

Patrícia Espinheira de Sá Maciel

Estudos de Genética Molecular da Doença de Machado-Joseph

**Contribuição para a compreensão das bases moleculares
da variabilidade fenotípica da doença e da
instabilidade mitótica e meiótica da mutação.**

Patrícia Espinheira de Sá Maciel

**Estudos de Genética Molecular da
Doença de Machado-Joseph**

**Contribuição para a compreensão das bases moleculares
da variabilidade fenotípica da doença
e da instabilidade mitótica e meiótica da mutação.**

Instituto de Ciências Biomédicas Abel Salazar
Universidade do Porto
1997

ÍNDICE

RESUMO	1
I. INTRODUÇÃO	5
Definição clínica e patológica da doença de Machado-Joseph	7
Apresentação clínica	
Quadro patológico	
Variabilidade de apresentação clínica na DMJ - subfenótipos	
Diagnóstico diferencial	
Testes laboratoriais	
Tratamento	
Unicidade da doença e variabilidade da expressão fenotípica	
Genética da DMJ	14
Modo de transmissão	
Penetrância dependente da idade	
Antecipação na DMJ	
Necessidade da detecção pré-sintomática da doença	
Identificação da mutação causadora da doença	18
Estudos de genética reversa (clonagem posicional)	
Estabelecimento do diagnóstico molecular da DMJ - confirmação da presença da expansão do (CAG) _n em famílias de origens étnicas diversas	
Desenvolvimento do teste preditivo da DMJ	20
Diagnóstico pré-sintomático	
Diagnóstico pré-natal	

Mecanismo patogénico da mutação da DMJ	22
Identificação da função normal da proteína - estudos de homologia	
Estudos de expressão do gene MJD1 normal e mutado	
A DMJ como doença associada à expansão de uma sequência (CAG) _n	
Modelos propostos para o mecanismo patogénico dos <i>repeats</i>	
CAG/poliglutamina expandidos	
Selectividade da morte neuronal associada à expansão de <i>repeats</i>	
CAG/poliglutamina	
Instabilidade meiótica e mitótica dos <i>repeats</i> de trinucleotídeos expandidos.	32
Factores que afectam a instabilidade genética dos <i>repeats</i> de trinucleotídeos	
Instabilidade dos <i>repeats</i> CAG expandidos em modelos animais	
Modelos propostos para a instabilidade genética dos <i>repeats</i> expandidos	
Instabilidade genética do <i>repeat</i> CAG e origem da mutação DMJ	
Objectivos do estudo	41
II. TRABALHO EXPERIMENTAL	43
CAPÍTULO 1. Estudos de ligação genética e esforço de clonagem do gene	45
Estudos de ligação genética do <i>locus</i> DMJ a marcadores do cromossoma 14q	47
Introdução	
Material e métodos	
Resultados	
Discussão	
Estudo de <i>repeats</i> de trinucleotídeos do cromossoma 14 em famílias com DMJ	55
Introdução	
Material e métodos	
Resultados	
Discussão	

Clonagem e elucidação da estrutura genómica do gene MJD1 62

Introdução

Material e métodos

Resultados

Discussão

CAPÍTULO 2. Correlações genótipo-fenótipo: quadro clínico e
antecipação 71

Resumo dos artigos

Artigo I: *Correlation between CAG repeat length and clinical features in Machado-Joseph disease*

Artigo II: *Gender equality in Machado-Joseph disease*

Artigo III: *Familial factor independent of repeat length influences age at onset of Machado-Joseph disease*

CAPÍTULO 3. Estudos de mosaïcismo somático do gene MJD1 no
sistema nervoso central 81

Resumo dos artigos

Artigo IV: *Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado-Joseph disease.*

Artigo V: *Mosaicism of the CAG repeat in CNS tissue in relation to age at death in SCA1 and MJD patients.*

CAPÍTULO 4. Instabilidade intergeracional do *repeat* CAG
expandido 89

Resumo do artigo

Artigo VI: *Effect of the normal chromosome on intergenerational instability of the expanded (CAG)_n in the Machado-Joseph disease gene.*

III. DISCUSSÃO GLOBAL	95
Bases moleculares da variabilidade clínica	98
Efeito do tamanho do <i>repeat</i> CAG expandido na idade de início	
Outros factores moduladores da idade de início na DMJ	
Homozigotia e idade de início	
Rapidez de progressão da DMJ e tamanho do <i>repeat</i> expandido	
Tamanho do <i>repeat</i> CAG expandido e subfenótipo	
Redefinição nosológica e fenotípica das ataxias espinocerebelosas	
Bases moleculares da especificidade de morte neuronal na DMJ	107
Exclusão do mosaicismo somático do (CAG) _n expandido no gene MJD1 como causa da especificidade da degenerescência na DMJ	
Interações proteicas como mecanismo determinante da morte celular na DMJ	
Mosaicismo somático do <i>repeat</i> CAG expandido - contribuições dos diferentes tipos celulares e progressão ao longo da vida	
Bases moleculares do fenómeno de antecipação - instabilidade do <i>repeat</i> CAG expandido	112
Instabilidade genética do <i>repeat</i> CAG expandido no gene MJD1: correlações com a antecipação	
Factores que contribuem para a instabilidade genética do (CAG) _n expandido na DMJ	
Instabilidade genética do <i>repeat</i> CAG: pistas possíveis para a origem da mutação DMJ	
Desenvolvimento do teste preditivo da DMJ	116
Interpretação do resultado do teste molecular: tamanhos dos (CAG) _n expandidos	
Homoalelismo para o alelo normal: possibilidades de eliminação da incerteza do resultado do teste molecular	
Comportamento não-mendeliano da mutação DMJ: relevância para o aconselhamento genético	
Necessidade de envolvimento da componente de investigação no aconselhamento genético	
IV. CONCLUSÕES E DIRECÇÕES FUTURAS DE INVESTIGAÇÃO	123

ABREVIATURAS UTILIZADAS

ANEXO: Descrição das famílias e generalidades sobre os métodos utilizados

AGRADECIMENTOS

RESUMO

RESUMO

Estudos de Genética Molecular da Doença de Machado-Joseph. Contribuição para a compreensão das bases moleculares da variabilidade fenotípica da doença e da instabilidade mitótica e meiótica da mutação.

A doença de Machado-Joseph é uma doença neurológica hereditária autossómica dominante de início na idade adulta, que se caracteriza por um quadro clínico muito diversificado, com ataxia espinocerebelosa e oftalmoplegia, associadas em grau variável a múltiplos outros sintomas, tais como distonia, espasticidade, amiotrofias e nistagmo. Desconhece-se o mecanismo biológico causador desta doença, assim como possíveis alterações celulares ou metabólicas que sirvam como um marcador precoce (pré-sintomático) da DMJ. Descrevem-se aqui os trabalhos realizados com vista à localização e identificação do gene causador da doença de Machado-Joseph, e os estudos de correlação genótipo-fenótipo, em termos da apresentação clínica, patológica e das características invulgares de transmissão desta doença, tais como a antecipação.

O gene associado à DMJ (MJD1) localiza-se no cromossoma 14q32.1 e contém na região codificante uma sequência repetitiva (CAG)_n altamente polimórfica, que se encontra expandida em indivíduos afectados. Definiram-se neste estudo os limites de tamanho do (CAG)_n na população normal e numa população alargada de doentes, de origem predominantemente portuguesa. À variabilidade de tamanhos da sequência repetitiva está associada parte da variabilidade fenotípica observada nesta doença, em termos de idade de início e de apresentação clínica: existe uma correlação inversa entre o tamanho do *repeat* CAG e a idade de início da doença. No entanto, esta correlação é incompleta, pelo que o tamanho do *repeat* não é utilizável no aconselhamento genético, na determinação do prognóstico (idade de início e gravidade da doença). Estudos mais aprofundados revelaram a existência de outros factores familiares moduladores da idade a que se inicia a doença. Também o quadro clínico apresentado pelos doentes DMJ se

relaciona com o tamanho do (CAG)_n no gene MJD1, estando os *repeats* maiores associados às formas mais graves da doença.

Em estudos de correlação entre o mosaïcismo somático e a patologia, o mosaïcismo observado no SNC de portadores da expansão do (CAG)_n no gene MJD1 foi excluído como base da especificidade de morte neuronal na DMJ, concluindo-se que outros factores celulares específicos estarão envolvidos na determinação da selectividade do efeito nocivo da mutação. O grau de mosaïcismo somático parece relacionar-se antes com a actividade mitótica dos diversos tecidos, e, portanto, com a sua composição celular e a história do seu desenvolvimento.

Foi também estudada a instabilidade intergeracional do *repeat* CAG expandido no gene MJD1, provável base molecular da antecipação observada na DMJ, assim como os factores que a influenciam, nomeadamente o sexo do progenitor que transmite o gene mutado e polimorfismos intragénicos presentes no cromossoma normal. Excluiu-se, por estudo de haplótipos intragénicos, o efeito desestabilizador de sequências flanqueantes do *repeat* CAG, na população normal e em portadores da mutação DMJ.

Finalmente, discutem-se de forma breve as aplicações práticas destes conhecimentos recém-adquiridos, em particular à realização de testes preditivos em indivíduos assintomáticos pertencentes a famílias com a doença, assim como nos primeiros casos de diagnóstico pré-natal, no âmbito do Programa Nacional de Teste Preditivo e Aconselhamento Genético da DMJ e outras Ataxias Dominantes.

I. INTRODUÇÃO

I. INTRODUÇÃO

DEFINIÇÃO CLÍNICA E PATOLÓGICA DA DMJ

Apresentação Clínica

A doença de Machado-Joseph (DMJ) faz parte do grupo das ataxias hereditárias espinocerebelosas. Nas primeiras publicações, foi descrita como três doenças distintas [Nakano *et al.*, 1972; Woods e Schaumburg, 1972; Rosenberg *et al.*, 1976]. Foi somente com os trabalhos pioneiros de Coutinho & Andrade [1978] que a DMJ foi reconhecida como uma entidade única, com características clínicas e patológicas próprias (ver tabelas 1 e 2, nas páginas 16 e 18, respectivamente).

A doença, que apresenta uma grande variabilidade, inicia-se geralmente na idade adulta, em média por volta dos 40 anos de idade, mas com extremos de 5 e 73 anos. A ataxia cerebelosa é a primeira e a mais frequente manifestação clínica da DMJ. Os primeiros sintomas são, quase invariavelmente, alterações do equilíbrio durante a marcha; surgem depois dificuldades de articulação verbal e, anos mais tarde, incoordenação dos membros superiores. A segunda manifestação clínica mais comum é a oftalmoparesia externa, caracterizada por uma limitação do olhar vertical para cima, primeiro, seguida depois pela limitação da convergência e do olhar lateral. Um achado ocular muito característico na DMJ, mas presente apenas em cerca de 25% dos doentes portugueses, é a retracção palpebral (aspecto de "olhos arregalados"). O síndrome piramidal, o outro elemento da tríada que caracteriza a DMJ, é observado em mais de 80% dos doentes e inclui hiperreflexia osteotendinosa, sinal de Babinski e espasticidade. Outro sintoma, que ocorre aproximadamente em 65% dos doentes, é a disfagia, que pode ter consequências graves levando a morte por aspiração maciça ou por broncopneumonias repetidas.

Outras manifestações clínicas que também aparecem na DMJ são (1) um síndrome periférico, com sinais que vão desde a simples abolição dos reflexos aquilianos até atrofias musculares, mais ou menos pronunciadas; (2) sinais extrapiramidais,

Tabela 1. Critérios para o diagnóstico clínico da DMJ

(modificados de Coutinho, 1992)

1. Hereditariedade autossômica dominante
2. Início mais frequente na vida adulta
3. Presença de ataxia, oftalmoparesia externa progressiva e sinais piramidais
4. Associação frequente de sinais extrapiramidais ou periféricos
5. Presença de sinais menores, mas muito característicos, como fasciculações de contracção da face e língua e retracção palpebral
6. Ausência de alterações cognitivas
7. Sobrevida média de aproximadamente 21 anos.

principalmente posturas distónicas dos membros e alguns (raros) casos de quadro parkinsoniano, caracterizado por amimia, com pobreza dos movimentos da face e expressão, e lentificação de todos os movimentos voluntários; (3) fasciculações de contracção da face e da língua.

A inteligência está geralmente bem preservada, mesmo em doentes de evolução prolongada, servindo como um critério chave no diagnóstico diferencial com outras formas de ataxia hereditária que apresentam um comprometimento cognitivo importante. A sobrevida média é de aproximadamente 21 anos e a morte ocorre devido a complicações advindas da imobilidade prolongada e da incapacidade motora, que levam a quadros pulmonares agudos ou crónicos [Coutinho, 1992].

Quadro patológico

Os critérios patológicos propostos para o diagnóstico da DMJ são apresentados na tabela 2. Tipicamente, observa-se nesta doença a degenerescência dos núcleos espinocerebeloso, dentado, pôntico e vestibular, bem como de estruturas extrapiramidais, como a *substantia nigra*, o *locus coeruleus* e o complexo palidolusiano. Há também redução da densidade neuronal nos núcleos dos nervos cranianos motores, dos cornos anteriores da medula, nos gânglios raquidianos e cordões posteriores. Estruturas como o córtex cerebral, o córtex cerebeloso e as olivas bulbares são normalmente poupadas [Rosenberg *et al.*, 1976; Coutinho e Andrade, 1978; Coutinho *et al.*, 1982; Sakai *et al.*, 1983; Kanda *et al.*, 1989; Coutinho, 1992].

Variabilidade de apresentação clínica na DMJ - Subfenótipos

Desde as primeiras descrições da DMJ foi notória uma grande variabilidade clínica entre diferentes famílias, observada também (em menor grau) em doentes pertencentes à mesma família. Esta variabilidade foi organizada numa classificação em três subtipos (ou subfenótipos) clínicos [Lima e Coutinho, 1980]. O *tipo 1* compreende os doentes com

Tabela 2. Critérios patológicos para o diagnóstico da DMJ
*(propostos por Coutinho, 1992, a partir da revisão
de 18 autópsias descritas na literatura até 1990)*

1. Estruturas afectadas na quase totalidade dos casos descritos:

- *Locus niger* e núcleo subtalâmico
- Núcleo vermelho
- Núcleos pânticos, dentados / pedúnculos cerebelosos médios e superiores
- Feixe longitudinal médio
- Núcleos vestibulares
- Núcleos dos nervos cranianos motores
- Cornos anteriores da medula / raízes anteriores
- Colunas de Clarke / feixes espino-cerebelosos
- Cordões posteriores, gânglios raquidianos

2. Estruturas quase sempre poupadas

- Córtex cerebral
- Estriado
- Córtex cerebeloso
- Olivas bulbares
- Feixe piramidal

início mais precoce e evolução mais grave, em quem os sinais piramidais são muito marcados e que apresentam, além da tríada típica da DMJ, sinais extrapiramidais importantes, geralmente na forma de posturas distónicas. O *tipo 2*, o mais comum, é intermédio em termos de idade de início; o seu quadro clínico está limitado à tríada clássica de sintomas (ataxia, oftalmoplegia e sinais piramidais) e pode ser uma forma transitória, já que todos os doentes começam como tal e, após alguns anos, podem ou não evoluir para os tipos 1 ou 3. O *tipo 3* tem início mais tardio e engloba doentes com sinais periféricos muito marcados, além dos sinais clássicos principais. A evolução da doença é mais benigna e a sobrevivência menos afectada nestes doentes. Embora se trate, como acontece com qualquer classificação, de uma divisão arbitrária do que na realidade é um contínuo de sintomas observados nos doentes DMJ, a classificação em três subtipos mostra-se muito relevante para a prática clínica, com valor comprovado quer em termos de diagnóstico quer de prognóstico [Sequeiros, 1989], tendo sido adoptada em geral nos estudos desta doença [Rosenberg, 1977; Barbeau *et al.*, 1984].

Diagnóstico diferencial

Uma questão problemática é a do diagnóstico diferencial com outras formas de ataxia espinocerebelosa de início na vida adulta. Muito se tem debatido sobre os aspectos clínicos e a classificação desse grupo de doenças heredo-degenerativas, cujos quadros clínicos e patológicos se sobrepõem, dificultando a diferenciação de cada uma delas, sobretudo em indivíduos isolados ou pertencentes a famílias pequenas. As principais dificuldades residem não só nesta sobreposição fenotípica, mas também na existência de uma grande variabilidade na apresentação clínica da DMJ, por vezes mesmo no seio de uma mesma família. A identificação dos genes causadores das diversas formas de ataxia possibilitará a reavaliação dos critérios clínicos e patológicos na presença dos resultados moleculares, o que permitirá possivelmente uma melhor classificação clínica e genética das ataxias espinocerebelosas.

Testes laboratoriais

As anomalias bioquímicas que foram encontradas na DMJ e em outras ataxias espinocerebelosas [Duvoisin *et al.*, 1983; Kish *et al.*, 1987; Blass *et al.*, 1988] são inespecíficas e provavelmente reflectem apenas os processos finais de degenerescência neuronal comuns a quase todas estas doenças. Daí a dificuldade em se desenvolver um teste bioquímico que permitisse detectar alterações mais específicas desta doença. Assim, o diagnóstico de genética molecular, ou seja, a detecção directa da mutação causadora da doença por estudo de ADN [Kawaguchi *et al.*, 1994; Maciel *et al.*, 1995; Lopes-Cendes *et al.*, 1996b] surgiu como a única possibilidade de diagnóstico de certeza da DMJ.

Tratamento

Não existe, até ao momento, nenhum tratamento específico para a DMJ. A medicação sintomática para alívio da espasticidade ou de sintomas parkinsonianos, pode beneficiar alguns doentes com essas manifestações clínicas [Subramony *et al.*, 1990]. Existem relatos do uso de agonistas colinérgicos e precursores serotoninérgicos com o objectivo de tentar aliviar os sintomas cerebelosos, mas os resultados são muito controversos [Kark *et al.*, 1977; Sorbi *et al.*, 1988; Trouillas *et al.*, 1988]. Ensaios terapêuticos com a combinação de sulfametoxazol e trimetoprim (Bactrim®) têm sido realizados por vários grupos com resultados promissores [Mello e Abbot, 1988; Sangla *et al.*, 1990; Coutinho, 1992].

Unicidade da doença, variabilidade da expressão fenotípica

Os subtipos definidos por Coutinho e Andrade [1978] correspondem aproximadamente aos fenótipos mais frequentemente observados nas três famílias originalmente descritas: a família Joseph [Rosenberg *et al.*, 1976], em que o quadro clínico predominante corresponderia ao tipo 1, a família Thomas [Woods e Schaumburg,

1972], com fenótipo correspondente ao tipo 2, e a família Machado [Nakano *et al.*, 1972], com predominância das formas mais benignas correspondentes ao tipo 3. Estas famílias foram descritas separadamente e as suas manifestações neurológicas consideradas como entidades clínicas diversas. No entanto, o facto de por vezes se observarem doentes de sub-tipos diferentes, nomeadamente dos dois subtipos extremos (1 e 3), na mesma família, assim como as modificações fenotípicas observadas ao longo do tempo num mesmo doente, sugeriram que se tratava de uma única doença, com expressão variável de um gene principal, e não de doenças distintas [Coutinho e Andrade, 1978].

A extrema diversidade na expressão clínica da doença de Machado-Joseph é um caso particularmente interessante da variabilidade fenotípica frequentemente observada em doenças genéticas, sobretudo em doenças autossómicas dominantes. Embora os mecanismos que levam a esta variabilidade fenotípica em portadores comprovados de um mesmo gene mutante não sejam compreendidos na maioria dos casos, poderiam ser de esperar influências (a) do alelo normal do mesmo gene, (b) do restante genoma, e (c) do meio ambiente, na expressão fenotípica do gene mutado principal; é compreensível que nas doenças autossómicas dominantes, que dependem da presença de um único gene mutado, haja mais lugar para variação, a nível da variante normal do mesmo gene ou do resto do genoma, o que pode condicionar a expressão fenotípica do gene mutado [Sequeiros, 1989]. A possibilidade de existirem genes modificadores do fenotipo na DMJ foi levantada por Sequeiros [1989], com base nos valores das correlações de idades de início em pares de familiares.

GENÉTICA DA DMJ

Modo de transmissão

A DMJ é transmitida de modo autossómico dominante, o que significa que basta uma cópia do gene mutado para o fenótipo mutante se revelar; ou seja, quer os heterozigotos quer os homozigotos para a mutação manifestam a doença. Assim, se um dos progenitores for doente ou portador comprovado da mutação, os seus descendentes têm, *a priori*, um risco de 50% de virem também a ser afectados. É possível, no entanto, que os (raros) homozigotos para a mutação possam ter um início mais precoce e um fenótipo mais grave que os heterozigotos [Coutinho *et al.*, 1982; Coutinho, 1992; Lang *et al.*, 1994; Kawakami *et al.*, 1995; Gadoth *et al.*, 1996].

Penetrância dependente da idade

Como se referiu, a DMJ é uma doença com início geralmente na vida adulta (média 37.4 ± 14.1 anos), mas em que a idade de manifestação dos primeiros sintomas é muito variável (extremos de 5 e 73 anos em doentes portugueses) [Sequeiros e Coutinho, 1993]. Conhecem-se alguns casos de não-penetrância (cerca de 2%) [Sequeiros, 1989], isto é, de portadores obrigatórios do gene que são ainda assintomáticos em idades muito avançadas (por vezes até aos 90 anos). Existe pois uma penetrância dependente da idade, ou seja, uma distribuição probabilística (de base empírica) das idades de início da doença.

Não existem diferenças entre os sexos na manifestação da DMJ, quer quanto ao número de homens e de mulheres afectadas, quer quanto à idade de início num e noutra sexo [Coutinho, 1992; Sequeiros e Coutinho, 1993]. Também não foi detectado qualquer efeito do sexo do progenitor na idade de início da doença.

Antecipação na DMJ

A antecipação, isto é, o início progressivamente mais precoce de uma doença hereditária ao longo das gerações, normalmente acompanhado de um aumento da sua gravidade, foi até há pouco tempo um fenómeno genético controverso, cuja aceitação pela comunidade científica levou bastantes anos e para o qual apenas recentemente se encontrou uma base biológica - a instabilidade dos *repeats* de trinucleotídeos expandidos [McInnis, 1996]. Já no século XIX, Benedict Auguste Morel [1857], no seu tratado sobre a teoria da degeneração, "*Traité des Dégénérescences*", descrevia a deterioração progressiva das doenças de pais para filhos, ocorrendo estas de forma mais grave e numa idade mais precoce com as gerações sucessivas, levando finalmente à extinção da linhagem. Morel, cujos estudos principais foram na área do bócio e do cretinismo, citava vários tipos de causas (ambientais e hereditárias) que contribuiriam para essa "degeneração".

No princípio do século XX, a "degeneração", então eufemizada e designada "antecipação", adquiriu um papel central nos estudos de genética psiquiátrica. Em 1911, Mott referia-se à "lei da antecipação", descrevendo o início mais precoce das doenças psiquiátricas ao longo de gerações sucessivas, concluindo que a predisposição hereditária era o factor mais importante na genética destas doenças e que a presença de antecipação demonstrava a heritabilidade de uma "tara neuropática", a tendência para a doença mental [Mott, 1911]. A forma de se voltar ao "estado normal" nestes casos seria ou através do casamento com linhagens normais ou pela eliminação natural dos elementos patológicos naquelas linhagens, devida à antecipação. Os movimentos eugenistas, empenhados no "melhoramento" da raça humana, aderiram com entusiasmo a estas teorias.

As primeiras observações de antecipação na distrofia miotónica foram efectuadas por um oftalmologista alemão, Fleischer, que identificou casos de indivíduos com cataratas cuja descendência apresentava distrofia miotónica, sem que eles próprios tivessem quaisquer manifestações musculares da doença, e levantou a hipótese de uma conexão hereditária entre estes fenótipos [Fleischer, 1918]. Muitas objecções foram no

entanto levantadas a esta ideia. Em particular, Lionel Penrose, um empenhado anti-eugenista, argumentou que seria possível encontrar antecipação em quase todos os estudos de doenças que ocorram frequentemente em duas gerações de uma mesma família, devido a diversos tipos de enviesamento inerentes ao estudo destas [Penrose, 1948].

As "causas" da aparente antecipação seriam, segundo Penrose, a variação do espectro das idades de início da doença na população, em combinação com os enviesamentos de amostragem, resultante da selecção preferencial (1) de pais afectados com início da doença mais tardio, (2) de descendentes afectados de início precoce e (3) de casos de início mais ou menos simultâneo da doença em pais e filhos. Estes casos atrairiam mais facilmente, à partida, a atenção médica, levando mais facilmente à suspeita de hereditariedade por parte do clínico ou da família, na qual se cria uma consciência aumentada para a existência da doença e sua transmissão, que levaria à detecção desta aos primeiros sinais, gerando-se assim a ilusão de antecipação [Penrose, 1948; Harper *et al.*, 1992].

Embora estudos prospectivos recentes, em que se evitaram cuidadosamente estes enviesamentos possíveis [Howeler *et al.*, 1989], tenham confirmado a existência real de antecipação na distrofia miotónica, só com a identificação da mutação causadora desta doença, a expansão de um *repeat* CTG [Brook *et al.*, 1992], se eliminaram as dúvidas restantes quanto ao significado biológico destas observações: o *repeat* expandido era geneticamente instável, variando de tamanho durante a transmissão do progenitor ao descendente e, a esta variação de tamanho (mais frequentemente uma expansão), correspondia uma variação na idade de início e gravidade da doença. A descoberta deste novo tipo de mutação, que foi designada "mutação dinâmica" [Richards e Sutherland, 1992], veio abrir um novo capítulo nos estudos da genética, por permitir explicar as observações prévias, dificilmente aceites por não encaixarem no esquema conceptual de um genoma estável (de uma geração para a seguinte), e por ter levado à identificação de um mesmo tipo de mutação em várias outras doenças genéticas em que observava antecipação: o síndrome do X frágil [Fu *et al.*, 1991; Kremer *et al.*, 1991; Verkerk *et al.*,

1991; Yu *et al.*, 1991], a doença de Huntington [The Huntington Disease Collaborative Research Group, 1993], a atrofia espino-bulbar [La Spada *et al.*, 1991], a atrofia dentatorubropalidoluisiana [Koide *et al.*, 1994], a ataxia espinocerebelosa tipo 1 [Orr *et al.*, 1993], a doença de Machado-Joseph [Kawaguchi *et al.*, 1994; Imbert *et al.*, 1996; Pulst *et al.*, 1996; Sanpei *et al.*, 1996] e a ataxia espinocerebelosa tipo 2.

Na DMJ, ainda antes da identificação da alteração genética subjacente à doença, ou mesmo do conhecimento das expansões de trinucleotídeos como mutação causadora de doenças humanas, foi descrita a existência de "antecipação", embora não fosse uma antecipação tão marcada como na distrofia miotónica e se observasse apenas em algumas famílias [Sequeiros e Coutinho, 1981; Oliveira *et al.*, 1994; Sequeiros *et al.*, 1994a], sugerindo desde logo que esta doença poderia também estar associada a uma expansão de trinucleotídeos. Estudos mais aprofundados, em que se procurou eliminar o efeito dos enviesamentos de amostragem descritos por Penrose, confirmaram recentemente a existência de antecipação na DMJ [Sousa *et al.*, 1997].

Necessidade da detecção pré-sintomática da doença

Apesar da sua gravidade, o início tardio da DMJ permite a transmissão do gene mutante de pais para filhos, uma vez que os sintomas se manifestam, na maioria dos casos, durante ou após a fase reprodutiva. A incidência desta doença, altamente incapacitante e sem cura, levando a uma morte vezes precoce, após um longo período de sofrimento para o indivíduo e para a família, tem consequências emocionais devastadoras para as pessoas afectadas e para todos os que estão em risco de vir a desenvolvê-la. Como muitas doenças genéticas, a DMJ pode levar à estigmatização e segregação das famílias afectadas [Boutté, 1987], bem como ao agravamento das suas condições económicas, muitas vezes já deficientes, devido à perda de membros produtivos da família. O aconselhamento genético é pois uma arma fundamental no auxílio às famílias afectadas por esta doença. No entanto, e dada a inexistência de sinais precoces que permitam a detecção pré-sintomática da (clínica ou bioquímica) da DMJ, até 1993 a única

possibilidade de previsão do risco de um dado indivíduo (assintomático) de uma família vir a desenvolver a doença era o cálculo pelo método Bayesiano, entrando em conta com a distribuição empírica das idades de início [Sequeiros, 1989]. Este método, embora permitindo estimativas de risco correctas e mais baixas do as que não levassem em linha de conta as idades de início, não tinha muita utilidade prática em termos de planeamento familiar já que, dado o início tardio da doença, o risco só se tornava suficientemente baixo após a idade reprodutiva. Tornava-se assim necessária a identificação do gene mutado na DMJ e o consequente desenvolvimento de um teste molecular adequado, a fim de se poder levar a cabo o aconselhamento genético de forma mais eficaz.

IDENTIFICAÇÃO DA MUTAÇÃO CAUSADORA DA DMJ

Estudos de genética reversa (clonagem posicional)

A necessidade de se identificar a anomalia bioquímica causadora da doença de Machado-Joseph, com o objectivo de se compreender o mecanismo patológico subjacente a esta doença, de se iniciar a investigação de uma forma de tratamento eficaz e de se desenvolver um teste diagnóstico, levou aos estudos de *genética reversa* ou *clonagem posicional*, que permitem identificar directamente o gene potencialmente alterado numa dada doença hereditária, ultrapassando a etapa da identificação da alteração bioquímica. A vantagem deste tipo de abordagem é que torna possível, uma vez identificado o gene causador da doença e antes mesmo de se chegar a compreender na totalidade a sua etiologia, desenvolver o teste directo da mutação em indivíduos afectados ou portadores assintomáticos, para fins de diagnóstico molecular ou teste preditivo e aconselhamento genético.

Numa primeira fase desta estratégia, diversos grupos a nível internacional levaram a cabo a procura da localização cromossómica do gene mutado na DMJ. Este acabou por ser mapeado primeiro em famílias japonesas [Takiyama *et al.*, 1993] no braço longo do

cromossoma 14. Os estudos de ligação a marcadores do cromossoma 14q32.1 em famílias DMJ portuguesas (capítulo 1) permitiram confirmar essa mesma localização cromossómica e demonstrar a homogeneidade genética entre as famílias de ambas as origens [Sequeiros *et al.*, 1994b]. O refinamento da região cromossómica de interesse a um intervalo de 2 cM através do estudo de marcadores genéticos adicionais nas famílias DMJ de origem portuguesa e em famílias DMJ de outras origens, também ligadas ao cromossoma 14q, permitiu o início do mapeamento físico da região candidata, em paralelo com estratégias alternativas para identificação de genes candidatos.

Em particular, certas características clínicas e genéticas da DMJ sugeriam desde logo que esta doença pudesse ser causada pelo mesmo tipo de mutação que várias outras doenças degenerativas do sistema nervoso, progressivas e de início na idade adulta, como a doença de Huntington, a doença de Kennedy (atrofia espino-bulbar), a ataxia espinocerebelosa tipo 1 (SCA 1) e a atrofia dentatorubropalidoluisiana (ADRPL), que se provou estarem associadas à expansão de sequências repetitivas (CAG)_n, localizadas na região codificante dos respectivos genes. Com o objectivo de testar esta hipótese, e como via alternativa para a identificação do gene causador da DMJ, iniciou-se a identificação de sequências repetitivas de trinucleotídeos na região de interesse e o estudo destas em doentes DMJ (capítulo 1). Nesse estudo, porém, não foi identificada nenhuma sequência de trinucleotídeos expandida.

A clonagem do ADNc do gene MJD1 por Kawaguchi e colaboradores [1994], seguindo uma estratégia semelhante de identificação de *repeats* CAG, em bancos de ADNc de cérebro, veio demonstrar que, de facto, este tipo de mutação estava também associado à DMJ: o gene MJD1 contém uma sequência repetitiva (CAG)_n, altamente polimórfica, que na população controlo testada variava entre 13 e 36 tripletos CAG [Kawaguchi *et al.*, 1994], mas que nos indivíduos afectados de uma família com diagnóstico clínico e patológico de DMJ se apresentava expandida, contendo 68 a 79 tripletos CAG. Neste estudo, a existência de uma expansão do (CAG)_n no gene MJD1 foi demonstrada apenas em 11 dos 12 doentes japoneses testados. Assim, permanecia em

aberto a possibilidade de nem todos os casos de DMJ serem causados pelo mesmo tipo de mutação, ou seja, de existir heterogeneidade genética, alélica ou não-alélica (de *locus*).

ESTABELECIMENTO DE UM DIAGNÓSTICO MOLECULAR DA DMJ

Confirmação da expansão do *repeat* CAG no gene MJD1 como mutação causadora da doença em famílias DMJ de origens diversas

O nosso estudo (artigo I) confirmou que a mesma expansão do (CAG)_n no gene MJD1 estava presente na grande maioria das famílias DMJ açorianas, brasileiras e norte-americanas (de origem açoriana) previamente descritas. Apenas quatro das 38 famílias clinicamente diagnosticadas como DMJ, não se confirmaram como portadoras desta mutação: uma delas demonstrou-se posteriormente ser portadora da mutação SCA2 [Silveira et al, submetido a publicação], enquanto as mutações das restantes permanecem por identificar, após terem sido excluídas também as mutações SCA1, SCA2, SCA6 e ADRPL. A reavaliação clínica detalhada destas e de outras famílias, após realização destes testes moleculares, poderá contribuir para a redefinição do fenótipo clínico da DMJ (ver Discussão).

Da mesma forma, na quase totalidade das famílias DMJ de outras origens geográficas e étnicas previamente descritas [Coutinho, 1994] foi encontrada a expansão do *repeat* CAG no gene MJD1: no Japão [Maruyama *et al.*, 1995; Watanabe *et al.*, 1996], no Brasil [Lopes-Cendes *et al.*, 1996d], na China [Zhou *et al.*, 1997], em famílias aborígenes da Austrália [Burt *et al.*, 1996] e em famílias de judeus iemenitas [Gadoth *et al.*, 1996]. Também em famílias francesas com ataxia espinocerebelosa dominante, designadas SCA3, se encontrou esta mutação [Stevanin *et al.*, 1995b], bem como em numerosas outras famílias inglesas [Giunti *et al.*, 1995], alemãs [Haberhausen *et al.*, 1995; Schols *et al.*, 1995], holandesas [Brunt *et al.*, 1996], indianas [Giunti *et al.*, 1995; Matilla *et al.*, 1995], chinesas e de Taiwan [Soong *et al.*, 1997] e norte-americanas de

origem não-portuguesa [Matilla *et al.*, 1995; Ranum *et al.*, 1995; Higgins *et al.*, 1996]. As famílias indianas descritas por Jain e colaboradores [1986] foram excluídas deste grupo, demonstrando-se posteriormente serem portadoras de uma expansão no gene SCA2.

Não foram descritas até à data quaisquer outras mutações no gene MJD1 associadas a doença; é provável pois que a expansão do (CAG)_n seja a causa da DMJ em todas as populações, sendo os casos em que ao diagnóstico clínico de DMJ não corresponde a presença de uma expansão explicáveis por sobreposição fenotípica com mutações em outros genes, tais como as expansões de um (CAG)_n nos genes SCA1, SCA2 ou SCA6, ou outras ainda não identificadas.

Desenvolvimento do teste preditivo da DMJ

Com a disponibilidade do teste molecular directo, tornou-se possível identificar portadores da mutação da DMJ, ainda em fase pré-sintomática, para fins de aconselhamento genético. Não são pois necessários estudos familiares para determinar (por análise de ligação) o estatuto genético de um indivíduo no que respeita à DMJ. O teste é feito por amplificação do segmento de gene contendo a sequência repetitiva CAG, por *polymerase chain reaction (PCR)*, a partir de ADN genómico do indivíduo (extraído, normalmente, de linfócitos do sangue periférico), seguida da determinação do tamanho do (CAG)_n nos dois alelos. Pode assim verificar-se se estes se encontram na gama correspondente aos alelos normais ou na dos alelos expandidos, sendo suficiente um alelo expandido para se manifestar a doença (por definição de doença dominante).

