fatores moleculares determinantes para a modulação da agregação e toxicidade das diversas proteínas poliQ. Além disso, esse artigo explorou a hipótese de que diferenças fenotípicas em poliglutaminopatias distintas possam ser devidas à variabilidade de ação dos sistemas de controle de qualidade proteica (e das chaperonas, especificamente) entre pacientes. De acordo com esse cenário, diferenças no padrão de expressão de certas chaperonas, por exemplo, poderiam ajudar a explicar a variabilidade da AO na DMJ/SCA3 e em outras poliglutaminopatias (Zijlstra et al. 2010).

Essa hipótese foi testada no manuscrito que encerra a presente tese, intitulado "*The molecular chaperone DNAJB6 as a phenotypic modulator of spinocerebellar ataxia type 3/ Machado-Joseph disease*". Nosso objetivo foi estudar o potencial efeito modulador das chaperonas moleculares sobre dois fenótipos da DMJ/SCA3: a AO e a velocidade de progressão da doença. Testamos o efeito de diversas chaperonas, mas houve um foco especial na co-chaperona DNAJB6. Como discutido no capítulo 1 desta tese, estudos recentes sugerem que a DNAJB6 é uma poderosa supressora da agregação e toxicidade de proteínas amiloidogênicas, como é o caso das proteínas com tratos poliQ expandidos (Hageman et al. 2010; Månsson et al. 2014; Kakkar et al. 2016; Reidy et al. 2016; Aprile et al. 2017). Além disso, observações anteriores do grupo holandês colaborador deste manuscrito sugeriram que diferenças nos níveis de certas chaperonas podem ajudar a explicar a variabilidade da AO na DMJ/SCA3 (Zijlstra et al. 2010). Essa publicação demonstrou maiores níveis da chaperona DNAJB1

em fibroblastos de pacientes com AO significativamente mais tardia do que o esperado, em comparação a pacientes com AO na média ou mais precoce do que o esperado.

Desse modo, o referencial teórico do manuscrito 5 foi centrado em dois pontos principais: (i) tanto a DNAJB1 como a DNAJB6 pertencem à mesma família de chaperonas (HSP40/DNAJ) (Hageman and Kampinga 2009), sugerindo mecanismos e rotas de ação potencialmente similares; e (ii) a DNAJB6 é mais efetiva do que a DNAJB1 na inibição da agregação e toxicidade de proteínas poliQ (Hageman et al. 2010). Supondo que os potenciais benefícios advindos dos maiores níveis da DNAJB1 na DMJ/SCA3 (Zijlstra et al. 2010) são relacionados à supressão da agregação e toxicidade da proteína ataxina-3 expandida, é de se esperar que um efeito similar, porém de maior magnitude, ocorra na presença de maiores níveis da DNAJB6.

O ponto de partida do manuscrito 5 foi o recrutamento de pacientes com DMJ/SCA3 da coorte Rio Grande do Sul que eram *outliers* extremos para a AO. Esse método de amostragem de fenótipos extremos é extremamente vantajoso e classicamente utilizado em estudos de associação de condições multifatoriais (Li et al. 2011; Barnett et al. 2013; Peloso et al. 2015). Ao se recrutar indivíduos que estão nos extremos de um espectro fenotípico, aumentam-se as chances de detecção de variantes genéticas raras de alto impacto sobre o fenótipo em questão. De modo análogo, quaisquer diferenças na variabilidade de um fenótipo específico de uma condição monogênica, como a DMJ/SCA3, são potencialmente exacerbadas quando indivíduos nos extremos do fenótipo são comparados. Particularmente na DMJ/SCA3, esse procedimento amostral pode garantir que quaisquer associações detectadas são independentes do grande efeito da mutação causal (CAGexp) sobre a AO (de Castilhos et al. 2014; Tezenas du Montcel et al. 2014a; Raposo et al. 2015; Chen et al. 2016) e, potencialmente, sobre a progressão da doença (Jardim et al. 2010; Donis et al. 2016).

Dado o grande número de pacientes da coorte Rio Grande do Sul (Souza et al. 2016), nosso grupo de pesquisadores pôde recrutar, em pouco mais de um ano,

um número expressivo de pacientes *outliers* para a AO que compreenderam a coorte brasileira do manuscrito 5. Além disso, a colaboração com o grupo holandês possibilitou a utilização das mesmas amostras em que a associação entre DNAJB1 e AO foi descoberta (Zijlstra et al. 2010). Essa coorte holandesa serviu como grupo de validação da coorte brasileira. Como discutido no manuscrito 5, a AO é um parâmetro muito mais simples de ser obtido do que a velocidade de progressão da doença, ainda que ambas sejam mensurações do estado da DMJ/SCA3. O recrutamento de *outliers* para a progressão da doença exigiria o acompanhamento de indivíduos durante um longo intervalo de tempo antes de categorizá-los, o que tornaria o estudo menos viável, pelo menos neste momento. Assim, ainda que os sujeitos tenham sido recrutados com base em AOs extremas, a velocidade da progressão da DMJ/SCA3 foi também mensurada na coorte brasileira pelas escalas neurológicas SARA (Schmitz-Hübsch et al. 2006) e NESSCA (Kieling et al. 2008), o que possibilitou as correlações entre os níveis de chaperonas e a taxa de progressão da doença.

