

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

AMANDA DE SOUZA

**DOENÇA DE HUNTINGTON: VARIANTE NO GENE DO SISTEMA DE REPARO
DO DNA COMO POSSÍVEL MODIFICADOR DA IDADE DE INÍCIO**

Porto Alegre, 2021

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA DO ESTADO DO RIO GRANDE DO SUL
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Doença de Huntington: variante no gene do sistema de reparo do DNA como
possível modificador da idade de início

Dissertação submetida ao Programa
de Pós-Graduação em Biologia
Celular e Molecular do Centro de
Biotecnologia da UFRGS como
requisito parcial para obtenção do
título de Mestre em Ciências.

Amanda de Souza

Orientadora: Dra. Maria Luiza Saraiva Pereira

Porto Alegre, maio de 2021

INSTITUIÇÃO E FONTES FINANCIADORAS

Este trabalho foi realizado no Laboratório de Neurogenética Translacional (NeuroGeT) localizado no Centro de Pesquisa Experimental do Hospital de Clínicas de Porto Alegre (HCPA) em parceria com o Serviço de Genética Médica, contando com apoio financeiro do Fundo de Incentivo à Pesquisa e Eventos – Hospital de Clínicas de Porto Alegre e da Fundação de Amparo à Pesquisa do Rio Grande do Sul (FAPERGS). Durante o período de realização do trabalho a bolsa de estudos foi concedida pela Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES).

AGRADECIMENTOS

Agradeço, primeiramente e diariamente ao Universo, por todas as possibilidades, pela disposição para ir atrás de meus sonhos e por me manter firme e forte para sempre seguir em frente.

Agradeço infinitamente a minha família. Toda minha trajetória foi facilitada e ancorada por eles, em todos os momentos, em todas as situações. Ao meu pai, Beto, por ser o meu maior parceiro, incentivador, exemplo e minha maior referência de vida. Por sempre fazer o possível e impossível para que eu alcance meus sonhos e objetivos e, principalmente, por me ensinar e me guiar no caminho da luz, da bondade e da esperança. Agradeço a minha mãe, Mônica, por sempre me motivar a ser uma pessoa melhor, por todo apoio ao longo de todos esses anos e por ser minha maior referência e maior exemplo de força e resiliência. Ao meu irmão, Arthur, por ser meu melhor amigo, meu companheiro de tantos momentos, por ser o maior e melhor presente que a vida me deu. A conclusão desta etapa, bem como todas as minhas conquistas são por vocês e para vocês!

Agradeço à minha orientadora, Professora Maria Luiza Saraiva-Pereira, pela oportunidade, por toda confiança e pelos inúmeros ensinamentos. Tais lições vão muito além da bancada e me impulsionam a ser uma pessoa melhor. Obrigada por toda paciência, por me incentivar a superar meus medos e obstáculos e, principalmente, por abrir e iluminar meus caminhos, por me nortear e acalmar em momentos agitados. Obrigada por tudo, obrigada por tanto!

Agradeço a todos os colegas e membros do laboratório NeuroGeT por toda ajuda, risadas e parceria. Em especial a Rafaella Mergener, por toda ajuda, carinho e amizade. Uma das flores mais lindas do meu jardim e o melhor presente que o lab. poderia ter me dado, essa realização não seria possível sem você. Ao Rafael, Luís, Ana Carolina e Lucas por tornarem meus dias mais coloridos e divertidos. Um agradecimento muito especial a estatística Vanessa, pela assistência e ajuda neste trabalho. Agradeço a todas as pessoas, familiares, amigos, que cruzaram meu caminho ao longo desta jornada e que, direta ou indiretamente, agregaram, torceram e possibilitaram meus passos ao longo desta etapa. Por fim, obrigada à vida que me inspira, me renova e me dá chances de evoluir diariamente.

LISTA DAS PRINCIPAIS ABREVIATURAS

A	Adenina
AD	Autossômica dominante
AO	Idade de início (<i>Age of onset</i>)
<i>ATN1</i>	Gene da Atrofina 1 (<i>Atrophin 1</i>)
<i>ATXN1</i>	Gene da Ataxina 1 (<i>Ataxin 1</i>)
<i>ATXN2</i>	Gene da Ataxina 2 (<i>Ataxin 2</i>)
<i>ATXN3</i>	Gene da Ataxina 3 (<i>Ataxin 3</i>)
<i>ATXN7</i>	Gene da Ataxina 7 (<i>Ataxin 7</i>)
C	Citosina
<i>CACNA1A</i>	Subunidade 1 α do canal de cálcio voltagem dependente do tipo P/Q (<i>Calcium channel, voltage dependent, P/Q type, α1A subunit</i>)
DH	Doença de Huntington
DNA	Ácido desoxirribonucleico (<i>Deoxyribonucleic acid</i>)
DRPLA	Atrofia dentatorubro-palidoluisiana (<i>Dentatorubral-pallidoluisian atrophy</i>)
EHW	Equilíbrio de Hardy-Weinberg
FA	Anemia Fanconi (<i>Fanconi anemia</i>)
G	Guanina
GWAS	Estudos de associação genômica (<i>Genome-wide association studies</i>)
Htt	Proteína Huntingtina
<i>HTT</i>	Gene da huntingtina (<i>HTT gene</i>)
ICL	Ligações cruzadas intercadeia (<i>Interstrand crosslinks</i>)
iPS	Células tronco pluripotente induzidas (<i>Induced pluripotent stem cells</i>)
kb	Kilobase
kDA	Kilodalton
MAF	Frequência do alelo menor (<i>Minor allele frequency</i>)
MJD	Doença de Machado-Joseph (<i>Machado-Joseph disease</i>)
MTMR10	Proteína 10 relacionada à miotubularina (myotubularin-related protein 10)

Poli-Q	Poliglutamina
R ²	Coeficiente de determinação
RA	Receptor de androgênio (<i>Androgen receptor</i>)
SBMA	Atrofia muscular espinobulbar (<i>Spinal-bulbar muscular atrophy</i>)
SCA	Ataxia espinocerebelar (<i>Spinocerebellar ataxia</i>)
SNV	Variante de nucleotídeo simples (<i>Single nucleotide variant</i>)
TBP	Proteína ligadora de TATA-box (<i>TATA-box binding protein</i>)

RESUMO

A doença de Huntington (DH) é um distúrbio neurodegenerativo raro de início tardio, de herança autossômica dominante, cujos sintomas incluem perda da coordenação motora, alterações psiquiátricas e demência progressiva. A DH é causada por uma expansão do trinucleotídeo CAG no éxon 1 do gene *HTT*. A determinação da idade de início dos sintomas é parcialmente associada ao número de repetições CAG, sendo o restante associado a fatores genéticos e ambientais. Dentre os fatores que podem influenciar a idade de início da doença, o alelo C da variante rs3512, localizada no gene *FAN1*, foi associado com um atraso de 1,2 a 1,5 anos na idade de início esperada da doença conforme o tamanho da expansão CAG em um estudo europeu. Baseado nessas informações, o objetivo deste trabalho foi identificar a variante rs3512 em um grupo de pacientes com DH e avaliar a possível influência na idade de início desses pacientes. Na população em estudo, 57 pacientes foram avaliados e o número de repetições CAG no alelo mutante do gene *HTT* foi responsável por 56,1% da variação da idade de início ($p < 0,001$). A distribuição das frequências alélicas ($p = 0,645$) e genotípicas ($p = 0,491$) foram estabelecidas no grupo de pacientes com DH e comparadas com os controles. No grupo de pacientes, a frequência do alelo C foi de 18,4% e, no grupo controle foi de 22,0%. O alelo G apresentou uma frequência de 81,6% no grupo de pacientes e de 78,0% no grupo controle. A distribuição das frequências genotípicas foi de 5,2% e 6,0% de homozigotos para o alelo C, 26,3% e 32,0% de heterozigotos e 68,4% e 62,0% de homozigotos para o alelo G nos grupos de pacientes com DH e controles locais, respectivamente. Esses dados foram comparados a bancos de dados internacionais e, encontram-se dentro do esperado, baseado em nossa composição populacional. Dentro da população estudada, a associação entre o alelo C da variante rs3512, com um atraso na manifestação dos primeiros sintomas ($p = 0,643$) dos pacientes, não foi replicada. A falha nesta replicação pode ser atribuída à heterogeneidade genética entre as populações e ao pequeno tamanho amostral. Sendo assim, mais estudos devem ser realizados, estendendo a análise para outras regiões do genoma em um número amostral maior e com histórico étnico semelhante.