Foi iniciado, em finais de 1995, um Programa Nacional de Teste Preditivo e Aconselhamento Genético na DMJ e outras Ataxias Dominantes, que inclui seis centros locais (Vila Real, Porto, Coimbra, Lisboa, Ponta Delgada e Angra do Heroísmo), cada um com uma equipa idealmente formada por pelo menos um neurologista, um geneticista, um psicólogo clínico e uma assistente social, e uma equipa central que inclui ainda geneticistas moleculares, socio-psicólogos, psiquiatras, antropólogos e eticistas, entre

outros [Sequeiros, 1996]. Este Programa foi elaborado com base na experiência com o diagnóstico pré-sintomático da doença de Huntington nos EUA, Canadá e Europa, que mostrou que este tipo de acompanhamento pode ajudar a prevenir as reacções psicológicas desajustadas ao diagnóstico precoce duma doença incurável, mas que é também necessário o mesmo tipo de acompanhamento para os indivíduos que recebem o diagnóstico de não-portadores [Paúl, 1996].

Diagnóstico Pré-Natal da DMJ

Sendo a DMJ uma doença de início tardio, com algumas formas menos graves, as questões éticas levantadas pela realização de um diagnóstico pré-natal (DPN) são bastante complexas, na ausência de uma abordagem terapêutica para a doença. No entanto, a determinação do estatuto do feto em relação à mutação DMJ é possível, pelos mesmos métodos laboratoriais utilizados no teste pré-sintomático, utilizando ADN extraído de amniócitos ou de vilosidades coriônicas. O DPN pode pois ser oferecido aos casais que o procurem e em que um dos elementos seja afectado ou um portador comprovado da mutação, detectado no âmbito do Programa de Teste Preditivo. Na sequência daquele Programa, foi já feito o diagnóstico molecular em dois fetos [Sequeiros *et al.*, submetido a publicação].

MECANISMO PATOGENICO DA MUTAÇÃO DA DMJ

Um dos objectivos principais da identificação do gene causador da DMJ seria, como se disse, a compreensão do mecanismo patogénico desta doença. Em alguns casos de isolamento de genes associados a doenças por métodos de clonagem posicional, muita informação sobre o mecanismo patogénico pôde ser extraída a partir unicamente do que se sabia sobre o gene. Nomeadamente, sendo identificadas regiões de homologia significativa da proteína codificada pelo gene em questão com outras proteínas de função

conhecida, é por vezes possível prever a função dessa proteína na célula e procurar testar a(s) hipótese(s) com experiências adequadas.

Identificação da função normal da proteína - estudos de homologia

No caso do gene MJD1, não foram detectadas homologias com nenhuns genes conhecidos, além da sequência de poliglutamina, que é comum a outros genes envolvidos em doenças degenerativas do sistema nervoso. Destes, apenas se conhece a função para o gene mutado na doença de Kennedy (SBMA), o receptor de androgénios: trata-se de um factor de transcrição, com uma região C-terminal de ligação aos androgénios e uma região central de ligação ao ADN [Mhatre *et al.*, 1993]. A sequência poliglutamina situa-se na região N-terminal do receptor de androgénios, que se pensa estar envolvida na modulação da transcrição. As funções dos genes associados à doença de Huntington, às ataxias espinocerebelosas tipo 1 ou 2 e à atrofia dentatorubropalidoluisiana são desconhecidas, embora para a huntingtina se tenha especulado tratar-se de uma proteína envolvida no transporte de vesículas nos neurónios [DiFiglia *et al.*, 1995].

Existem outros exemplos de proteínas com regiões ricas em glutaminas, em geral factores de transcrição [Perutz *et al.*, 1994], tais como as proteínas *homeobox* na *Drosophila* e o factor de transcrição Sp1 em humanos. A proteína que se liga à *TATA box*, permitindo o início da transcrição de numerosos genes, designada *TATA binding protein* (TBP), contém uma sequência poliglutamina de comprimento variável [Imbert *et al.*, 1994], que vai de 25 a 42 unidades na população normal [Gostout *et al.*, 1993; Imbert *et al.*, 1994]; pensa-se que esta sequência aumenta a afinidade dos domínios de activação transcripcional para as outras proteínas envolvidas no processo de activação [Seipel *et al.*, 1994]. Outros exemplos são uma RNA polimerase mitocondrial no fungo *Neurospora crassa*, codificada pelo gene *cyt-5* [Chen *et al.*, 1996], e a interleuquina-2 do ratinho, em que se demonstrou uma correlação inversa entre o comprimento da sequência poliglutamina e a actividade biológica da proteína (nomeadamente a capacidade de indução de crescimento em células humanas) [Matesanz e Alcina, 1996], bem como uma

associação preferencial dos alelos de maior tamanho com a diabetes auto-imune nos ratinhos NOD [Ghosh *et al.*, 1993].

Estudos adicionais de caracterização da proteína codificada pelo gene MJD1, nomeadamente da sua localização intracelular e tecidual e de modificações pós-traducionais, assim como estudos funcionais e de interação com outras proteínas, podem permitir uma melhor compreensão da função desta proteína a nível celular e fisiológico, bem como da forma como está envolvida na patologia da DMJ.

Estudos de expressão do gene MJD1 normal e mutado

Os estudos de expressão do ARN do gene MJD1, por hibridação *in situ*, apontam para a expressão ubíqua deste gene [Nishiyama *et al.*, 1996]. A análise da distribuição da proteína codificada pelo gene MJD1, por análise de imunohistoquímica [Paulson *et al.*, 1997a], revelou que esta é expressa em geral por todo o organismo e em todas as regiões do cérebro examinadas, embora nalgumas regiões do cérebro, como o estriado, a expressão se limite a alguns grupos de neurónios (possivelmente interneurónios). A localização da proteína na célula é predominantemente citoplasmática, com uma distribuição somatodendrítica, normalmente excluindo o núcleo. Há também indicação da possível localização da proteína a nível mitocondrial [Trottier *et al.*, 1997].

Em doentes com DMJ [Paulson *et al.*, 1997a], a proteína com a expansão do segmento poliglutamina é traduzida a níveis normais, nos mesmos tecidos que a proteína normal e com a mesma distribuição subcelular somatodendrítica; observações preliminares sugerem, no entanto, um aumento da quantidade de proteína no núcleo, o que, a confirmar-se, poderá estar de alguma forma relacionado com a patologia celular.

A DMJ como doença associada à expansão de uma sequência (CAG)_n/poliglutamina

Outras pistas para a compreensão do mecanismo patogénico da DMJ poderão provir da informação adquirida sobre outras doenças causadas pelo mesmo tipo de

mutação. Como se disse, à data da clonagem do gene eram já conhecidas quatro outras doenças degenerativas do sistema nervoso causadas pela expansão de um *repeat* CAG/poliglutamina: a doença de Huntington [The Huntington Disease Collaborative Research Group, 1993], a doença de Kennedy [La Spada *et al.*, 1991], a ataxia espinocerebelosa tipo 1 [Orr *et al.*, 1993] e a atrofia dentatorubropalidoluisiana [Koide *et al.*, 1994]. Posteriormente, demonstrou-se a presença de expansões de (CAG)_n em outras doenças similares, nomeadamente em duas outras formas de ataxia espinocerebelosa: a ataxia espinocerebelosa tipo 2 (SCA2) [Imbert *et al.*, 1996; Pulst *et al.*, 1996; Sanpei *et al.*, 1996] e uma forma particular de ataxia com degenerescência da retina (SCA7) [Gouw *et al.*, 1995; Stevanin *et al.*, 1996]; nesta última, a demonstração inicial da presença deste tipo de mutação foi feita por estudos com um anticorpo específico para poliglutaminas expandidas [Stevanin *et al.*, 1996], tendo-se posteriormente clonado o gene mutado [David *et al.*, 1997], de que se desconhece ainda a função. Embora o mecanismo de morte celular nestas doenças seja desconhecido, os resultados de diversos estudos parecem sugerir que se trata não de uma perda da função normal das proteínas em questão, mas sim o ganho de uma nova "função" devido à expansão da poliglutamina [Housman, 1995], com efeitos tóxicos para as células, em particular alguns neurónios. Isto estaria de acordo, quer com o modo de transmissão autossómico dominante destas doenças, quer com o facto de a expressão de proteínas contendo sequências poliglutaminas expandidas em modelos animais (ratinhos transgénicos) levar à degenerescência neuronal nestes. Por outro lado, no caso do receptor de androgénios são conhecidos portadores de mutações pontuais que apresentam actividade diminuída deste receptor, com conseqüente feminização testicular, mas sem o fenótipo neurológico observado na presença de uma expansão da sequência poliglutamina, sugerindo que a esta última está associada um ganho de uma nova função, tóxica para os neurónios, e não apenas uma redução da actividade normal da proteína.

É interessante notar também que sequências longas de poliglutamina (59-81 unidades) são tóxicas quando introduzidas em *Escherichia coli*, inibindo o crescimento destas bactérias, enquanto sequências mais pequenas de poliglutamina (10-35 unidades)

ou sequências longas de polialanina (61 unidades) não têm qualquer efeito [Onodera *et al.*, 1996].

Modelos propostos para o mecanismo patogénico dos *repeats* CAG/poliglutamina expandidos

Embora não se pretenda fazer uma revisão detalhada desta matéria, impõe-se aqui uma referência aos principais modelos especulativos, propostos até à data, para explicar o mecanismo patogénico das poliglutaminas expandidas:

1. Poliglutaminas como substratos da transglutaminase [Green, 1993]: neste modelo propõe-se que a enzima transglutaminase, que cataliza a formação das ligações isopeptídicas { ϵ (γ glutamil-lisina)} entre os seus substratos, ricos em glutaminas, e outras proteínas, que actuam como dadoras do grupo lisil, possa actuar de forma inusitada nas proteínas com sequências de poliglutamina excessivamente longas (geralmente acima de um limite de aproximadamente 40 unidades), formando agregados de proteínas ligados covalentemente. Embora estes agregados fossem possivelmente degradados por proteólise, o isopéptido residual não pode ser degradado por este processo, acumulando-se na célula, com potencial efeito tóxico para esta. A especificidade neuronal deste efeito tóxico seria explicável pela diminuída capacidade dos neurónios de degradar os produtos da transglutaminase, assim como pela ausência de renovação celular nestes tipos de células. Em conformidade com esta hipótese, foi demonstrado que peptídeos contendo *repeats* de glutamina flanqueados por sequências parciais das proteínas associadas à ataxia espinocerebelosa tipo 1, doença de Machado-Joseph e atrofia dentatorubropalidolusiana são excelentes substratos para a transglutaminase *in vitro*, formando agregados insolúveis com proteínas de extractos de cérebro [Kahlem *et al.*, 1996]. O aumento do tamanho da sequência repetitiva aumentava a reactividade dos peptídeos. No entanto, não foi possível até à data observar agregados proteicos *in vivo* em nenhuma das doenças associadas à expansão de uma poliglutamina, o que reduz a aplicabilidade deste modelo.

2. Funcionamento das poliglutaminas como *polar zippers* [Perutz *et al.*, 1994]: demonstrou-se recorrendo a modelização molecular das possíveis estruturas adoptadas por proteínas contendo sequências repetitivas de resíduos polares, como é o caso da glutamina, que estas eram capazes de ligar β -strands entre si, formando estruturas em barril ou folha pregueada β , através do estabelecimento de pontes de hidrogénio entre as amidas da cadeia principal e os seus grupos polares. Perutz e colaboradores designaram este tipo de estrutura *polar zipper*. Estudos de dicroísmo circular de ultravioletas, microscopia de polarização e microscopia electrónica, demonstraram que um péptido Asp₂-Gln₁₅-Lys₂ (em que os resíduos aspartato e lisina foram adicionados com o objectivo de conferir solubilidade em meio aquoso ao péptido) formava agregados de elevado peso molecular, que apresentavam de facto um tipo de estrutura compatível com a presença de um *polar zipper*. Seria então possível que ao aumento do comprimento do segmento poliglutamina nas proteínas mutadas correspondesse uma afinidade excessivamente elevada para as proteínas a que normalmente se ligam, ou uma afinidade não-específica para outras proteínas. Alternativamente, a extensão da poliglutamina acima de um dado limite de tamanho poderia levar à aglomeração e precipitação das proteínas mutadas nos neurónios.

3. Efeito excitotóxico: neste modelo propõe-se que a expansão da poliglutamina poderia causar o metabolismo anormal das proteínas em que se inseriu, levando à acumulação intracelular de poliglutaminas [Cha e Dure, 1994]. Estas teriam um potencial efeito excitotóxico, devido à actividade das poliaminas como moduladores dos receptores de aminoácidos excitatórios, ou um efeito perturbador no metabolismo energético da célula, por interferência com o influxo de cálcio na mitocôndria. De acordo com outra hipótese proposta, o metabolismo celular de síntese e degradação das proteínas contendo poliglutaminas expandidas levaria a alterações subtis, mas crónicas, nos níveis intracelulares de glutamina, o que se reflectiria na concentração intracelular e vascular de glutamato (obtido a partir da glutamina por acção da enzima transglutaminase), com a possível consequência de morte celular por excitotoxicidade crónica, em alguns grupos de neurónios susceptíveis [Hannan, 1996].

4. Interferência com o metabolismo energético : referida por vários autores, a hipótese de que uma deficiência na capacidade da célula gerar energia poderia causar a degenerescência neuronal foi reforçada pelos resultados obtidos por Burke e colaboradores [1996], que demonstraram que péptidos poliglutamina sintéticos, assim como as proteínas codificadas pelos genes da doença de Huntington e da ADRPL, apresentavam uma interacção com a enzima central da glicólise, a gliceraldeído 3-fosfato desidrogenase (GAPDH); esta interacção era tanto mais forte quanto maior a sequência poliglutamina [Burke *et al.*, 1996]. O mesmo tipo de interacção com a GAPDH, específica e dependente da extensão da poliglutamina, foi demonstrado também para o receptor de androgénios e para a ataxina 1 [Koshy *et al.*, 1996]. Assim, é possível que as proteínas com poliglutaminas expandidas perturbem a produção de energia da célula por interferirem com a GAPDH e que o *stress* metabólico seja o mecanismo patogénico comum a todas as doenças associadas a este tipo de mutação. O facto de o cérebro, ao contrário de outros tecidos, não poder utilizar energia dos lípidos, mas apenas da glucose, torna qualquer interferência com as enzimas da glicólise particularmente nociva nesta estrutura. Mesmo uma alteração subtil na capacidade de utilização da energia pode, a longo prazo, afectar células como os neurónios, com reduzida capacidade de renovação. Por outro lado, além do seu papel na glicólise, a GAPDH apresenta múltiplas outras funções e interacções com outras proteínas possivelmente relevantes para a toxicidade neuronal: actua como uracil-DNA-glicosilase (função na reparação do ADN), liga-se ao ARN, ao ATP, à calciclina, à actina, à tubulina e à proteína precursora da amilóide (APP). Assim, há várias vias possíveis pelas quais uma interacção anormal das poliglutaminas expandidas com esta enzima poderia levar a um efeito nocivo para as células, sobretudo em células neuronais especialmente susceptíveis. No entanto, resta ainda demonstrar o efeito da interacção das proteínas com poliglutaminas expandidas na actividade da GAPDH.

Há ainda outras pistas que apontam para a alteração do metabolismo energético neuronal como base molecular da degenerescência, em particular no caso da DMJ, uma vez que foi identificada independentemente por dois grupos a presença da ataxina-3 a

nível da mitocôndrias [Paulson *et al.*, 1997b; Trotter *et al.*, 1997]; por outro lado, a frataxina, o produto do gene que está alterado na ataxia de Friedreich; e que está praticamente ausente nas células dos indivíduos com esta doença, apresenta normalmente uma localização mitocondrial, pensando-se que a sua função se relaciona com a homeostase do ferro e com a actividade respiratória [Babcock *et al.*, 1997].

5. Interferência com o processo de apoptose: A apoptose é uma forma morfológica e bioquimicamente distinta de morte celular, que ocorre em vários tipos de células, em diferentes circunstâncias, fisiológicas e/ou patológicas. Vários estudos apontam para a desregulação dos mecanismos apoptóticos como componente importante na morte celular precoce das células nervosas nas doenças degenerativas progressivas. No caso particular da doença de Huntington, foram identificados no neo-estriado de indivíduos afectados sinais característicos de células em processo apoptótico (normalmente presentes apenas durante o desenvolvimento embrionário do sistema nervoso), nomeadamente a fragmentação do ADN por endonucleases, com positividade para a coloração TUNEL (*terminal transferase-mediated deoxyuridine triphosphate-biotin nick end labelling*) [Gavrieli *et al.*, 1992] e a presença de anomalias a nível da ultraestrutura celular, tais como vesiculação das membranas, condensação perinuclear da cromatina e aumento do volume dos organelos [Portera-Cailliau *et al.*, 1995]. Por outro lado, ratinhos nulizigóticos (*knock-out*) para o gene da doença de Huntington apresentam letalidade embrionária precoce, com aumento da apoptose, sugerindo que a huntingtina poderá ter influência nas vias apoptóticas.

Mais recentemente, foi demonstrado que a huntingtina é eficientemente clivada por uma protease com papel chave na apoptose, a apopaina, e que a eficiência desta digestão é dependente do comprimento da poliglutamina, sendo as sequências mais longas clivadas mais eficazmente pela enzima [Goldberg *et al.*, 1996; Rosen, 1996]. Sabendo-se que a proteólise específica de certos substratos é de importância central no desencadeamento da apoptose, torna-se necessário agora esclarecer as consequências específicas da digestão da huntingtina pela apopaina e, em particular, qual o seu efeito na cascata molecular que leva à morte celular programada.

6. Interação com RNA-binding proteins [McLaughlin *et al.*, 1996]: As interações entre o ARN e as proteínas de ligação ao ARN (*RNA-binding proteins: RBP*) desempenham um papel importante na expressão de vários genes eucarióticos. Um exemplo é o ARN codificante para a ferritina, que adota estruturas tridimensionais características (*stem-loop*) e interage com proteínas que regulam a sua própria expressão. Por outro lado, há RBPs que se ligam aos ARN_m e os transportam do núcleo para o citoplasma e que se pensa serem importantes na regulação da localização subcelular dos ARN_m, actuando como mediadoras em processos complexos de regulação da expressão, nomeadamente em células que apresentam polaridade funcional, como alguns neurónios. As estruturas tridimensionais dos ARN_m têm também influência na expressão dos genes a nível proteico, porque podem impedir a translocação eficiente dos ribossomas através da cadeia a traduzir. Assim, a demonstração, por McLaughlin e colaboradores [1996], da capacidade de associação de sequências de ARN_m contendo sequências repetitivas CAG com proteínas citoplasmáticas (RBPs), sendo essa associação específica do tecido e dependente do comprimento da sequência poliglutamina, atraiu o interesse para a possibilidade de o mecanismo patológico nas doenças associadas à expansão de CAGs envolver não as proteínas mas os ARN_m codificados por esses genes. Os *repeats* expandidos de CAGs poderiam "titular" proteínas citoplasmáticas, que normalmente se ligassem a outros ARN_m. A ligação excessiva dos (CAG)_n expandidos a proteínas reguladoras da expressão de outros genes poderia levar à desregulação da expressão destes; alternativamente, caso se tratasse de proteínas com função de *chaperone*, envolvidas na regulação da localização subcelular da expressão, ocorreria a distribuição subcelular alterada dos produtos génicos, com efeitos nocivos para a célula.

Estes modelos apresentam alguma sobreposição entre si, e não são de todo mutuamente exclusivos, sendo possível que vários dos efeitos descritos se combinem entre si para levar à toxicidade das poliglutaminas expandidas no sistema nervoso.

Recentemente, uma outra forma de ataxia hereditária progressiva, designada SCA6, foi também associada à presença de uma pequena expansão de uma sequência

CAG no gene CACNL1A4, que se localiza no cromossoma 19 e codifica para a subunidade central de um canal de cálcio dependente da voltagem (subunidade α_{1A}) [Zhuchenko *et al.*, 1996]. Não há ainda certeza no que respeita à tradução da sequência poliglutamina em questão: aparentemente, em algumas isoformas do gene esta sequência localiza-se na região 3' não-codificante e, noutras, na região codificante, sendo traduzida como uma sequência poliglutamina. No entanto, a diferença de tamanho do (CAG)_n entre os indivíduos normais e os afectados por esta doença é muito pequena, além de que os tamanhos dos *repeats* nos indivíduos afectados (21-27 unidades CAG) estão dentro dos limites da normalidade definidos para todas as outras doenças degenerativas do sistema nervoso associadas à expansão de uma sequência (CAG)_n; é por isso provável que a SCA6 seja causada por um mecanismo patológico diferente, possivelmente a ruptura do funcionamento do canal de cálcio, por um efeito dominante-negativo. Curiosamente, mutações pontuais *missense* em outras posições deste mesmo gene estão envolvidas em casos familiares de enxaqueca com aura hemiplégica, enquanto deleções e substituições que levam a um *splicing* anormal do gene estão associadas a ataxia episódica tipo 2 [Ophoff *et al.*, 1996].

Selectividade da morte neuronal associada à expansão de *repeats*

CAG/poliglutamina

Muito embora se assuma que o mecanismo de morte celular é o mesmo, ou pelo menos semelhante, em todas as doenças degenerativas do sistema nervoso associadas a sequências poliglutamina expandidas, dadas as semelhanças aparentes entre estas doenças, os grupos de neurónios afectados pelo efeito tóxico da poliglutamina diferem de doença para doença, levando a manifestações neurológicas diversas. Uma explicação possível para estas observações seria a expressão selectiva dos genes mutados em diferentes populações celulares, em que as células mais afectadas seriam aquelas em que se observa uma expressão mais elevada do gene. No entanto, os estudos de expressão destes genes revelaram em todos os casos uma expressão generalizada à maioria dos

tecidos do organismo, e em todas as regiões do SNC, afectadas ou não na doença [Strong *et al.*, 1993; The Huntington Disease Collaborative Research Group, 1993; Banfi *et al.*, 1994; Nagafuchi *et al.*, 1994; Landwehrmeyer *et al.*, 1995; Sharp *et al.*, 1995; Stine *et al.*, 1995], excluindo a especificidade de expressão como base da selectividade de morte neuronal.

INSTABILIDADE MITÓTICA E MEIÓTICA DOS *REPEATS* DE TRINUCLEOTÍDEOS EXPANDIDOS

As sequências repetitivas de trinucleotídeos expandidas apresentam uma instabilidade característica, variando de tamanho durante as divisões celulares, mitóticas ou meióticas. A instabilidade mitótica origina o mosaicismo somático do *repeat*, sendo que diferentes grupos de células apresentam diferentes tamanhos do *repeat*, e a instabilidade meiótica, possivelmente combinada com a instabilidade mitótica, pelo menos durante os primeiros dias de vida do embrião [Malter *et al.*, 1997], origina a instabilidade intergeracional do *repeat*: os descendentes de um indivíduo portador de uma expansão de um dado tamanho podem ter um *repeat* de tamanho diferente do do progenitor, maior ou menor, observando-se em geral uma tendência aumentada para a expansão. Embora estes efeitos, mosaicismo somático e instabilidade intergeracional, sejam mais marcados nas expansões, de maior tamanho, dos *repeats* CTG, CGG ou GAA (na distrofia miotónica, síndrome do X frágil ou ataxia de Friedreich), são também detectáveis, em menor grau, nas doenças associadas à expansão de *repeats* CAG/poliglutamina.

Mosaicismo somático do *repeat* CAG expandido

Embora tenha existido alguma controvérsia inicial sobre a existência ou não de mosaicismo somático na doença de Huntington, com vários estudos negando a instabilidade somática do *repeat* CAG expandido nesta doença [MacDonald *et al.*, 1993;

Zuhlke *et al.*, 1993], esta foi posteriormente demonstrada, em particular nos estudos *post-mortem* de casos juvenis desta doença, em que eram bastante evidentes as diferenças de tamanho e de grau de mosaicismos entre diferentes estruturas do sistema nervoso central [Schaefer *et al.*, 1994; Stine *et al.*, 1994; Telenius *et al.*, 1994]. Foi também demonstrado mosaicismos somáticos do tamanho do *repeat* CAG no SNC na ADRPL [Takano *et al.*, 1996] e na SCA1 [Chong *et al.*, 1995]. O nosso estudo (artigo IV) permitiu demonstrar a existência de mosaicismos somáticos na doença de Machado-Joseph. A possibilidade, levantada relativamente à doença de Huntington [Telenius *et al.*, 1994], de o mosaicismos ser a base da especificidade de morte neuronal, com as células portadoras de *repeats* maiores mais afectadas do que as portadoras de *repeats* de menor tamanho, foi também testada neste estudo, em que se tentou estabelecer a correlação entre o padrão de mosaicismos e o padrão de degenerescência neuronal observado em estudos detalhados de patologia no sistema nervoso central de doentes com MJD e SCA1. Procurou-se ainda elucidar o comportamento do *repeat* CAG expandido ao longo do tempo de vida dos portadores de expansões do (CAG)_n e esclarecer as contribuições dos diferentes tipos de células para o mosaicismos somáticos (artigo V).

Factores determinantes da instabilidade intergeracional dos *repeats* de trinucleotídeos expandidos

Apesar de não serem ainda compreendidos os mecanismos moleculares que originam a instabilidade dos *repeats* de trinucleotídeos, tem vindo a ser demonstrada a importância de diversos factores que contribuem para a instabilidade intergeracional nas doenças associadas a *repeats* expandidos. A importância da identificação de elementos que contribuam para a instabilidade genética dos *repeats* é, por um lado, a sua aplicação no processo de aconselhamento genético (permitindo estabelecer modelos de previsão da idade de início e gravidade da doença) e, por outro, levar a uma melhor compreensão dos mecanismos moleculares que estão na base desta instabilidade. Qualquer modelo teórico

proposto para explicar o fenómeno da instabilidade genética das sequências repetitivas de trinucleotídeos terá que levar em conta estas observações.

Os principais factores com efeito na instabilidade intergeracional dos *repeats* de trinucleotídeos são:

- O tipo de sequência repetitiva, com diferenças nítidas entre o comportamento genético dos *repeats* CAG expandidos (que variam de um pequeno número de unidades) e os *repeats* CGG, CTG ou GAA (cujas variações são muito maiores, podendo atingir as centenas ou milhares de unidades).

- O tamanho do *repeat* no progenitor [Fu *et al.*, 1991; Hunter *et al.*, 1992; Yu *et al.*, 1992; Brunner *et al.*, 1993; Duyao *et al.*, 1993; Lavedan *et al.*, 1993; Snow *et al.*, 1993; Fisch *et al.*, 1995; Pianese *et al.*, 1997]: *repeats* maiores são mais susceptíveis a instabilidade.

- O sexo do transmissor da mutação: em geral, transmissões paternas estão associada a uma maior instabilidade, com excepção do síndrome do X frágil, em que as grandes transições de pré-mutação a mutação ocorrem em transmissões maternas [Fu *et al.*, 1991; Brunner *et al.*, 1993; Chung *et al.*, 1993; De Rooij *et al.*, 1993; Duyao *et al.*, 1993; Goldberg *et al.*, 1993; Lavedan *et al.*, 1993; Snow *et al.*, 1993; Fisch *et al.*, 1995; Maciel *et al.*, 1995; Maruyama *et al.*, 1995; Ranen *et al.*, 1995; Igarashi *et al.*, 1996a; Pianese *et al.*, 1997].

- A presença de interrupções da sequência repetitiva [Chung *et al.*, 1993; Eichler *et al.*, 1994; Kunst e Warren, 1994; Snow *et al.*, 1994; Leeflang e Arnheim, 1995; Nancarrow *et al.*, 1995], que têm um efeito estabilizador do *repeat*.

- As sequências flanqueantes em que se insere o *repeat*: determinados haplótipos estão associados a maiores tamanhos dos *repeats* (CAG, CGG ou CTG) na população normal e/ou a uma predisposição dos alelos de tamanho intermédio para uma maior instabilidade, verificando-se uma frequência aumentada de transições para a expansão em portadores desses haplótipos, nos genes da distrofia miotónica [Imbert *et al.*, 1993; Neville *et al.*, 1994; Rubinsztein *et al.*, 1994], da doença de Huntington [Squitieri *et al.*,

1994; Almqvist *et al.*, 1995; Rubinsztein *et al.*, 1995b], do síndrome do X frágil A [Richards *et al.*, 1992; MacPherson *et al.*, 1994; Snow *et al.*, 1994] e E [Murray *et al.*, 1997], da ADRPL [Yanagisawa *et al.*, 1996] e da doença de Machado-Joseph [Igarashi *et al.*, 1996a].

Modelos propostos para a instabilidade genética dos *repeats* expandidos

Para tentar compreender os mecanismos moleculares que levam à instabilidade das chamadas mutações dinâmicas, é importante ter em consideração os mecanismos de manutenção da fidelidade de replicação nas células eucariotas e de que forma os *repeats* expandidos escapam a este controlo.

Existem, por um lado, mecanismos celulares de verificação da exactidão da replicação que actuam durante a replicação propriamente dita, nomeadamente a actividade exonucleolítica de *proofreading* das DNA polimerases. Por outro lado, e após a replicação, o mecanismo de *mismatch repair* contribui para a fidelidade da replicação por excisão de ADN mal-emparelhado. Leveduras (*S. cerevisiae*) mutantes para as proteínas envolvidas no processo de *mismatch repair* (*pms1*, *mlh1* e *msh2*) apresentam uma taxa de mutação mais elevada (100 a 700 vezes) dos *repeats* de di-, tri- ou tetra-nucleotídeos [Strand *et al.*, 1993; Schweitzer e Livingston, 1997], enquanto mutações afectando a actividade de *proof-reading* das polimerases têm pouco efeito na estabilidade mitótica ou meiótica destes *repeats*, mas aumentam a frequência de substituições. Estes resultados sugerem que a actividade de *proofreading* da polimerase é mais relevante para corrigir substituições pontuais do que para o tipo de mutações que ocorrem nos microsatélites - normalmente a variação (aumento ou diminuição) do número de unidades repetitivas.

Em humanos, as mutações no gene *hMSH2* estão associadas a uma instabilidade generalizada a nível somático, com taxas de mutação elevadas, e causam cancro colo-rectal (não-poliposo). No entanto, a comparação da estabilidade de 10 *loci* de microsatélites, incluindo os *repeats* de trinucleotídeos nos genes da doença de Huntington, *SCA1* e da *TATA-binding protein* (*TBP*), em indivíduos com doença de

Huntington e com cancro colo-rectal [Goellner *et al.*, 1997] revelou que estes últimos apresentavam um perfil característico de instabilidade, generalizada a todos os *loci* estudados, enquanto os indivíduos com doença de Huntington só apresentavam instabilidade no *locus* da doença. Por outro lado, o tipo de instabilidade observado não era completamente semelhante: enquanto na doença de Huntington se observava uma maior proporção de expansões que de contracções do *repeat* CAG, no cancro colo-rectal as deleções de unidades dos *repeats* eram o tipo de mutação mais observado. Pode pois concluir-se que estas doenças correspondem a problemas diferentes em termos de instabilidade, um com origem em factores celulares deficientes, outro com origem na própria estrutura do *repeat* e sequências flanqueantes.

Foi demonstrado que a sequência e comprimento dos *repeats* determinam a taxa de *slippage* da polimerase [Schlotterer e Tautz, 1992]; o tipo de *repeat*, a presença de interrupções e as sequências flanqueantes são também, como se disse, factores moduladores da estabilidade genética.

É possível que as estruturas secundárias e conformações terciárias alternativas adoptadas pelo ADN em sequências repetitivas (*hairpin loops*, *triplex DNA*, estruturas cruciformes, *toroids*, etc) afectem de alguma forma a sua replicação ou reduzam a capacidade de correcção de possíveis erros pelos mecanismos celulares habituais [McMurray, 1995; Wells, 1996]. Os *repeats* CTG, CAG e CGG nos genes da distrofia miotónica, doença de Huntington, SCA1, DRPLA e síndrome do X frágil têm em comum a capacidade de formar estruturas em *hairpin loop* [Chen *et al.*, 1995; Gacy *et al.*, 1995] com uma estrutura e energia de estabilização semelhantes. A energia de estabilização é influenciada quer pela sequência (tipo de *repeat*, presença de interrupções) e tamanho do *repeat* propriamente dito, quer por sequências flanqueantes. Só o ADN com determinadas sequências e a partir de um determinado tamanho pode formar *hairpins* cuja energia esteja abaixo do limiar de estabilidade (aproximadamente -50kcal/*hairpin*). Curiosamente, existe uma correlação entre a capacidade de estas sequências formarem *hairpins* de estabilidade apropriada e as observações genéticas em doenças associadas a

repeats expandidos, em termos da selectividade de sequência e do efeito do tamanho do *repeat* na instabilidade [McMurray, 1995].

A formação das estruturas em *hairpin loop* por cadeias simples de ADN durante a replicação poderia levar quer ao deslizamento aumentado da DNA polimerase, devido à estabilização energética das cadeias simples ou ao bloqueio estrutural à passagem da polimerase, quer à introdução de erros no realinhamento da cadeia. A combinação destes efeitos poderá contribuir para a instabilidade observada nas sequências repetitivas de trinucleotídeos.

O *repeat* GAA no gene da ataxia de Friedreich forma estruturas secundárias diferentes, em particular triplexes de ADN, com emparelhamento de uma terceira cadeia de ADN que pode ser de origem externa ou intra-molecular, bem como outras estruturas alternativas, no caso de existir uma expansão [Ohshima e Wells, 1997]. Foi demonstrado que os triplexes de ADN causam pausas prolongadas da DNA polimerase [Dayn *et al.*, 1992], dificultando a replicação das sequências que adoptam esta estrutura.

Instabilidade dos *repeats* CAG expandidos em modelos animais

Surpreendentemente, muitos dos modelos animais obtidos até à data para as doenças associadas à expansão de trinucleotídeos, em particular ratinhos transgénicos contendo o ADN_C dos genes com CAGs expandidos, apresentavam (além de uma ausência notória de fenótipo, na maior parte dos casos) uma estabilidade absoluta do *repeat* durante as transmissões intergeracionais deste [Bingham *et al.*, 1995; Goldberg *et al.*, 1996; Ikeda *et al.*, 1996].

Só recentemente foram descritos modelos animais com instabilidade meiótica (e mitótica) de *repeats* de trinucleotídeos [Korneluk e Narang, 1997]: os ratinhos transgénicos portadores de um *repeat* CAG expandido (82 unidades) no primeiro exão do gene da doença de Huntington [Mangiarini *et al.*, 1997] e os ratinhos transgénicos contendo um fragmento de ADN genómico incluindo a região não-traduzida 3' do gene da distrofia miotónica, com um *repeat* de 162 tripletos CTG [Monckton *et al.*, 1997], ou

um *clone* de cosmídeo contendo um fragmento de ADN genómico que incluía a totalidade do gene da distrofia miotónica de um indivíduo com um alelo intermédio (55 CTGs), e ainda dois genes flanqueantes deste [Gourdon *et al.*, 1997].

O facto de se observar instabilidade nestes modelos e não nos precedentes pode estar relacionado com efeitos das sequências flanqueantes, incluídas no transgene juntamente com os genes contendo os *repeats* expandidos, ou presentes nos locais de inserção deste, o que estaria em conformidade com os efeitos de haplótipo descritos em humanos e seria explicável, por exemplo, por efeitos de estruturas tridimensionais adoptadas pelo ADN [McMurray, 1995; Wells, 1996].

Uma explicação alternativa, avançada por Korneluk e Narang [1997], é que exista uma relação entre a actividade transcripcional dos genes contendo os *repeats* e a existência de instabilidade. Nos modelos anteriores, a expressão dos transgenes era bastante reduzida ou ausente, possivelmente devido à escolha inapropriada de promotores ou à ausência de elementos essenciais de regulação da expressão, localizados nas regiões genómicas flanqueantes ou nos intrões do gene (que estão ausentes nos modelos baseados na inserção de ADN_{CS}), problemas esses que foram ultrapassados nos três casos acima referidos. É de salientar que, no estudo por Mangiarini e colaboradores [1997], na única linha de ratinhos que não apresentava expressão significativa do transgene (possivelmente devido a efeitos repressores de sequências flanqueantes no local de inserção), o *repeat* CAG expandido era estável.

Apesar de se observar instabilidade intergeracional do *repeat* nestes modelos e de no caso do transgene com o exão 1 do gene HD se ter observado um fenótipo neurológico progressivo semelhante ao da doença de Huntington [Mangiarini *et al.*, 1996], não foi possível recriar um efeito de antecipação como seria de esperar. Os ratinhos portadores de pré-mutações no gene da distrofia miotónica não apresentavam fenótipo correspondente à doença, mas as expansões detectadas eram ainda pequenas, esperando-se no entanto que na sua descendência surjam expansões de grande tamanho, possivelmente associadas a um fenótipo anormal [Korneluk e Narang, 1997].