Na coorte holandesa, foi possível validar o achado anteriormente publicado de maiores níveis da DNAJB1 em pacientes com AO mais tardia do que o esperado (Zijlstra et al. 2010). Além disso, maiores níveis da isoforma a da DNAJB6 (DNAJB6a), da HSP60 e da HSPA8 também mostraram correlação significativa com a AO mais tardia em pacientes holandeses. Contudo, nenhuma dessas observações foi replicada na coorte brasileira. Discutiu-se no manuscrito 5 que essa falta de replicação do efeito modificador de chaperonas sobre a AO entre as coortes pode ter diversas causas. O tamanho do CAGexp determina a maior parte da AO na DMJ/SCA3, e o comprimento médio do CAGexp é significativamente maior na coorte Rio Grande do Sul do que nas coortes europeias, como demonstrou a metanálise do artigo 1. Assim, é possível que o efeito das chaperonas sobre a AO tenha sido mascarado na coorte brasileira devido a um efeito mais “dominante” do CAGexp. De fato, alguns autores sugerem que o efeito do CAGexp sobre a AO na DMJ/SCA3 segue uma distribuição quadrática, com comprimentos maiores do CAGexp tendo uma influência muito mais marcante sobre a AO, em relação a tratos CAGexp menores (Tezenas du Montcel et al. 2014a; Chen et al. 2016). Por outro lado, é também possível que o efeito das chaperonas sobre a AO seja população-

específico. Estudos futuros em coortes maiores serão necessários para testar essas hipóteses. De qualquer modo, os dados do manuscrito 5 sugerem um papel importante para as chaperonas moleculares, especialmente DNAJB6a, na determinação da AO na DMJ/SCA3, pelo menos em pacientes de origem europeia.

Contudo, o achado mais interessante do manuscrito 5 foi a forte correlação entre os níveis proteicos da DNAJB6 (tanto da isoforma a como da b) e a taxa de progressão da DMJ/SCA3, medida pela escala SARA, na coorte brasileira. Similarmente, detectou-se uma correlação significativa entre maiores níveis da DNAJB6a e progressão mais lenta, medida pela escala SCAFI. Apesar da baixa amplitude de variação dos níveis proteicos da DNAJB6a e b entre pacientes, do pequeno tamanho amostral e das inconsistências inerentes às escalas clínicas, os níveis da DNAJB6 em fibroblastos foram excelentes preditores da progressão das escalas SARA e SCAFI. Foram discutidas no manuscrito 5 as razões pelas quais isso não foi observado para a NESSCA, a outra escala neurológica utilizada no estudo. Muito provavelmente, a falta de correlação entre níveis da DNAJB6 e a progressão da NESSCA tenha sido devida ao maior número de indivíduos com escores de NESSCA não informativos (ou seja, escore menor no seguimento do que no início do estudo, o que é uma impossibilidade em uma doença neurodegenerativa progressiva e ainda sem tratamento, como a DMJ/SCA3) do que para a SARA e SCAFI. Por outro lado, as escalas SARA, SACFI e NESSCA não são inteiramente complementares, já que a SARA e a SCAFI avaliam apenas sintomas atáxicos (Schmitz-Hübsch et al. 2006; Schmitz-Hübsch et al. 2008), enquanto a NESSCA leva em conta também os sintomas não-atáxicos (Kieling et al. 2008). Entre os sintomas não-atáxicos, sobressaem-se as manifestações extrapiramidais na SCA3/MJD, que sabidamente evoluem bem mais lentamente do que as atáxicas (Jardim et al. 2010; Jacobi et al. 2012). Talvez os efeitos benéficos de maiores níveis da DNAJB6 sejam mais expressivos para os sintomas atáxicos, do que para outros achados clínicos. Ou talvez a progressão dos sintomas atáxicos seja rápida o suficiente para permitir a distinção entre os grupos com mais e menos expressão da DNAJB6. Estudos subsequentes serão necessários para esclarecer essa singularidade.

Assim, a presente tese traz fortes evidências em favor da modulação de fenótipos da DMJ/SCA3 pelas chaperonas moleculares, especialmente pela DNAJB6, uma poderosa supressora de agregação e toxicidade de espécies amiloidogênicas. Esse achado é corroborado pela plausibilidade biológica do efeito da DNAJB6 e de outras chaperonas sobre proteínas com tratos poliQ expandidos (Hageman et al. 2010; Kakkar et al. 2016; Reis et al. 2016). Mais do que isso, o manuscrito 5 argumenta em favor da exploração terapêutica da DNAJB6 e, possivelmente, de outras chaperonas na modulação da DMJ/SCA3. Seja através de intervenções no período assintomático ou na fase sintomática da doença, a DNAJB6 pode ter um efeito muito significativo no retardamento da AO ou na lentificação da progressão, respectivamente, da DMJ/SCA3 e de outras doenças neurodegenerativas com bases moleculares semelhantes.

Em conclusão, essa tese avaliou, através de abordagens distintas, diversos potenciais moduladores do fenótipo na DMJ/SCA3. Uma importante contribuição de fatores população-específicos que modulam a AO na DMJ/SCA3 foi descoberta (artigo 1), testada e validada (artigo 2). Além disso, a variabilidade do gene para a apolipoproteína E (*apoE*) foi significativamente correlacionada com a AO, detectando-se AOs mais precoces entre indivíduos homozigotos para o alelo de risco ε4 (manuscrito 3). Por fim, abordou-se o papel das chaperonas moleculares na determinação da AO e da velocidade de progressão da DMJ/SCA3 (artigo 4 e manuscrito 5). Os dados aqui apresentados argumentam fortemente em favor da manipulação dos níveis de expressão de diversas chaperonas – e da DNAJB6 em particular – para a lentificação da progressão da DMJ/SCA3 e, possivelmente, de outras doenças semelhantes.

Referências Bibliográficas Gerais

Aprile FA, Källstig E, Limorenko G, Vendruscolo M, Ron D and Hansen C (2017) The molecular chaperones DNAJB6 and Hsp70 cooperate to suppress α-synuclein aggregation. Sci Rep 7:1–11.

Arrasate M, Mitra S, Schweitzer ES, Segal MR and Finkbeiner S (2004) Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. Nature 431:805–810.