ABSTRACT

Huntington's disease (HD) is a rare, late-onset neurodegenerative disorder with autosomal dominant inheritance, whose symptoms include loss of motor coordination, psychiatric disorders, and progressive dementia. HD is caused by an expansion of the CAG repeat in exon 1 of the *HTT* gene. The determination of the age of onset (AO) of the first symptoms is partly associated with the length of the CAG repeats of the mutant *HTT* allele, the remaining variation is probably due to genetic and/or environmental factors. A possible factor that might influence the age of onset of the disease is the C allele of the variant rs3512, located in the *FAN1* gene, which was associated with a delay of 1.2 to 1.5 years on expected AO of the disease according to the length of the CAG repeats in a European study. Based on this information, the aim of this study was to identify the distribution of the variant rs3512 in a group of HD patients, comparing with a sample of local controls, and to evaluate the possible effects on the age of onset of these patients. In our population, 57 HD patients were evaluated and the CAG repeats length in mutant *HTT* alleles explained 56.1% of the variation in the age of onset ($p < 0.001$). The distribution of the allelic ($p = 0.645$) and genotypic ($p = 0.491$) frequencies of the variant rs3512 were established in both groups, local controls and HD patients. In the group of patients, the frequency of the C allele was established at 18.4% and 22.0% in the control group. The G allele frequency was 81.6% in HD patients and 78.0% in the control group. The distribution of genotypic frequencies was 5.2% and 6.0% homozygous for the C allele, 26.3% and 32.0% heterozygous and 68.5% and 62.0% homozygous for the G allele in the HD group and control group, respectively. The data was compared to international databases and it is within the expected, based on our population composition. In the studied population no association was found between the C allele of the variant rs3512 and a delay in the AO of the patients. The failure in this replication can be attributed to the genetic heterogeneity between the populations and the small sample size. Therefore, further studies should be carried out extending the analysis to other regions of the genome in a larger sample size and with a similar ethnic background.

SUMÁRIO

1 Introdução	9
1.1 Doença de Huntington	9
1.1.1 Aspectos Históricos	10
1.1.2 Epidemiologia	12
1.1.3 Manifestações clínicas	13
1.1.4 Identificação e isolamento do gene	15
1.1.5 Proteína Huntingtina	17
1.2 Modificadores Genéticos	20
1.2.1 Estudos Genéticos de Associação	20
1.2.2 Gene <i>FAN1</i>	26
1.2.3 Variante rs3512	27
2 Objetivos	30
2.1 Objetivo Geral	30
2.2 Objetivos específicos	30
3 Resultados	31
3.1 Artigo: “Huntington’s disease: variation in DNA repair system gene as a possible modifier of age of onset”	32
4 Discussão	51
5 Referências	55

1 Introdução

1.1 Doença de Huntington

A doença de Huntington (DH) é um distúrbio neurodegenerativo raro de herança autossômica dominante que, usualmente, tem um início sutil por volta da quarta década de vida e piora gradualmente até a morte. A progressão da doença varia em torno de 10 a 20 anos de forma irreversível (The Huntington 's Disease Collaborative Research Group, 1993).

Embora, normalmente, apresente o início das manifestações clínicas na vida adulta, a DH também se manifesta em jovens. Cerca de 10% dos pacientes manifestam os sintomas antes dos 20 anos de idade. O início precoce da doença está associado a sintomas mais graves e curso mais rápido. Enquanto o início tardio, acima de 60 anos, está associado a um fenótipo clínico mais brando da doença (Kremer *et al.*, 2002; Gosh & Tabrizi, 2018).

Um padrão característico de atrofia cerebral, principalmente na região do estriado no cérebro é observado nos pacientes com a DH. Nas regiões dos núcleos, putâmen e caudado, a atrofia se estabelece de forma precoce, no entanto, com a evolução da doença, todo o cérebro acaba sendo acometido, resultando em uma diminuição de pesos e volumes cerebrais (Figura 1). A DH causa perda neuronal em várias regiões do cérebro, incluindo neocórtex, cerebelo, hipocampo, substância negra e núcleo do tronco cerebral (Vonsattel *et al.*, 1998; Aylward *et al.*, 2000; Kloppel *et al.*, 2009).

Atualmente, não existe nenhum tratamento curativo para a DH. Todos os tratamentos são baseados no controle e alívio dos sintomas, isto é, são tratamentos paliativos, não impedem a progressão da doença, o que reflete em um ciclo de vida muito curto após as manifestações dos primeiros sintomas.

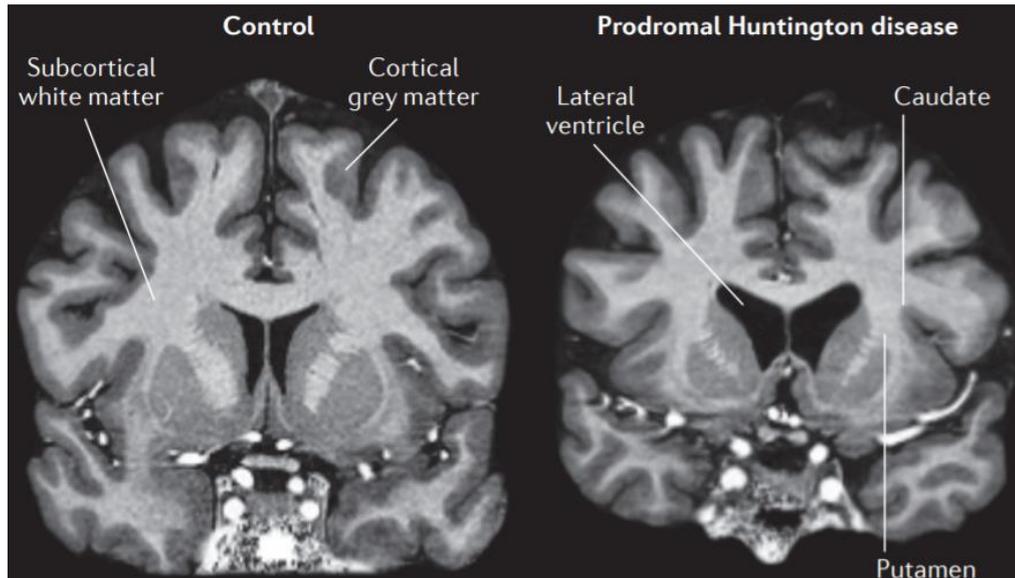


Figura 1. Atrofia na fase prodrômica da DH. Atrofia bilateral dos núcleos, caudado e putâmen, e aumento no tamanho do ventrículo lateral em um indivíduo com DH (direita) em comparação a um indivíduo normal (esquerda). Fonte: Bates *et al.* (2015).