Instabilidade genética do *repeat* CAG e origem da mutação DMJ

Uma das questões de relevo no estudo da doença de Machado-Joseph tem sido, desde sempre, a origem da mutação causadora desta afecção. Estes estudos basearam-se, até à identificação e clonagem do gene responsável, em dados epidemiológicos e históricos que sugeriam que os casos portugueses de DMJ pudessem ser derivados de uma mesma mutação e que essa mutação tenha sido introduzida mais tarde em algumas famílias de origem não portuguesa; pelo menos parte da distribuição mundial actual da DMJ poderia ser explicada pelos laços históricos estabelecidos pelos portugueses nas suas expedições marítimas desde meados do século XV e nas sucessivas vagas de emigração portuguesa mais recentes [Sequeiros e Coutinho, 1993]. Uma doença dominante de início tardio como a DMJ consegue facilmente atingir uma frequência relativamente alta em populações muito diversas. No entanto, em algumas famílias DMJ a mutação poderia ter tido origem num acontecimento mutacional independente, sendo esta uma explicação mais razoável para a presença da doença nas famílias em que não se conseguiu encontrar nenhuma ligação com doentes portugueses.

Com a identificação do gene responsável pela DMJ começam a gerar-se dados moleculares que dão resposta à questão de um possível efeito fundador na DMJ. A existência de uma mutação fundadora poderia, por exemplo, explicar o grande intervalo entre a distribuição do tamanho dos *repeats* CAG nos alelos normais e nos alelos expandidos que se verifica na DMJ [Cancel *et al.*, 1995; Maciel *et al.*, 1995; Takiyama *et al.*, 1995]. Esta diferença entre as duas distribuições sugeria uma única mutação ancestral, com transição directa da gama de tamanhos normais à gama de tamanhos associada à doença, em oposição à ocorrência de numerosas expansões a partir de alelos de tamanho pré-mutação (alelos intermédios).

Os estudos recentes de associação alélica entre o locus *MJD1* e marcadores flanqueantes apontam, no entanto, para a existência de mais do que um haplótipo associado à doença em diversos grupos étnicos [Stevanin *et al.*, 1995a; Takiyama *et al.*, 1995; Gaspar *et al.*, 1996]. Para clarificar e aprofundar estes resultados levou-se a cabo

um estudo colaborativo detalhado de co-segregação de alelos de marcadores (flanqueantes e intragénicos), e a determinação dos haplótipos associados à mutação num vasto número de famílias, representativas dos diferentes grupos étnicos em que a doença foi identificada [Gaspar *et al.*, 1997]. Resultados preliminares apontam para a presença de pelo menos dois haplótipos na população portuguesa e de haplotipos adicionais na população mundial. Estes resultados, assim como o facto de a DMJ ser a causa mais frequente de ataxia espinocerebelosa autossómica dominante [Ranum *et al.*, 1995; Schols *et al.*, 1995; Silveira *et al.*, 1996], sugerem fortemente a existência de mais do que uma mutação original responsável pela DMJ na população mundial.

No entanto, a grande maioria das famílias DMJ (70%) [Gaspar *et al.*, 1997], mesmo de etnias cuja relação entre si ou com os portugueses não é à partida evidente, apresentam um mesmo haplótipo intragénico comum. Existem duas explicações possíveis: ou (1) se trata de uma mutação ancestral muito antiga a segregar nestas famílias ou (2) de mais que um evento mutacional ocorrendo no mesmo haplótipo, o que implicaria uma "desestabilização" do *repeat* CAG na população por parte deste haplótipo, que poderia ser devida, conforme discutido previamente, à influência das sequências flanqueantes na estrutura secundária e terciária adoptada pelo *repeat*, com efeito na fidelidade de replicação deste.

Para tentar esclarecer a questão da existência ou não de haplótipos "desestabilizadores" na DMJ, compararam-se os tamanhos de *repeats* CAG no gene MJD1 associados a diferentes haplótipos, na população normal (artigo VI), e o grau de instabilidade intergeracional do *repeat* expandido associado em *cis* aos diferentes haplótipos.

Objectivos do estudo

Assim, os objectivos principais deste estudo, a que adiante se procura dar resposta, foram os seguintes:

1. Localização e identificação do gene causador da doença de Machado-Joseph
2. Contribuição para o estabelecimento de um teste genético para a DMJ em indivíduos afectados e do teste preditivo e do diagnóstico pré-natal para os seus familiares em risco (para fins de aconselhamento genético).
3. Estudo das bases moleculares da variabilidade clínica na DMJ, em termos de idade de início e subtipo, com vista ao esclarecimento do mecanismo através do qual a mutação leva à manifestação da doença.
4. Tentativa de esclarecimento das bases moleculares da especificidade da degenerescência neuronal na DMJ: estudo do mosaicismo somático do $(CAG)_n$ expandido no SNC de doentes DMJ e SCA1, como possível causador desta especificidade a nível patológico.
5. Estudo dos mecanismos moleculares associados a algumas características não-mendelianas de transmissão desta doença, nomeadamente o fenómeno de antecipação: estudos da instabilidade intergeracional do $(CAG)_n$ expandido no gene MJD1.

II. TRABALHO EXPERIMENTAL

CAPÍTULO 1

Estudos de ligação genética e esforço de clonagem do gene

Estudos de ligação genética do *locus* DMJ a marcadores do cromossoma 14q

Introdução

Os primeiros estudos de ligação genética da DMJ foram levados a cabo recorrendo a polimorfismos celulares e séricos convencionais (marcadores proteicos) [Myers *et al.*, 1993; Amorim *et al.*, 1995]. Nestes estudos excluiu-se a localização em certas regiões de diversos cromossomas (2, 8, 9, 13, 16 e, sobretudo, do cromossoma 1), mas não se pôde identificar a localização do gene causador da DMJ.

Adoptou-se posteriormente uma estratégia de rastreio do genoma por análise de marcadores de ADN (RFLPs e, posteriormente, VNTRs), tendo sido estabelecida uma ordem de prioridade para o estudo dos diferentes cromossomas, de acordo com as análises de exclusão prévias e com o tamanho dos cromossomas: dado que os cromossomas humanos são ordenados por ordem decrescente de tamanho (com excepção dos cromossomas 21 e 22), a probabilidade de um *locus* desconhecido se encontrar nos primeiros cromossomas é sempre maior. Foram seleccionados marcadores genéticos polimórficos distribuídos por todo o genoma, separados entre si, em média, de 20 cM.

As localizações cromossómicas dos genes causadores de duas formas de ataxia espinocerebelosa dominante, a OPCA clássica [Jackson *et al.*, 1977] e a ataxia de Holguin [Auburger *et al.*, 1990], previamente estabelecidas nos cromossomas 6 e 12, respectivamente (que passaram mais tarde a ser designadas SCA1 e SCA2), foram excluídas para o gene da DMJ [Carson *et al.*, 1992; Silveira *et al.*, 1993; Twist *et al.*, 1994]. Foi também excluída a região onde se localizava o gene da doença de Huntington [Forse *et al.*, 1989]. Prosseguiu-se então o rastreio do genoma com o objectivo de identificar a posição do *locus* DMJ. Aproximadamente um terço do genoma foi excluído pelo consórcio de estudos de ligação da DMJ estabelecido entre os grupos portugueses e os grupos francês e canadiano [Sequeiros, resultados não publicados].

Em 1993, foi identificada a localização do gene causador da DMJ em famílias japonesas no cromossoma 14q32.1 [Takiyama *et al.*, 1993], num intervalo de cerca de 30 cM entre os marcadores D14S51 e D14S43. Com base em dados históricos [Sequeiros e Coutinho, 1993] e em características clínicas comuns [Yuasa *et al.*, 1990], havia razões para crer que o locus DMJ pudesse situar-se também nesta região nas famílias de origem portuguesa. Para testar esta hipótese realizaram-se estudos de ligação genética a marcadores do cromossoma 14q em 16 famílias açorianas com DMJ [Sequeiros *et al.*, 1994b].

Material e métodos

Foram incluídas neste estudo 16 famílias com diagnóstico clínico de DMJ, cujas porções relevantes para a análise de ligação incluíam 435 indivíduos, 166 afectados e 14 codificados como tendo "diagnóstico suspeito", porque embora em risco e ainda assintomáticos, apresentavam pelo menos dois dos seguintes sinais: hiperreflexia, clonus, nistagmo ou retracção palpebral. Colheu-se sangue e extraiu-se ADN de linfócitos segundo os protocolos habituais [Sambrook *et al.*, 1989]. Utilizaram-se os marcadores D14S53, D14S55, D14S48 e D14S45, cujos *primers* foram sintetizados de acordo com as sequências publicadas [Luty e Litt, 1991; Wang e Weber, 1992] utilizando um sintetizador de oligonucleotídeos (*Applied Biosystems 381A*). A amplificação por PCR dos segmentos de DNA contendo os microsatélites foi realizada a partir do ADN genómico conforme descrito previamente [Weber e May, 1989]. Amostras de ADN de indivíduos das famílias japonesas com DMJ foram-nos cedidas pelo grupo da Universidade de Niigata, a fim de permitir a comparação do tamanho dos alelos. As análises de ligação genética foram realizadas utilizando os programas de LINKAGE (versão 5.10) [Lathrop e Lalouel, 1984] num PC 486/66. Para o locus DMJ, utilizámos os parâmetros previamente definidos [Silveira *et al.*, 1993] e um modelo de transmissão autossómica dominante com 5, 10 ou 20 classes de risco, para corrigir o facto de a penetrância ser dependente da idade, e considerando o risco de início mais precoce da

doença para nos homozigotos para a mutação. Para as análises de *multipoint lod* e de *location score*, os alelos dos marcadores foram recodificados para reduzir o seu número a um máximo de 4, utilizando as estratégias recomendadas [Ott, 1991]. Para fazer análises de *affected-only*, em que a descendência normal não contribui com informação respeitante à doença, mas apenas com informação sobre o genótipo e a fase de *linkage* dos marcadores nos seus progenitores (muitas vezes não disponíveis para estudo), reduziu-se o risco das várias classes a um centésimo do seu valor, tornando assim irrelevante o fenótipo dos sujeitos em risco. Para cada marcador, as frequências alélicas foram estimadas utilizando os indivíduos colaterais das famílias (60 cromossomas), na ausência de outros controlos relevantes para a população utilizada. Utilizaram-se os programas HOMOG versão 3.33 [Ott, 1991] para testar a possibilidade de heterogeneidade genética entre 14 famílias portuguesas e 5 famílias japonesas informativas para o marcador D14S48. O programa MTEST foi usado para comparar dois grupos, considerando-se cada família como um grupo [Sequeiros *et al.*, 1994b].

Resultados

Na tabela 3 podem ver-se os resultados das análises *pairwise*. Os *lod scores* mais elevados foram obtidos com o marcador D14S48, o marcador previamente identificado como o mais próximo do *locus* da DMJ (estimado a zero cM nas famílias japonesas). A tabela 4 mostra os *lod scores* obtidos com o marcador D14S48 para cada família, considerando 20 classes de risco, frequências estimadas dos alelos e análise só de afectados. Na análise de *multipoint*, obteve-se um valor de $Z_{\max} = 3.19$, para o intervalo entre D14S48 e D14S45 (figura 1), embora não se pudessem excluir as outras localizações. A análise com o programa HOMOG não revelou a existência de heterogeneidade genética entre as famílias japonesas e portuguesas ($X_1^2 = 0.62$, $p=0.43$, $L_{\text{ratio}}=1.36$). Com o MTEST, $X_{18}^2 = 18.28$ ($p=0.44$) quando se considerou cada família um grupo.

TABELA 3

Resultados das análises de ligação *pairwise*
entre os marcadores do cromossoma 14 estudados e o locus da DMJ

Lod scores para valores de θ de:

<i>Locus</i>	0.00	0.01	0.05	0.1	0.2	0.3	0.4	<i>Zmax</i> θ
<i>Todos os indivíduos, alelos equifrequentes, 5 classes de risco</i>								
D14S53	-24.97	-8.18	-2.22	0.06	1.35	1.18	0.58	1.39 (0.226)
D14S55	-4.75	-2.23	-0.15	0.63	0.86	0.55	0.19	0.89 (0.171)
D14S48	-5.98	0.59	3.23	3.80	3.24	2.02	0.81	3.81 (0.110)
D14S45	-22.13	-6.59	-0.66	1.33	1.97	1.23	0.30	2.01 (0.177)
<i>Affected only, alelos equifrequentes, 5 classes de risco</i>								
D14S53	-7.61	-2.53	1.28	2.39	2.38	1.55	0.65	2.58 (0.144)
D14S55	1.67	1.97	2.23	2.14	1.49	0.76	0.24	2.23 (0.059)
D14S48	1.61	3.26	4.57	4.66	3.62	2.15	0.83	4.72 (0.079)
D14S45	-10.52	-4.58	-0.28	1.26	1.69	1.00	0.22	1.75 (0.169)
<i>Affected only, frequências de alelos estimadas, 5 classes de risco</i>								
D14S53	-8.78	-3.90	0.02	1.41	1.83	1.30	0.58	1.87 (0.175)
D14S55	-4.36	-3.26	-1.55	-0.60	0.08	0.18	0.11	0.18 (0.285)
D14S48	-3.02	-1.27	0.46	1.10	1.14	0.71	0.26	1.23 (0.147)
D14S45	-13.83	-7.82	-2.92	-0.90	0.28	0.22	-0.05	0.28 (0.200)
<i>Affected only, frequências de alelos estimadas, 20 classes de risco</i>								
D14S53	-9.04	-4.01	-0.01	1.40	1.84	1.31	0.58	1.86 (0.176)
D14S55	-4.48	-3.37	-1.62	-0.64	0.06	0.17	0.10	0.17 (0.289)
D14S48	-3.13	-1.35	0.45	1.10	1.14	0.71	0.26	1.23 (0.147)
D14S45	-14.10	-7.92	-2.95	-0.91	0.28	0.22	-0.05	0.28 (0.200)

TABELA 4

Lod Scores para famílias informativas com D14S48

<i>Pedigree</i>	θ							<i>Zmax</i> (θ)
	0.00	0.01	0.05	0.1	0.2	0.3	0.4	
1	-0.36	-0.32	0.10	0.35	0.40	0.28	0.13	
8	0.64	0.62	0.55	0.46	0.31	0.17	0.06	
10	-0.01	-0.01	-0.01	-0.01	-0.00	-0.00	-0.00	
14	-1.46	-0.71	-0.17	-0.01	0.06	0.05	0.02	
16	0.72	0.71	0.63	0.53	0.34	0.17	0.04	
17	0.55	0.54	0.48	0.40	0.25	0.13	0.04	
21	-0.12	-0.11	-0.09	-0.07	-0.04	-0.02	-0.00	
25	-0.23	-0.22	-0.15	-0.1	-0.04	-0.01	-0.00	
26	-1.40	-1.17	-0.75	-0.50	-0.25	-0.12	-0.04	
33	0.18	0.18	0.15	0.13	0.08	0.03	0.01	
35	0.17	0.17	0.14	0.12	0.07	0.03	0.01	
36	-2.03	-1.21	-0.60	-0.35	-0.13	-0.04	-0.01	
44	0.21	0.20	0.17	0.14	0.09	0.04	0.01	
Total	-3.13	-1.35	0.45	1.10	1.14	0.71	0.26	1.23 (0.147)

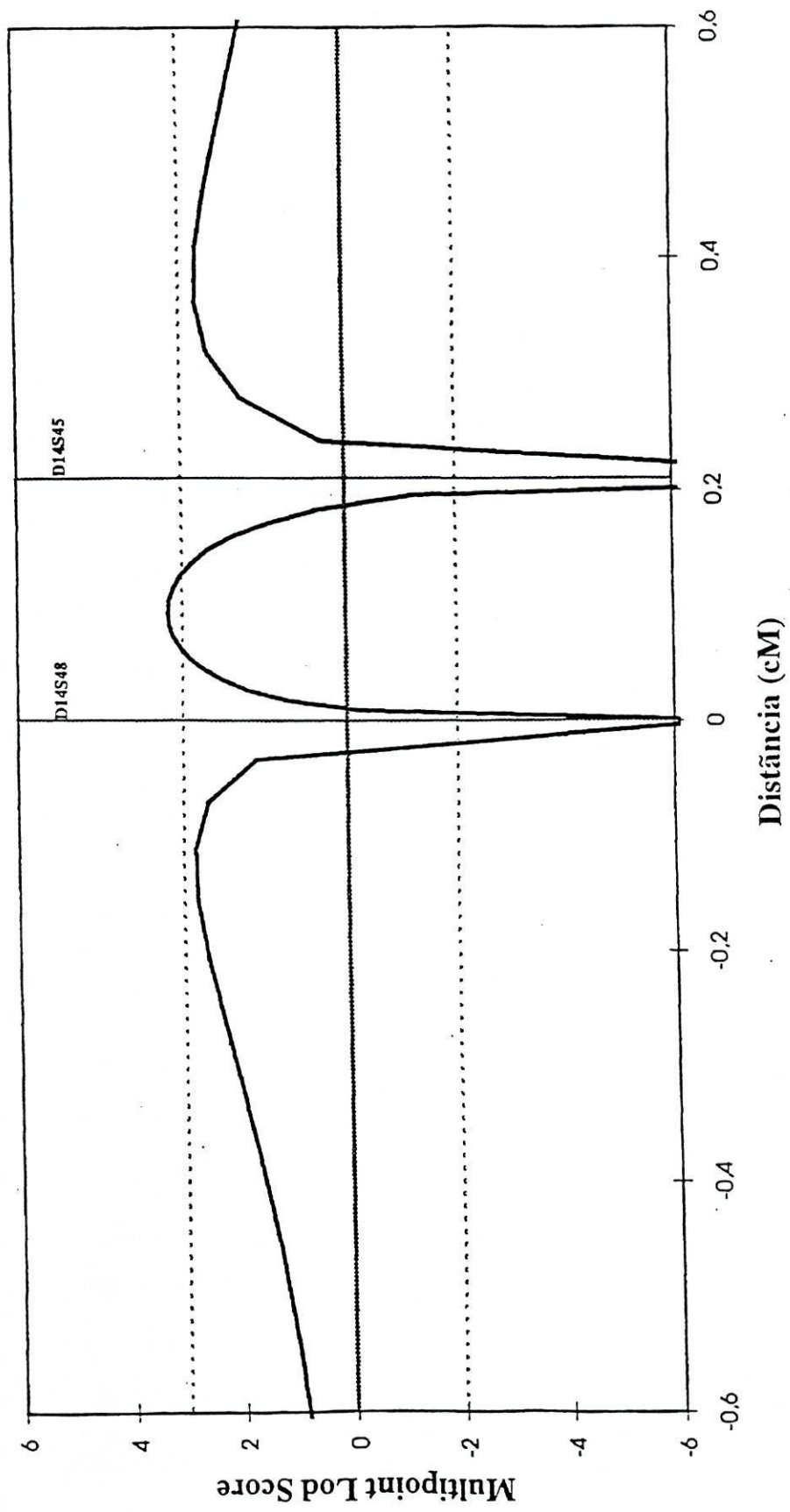


Fig. 1

Discussão

O estudo de marcadores do cromossoma 14q em famílias DMJ portuguesas confirma a localização do gene causador desta doença na mesma região para a qual se demonstrou ligação em famílias japonesas. No entanto, dado o poder teórico de *linkage* das famílias utilizadas, os *lod scores* obtidos neste estudo foram surpreendentemente baixos.

Diferenças em termos de idade de início e distribuição dos subfenótipos podem ter contribuído para estes resultados. Por um lado, a frequência das apresentações clínicas mais graves e clinicamente mais distinguíveis das de outras formas de ataxia dominante (as formas distónicas, por exemplo, características do subfenótipo 1) é mais elevada entre os japoneses, enquanto os subfenótipos 2 e 3, que são de mais difícil diagnóstico em casos menos graves ou incipientes, predominam nos doentes portugueses. Em certos casos, em indivíduos de idade mais avançada, com sintomas inespecíficos, não era possível estabelecer um diagnóstico seguro; estes indivíduos foram codificados como tendo diagnóstico "suspeito". Por outro lado, a distribuição das idades de início é muito mais ampla e de média significativamente mais elevada nas famílias de origem portuguesa (39.1 ± 0.8 anos) do que nas famílias japonesas (26.5 ± 3.3 anos) [Sequeiros e Coutinho, 1993], existindo nas nossas famílias portadores obrigatórios do gene que, com idades entre os 63 e 70 anos, são ainda assintomáticos. A idade de início mais tardia provoca uma redução da informatividade das famílias devido à penetrância incompleta (dificuldade que se procurou ultrapassar com a definição de 20 classes de risco e com as análises de *affected only*, as quais permitiram em certos casos obter *lod scores* mais elevados), mas sobretudo intervem também na dificuldade acrescida de encontrar indivíduos afectados em duas ou mais gerações nas famílias, essenciais para estudos de análise de ligação.

Posteriormente à clonagem do ADN_C do gene MJD1 e à possibilidade de se realizar o teste directo da mutação, veio-se ainda a verificar que existia heterogeneidade genética nas famílias utilizadas nos estudos de ligação genética. Em particular, numa das famílias seleccionadas para os estudos de ligação a marcadores do cromossoma 14q, a

família P33, os indivíduos afectados não apresentavam expansão do *repeat* CAG no gene MJD1, demonstrando-se posteriormente que eram portadores de mutações no gene SCA2; curiosamente, os *lod scores* obtidos com marcadores do cromossoma 14q apenas para esta família não permitiam a exclusão desta localização. Além desta dificuldade, inultrapassável dada a sobreposição fenotípica entre os diversos tipos de ataxias dominantes, e inerente a este tipo de abordagem, foram também detectados *a posteriori* alguns erros de codificação e troca de amostras, já que alguns indivíduos notoriamente afectados e pertencentes a famílias com portadores comprovados da mutação DMJ não eram portadores desta. Estes dois factores podem ter contribuído adicionalmente para baixar o valor dos *lod scores* obtidos com os marcadores do cromossoma 14q neste conjunto de famílias.

A localização no cromossoma 14q foi posteriormente confirmada em famílias brasileiras e norte-americanas de ascendência portuguesa por Twist e colaboradores [1995] que, recorrendo à análise de recombinantes, refinaram a localização do gene da DMJ a um intervalo de 11 cM, demonstrando-se também que a distância genética entre o marcador D14S48 e o *locus* da doença era mais elevada do que fora estimado nos estudos em famílias japonesas.

Uma região cromossómica de 11cM não é suficientemente restrita para justificar o início do mapeamento físico, tornando-se necessário identificar mais marcadores polimórficos na região, a utilizar na identificação de recombinantes raros e em estudos de desequilíbrio de ligação que permitissem reduzir a região crítica. No entanto, assumindo que as famílias francesas com uma ataxia ligada à mesma região do cromossoma 14, que foram designadas como "SCA3" [Stevanin *et al.*, 1994], fossem portadoras da mesma mutação que as famílias DMJ [Twist *et al.*, 1995], a combinação da informação obtida nas nossas famílias com a informação obtida nas famílias SCA3 permitiria reduzir a região a um intervalo de 2 cM, o que tornaria possível iniciar o mapeamento físico da região, e a identificação de genes candidatos. Na secção seguinte descreve-se a estratégia utilizada na identificação de potenciais genes candidatos para a DMJ.

Estudo de *repeats* de trinucleotídeos do cromossoma 14 em famílias com DMJ

Introdução

Várias das características genéticas da DMJ sugeriam que a mutação causadora desta doença fosse uma expansão de uma sequência repetitiva de trinucleotídeos, possivelmente uma expansão de uma sequência (CAG)_n na região codificante. São neste sentido de realçar a variabilidade de apresentação clínica, a antecipação, e a semelhança fenotípica com outras doenças causadas por este tipo de mutação, nomeadamente com outra ataxia espinocerebelosa de transmissão dominante, a SCA1, cujas manifestações clínicas são muito semelhantes às da DMJ, e com a ADRPL (atrofia dentatorubropalidolusiana), uma ataxia dominante em que há epilepsia mioclónica e outros sintomas. Considerou-se assim provável que este tipo de mutação estivesse também na base do mecanismo patogénico da DMJ e iniciou-se, em paralelo com o mapeamento físico da região de interesse (levado a cabo por outros elementos do nosso grupo), o estudo de sequências repetitivas de trinucleotídeos do cromossoma 14 em doentes com DMJ.

Material e métodos

Em colaboração com o grupo de Tom Hudson, Universidade de Harvard, que identificou *repeats* de trinucleotídeos num banco de ADN genómico, utilizando o protocolo descrito na bibliografia [Hudson *et al.*, 1992] e os localizou, recorrendo a painéis de híbridos celulares somáticos monocromossómicos, testámos 6 *repeats* localizados no cromossoma 14 em doentes DMJ e em controlos, a fim de identificar potenciais expansões do número de unidades trinucleotídicas presentes nos doentes mas não nos controlos. Para isso, seleccionámos 12 doentes, de origem portuguesa (famílias P1, P4, P33 e P36), brasileira (família B1) e açoriana/americana (família MJD2), com

início mais precoce da doença (entre 15 e 26 anos) e 18 controlos de origens caucasianas muito variadas. As designações e características dos *repeats* testados estão representadas na tabela 6.

Amplificaram-se os *repeats* a partir de DNA genómico, em condições de PCR habituais: 10 min desnaturação a 94°C seguida de 30 ciclos de 1 min de desnaturação a 94°C, 1 min de *annealing* a temperaturas entre 50 e 65°C e 1 min de elongação a 72°C e um período de elongação final a 72°C por 5 min; variou-se a temperatura de *annealing* para obter em cada caso a melhor amplificação possível, sem bandas inespecíficas que poderiam, neste caso em particular mais ainda que noutras situações, complicar a análise dos resultados. A amplificação foi levada a cabo numa solução contendo 1X tampão de PCR (*Perkin Elmer*), dNTP-A 0.2 mM, dATP 0.03 mM, 8 ng/ μ l de cada um dos *primers*, 0.2 μ Ci de 35 S dATP, 0.1 μ g/ μ l de BSA, 0.12 U AmpliTaq (*Perkin Elmer*). Na tabela 5 apresentam-se as sequências dos *primers* utilizados em cada caso, assim como as temperaturas de *annealing* e possíveis reagentes adicionais necessários para a amplificação eficaz da sequência repetitiva.

Utilizaram-se dois métodos de detecção diferentes: incorporação de 35 S no produto de PCR e migração em gel desnaturante de poliacrilamida 6% seguida de autorradiografia, ou PCR a frio seguida de migração em gel desnaturante de poliacrilamida 6%, transferência para membrana *Hybond+* e hibridação a 42°C com uma sonda (CAG)₁₅ marcada com γ 32 P-dATP utilizando a enzima terminal transferase.

Para confirmar a localização dos *repeats* testados no cromossoma 14, amplificaram-se estes por PCR em ADN de um híbrido monocromossómico contendo o cromossoma 14, e de um painel de células somáticas híbridas humano/roedor do *National Institute of General Medical Sciences - NIGMS* [Drwinga *et al.*, 1993].

A fim de se identificarem ainda outros *repeats* de trinucleotídeos, em particular *repeats* CAG, localizados no cromossoma 14, levou-se a cabo o rastreio de um banco de cosmídeos contendo DNA genómico do cromossoma 14 (obtido do Los Alamos National Laboratory), recorrendo a um método de hibridação apropriado para sondas repetitivas, que permite diminuir a hibridação não específica: realizou-se uma pré-hibridação da

Tabela 5. Sequência dos primers e condições de amplificação por PCR (temperatura de *annealing* e reagentes adicionais quando necessário) dos repeats de trinucleotídeos testados. Os repeats que apresentavam polimorfismo na população controle estão marcados com #.

Designação	Sequência dos primers	TA (°C)	Outros reagentes
GCT2C07	5'-GGACAAACTCGTACTGCTGC 5'-TAATGCCTAGCACCCAGAAA	62	-
GCT8B05#	5'-CACCCATCACTTACCTTCCA 5'-GGTCTTGGAGCAGCCTTC	58	-
GCT8B04	5'-AACAACTTGTGCTTTTTCTTCA 5'-ATAATATAATAGCACCTGCCTTCG	58	-
ACT2FO3#	5'-CACTGCTCCACCTTCCTAAA 5'-ATATGACATTTTCAAACACAGAGC	58	-
GCT11B02	5'-AAACCTTCAGTATTCAGTTGGC 5'-TTAGCTGTGTGCGAGTGTTTC	57	2% DMSO
GCT8DO3	5'-AGGGTGAGACTCCATCTCAA 5'-CTTCTACTCTGTGGGTTGTCTG	57	2% DMSO
CTG-B37# (DRPLA)	5'-CCTCCAGTGGGTGGGGAAATGCTC 5'-CACCAAGTCTCAACACATCACCATC	62	-

membrana em 4SSC, 7% N-laurilsarcosina à temperatura ambiente durante 24 horas, seguida de hibridação por 12 horas à temperatura ambiente com a sonda (CAG)₁₅, marcada radioactivamente por incorporação de $\gamma^{32}\text{P}$ -dATP com a enzima T4-kinase e de 3 lavagens por 15 minutos com uma solução 50mM Tris pH 7.6, 2 mM EDTA, 0.1% SDS e 3M cloreto de tetrametilamónio, a temperaturas progressivamente mais elevadas (50°C, 55°C e 60°C).

Resultados

Conforme se mostra na tabela 6 nenhum dos sete *repeats* testados, cuja localização se confirmou ser no cromossoma 14 (excepto no caso do *repeat* CAG do gene DRPLA, que se localiza no cromossoma 12), se apresentava expandido nos doentes DMJ estudados.

Dada a ausência de resultados positivos com os *repeats* do cromossoma 14 testados desta forma, prosseguiu-se para a identificação de novos *repeats* localizados neste cromossoma, em particular *repeats* CAG/CTG. A hibridação do banco de cosmídeos do cromossoma 14 com a sonda (CAG)₁₅ originou numerosos positivos, tendo-se seleccionado 46 (com sinal mais forte e presente nas duas cópias do banco de cosmídeo testadas), de que se iniciou a caracterização a fim de confirmar quais continham de facto *repeats* CAG e quais se localizavam na região de interesse, determinada por estudos genéticos. No entanto, a confirmação destes cosmídeos como positivos e posterior caracterização não foi concretizada devido à publicação, entretanto, da clonagem do gene MJD1 pelo grupo japonês de Kakizuka e colaboradores [Kawaguchi *et al.*, 1994], que nos levou a modificar a estratégia de estudo.

Tabela 6. Características dos *repeats* de trinucleotídeos testados (designação, unidade repetida e tamanho do fragmento obtido por amplificação por PCR), resultados obtidos para os doentes DMJ testados e confirmação, em caso disso, da localização no cromossoma 14, utilizando um híbrido monocromossómico. Os *repeats* que eram polimórficos na população testada estão marcados com o símbolo #.

Designação	Unidade repetida	Tamanho do fragmento (bp)	Expansão em doentesDMJ	Hib14
GCT2C07	TAC/ CTG×CTA	411	ausente	+
GCT8B05#	GCT	148	ausente	+
GCT8B04	GCT	201	ausente	+
ACT2FO3#	AGT	137	ausente	+
GCT11B02	CAG	151	ausente	+
GCT8D03	CAG	91	ausente	+
CTG-B37# (DRPLA)	CAG	116-161	ausente	-

Discussão

A estratégia adoptada para a identificação de potenciais genes candidatos para a DMJ baseou-se em características clínicas e genéticas partilhadas por esta doença e por outras doenças degenerativas do sistema nervoso de início na idade adulta. A escolha dos doentes a testar para possíveis expansões dos *repeats* de trinucleotídeos fez-se entre indivíduos de várias famílias, de origens geográficas diversas, prevendo-se a possível existência de heterogeneidade genética entre as famílias, o que de facto se veio a verificar posteriormente (uma das famílias testadas, a P33, apresentava, como se referiu atrás, a mutação SCA2). Escolheram-se os doentes de início mais precoce, porque se esperavam que a estes correspondessem a diferenças de tamanho entre o alelo normal e o alelo expandido mais significativas, de acordo com as observações anteriores em doenças associadas à expansão de trinucleotídeos. Não se identificaram, no entanto, expansões nos sete *repeats* do cromossoma 14 testados. Embora a possibilidade de os *repeats* expandidos não serem passíveis de amplificação por PCR não pudesse nunca ser excluída completamente, recorreu-se a vários dos protocolos normalmente utilizados na amplificação de outros *repeats* expandidos, e utilizou-se a hibridação com sondas marcadas radioactivamente, para se obter uma maior sensibilidade a produtos de PCR menos abundantes, numa tentativa de excluir esta possibilidade.

É de interesse notar que, dos *repeats* testados, só dois (além do *repeat* CAG no gene DRPLA) apresentavam polimorfismo na população controlo, ao contrário do que se observa em geral nos *repeats* associados a doenças identificados até à data, cuja elevada taxa de mutação leva à existência de uma grande variedade de alelos na população normal.

Apesar dos *repeats* testados não se terem revelado de interesse para a DMJ, e uma vez que estão já desenvolvidas as condições experimentais e que está confirmada a sua localização cromossómica, seria interessante testá-los em outras doenças neurológicas ligadas ao cromossoma 14, mesmo na ausência de características clínicas ou genéticas evidentes que sugiram este tipo de mutação. De facto, realizou-se já um estudo semelhante ao descrito acima, em que se testaram os mesmos sete *repeats* em doentes

com distrofia muscular oculo-faríngea, cujo gene se localiza no cromossoma 14p [Brais *et al.*, 1995], e em doentes com distonia de torsão, ligada a marcadores do cromossoma 14q [Nygaard *et al.*, 1993] e também nestes doentes não se encontraram expansões dos *repeats* testados [Maciel, resultados não-publicados]. É no entanto possível que algum destes *repeats* seja relevante na patologia de outras doenças neurológicas, ou mesmo de doenças psiquiátricas, como a esquizofrenia ou a psicose maníaco-depressiva, o que está presentemente a ser estudado [Joobher, comunicação pessoal]. Pela mesma razão, seria interessante completar a caracterização dos outros *repeats* do cromossoma 14 identificados, alguns dos quais poderão ser utilizados como marcadores genéticos ou mesmo constituir genes candidatos para outras doenças ligadas a este cromossoma. No caso das doenças psiquiátricas, que se prevêm poligénicas, com múltiplos *loci* a contribuírem de forma variável para o fenótipo, é possível que em alguns dos genes envolvidos existam expansões de trinucleotídeos. Apontam neste sentido estudos que demonstram a existência de antecipação em famílias com esquizofrenia [Bassett e Honer, 1994], assim como resultados recentes utilizando uma técnica directa de detecção de *repeats* expandidos na ausência de conhecimento sobre a localização cromossómica ou sequência flanqueante destes [Schalling *et al.*, 1993], em que se demonstrou a existência de expansões em doentes com esquizofrenia e em doentes com doença afectiva bipolar e unipolar [Lindblad *et al.*, 1995; O'Donovan *et al.*, 1995; Oruc *et al.*, 1997].

A identificação de uma expansão do *repeat* CAG no gene MJD1 em doentes DMJ [Kawaguchi *et al.*, 1994] veio pois confirmar as previsões teóricas, baseadas na observação das características clínicas e genéticas da doença e na comparação da DMJ com outras doenças causadas por este tipo de mutação. Além da sequência repetitiva CAG, o ADN_C do gene MJD1 não apresentava no entanto qualquer homologia com os genes causadores destas doenças nem com outros genes conhecidos que permitissem prever a sua função na célula ou o mecanismo patológico envolvido na DMJ. Na secção seguinte descreve-se a clonagem do ADN_C e do ADN genómico do gene MJD1 e os resultados preliminares da caracterização e elucidação da estrutura genómica deste gene.