Balchin D, Hayer-Hartl M and Hartl FU (2016) In vivo aspects of protein folding and quality control. Science 353(6294):aac4354.

Barnett IJ, Lee S and Lin X (2013) Detecting Rare Variant Effects Using Extreme Phenotype Sampling in Sequencing Association Studies. Genet Epidemiol 37:142–151.

Bettencourt C, Raposo M, Kazachkova N, Cymbron T, Santos C, Kay T, Vasconcelos J, Maciel P, Donis KC, Saraiva-Pereira ML et al. (2011) The APOE ε2 allele increases the risk of earlier age at onset in Machado-Joseph disease. Arch Neurol 68:1580–1583.

Bettencourt C, Santos C, Montiel R, Costa MDC, Cruz-Morales P, Santos LR, Simões N, Kay T, Vasconcelos J, Maciel P et al. (2010a) Increased transcript diversity: novel splicing variants of Machado-Joseph disease gene (ATXN3). Neurogenetics 11:193–202.

Bettencourt C, Santos C, Montiel R, Kay T, Vasconcelos J, Maciel P and Lima M (2010b) The (CAG)n tract of Machado-Joseph Disease gene (ATXN3): a comparison between DNA and mRNA in patients and controls. Eur J Hum Genet 18:621–623.

Bird TD. In: Adam MP, Ardinger HH, Pagon RA, Wallace SE, Bean LJH, Stephens K, Amemiya A, editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2018. [updated 2016 Nov 3]

Braten O, Livneh I, Ziv T, Admon A, Kehat I, Caspi LH, Gonen H, Bercovich B, Godzik A, Jahandideh S et al. (2016) Numerous proteins with unique characteristics are degraded by the 26S proteasome following

monoubiquitination. Proc Natl Acad Sci 113(32):E4639-4647.

Brehme M and Voisine C (2016) Model systems of protein-misfolding diseases reveal chaperone modifiers of proteotoxicity. Dis Model Mech 9:823–838.

Bukau B, Weissman J and Horwich A (2006) Molecular chaperones and protein quality control. Cell 125:443–451.

Burnett B, Li F and Pittman RN (2003) The polyglutamine neurodegenerative protein ataxin-3 binds polyubiquitylated proteins and has ubiquitin protease activity. Hum Mol Genet 12:3195–3205.

Carra S, Sivilotti M, Chávez Zobel AT, Lambert H and Landry J (2005) HspB8, a small heat shock protein mutated in human neuromuscular disorders, has in vivo chaperone activity in cultured cells. Hum Mol Genet 14:1659–1669.

Chai Y, Koppenhafer SL, Bonini NM and Paulson HL (1999a) Analysis of the role of heat shock protein (Hsp) molecular chaperones in polyglutamine disease. J Neurosci 19:10338–10347.

Chai Y, Koppenhafer SL, Shoesmith SJ, Perez MK and Paulson HL (1999b) Evidence for proteasome involvement in polyglutamine disease: localization to nuclear inclusions in SCA3/MJD and suppression of polyglutamine aggregation in vitro. Hum Mol Genet 8:673–682.

Chen Z, Zheng C, Long Z, Cao L, Li X, Shang H, Yin X, Zhang B, Liu J, Ding D et al. (2016) (CAG)n loci as genetic modifiers of age-at-onset in patients with Machado-Joseph disease from mainland China. Brain 139:e41.

Ciechanover A and Kwon YT (2017) Protein Quality Control by Molecular Chaperones in Neurodegeneration. Front Neurosci 11:1–18.

Clague MJ, Heride C and Urbé S (2015) The demographics of the ubiquitin system. Trends Cell Biol 25(7):417-426.

Collins GA and Goldberg AL (2017) The Logic of the 26S Proteasome. Cell 169:792–806.

Corder EH, Saunders AM, Risch NJ, Strittmatter WJ, Schmechel DE, Gaskell PC, Rimmler JB, Locke PA, Conneally PM and Schmader KE (1994) Protective effect of apolipoprotein E type 2 allele for late onset Alzheimer disease. Nat

Genet 7:180–184.

Coutinho P, Guimarães A, Pires MM and Scaravilli F (1986) The peripheral neuropathy in Machado-Joseph disease. Acta Neuropathol 71:119–124.

da Silva Carvalho G, Saute JAM, Haas CB, Torrez VR, Brochier AW, Souza GN, Furtado GV, Gheno T, Russo A, Monte TL et al. (2016) Cytokines in Machado Joseph Disease/Spinocerebellar Ataxia 3. Cerebellum 15:518–525.

Dantuma NP and Bott LC (2014) The ubiquitin-proteasome system in neurodegenerative diseases: precipitating factor, yet part of the solution. Front Mol Neurosci 7:1–18.

de Castilhos RM, Furtado GV, Gheno TC, Schaeffer P, Russo A, Barsottini O, Pedroso JL, Salarini DZ, Vargas FR, de Lima MA de FD et al. (2014) Spinocerebellar ataxias in Brazil--frequencies and modulating effects of related genes. Cerebellum 13:17–28.

Dikic I (2017) Proteasomal and Autophagic Degradation Systems. Annu Rev Biochem 86:193-224.

Donis KC, Saute JAM, Krum-Santos AC, Furtado GV, Mattos EP, Saraiva-Pereira ML, Torman VL and Jardim LB (2016) Spinocerebellar ataxia type 3/Machado-Joseph disease starting before adolescence. Neurogenetics 17:107–113.

Doyle SM, Genest O and Wickner S (2013) Protein rescue from aggregates by powerful molecular chaperone machines. Nat Rev Mol Cell Biol 14:617–629.

Dürr A, Stevanin G, Cancel G, Duyckaerts C, Abbas N, Didierjean O, Chneiweiss H, Benomar A, Lyon-Caen O, Julien J et al. (1996) Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. Ann Neurol 39:490–499.