1.1.1 Aspectos Históricos

A primeira referência a DH foi feita por Charles Oscar Waters em 1842 através de uma carta publicada na primeira edição do livro *“Practice of Medicine”*. Posteriormente, em 1846, foi observado, por Charles Gorman, uma elevada incidência da doença em determinadas regiões (Lanska, 2000).

No entanto, a primeira descrição mais consolidada foi feita em 1872 pelo médico George Huntington, na época, com apenas 22 anos de idade. Huntington, desde pequeno, acompanhava seu pai e avô, ambos médicos, em suas atividades profissionais em uma comunidade rural na Nova Inglaterra e, durante essas visitas, notou uma família com sintomas semelhantes durante várias gerações e concluiu que essa sintomatologia poderia estar, de alguma forma, interligada. Essa experiência proporcionou à Huntington uma perspectiva única sobre a doença (Bates, 2005; Wexler *et al.*, 2016).

No decorrer de sua carreira profissional, Huntington notou três características peculiares da doença: sua natureza hereditária, tendência à insanidade ou suicídio e início tardio da manifestação dos primeiros sintomas. Em 15 de fevereiro de 1872, Huntington publicou o artigo “*On Chorea*”, no “*The Medical and Surgical Reporter of Philadelphia*” (Figura 2), no qual descreveu, com precisão, as características clínicas da doença, que na época era conhecida como Coreia de Huntington, bem como seu padrão de herança autossômica dominante. A publicação de Huntington despertou grande interesse no meio científico e impulsionou uma série de trabalhos posteriores. No final do século XIX, a doença foi, finalmente, reconhecida como uma doença global (Huntington, 1872; Lanska, 2000; Bates, 2005; Wexler *et al.*, 2016).

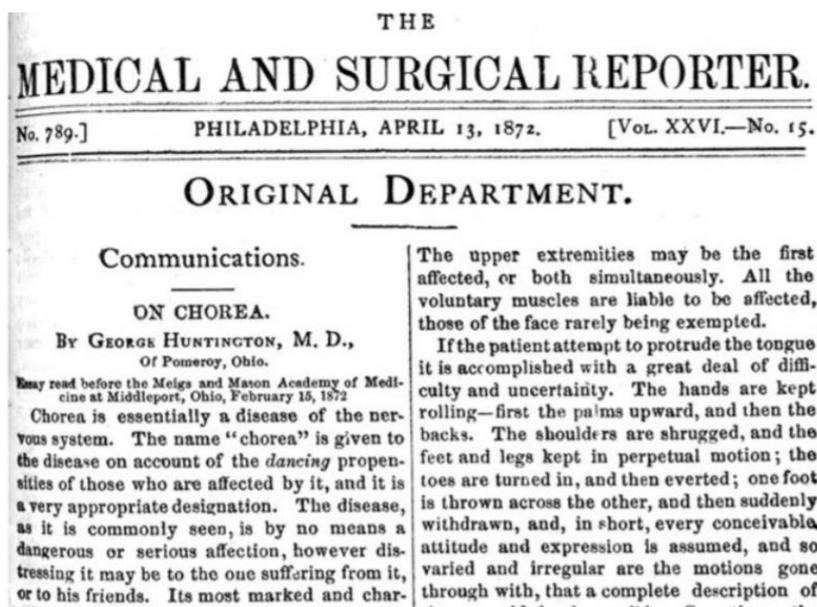


Figura 2. Título do artigo clássico de George Huntington, “ON CHOREA” publicado em 1872, no *Medical and Surgical Reporter*. Fonte: Wexler *et al.* (2016).

1.1.2 Epidemiologia

Embora a DH já tenha sido descrita em vários países e/ou populações, sua prevalência global varia muito de acordo com a região e a etnia, mas estima-se que ocorra com maior frequência em pessoas de origem euro-descendente (Harper, 1992).

A prevalência global da doença é de 5,70 indivíduos afetados para cada 100.000 habitantes nas populações da Europa Ocidental, América do Norte e Austrália e uma prevalência, significativamente menor, de 0,40 indivíduos afetados para cada 100.000 habitantes foi encontrada na Ásia (Pringsheim *et al.*, 2012). A prevalência nas populações da Índia e da África ainda não estão muito bem definidas, pois são poucos os estudos e dados epidemiológicos nessas regiões.

Por ser um continente caracterizado por grandes misturas étnicas, os dados epidemiológicos encontrados na América Latina são escassos. Diante desse cenário, uma revisão sistemática abrangente foi realizada em pacientes latino-americanos e a prevalência mínima da doença foi estimada em 0,5 indivíduos afetados para cada 100.000 habitantes nas populações da Venezuela e do México (Castilhos *et al.*, 2015).

Alguns isolados geográficos apresentam uma prevalência média bem mais alta, como, por exemplo, a região do Lago Maracaibo na Venezuela, considerada como a maior concentração mundial de pacientes com DH e, cuja prevalência mínima estimada é de 700 indivíduos afetados para cada 100.000 habitantes, o que pode indicar a influência de um efeito fundador (Ávila-Girón, 1973; Moscovich *et al.*, 2011; Castilhos *et al.*, 2019).

Ainda não existem dados concretos sobre prevalência da população geral no Brasil, mas a prevalência mínima da DH no estado do Rio Grande do Sul foi descrita em 1,85 indivíduos afetados para cada 100.000 habitantes, número bem menor do que o observado em populações predominantemente europeias (Castilhos *et al.*, 2019).

Contudo, é importante ressaltar que, até muito pouco tempo atrás, o diagnóstico era feito, unicamente, baseado nas manifestações clínicas. Com isso, muitas famílias, por desconhecimento ou preconceito, esconderam seus históricos e sintomas da sociedade. Portanto, as estimativas de prevalência devem ser repetidas a partir de múltiplas fontes, pois os números de prevalência tendem a serem maiores, uma vez que, com o avanço científico, técnicas moleculares mais precisas estão disponíveis para o diagnóstico dos pacientes (Morrison *et al.*, 2010; Wexler, 2010).

1.1.3 Manifestações clínicas

A DH é clinicamente caracterizada por uma combinação de alterações motoras, cognitivas, psiquiátricas e/ou comportamentais. Outros sintomas menos frequentes incluem perda de peso involuntária, arritmia cardíaca e distúrbios do sono. Com a progressão da doença esses sintomas se agravam e tornam os pacientes cada vez mais dependentes, incapacitando-os de realizar as tarefas cotidianas de forma independente e, os levando, gradualmente à morte. Estima-se que as causas mais comuns de morte sejam por pneumonia e/ou suicídio (Roos, 2010).

As anormalidades motoras podem ser observadas em dois estágios: inicial, caracterizado por movimentos involuntários e, na fase mais avançada da doença, caracterizado por movimentos voluntários (Ghosh & Tabrizi, 2018). Dentre as alterações na fase inicial da doença, a mais evidente e comum é a coreia, que se refere a movimentos involuntários, bruscos e irregulares, frequentemente iniciados nas extremidades distais, como dedos dos pés e das mãos, progredindo para outros segmentos até chegar nos pequenos músculos faciais. A primeira descrição da coreia foi feita pelo médico George Huntington em seu artigo publicado em 1872, onde ele cita a alteração como sendo um sinal típico e marcante de espasmo que, comumente, se inicia nos músculos do rosto e aumenta progressivamente em velocidade e intensidade. Não existe um padrão único para os movimentos de

coreia, isto é, variam de um indivíduo para o outro, mas estão presentes o tempo todo em que o paciente está desperto e se agravam com a progressão da doença (Huntington, 1872; Roos, 2010).