Clonagem e elucidação da estrutura genómica do gene MJD1

Introdução

O ADN_C do gene DMJ foi clonado em 1994 [Kawaguchi *et al.*, 1994], por rastreio de um banco de ADN_C de cérebro utilizando uma sonda (CTG)₁₃, complementar aos *repeats* CAG. Um dos *clones* identificados com esta estratégia, designado MJD1, continha um *repeat* CAG e foi estudado em mais detalhe, demonstrando-se a sua expressão no cérebro por RT-PCR e a sua localização cromossómica na banda 14q32.1 por análise de hibridação de fluorescência *in situ* (FISH), o que sugeriu que este ADN_C fosse um bom gene candidato para a DMJ. A análise do ADN de 12 doentes clinicamente identificados com DMJ confirmou a existência de um mesmo tipo de mutação no gene em questão - a expansão do segmento (CAG)_n - em 11 destes doentes: enquanto os indivíduos controlo testados apresentavam segmentos de 13 a 36 unidades CAG, os doentes eram portadores de 68 a 79 unidades CAG.

A análise da sequência do ADN_C MJD1 revelou que se tratava de um gene pequeno (ADN_C de 1.2kb), prevendo-se a existência de quatro exões, e que o *repeat* CAG se localizava na região codificante, sendo potencialmente traduzido como um segmento de poliglutamina.

A sequência prevista para a proteína codificada por este ADN_C não apresentava homologia com nenhuma outra proteína conhecida, com excepção da componente poliglutamina, nem motivos estruturais reconhecíveis, não se podendo portanto inferir sobre a sua função bioquímica. De acordo com a sequência proteica prevista com base no ADN_C clonado, a proteína teria um perfil maioritariamente hidrofílico e não apresentaria nem sequências-sinal identificáveis, nem domínios transmembranares, sugerindo uma localização intracelular, possivelmente citoplasmática.

Por hibridação de um banco de ADN genómico com uma sonda derivada do ADN_C do gene MJD1, ao qual se retirou a sequência (CAG)_n e a sequência Alu na região 3' [Kawaguchi *et al.*, 1994], foram identificados três genes homólogos daquele,

designados MJD2, MJD3 e MJD4, que foram localizados por análise de FISH nos cromossomas 8q23, 14q21 e Xp22.1. Os genes MJD3 e MJD4 não continham o segmento em que se encontra o *repeat* CAG, enquanto o gene MJD2 continha uma sequência repetitiva CAGCGG na região correspondente.

Dado que se desconhecia ainda a estrutura genómica do gene MJD1, assim como a estrutura de possíveis formas de *splicing* alternativo deste gene, levámos a cabo uma estratégia de estudo envolvendo a clonagem do ADN_C e, posteriormente, do ADN genómico correspondente, que se procurou caracterizar. Embora este estudo não esteja ainda completo, apresentam-se aqui os resultados preliminares.

Material e métodos

Extraíu-se ARN total, utilizando o método de extracção por cloreto de céσιο [Sambrook *et al.*, 1989], a partir de linhas celulares de dois indivíduos portadores da mutação DMJ (indivíduo 15-2, com um alelo normal de 24 *repeats* CAG e um alelo expandido de 80, e indivíduo 37-2, com um alelo normal de 24 e um alelo expandido de 66).

O ARN assim obtido foi utilizado na amplificação total ou parcial do transcripto do gene MJD1 por RT-PCR (GeneAmp® RNA PCR kit, Perkin Elmer Cetus) utilizando-se para esse fim as combinações de *primers* 2050-2049 (totalidade do ADN_C), 2050-MJDseq1 (exão 1), 2050-MJDseq2 (exões 1 e 2) e 2050-MJDseq3 (exões 1 a 3, excluindo-se assim o *repeat*). Após purificação do produto de RT-PCR em gel de agarose e extracção do ADN, clonaram-se estes produtos directamente num vector, o plasmídeo pCR™ 2.1 (TA Cloning® kit Invitrogen), e transformaram-se bactérias competentes [Sambrook *et al.*, 1989], da estirpe INVαF, com este ADN, sendo os transformantes confirmados como contendo o ADN_C do gene MJD1 por sequenciação (*Sequenase version 2.0 DNA sequencing kit, United States Biochemical*). Em paralelo, procedeu-se à clonagem de ADN_Cs do gene MJD1, por rastreio de um banco de ADN_C de cérebro

humano (*Clontech*) e obteve-se um clone contendo uma EST - *expressed sequence tag* - (projecto IMAGE-Merck), homólogo do gene MJD1.

Parte do produto de RT-PCR obtido com os *primers* 2050-MJDseq3 a partir do ARN do indivíduo 15-2 foi marcado radioactivamente com α ^{32}P -dATP, por *random priming* [Sambrook *et al.*, 1989], e utilizado directamente para hibridação do banco de cosmídeos com ADN genómico do cromossoma 14. O ADN dos cosmídeos identificados como positivos para esta sonda foi extraído por *miniprep* [Sambrook *et al.*, 1989] e digerido com a enzima *Eco* RI, sendo os fragmentos resultantes analisados por electroforese em gel de agarose a 1% e transferidos para uma membrana *Hybond+*, que foi em seguida re-hibridada sequencialmente com sondas obtidas por RT-PCR (com os *primers* 2050-MJDseq1 e 2050-MJDseq3) e com oligonucleotídeos específicos de cada exão. Uma segunda hibridação do banco de cosmídeos foi realizada após a identificação de um exão adicional, não descrito por Kawaguchi e colaboradores (figura 2b), com uma sonda de ADN_C contendo esse novo exão, a fim de identificar os clones que continham esta porção do gene MJD1.

Os cosmídeos foram utilizados para localização cromossómica por *FISH* e por *fiber-FISH*, uma variante desta técnica em que se estiram os cromossomas, obtendo-se uma maior resolução em termos da localização cromossómica, segundo métodos descritos [Wiegant *et al.*, 1992; Florijn *et al.*, 1995].

Resultados

Por RT-PCR só se conseguiu amplificar o alelo normal do gene MJD1, pelo que só se clonou o alelo normal, do indivíduo 15-2. Os clones de ADN_C (14C6, 14C7 e 14C9) assim obtidos eram idênticos em sequência ao clonado pelo grupo japonês [Kawaguchi *et al.*, 1994], com excepção das substituições pontuais nos nucleotídeos C⁹⁸⁷GG/G⁹⁸⁷GG, TAA¹¹¹⁸/TAC¹¹¹⁸ e A⁶⁶⁹TG/G⁶⁶⁹TG, que se considerou como possíveis erros introduzidos durante a PCR, mas que posteriormente se veio a demonstrar serem polimorfismos correntes na população normal.

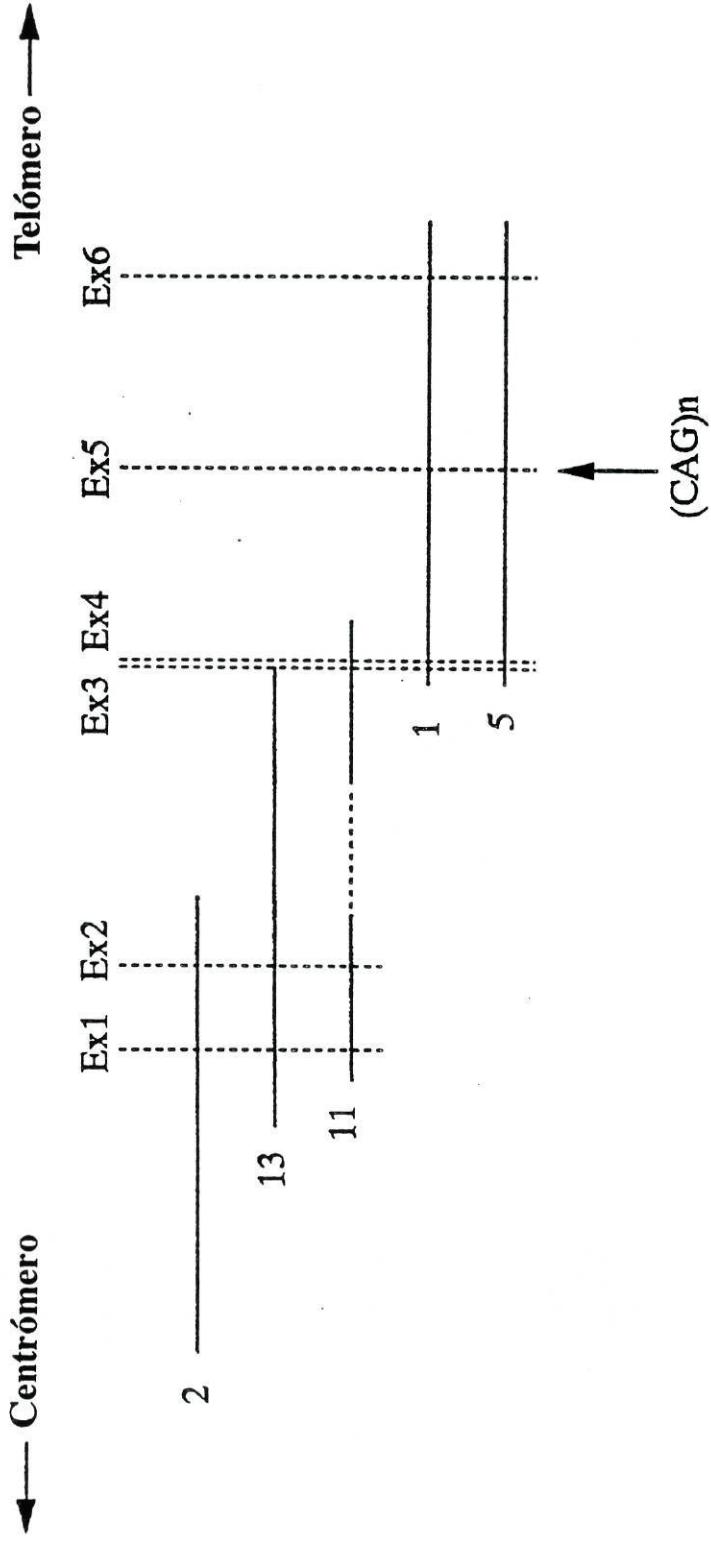
Obtiveram-se ainda três *clones* de ADN_C a partir de dois bancos de ADN_C de cérebro: um contendo apenas um fragmento do ADN_C (*clone* 1211) com aproximadamente 800 bp, que estendia na extremidade 3' 100 bp mais além da sequência publicada por Kawaguchi e colaboradores [1994], mas não continha a extremidade 5' e dois outros ADN_C completos dos quais se apresenta a sequência na figura 2. Um dos ADN_C era similar ao publicado por Kawaguchi e colaboradores (figura 2.a) e outro era diferente a partir do nucleotídeo 1049 (figura 2.b).

Identificaram-se 12 cosmídeos positivos para a sonda de ADN_C contendo os exões 1 a 3. A carta de restrição destes cosmídeos com a enzima EcoRI permitiu identificar algumas sobreposições entre os diferentes cosmídeos, embora sugerisse que o gene não estava contido na sua totalidade em nenhum dos cosmídeos, mas sim parcialmente distribuído por pelo menos três dos cosmídeos em questão. Esta informação foi complementada (a) com os resultados da hibridação dos *Southern blots* com sondas obtidas por RT-PCR (contendo o exões 1 e os exões 1 a 3, de acordo com o publicado por Kawaguchi e colaboradores) e com oligonucleotídeos localizados nos diferentes exões, e (b) com os resultados da amplificação por PCR levando a um mapa tentativo da região contida nos cosmídeos. Utilizando a sonda de ADN_C contendo o novo exão identificaram-se dois *clones* positivos, um já identificado previamente (cosmídeo #1) e outro novo (cosmídeo #13), que foi adicionado ao estudo. O *contig* formado pelos cosmídeos que se confinam o gene MJD1 (cosmídeos 1, 2, 3, 5, 10, 11, 13) está representado na figura 3. A análise de FISH localiza estes cinco *clones* na região descrita por Kawaguchi e colaboradores para o gene MJD1. Os restantes cosmídeos (4, 6, 8, 9, 12) parecem conter um gene homólogo ao gene MJD1, que se localiza na proximidade da banda 14q21, e que poderia corresponder ao gene MJD2 [Kawaguchi *et al.*, 1994]. O facto de nestes cosmídeos todas as sondas utilizadas para localizar os diferentes exões hibridarem com um mesmo fragmento de restrição EcoRI sugere que se trata de um gene sem intrões (pseudogene?), possivelmente resultante da retrotranscrição de um ARNm já processado pela maquinaria celular de *splicing*.

GTT GGC TCC AGA CAA ATA AAC ATG GAG TCC ATC TTC CAC GAG AAA CAA GAA GGC TCA CTT	60
Met Glu Ser Ile Phe His Glu Lys Gln Glu Gly Ser Leu	13
TGT GCT CAA CAT TGC CTG AAT AAC TTA TTG CAA GGA GAA TAT TTT AGC CCT GTG GAA TTA	120
Cys Ala Gln His Cys Leu Asn Asn Leu Leu Gln Gly Glu Tyr Phe Ser Pro Val Glu Leu	33
TCC TCA ATT GCA CAT CAG CTG GAT GAG GAG GAG AGG ATG AGA ATG GCA GAA GGA GGA GTT	180
Ser Ser Ile Ala His Gln Leu Asp Glu Glu Glu Arg Met Arg Met Ala Glu Gly Gly Val	58
ACT AGT GAA GAT TAT CGC ACG TTT TTA CAG CAG CCT TCT GGA AAT ATG GAT GAC AGT GGT	240
Thr Ser Glu Asp Tyr Arg Thr Phe Leu Gln Gln Pro Ser Gly Asn Met Asp Asp Ser Gly	78
TTT TTC TCT ATT CAG GTT ATA AGC AAT GCC TTG AAA GTT TGG GGT TTA GAA CTA ATC CTG	300
Phe Phe Ser Ile Gln Val Ile Ser Asn Ala Leu Lys Val Trp Gly Leu Glu Leu Ile Leu	98
TTC AAC AGT CCA GAG TAT CAG AGG CTC AGG ATC GAT CCT ATA AAT GAA AGA TCA TTT ATA	360
Phe Asn Ser Pro Glu Tyr Gln Arg Leu Arg Ile Asp Pro Ile Asn Glu Arg Ser Phe Ile	118
TGC AAT TAT AAG GAA CAC TGG TTT ACA GTT AGA AAA TTA GGA AAA CAG TGG TTT AAC TTG	420
Cys Asn Tyr Lys Glu His Trp Phe Thr Val Arg Lys Leu Gly Lys Gln Trp Phe Asn Leu	138
AAT TCT CTC TTC ACG GGT CCA GAA TTA ATA TCA GAT ACA TAT CTT GCA CTT TTC TTG GCT	480
Asn Ser Leu Leu Thr Gly Pro Glu Leu Ile Ser Asp Thr Tyr Leu Ala Leu Phe Leu Ala	158
CAA TTA CAA CAG GAA GGT TAT TCT ATA TTT GTC GTT AAG GGT GAT CTG CCA GAT TGC GAA	540
Gln Leu Gln Gln Glu Gly Tyr Ser Ile Phe Val Val Lys Gly Asp Leu Pro Asp Cys Glu	178
GCT GAC CAA CTC CTG CAG ATG ATT AGG GTC CAA CAG ATG CAT CGA CCA AAA CTT ATT GGA	600
Ala Asp Gln Leu Leu Gln Met Ile Arg Val Gln Gln Met His Arg Pro Lys Leu Ile Gly	198
GAA GAA TTA GCA CAA CTA AAA GAG CAA AGA GTC CAT AAA ACA GAC CTG GAA CGA GTG TTA	660
Glu Glu Leu Ala Gln Leu Lys Glu Gln Arg Val His Lys Thr Asp Leu Glu Arg Val Leu	218
GAA GCA AAT GAT GGC TCA GGA ATG TTA GAC GAA GAT GAG GAG GAT TTG CAG AGG GCT CTG	720
Glu Ala Asn Asp Gly Ser Gly Met Leu Asp Glu Asp Glu Glu Asp Leu Gln Arg Ala Leu	238
GCA CTA AGT CGC CAA GAA ATT GAC ATG GAA GAT GAG GAA GCA GAT CTC CGC AGG ACT ATT	780
Ala Leu Ser Arg Gln Glu Ile Asp Met Glu Asp Glu Glu Ala Asp Leu Arg Arg Thr Ile	258
CAG CTA AGT ATG CAA GGT AGT TCC AGA AAC ATA TCT CAA GAT ATG ACA CAG ACA TCA GGT	840
Gln Leu Ser Met Gln Gly Ser Ser Arg Asn Ile Ser Gln Asp Met Thr Gln Thr Ser Gly	278
ACA AAT CTT ACT TCA GAA GAG CTT CGG AAG AGA CGA GAA GCC TAC TTT GAA AAA CAG CAG	900
Thr Asn Leu Thr Ser Glu Glu Leu Arg Lys Arg Arg Glu Ala Tyr Phe Glu Lys Gln Gln	298
CAA AAG CAG CAA CAG CAG CAG CAG CAG CAG GGG GAC CTA TCA GGA CAG AGT TCA	960
Gln Lys Gln Gln Gln Gln Gln Gln Gln Gln Gln Gly Asp Leu Ser Gly Gln Ser Ser	318
CAT CCA TGT GAA AGG CCA GCC ACC AGT TCA GGA GCA CTT GGG AGT GAT CTA GGT AAG GCC	1020
His Pro Cys Glu Arg Pro Ala Thr Ser Ser Gly Ala Leu Gly Ser Asp Leu Gly Lys Ala	338
TGC TCA CCA TTC ATC ATG TTC GCT ACC TTC ACA CTT TAT CTG ACA TAC GAG CTC CAT GTG	1080
Cys Ser Pro Phe Ile Met Phe Ala Thr Phe Thr Leu Tyr Leu Thr Tyr Glu Leu His Val	358
ATT TTT GCT TTA CAT TAT TCT TCA TTC CCT CTT TAA TCA TAT TAA GAC TCT TAA GTA AAT	1140
Ile Phe Ala Leu His Tyr Ser Ser Phe Pro Leu ***	369
TTG TAA TCT ACT AAA TTT CCC TGG ATT AAG GAG CAG TTA CCA AAA AAA AAA AAA AAA	1200
AAA	

Fig. 2A

Mapa de cosmídeos da região do gene *MJD1*



Escala arbitrária

Discussão

Os resultados obtidos neste estudo da estrutura genômica do gene MJD1, embora ainda preliminares, são bastante interessantes e relevantes para outros aspectos do estudo da DMJ. A primeira observação de interesse é que o gene, apesar da pequena dimensão dos ADN_C clonados, se apresenta distribuído por uma região de pelo menos 100kb.

A identificação de novos exões e de variantes de *splicing* abre também novas perspectivas, já que é possível que existam formas específicas de determinados tecidos e que a relevância dessas diferentes formas para a patogénese da doença seja diferente. As análises de *Northern blot* não permitiram esclarecer esta questão (Goto, comunicação pessoal), pois não se identificam os transcritos correspondentes aos ADN_C clonados até à data, mas sim transcritos de maior tamanho que o previsto (cerca de 8kb), o que sugere que a totalidade dos exões deste gene pode não estar ainda identificada. No entanto, em *Western blots* de cérebro e de diversos outros tecidos humanos [Paulson *et al.*, 1997a], realizados utilizando anticorpos desenvolvidos contra a proteína MJD1 expressa em bactérias (com a sequência da variante de ADN_C originalmente descrita), o principal produto observado parece corresponder aproximadamente ao tamanho previsto para esta proteína, tendo em conta diferenças na velocidade de migração normalmente associadas a proteínas com poliglutaminas [Trottier *et al.*, 1997]. Observam-se múltiplas bandas secundárias, de maior e menor tamanho, que parecem relevantes noutros tecidos mas de fraca intensidade em todas as regiões do cérebro estudadas [Paulson *et al.*, 1997a].

A identificação de variantes de sequência em vários pontos da sequência do ADN_C - polimorfismos intragénicos - permitiu a sua aplicação na determinação de haplótipos, com relevância para várias questões, desde o estudo da origem da mutação (Gaspar, manuscrito em preparação), a estudos de factores envolvidos na instabilidade intergeracional do *repeat* CAG no gene MJD1 (ver Capítulo 4). Seria importante caracterizar todos os polimorfismos intragénicos por sequenciação. Por um lado, estes podem estar associados a diferenças fenotípicas, em termos de idade de início da doença e gravidade do quadro clínico. Por outro lado, existe o precedente da doença de

Huntington, em que foram descritos casos de falsos negativos nos testes moleculares devido a polimorfismos (raros) na região de emparelhamento dos *primers* usados na amplificação do *repeat* por PCR [Cross *et al.*, 1994]; seria útil excluir este tipo de dificuldade no diagnóstico molecular da DMJ.

No que respeita aos genes homólogos do gene MJD1, a caracterização detalhada destes genes será muito importante, por diversas razões: uma delas é a possível existência de regiões de elevada homologia entre os vários genes, cuja identificação, juntamente com o estudo da sua expressão e localização intracelular e tecidual, permita uma melhor compreensão da função do gene MJD1. A identificação destas regiões de homologia pode também ter relevância prática mais imediata: se ocorrerem nas sequências a que se emparelham os *primers* utilizados na amplificação por PCR, quer do *repeat* ou de outros segmentos do gene MJD1 (nomeadamente os que contenham polimorfismos intragénicos a testar), podem levar a resultados erróneos.

Por outro lado, seria sem dúvida interessante estudar a possível existência de mutações nestes genes em outras doenças neurológicas (tais como as formas de ataxia para as quais não se conhece ainda o gene causador) ou outras doenças ligadas às regiões em que se situam estes genes. Em particular, o gene MJD2, que contém uma sequência repetitiva, poderia ser um bom gene candidato a testar em doenças neurológicas.

CAPÍTULO 2

Correlações genótipo-fenótipo: quadro clínico e antecipação

Correlações genótipo-fenótipo: quadro clínico e antecipação

A clonagem do ADN_C do gene MJD1 [Kawaguchi *et al.*, 1994] e a demonstração da existência de uma expansão de uma sequência (CAG)_n neste gene, em 11 de 12 doentes japoneses, vieram levantar várias questões, nomeadamente, se a expansão da sequência (CAG)_n seria a causa da DMJ em todas as famílias descritas com esta doença e a necessidade de definir, numa população alargada, os tamanhos dos alelos associados à doença e dos alelos normais.

Nos trabalhos aqui apresentados, demonstrou-se que esta mesma mutação estava presente em 156 doentes, pertencentes a 33 famílias DMJ (15 famílias açorianas, 2 brasileiras e 16 norte-americanas de origem açoriana): enquanto os cromossomas normais continham entre 12 e 37 unidades CAG no gene MJD1, os cromossomas associados à doença continham entre 62 e 84 (artigo I).

Procurou-se ainda uma base molecular para a grande diversidade clínica característica da DMJ, quer em termos de idade de início, quer em termos dos principais sintomas apresentados, que levou à classificação dos quadros clínicos em 3 subfenótipos (artigo I). Observou-se uma correlação inversa muito significativa entre o tamanho do *repeat* CAG expandido e a idade de início da doença, assim como uma diferença significativa entre os subfenótipos em termos de tamanhos médios do *repeat* expandido, demonstrando que grande parte desta heterogeneidade de apresentação clínica é atribuível à diversidade de tamanhos do *repeat* CAG expandido. No entanto, esta correlação era incompleta, o que sugeria que outros factores, genéticos ou ambientais, afectavam também a idade de início e apresentação clínica da doença.

De acordo com as nossas observações, o tamanho do *repeat* CAG no cromossoma normal não apresentava correlação com a idade de início da DMJ (artigo I) e o efeito do sexo do indivíduo, que, de acordo com o sugerido por Kawakami e colaboradores seria um factor com influência na idade de início, com efeito aditivo ao do

tamanho do *repeat* CAG [Kawakami *et al.*, 1995], não se demonstrou relevante na nossa população de doentes (artigo II).

Assim, foi-se investigar a existência de outros factores, além do tamanho do *repeat* CAG expandido, com efeito na idade de início da DMJ (artigo III), usando métodos estatísticos especificamente desenvolvidos para este fim, que permitiram controlar para o efeito deste factor. Identificou-se a existência de um efeito familiar, independente do tamanho do (CAG)_n, originando uma maior semelhança de idades de início entre indivíduos de uma mesma fratria do que entre indivíduos com grau de parentesco mais afastado ou não relacionados. Não foi, no entanto, ainda possível distinguir se se tratava de um efeito ambiental ou de um efeito genético, sendo necessários estudos adicionais, com um número de indivíduos substancialmente maior, para se poder esclarecer esta questão.

Fez-se também um estudo inicial (estudos adicionais descritos no capítulo 4, artigo VI) da instabilidade intergeracional do *repeat* CAG expandido no gene MJD1, e dos factores que poderiam de algum modo afectar esta instabilidade, em particular o sexo do progenitor e o tamanho inicial do *repeat* expandido, tendo-se observado uma maior instabilidade intergeracional associada às transmissões paternas (expansões e contracções de maior amplitude), mas uma ausência de correlação entre o tamanho inicial do *repeat* CAG no progenitor e a sua instabilidade durante a transmissão. Por outro lado, não se detectou uma correlação linear significativa entre a variação de tamanho do (CAG)_n expandido no gene MJD1 e a antecipação observada entre pares pais-filhos afectados, o que pode ser devido a enviesamentos de amostragem.

ARTIGO I

Correlation between CAG Repeat Length and Clinical Features in Machado-Joseph Disease

Patrícia Maciel,^{1,2} Claudia Gaspar,^{1,2} Anita L. DeStefano,³ Isabel Silveira,^{1,2} Paula Coutinho,⁵ João Radvany,⁶ David M. Dawson,^{4,7} Lewis Sudarsky,^{4,7} João Guimarães,⁸ Jose E. L. Loureiro,⁹ Marjan M. Nezarati,¹⁰ Lee I. Corwin,¹¹ Iscia Lopes-Cendes,¹ Karen Rooke,¹ Roger Rosenberg,¹² Patrick MacLeod,¹³ Lindsay A. Farrer,³ Jorge Sequeiros,² and Guy A. Rouleau¹

¹Centre for Research in Neuroscience, McGill University and the Montreal General Hospital Research Institute, Montreal; ²Lab. Genética Médica, ICBAS and UniGENe, IBMC Universidade do Porto, Portugal; ³Department of Neurology, Boston University School of Medicine, and ⁴Neurology Division, Brigham and Women's Hospital, Boston; ⁵Serviço de Neurologia, Hospital Santo, António, Porto, Portugal; ⁶Neurologia, Hospital Israelita Albert Einstein, São Paulo; ⁷Neurology Service, Veterans Administration Medical Center, West Roxbury, Massachusetts; ⁸Hospital Egas Moniz, Lisboa; ⁹Hospital de Vila Real, Vila Real, Portugal; ¹⁰Faculty of Health Sciences, McMaster University, Hamilton, Ontario; ¹¹Southeastern Neurological Associates, Plymouth, Massachusetts; ¹²Department of Neurology, University of Texas South Western Medical Center, Dallas; and ¹³Section of Genetics, Department of Laboratory Medicine, Victoria General Hospital, Victoria, British Columbia

Summary

Machado-Joseph disease (MJD) is associated with the expansion of a CAG trinucleotide repeat in a novel gene on 14q32.1. We confirmed the presence of this expansion in 156 MJD patients from 33 families of different geographic origins: 15 Portuguese Azorean, 2 Brazilian, and 16 North American of Portuguese Azorean descent. Normal chromosomes contain between 12 and 37 CAG repeats in the MJD gene, whereas MJD gene carriers have alleles within the expanded range of 62–84 CAG units. The distribution of expanded alleles and the gap between normal and expanded allele sizes is either inconsistent with a premutation hypothesis or most (if not all) of the alleles we studied descend from a common ancestor. There is a strong correlation between the expanded repeat size and the age at onset of the disease as well as the clinical presentation. There is mild instability of the CAG tract length with transmission of the expanded alleles; both increase and decrease in size between parents and progeny occur, with larger variations in male than in female transmissions. Together, these effects can partly explain the variability of age at onset and of phenotypic features in MJD; however, other modifying factors must exist.

Introduction

Machado-Joseph disease (MJD) is an autosomal dominant spinocerebellar degeneration characterized by a

wide range of clinical manifestations, including ataxia, progressive external ophthalmoplegia, pyramidal and extra pyramidal signs, dystonia with rigidity, and distal muscular atrophies (Coutinho et al. 1977; Lima and Coutinho 1980). Typically, patients will become confined to a wheelchair and will later be bedridden; the median survival time after onset is 20 years. Disease manifestations usually start during adulthood, with a mean age at onset of 37.4 years (SD 14.1), but the distribution of age at onset is very wide, ranging from 5 to 73 years (Sequeiros and Coutinho 1993).

Pathologically, the disease is characterized by degeneration of the spinocerebellar, dentate, pontine, and vestibular nuclei as well as extra pyramidal structures, such as the substantia nigra, locus coeruleus, and the pallidolusian complex. In addition, there is neuronal loss in motor cranial nerve nuclei, anterior horn cells, and posterior root ganglia. Structures such as the cerebral and cerebellar cortex and inferior olives are spared (Rosenberg et al. 1976; Coutinho and Andrade 1978; Coutinho et al. 1982; Sakai et al. 1983; Kanda et al. 1989).

Detailed clinical investigation of the variability in MJD led to its classification into three subphenotypes, occasionally present in the same family: type 1 patients, with earlier onset (mean 24.3 years) and marked pyramidal and extrapyramidal signs, in addition to the common features of cerebellar ataxia and ophthalmoplegia, as well as a rapid and more severe clinical course; type 2 patients have a mean age at onset of 40.5 years and show mainly cerebellar ataxia and ophthalmoplegia; type 3 patients show a later onset (mean 46.8 years), have marked peripheral signs, weakness, and amyotrophy, along with cerebellar ataxia and ophthalmoplegia. Type 2 may be transitional, since all patients start as a type 2; some later evolve into either type 1 or type 3, while others remain as a type 2 all their lives.

The disease locus was mapped to chromosome 14q32.1 in Japanese families (Takiyama et al. 1993),

Received February 13, 1995; accepted for publication April 6, 1995.

Address for correspondence and reprints: Dr. Guy A. Rouleau, Centre for Research in Neuroscience, McGill University, Montreal General Hospital Research Institute, 1650 Cedar Avenue, Montreal, Québec H3G 1A4, Canada.

© 1995 by The American Society of Human Genetics. All rights reserved.
0002-9297/95/5701-0007\$02.00

and linkage was confirmed in families of Portuguese Azorean origin as well as in Portuguese American and Brazilian pedigrees (Sequeiros et al. 1994; Twist et al. 1995). The gene was recently identified (Kawaguchi et al. 1994) and shown to contain a CAG repeat motif in the 5' region of the coding sequence, which is selectively expanded in MJD patients. Therefore, MJD is the most recent addition to the growing number of neurodegenerative disorders for which the causative mutation is the unstable expansion of a CAG repeat.

Correlation of repeat length with age at onset of the disease has been demonstrated in Huntington disease (HD) (Andrew et al. 1993; Duyao et al. 1993; Snell et al. 1993), dentatorubropallidolusian atrophy (DRPLA) (Koide et al. 1994), spinocerebellar ataxia type 1 (SCA1) (Jodice et al. 1994; Ranum et al. 1994), and X-linked spinal and muscular atrophy or Kennedy disease (La Spada et al. 1992). Yet, the correlation of repeat size with age at onset is not perfect, and some clinical features do not correlate with repeat length, suggesting that factors other than repeat size may influence the clinical presentation in these disorders (Andrew et al. 1993; Ranum et al. 1994).

We have studied 33 families of different geographic origins diagnosed as MJD, in order to confirm the presence of the CAG repeat expansion on 14q32.1, to compare the size range of this repeat in MJD and control populations, and to investigate the relationship between the size of the expanded allele and the clinical presentation. Furthermore, we analyzed the instability of the repeat size in its transmission from parent to progeny, in an attempt to determine the molecular basis for the previously described effect of anticipation in this disease (Sequeiros and Coutinho 1981).

Subjects and Methods

Subjects

This study was performed using DNA samples from 212 members of 33 MJD families (Coutinho and Andrade 1978; Sudarsky et al. 1992; Radvany et al. 1993) of different geographical origins: 15 Portuguese Azorean, 2 Brazilian, and 16 North American. All families had traceable Portuguese ancestry; 23 of these families previously had been used for linkage studies with 14q32.1 markers (Sequeiros et al. 1994; Twist et al. 1995). The diagnosis of MJD was determined by clinical examination by an experienced neurologist using established diagnostic criteria (Lima and Coutinho 1980). Ages at onset were based on information provided by the patient and/or a close relative. For determination of the frequency of the normal alleles, a total of 83 normal controls were used, including a group of unrelated unaffected spouses from MJD families and a group of controls from different ethnic origins. Since both groups

had similar distributions of allele sizes, the data were pooled and the controls were considered a single group.

Methods

Genomic DNA was isolated from peripheral lymphocytes by standard methods (Sambrook et al. 1989) or from lymphoblastoid cell lines established by transformation with Epstein-Barr virus (Anderson and Gusella 1984). The CAG-containing fragment of the MJD gene was amplified by PCR using primers MJD52 and MJD25a (a slight modification of MJD25, of sequence ATCCATGTGCAAAGGCCAGCC) (Kawaguchi et al. 1994). PCR was performed in a final volume of 20 μ l, containing 40 ng of genomic DNA; 10 mM Tris-HCl (pH 8.8); 1.5 mM MgCl₂; 50 mM KCl; 0.1% Triton X-100; 10% dimethylsulfoxide; 250 μ M each dCTP, dGTP, and dTTP; 25 μ M dATP; 1.5 μ Ci 35S alpha-dATP; 125 ng each primer; and 3 units of Taq polymerase (Perkin Elmer). The DNA was denatured at 94°C for 5 min; then 32 cycles at 94°C for 1 min, 60°C for 1 min, and 72°C for 1 min were performed, followed by a final extension at 72°C for 5 min.

For determination of allele sizes, PCR products were analyzed on denaturing 5% polyacrilamide gels in parallel with an M13 sequencing ladder (-40 primer) and were visualized by autoradiography. Allele sizes (*S*) were determined by comparison with the sequence and were converted to CAG unit numbers (*N*), using the formula $N = (S - 121)/3$, assuming that the variation in size of the PCR product occurred within the repetitive CAG stretch.

Linear regression techniques were used to determine the association between repeat number in the MJD or normal allele and age at onset, association between variation in repeat number during transmission and anticipation, and the association between variation in repeat number during transmission and initial size of the allele (Neter et al. 1985). Differences in repeat number between the three clinical subtypes were assessed by analysis of variance and Tukey's multiple comparison of means test (Neter et al. 1985). Chi-square and Fisher's exact tests were used to determine whether variation in repeat number among parent and offspring, expressed as raw data or as a trichotomous variable (expansion, contraction, and stable transmission), is associated with sex of transmitting parent or affection status (affected or asymptomatic) of MJD allele carriers. All analyses were carried out using procedures of the Statistical Analysis System version 6 (1989).

Results

Distribution of Repeat Length in Normal and MJD Chromosomes

Figure 1 shows the frequencies of different CAG repeat lengths in 186 normal and 156 MJD chromosomes.

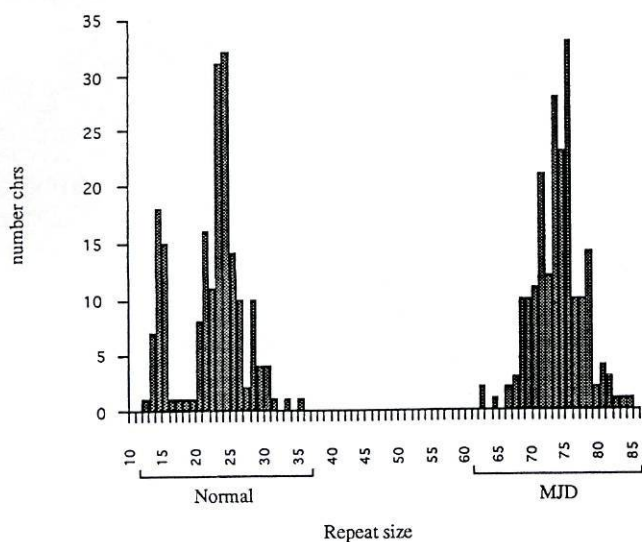


Figure 1 Distribution of CAG repeat sizes in unaffected control individuals and in MJD alleles from affected individuals.