Emamzadeh FN, Aojula H, McHugh PC and Allsop D (2016) Effects of different isoforms of apoE on aggregation of the α-synuclein protein implicated in Parkinson’s disease. Neurosci Lett 618:146–151.

Escusa-Toret S, Vonk WIM and Frydman J (2013) Spatial sequestration of misfolded proteins by a dynamic chaperone pathway enhances cellular fitness during stress. Nat Cell Biol 15:1231–1243.

Frontali M, Sabbadini G, Novelletto A, Jodice C, Naso F, Spadaro M, Giunti P, Jacopini AG, Veneziano L, Mantuano E et al. (1996) Genetic fitness in Huntington's Disease and Spinocerebellar Ataxia 1: a population genetics model for CAG repeat expansions. Ann Hum Genet 60:423–435.

Fujigasaki H, Uchihara T, Koyano S, Iwabuchi K, Yagishita S, Makifuchi T, Nakamura a, Ishida K, Toru S, Hirai S et al. (2000) Ataxin-3 is translocated into the nucleus for the formation of intranuclear inclusions in normal and Machado-Joseph disease brains. Exp Neurol 165:248–256.

Garrido C, Paul C, Seigneuric R and Kampinga HH (2012) The small heat shock proteins family: the long forgotten chaperones. Int J Biochem Cell Biol 44:1588–1592.

Gaspar C, Lopes-Cendes I, Hayes S, Goto J, Arvidsson K, Dias a, Silveira I, Maciel P, Coutinho P, Lima M et al. (2001) Ancestral origins of the Machado-Joseph disease mutation: a worldwide haplotype study. Am J Hum Genet 68:523–528.

Gillis J, Schipper-Krom S, Juenemann K, Gruber A, Coolen S, van den Nieuwendijk R, van Veen H, Overkleeft H, Goedhart J, Kampinga HH et al. (2013) The DNAJB6 and DNAJB8 protein chaperones prevent intracellular aggregation of polyglutamine peptides. J Biol Chem 288:17225–17237.

Globas C, du Montcel ST, Baliko L, Boesch S, Depondt C, DiDonato S, Durr A, Filla A, Klockgether T, Mariotti C et al. (2008) Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. Mov Disord 23:2232–2238.

Goto J, Watanabe M, Ichikawa Y, Yee S-B, Ihara N, Endo K, Igarashi S, Takiyama Y, Gaspar C, Maciel P et al. (1997) Machado–Joseph disease gene products carrying different carboxyl termini. Neurosci Res 28:373–377.

Gusella JF and MacDonald ME (2000) Molecular genetics: unmasking polyglutamine triggers in neurodegenerative disease. Nat Rev Neurosci 1:109–115.

Hageman J and Kampinga HH (2009) Computational analysis of the human HSPH/HSPA/DNAJ family and cloning of a human HSPH/HSPA/DNAJ expression library. Cell Stress Chaperones 14:1–21.

Hageman J, Rujano M a, van Waarde M a WH, Kakkar V, Dirks RP, Govorukhina N, Oosterveld-Hut HMJ, Lubsen NH and Kampinga HH (2010) A DNAJB chaperone subfamily with HDAC-dependent activities suppresses toxic protein aggregation. Mol Cell 37:355–369.

Hara T, Nakamura K, Matsui M, Yamamoto A, Nakahara Y, Suzuki-Migishima R, Yokoyama M, Mishima K, Saito I, Okano H et al. (2006) Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice. Nature 441:885–889.

Hetz C, Chevet E and Oakes SA. (2015) Proteostasis control by the unfolded protein response. Nat Cell Biol 17:829–838.

Huynh T-P V., Davis AA, Ulrich JD and Holtzman DM (2017a) Apolipoprotein E and Alzheimer's disease: the influence of apolipoprotein E on amyloid-β and other amyloidogenic proteins. J Lipid Res 58:824–836.

Huynh T-P V., Liao F, Francis CM, Robinson GO, Serrano JR, Jiang H, Roh J, Finn MB, Sullivan PM, Esparza TJ et al. (2017b) Age-Dependent Effects of apoE Reduction Using Antisense Oligonucleotides in a Model of β-amyloidosis. Neuron 96:1013–1023.e4.

Ichikawa Y, Goto J, Hattori M, Toyoda a, Ishii K, Jeong SY, Hashida H, Masuda N, Ogata K, Kasai F et al. (2001) The genomic structure and expression of MJD, the Machado-Joseph disease gene. J Hum Genet 46:413–422.

Igarashi S, Takiyama Y, Cancel G, Rogaeva E a, Sasaki H, Wakisaka a, Zhou YX, Takano H, Endo K, Sanpei K et al. (1996) 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 5:923–932.

Ikeuchi T, Igarashi S, Takiyama Y, Onodera O, Oyake M, Takano H, Koide R, Tanaka H and Tsuji S (1996) Non-Mendelian transmission in dentatorubral-pallidoluysian atrophy and Machado-Joseph disease: the mutant allele is preferentially transmitted in male meiosis. Am J Hum Genet 58:730–733.

Infante J, Combarros O, Volpini V, Corral J, Llorca J and Berciano J (2005)

Autosomal dominant cerebellar ataxias in Spain: Molecular and clinical correlations, prevalence estimation and survival analysis. Acta Neurol Scand 111:391–399.

Jacobi H, Bauer P, Giunti P, Labrum R, Sweeney MG, Charles P, Dürr A, Marelli C, Globas C, Linnemann C et al. (2011) The natural history of spinocerebellar ataxia type 1, 2, 3, and 6: a 2-year follow-up study. Neurology 77:1035–1041.