Embora a coreia seja o sintoma mais acentuado na doença, os pacientes sofrem um amplo espectro de alterações motoras. Com o avanço da doença nota-se a presença de disartria e disfagia, uma vez que conversar e engolir se tornam processos de alta complexidade para os pacientes que, ocasionalmente, podem chegar à mudez total. Com o tempo e progressão da doença, todos os pacientes, eventualmente, acabam desenvolvendo bradicinesia, hipocinesia, acinesia, distonia, tiques, movimento anormal dos olhos, movimentos excessivos e/ou inadequados e rigidez, levando o paciente a apresentar um ritmo mais lento nas atividades cotidianas. A influência do distúrbio motor progride com o tempo, desta forma, a mobilidade vai sendo gradualmente reduzida e quedas são muito frequentes, ao final da doença o paciente torna-se incapaz de se movimentar sozinho (Roos, 2010).

As alterações e/ou declínio cognitivos podem estar presentes antes mesmo das disfunções motoras e são características marcantes na DH (Paulsen *et al.*, 2008). Os primeiros sinais de disfunção cognitiva podem ser notados nas funções executivas, onde o indivíduo sofre uma perda na habilidade de organização e planejamento e, um comprometimento da memória e concentração. Com o avanço da doença, esses sinais se agravam e as tarefas usuais tornam-se muito mais difíceis. A capacidade cognitiva dos pacientes fica cada vez mais comprometida, até chegar ao ponto de o indivíduo apresentar um quadro de demência progressiva e alterações na personalidade (Tabrizi *et al.*, 2009; Rupp *et al.*, 2010; Roos, 2010).

Os pacientes apresentam um amplo espectro de alterações psiquiátricas que, muitas vezes, não são tão evidentes quanto as disfunções motoras e cognitivas, mas estão presentes. A depressão é o sintoma mais comum, sendo um fator de impacto direto na qualidade de vida e disposição dos pacientes afetados pela DH. Dentre as demais alterações psiquiátricas e/ou comportamentais descritas estão: síndrome de ansiedade, apatia, irritabilidade, agressividade, bipolaridade, hiper e hipossexualidade, sintomas obsessivos compulsivos e, em estágios mais

avançados, psicose. O risco de suicídio é relativamente alto nesses indivíduos, sendo a segunda causa mais comum de morte na doença (Paulsen *et al.*, 2001; van Duijn *et al.*, 2007; Ghosh & Tabrizi, 2018).

Embora a amplitude e intensidade dos sintomas varia de paciente a paciente, há três estágios principais de desenvolvimento da DH. Estágio inicial, onde as alterações são leves e sutis, estágio intermediário, onde há uma perda mais severa na coordenação motora e no equilíbrio, os sintomas se intensificam, a execução das tarefas diárias se torna de maior complexidade e há uma perda gradual na capacidade cognitiva. E, por fim, o estágio avançado, onde os pacientes se tornam totalmente afetados pela DH e dependentes, incapazes de andar, comer ou falar sozinhos.

1.1.4 Identificação e isolamento do gene

Desde a primeira descrição da DH, houve uma grande mobilização na comunidade científica a fim de entender melhor o percurso e a causa da doença, com isso inúmeros avanços foram feitos. Em 1983, o gene associado à DH foi mapeado, mas, somente 10 anos depois, o gene foi identificado através de técnicas de clonagem (Gusella *et al.*, 1983; The Huntington's Disease Collaborative Research Group, 1993; Bates, 2005).

O gene associado à DH foi localizado próximo a extremidade telomérica do braço curto do cromossomo 4, na região 4p16.3, está dividido em 67 éxons que compreendem uma ampla região de 210 Kb, o gene foi denominado como *IT15*, mas atualmente é mais conhecido como *HTT*. Este gene é caracterizado por uma região polimórfica no éxon 1 composta por repetições do trinucleotídeo citosina-adenina-guanina (CAG). A trinca de nucleotídeos CAG codificam o aminoácido glutamina, o que resulta na produção de uma cadeia de poliglutamina (poli-Q) na proteína (Gusella *et al.*, 1983; The Huntington's Disease Collaborative Research Group, 1993).

A manifestação do fenótipo clínico da doença está associada ao comprimento da cauda poli-Q, desta forma, indivíduos normais apresentam uma faixa de 6 a 35 repetições CAG, enquanto indivíduos afetados pela DH, isto é, aqueles que irão manifestar os sintomas clássicos da doença, apresentam uma penetrância completa em 41 ou mais repetições CAG. A faixa de 27 a 35 repetições trinucleotídicas corresponde aos alelos intermediários, que são aqueles que podem sofrer instabilidade somática durante a reprodução e, eventualmente, se expandirem em uma próxima geração. A penetrância incompleta é definida pela faixa de 36 a 39 repetições CAG, onde o início dos sintomas ocorre mais tardiamente e a progressão da doença é mais lenta, em alguns raros casos, a manifestação dos sintomas é tão tardia que quase nem é notada pelo paciente (The Huntington's Disease Collaborative Research Group, 1993; Ross *et al.*, 2014).

O número de repetições CAG implica, em partes, na determinação da idade de início da manifestação dos primeiros sintomas e, na extensão e/ou progressão dos danos causados pela DH. Existe uma correlação inversamente proporcional entre esses dois fatores, ou seja, quanto maior for o número de repetições CAG, mais cedo o paciente irá manifestar os primeiros sintomas da doença (Figura 3). Sendo assim, repetições muito longas, acima de 60, resultam em um início mais precoce dos sintomas e uma maior velocidade de progressão da doença. Essa condição, denominada DH juvenil, ocorre em torno de 10% dos pacientes que desenvolvem os sintomas antes dos 20 anos de idade (Gusella *et al.*, 1983; Roos *et al.*, 2014).

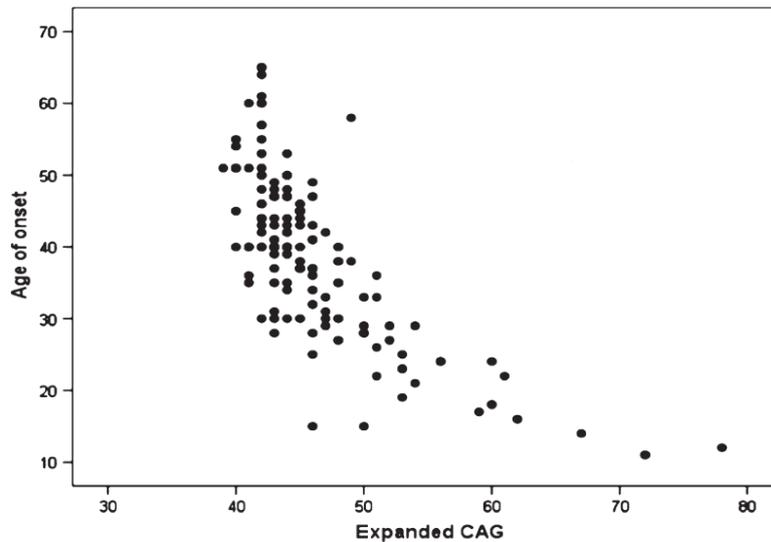


Figura 3: Idade de início dos primeiros sintomas x comprimento de repetições CAG. Representação gráfica da correlação inversa entre a idade de início (AO) e o número de repetições CAG em pacientes brasileiros afetados pela DH. Fonte: Castilhos *et al.* (2015).