The normal chromosomes have alleles containing from 12 to 37 repeats, whereas affected individuals have at least one expanded allele, with sizes varying between 66 and 84. Expanded alleles of 62 and 64 repeats were found in two asymptomatic carriers, of ages 38 and 45 years, respectively. Thus, there was a wide gap between the size ranges of normal and MJD alleles, and no intermediate alleles (37–62 repeats) were found in the control population. There was no significant difference between mean allele size for paternally (72.8 ± 3.4) and maternally (73.1 ± 4.3) inherited alleles ($P = .51$).

Family Studies with the MJD Repeat

An expanded allele was found in 156/165 individuals considered to be affected. Among the nine cases having two normal alleles and no expansion, five did not show the disease haplotype in previous studies with markers on 14q32.1 and haplotype data for four patients were not available. On retrospective review of clinical records, two of the nine cases were found not to be affected, and two were found to have had diagnostic difficulties (incipient signs, other diseases as possible confounding factors). In the remaining five cases, however, no explanation was found, and the results may be due to sample mix-ups, misdiagnosis, or clerical errors (Andrew et al. 1994).

Trinucleotide Repeat Length and Clinical Features

The relationship between the trinucleotide repeat length in the MJD allele of affected individuals and age at onset of the disease is shown in figure 2. A correlation coefficient of $-.68$ was obtained ($P = .0001$), assuming a linear relationship between age at onset and repeat length for the MJD chromosome. This suggests a ten-

dency for age at onset to decrease as the CAG repeat length increases. No significant correlation was found between the normal allele repeat length and age at onset ($r = .006$; $P = .95$).

Data were available on 10 cases of juvenile onset of MJD (between 10 and 20 years of age). In this group, the CAG repeat size ranged from 73 to 84, with a mean of 80.4 ± 2.5 repeats. This is significantly different from the adult onset mean of 72.6 ± 3.3 ($P = .0001$). Of these patients, 7 (70%) inherited the expanded allele from the mother and 3 (30%) from the father, but, given the small sample, the apparent excess of female transmission in juvenile onset cases when compared to the rest of the population is not significant ($P = .13$, χ^2 test).

When repeat sizes of patients of different clinical subphenotypes were compared, a significant difference was observed between subtype 1 and the other two subtypes, with a tendency for increased severity with larger repeat size ($P < .05$). Type 1 patients tend to have larger alleles (mean 76.2 ± 5.08 , $N = 11$), whereas milder forms (types 2 and 3) tend to show smaller sizes (mean 73.0 ± 3.64 , $N = 88$, and 72.2 ± 2.97 , $N = 31$, respectively). This relationship between repeat size and subtype may reflect the fact that age at onset, which is correlated with repeat size, is tightly associated with the clinical classification of subtype. Since type 2 may be a transitional form and the group of type 2 patients could include some individuals that will later become either type 1 or 3, the data were reanalyzed using only patients who have continued to have a type 2 disease presentation for ≥ 10 years. We obtained the same results as described above.

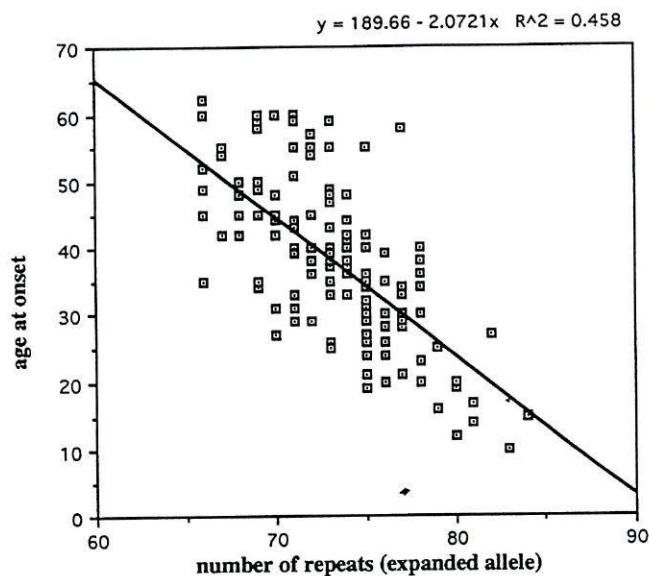


Figure 2 Correlation of CAG repeat length in the MJD chromosomes of 156 affected individuals with age at onset of disease. A squared linear correlation coefficient of $r^2 = .458$ was obtained.

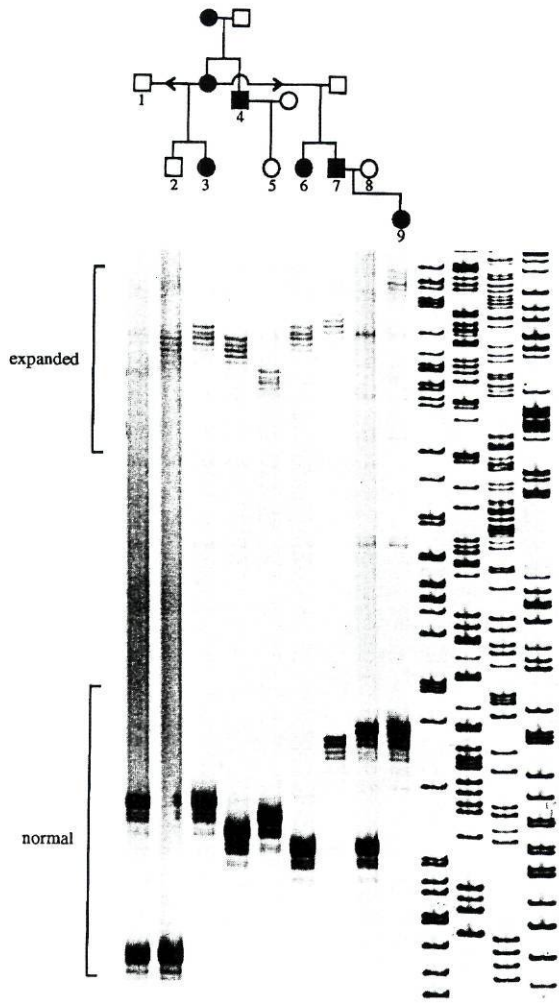


Figure 3 PCR analysis of the CAG repeat in the MJD gene in a branch of family N3. The four tracks to the right contain M13 sequence, which was used to determine the sizes of alleles.

Instability of CAG Repeat Length

Instability of the CAG repeat number during transmission from parent to offspring has been observed for HD, SCA1, spinal bulbar muscular atrophy, and DRPLA. The instability of the CAG repeat size in transmission of the MJD allele is illustrated in figure 3. Expansion, as well as contraction, of repeat tract length occurred in this family.

Within our cohort group, there were 58 individuals who had an expanded allele and for whom data on the expanded allele in the affected parent was available. Of these 58 pairs, 24 of the children were diagnosed as affected and the remaining 34 are currently at risk and presumed to be presymptomatic carriers. Figure 4A shows the distribution of variation in repeat number during transmission. In 41 (55%) of these transmissions, there was a variation in the size of the repeat, with 32 (78%) increasing in size and 9 (22%) contracting. The mean variation was +0.84 repeats, with a range from

-5 to 9. The mean amplitude of variation (absolute value) was 1.6 repeat units for each transmission. Each variation of absolute value >4 occurred in transmission from unique affected parents, from two families, B1 and N3. These two families contribute 30% and 8.6% of the pairs in this analysis, respectively. Comparing the distribution of variation in repeat size by disease status (fig. 4B), we observed that affected individuals tend to have inherited expansions, while the asymptomatic carriers tend to have inherited contractions. This difference in distribution is significant ($P = .0002$, Fisher's exact test).

The distribution of instability of repeat size by sex of

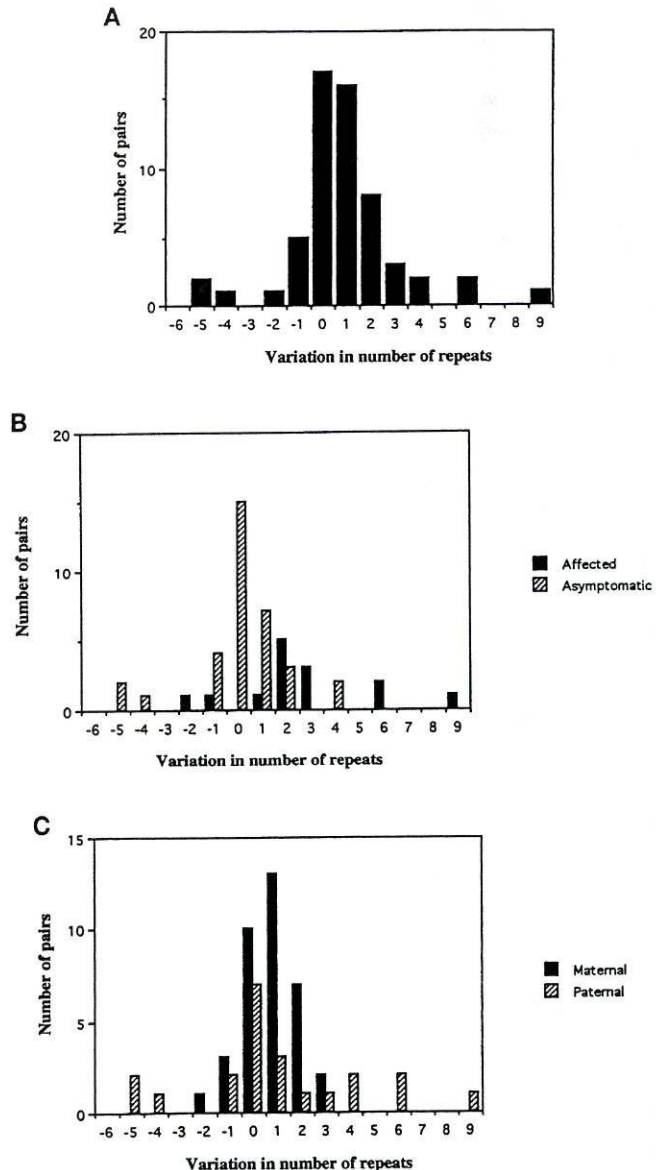


Figure 4 Distribution of the variation in repeat number during transmission of the MJD allele from parent to progeny for 58 parent/child pairs (A) for all transmissions, (B) for affected and asymptomatic carrier offspring, and (C) by sex of parent.

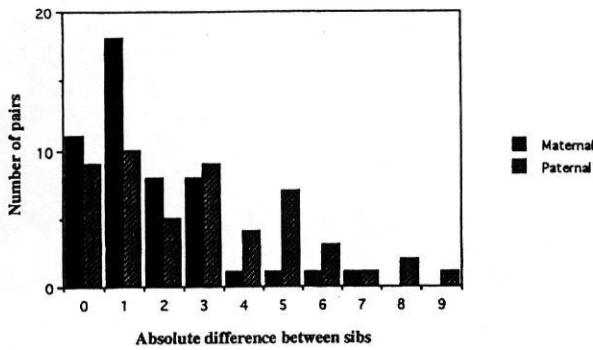


Figure 5 Distribution of absolute difference in size of repeat between sib pairs by sex of transmitting parent.

transmitting parent is shown in figure 4C. The observations at either tail of the distribution are due to alleles inherited from different affected fathers, suggesting that paternally inherited alleles are more volatile than maternal alleles. Although, when the instability data was classified as contractions, stable inheritance, or expansions we observed no significant difference in frequency of each classification for male or female meioses ($P = .393$), the mean amplitude of variation was significantly different between male (2.4 ± 2.6) and female transmissions (1.0 ± 0.86) ($P = .026$). Interestingly, the four juvenile cases for which we typed parent and child corresponded to expansions of the CAG repeat from the transmitting parent, with increases of 9, 6, 5, and 2 repeats.

The number of repeats in the transmitting parent does not correlate with the absolute value of the change in repeat during transmission ($r = -.0013$; $P = .99$). In other words, there is no effect of the initial size of allele on its stability.

Sib pairs provide another opportunity to examine instability of the CAG repeat. In our cohort there were 53 sibships containing two or more individuals carrying the expanded allele (affected and at risk). The majority of sibships consisted of a pair of individuals; however, there was 1 sibship of five, 3 sibships of four, and 13 sibships consisting of three persons with the MJD allele. Figure 5 shows the distribution of the absolute difference between sibs by sex of affected parent. There is a significant difference in mean absolute difference for paternal (2.9 ± 2.4 , $N = 51$) and maternal (1.6 ± 1.6 , $N = 49$) transmissions ($P = .002$). This supports the evidence presented earlier indicating paternal transmissions may be less stable than maternal transmissions. Among 52 sib pairs in which both sibs were affected, there was a marginal correlation between difference in age at onset between sib pairs and difference in repeat number ($r = .027$; $P = .07$). For nine sib pairs with identical repeat number, difference in age at onset ranged from 0 to 12 years. In contrast, a sib pair with a difference in repeat number of 9 had only a 6-year difference in age at onset.

Anticipation data were available on 22 of the affected parent-child pairs for which molecular data were also available. Anticipation is shown plotted against variation in repeat size, in figure 6. There was no significant correlation between change in repeat number between parent and offspring and anticipation ($r = .10$; $P = .66$).

Discussion

The CAG repeat within the coding region of the MJD gene was found to be highly polymorphic, both in the normal size range (12–37 repeat units) and in the expanded MJD size range (62–84 repeat units). There is a bimodal distribution of the normal alleles, with peaks at 14 and 24 repeats. Frequencies of normal alleles were not significantly different in the control populations we studied—spouses of Azorean, Brazilian, and Portuguese American patients and other controls of multiple ethnic origins (data not shown).

No overlap was found between the ranges of normal and MJD repeat size. Unlike other disorders caused by an expansion of a CAG repeat, there was a wide gap between the normal and MJD allele sizes, and no intermediate alleles were found in our control population. This may be just a matter of sample size, and the study of additional families may later change these figures. However, if this observation is confirmed in a sufficiently large control population, it will greatly facilitate presymptomatic testing for MJD, since the existence of overlap between the normal and disease size ranges was one of the difficulties encountered in HD (Andrew et al. 1993; Snell et al. 1993) and SCA1 (Goldfarb et al. 1994), although in SCA1 the distinction between normal (interrupted CAG repeat) and the disease (pure CAG repeat) intermediate size alleles can be made using a

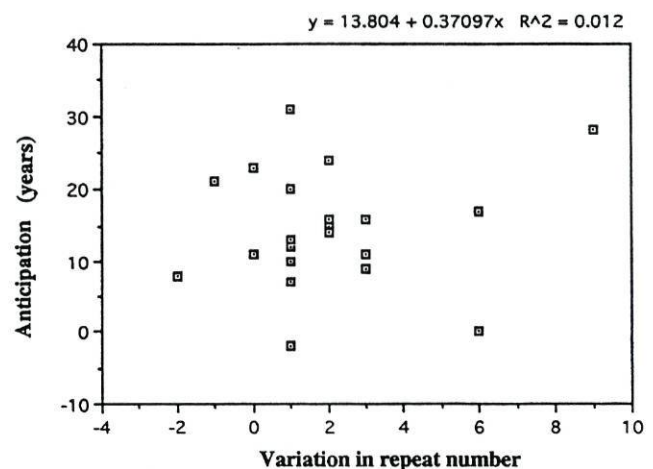


Figure 6 Correlation between variation in size of repeat and anticipation for 23 affected parent-child pairs. A correlation coefficient of $r = .045$ ($P = .83$) was found.

restriction polymorphism (Chung et al. 1993). The clustering of expanded repeat sizes is also suggestive of a unique ancient founder mutation rather than multiple expansions from the intermediate "premutation" size range (Barceló et al. 1993; Goldberg et al. 1993; Rubisztein et al. 1994). This question can be addressed by detailed haplotype analysis. The mean variation in size of +0.84 repeats on transmission with each generation suggests that, although both contractions and expansions of variable magnitude occur, the general tendency is toward a slow increase in expanded repeat size in the population.

The mild instability of the expanded CAG repeat provides a molecular basis to explain this disease's great variability in age at onset and clinical presentation. The highly significant correlation found in this study between expanded repeat size in the MJD gene with age at onset supports the influence of the CAG tract length in the determination of the disease presentation and progression. The repeat size accounts for ~50% of the variation in the age at onset ($r^2 = .485$); other factors—genetic, environmental, or both—therefore influence the natural history of the disorder. Similar patterns have been reported for other diseases caused by the same mutational mechanism (La Spada et al. 1992; Andrew et al. 1993; Duyao et al. 1993; Snell et al. 1993; Jodice et al. 1994; Ranum et al. 1994). However, the exact contribution of the repeat size to the of age at onset may not be precise, since several difficulties may affect the correlation between repeat size and age at onset. First, the age at onset of disease is never a precise figure, because the information given by the patients or their relatives may be inaccurate; namely, awareness of the disease in the family facilitates detection of first symptoms, creating a bias toward earlier onset in younger generations. On the other hand, the determination of the exact repeat size may also be inaccurate, because of the presence of multiple bands in the expanded allele, as seen in figure 4. Finally, the size of the expanded CAG repeat in lymphocytes can be different from the size in cells of involved structures that would be the actual determinants of the development of disease. This type of mosaicism has been described for HD, DRPLA, and SCA1 (Aoki et al. 1994; Chong et al. 1994; Telenius et al. 1994). Further studies are necessary to determine whether the phenomenon of somatic mosaicism occurs in MJD and whether repeat size in cells other than lymphocytes correlates more closely with age at onset of the disease.

The observed correlation between disease subphenotypes and repeat expansion is not surprising. Our results of this analysis and of juvenile onset cases suggest that the greater CAG expansions result in a more severe phenotype, which is qualitatively different from latter onset forms. This may represent cell-specific variable susceptibility, perhaps age dependent, of the target structures in the CNS.

The distribution of expanded alleles in the MJD popu-

lation appears symmetric. This is in contrast to the skewed distribution observed for the expanded allele in HD and SCA1. In both HD and SCA1, the largest alleles are associated with paternal transmission. In HD, severe instability of the allele during selected paternal transmission with unusually large expansions of up to 41 repeat units may explain the skewed distribution. Although our studies of variation in repeat size between both parent-offspring and sib pairs suggest increased volatility of paternal transmissions, large expansions have not been observed in MJD; the largest observed expansion was 9 repeats. This suggests that the mechanism of expansion may be different for these two diseases or may indicate the existence of genetic or environmental factors influencing some paternal transmissions in HD but not in MJD. Length of uninterrupted CAG repeat sequence is one factor that has been demonstrated to affect stability of the allele during transmission in fragile X syndrome (Eichler et al. 1994) and SCA1 (Chung et al. 1993). For SCA1, a case of an affected individual who had an apparently "normal" size allele consisting of an uninterrupted repeat has been reported (Goldfarb et al. 1994). It is known that a polymorphism CAG/CAA exists within the CAG repeat tract in the MJD gene (Kawaguchi et al. 1994). Further studies are required to determine whether this interruption of the CAG tract has an effect on stability during transmission and on clinical presentation of disease.

In contrast with myotonic dystrophy (Harley et al. 1993; Lavedan et al. 1993), there is no increased instability in the trinucleotide repeat with larger size of repeat in the transmitting parent. This suggests a different mechanism may be involved in instability during transmission for these two diseases.

The absence of association between variation in repeat number and anticipation, a finding also observed in HD (Andrew et al. 1993), remains unexplained. For MJD and other late onset trinucleotide diseases, it is important to consider that the data used to examine the relationship between anticipation and change in repeat size are subject to ascertainment bias. In retrospective studies like ours, one is likely to have large anticipation (i.e., the age at onset of the children will be much earlier than that of the parents) in order to obtain blood from affected offspring while the affected parent is still alive. In our analysis, mean age at onset in children of the 23 pairs was significantly lower than the age at onset of the entire population of affected individuals. Ascertainment bias is also suggested by comparison of variation in allele size during transmission between the affected and presymptomatic offspring, as shown in figure 4B. A correlation between anticipation and variation in allele size may exist but be masked by the nonrandom selection of the data used in the analysis.

In conclusion, our results validate the use of a direct

analysis of the CAG expansion in the MJD gene in the diagnosis of this disease. The significant correlation between CAG expansion size and disease severity/age at onset may also be of clinical value, though further studies will be required to clarify other genetic or environmental factors involved in the determination of the clinical presentation in MJD patients.

Acknowledgments

We would like to thank the families for their willingness to participate in this study, all the clinicians in the International Joseph Disease Foundation who provided clinical information for the study, Drs. B. White and J. Holden for their assistance. This project is supported by grants from the National Institutes of Health (NS 31687), the Fonds de Recherche en Santé du Québec, the Network of Centres of Excellence (Canadian Genetic Diseases Network) and the European Community Program STRIDE. P.M., C.G., and I.S. are recipients of a scholarship by the Junta Nacional de Investigaçã Científica e Tecnológica, Portugal.

References

- Anderson MA, Gusella JF (1984) Use of cyclosporin A in establishing Epstein-Barr virus-transformed human lymphoblastoid cell lines. *In Vitro* 20:856-858
- Andrew SE, Goldberg YP, Kremer B, Squitieri F, Theilmann J, Zeisler J, Telenius H, et al (1994) Huntington disease without CAG expansion: phenocopies or errors in assignment? *Am J Hum Genet* 54:852-863
- Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilmann J, Adam S, Starr E, et al (1993) The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat Genet* 4:398-403
- Aoki M, Abe K, Kameya T, Watanabe M, Nagata T, Itoyama Y (1994) Maternal anticipation of DRPLA and somatic heterogeneity of CAG repeats. *Am J Hum Genet Suppl* 55:A210
- Barceló JM, Mahadevan MS, Tsilfidis C, MacKenzie AE, Korneluk RG (1993) Intergenerational stability of the myotonic dystrophy protomutation. *Hum Mol Genet* 2:705-709
- Chong SS, McCall AE, Subramony S, Cota J, Orr HT, Hughes MR, Zogby HY (1994) Single cell analysis reveals gametic and tissue-specific instability of the SCA1 CAG repeat. *Am J Hum Genet Suppl* 55:A213
- Chung M, Ranum LPW, Duvick LA, Servadio A, Zoghbi HY, Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat Genet* 5:254-258
- Coutinho P, Andrade C (1978) Autosomal dominant system degeneration in Portuguese families of the Azores Islands. *Neurology* 28:703-709
- Coutinho P, Calheiros JM, Andrade C (1977) Sobre uma nova doença degenerativa do sistema nervoso central transmitida de modo autossômico dominante e aspectos familiares originários dos Açores. *O Médico* 82:1-3
- Coutinho P, Guimarães A, Scaravilli F (1982) The pathology of Machado-Joseph disease: report of a possible homozygous case. *Acta Neuropathol* 58:48-54
- Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Fontali M, Folstein S, et al (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nat Genet* 4:387-392
- Eichler E, Holden J, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, et al (1994) Length of uninterrupted CGG repeats determines instability in the FMR-1 gene. *Nat Genet* 8:88-94
- Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, Telenius H, et al (1993) Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. *Nat Genet* 5:174-179
- Goldfarb LG, Lunke A, Vasconcelos O, Platonov FA, Nagle J, Cervenakova L, Kononova SK, et al (1994) Mutation analysis of spinocerebellar ataxia type 1 in a large Yakut kinship of eastern Siberia. *Am J Hum Genet Suppl* 55:A221
- Harley HG, Rundle SA, MacMillan JC, Myring J, Brook JD, Crow S, Reardon W, et al (1993) Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *Am J Hum Genet* 52:1164-1174
- Jodice C, Malaspina P, Persichetti F, Novelletto A, Spadaro M, Giunti P, Morocutti C, et al (1994) Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia 1. *Am J Hum Genet* 54:959-965
- Kanda T, Isozaki E, Kato S, Tanabe H, Oda M (1989) Type III Machado-Joseph disease in a Japanese family: a clinicopathological study with special reference to the peripheral nervous system. *Clin Neuropathol* 8:134-141
- Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, et al (1994) CAG expansions in a novel gene from Machado-Joseph Disease at chromosome 14q32.1. *Nat Genet* 8:221-227
- Koide R, Ikeuchi T, Onodera O, Tanaka h, Igarashi S, Endo K, Takahashi H, et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolysian atrophy (DRPLA). *Nat Genet* 6:9-13
- La Spada AR, Roling DB, Harding AE, Warner CL, Spiegel R, Petruszewicz IH, Yee W-C, et al (1992) Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nat Genet* 2:301-304
- Lavedan C, Radvanyi HH, Shelbourne P, Rabes JP, Duros C, Savoy D, Dehaupas I, et al (1993) Myotonic dystrophy: size- and sex-dependent dynamics of CTG meiotic instability and somatic mosaicism. *Am J Hum Genet* 52:875-883
- Lima L, Coutinho P (1980) Clinical criteria for diagnosis of Machado-Joseph disease: report of a non-Azorean Portuguese family. *Neurology* 30:319-322
- Neter J, Wasserman W, Kutner MH (1985) Applied linear statistical models. Irwin, Homewood, IL
- Radvanyi J, Camargo CHP, Costa ZM, Fonseca NC, Nascimento ED (1993) Machado-Joseph disease of Azorean ancestry in Brazil: the Catarina kindred. *Arq Neuropsiquiatr* 51:21-30
- Ranum LPW, Chung M, Banfi S, Bryer A, Schut LJ, Ramesar R, Duvick LA, et al (1994) Molecular and clinical correlations in spinocerebellar ataxia type 1: evidence for familial effects on the age at onset. *Am J Hum Genet* 55:244-252
- Rosenberg RN, Nyhan WL, Bay C, Shore P (1976) Autosomal

- dominant striatonigral degeneration. *Neurology* 26:703-714
- Rubisztein DC, Amos W, Leggo J, Goodburn S, Ramesar RS, Old J, Bontrop R, et al (1994) Mutational bias provides a model for the evolution of Huntington's disease and predicts a general increase in disease prevalence. *Nat Genet* 7:525-530
- Sambrook J, Fritsch EF, Maniatis (eds) (1989) *Molecular cloning: a laboratory manual*, 2d ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sakai T, Ohta M, Ishino H (1983) Joseph disease in a non-Portuguese family. *Neurology* 33:74-80
- Sequeiros J, Coutinho P (1993) Epidemiology and clinical aspects of Machado-Joseph disease. In: Harding A, Deufel T, Chamberlain S (eds) *Advances in neurology*. Raven Press, New York, pp 139-153
- Sequeiros J, Coutinho P (1981) Genetic aspects of Machado-Joseph disease. *Brotéria-Genética* 2:137-147
- Sequeiros J, Silveira I, Maciel P, Coutinho P, Manaia A, Gaspar C, Burlet P, et al (1994) Genetic linkage studies of Machado-Joseph disease with chromosome 14q STRPs in 16 Portuguese-azorean kindreds. *Genomics* 21:645-648
- Snell RG, MacMillan JC, Cheadle JP, Fenton I, Lazarou LP, Davis P, MacDonald ME, et al (1993) Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nat Genet* 4:393-397
- Sudarsky L, Corwin L, Dawson D (1992) Machado-Joseph disease in New England: clinical description and distinction from the olivopontocerebellar atrophies. *Mov Disord* 7:204-208
- Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, Shimazaki H, et al (1993) The gene for Machado-Joseph disease is mapped to chromosome 14q. *Nat Genet* 4:300-304
- Telenius H, Kremer B, Goldberg YP, Jane Theilmann J, Andrew SE, Zeisler J, Adam S, et al (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet* 6:409-414
- Twist EC, Causaubon LK, Ruttledge M, Rao VS, MacLeod PM, Radvany J, Zhao Z, et al (1995) Machado-Joseph disease maps to the same region of chromosome 14 as the spinocerebellar ataxia type 3 locus. *J Med Genet* 31:823-829

ARTIGO II

Gender equality in Machado-Joseph disease

Sir—In the April 1995 issue of *Nature Genetics*, Kawakami *et al.*¹ provided evidence for a gender effect in Machado-Joseph disease, a form of autosomal dominant cerebellar ataxia type I linked to chromosome 14q, which is allelic to spinocerebellar ataxia 3 (SCA3) (ref. 2). They analyzed 14 affected sib pairs of which seven were of different sexes and found that brothers develop symptoms about ten years earlier than sisters. This prompted us to look for a gender effect in our series of five sib pairs and

six triplets of different sexes from 11 families with the MJD/SCA3 CAG expansion. We averaged the age at onset between sibs of the same sex in triplets. Mean ages at onset in brothers and sisters were not significantly different, 37.2 ± 11 (22–57) and 39.6 ± 15 (22–70), respectively. Onset in brothers occurred at a mean of -2.45 years earlier than in sisters but this was not significant ($P=0.16$, paired *t*-test). The mean difference between the number of CAG repeats in males (72.8 ± 3) and females (73.5 ± 5) was also not significant (-0.64 , $P=0.35$). The difference between the actual age at onset and the expected age at onset, determined from the linear regression correlating age at onset and number of CAG repeats in 89 patients, was not significantly different in males (-0.6 ± 5 years) and in females (2.9 ± 7 years) from the discordant sibships.

We found no statistically significant differences between 41 and 48 affected men and women from 28 families with the MJD/SCA3 expansion. The mean age at onset was 38 ± 11 years for men and 36 ± 13 years in women, and the mean number of CAG repeats

72.6 ± 4 and 74.0 ± 4 , respectively. The regression and correlation coefficients between age at onset and CAG repeat number were similar in men ($r = -0.82$; slope -2.61) and in women ($r = -0.79$; slope -2.79) (Fig. 1). Furthermore, the covariance analysis using the repeat number as the variant to control for its effect on age at onset was not significant. In these families, the number of CAG repeats accounts for approximately 60% of the variability in age at onset, a value which is similar to that found in spinocerebellar ataxia 1 (SCA1) (ref. 3).

Our results indicate that gender has no sizeable effect on age at onset. Previous results suggesting a gender effect may be attributable to a bias due to the smaller number of sibs analyzed. Our finding is reinforced by the absence of a gender effect on age at onset in MJD in Brazilian families of Portuguese ancestry⁴. As yet unidentified genetic or environmental factors could be responsible for the residual age at onset variance in autosomal dominant cerebellar ataxias.

Alexandra Dürr
Giovanni Stevanin
Géraldine Cancel
Nacer Abbas
Hervé Chneiweiss
Yves Agid
Josué Feingold
Alexis Brice

INSERM U 289 and Fédération de Neurologie, Hôpital de la Salpêtrière, 47 boulevard de l'Hôpital, 75651 Paris Cedex 13, France

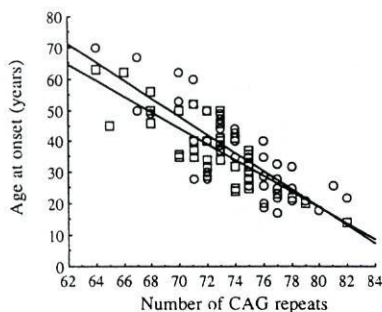


Fig. 1 Correlation between age at onset and the number of CAG repeats according to sex. Circles indicate women, squares indicate men.

Sir—Age at onset of Machado-Joseph disease (MJD) is inversely correlated with number of CAG repeats in the *MJD1* gene⁵. We have confirmed these results in a data set of 156 affected individuals from North America, Brazil and Portugal⁶. The number of CAG repeats accounts for 50% of the variation in age at onset suggesting other factors influence onset⁶. Kawakami *et al.*¹ recently proposed that gender and gene dosage affect onset age. Our studies suggest that gender has little effect, but that variation in onset age is partially accounted for by an effect common to sibships that is independent of CAG repeat length.

We found no significant difference in age at onset between males (38.1 ± 1.3) and females (37.4 ± 1.5)

($P=0.72$). In a multiple regression model with age at onset as the dependent variable, CAG repeat length was significantly correlated with onset ($P=0.0001$) while patient gender was not ($P=0.30$). However, familial factors other than CAG repeat length might mask a gender effect in age at onset. To control for familial factors, we analyzed 27 brother–sister pairs. Mean difference in onset ages for sister minus brother did not differ significantly from zero (-2.3 ± 1.9 , $P=0.23$). Although Kawakami *et al.*¹ saw a large positive difference in onset age for sister–brother pairs, only seven unlike-sex pairs were available. These differences do not account for the influence of CAG repeat length on age at onset and may not be an accurate measure of a gender effect.

Kawakami *et al.*¹ controlled for CAG repeat length by examining sib pairs with identical CAG repeat length. However, determination of exact repeat size may be inaccurate due to the presence of multiple bands in the expanded allele. Therefore, a small range of repeat differences should be considered. In our data set, among 16 sib pairs of different sex whose CAG repeat length differed by two units or less, the mean difference in onset age for sisters minus brothers was 0.9 ± 2.0 ($P=0.47$). Brothers had onset four years earlier on average than sisters when limiting the sample to the 11 sib pairs that differ by 1 CAG repeat or less ($P=0.02$).

A more rigorous method to account for CAG repeat length is analysis of covariance, with onset age as the dependent variable, family ID and sex as independent variables and CAG repeat length included as a covariate. In our data set of 44 individuals in 18 sibships, sex was a nonsignificant ($P=0.42$) predictor of onset age while CAG repeat number was significant ($P=0.0007$).

The evidence for a gender effect in our data set is contradictory. An effect was apparent only among brothers and sisters whose CAG repeat length differed by 1 unit at the most. However, caution is required for small numbers of observations; our analysis of covariance results from a larger sample argue against a gender effect on age at onset.

Kawakami *et al.*¹ also describe a homozygote with 67 CAG-repeats on both chromosomes and onset at age 28. The 95% confidence interval (CI) for onset age for a heterozygote with 67 repeats is 35–68 years, based on our regression analysis. Although the onset age for the homozygote is seven years outside this interval, we have found several individuals with onset ages outside the predicted range. An individual homozygous for the expanded *MJD1* allele (70 and 72 repeats) was recently reported⁷. The onset age of 16 years is again outside the 95% CI for predicted age at onset. Four heterozygotes also had onset ages outside the predicted range, indicating variability in onset age

occurs regardless of the number of expanded alleles. Therefore, while these two confirmed homozygotes and a putative homozygote^{8, 9} are consistent with a gene dosage effect, additional studies are required to confirm such an effect.

Anita L. DeStefano

Lindsay A. Farrer

Department of Neurology, Boston University School of Medicine, Boston, Massachusetts 02118 USA

Patricia Maciel

Claudia Gaspar

Guy A. Rouleau

Centre for Research in Neuroscience, McGill University and the Montreal General Hospital Research Institute, Montreal, Quebec, Canada

Paula Coutinho

Jorge Sequeiros

Lab. Genetica Medica, ICBAS and UnIGENE, IBMC Universidade do Porto, Portugal

1. Kawakami, H. *et al.* *Nature Genet.* **9**, 344–345 (1995).
2. Cancel, G. *et al.* *Am. J. hum. Genet.* (in the press).
3. Dubourg, O. *et al.* *Ann. Neurol.* **37**, 176–180 (1995).
4. Stevanin, G. *et al.* *J. med. Genet.* (in the press).
5. Kawaguchi, Y. *et al.* *Nature Genet.* **8**, 221–228 (1994).
6. Maciel, P. *et al.* *Am J. hum. Genet.* **57**, 54–61 (1995).
7. Takiyama, Y. *et al.* *Hum. molec. Genet.* **4**, 1137–1146 (1995).
8. Coutinho, P. & Andrade, C. *Neurol.* **28**, 703–708 (1978).
9. Coutinho, P., Gunimaraes, A. & Scaravilli, F. *Acta Neuropathol.* **58**, 48–54 (1982).