Jacobi H, Hauser T-K, Giunti P, Globas C, Bauer P, Schmitz-Hübsch T, Baliko L, Filla A, Mariotti C, Rakowicz M et al. (2012) Spinocerebellar ataxia types 1, 2, 3 and 6: the clinical spectrum of ataxia and morphometric brainstem and cerebellar findings. Cerebellum 11:155–166.

Jardim LB, Hauser L, Kieling C, Saute JAM, Xavier R, Rieder CRM, Monte TL, Camey S and Torman VBL (2010) Progression rate of neurological deficits in a 10-year cohort of SCA3 patients. Cerebellum 9:419–428.

Jardim LB, Pereira ML, Silveira I, Ferro a, Sequeiros J and Giugliani R (2001a) Machado-Joseph disease in South Brazil: clinical and molecular characterization of kindreds. Acta Neurol Scand 104:224–231.

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

Kaganovich D, Kopito R and Frydman J (2008) Misfolded proteins partition between two distinct quality control compartments. Nature 454:1088–1095.

Kakkar V, Månsson C, de Mattos EP, Bergink S, van der Zwaag M, van Waarde MAWH, Kloosterhuis NJ, Melki R, van Cruchten RTP, Al-Karadaghi S et al. (2016) The S/T-Rich Motif in the DNAJB6 Chaperone Delays Polyglutamine Aggregation and the Onset of Disease in a Mouse Model. Mol Cell 62:272–283.

Kakkar V, Prins LCB and Kampinga HH (2012) DNAJ proteins and protein aggregation diseases. Curr Top Med Chem 12:2479–2490.

Kampinga HH and Craig EA (2010) The HSP70 chaperone machinery: J proteins as drivers of functional specificity. Nat Rev Mol Cell Biol 11:579–592.

Kampinga HH and Garrido C (2012) HSPBs: small proteins with big implications in

human disease. Int J Biochem Cell Biol 44:1706–1710.

Kaur J and Debnath J (2015) Autophagy at the crossroads of catabolism and anabolism. Nat Rev Mol Cell Biol 16:461–472.

Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M and Akiguchi I (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. Nat Genet 8:221–228.

Khaminets A, Behl C and Dikic I (2015) Ubiquitin-Dependent And Independent Signals In Selective Autophagy. Trends Cell Biol 26:6–16.

Kieling C, Rieder CRM, Silva ACF, Saute JAM, Cecchin CR, Monte TL and Jardim LB (2008) A neurological examination score for the assessment of spinocerebellar ataxia 3 (SCA3). Eur J Neurol 15:371–376.

Kim M, Sandford E, Gatica D, Qiu Y, Liu X, Zheng Y, Schulman BA, Xu J, Semple I, Ro SH et al. (2016a) Mutation in ATG5 reduces autophagy and leads to ataxia with developmental delay. Elife 5:1–18.

Kim YE, Hosp F, Frottin F, Ge H, Mann M, Hayer-Hartl M and Hartl FU (2016b) Soluble Oligomers of PolyQ-Expanded Huntingtin Target a Multiplicity of Key Cellular Factors. Mol Cell 63:950–964.

Komatsu M, Waguri S, Chiba T, Murata S, Iwata JI, Tanida I, Ueno T, Koike M, Uchiyama Y, Kominami E et al. (2006) Loss of autophagy in the central nervous system causes neurodegeneration in mice. Nature 441:880–884.

Kraut DA, Israeli E, Schrader EK, Patil A, Nakai K, Nanavati D, Inobe T and Matouschek A (2012) Sequence- and species-dependence of proteasomal processivity. ACS Chem Biol 7:1444–1453.

Kuiper EFE, de Mattos EP, Jardim LB, Kampinga HH and Bergink S (2017) Chaperones in Polyglutamine Aggregation: Beyond the Q-Stretch. Front Neurosci 11:145.

Langbehn DR, Brinkman RR, Falush D, Paulsen JS, Hayden MR and International Huntington's Disease Collaborative Group (2004) A new model for prediction of the age of onset and penetrance for Huntington's disease based on CAG length.

Clin Genet 65:267–277.

Langbehn DR, Hayden MR, Paulsen JS and and the PREDICT-HD Investigators of the Huntington Study Group (2010) CAG-repeat length and the age of onset in Huntington disease (HD): a review and validation study of statistical approaches. Am J Med Genet B Neuropsychiatr Genet 153B:397–408.

Lautner R, Insel PS, Skillbäck T, Olsson B, Landén M, Frisoni GB, Herukka SK, Hampel H, Wallin A, Minthon L et al. (2017) Preclinical effects of APOE ϵ4 on cerebrospinal fluid Aβ42 concentrations. Alzheimer's Res Ther 9:1–7.

Li D, Lewinger JP, Gauderman WJ, Murcray CE and Conti D (2011) Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. Genet Epidemiol 35:790–799.

Lilienbaum A (2013) Relationship between the proteasomal system and autophagy. Int J Biochem Mol Biol 31;4(1):1-26.

Lim J and Yue Z (2015) Neuronal Aggregates: Formation, Clearance, and Spreading. Dev Cell 32:491–501.

Limprasert P, Nouri N, Heyman R a, Nopparatana C, Kamonsilp M, Deininger PL and Keats BJ (1996) Analysis of CAG repeat of the Machado-Joseph gene in human, chimpanzee and monkey populations: a variant nucleotide is associated with the number of CAG repeats. Hum Mol Genet 5:207–213.

Liu C-C, Zhao N, Fu Y, Wang N, Linares C, Tsai C-W and Bu G (2017) ApoE4 Accelerates Early Seeding of Amyloid Pathology. Neuron 96:1024–1032.e3.