1.1.5 Proteína Huntingtina

O gene da DH codifica uma proteína com múltiplos domínios de 348 kDa chamada huntingtina (htt), cuja exata função ainda não é conhecida. A huntingtina é expressa em vários tecidos, como, por exemplo, pulmões, fígado e coração, mas em maiores concentrações no cérebro e nos testículos. A região N-terminal é caracterizada por uma extensão de poli-Q proveniente das repetições trinucleotídicas CAG presentes no gene *HTT* (Saudou & Humbert, 2016).

Essa cauda poli-Q, presente na huntingtina mutante, é responsável por induzir alterações na conformação estrutural original da proteína. Essas mudanças impedem o enovelamento correto e ocasiona um acúmulo de agregados proteicos intracelulares, que na maioria dos casos manifestam-se no citoplasma e no núcleo dos neurônios. Esses agregados podem ser tóxicos para as células e estar envolvidos em mecanismos como excitotoxicidade, danos por radicais livres, ativação de caspase e apoptose (Landles & Bates, 2004).

A huntingtina mutante foi descrita nos pacientes da DH como a principal causa da neurodegeneração, de alterações de determinadas funções e da manifestação dos primeiros sinais clínicos da doença. Embora não se saiba ao certo o real mecanismo da causa dessa toxicidade, acredita-se que essas alterações sejam explicadas por duas possíveis hipóteses: ganho de função tóxico oriundo da extensa cauda poli-Q anormal ou perda de função proveniente das disfunções conformacionais estruturais (Landles & Bates, 2004; Takahashi et al., 2010; Ross & Tabrizi, 2011).

A DH é apenas uma dentro de um grupo total de nove doenças degenerativas causadas por mutações nas quais um trato de CAG expandido resulta em longos trechos de poli-Q na proteína codificada. Cada uma dessas doenças está associada a um gene específico e a uma proteína mutante diferente, o que implica na perda de neurônios específicos, com pouca sobreposição entre as regiões do cérebro afetadas nessas doenças. Dentro da família das doenças de poliglutaminas, além da DH, podemos citar: a atrofia dentatorubro-palidoluisiana (DRPLA), a atrofia muscular espinobulbar (SBMA) e as ataxias espinocerebelares 1, 2, 3, 6, 7 e 17 (Tabela 1). Embora não esteja muito bem elucidado, acredita-se na existência de mecanismos compensatórios que sejam capazes de neutralizar, por algum tempo, o efeito tóxico da proteína, uma vez que os indivíduos já nascem com a mutação, mas só em um determinado ponto da vida são acometidos pela toxicidade proteica (Orr, 2001; Landles & Bates, 2004; Takahashi et al., 2010; Orr, 2012).

Tabela 1: Características clínicas das doenças de poliglutaminas.

Doença	Gene Causador	Tipo de Herança	Características clínicas
SBMA	Receptor de androgênio (<i>RA</i>)	Recessiva ligada ao X	Fraqueza, atrofia da musculatura bulbar, tremores
DH	Huntingtina (<i>HTT</i>)	Autossômica Dominante (AD)	Coreia, demência, distonia
DRPLA	Atrofina1 (<i>ATN1</i>)	AD	Epilepsia mioclônica, ataxia, coreia, demência
SCA1	Ataxina1 (<i>ATXN1</i>)	AD	Ataxia, atrofia da musculatura bulbar, espasticidade, polineuropatia
SCA2	Ataxina2 (<i>ATXN2</i>)	AD	Ataxia, polineuropatia, parkinsonismo
SCA3	Ataxina3 (<i>ATXN3</i>)	AD	Ataxia, espasticidade, polineuropatia, disfagia, distonia, diplopia
SCA6	<i>CACNA1A</i> (Subunidade α 1A do canal de cálcio voltagem dependente tipo P/Q)	AD	Ataxia, disartria, nistagmo
SCA7	Ataxina7 (<i>ATXN7</i>)	AD	Ataxia, degeneração retiniana, oftalmoplegia
SCA17	<i>TBP</i> (proteína de ligante de TATA-box)	AD	Ataxia, demência, psicose,

Fonte: Adaptado de Takahashi *et al.* (2010)

Todas essas doenças de poliglutaminas apresentam o acúmulo de agregados proteicos intracelulares como característica em comum. Por muito tempo todos os estudos e investigações apontavam esses agregados como a causa patogênica da doença, porém, recentemente, alguns estudos demonstraram o contrário, que os agregados poderiam ser potenciais protetores da toxicidade da proteína mutante, reduzindo os níveis da htt difusa e, conseqüentemente, diminuindo o risco de morte celular e, aumentando a sobrevivência neuronal. (Davies *et al.*, 1998; Arrasate *et al.*, 2004; Miller *et al.*, 2010; Arrasate & Finkbeiner, 2012). Desta forma, essas descobertas fornecem novas perspectivas na compreensão do mecanismo da doença e abrem portas para maiores investigações sobre o assunto, trazendo esperança e clareza para essas doenças neurodegenerativas incuráveis.

1.2 Modificadores Genéticos

Uma das peculiaridades da DH é a variância na idade de início, embora tenha sido previamente descrito e estabelecido o efeito da correlação inversa entre a idade de início e o número de repetições CAG, isto explica apenas cerca de 70% na variabilidade da idade de início dos primeiros sintomas. Pacientes da mesma família e com o mesmo número de repetições CAG apresentam idades de início distintas. Essa variação sugere que além da correlação inversa existam outros fatores genéticos e/ou ambientais atuando na idade de início dos indivíduos afetados pela doença. Algumas evidências sugerem que grande parte dessa variabilidade na idade de início decorre de variantes genéticas comuns que podem modificar a estrutura e a função da proteína, impactando várias redes metabólicas e influenciando na expressão gênica. Essa interação poderia influenciar no tempo em que a primeira característica clínica se torna evidente (Duyao *et al.*, 1993; Rosenblatt *et al.*, 2001; Wexler *et al.*, 2004; Arning & Epplen, 2013).

Há várias décadas busca-se por fatores genéticos que possam influenciar na idade de início de pacientes com a DH. Neste contexto, a identificação desses modificadores genéticos pode ampliar a compreensão dos mecanismos biológicos, que são subjacentes à patogênese da doença e, proporcionar algum, potencial, tipo de abordagem terapêutica no futuro (Gusella *et al.*, 2014).

1.2.1 Estudos Genéticos de Associação

Um dos principais objetivos das análises genéticas em humanos é entender melhor os processos patogênicos da doença para o desenvolvimento de possíveis tratamentos e intervenções terapêuticas. Desta forma, estudos de associação genômica (*Genome Wide Association Study - GWAS*) fazem uma varredura por todo o genoma em busca de regiões polimórficas que são, usualmente, caracterizadas por uma variante de nucleotídeo simples (*single nucleotide variant - SNVs*), que sozinhos não são a causa da doença, mas funcionam como

marcadores genéticos que estabelecem pontos de referências no genoma. As SNVs são a forma mais comum de variação no genoma e são oriundos da troca de um determinado nucleotídeo por outro em um mesmo *locus* cromossômico. Sendo assim, investiga-se uma associação entre o genótipo e fenótipo de centenas ou milhares de SNVs, tal associação pode indicar uma relação direta, refletindo nos mecanismos e na progressão da doença, ou indireta, podendo ser um indicador para a localização da variante causal. Variantes genéticas de inúmeras doenças complexas já foram localizadas a partir de estudos de GWAS (Arning, 2016; Holmans & Stone, 2018).