ARTIGO III

A Familial Factor Independent of CAG Repeat Length Influences Age at Onset of Machado-Joseph Disease

Anita L. DeStefano,¹ L. Adrienne Cupples,² Patricia Maciel,^{5,6} Claudia Gaspar,^{5,6} Joao Radvany,⁸ David M. Dawson,^{3,9} Lewis Sudarsky,^{3,9} Lee Corwin,^{1,4} Paula Coutinho,⁷ Patrick MacLeod,¹⁰ Jorge Sequeiros,⁶ Guy A. Rouleau,⁵ and Lindsay A. Farrer^{1,2}

Departments of ¹Neurology and ²Epidemiology and Biostatistics, Boston University School of Medicine, ³Neurology Division, Brigham and Women's Hospital, and ⁴Department of Neurology, Harvard Medical School, Boston; ⁵Centre for Research in Neuroscience, McGill University and the Montreal General Hospital Research Institute, Montreal; ⁶Lab. Genética Médica, ICBAS, and UniGENe, IBMC, Universidade do Porto, and ⁷Servico de Neurologia, Hospital Santo Antonio, Porto, Portugal; ⁸Neurology Service, Veterans Administration Medical Center, West Roxbury, MA; ⁹Neurologia, Hospital Israelita Albert Einstein, Sao Paulo; and ¹⁰Section of Genetics, Department of Laboratory Medicine, Victoria General Hospital, Victoria

Summary

Machado-Joseph disease (MJD) is a late-onset, progressive, neurodegenerative disorder caused by the expansion of an unstable trinucleotide (CAG) repeat sequence in a novel gene (*MJD1*) on chromosome 14. Previous studies showed that age at onset is negatively correlated with the number of CAG repeat units, but only part of the variation in onset age is explained by CAG repeat length. Ages at onset and CAG repeat lengths of 136 MJD patients from 23 kindreds of Portuguese descent were analyzed, to determine whether familial factors independent of CAG repeat length modulate age at onset of MJD. Correlation among sibs for onset age adjusted for CAG repeat length was .43, which indicates that an environmental or genetic factor common to sibs influences onset age. Positive correlations were also observed for avuncular ($r = .22$) and first-cousin pairs ($r = .28$), which supports the hypothesis that a genetic factor is influencing age at onset. Commingling analysis of onset ages adjusted for CAG repeat length identified three distributions in this population of affected individuals. Further studies of a much larger sample are needed to determine whether these distributions represent the influence of a genetic or environmental factor.

guese-Azorean families (Nakano et al. 1972; Woods and Schaumburg 1972; Rosenberg et al. 1976). Further studies led to the recognition of these cases as manifestations of the same disease (Romanul et al. 1995; Coutinho and Andrade 1978), and, since then, numerous Portuguese MJD families (from the Azores, mainland Portugal, Brazil, the United States, and Canada) and presumed non-Portuguese families (Japanese, east Indian, Italian, Australian aboriginal, and African-American) have been reported (Healton et al. 1980; Sakai et al. 1983; Livingstone and Sequeiros 1984; Bharucha et al. 1986; Sudarsky et al. 1992; Burt et al. 1993; Sequeiros and Coutinho 1993). On the basis of clinical presentation and age at onset, MJD patients can be classified as having one of three subtypes (Coutinho and Andrade 1978). Type II is the most frequent presentation, with progressive ataxia, pyramidal signs, ophthalmoplegia, and mean age at onset of ~40 years. Type I is characterized by early onset with marked pyramidal and extrapyramidal signs in addition to those seen in type II. Type III is associated also with marked distal atrophies and sensory loss and onset between ages 40 and 60 years. Prior to the identification of the genetic defect for MJD, the existence of environmental or genetic modifiers, which influence the expression of the main genetic defect, was proposed by numerous authors to explain the heterogeneity of phenotype and onset age (Woods and Schaumburg 1972; Bharucha et al. 1986; Rosenberg 1992; Sudarsky et al. 1992; Sequeiros 1993).

Linkage analysis in Japanese families (Takiyama et al. 1993) and, subsequently, in families of Portuguese descent (Sequeiros et al. 1994; St George-Hyslop et al. 1994; Twist et al. 1995) localized the MJD gene to chromosome 14q32.1. Identification of an expanded trinucleotide repeat sequence as the gene defect in other inherited degenerative diseases led investigators to suspect that such a defect may be responsible for MJD. Recently, this hypothesis was confirmed by the discovery of an unstable expansion in MJD patients of a CAG repeat in a novel gene (*MJD1*) at this chromosomal location (Kawaguchi et al. 1994).

Introduction

Machado-Joseph disease (MJD) is a progressive neurodegenerative disorder with an autosomal dominant pattern of inheritance. The clinical features of MJD were initially described as distinct disorders in three Portu-

Received November 28, 1995; accepted for publication April 11, 1996.

Address for correspondence and reprints: Dr. Lindsay A. Farrer, Department of Neurology, Boston University School of Medicine, 80 East Concord Street, Boston, MA 02118.

© 1996 by The American Society of Human Genetics. All rights reserved.
0002-9297/96/5901-0017\$02.00

Linkage studies have suggested that the spinocerebellar ataxia type 3 (SCA3) locus is the same as or lies close to MJD1 (Stevanin et al. 1995). Subsequent screening of a German SCA patient population revealed that the majority of SCA families carried the expansion at MJD1, providing strong evidence that SCA3 and MJD are due to the same defect and may be the same disease (Schöls et al. 1995). Genetic diagnosis of MJD (Kawaguchi et al. 1994), SCA1 (Orr et al. 1993), and dentatorubropallidolusian atrophy (DRPLA) (Koide et al. 1994) has already increased our understanding of the epidemiology of the autosomal dominant spinocerebellar ataxias (Cancel et al. 1995; Ranum et al. 1995; Silveira et al. 1996) and will likely lead to a revised nosology of this group of disorders.

Identification of the (CAG)_n repeat motif in the MJD1 gene provides some explanation for the clinical heterogeneity of MJD. Age at onset has been shown to be negatively correlated with the number of CAG repeats. This correlation ranged from $-.87$ to $-.67$ in various studies (Kawaguchi et al. 1994; Maciel et al. 1995; Maruyama et al. 1995; Ranum et al. 1995; Takiyama et al. 1995). It is evident, however, that CAG repeat length does not account for all of the variation in age at onset. In fact, the range of onset age for a specific CAG repeat length is large enough to preclude prediction of age at onset when counseling presymptomatic carriers.

Negative correlation between onset and CAG repeat length has also been observed for Huntington disease (HD) (Andrew et al. 1993; Duyao et al. 1993), DRPLA (Koide et al. 1994), and SCA1 (Orr et al. 1993; Ranum et al. 1994). It has been suggested that additional genetic factors influence age at onset of these diseases because variation in onset age remains after accounting for CAG repeat length (Duyao et al. 1993; Kremer et al. 1993; Burke et al. 1994; Ranum et al. 1994). Ranen et al. (1995) found that sex and age at onset of transmitting parent accounted for 22% of the variation in age at onset in HD not explained by the CAG repeat length. In a study of 87 SCA1 kindreds, Ranum et al. (1994) found a linear correlation of $-.77$ between onset age and number of CAG repeats, which indicates that CAG repeat length accounts for 60% of the variation in onset age. Analysis of variance on the seven largest kindreds showed significant differences between families after correcting for number of repeats. CAG repeat length and interfamilial differences accounted for 66% and 5%, respectively, of the variation in onset age, which suggests that a genetic or environmental effect common to families also influences age at onset in SCA1.

The aim of this study was to determine whether familial factors independent of CAG repeat length modulate age at onset of MJD. Two methods for separating the effect of CAG repeat length from other familial effects on age at onset were employed. In one approach, linear regression was used to adjust onset age for CAG repeat

length prior to analyses. In addition, a method that allows for simultaneous estimation of the effect due to CAG repeat and a second effect due to sibship was applied to the data.

Subjects and Methods

Subjects

DNA samples and age-at-onset information were obtained from 136 affected members of 9 Azorean, 12 North American, and 2 Brazilian kindreds, all of Portuguese ancestry. The diagnosis of MJD was determined by clinical examination by experienced neurologists using established diagnostic criteria (Lima and Coutinho 1980). Age at onset was based on information provided by the patient and/or a close relative. Determination of number of CAG units for these individuals was described elsewhere (Maciel et al. 1995). Twenty patients from 10 families included in the original report were excluded because DNA samples and onset age information were available for only a single individual in the family or the affected individuals were more distantly related than first-cousins and thus did not contribute any information to the present study.

Statistical Methods

Two methods for estimation of familial correlation were employed. In the first approach, to investigate sibship effects for age at onset, onset age was modeled as

$$y_{ij} = \mu + bx_{ij} + \varepsilon_{ij},$$

where y_{ij} is observed age at onset of MJD for sibling j in sibship i , μ is mean age at onset in all subjects, x_{ij} is the number of CAG repeats for sibling j in sibship i , b is the effect of CAG repeat number, and ε_{ij} is a random residual for sibling j in sibship i . Ages at onset for members of the same sibship were treated as repeated measures. The (co)variance matrix for members of the same sibship was modeled as having compound symmetry. This structure defines an equal covariance between siblings in addition to the usual variance element on the diagonal. Therefore, the matrix of covariance for residuals is a block diagonal matrix, with each block for subjects in the same sibship as follows

$$\begin{bmatrix} \sigma^2 + \sigma_1 & \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma^2 + \sigma_1 & \sigma_1 \\ \sigma_1 & \sigma_1 & \sigma^2 + \sigma_1 \end{bmatrix}.$$

The dimension of each block corresponds to the number of siblings within a particular sibship. The off-diagonal elements, σ_1 , estimate the covariance among siblings and an estimate of the common correlation coefficient among sibs is

$$\frac{\hat{\sigma}_1}{\hat{\sigma}^2 + \hat{\sigma}_1}$$

Estimates of the residual (co)variance were obtained using the repeated measures options in PROC MIXED, which is part of the Statistical Analysis System (SAS) (SAS Institute 1992). In this model, this residual correlation indicates the association among siblings after the effect of CAG repeat is removed. A repeated measure model that included sex of transmitting parent and CAG repeat length was also fit.

In the second approach, residuals from the regression of age at onset on CAG repeat number for all affected individuals in the population were computed. Correlations among the residuals and age at onset for parent-offspring, sib, avuncular (aunt-niece, aunt-nephew, uncle-niece, uncle-nephew), and first-cousin pairs were calculated using FCOR in SAGE (SAGE 1994). For pairs for which no natural ordering exists, such as those between sibs and cousins, FCOR computes the intraclass correlation by forming all possible pair orderings and computing the Pearson correlation coefficient (double pairwise method). Three methods of weighting pairs included in the correlation are used in FCOR for both intraclass and interclass correlations. Equal weight is given to each pair, to each pedigree, or to each nuclear family.

Standard methods of significance testing and confidence-interval construction cannot be applied to correlations if the pairs are not independent, as often is the case in familial studies. Confidence intervals (CI) were constructed using Fisher's z transformation (Snedecor and Cochran 1980, pp. 184–86). To adjust for dependencies in the data, the effective number of pairs in the sample, K_e , was used to compute CIs for all correlations. K_e was computed as follows

$$K_e = \sum_{i=1}^N \frac{k_i}{1 + (k_i - 1)r_{\text{sib}}} \quad (1)$$

For parent-offspring correlations, k_i is the number of offspring of parent i , N is the total number of parents, and r_{sib} is the intraclass correlation among siblings (Rosner et al. 1979). For sibling correlations, k_i is the number of unique pairs in sibship i , N is the total number of sibships, and r_{sib} is the intraclass correlation among sibs (Sepulveda et al. 1994). Because methods for construction of CIs for avuncular or first-cousin pairs have not been developed, to our knowledge, the above methods were modified to allow determination of K_e for cousin and avuncular pairs. The effective number of first-cousin pairs was calculated using (1) and letting k_i equal the number of first-cousin pairs in a "cousinship" and letting N be the total number of cousin groups. Distinct cousin groups within the same pedigree were

assumed to be independent. Individuals included in cousin pairs within a group are related to each other as cousins or as siblings. Because the sibling correlation is expected to be larger than the first-cousin correlation, r_k was conservatively set equal to the sibling correlation. Choosing a larger value for r_k will increase the denominator in (1), resulting in a smaller K_e and hence a larger, more conservative CI.

The equation for K_e has been extended to allow for a variable number of individuals in each class when computing correlations between two classes (Rosner 1982). Although illustrated for parent-offspring correlations where either one or both parents and a variable number of offspring were sampled, it was applied here to compute K_e for avuncular pairs. Given

$$K_e = \sum_{i=1}^N \frac{k_i l_i}{[1 + (k_i - 1)r_k][1 + (l_i - 1)r_l]} \quad (2)$$

we let k_i equal the number of nieces or nephews in pedigree i , l_i equal the number of aunts or uncles in pedigree i , r_k equal the correlation among the nieces and nephews in a pedigree, and r_l equal the correlation among the aunts or uncles in a pedigree. Individuals included as aunts or uncles in avuncular pairs within a single avuncular group are siblings, so r_l was set equal to the sibling correlation. Individuals included as nieces and nephews in avuncular pairs within an avuncular group are either siblings, half-siblings, or first-cousins. Therefore, as a conservative measure, r_k was set equal to the sibling correlation. While the majority of pedigrees contained a single avuncular group, 3 of the 10 pedigrees that contributed avuncular pairs had multiple avuncular groups; avuncular groups within a pedigree were assumed to be independent.

Commingle analyses were performed on standardized residuals from the regression of age at onset on CAG repeat number using the REGC program of SAGE (SAGE 1994) to determine whether the population of residuals represented a single distribution or was a mixture of multiple distributions (MacLean et al. 1976). Models with one, two, or three distributions were fit. Models for residuals transformed by the standardized generalized modulus power transformation were also fit to test for skewness in the distribution(s). Parameters estimated in one or more of these analyses include the percent of subjects in population A (q^2), the phenotypic mean (μ_i), for each population ($i = A, B, C$), and a variance common to all populations (σ^2). Populations A, B, and C correspond to the distributions being fit in each model. The percent of subjects in populations B and C were a function of the proportion in population A: $2(q)(1 - q)$ and $(1 - q)^2$, respectively. An environmental model was specified by setting familial correlations to 0 and setting $\tau_A = \tau_B = \tau_C = q$, where τ_i is the

transmission probability for each population. Therefore, although REGC utilizes the pedigree structure of each kindred in analysis of the data, the model specifies no genetic transmission. In a genetic model, the three populations would correspond to the three major locus types (AA, AB, and BB) and q to the frequency of the A allele. The relationship between the proportion of subjects in each population assumes Hardy-Weinberg equilibrium for a genetic model. The power parameter (λ_1) and the shift parameter (λ_2) were also estimated for models analyzing power transformed data. Hypotheses regarding the distribution of the observed data were tested by comparing nested models via a likelihood ratio test.

The clinical subtype was available for 119 affected individuals. Differences in standardized residuals from the regression of age at onset on CAG repeat number between patients classified as having clinical subtype I, II, or III were compared by analysis of variance.

Results

Ages at onset for the 136 MJD patients ranged from 10 to 62 years, with a mean onset of $37.9 \pm .9$ years. The mean number of CAG repeat units was $73.2 \pm .3$, with a minimum and maximum of 66 and 84 years, respectively.

Adjusting for CAG repeat, the estimated covariance among siblings when treating sibs as repeated measures was significantly >0 ($\hat{\sigma}_1 = 28.8$; $P = .009$). The estimated common correlation among sibs was .43. Because the number of CAG repeats was included in the model, the significant effect ($P = .001$) of CAG repeat length on age at onset is accounted for in this correlation estimate. This estimate is based on 81 individuals representing 34 sibships with multiple affected siblings. The square of the correlation, $r^2 = .18$, is an estimate of the amount of variation in onset age explained by being in the same sibship.

Sex of transmitting parent was not significant in the repeated measures model ($P = .14$), and the estimated correlation among sibs was unchanged when this variable was included in the model. Therefore, sex of transmitting parent was not included in subsequent analyses.

The Pearson correlation among sibs, as well as other types of relative pairs for age at onset and for residuals obtained when age at onset is regressed on number of CAG repeats, are shown in table 1. Age at onset was positively correlated for parent-offspring, sibling, cousin, and avuncular pairs. A positive correlation was also found for sib, avuncular, and cousin pairs when age at onset was adjusted for number of CAG repeats. In contrast to all other relative pairs, the correlation between residuals for parent-offspring pairs was negative. Less than half the number of parent-offspring pairs were available ($n = 24$) compared to the other types of relatives considered.

Table 2 indicates that some pedigrees provided much more information than others. Moreover, many of the relative pairs were contributed by a single pedigree and therefore are not independent. For example, the 15 sib pairs in pedigree 303 belong to a single sibship of six affected individuals. As a result of such dependencies in the data, K_e values (table 1) were less than the total number of observed pairs (table 2). Approximate 95% CIs based on K_e included 0 for parent-offspring, avuncular, and cousin pairs for adjusted onset age and for avuncular and cousin pairs for unadjusted onset age, which indicates that these correlations may not be significant.

Commingle analysis of standardized residuals (table 3) showed that, although a single distribution fitted the data as well as two (model 1a vs. 2a: $\chi^2_{(2)} = 2.42$; $P > .25$), there was evidence for a model of three distributions (model 1a vs. 3a: $\chi^2_{(3)} = 8.56$; $P < .05$). Use of the power transformation yielded a significantly better fit (model 3a vs. 3b: $\chi^2_{(2)} = 8.42$; $P < .05$) for the three distribution model, which suggests that skewness may be present in the data.

Figure 1 shows the three populations, as defined by the parameter estimates obtained from the three distribution model with no transformation (table 3), overlaid on a histogram of the standardized residuals. The mean for each distribution can be scaled to years by multiplying by 8.7, which is the SD of the residuals. Population A in figure 1 corresponds to a distribution of patients with onset age, on average, 11.8 years above the onset age predicted by regression of age at onset on CAG repeat length. Population C corresponds to a distribution of patients with average onset 11.2 years earlier than predicted on the basis of CAG repeat length. A third distribution of patients with onset age similar to the predicted value is population B.

To illustrate how MJD patients may be described by these three populations, consider an affected individual with 71 CAG repeat units and onset age of 44 years. Predicted age at onset for an individual with 71 CAG repeats is 45.5 years, on the basis of the regression of age at onset on CAG repeat length in this population. The predicted and observed onset age differ by only 1.5 years for this individual. It is interesting to note that an affected sibling of this patient with 74 CAG repeat units had an onset age of 38 years, which differs from the predicted onset by only 1.93 years. These siblings are both part of population B, which consists of subjects with onset ages similar to the predicted value. In contrast, we observed an individual with 75 CAG repeat units and onset of symptoms at 21 years of age. This age at onset is 12.9 years earlier than predicted for an individual with 75 repeat units, i.e., 33.9 years. This patient therefore would be classified as a member of population C, which contains patients with an average onset 11.2 years earlier than predicted by regression of onset age on CAG repeat length.

Table 1

Correlation for Age at Onset of MJD and Age at Onset Adjusted for Number of CAG Repeats (Residual) among Various Types of Relative Pairs

RELATIONSHIP	ONSET AGE				RESIDUAL			
	K_e	Pairs	Pedigree	Nuclear	K_e	Pairs	Pedigree	Nuclear
Parent-Offspring	19.5	.55 (.14, .80)*	.61 (.23, .83)	.60 (.21, .81)	20.2	-.24 (-.62, .24)	-.33 (-.68, .14)	-.22 (-.65, .25)
Sibling	38.7	.58 (.33, .76)	.63 (.39, .79)	.57 (.31, .75)	41.7	.43 (.15, .65)	.47 (.19, .68)	.41 (.12, .64)
Avuncular	22.4	.27 (-.17, .62)	.40 (-.02, .70)	...	27.6	.22 (-.17, .55)	.26 (-.13, .58)	...
Cousin	23.2	.23 (-.20, .59)	.41 (.001, .70)	...	26.7	.28 (-.11, .60)	.30 (-.09, .61)	...

NOTE.—Correlations were computed with equal weighting given to each pair, each pedigree, or each nuclear family by using FCOR in SAGE (SAGE 1994).

* 95% confidence interval based on effective number of pairs, K_e . Formulae for K_e given in text.

Histograms of the standardized residuals for populations of patients classified by clinical subtype are shown in figure 2. Type II patients, who represent the largest group, had a wide range of standardized residuals, whereas clustering of residuals for type I and type III patients were observed. The mean standardized residuals are $-.80$, $-.19$, and $.83$ for types I, II, and III, respectively. The mean for type III is significantly larger than those for the other types ($P < .05$). Comparison of figures 1 and 2 shows a correspondence between population A identified by commingling analysis and the population of patients with type III. Although clustered, type I patients appear equally distributed in populations B

and C. Classification of patients into these clinical types is based to a large extent on age at onset. The clustering of the subtypes within a range of residuals therefore reflects the tendency of patients with older onset ages to have positive residuals and those with very young onset ages to have negative residuals.

Discussion

The moderate, positive correlation between sibs for age at onset adjusted for CAG repeat length, obtained by two methods, demonstrates that familial factors other than CAG repeat length influence onset age of MJD. The two methods gave comparable results for these data, although the observations were handled differently by each. When PROC MIXED is used, the observations are described by a statistical model that estimates the correlation among sibs. Treating onset ages for siblings as repeated measures and including CAG repeat length as a covariate allowed for simultaneous estimation of correlation and effect of the CAG repeat length. In contrast, computation of correlation among siblings for residual onset age required prior adjustment for the effect of CAG repeat number.

A desirable feature of the repeated measure analysis is that it correctly handles sibships with more than two affected sibs. This is not the case when one simply computes the Pearson correlation over all sib pairs. This computation assumes the pairs are independent. If sibships with more than two affected siblings are present, however, individuals will be included more than once when forming all possible pairs and disproportionate emphasis will be given to data from individuals that appear more than once. Three weighting procedures were implemented to reduce the effect of repeated individuals: (1) equal weight given to each pair; (2) equal weight given to each pedigree; and (3) equal weight given to each nuclear family. Giving equal weight to each nuclear family addressed the issue of lack of independent pairs, but it is only an approximate adjustment.

Table 2

Number of Relative Pairs Contributed by Each Pedigree

Pedigree	Parent-Offspring	Sibling	Avuncular	Cousin
101	0	7	9	23
102	2	7	12	2
103	3	4	11	2
104	1	0	0	1
105	2	1	1	2
106	1	0	1	1
110	1	0	0	0
111	0	2	4	0
113	2	0	0	1
115	0	3	0	0
201	0	4	0	6
208	0	5	0	8
214	5	8	6	3
216	0	1	0	2
217	0	1	3	2
225	2	1	4	2
226	0	1	2	0
235	0	1	0	0
244	0	1	0	0
301	5	9	14	13
303	0	15	0	0
Total	24	71	67	68

Table 3

Commingling Analysis of Standardized Residuals Obtained from Regression of Onset Age on Number of CAG Repeats

Model	Transformed	q	μ_A	μ_B	μ_C	σ^2	λ_1	λ_2	N^a	$-2\ln(L)$
1a	No	1.0	.0099	2	384.95
1b	Yes	1.0	1.6798	.79	1.60	4	383.77
2a	No	.50	-.36	1.1059	4	382.53
2b	Yes	.11	2.48	.5446	1.57	.96	6	377.96
3a	No	.50	1.36	-.08	-1.29	.19	5	376.39
3b	Yes	.53	1.84	.63	-.90	.13	.22	-.13	7	367.97

NOTE.—Models assuming a single, two, or three distributions were fit both with and without transformation. Parameters are defined in the text.

^a Number of independent parameters.

Therefore, a repeated measures model may be more suitable than the nuclear family-weighted procedure when estimating correlation among sibs. Neither the repeated measures model nor the weighted correlations, as used in this study, correctly accounted for the occurrence of multiple sibships within the same kindred.

A limitation of the repeated measure approach for analysis of family data is that modeling and obtaining

estimates of the covariance matrix become complex when relatives who are not equally related are included. Each block of the block-diagonal (co)variance matrix is then a unique covariance matrix that must be specified for every pedigree, and a covariance element for each relationship (e.g., parent-offspring, parent-parent, sibling, half-sib, first-cousin, etc.) must be estimated. Specification of such a (co)variance matrix for estimation of

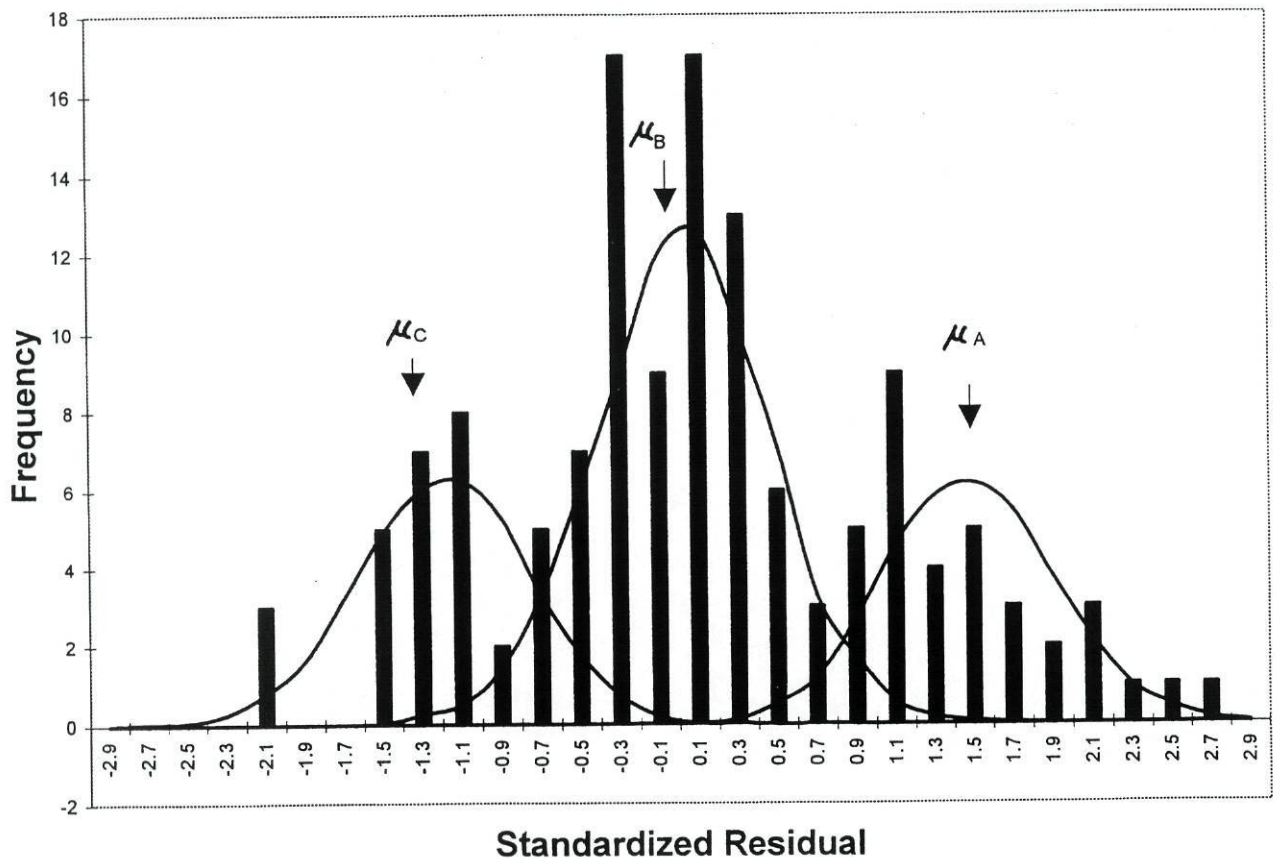


Figure 1 Histogram of standardized residuals from regression of onset age on CAG repeat length. Three distributions identified by commingling analysis corresponding to populations A, B, and C with means μ_A , μ_B , and μ_C , respectively, are indicated. The units of the x-axis are SDs. The SD of the residuals was 8.7 years.

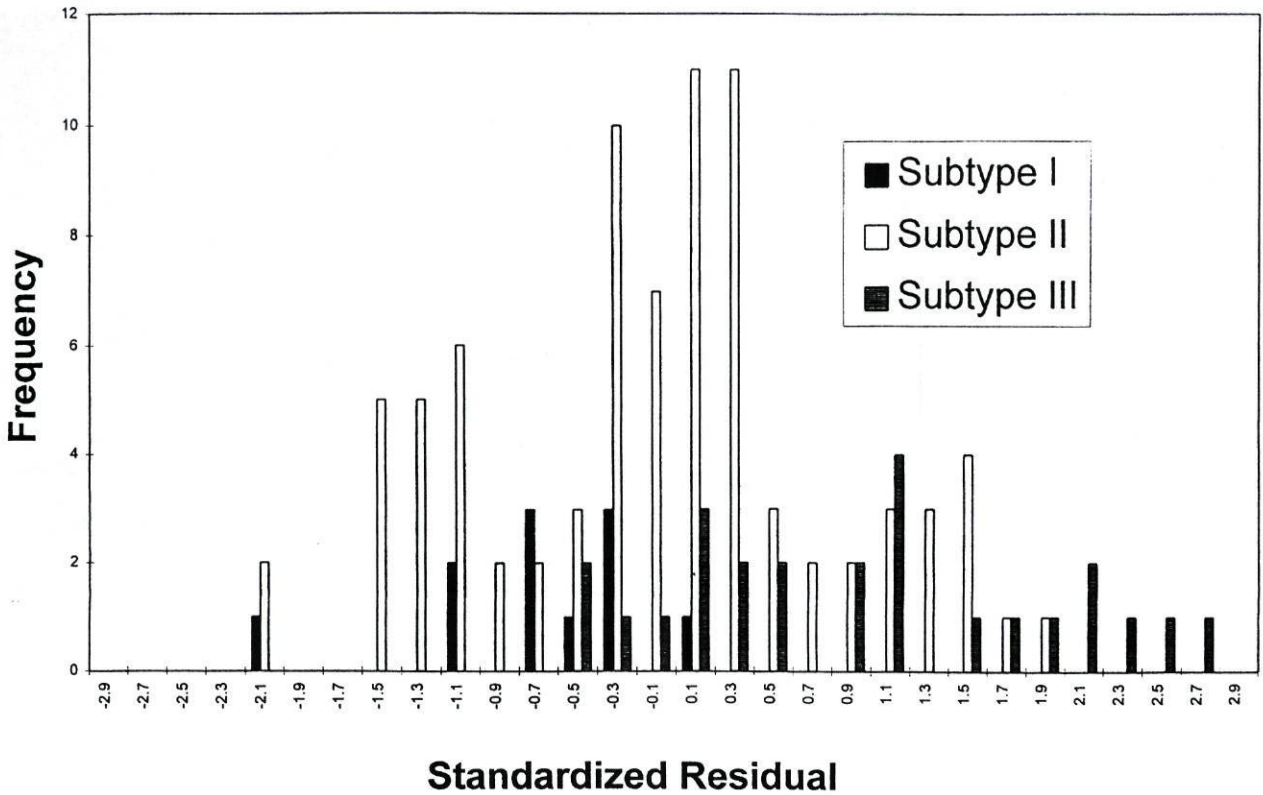


Figure 2 Histogram of standardized residuals from regression of onset age on CAG repeat length by clinical subtype

the covariance parameters is currently not possible using PROC MIXED.

Although the majority of sib pairs in this study have an absolute difference of ≤ 2 CAG repeats, differences of up to 9 units were observed in these sib pairs (Maciel et al. 1995). A moderate correlation for age at onset was still present after adjustment for the number of CAG repeats. Because sibs generally share environmental factors as well as genes, which class of factors is influencing age at onset of MJD cannot be resolved from examining sibs alone.

A positive correlation between parent and offspring residuals is consistent with a genetic factor influencing age at onset for MJD; however, our results revealed a negative correlation. This negative correlation may be an artifact of biased ascertainment. As discussed by Maciel et al. (1995), there is a tendency to sample parents with late onset and offspring with early onset, in studies requiring DNA from both affected parents and offspring. The mean residuals for parents and offspring in the sample studied here were 5.37 and -5.96 , respectively. These residuals, which take into account CAG repeat length, indicate a biased selection of parents with late onset age and offspring with early onset age. Hence, the correlation for residuals among parent-offspring pairs was predictably negative. In contrast, the mean residual for the sibs was .56, which is close to the ex-

pected value of 0 in a sample consisting of an equal number of early- and late-onset individuals.

The positive correlations for adjusted age at onset observed among avuncular and first-cousin pairs are consistent with a genetic component. For a polygenic trait the correlation among avunculars (second-degree relatives) is expected to be twice the correlation among cousins (third-degree relatives) and at least one-half the correlation among sibs (first-degree relatives). The relative magnitudes of the correlations observed for cousins ($r = .28$) and avunculars ($r = .22$) are not consistent with the expectation for a polygenic trait, whereas the relative magnitude of avuncular pairs relative to sibling pairs is consistent with this genetic explanation for the adjusted age at onset. However, the wide 95% CIs for those correlations (table 1) suggest that larger numbers of relative pairs are needed to obtain precise estimates.

Alternate genetic models may explain the familial effect observed for onset age. Commingling of distributions for a quantitative trait may result from a single genetic locus with a major effect on that trait. The three distributions detected in our sample are consistent with a major-gene hypothesis. However, the mixture of distributions may also result from a single environmental factor influencing onset age. Segregation analysis, which would require a larger number of families than available in this study, could be used to determine whether the

transmission of the observed effect was consistent with the expectation for a single major gene.

One potential genetic effect, the number of CAG repeat units on the normal chromosome of MJD individuals, has been found to have no association with age at onset (Maciel et al. 1995; Takiyama et al. 1995). CAG repeat length was treated as a uniform effect in the general population in these studies and was not examined as a sibship-specific effect. Variations within the (CAG)_n sequence or within flanking regions should also be examined for any influence on onset age. If consistent differences in age at onset were observed in individuals with the same CAG repeat length but different alleles in or near the CAG track, it would suggest such variations are contributing to the residual correlations observed. However, if these variations influence the stability of the CAG repeat length during transmission, the effect would be independent of a residual familial factor because CAG repeat length was included in the regression model. Alternatively, a polymorphism at a different chromosomal location may be associated with the familial effect described.

In summary, we demonstrated that age at onset of MJD in a population of families of Portuguese descent is influenced by a familial factor independent of CAG repeat length. Positive correlations after adjustment for CAG repeat length among siblings, first-cousins, and avunculars suggest a factor that may be genetic. The negative correlation between parents and offspring pairs is inconsistent with a genetic effect but may have resulted from selection bias. This data set does not allow precise determination of a genetic model to account for the familial patterns of the age-at-onset residuals. Some of the correlations are inconsistent with a polygenic model. Although the commingling analysis results are consistent with a major gene for age at onset, other explanations, including a shared familial environmental factor and an oligogenic model, are plausible.

The previously reported findings of interfamilial differences for onset age in SCA1 and the association of age at onset of the transmitting parent with onset age in HD suggest that a familial factor may also be influencing age at onset in these CAG repeat diseases. Sex of transmitting parent appears to have a stronger association with onset age in HD than in MJD, which indicates that differences in determination of onset age may exist among these diseases. Identification of genetic and environmental factors influencing age at onset of MJD and other degenerative diseases will have important implications in understanding the pathogenesis and predicting onset age in presymptomatic individuals, and ultimately in preventing onset of symptoms.

Acknowledgments

We would like to thank the families for their participation in this study. This study was supported by National Institutes

of Health grant NS31687, the Fonds de Recherches en Santé du Quebec, the Network of Centers of Excellence (Canadian Genetic Diseases Network), and the Medical Research Council of Canada. C.G., I.S., and P.M. are recipients of a scholarship by the Junta Nacional de Investigação Científica e Tecnológica, Portugal. Some of the results in this paper were obtained by the program package SAGE, which is supported by a U.S. Public Health Service resource grant (1 P41 RR03655) from the Division of Research Resources.