Luo L, Wang J, Lo RY, Figueroa KP, Pulst SM, Kuo PH, Perlman S, Wilmot G, Gomez CM, Schmahmann JD et al. (2017) The Initial Symptom and Motor Progression in Spinocerebellar Ataxias. Cerebellum 16:615–622.

Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, Dawson DM, Sudarsky L, Guimarães J and Loureiro JE (1995) Correlation between CAG repeat length and clinical features in Machado-Joseph disease. Am J Hum Genet 57:54–61.

Maciel P, Gaspar C, Guimarães L, Goto J, Lopes-Cendes I, Hayes S, Arvidsson K, Dias A, Sequeiros J, Sousa A et al. (1999) 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 7:147–156.

Malinovska L, Kroschwald S, Munder MC, Richter D and Alberti S (2012) Molecular chaperones and stress-inducible protein-sorting factors coordinate the spatiotemporal distribution of protein aggregates. Mol Biol Cell 23:3041–3056.

Mannini B and Chiti F (2017) Chaperones as Suppressors of Protein Misfolded Oligomer Toxicity. Front Mol Neurosci 10:98.

Månsson C, Arosio P, Hussein R, Kampinga HH, Hashem RM, Boelens WC, Dobson CM, Knowles TPJ, Linse S and Emanuelsson C (2014) Interaction of the molecular chaperone DNAJB6 with growing amyloid-beta 42 (Aβ42) aggregates leads to sub-stoichiometric inhibition of amyloid formation. J Biol Chem 289:31066–31076.

Månsson C, Kakkar V, Monsellier E, Sourigues Y, Härmark J, Kampinga HH, Melki R and Emanuelsson C (2013) DNAJB6 is a peptide-binding chaperone which can suppress amyloid fibrillation of polyglutamine peptides at substoichiometric molar ratios. Cell Stress Chaperones 19(2):227-239.

Martin DDO, Ladha S, Ehrnhoefer DE and Hayden MR (2014) Autophagy in Huntington disease and huntingtin in autophagy. Trends Neurosci 38(1):26-35.

Martins S, Calafell F, Gaspar C, Wong VCN, Silveira I, Nicholson G a, Brunt ER, Tranebjaerg L, Stevanin G, Hsieh M et al. (2007) Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. Arch Neurol 64:1502–1508.

Martins S, Coutinho P, Silveira I, Giunti P, Jardim LB, Calafell F, Sequeiros J and Amorim A (2008) Cis-acting factors promoting the CAG intergenerational instability in Machado-Joseph disease. Am J Med Genet B Neuropsychiatr Genet 147B:439–446.

Martins S, Pearson CE, Coutinho P, Provost S, Amorim A, Dubé M-P, Sequeiros J and Rouleau GA (2014) Modifiers of (CAG)(n) instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. Hum Genet 133:1311–1318.

Martins S, Soong B-W, Wong VCN, Giunti P, Stevanin G, Ranum LPW, Sasaki H, Riess O, Tsuji S, Coutinho P et al. (2012) Mutational origin of Machado-Joseph disease in the Australian Aboriginal communities of Groote Eylandt and Yirrkala. Arch Neurol 69:746–751.

Matilla-Dueñas A, Ashizawa T, Brice A, Magri S, McFarland KN, Pandolfo M, Pulst SM, Riess O, Rubinsztein DC, Schmidt J et al. (2014) Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. Cerebellum 13:269–302.

Matsumura R, Takayanagi T, Murata K, Futamura N, Hirano M and Ueno S (1996) Relationship of (CAG)nC configuration to repeat instability of the Machado-Joseph disease gene. Hum Genet 98:643–645.

Mittal U, Srivastava AK, Jain S, Jain S and Mukerji M (2005) Founder haplotype for Machado-Joseph disease in the Indian population: novel insights from history and polymorphism studies. Arch Neurol 62:637–640.

Nagai Y, Inui T, Popiel HA, Fujikake N, Hasegawa K, Urade Y, Goto Y, Naiki H and Toda T (2007) A toxic monomeric conformer of the polyglutamine protein. Nat Struct Mol Biol 14:332–340.

Nakano KK, Dawson DM and Spence A (1972) Machado disease. A hereditary ataxia in Portuguese emigrants to Massachusetts. Neurology 22:49–55.

Nillegoda NB, Kirstein J, Szlachcic A, Berynskyy M, Stank A, Stengel F, Arnsburg K, Gao X, Scior A, Aebersold R et al. (2015) Crucial HSP70 co-chaperone complex unlocks metazoan protein disaggregation. Nature 524(7564):247-251.

Nishiyama K, Murayama S, Goto J, Watanabe M, Hashida H, Katayama S, Nomura Y, Nakamura S and Kanazawa I (1996) Regional and cellular expression of the Machado-Joseph disease gene in brains of normal and affected individuals. Ann Neurol 40:776–781.

Nuriel T, Peng KY, Ashok A, Dillman AA, Figueroa HY, Apuzzo J, Ambat J, Levy E, Cookson MR, Mathews PM et al. (2017) The Endosomal–Lysosomal Pathway Is Dysregulated by APOE4 Expression in Vivo. Front Neurosci 11:702.

Opal P and Zoghbi HY (2002) The role of chaperones in polyglutamine disease.

Trends Mol Med 8:232–236.

Orr HT (2012) Cell biology of spinocerebellar ataxia. J Cell Biol 197:167–177.

Øverbye A, Fengsrud M and Seglen PO (2007) Proteomic analysis of membrane-associated proteins from rat liver autophagosomes. Autophagy 3:300–322.

Paulson HL (1999) Protein fate in neurodegenerative proteinopathies: polyglutamine diseases join the (mis)fold. Am J Hum Genet 64:339–345.

Paulson HL, Das SS, Crino PB, Perez MK, Patel SC, Gotsdiner D, Fischbeck KH and Pittman RN (1997a) Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. Ann Neurol 41:453–462.