Nesse âmbito, uma ampla análise de GWAS foi realizada em um total de 4.082 indivíduos afetados pela DH a fim de identificar, não fatores de risco, mas sim, possíveis modificadores genéticos que pudessem estar atuando na idade de início da manifestação dos primeiros sintomas da doença. A partir desta varredura genômica o sinal mais significativo foi encontrado no cromossomo 15 (Figura 4) (GeM-HD Consortium, 2015).

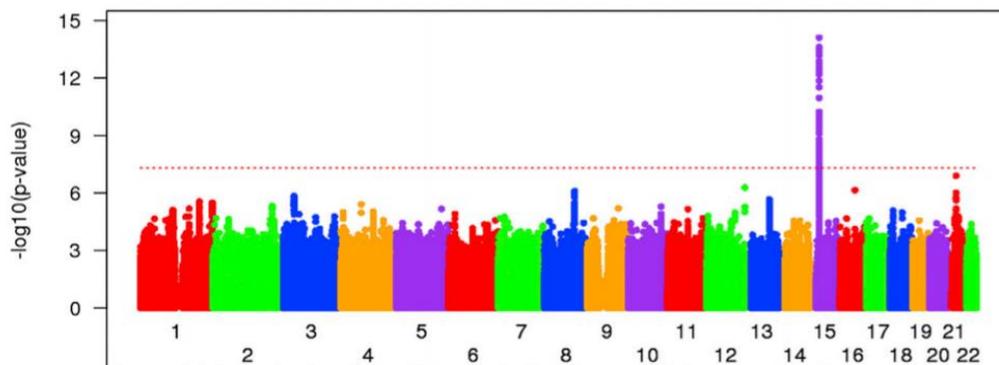


Figura 4: Análise de Associação Genômica em pacientes com DH. *Manhattan plot* da análise de GWAS evidenciando o pico de maior significância no cromossomo 15. Fonte: GeM-HD Consortium (2015).

Em um primeiro momento, dentro deste *locus* cromossômico, foram evidenciados dois grupos significativos de SNVs, o primeiro, compartilhado pelos genes *MTMR10* e *FAN1* e, o segundo, localizado no gene *HERC2P10*. Essas análises de associação foram feitas a partir do efeito de SNVs independentes

(rs2140734 e rs146353869). Após correções estatísticas, para a remoção do efeito dessas variantes específicas, foi notado que apenas a significância dos genes *FAN1* e *MTMR10* foi mantida. Ambos os genes fazem parte do sistema de reparo do DNA. (Figura 5) (GeM-HD Consortium, 2015).

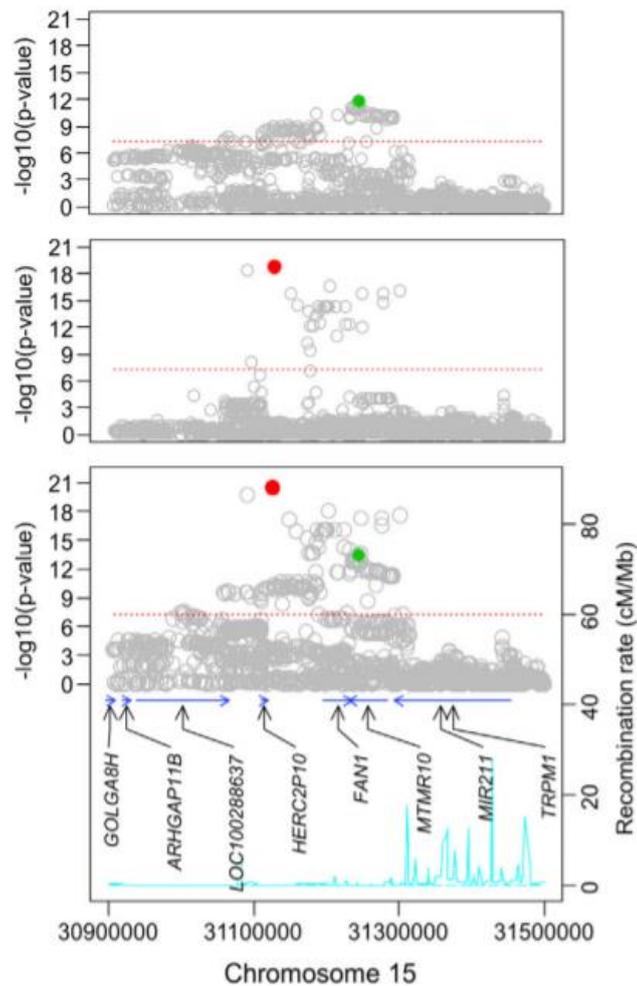


Figura 5: Análise de associação condicional no locus cromossômico principal. Painel inferior: Os círculos vermelhos e verdes representam o grupo de SNVs independentes mais significativos; Painel do meio: Análise de associação condicionada por uma variante específica (rs2140734). Ambos os grupos (vermelhos e verdes) são mantidos após a remoção do efeito associado desta variante; Painel superior: análise de associação condicionada por outra variante (rs146353869), apenas o grupo verde permanece significativo após a remoção do efeito desta variante. Fonte: GeM-HD Consortium (2015).

Dentro do grupo de SNVs significativos encontrados no gene *FAN1*, a variante rs3512, apresentou um retardo, estatisticamente significativo ($p < 0.001$), na idade de início da manifestação dos primeiros sintomas da doença. Foi demonstrado que os pacientes portadores do alelo C, desta variante, apresentavam os primeiros sintomas da doença 1,31 anos mais tarde que o esperado, em relação ao número de repetições CAG presentes no gene *HTT*. Desta forma, ao que tudo indica, o alelo C teria um papel protetor na idade de início em indivíduos americanos afetados pela DH (GeM-HD Consortium, 2015).

Este estudo de GWAS abriu caminhos e foi fonte de inspiração para uma série de publicações posteriores, os resultados apresentados tiveram um impacto profundo em várias áreas e forneceram, principalmente, esperança de um melhor entendimento da patogênese da doença.

A partir dessas informações e resultados, um estudo multicêntrico, envolvendo diversos grupos de pesquisa europeus, avaliaram 22 SNVs visando encontrar uma associação entre os genes envolvidos no sistema de reparo do DNA e a idade de início dos pacientes. Essas variantes genéticas foram avaliadas para várias doenças de poliglutaminas, incluindo DH, DRPLA, SBMA e as SCAS 1, 2, 3, 6, 7 e 17. Um total de 445 indivíduos foram analisados para a DH e, dentre todas as SNVs avaliadas, a SNV individual mais significativa foi a variante rs3512. Seguindo a mesma direção do trabalho citado anteriormente, foi demonstrado que os pacientes portadores do alelo C, da variante rs3512, apresentaram um atraso de 1,40 anos na idade de início dos primeiros sintomas ($p < 0.001$), em comparação aos portadores do alelo G. Assim sendo, foi demonstrado que os genes de reparo do DNA são potenciais modificadores da idade de início em DH e em SCAs, sugerindo um mecanismo patogênico comum que pode vir a oferecer novas oportunidades terapêuticas para várias doenças (Bettencourt *et al.*, 2016).