References

- Andrew SE, Goldberg YP, Kremer B, Telenius H, Theilman J, Adam S, Starr E, et al (1993) The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nat Genet* 4:398–403
- Bharucha NE, Bharucha EP, Bhabha, SK (1986) Machado-Joseph disease in India. *Arch Neurol* 43:142–144
- Burke JR, Wingfield MS, Lewis KE, Roses AD, Lee JE, Hulette C, Pericak-Vance MA, Vance JM (1994) The Haw river syndrome: Dentatorubropallidolusian atrophy (DRPLA) in an African-American family. *Nat Genet* 7:521–524
- Burt T, Blumbergs P, Currie B (1993) A dominant hereditary ataxia resembling Machado-Joseph disease in Arnhem Land, Australia. *Neurology* 43:1750–1752
- Cancel G, Abbas N, Stevanin G, Dürr A, Chneiweiss H, Néri C, Duyckaerts C, et al (1995) Marked phenotypic heterogeneity associated with expansion of a CAG repeat sequence at the spinocerebellar ataxia 3/Machado-Joseph disease locus. *Am J Hum Genet* 57:809–816
- Coutinho P, Andrade C (1978) Autosomal dominant system degeneration in Portuguese families of the Azores Islands. *Neurology* 28:703–708
- Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Frontali M, Folstein S, et al (1993) Trinucleotide repeat length instability and age at onset in Huntington's disease. *Nat Genet* 4:387–392
- Healton EB, Brust JCM, Kerr DL, Resor S, Penn A (1980) Presumably Azorean disease in a presumably non-Portuguese family. *Neurology* 30:1084–89
- Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, et al (1994) CAG expansion in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* 8:221–228
- Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolusian atrophy (DRPLA). *Nat Genet* 6:9–13
- Kremer B, Squitieri F, Telenius H, Andrew SE, Theilmann J, Spence N, Goldberg YP, Hayden MR (1993) Molecular analysis of late onset Huntington's disease. *J Med Genet* 30:991–995
- Lima L, Coutinho P (1980) Clinical criteria for diagnosis of Machado-Joseph disease: report of a non-Azorean Portuguese family. *Neurology* 30:319–322
- Livingstone IR, Sequeiros J (1984) Machado-Joseph disease in an American-Italian family. *J Neurogenet* 1:185–186
- MacLean CJ, Morton NE, Lew R (1976) Skewness in commingled distributions. *Biometrics* 32:695–699
- Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P,

- Radvany J, Dawson DM, et al (1995) Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am J Hum Gen* 57:54-61
- Maruyama H, Nakamura S, Matsuyama Z, Sakai T, Doyu M, Sobue G, Seto M, et al (1995) Molecular features of the CAG repeats and clinical manifestations of Machado-Joseph disease. *Hum Mol Genet* 4:807-812
- Nakano KK, Dawson DM, Spence A (1972) Machado disease: a hereditary ataxia in Portuguese emigrants to Massachusetts. *Neurology* 22:49-55
- Orr HT, Chung M, Banfi S, Kwiatkowski Jr. TJ, Servadio A, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 4:221-226
- Ranen NG, Stine OC, Abbott MH, Sherr M, Codori A-M, Franz ML, Chao NI, et al (1995) Anticipation and instability of IT-15 (CAG)_N repeats in parent-offspring pairs with Huntington disease. *Am J Hum Genet* 57:593-602
- Ranum LPW, Chung M-y, Banfi S, Bryer A, Schut LJ, Ramesar R, Duvick LA, 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
- Ranum LPW, Lundgren JK, Schut LJ, Ahrens MJ, Perlman S, Aita J, Bird TD, et al (1995) Spinocerebellar ataxia type I and Machado-Joseph disease: incidence of CAG expansion among adult-onset ataxia patients from 311 families with dominant, recessive, or sporadic ataxia. *Am J Hum Genet* 57:603-608
- Romanul FCA, Fowler HL, Radvany J, Feldman RG, Feingold M (1977) Azorean disease of the nervous system. *N Engl J Med* 26:1505-1508
- Rosenberg RN (1992) Machado-Joseph disease: An autosomal dominant motor system degeneration. *Mov Disord* 7:193-203
- Rosenberg RN, Nyhan WL, Bay C, Shore P (1976) Autosomal dominant striatonigral degeneration. *Neurology* 26:703-714
- Rosner B (1982) On the estimation and testing of interclass correlations: the general case of multiple replicates for each variable. *Am J Epidemiol* 116:722-730
- Rosner B, Donner A, Hennekens CH (1979) Significance testing of interclass correlations from familial data. *Biometrics* 35:461-471
- SAGE (1994) Statistical Analysis for Genetic Epidemiology, Release 2.2 Computer program package available from the Department of Biometry and Genetics, LSU Medical Center, New Orleans
- Sakai T, Ohta M, Ishino H (1983) Joseph disease in a non-Portuguese family. *Neurology* 33:74-80
- SAS Institute Inc (1992) SAS tech rep P-229. SAS/STAT software: changes and enhancements, release 6.07. SAS Institute, Cary, NC, pp 287-368
- Sepulveda RL, Heiba IM, King A, Gonzalez B, Elston RC, and Sorensen RU (1994) Evaluation of tuberculin reactivity in BCG-immunized siblings. *Am J Respir Crit Care Med* 149:620-624
- Sequeiros J (1993) Machado-Joseph disease: epidemiology, genetics and genetic epidemiology. In: Lechtenberg R (ed) *Handbook of cerebellar diseases*. Marcel Dekker, New York, pp 345-351
- Sequeiros J, Coutinho P (1993) Epidemiology and clinical aspects of Machado-Joseph disease, In: Harding A, Deufel T, Chamberlain S (eds) *Advances in neurology*. Raven, New York, pp 139-153
- Sequeiros J, Silveira I, Maciel P, Coutinho P, Manaia A, Gaspar C, Burlet P, et al (1994) Genetic linkage studies of Machado-Joseph disease with chromosome 14q STRPs in 16 Portuguese-azorean kindreds. *Genomics* 21:645-648
- Schöls L, Vieira-Saecker AMM, Schöls S, Przuntek H, Epplen JT, Riess O (1995) Trinucleotide expansion within the MJD1 gene presents clinically as spinocerebellar ataxia and occurs most frequently in German SA patients. *Hum Mol Genet* 4:1001-1005
- Silveira I, Lopes-Cendes I, Kish S, Maciel P, Gaspar C, Coutinho P, Botez MI, et al (1996) Frequency of spinocerebellar ataxia type 1, dentaorubropallidolusian atrophy and Machado-Joseph disease mutations in a large group of spinocerebellar ataxia patients. *Neurology* 46:214-218
- Snedecor GW, Cochran WG (1980) *Statistical methods*, 7th ed. Iowa State University Press, Ames
- St George-Hyslop P, Rogaeva E, Huterer J, Tsuda T, Santos J, Haines JL, Schlumpf K, et al (1994) Machado-Joseph disease in pedigrees of Azorean descent is linked to chromosome 14. *Am J Hum Genet* 55:120-125
- Stevanin G, Cancel G, Dürr A, Chneiweiss H, Dubourg O, Weissenbach J, Cann HM, et al (1995) The gene for spinal cerebellar ataxia (SCA3) is located in a region of ~3 cM on chromosome 14q24.3-q32.2. *Am J Hum Genet* 56:193-201
- Sudarsky L, Corwin L, Dawson DM. (1992) Machado-Joseph disease in New England: Clinical description and distinction from the olivopontocerebellar atrophies. *Mov Disord* 7:204-208
- Takiyama Y, Igarashi S, Rogaeva EA, Endo K, Rogaev EI, Tanaka H, Sherrington R, et al (1995) Evidence for intergenerational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. *Hum Mol Genet* 4:1137-1146
- Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, Shimazaki H, et al (1993) The gene for Machado-Joseph disease is mapped to chromosome 14q. *Nat Genet* 4:300-304
- Twist EC, Causaubon LK, Ruttledge M, Rao VS, MacLeod PM, Radvany J, Zhao A, et al (1995) Machado-Joseph disease maps to the same region of chromosome 14 as the spinocerebellar ataxia type 3 locus. *J Med Genet* 31:823-829
- Woods BT, Schaumburg HH (1972) Nigro-spino-dentatal degeneration with nuclear ophthalmoplegia. *J Neurol Sci* 17:149-166

CAPÍTULO 3

Estudos de mosaïcismo somático do gene MJD1 no sistema nervoso central

Estudos de mosaïcismo somático do gene MJD1 no sistema nervoso central

As alterações patológicas observadas na doença de Machado-Joseph são bastante específicas e restritas a determinadas populações de neurónios. No primeiro trabalho apresentado neste capítulo (artigo IV) procurou-se explorar as bases moleculares desta especificidade de degenerescência neuronal.

Os dados disponíveis até à data relativos à expressão do gene MJD1 sugeriam que a expressão deste gene fosse ubíqua, não permitindo justificar a morte celular limitada a apenas alguns tecidos. O mesmo se observava em todas as outras doenças associadas à expansão de uma sequência poliglutamina, nomeadamente na doença de Huntington.

A instabilidade mitótica do *repeat* CAG expandido leva a que diferentes tecidos apresentem diferentes tamanhos do *repeat* expandido. Este fenómeno - mosaïcismo somático - pode ser observado em todas as doenças associadas a poliglutaminas expandidas. Foi assim sugerido que o mosaïcismo somático no sistema nervoso central pudesse ser a base da selectividade de morte celular, sendo as células mais afectadas as que fossem portadoras de maiores *repeats* expandidos. A fim de testar esta hipótese, determinaram-se os tamanhos dos *repeats* CAG expandidos em 20 regiões do sistema nervoso central (SNC) de doentes DMJ e SCA1, que foram, paralelamente, estudadas em detalhe do ponto de vista patológico. Observou-se, em geral, uma grande homogeneidade em termos do tamanho médio dos *repeats* e do grau de mosaïcismo presentes nas diferentes regiões do SNC estudadas, com excepção do córtex cerebeloso, que apresentava, em todos os casos, *repeats* mais pequenos e um menor grau de mosaïcismo somático. Este padrão de mosaïcismo era, contudo, semelhante em *specimens* de doentes DMJ e SCA1, apesar das suas diferentes manifestações patológicas, em particular a nível do cerebelo, demonstrando-se assim que não existe uma correlação entre o grau de degenerescência observado em cada região e o tamanho dos *repeats* detectados. O padrão

de mosaïcismo parece antes reflectir a composiço celular dos diferentes tecidos e, possivelmente, a sua histria em termos de desenvolvimento, embrionrio e ps-natal.

Uma possvel explicaço para a observaço de que o crtex cerebeloso apresenta *repeats* menores e menor grau de mosaïcismo, no so na DMJ e na SCA1, mas tambm em outras doenças associadas a poliglutaminas expandidas (a doença de Huntington, a ADRPL e a ataxia de Friedreich), seria a menor proporço de clulas gliais nesta regio, particularmente rica em neurnios. As clulas que mantm potencial mittico durante a vida do indivduo, como as clulas gliais, poderiam contribuir maioritariamente para o mosaïcismo observado no SNC de indivduos portadores de uma expanso, que tenderia a aumentar ao longo da vida, enquanto em clulas ps-mitticas, tais como os neurnios, os *repeats* se manteriam mais estveis.

No estudo apresentado no artigo V, determinou-se o grau de mosaïcismo somtico do *repeat* expandido em *specimens* de crtex e de substncia branca cerebral de doentes DMJ e SCA1, a fim de se comparar o grau de mosaïcismo somtico nestas estruturas, e de se estudar uma possvel correlaço entre a idade de morte dos indivduos e o grau de mosaïcismo somtico observado. Em contraste com observaçes prvias em doentes com ADRPL [Takano *et al.*, 1996], no se observou uma correlaço significativa entre a idade de morte e o grau de mosaïcismo observado no SNC dos doentes com DMJ ou SCA1. A diferenç entre o grau de mosaïcismo observado no crtex e na substncia branca tambm no era significativa em nenhum dos casos. Estes resultados sugerem que o comportamento dos *repeats* CAG expandidos, em termos de mosaïcismo somtico, pode diferir entre estas duas doenças e a ADRPL. No entanto, no se exclui completamente a possibilidade de as clulas gliais serem os principais contribuintes para o mosaïcismo somtico no SNC: estudos mais detalhados, recorrendo a microdissecço e *single-cell PCR*, podero permitir esclarecer definitivamente esta questo.

ARTIGO IV

Somatic Mosaicism in the Central Nervous System in Spinocerebellar Ataxia Type 1 and Machado-Joseph Disease

Iscia Lopes-Cendes,^{*†} Patrícia Maciel,^{*†‡} Stephen Kish,[‡] Claudia Gaspar,^{*†} Yves Robitaille,[§] H. Brent Clark,[‡] Arnulf H. Koeppen,^{**} Martha Nance,^{**} Lawrence Schut,^{††} Isabel Silveira,^{*†} Paula Coutinho,^{‡‡} Jorge Sequeiros,[†] and Guy A. Rouleau, MD, PhD*

Spinocerebellar ataxia type 1 and Machado-Joseph disease are two autosomal dominant cerebellar ataxias caused by expansions of unstable CAG repeats in the coding region of the causative genes. The selectivity of cell death and the resulting characteristic neuropathological features in each of these diseases are not explained by the gene expression patterns. Since the repeat size correlates with age at onset and severity of these diseases, somatic mosaicism, the result of mitotic instability of the CAG repeat, could be the basis for specificity of neurodegeneration; brain structures with larger expanded repeats would be more severely affected. To study the association between neuropathological changes and somatic mosaicism of the CAG repeat size in the central nervous system of patients with these two ataxias, we determined the size of the (CAG)_n expansion in 20 different regions of the brain, brainstem, cerebellum, and spinal cord from 3 patients with spinocerebellar ataxia type 1 and 3 with Machado-Joseph disease; these regions were selected for their differential neuropathological involvement in the two disorders. We observed a considerable homogeneity of repeat size ranges in all but 1 of the 20 regions examined: The cerebellar cortex showed slightly smaller (CAG)_n tracts in all specimens from both groups of patients. Our results suggest that the pattern of repeat size mosaicism, similar in spinocerebellar ataxia type 1 and Machado-Joseph disease, reflects the developmental pathways and cell composition of different central nervous system regions and is not the cause of selective cell death in these disorders.

Lopes-Cendes I, Maciel P, Kish S, Gaspar C, Robitaille Y, Clark HB, Koeppen AH, Nance M, Schut L, Silveira I, Coutinho P, Sequeiros J, Rouleau GA. Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado-Joseph disease. *Ann Neurol* 1996;40:199–206

Spinocerebellar ataxia type 1 (SCA1) and Machado-Joseph disease (MJD) are autosomal dominant spinocerebellar ataxias with onset in adult life [1, 2]. Both disorders are characterized clinically by progressive cerebellar ataxia with associated features including ophthalmoplegia and pyramidal signs. Autopsy studies in SCA1 and MJD patients showed symmetrical degeneration of circumscribed groups of neurons, which differ in each of these disorders [2–8]. These two types of spinocerebellar ataxia are associated with expansion of a (CAG)_n tract in the coding region of the causative genes [9, 10], both of unknown function and without homology to other known genes.

Some features are shared by these and other progressive neurodegenerative disorders of adult onset associated with (CAG)_n tract expansion—Kennedy's disease or spinobulbar muscular atrophy (SBMA) [11], Huntington's disease (HD) [12], and dentatorubropallidolusian atrophy (DRPLA) [13, 14]—suggesting a common pathogenic mechanism. The autosomal dominant transmission, together with evidence for the actual translation of the mutant protein in HD [15, 16], SCA1 [17], and DRPLA [18] and the transgenic mouse model of SCA1 [19], suggests a toxic gain of function by the proteins containing an expanded glutamine tract. In addition, the observed inverse correla-

From the *Centre for Research in Neuroscience, McGill University and the Montreal General Hospital Research Institute, Montreal, Québec; †Lab. Genética Médica, ICBAS, and UnIGENE, IBMC, Universidade do Porto, Porto; ‡Human Neurochemical Pathology Laboratory, Clarke Institute of Psychiatry, Toronto, Ontario; §Department of Neuropathology, University of Montreal, Montreal, Québec, Canada; †Department of Laboratory Medicine and Pathology, University of Minneapolis, Minneapolis; **Neurology Service (127), Veterans Administration Medical Center, Albany, NY; ††Department of Neurology, University of Minnesota, Minneapolis, MN; and ‡‡Serviço de Neurologia, Hospital de Sto. António, Porto, Portugal. †I.L.-C. and P.M. contributed equally to this paper.

Received Jan 18, 1996, and in revised form Feb 22. Accepted for publication Mar 4, 1996.

Address correspondence to Dr Rouleau, Centre for Research in Neuroscience, McGill University, Montreal General Hospital Research Institute, 1650 Cedar Avenue, Montreal, Québec, Canada H3G1A4.

tion between repeat size and age at onset of disease [14, 20–27] indicates that this toxic effect is proportional to the length of the polyglutamine tract. Several interesting hypotheses have been advanced [28], but the mechanism of pathogenesis of the abnormal polyglutamine stretch is still not understood.

Even if one assumes the existence of a common pathogenic mechanism for the expanded polyglutamine tract in all these CAG-related disorders, the explanation for the very selective patterns of neuronal loss remains unknown. Expression studies of the Huntington, DRPLA, and SCA1 genes in humans showed widespread expression, at the messenger RNA (mRNA) [12, 13, 29–32] and protein levels [17, 18, 33], that do not explain the differential neuropathological findings characterizing each of these diseases.

Individuals affected with HD [34–36], DRPLA [37, 38], or SCA1 [39] can have different sizes of the expanded CAG repeat in different cells, as the result of mitotic instability of the repetitive sequence during embryogenesis. Since the age at onset and severity of disease presentation in patients are correlated with repeat size, one hypothesis is that somatic mosaicism is the basis for specific cell death, with cells containing larger (CAG) n tracts being more severely affected by the toxic effect of the expanded polyglutamine. This mechanism was previously suggested for HD [34].

To investigate whether any association exists between the size of the CAG repeat and neuropathological involvement in SCA1 and MJD, we determined the length of the (CAG) n tract in 20 different central nervous system (CNS) regions of 6 patients, 3 with SCA1 and 3 with MJD.

Materials and Methods

Pathological Material

At autopsy, brains and spinal cord were obtained from 3 patients with end-stage SCA1 and 3 with MJD; the 3 SCA1 patients and 1 MJD patient were members of previously published pedigrees [40, 41]. All patients were wheelchair bound or bedridden and died of bronchopneumonia. At autopsy the brains were split into two halves. One half-brain was frozen and stored at -80°C , whereas the other half-brain was fixed in 10% buffered formalin. Representative cortical and subcortical sections were embedded with paraffin and stained with hematoxylin and eosin, Klüver-Barrera, modified Bielschowsky, and Holzer preparations. Peripheral blood samples were not available for any of the 6 patients.

DNA Analysis

Twenty different regions of interest were isolated from the brain, brainstem, cerebellum, and spinal cord of the 6 patients (Table). The regions studied were selected based on their differential pathological involvement in SCA1 and MJD. Fifty to 100 mg of tissue was obtained from each region by careful dissection. Genomic DNA was extracted

from the autopsy material following standard techniques [42].

The (CAG) n -containing fragments of the SCA1 and MJD genes were amplified by polymerase chain reaction (PCR), using the previously published primers Rep1 and Rep2 for SCA1 [9] and MJD52 and MJD25a (a slight modification of MJD25, of sequence ATCCATGTGCAAAGG CCAGCC) for MJD [10], using PCR conditions described elsewhere [10, 43]. PCR products were separated on 6% polyacrylamide gels and, for determination of allele sizes, an M13 sequencing ladder (-40 primer) was run in parallel and used for comparison. For MJD, ^{35}S -deoxyadenosine triphosphate (dATP) labeling was used, whereas for SCA1 the nonlabeled PCR products were transferred from the gel into Hybond N+ membranes and hybridized with an α - ^{32}P 3'-end-labeled (CAG) $_{15}$ probe. Detection was done by autoradiography in both cases. For the SCA1 gene, normal alleles range in size from 19 to 37 CAG repeats, while affected alleles range in size from 42 to 81 [20, 21]. The MJD gene contains 12 to 37 CAG repeats in normal individuals, and 62 to 84 in patients [23].

Results

Neuropathology

Pathological specimens from all 6 patients showed severe neuronal loss. The 3 patients with SCA1 had abnormalities characteristic of end-stage olivopontocerebellar atrophy (OPCA) with severe neuronal loss in the inferior olives, pons, and Purkinje cell layer of the cerebellar cortex. By contrast, the 3 patients with MJD had no cell loss in the inferior olives, only mild to moderate neuronal loss in the cerebellar cortex, and moderate to severe cell loss in the substantia nigra and basal pons. The Table summarizes the pathological abnormalities for each of the 6 patients studied.

Molecular Analysis of the (CAG) n Repeat

The normal allele had the same size in all CNS regions studied for the 3 SCA1 and the 3 MJD patients, with a single strong band distinctively seen on the autoradiographs (Fig). The expanded allele showed numerous bands, with a larger number of bands present in specimens from SCA1 patients. This was seen even when SCA1 and MJD samples were electrophoresed in parallel and the same method of detection (^{32}P -dATP labeling) was used for both. For each patient, the size range of the expanded allele was homogeneous in all different regions, except for the cerebellar cortex where the expanded repeat was smaller than the strongest band for the other areas studied, in both SCA1 and MJD. In addition, the number of bands observed for the expanded allele in the cerebellar cortex was smaller than that in all other regions (more obviously so in the SCA1 specimens), indicating that less somatic mosaicism is present in the cerebellar cortex. Other apparent differences in the number of bands between the various regions observed in any one specific experiment were

*Neuropathological Abnormalities Found in CNS Specimens of Patients with Spinocerebellar Ataxia Type 1 (SCA1) and Machado-Joseph Disease (MJD)**

Anatomical Regions Studied	SCA1 A	SCA1 B	SCA1 C	MJD D	MJD E	MJD F
1. Area 9—frontal cortex	+	—	—	—	—	—
2. Area 22—temporal cortex	+	—	—	—	—	—
3. Area 7a—parietal cortex	+	—	—	—	—	—
4. Area 18—occipital cortex	—	—	—	—	—	—
5. Caudate nucleus	—	—	—	—	NE	—
6. Putamen	—	—	—	—	NE	—
7. Hippocampus	+	—	—	+	—	—
8. Medial dorsal nucleus of the thalamus	—	—	—	—	—	—
9. Nucleus lateralis of the thalamus	—	—	—	—	—	—
10. Anterior dorsal nucleus of the thalamus	—	—	NE	—	—	—
11. Centromedian nucleus of the thalamus	—	—	—	—	—	—
12. Red nucleus	++	++	++	—	—	++
13. Internal globus pallidus	—	—	—	+	NE	+
14. External globus pallidus	—	—	—	+	NE	++
15. Area 23—cingulate cortex	NE	NE	NE	NE	—	NE
16. Cerebellar cortex	+++	+++	+++	+ / +++	+	+
17. Medial segment of the pulvinar nucleus of the thalamus	+	—	+	—	—	NE
18. White matter	—	—	—	—	—	—
19. Pons	+++	+++	+++	+++	+++	++
20. Cervical segment spinal cord	+++	+++	+++	+++	+++	++

Atrophy was classified into four categories: — = no atrophy; + = mild atrophy; ++ = moderate atrophy; +++ = severe atrophy. NE = not examined.

*Letters A–F next to disease designations (SCA1, MJD) in column headings correspond to patients represented in Fig A–F.

not reproducible, probably resulting from differences in the amount of DNA present or the efficiency of each specific PCR. The difference between the cerebellum and the remaining areas, regarding both the size of the expanded repeats and the amount of mosaicism, was consistent in at least four experiments for all 6 patients analyzed.

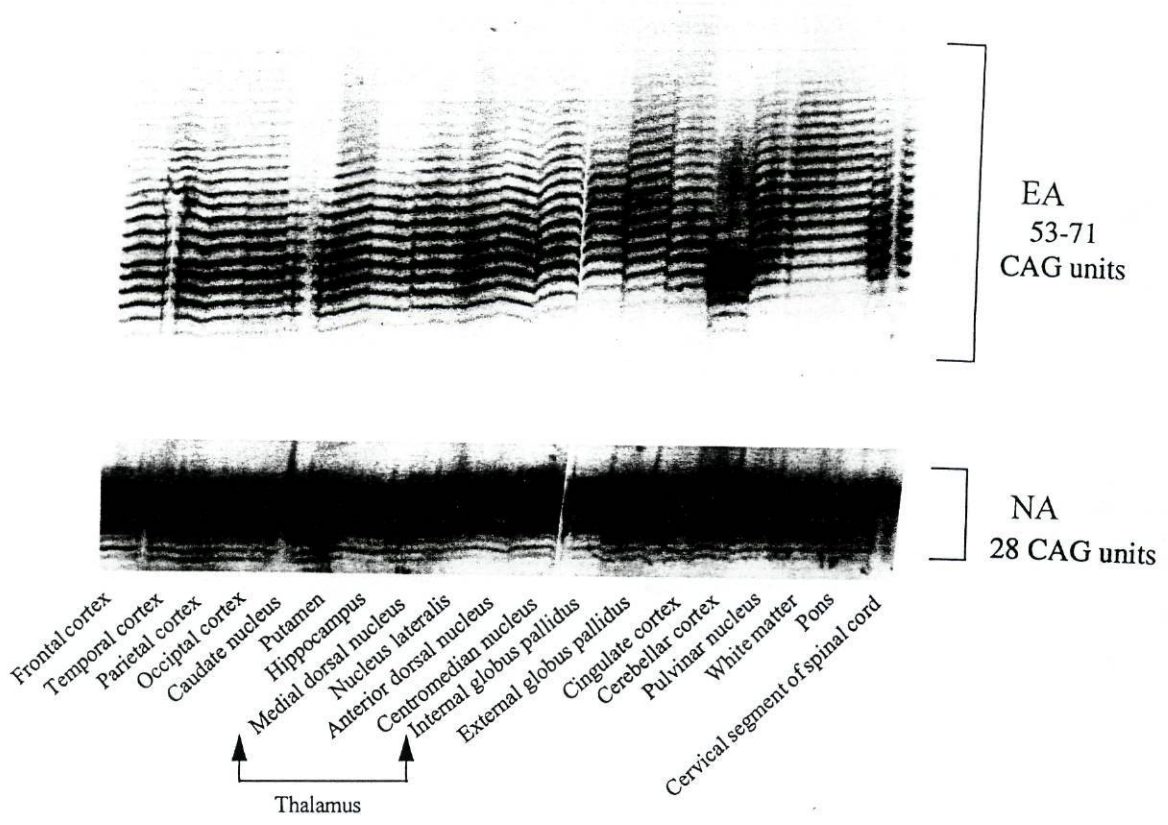
Discussion

Somatic mosaicism of the expanded CAG repeat was first found in HD [34–36], which initially suggested that the size of the (CAG)_n in different CNS regions correlated with pathological involvement, because a smaller (CAG)_n was observed in the cerebellar cortex, which is not typically involved in HD [34]. Other studies, however, did not detect significant differences in size between different brain regions in HD patients [44, 45]. Somatic mosaicism of the (CAG)_n tract was also observed in DRPLA and SCA1, with differences found between CNS and peripheral tissues, but no apparent association with neuropathological features [38, 39]. However, the SCA1 study was based on a limited number of specimens.

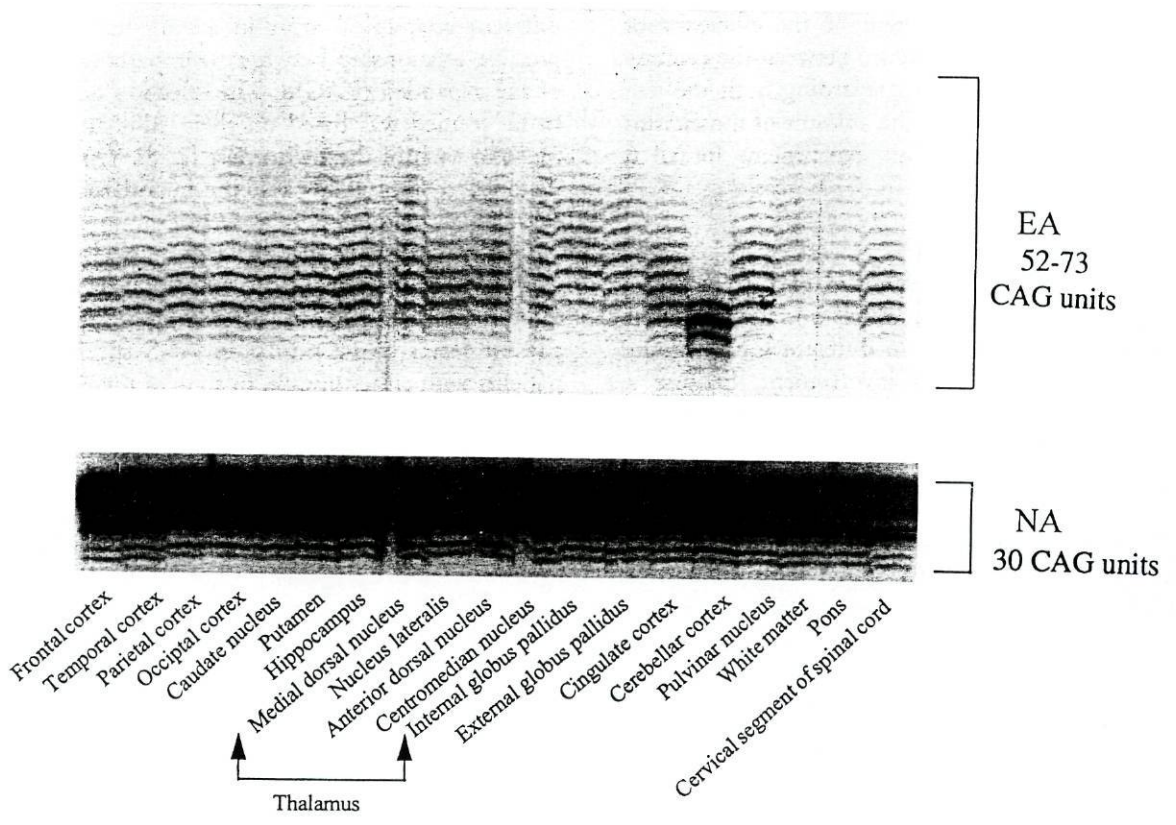
We investigated somatic mosaicism within the CNS in further detail in two autosomal dominant spinocerebellar ataxias for which the causative genes are known—SCA1 and MJD. Their similar phenotypes, with nonetheless distinctive clinical and pathological features, make these two disorders particularly interest-

ing for this type of comparative study. We chose 20 different anatomical areas in an attempt to detect a possible relationship between neuropathology and size of the expanded (CAG)_n. The selection of CNS areas to be studied was based on their differential involvement in each of these disorders [2–8]. Our study was designed to thoroughly test the hypothesis, as a large number of anatomical areas were examined.

Our results show that there is mosaicism in terms of the size of the expanded CAG repeat in the CNS of SCA1 and MJD patients. Interestingly, the mosaicism pattern seems to be similar in CNS specimens from patients with either disease, in spite of the distinct neuropathological abnormalities seen in SCA1 and MJD [2–8]. All regions studied were homogeneous in terms of repeat size range, except for the cerebellar cortex, which showed smaller (CAG)_n tracts in both diseases. The cerebellar cortex is one of the main structures degenerated in SCA1 patients, but it is usually preserved or only mildly affected in MJD. If mosaicism of the CAG repeat was the basis for the selectivity of neurodegeneration, one would expect the expanded (CAG)_n to be larger in the cerebellar cortex of patients with SCA1, but not in MJD patients. Therefore, the present findings exclude any association between repeat size and specificity of neurodegeneration in SCA1 and MJD. Other cell-specific factors must be responsible for the selectivity of neurotoxicity of the expanded polyglutamine tract in these diseases.

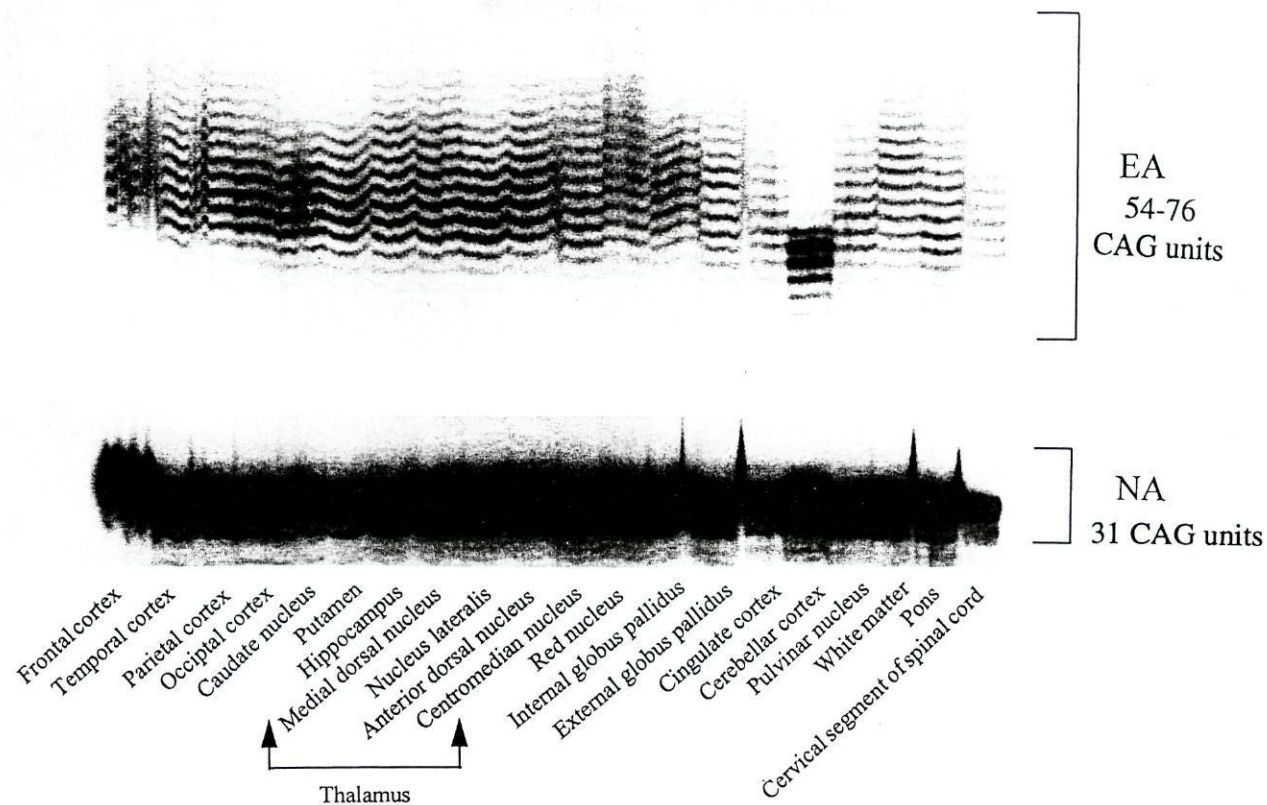


A

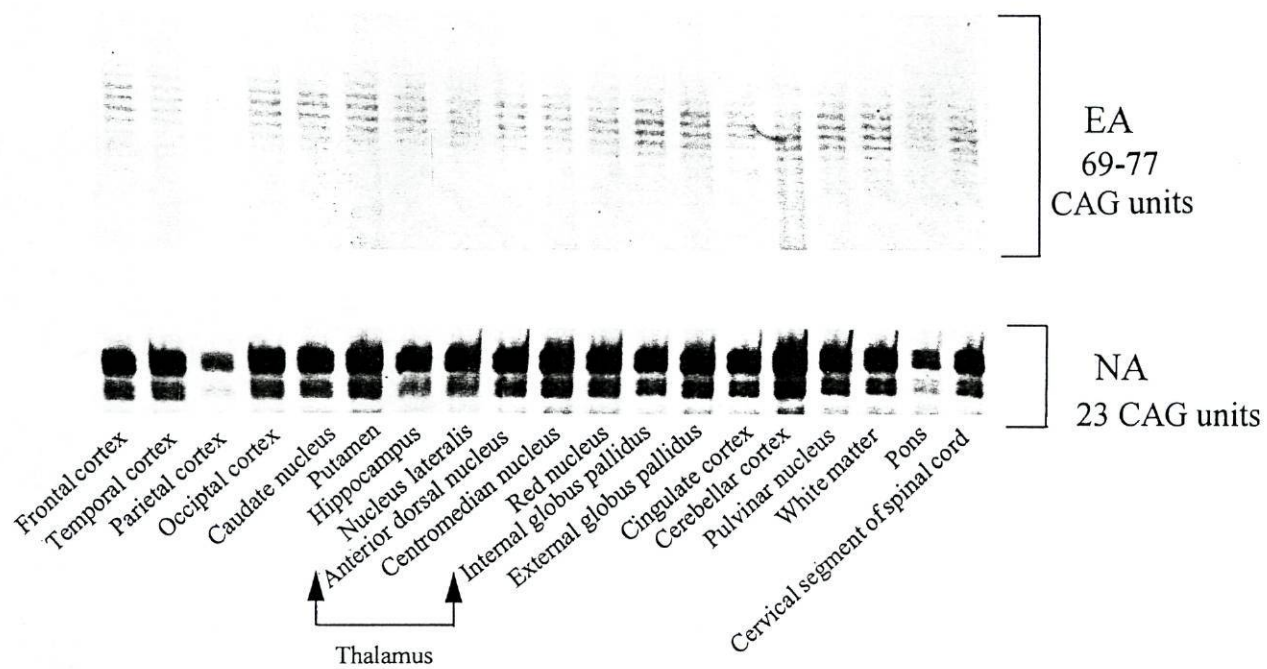


B

Polymerase chain reaction analysis of the CAG repeat at the spinocerebellar ataxia type 1 (SCA1) (A–C) and the Machado-Joseph disease (MJD) (D–F) loci in different regions of the CNS of 3 SCA1 and 3 MJD patients. The sizes of the normal (NA) and expanded alleles (EA) are indicated on the right. Note that for 2 SCA1 patients (A, B) the samples corresponding to the red nucleus, and for 1 MJD patient (D) the medial dorsal nucleus of the thalamus were not tested.