Paulson HL, Perez MK, Trottier Y, Trojanowski JQ, Subramony SH, Das SS, Vig P, Mandel JL, Fischbeck KH and Pittman RN (1997b) Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. Neuron 19:333–344.

Peloso GM, Rader DJ, Gabriel S, Kathiresan S, Daly MJ and Neale BM (2015) Phenotypic extremes in rare variant study designs. Eur J Hum Genet 24(6):924-30.

Peng H, Wang C, Chen Z, Sun Z, Jiao B, Li K, Huang F, Hou X, Wang J, Shen L et al. (2014) The 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 35:2179.e15-8.

Prestes PR, Saraiva-Pereira ML, Silveira I, Sequeiros J and Jardim LB (2008) Machado-Joseph disease enhances genetic fitness: a comparison between affected and unaffected women and between MJD and the general population. Ann Hum Genet 72:57–64.

Ranum LP, Lundgren JK, Schut LJ, Ahrens MJ, Perlman S, Aita J, Bird TD, Gomez C and Orr HT (1995) Spinocerebellar ataxia type 1 and Machado-Joseph disease: incidence of CAG expansions among adult-onset ataxia patients from 311 families with dominant, recessive, or sporadic ataxia. Am J Hum Genet 57:603–608.

Raposo M, Ramos A, Bettencourt C and Lima M (2015) Replicating studies of

genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid? Brain 138:e398.

Reidy M, Sharma R, Roberts B-L and Masison DC (2016) Human J-protein DnaJB6b Cures a Subset of Saccharomyces cerevisiae Prions and Selectively Blocks Assembly of Structurally Related Amyloids. J Biol Chem 291:4035–4047.

Reis SD, Pinho BR and Oliveira JMA (2016) Modulation of Molecular Chaperones in Huntington's Disease and Other Polyglutamine Disorders. Mol Neurobiol 54(8):5829-5854.

Riess O, Rüb U, Pastore A, Bauer P and Schöls L (2008) SCA3: neurological features, pathogenesis and animal models. Cerebellum 7:125–137.

Robert J, Button EB, Yuen B, Gilmour M, Kang K, Bahrabadi A, Stukas S, Zhao W, Kulic I and Wellington CL (2017) Clearance of beta-amyloid is facilitated by apolipoprotein E and circulating high-density lipoproteins in bioengineered human vessels. Elife 6. pii: e29595.

Rose JM, Novoselov SS, Robinson PA and Cheetham ME (2011) Molecular chaperone-mediated rescue of mitophagy by a Parkin RING1 domain mutant. Hum Mol Genet 20:16–27.

Rosenberg RN, Nyhan WL, Bay C and Shore P (1976) Autosomal dominant striatonigral degeneration. A clinical, pathologic, and biochemical study of a new genetic disorder. Neurology 26:703–714.

Rüb U, Schöls L, Paulson H, Auburger G, Kermer P, Jen JC, Seidel K, Korf H-W and Deller T (2013) Clinical features, neurogenetics and neuropathology of the polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7. Prog Neurobiol 104:38–66.

Rubinsztein DC (2007) Autophagy Induction Rescues Toxicity Mediated by Proteasome Inhibition. Neuron 54:854–856.

Saibil H (2013) Chaperone machines for protein folding, unfolding and disaggregation. Nat Rev Cancer 13:630–642.

Saute JAM, da Silva ACF, Muller AP, Hansel G, de Mello AS, Maeda F, Vedolin L, Saraiva-Pereira ML, Souza DO, Arpa J et al. (2011) Serum insulin-like system

alterations in patients with spinocerebellar ataxia type 3. Mov Disord 26:731–735.

Saute JAM, Donis KC, Serrano-Munuera C, Genis D, Ramirez LT, Mazzetti P, Pérez LV, Latorre P, Sequeiros J, Matilla-Dueñas A et al. (2012a) Ataxia rating scales--psychometric profiles, natural history and their application in clinical trials. Cerebellum 11:488–504.

Saute JAM and Jardim LB (2015) Machado Joseph disease: clinical and genetic aspects, and current treatment. Expert Opin Orphan Drugs 3(5):517–535.

Saute JAM, Silva ACF Da, Souza GN, Russo AD, Donis KC, Vedolin L, Saraiva-Pereira ML, Portela LVC and Jardim LB (2012b) Body mass index is inversely correlated with the expanded CAG repeat length in SCA3/MJD patients. Cerebellum 11:771–774.

Schmidt T, Lindenberg KS, Krebs A, Schöls L, Laccone F, Herms J, Rechsteiner M, Riess O and Landwehrmeyer GB (2002) Protein surveillance machinery in brains with spinocerebellar ataxia type 3: redistribution and differential recruitment of 26S proteasome subunits and chaperones to neuronal intranuclear inclusions. Ann Neurol 51:302–310.

Schmitt I, Evert BO, Khazneh H, Klockgether T and Wuellner U (2003) The human MJD gene: genomic structure and functional characterization of the promoter region. Gene 314:81–88.

Schmitt I, Linden M, Khazneh H, Evert BO, Breuer P, Klockgether T and Wuellner U (2007) Inactivation of the mouse Atxn3 (ataxin-3) gene increases protein ubiquitination. Biochem Biophys Res Commun 362:734–739.

Schmitz-Hübsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, Giunti P, Globas C, Infante J, Kang J-S et al. (2006) Scale for the assessment and rating of ataxia: development of a new clinical scale. Neurology 66:1717–1720.

Schöls L, Amoiridis G, Büttner T, Przuntek H, Epplen JT and Riess O (1997) Autosomal dominant cerebellar ataxia: phenotypic differences in genetically defined subtypes? Ann Neurol 42:924–932.