Posteriormente, inspirado e baseado nos resultados anteriores, as mesmas 22 SNVs, citadas previamente, foram avaliadas em pacientes com a doença de Machado-Joseph, que é uma doença de poliglutaminas, também conhecida como ataxia espinocerebelar do tipo 3 ou SCA3. O estudo teve como objetivo verificar o efeito da associação entre as variantes genéticas localizadas nos genes do sistema

de reparo do DNA e a idade de início dos pacientes com SCA3, em uma população asiática. Um total de 798 indivíduos com SCA3 foram analisados e nenhuma associação entre a variante rs3512 e a idade de início foi encontrada. Na verdade, nenhuma das SNVs testadas apresentaram resultados significativos e, dentre as justificativas citadas pelo autor, pela falha na replicação dos resultados está o fato de que a incidência de doenças poli-Q na Ásia é muito baixa, sendo necessário um número amostral muito grande para seguir as mesmas direções dos estudos prévios, que foram caracterizados a partir de coortes caucasianas. Por fim, o autor cita a ampla heterogeneidade genética entre as populações e sua influência na importância na identificação de potenciais modificadores genéticos (Wang *et al.*, 2018).

Nessa conjuntura, um estudo realizado pelo nosso grupo avaliou a associação entre a variante rs3512 e a idade de início dos primeiros sintomas em 144 pacientes afetados pela doença de Machado-Joseph. Foi demonstrado que os pacientes portadores do genótipo homozigoto para o alelo G (G/G), apresentavam uma antecipação de 2,44 anos na idade de início estimada, ou seja, esses indivíduos apresentaram os primeiros sinais clínicos da doença 2,44 anos antes do esperado em relação ao número de repetições CAG presentes no gene *HTT*. Quando esse resultado foi combinado ao gene *ATXN2*, previamente descrito como sendo um modificador genético da idade de início em pacientes com SCA3, foi demonstrado uma antecipação de 2,58 anos na idade de início desses pacientes. Todavia, quando o gene *ATXN2* foi combinado com o genótipo homozigoto para o alelo C (C/C) ou heterozigoto (C/G), da variante rs3512, foi demonstrado um atraso de 4,25 anos na idade de início dos primeiros sintomas da doença nestes pacientes, indo ao encontro dos resultados citados previamente (Figura 6) (Mergener *et al.*, 2020).

Portanto, como previamente elucidado, o alelo C da variante rs3512 atua como um fator protetor na idade de início dos pacientes. Não só em DH e SCA3, mas em outros distúrbios neurodegenerativos de início tardio. Estima-se que a neurodegeneração começa muito antes da manifestação do primeiro sinal clínico, nessa visão, a compreensão dinâmica do DNA, proteínas e vias envolvidas é de

extrema relevância para desenvolver terapias gênicas e prevenir e/ou melhorar os sintomas e a qualidade de vida dos pacientes.

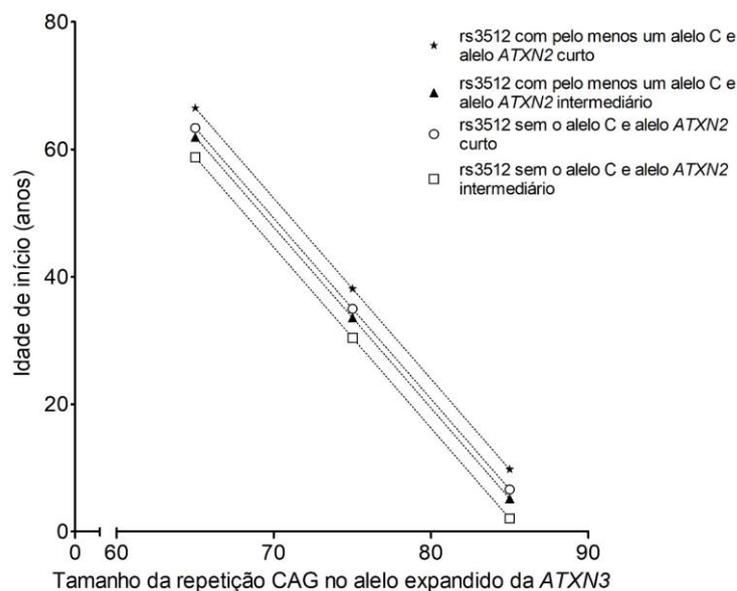


Figura 6: Análise estatística de diferentes genótipos da variante rs3512 associados ao gene ATXN2 em pacientes com SCA3.

Regressão linear: (★) representa pelo menos um alelo C combinado com o alelo curto da ATXN2 (até 22 repetições CAG); (▲) representa pelo menos um alelo C combinado com o alelo ATXN2 intermediário (entre 22 e 23 repetições CAG); (○) representa nenhum alelo C combinado com o alelo curto da ATXN2; (□) representa nenhum alelo C combinado com o alelo intermediário da ATXN2. Fonte: Mergener *et al.* (2020).

Por isso, a identificação de modificadores genéticos é de extrema importância na DH, e em várias outras doenças. A compreensão dos mecanismos que, de alguma forma, modulam a idade de início da manifestação dos primeiros sinais clínicos da doença, proporciona, além de um grande avanço na perspectiva e expectativa de vida, muita esperança para esses pacientes e familiares.

1.2.2 Gene *FAN1*

Várias publicações de estudos anteriores identificaram inúmeros genes como modificadores genéticos da idade de início na DH. Indivíduos sem a doença não sofrem os efeitos dos polimorfismos presentes nestes genes, no entanto, em pacientes podem alterar o curso e a progressão natural da doença.

O gene *FAN1*, um dos genes do sistema de reparo do DNA, foi descrito como um potencial modificador da idade de início na DH (GeM-HD Consortium, 2015; Bettencourt *et al.*, 2016). Esse gene está localizado no braço longo do cromossomo 15, no *locus* 15q13.3, é distribuído em 15 éxons que dão origem a 28 transcritos. A proteína codificada é altamente expressa no cérebro e leva o mesmo nome do gene, *FAN1*, é composta por 1.017 aminoácidos e, é caracterizada pelas atividades 5' endonuclease e 5'-3' exonuclease, tais atividades têm finalidade de reparar o DNA de danos ou lesões ocasionadas por ligações cruzadas intercadeia (*interstrand crosslinks* - ICLs) (Zerbino *et al.*, 2018; Yates *et al.*, 2020 - *Ensembl Human Genome Browser*).

ICLs bloqueiam a transcrição e a replicação do DNA e, como reflexo a isso, o sistema de reparo é acionado. *FAN1* interage com as proteínas do complexo dois da via da anemia Fanconi (*Fanconi anemia* - FA), para reparar os danos causados no DNA decorrentes desses ICLs. A via da FA é composta por três complexos proteicos que formam uma rede de interação a fim de corrigir os danos causados no DNA. O complexo central, denominado FANCI-FANCD2, é recrutado e ocorre uma interação entre *FAN1* e *FANCD2*, formando um heterodímero, com o complexo 1, denominado FAAP24. O complexo FANCM – FAAP24 desempenha um papel muito importante na via da FA, onde é capaz de identificar o dano, emitir sinais para o recrutamento do complexo central e estabilizar a forquilha de replicação para reparar a lesão causada no DNA (Smogorzewska *et al.*, 2010; Kim & D'Andrea, 2012; Lachaud *et al.*, 2016; Jin & Cho, 2017).

Alguns estudos prévios relataram a *FAN1* como potencial modificador genético, não apenas na DH. Foi demonstrado que a *FAN1*, em um modelo animal,

exerce um efeito protetor contra as repetições CGG no gene *FMR*, na Síndrome do X Frágil (Zhao & Usdin, 2018). E, posteriormente, em um modelo de células tronco pluripotentes induzidas (iPS), derivadas de pacientes, foi demonstrado que a superexpressão da *FAN1* estava associada a um retardo na idade de início e na progressão da DH, interagindo com as repetições CAG através de um mecanismo independente de nuclease. Isso fornece novas perspectivas para possíveis tratamentos terapêuticos na DH e, potencialmente, em outros distúrbios ocasionados por repetições trinucleotídicas (Goold *et al.*, 2019).