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

Sequeiros J, Martins S and Silveira I (2012) Epidemiology and population genetics of degenerative ataxias. Handb Clin Neurol 103:227–251.

Siebert M, Donis KC, Socal M, Rieder CRM, Emmel VE, Vairo F, Michelin-Tirelli K, França M, D'Abreu AC, Bettencourt C et al. (2012) Glucocerebrosidase gene variants in parkinsonian patients with Machado Joseph/spinocerebellar ataxia 3. Parkinsonism Relat Disord 18:185–190.

Sontag EM, Samant RS and Frydman J (2017) Mechanisms and Functions of Spatial Protein Quality Control. Annu Rev Biochem 86:97–122.

Soong BW and Liu RS (1998) Positron emission tomography in asymptomatic gene carriers of Machado-Joseph disease. J Neurol Neurosurg Psychiatry 64:499–504.

Souza GN, Kersting N, Krum-Santos AC, Santos ASP, Furtado G V, Pacheco D, Gonçalves TA, Saute JA, Schuler-Faccini L, Mattos EP et al. (2016) Spinocerebellar ataxia type 3/Machado-Joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. Clin Genet 90:134–140.

Sudarsky L and Coutinho P (1995) Machado-Joseph disease. Clin Neurosci 3:17–22.

Tait D, Riccio M, Sittler A, Scherzinger E, Santi S, Ognibene A, Maraldi NM, Lehrach H and Wanker EE (1998) Ataxin-3 is transported into the nucleus and associates with the nuclear matrix. Hum Mol Genet 7:991–997.

Takahashi T, Katada S and Onodera O (2010) Polyglutamine diseases: where does toxicity come from? what is toxicity? where are we going? J Mol Cell Biol 2:180–191.

Takahashi T, Kikuchi S, Katada S, Nagai Y, Nishizawa M and Onodera O (2008) Soluble polyglutamine oligomers formed prior to inclusion body formation are cytotoxic. Hum Mol Genet 17:345–356.

Takiyama Y, Sakoe K, Soutome M, Namekawa M, Ogawa T, Nakano I, Igarashi S, Oyake M, Tanaka H, Tsuji S et al. (1997) 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 6:1063–1068.

Tezenas du Montcel S, Charles P, Goizet C, Marelli C, Ribai P, Vincitorio C, Anheim M, Guyant-Maréchal L, Le Bayon A, Vandenberghe N et al. (2012) Factors influencing disease progression in autosomal dominant cerebellar ataxia and spastic paraplegia. Arch Neurol 69:500–508.

Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, Forlani S, Rakowicz M, Schöls L, Mariotti C et al. (2014a) Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain 137:2444–2455.

Tezenas du Montcel S, Durr A, Rakowicz M, Nanetti L, Charles P, Sulek A, Mariotti C, Rola R, Schols L, Bauer P et al. (2014b) Prediction of the age at onset in spinocerebellar ataxia type 1, 2, 3 and 6. J Med Genet 51:479–486.

Todi S V, Scaglione KM, Blount JR, Basrur V, Conlon KP, Pastore A, Elenitoba-Johnson K and Paulson HL (2010) Activity and cellular functions of the deubiquitinating enzyme and polyglutamine disease protein ataxin-3 are regulated by ubiquitination at lysine 117. J Biol Chem 285:39303–39313.

Trott A, Jardim LB, Ludwig HT, Saute JAM, Artigalás O, Kieling C, Wanderley HYC, Rieder CRM, Monte TL, Socal M et al. (2006) Spinocerebellar ataxias in 114 Brazilian families: clinical and molecular findings. Clin Genet 70:173–176.

Tsuang D, Leverenz JB, Lopez OL, Hamilton RL, Bennett DA, Schneider JA, Buchman AS, Larson EB, Crane PK, Kaye JA et al. (2013) APOE ε4 increases risk for dementia in pure synucleinopathies. JAMA Neurol 70:223–228.

van de Warrenburg BPC, Hendriks H, Dürr A, van Zuijlen MCA, Stevanin G, Camuzat A, Sinke RJ, Brice A and Kremer BPH (2005) Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort. Ann Neurol 57:505–512.

Venkatraman P, Wetzel R, Tanaka M, Nukina N and Goldberg AL (2004) Eukaryotic proteasomes cannot digest polyglutamine sequences and release them during degradation of polyglutamine-containing proteins. Mol Cell 14:95–104.

Williams AJ and Paulson HL (2008) Polyglutamine neurodegeneration: protein misfolding revisited. Trends Neurosci 31:521–528.

Wu C, Chen D-B, Feng L, Zhou X-X, Zhang J-W, You H-J, Liang X-L, Pei Z and Li X-H (2017a) Oculomotor deficits in spinocerebellar ataxia type 3: Potential biomarkers of preclinical detection and disease progression. CNS Neurosci Ther 23:321–328.

Wu X, Liao X, Zhan Y, Cheng C, Shen W, Huang M, Zhou Z, Wang Z, Qiu Z, Xing W et al. (2017b) Microstructural Alterations in Asymptomatic and Symptomatic Patients with Spinocerebellar Ataxia Type 3: A Tract-Based Spatial Statistics Study. Front Neurol 8:714.

Zhou Q, Ni W, Dong Y, Wang N, Gan S-R and Wu Z-Y (2014) The role of apolipoprotein E as a risk factor for an earlier age at onset for Machado-Joseph disease is doubtful. PLoS One 9:e111356.

Zijlstra MP, Rujano MA, Van Waarde MA, Vis E, Brunt ER and Kampinga HH (2010) Levels of DNAJB family members (HSP40) correlate with disease onset in patients with spinocerebellar ataxia type 3. Eur J Neurosci 32:760–770.

Zoghbi HY and Botas J (2002) Mouse and fly models of neurodegeneration. Trends Genet 18:463–471.