1.2.3 Variante rs3512

A variante rs3512 está localizada nos genes *FAN1* e *MTMR10*. Estudos prévios demonstraram que a SNV quando posicionada no gene *FAN1*, um dos genes do sistema de reparo do DNA, atuava como um potencial modificador genético na idade de início de doenças de poliglutaminas. Foi descrito um atraso na idade de início da manifestação dos primeiros sintomas clínicos em pacientes portadores do alelo C desta SNV, na DH e em SCA3 (GeM-HD Consortium, 2015; Bettencourt *et al.*, 2016; Mergener *et al.*, 2020).

O gene *MTMR10* é distribuído em 16 éxons que dão origem a 10 transcritos, faz parte da família de miotubularinas, que regulam diversos processos e formam heterodímeros com uma subunidade de fosfatase ativa para atuar nos fosfatos de fosfatidilinositol (Hnia *et al.*, 2012; Zerbino *et al.*, 2018; Yates *et al.*, 2020 - *Ensembl Human Genome Browser*). Até o momento não há nenhuma evidência de associação do alelo C da SNV rs3512 em relação a idade de início na DH, quando posicionado no gene *MTMR10*.

A SNV rs3512 está localizada na região regulatória, 3'UTR do gene *FAN1* e, em uma região intrônica do gene *MTMR10*, é caracterizado por uma variação do alelo ancestral G para C (G>C) e, não há nenhuma significância clínica reportada (Figura 7) (Zerbino *et al.*, 2018; Yates *et al.*, 2020 - *Ensembl Human Genome Browser National Center of Biotechnology Information*).

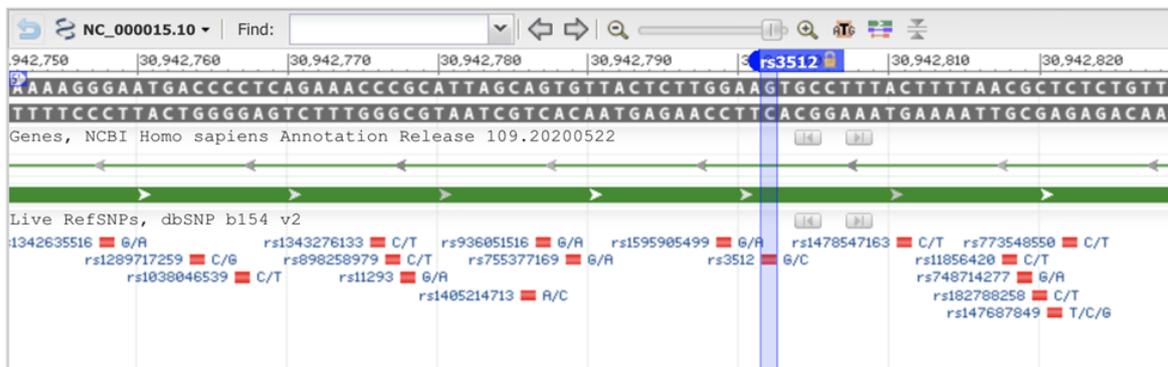


Figura 7: Visão genômica da variante rs3512. Fonte: Zerbino *et al.* (2018); Yates *et al.* (2020) - *Ensembl Human Genome Browser National Center of Biotechnology Information*.

De acordo com os bancos de dados internacionais *gnomAD* (Lek *et al.*, 2016) e *1000 genome* (Zerbino *et al.*, 2018), a frequência alélica global da variante rs3512 é de 0,733 para o alelo G e de 0,266 para o alelo C e, 0,812 para o alelo G e de 0,187 para o alelo C, respectivamente. A frequência do alelo menor (C) é mais alta na Europa, em torno de 30 à 33%, seguido da África, em torno de 19 à 21% e da América Latina (referente aos países Colômbia, México, Peru e Porto Rico), que apresenta a frequência do alelo C em torno de 17 à 18%, a menor frequência do alelo menor pode ser observada no Leste Asiático, com apenas 1% (Figura 8). A frequência genotípica global da SNV é de 0,673 para homocigotos para o alelo G (G/G), 0,280 para heterocigotos (C/G) e de 0,047 para homocigotos para o alelo C (C/C) (*1000 genome database*; Zerbino *et al.*, 2018). De uma forma geral, o genótipo mais frequente em todas as populações, exceto no Leste Asiático, é o G/G, seguido do genótipo C/G e, em menor frequência o genótipo C/C (Figura 8).

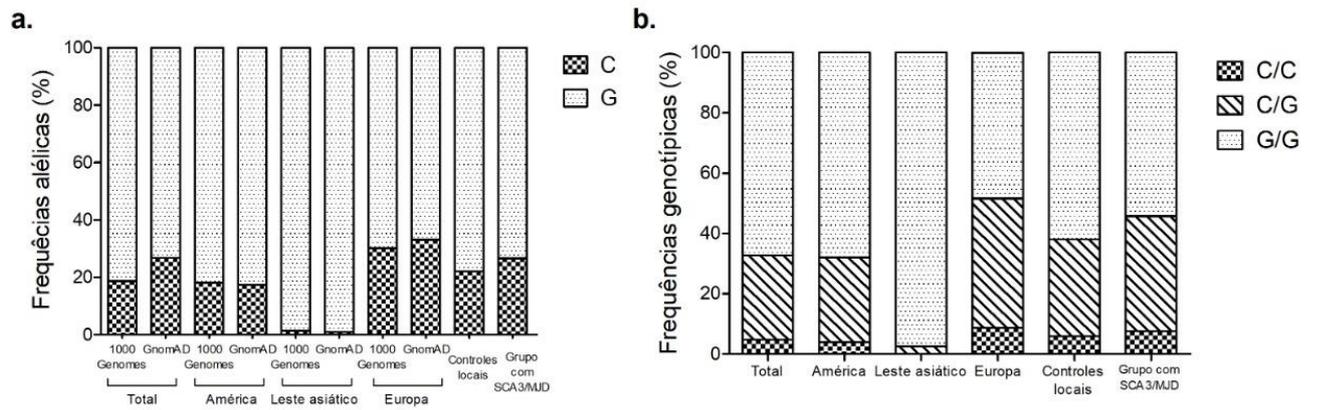


Figura 8: Distribuição das frequências alélicas e genotípicas da variante rs3512 em diferentes populações. a. Frequências alélicas extraídas de dois bancos de dados internacionais diferentes: *1000 genomas* e *GnomAD*. **b.** Frequências genotípicas de diferentes populações extraídas do Projeto *1000 genomas*. Fonte Adaptado de Mergerner *et al.* (2020).

2 Objetivos

2.1 Objetivo Geral

Avaliar possíveis efeitos do genótipo da variante rs3512 como modificador da idade de início em pacientes com a doença de Huntington.

2.2 Objetivos específicos:

- Identificar a variante rs3512 em um grupo de pacientes com DH e em amostras de controles normais;
- Determinar as frequências alélicas e genotípicas da variante rs3512 nos dois grupos;
- Comparar dados obtidos a partir da nossa população com bancos de dados internacional;
- Avaliar a possível influência da variante rs3512 com a idade de início dos pacientes com DH.