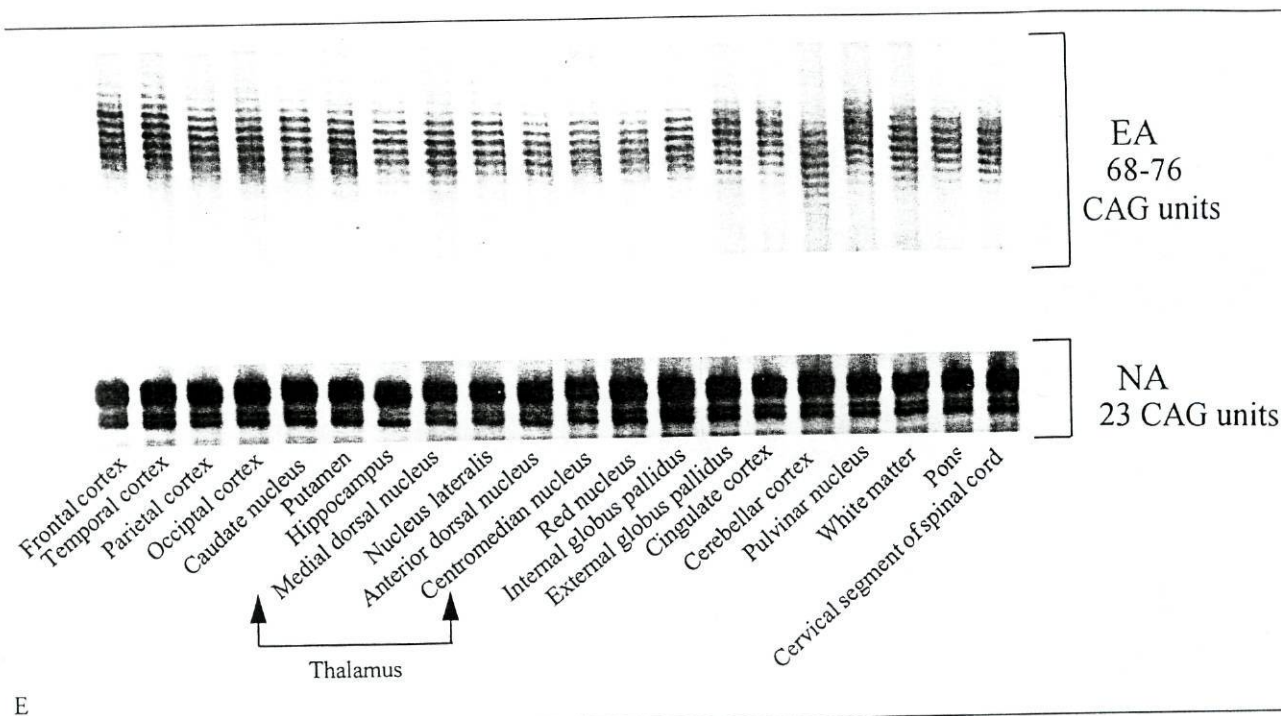


C

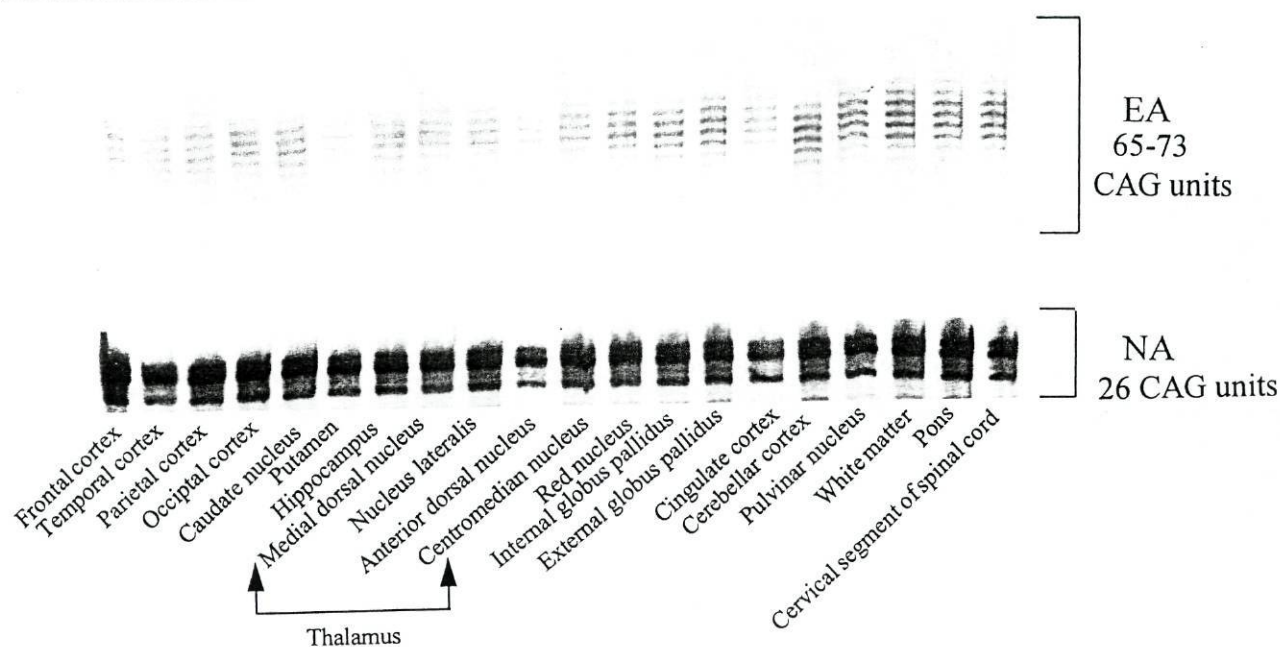


D

Fig (continued).



E



F

Fig (continued).

Recent reports for HD [34], DRPLA [38], and SCA1 [39] also showed that the cerebellum has a smaller (CAG)_n compared to other regions studied. It is possible that the mosaicism pattern is related to specific cell division rates during embryonic development of the different cell types. It has been suggested that the cerebellum, less rich in glia and other types of non-neuronal cells compared to other structures of the brain (therefore less rich in dividing cells), might undergo

less repeat expansion [39]. In agreement with this hypothesis are the results of a recent study of brain mosaicism in DRPLA, which showed that the number of bands observed for the expanded repeat within brain regions other than cerebellum increased with the age of the patient at the time of death [46]. This suggests that the dividing cells contribute largely to the somatic mosaicism, which increases during life due to continuing mitosis. Another possible explanation for the obser-

vation of a smaller expanded repeat in the cerebellar cortex in HD, DRPLA, SCA1, and MJD patients is that the cells in the cerebellar cortex may be particularly susceptible during development to the toxic effect of the expanded CAG repeat, and that a selection bias would eliminate the cells with larger repeats.

The fact that differences in expanded (CAG)_n size between the cerebellum and the rest of the CNS areas were not very large in the 6 brains studied could reflect the initial size of the expanded repeat inherited by each individual, with larger repeats being more unstable during mitotic cell divisions. In previous studies of HD brain specimens, the mosaicism was more pronounced in patients with juvenile-onset disease, carriers of a larger expanded allele [34]. All our SCA1 and MJD patients had what can be considered medium-sized expanded alleles in the CNS, and none had juvenile-onset disease.

There are some limitations to the type of study we performed, the most important one being the size of the tissue sample collected in each region: On the one hand, some of the structures degenerated in SCA1 or MJD (such as the dentate nucleus, inferior olives and substantia nigra) are often so small that it is difficult to dissect them apart from surrounding structures; on the other hand, the resolution of dissection may be insufficient to detect differences in repeat size that could exist between different types of cells. In fact, the observation of a large number of bands for the expanded allele in the samples studied suggests that there is mosaicism of the repeat size within each anatomical region sampled, with differences between individual cells or between different types of cells, or both. Another possible limitation of this kind of study is that even though to test the correlation between repeat size and specificity of neurodegeneration, the cells of greatest interest would be the ones that have undergone degeneration, the repeat size range determined for each region can only be the one corresponding to the surviving tissue. The only way to overcome this difficulty would be to study CNS specimens of carriers of the expanded repeat before the onset of neurodegeneration, who died from other causes. This type of material was not available to us.

In conclusion, our results suggest that although somatic mosaicism of (CAG)_n length is present in the CNS of SCA1 and MJD patients, the differences in repeat size between cells are the result of a cell-specific pattern of mitotic division during development, and are not the basis for the differential neuropathological involvement, and therefore the specificity of phenotype, in these two disorders.

Association Canadienne de l'Ataxie de Friedreich, the Medical Research Council of Canada, the Network of Centres of Excellence (Canadian Genetic Diseases Network), and a European Community Program STRIDE grant. Dr Kish is supported by National Institute of Neurological Disorders and Stroke grant 20634. Drs Maciel, Gaspar, and Silveira are the recipients of scholarships by J.N.I.C.T. (PRAXIS XXI), Portugal.

References

1. Harding AE. Idiopathic late-onset cerebellar ataxia. In: Harding, AE, ed. The hereditary ataxias and related disorders. London: Churchill-Livingstone, 1984:166-172
2. Coutinho P, Andrade C. Autosomal dominant system degeneration in Portuguese families of the Azores Islands. *Neurology* 1978;28:703-709
3. Schut JW, Haymaker W. Hereditary ataxia: a pathological study of five cases of common ancestry. *J Pathol Clin Neurol* 1951;1:183-213
4. Robitaille Y, Schut L, Kish SJ. Structural and immunocytochemical features of olivopontocerebellar atrophy caused by the SCA1 mutation define a unique phenotype. *Acta Neurol Scand* (in press)
5. Rosenberg RN, Nyhan WL, Bay C, Shore P. Autosomal dominant striatonigral degeneration. *Neurology* 1976;26:703-714
6. Coutinho P, Guimarães A, Scaravilli F. The pathology of Machado-Joseph disease: report of a possible homozygous case. *Acta Neuropathol (Berl)* 1982;58:48-54
7. Sakai T, Ohta M, Ishino H. Joseph disease in a non-Portuguese family. *Neurology* 1983;33:74-80
8. Kanda T, Iozaki E, Kato S, et al. Type III Machado-Joseph disease in a Japanese family: a clinicopathological study with special reference to the peripheral nervous system. *Clin Neuropathol* 1989;8:134-141
9. Orr HT, Chung M-Y, Banfi S, et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genet* 1993;4:221-226
10. Kawaguchi Y, Okamoto T, Taniwaki M, et al. CAG expansions in a novel gene from Machado-Joseph disease at chromosome 14q32.1. *Nature Genet* 1994;8:221-227
11. La Spada AR, Wilson EM, Lubahn DB, et al. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 1991;352:77-79
12. The Huntington's Disease Collaborative Research Group. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 1993;72:971-983
13. Nagafuchi S, Yanagisawa H, Ohsaki E, et al. Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidolysian atrophy (DRPLA). *Nature Genet* 1994;8:177-182
14. Koide R, Ikeuchi T, Onodera O, et al. Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolysian atrophy (DRPLA). *Nature Genet* 1994;6:9-13
15. Jou Y-S, Myers RM. Evidence from antibody studies that the CAG repeat in the Huntington disease gene is expressed in the protein. *Hum Molec Genet* 1995;4:465-469
16. Trotter Y, Devys D, Imbert G, et al. Cellular localization of the Huntington's disease protein and discrimination of the normal and mutated form. *Nature Genet* 1995;10:104-110
17. Servadio A, Koshy B, Armstrong D, et al. Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. *Nature Genet* 1995;10:94-98
18. Yazawa I, Nukina N, Hashida H, et al. Abnormal gene product identified in hereditary dentatorubral-pallidolysian atrophy (DRPLA) brain. *Nature Genet* 1995;10:99-103
19. Burreight EN, Clark HB, Servadio A, et al. SCA1 transgenic

This project was supported by a National Institute of Health grant (NS 31687), the Fonds de Recherche en Santé du Québec, the

- mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. *Cell* 1995;82:937-948
20. Jodice C, Malaspina P, Persichetti F, et al. Effect of trinucleotide repeat length and parental sex on phenotypic variation in spinocerebellar ataxia 1. *Am J Hum Genet* 1994;54:959-965
 21. Ranum LPW, Chung M, Banfi S, et al. Molecular and clinical correlations in spinocerebellar ataxia type 1: evidence for familial effects on the age at onset. *Am J Hum Genet* 1994;55:244-252
 22. Maruyama H, Nakamura S, Matsuyama Z, et al. Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Hum Molec Genet* 1995;4:807-812
 23. Maciel P, Gaspar C, DeStefano AL, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am J Hum Genet* 1995;57:54-61
 24. La Spada AR, Roling DB, Harding AE, et al. Meiotic stability and genotype-phenotype correlation of the trinucleotide repeat in X-linked spinal and bulbar muscular atrophy. *Nature Genet* 1992;2:301-304
 25. Andrew SE, Goldberg YP, Kremer B, et al. The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature Genet* 1993;4:398-403
 26. Duyao M, Ambrose C, Myers R, et al. Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genet* 1993;4:387-392
 27. Snell RG, MacMillan JC, Cheadle JP, et al. Relationship between trinucleotide repeat expansion and phenotypic variation in Huntington's disease. *Nature Genet* 1993;4:393-397
 28. Sutherland G, Richards RI. Simple tandem DNA repeats and human genetic disease. *Proc Natl Acad Sci USA* 1995;92:3636-3641
 29. Strong T, Tagle DA, Valdes JM, et al. Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues. *Nature Genet* 1993;5:259-265
 30. Landwehrmeyer GB, McNeil SM, Dure LS, et al. Huntington's disease gene: regional and cellular expression in brain of normal and affected individuals. *Ann Neurol* 1995;37:218-230
 31. Stine OC, Li S-H, Pleasant N, et al. Expression of the mutant allele of IT-15 (the HD gene) in striatum and cortex of Huntington's disease patients. *Hum Molec Genet* 1995;4:15-18
 32. Banfi S, Servadio A, Chung M-Y, et al. Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nature Genet* 1994;7:513-519
 33. Sharp AH, Loev SJ, Schilling G, et al. Widespread expression of Huntington's disease gene (IT15) protein product. *Neuron* 1995;14:1065-1074
 34. Telenius H, Kremer B, Goldberg YP, et al. Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nature Genet* 1994;6:409-414
 35. Schaefer FV, Calikoglu AS, Whetsell LH. Characterization of conservative somatic instability of the CAG repeat region in Huntington's disease. *Am J Hum Genet* 1994;55:A241
 36. Stine OC, Pleasant N, Wagster MV, et al. Somatic instability of the expanded allele of IT-15 from patients with Huntington disease. *Am J Hum Genet* 1994;55:A244
 37. Aoki M, Abe K, Kameya T, et al. Maternal anticipation of DRPLA and somatic heterogeneity of CAG repeats. *Am J Hum Genet* 1994;55:A210
 38. Ueno S, Kondoh K, Kotani Y, et al. Somatic mosaicism of CAG repeat in dentatorubral-pallidoluysian atrophy. *Hum Molec Genet* 1995;4:663-666
 39. Chong SS, McCall AE, Cora J, et al. Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nature Genet* 1995;10:344-349
 40. Schut JW. Hereditary ataxia: clinical study through six generations. *Arch Neurol Psychiatry* 1950;63:535-568
 41. Ranum LPW, Rich SS, Nance MA, et al. Autosomal dominant spinocerebellar ataxia: locus heterogeneity in a Nebraska kindred. *Neurology* 1992;42:344-347
 42. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: a laboratory manual*. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989
 43. Silveira I, Lopes-Cendes I, Kish S, et al. Frequency of spinocerebellar ataxia type 1, dentatorubropallidoluysian atrophy and Machado-Joseph disease mutations in the large group of spinocerebellar ataxia patients. *Neurology* 1996;46:214-218
 44. MacDonald ME, Barnes G, Srinidhi J, et al. Gametic but not somatic instability of CAG repeat length in Huntington's disease. *J Med Genet* 1993;30:982-986
 45. Zuhlke C, Riess O, Bockel B, et al. Mitotic stability and meiotic variability of the (CAG)_n repeat in the Huntington disease gene. *Hum Molec Genet* 1993;2:2063-2067
 46. Takano H, Onodera O, Takahashi H, et al. Somatic mosaicism of the CAG expansion in the central nervous system as a reflection of progressive neuronal cell loss in dentatorubral-pallidoluysian atrophy (DRPLA). *Am J Hum Genet* 1995;57:A9

ARTIGO V

Am. J. Hum. Genet. 60:993-996, 1997

Mosaicism of the CAG Repeat in CNS Tissue in Relation to Age at Death in Spinocerebellar Ataxia Type 1 and Machado-Joseph Disease Patients

To the Editor:

Machado-Joseph disease (MJD) and spinocerebellar ataxia type 1 (SCA1) are two autosomal dominant spinocerebellar ataxias caused by the expansion of a polyglutamine tract in the *MJD1* (Kawaguchi et al. 1994; Trotter et al. 1995) and *ataxin-1* (Chung et al. 1993; Banfi et al. 1994; Servadio et al. 1995) gene products, respectively. This specific type of dynamic mutation has been associated with other neurodegenerative diseases, namely Kennedy disease (La Spada et al. 1991), Huntington disease (HD) (Huntington's Disease Collaborative Research Group 1993), dentatorubropallidolysian atrophy (DRPLA) (Koide et al. 1994), and spinocerebellar ataxia type 2 (Imbert et al. 1996; Pulst et al. 1996; Sanpei et al. 1996), for which the mutated genes have been identified, and spinocerebellar ataxia type 7 (Trotter et al. 1995), where the causative gene is still unknown.

Instability of the expanded polyglutamine-coding (CAG)_n tracts during cell division leads to changes in repeat length, either contractions or expansions, resulting in certain populations of cells carrying different repeat lengths. Somatic mosaicism of the (CAG)_n tract has been demonstrated in MJD (Lopes-Cendes et al. 1996) and SCA1 (Chong et al. 1995) as well as in HD (Telenius et al. 1994) and DRPLA (Ueno et al. 1995). It is interesting that the pattern of repeat size distribution in CNS is similar in MJD and SCA1 patients (Lopes-Cendes et al. 1996) as well as in the other polyglutamine-related diseases (Telenius et al. 1994; Ueno et al. 1995). Therefore, somatic mosaicism seems to be the result of differences in the pattern of development and/or cell composition of each

structure in CNS and does not contribute directly to the specific pattern of cell death.

Given that in HD, DRPLA, SCA1 and MJD patients cerebellar cortex has the smallest repeat size and least mosaicism of all CNS regions studied (Telenius et al. 1994; Chong et al. 1995; Ueno et al. 1995; Lopes-Cendes et al. 1996), and because the ratio of neuronal to glial cells in cerebellar cortex is higher than elsewhere, it was suggested that the main source of expanded repeat instability in the CNS would be the glia that continue dividing, whereas neurons are postmitotic (Chong et al. 1995). This hypothesis is supported by a study in DRPLA postmortem brain specimens (Takano et al. 1996), where a wider range of repeat sizes was observed in older individuals. It was postulated that the continuous mitotic division of the glial cells throughout life leads to this increased somatic mosaicism.

We have determined the repeat size range in postmortem cerebral cortex specimens of 11 MJD and 7 SCA1 patients and in cerebral white-matter specimens of 9 MJD patients and 6 SCA1 patients (tables 1 and 2; fig. 1) and attempted to correlate the amount of somatic mosaicism with age at death of each individual. We studied cerebral cortex and cerebral white-matter samples because these regions remain unaffected in SCA1 and MJD (Schut and Haymaker 1951; Rosenberg et al. 1976; Coutinho and Andrade 1978; Coutinho et al. 1982; Sakai et al. 1983; Kanda et al. 1989; Robitaille et al. 1995). This has allowed us to study the behavior of the expanded (CAG)_n during the life of the individual independently of the disease process (neuronal death, possible gliosis) as well as to compare two regions with different proportions of glial and neuronal cells.

DNA was extracted from samples of ~100 mg of

Table 1

Age at Death and Molecular Characteristics of the 11 MJD Patients Tested

Patient	Age at Death (years)	Normal Allele (CAG units)	Expanded Allele (CAG units)	Mean No. of Bands in Cortex ^a (SD)	Mean No. of Bands in White Matter ^a (SD)
90-274	63	12	69	10 (.6)	9 (1.4)
94-547	63	20	74	10 (1.7)	12 (.7)
M-448	62	12	71	10 (1.5)	10 (1)
93-423	58	20	75	10 (0)	11 (1.7)
94-453	56	24	73	12 (1.5)	10 (1)
M-268	46	18	75	12 (1.4)	11 (1.2)
M-318	44	20	77	10 (.6)	10 (1.2)
M-420	43	20	78	10 (.7)	10 (1.2)
88-196	27	20	80	9 (0)	10 (1.7)
2303	56	12	75	10 (2)	... ^b
1965	52	24	74	9 (2.3)	... ^b

^a Average of three experiments.

^b Cerebral white matter sample was not available for typing.

postmortem cerebral cortex and cerebral white matter by use of standard techniques (Sambrook et al. 1989), and the CAG repeats in the *MJD1* and *SCA1* genes were amplified by PCR as described elsewhere (Orr et al. 1993; Kawaguchi et al. 1994). The size of PCR products was determined by electrophoresis in a denaturing polyacrylamide gel in parallel with an M13 sequencing ladder, and detection was performed by autoradiography. Each experiment was repeated three times to avoid inconstancy of results because of variation in intensity of signal. The results were analyzed by densitometry with a SciScan 5000 scanning densitometer (United States Biochemical). The repeat size, defined as the size of the strongest band and the mean number of bands (with $\geq 10\%$ of the intensity of the strongest band), for each sample was established. The correlation between the amplitude of the size range (mean number of bands) and the age at death in each of the two diseases was studied by regression analysis.

The degree of mosaicism (mean number of bands) of cerebral cortex and white-matter samples was compared using a paired *t*-test.

We observed the presence of marked somatic mosaicism in terms of the expanded (CAG)_n length in all samples tested (fig. 1). The quantitative results are shown in table 1 (MJD) and table 2 (SCA1). In the cerebral cortex, correlation coefficients of $r = .19$ ($P = .56$) and $r = -.37$ ($P = .42$) were obtained for MJD and SCA1, respectively. In cerebral white matter, the correlation coefficients were $r = .15$ ($P = .69$) for MJD and $r = -.5$ ($P = .31$) for SCA1. The comparison of the degree of mosaicism in white matter and cortex indicated that there is no significant difference between the two structures, either in MJD ($P = .47$) or in SCA1 ($P = .14$).

In contrast to the observations for DRPLA (Takano et al. 1996), no significant correlation was observed between age at death and the amount of mosaicism, in the

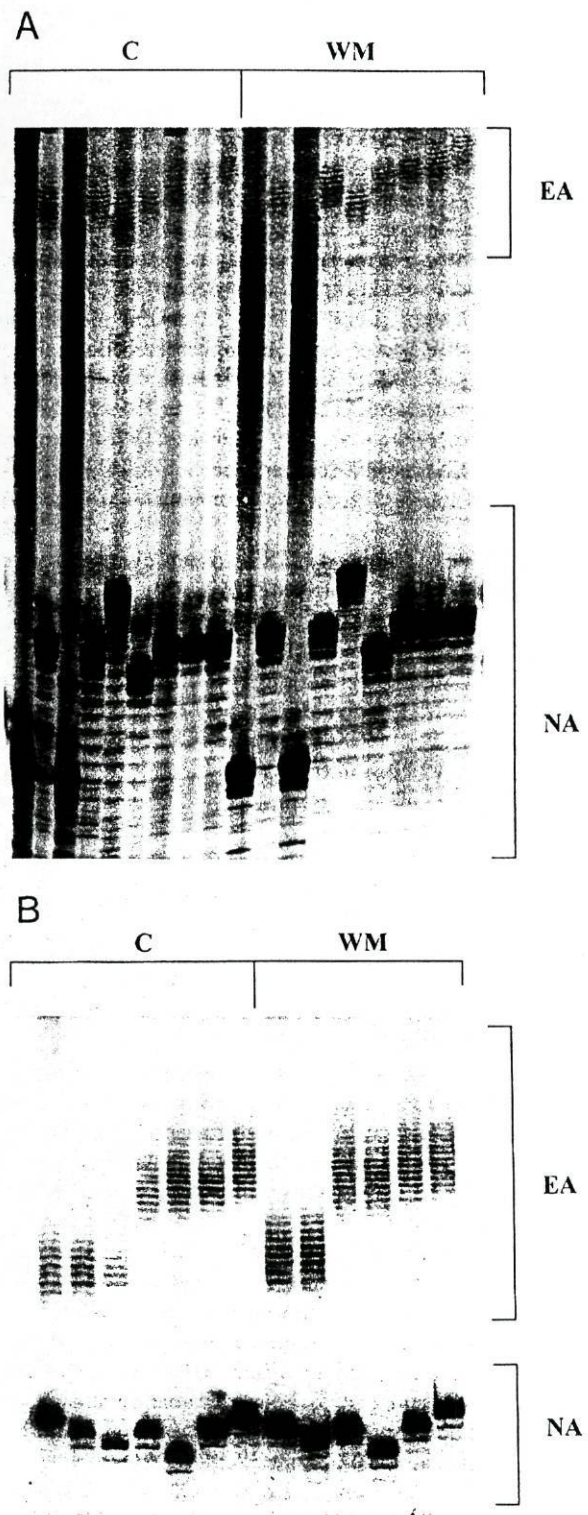
Table 2

Age at Death and Molecular Characteristics of the Seven SCA1 Patients Tested

Patient	Age at Death (years)	Normal Allele (CAG units)	Expanded Allele (CAG units)	Mean No. of Bands in Cortex ^a (SD)	Mean No. of Bands in White Matter ^a (SD)
94-538	77	31	44	11 (1.5)	... ^b
M-652	68	30	45	11 (.6)	13 (1.7)
M-378	65	29	45	11 (1)	12 (2.1)
91-288	57	30	54	15 (4)	15 (2.5)
93-441	53	28	54	16 (2.8)	15 (1.5)
90-276	45	30	56	11 (0)	14 (4)
92-271	38	31	56	13 (4)	14 (0)

^a Average of three experiments.

^b Cerebral white matter sample was not available for typing.



cerebral cortex or white matter, in our sample of MJD and SCA1 patients. In addition, a direct comparison of the cerebral cortex and white matter did not reveal a significant difference in the degree of mosaicism between these structures. These observations suggest that the expanded repeats involved in MJD and SCA1 might behave differently from DRPLA during cell division. This could be associated with the influence of polymorphisms or flanking sequences surrounding the *SCA1*, *MJD*, and *DRPLA* genes.

One limitation of our study is the small number of postmortem samples available for testing, for both MJD and SCA1. However, the MJD and SCA1 sample sizes in our study are equivalent to the one in the DRPLA study by Takano et al. (1996), and the range of ages at death is also comparable: 36 years for MJD and 30 years for SCA1. Nonetheless, it is possible that our sample is too small to detect a more subtle correlation between age at death and amount of mosaicism.

In conclusion, our results do not confirm the existence of a direct correlation between age at death and the degree of somatic mosaicism in CNS in MJD and SCA1, indicating that differences in terms of somatic instability exist between these two expanded CAG repeats and DRPLA. These results do not, however, completely exclude the possibility that glial cells might be the main contributors to the somatic mosaicism of expanded repeat size: more detailed studies, using microdissection and single-cell PCR, would be needed in order to improve our understanding of the origin and biological relevance of somatic mosaicism in the different CAG-related disorders.

PATRÍCIA MACIEL^{1,2,*} ISCIA LOPES-CENDES,^{1,*} STEPHEN KISH,³ JORGE SEQUEIROS,² AND GUY A. ROULEAU¹
¹Centre for Research in Neuroscience, McGill University and the Montreal General Hospital Research Institute, Montreal; ²Laboratório Genética Médica, ICBAS, and UnIGENE, IBMC, Universidade do Porto, Porto, Portugal; and ³Human Neurochemical Pathology Laboratory, Clarke Institute of Psychiatry, Toronto

Acknowledgments

This project was supported by an NIH grant (NS 31687), the Fonds de Recherche en Santé du Québec, the Association Canadienne de l'Ataxie de Friedreich, the Medical Research Council of Canada, the Network of Centres of Excellence (Canadian Genetic Diseases Network), and a European Community Program STRIDE grant. S.K. is supported by U.S. National Institute of Neurological Disorders and Stroke (NINDS) grant 20634. P.M. is the recipient of a scholarship by Junta Nacional de Investigação Científica e

Figure 1 PCR amplification of the CAG repeat in the *MJD1* (A) and *SCA1* (B) genes in postmortem specimens of cerebral cortex and cerebral white matter of affected individuals. Samples are ordered, from left to right, according to tables 1 and 2. Age at death increases from left to right. In panel A, samples belonging to patients 2303 and 1965 are not shown. NA = normal allele; EA = expanded allele; C = cerebral cortex; WM = cerebral white matter.

*These authors contributed equally to this work.

Tecnológica (PRAXIS XXI), Portugal. Some of the tissue specimens were obtained from National Research Specimen Bank, VAMC Wadsworth Division, Los Angeles, which is sponsored by NINDS/National Institute of Mental Health, National Multiple Sclerosis Society, Hereditary Disease Foundation, Comprehensive Epilepsy Program, Tourette Syndrome Association, Dystonia Medical Research Foundation, and Veterans Health Services and Research Administration, Department of Veterans Affairs.

References

- Banfi S, Servadio A, Chung M-Y, Kwiatkowski TJ, McCall AE, Duvick LA, Shen Y, et al (1994) Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nat Genet* 7:513-519
- Chong SS, McCall AE, Cota J, Subramony SH, Orr HT, Hughes MR, Zoghbi H (1995) Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 10:344-349
- Chung M, Ranum LPW, Duvick LA, Servadio A, Zoghbi HY, Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat Genet* 5:254-258
- Coutinho P, Andrade C (1978) Autosomal dominant system degeneration in Portuguese families of the Azores Islands. *Neurology* 28:703-709
- Coutinho P, Guimarães A, Scaravilli F (1982) The pathology of Machado-Joseph disease: report of a possible homozygous case. *Acta Neuropathol* 58:48-54
- Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72:971-983
- Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier J-M, Weber C, et al (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet* 14:285-291
- Kanda T, Isozaki E, Kato S, Tanabe H, Oda M (1989) Type III Machado-Joseph disease in a Japanese family: a clinicopathological study with special reference to the peripheral nervous system. *Clin Neuropathol* 8:134-141
- Kawaguchi Y, Okamoto, T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, et al (1994) CAG expansions in a novel gene from Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* 8:221-227
- Koide R, Ikeuchi T, Onodera O, Tanaka H, Igarashi S, Endo K, Takahashi H, et al (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidolusian atrophy (DRPLA). *Nat Genet* 6:9-13
- La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352:77-79
- Lopes-Cendes I, Maciel P, Kish S, Gaspar C, Robitaille Y, Clark HB, Koeppen AH (1996) Somatic mosaicism in CNS specimens of spinocerebellar ataxia type 1 (SCA1) and Machado-Joseph disease (MJD) patients. *Ann Neurol* 40:199-206
- Orr HT, Chung M-Y, Banfi S, Kwiatkowski TJJ, Servadio A, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 4:221-226
- Pulst S-M, Nechiporuk A, Nechiporuk T, Gispert S, Chen X-N, Lopes-Cendes I, Pearlman S, et al (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* 14:269-276
- Robitaille Y, Schut L, Kish SJ (1995) Structural and immunocytochemical features of olivopontocerebellar atrophy caused by the SCA1 mutation define a unique phenotype. *Acta Neuropathol* 90:572-581
- Rosenberg RN, Nyhan WL, Bay C, Shore P (1976) Autosomal dominant striatonigral degeneration: a clinical, pathological, and biochemical study of a new genetic disorder. *Neurology* 26:703-714
- Sakai T, Ohta M, Ishino H (1983) Joseph disease in a non-Portuguese family. *Neurology* 33:74-80
- Sambrook J, Fritsch EF, Maniatis T (eds) (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, Wakisaka A, et al (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* 14:277-291
- Schut JW, Haymaker W (1951) Hereditary ataxia: a pathological study of five cases of common ancestry. *J Pathol Clin Neurol* 1:183-213
- Servadio A, Koshy B, Armstrong D, Antalffy B, Orr HT, Zoghbi HY (1995) Expression analysis of the ataxin-1 protein in tissues from normal and spinocerebellar ataxia type 1 individuals. *Nat Genet* 10:94-98
- Takano H, Onodera O, Takahashi H, Igarashi S, Yamada M, Oyake M, Ikeuchi T, et al (1996) Somatic mosaicism of expanded CAG repeats in brains of patients with dentatorubral-pallidolusian atrophy: cellular population-dependent dynamics of mitotic instability. *Am J Hum Genet* 58:1212-1222
- Telenius H, Kremer B, Goldberg Y P, Theilmann J, Andrew SE, Zeisler J, Adam S, et al (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet* 6:409-414
- Trottier Y, Lutz Y, Stevanin G, Imbert G, Devys D, Cancel G, Saudou F, et al (1995) Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature* 378:403-406
- Ueno S, Kondoh K, Kotani Y, Komure O, Kuno S, Kawai J, Hazama F, et al (1995) Somatic mosaicism of CAG repeat in dentatorubral-pallidolusian atrophy. *Hum Mol Genet* 4:663-666

Address for correspondence and reprints: Dr. Guy A. Rouleau, Centre for Research in Neuroscience-McGill University, Montreal General Hospital Research Institute, 1650 Cedar Avenue, Montreal, H3G1A4 Québec, Canada.

© 1997 by The American Society of Human Genetics. All rights reserved.
0002-9297/97/6004-0029\$02.00

CAPÍTULO 4

Instabilidade intergeracional do *repeat* CAG expandido

Instabilidade intergeracional do *repeat* CAG expandido

A instabilidade intergeracional é uma das características mais importantes das expansões de *repeats* de trinucleotídeos associados a doenças, levando à variação de tamanho do *repeat* entre famílias e no seio de uma mesma família, que se manifestam sob a forma de idades de início e apresentações clínicas diferentes. É também a instabilidade genética dos *repeats* que constitui a base do fenómeno de antecipação observado nestas doenças.

Recentemente, demonstrou-se o efeito de um polimorfismo intragénico no gene causador da doença de Machado-Joseph, localizado imediatamente após o *repeat* CAG (C⁹⁸⁷GG/G⁹⁸⁷GG), na instabilidade intergeracional do *repeat* CAG expandido [Igarashi *et al.*, 1996a]. Surpreendentemente, observou-se o efeito não só do alelo deste polimorfismo presente no cromossoma associado à doença, mas também do alelo presente no cromossoma normal, sugerindo a existência de uma interacção entre os alelos normal e expandido que afecta de algum modo a fidelidade de replicação da sequência (CAG)_n. Este efeito poderia ser um efeito directo do polimorfismo em questão, ou, alternativamente, o polimorfismo poderia estar em desequilíbrio de ligação com uma outra sequência flanqueante que afecte a instabilidade do (CAG)_n expandido. A fim de confirmar estes resultados na nossa população e de distinguir entre um efeito directo ou indirecto do polimorfismo C⁹⁸⁷GG/G⁹⁸⁷GG, estudámos 70 pares transmissor-descendente, para os quais se conhecia a variação do tamanho do (CAG)_n expandido, determinando os alelos de três polimorfismos intragénicos: C⁹⁸⁷GG/G⁹⁸⁷GG, TAA¹¹¹⁸/TAC¹¹¹⁸ e A⁶⁶⁹TG/G⁶⁶⁹TG. Os nossos resultados confirmam a existência de um efeito significativo do cromossoma normal na instabilidade do (CAG)_n expandido do gene MJD1; no entanto, apenas os polimorfismos C⁹⁸⁷GG/G⁹⁸⁷GG e TAA¹¹¹⁸/TAC¹¹¹⁸, localizados a juzante do *repeat* CAG, parecem ter efeito em termos de instabilidade intergeracional do *repeat* CAG expandido, enquanto o polimorfismo A⁶⁶⁹TG/G⁶⁶⁹TG, situado a montante da sequência repetitiva CAG, não tem efeito no grau de instabilidade. Desconhecem-se ainda os mecanismos moleculares pelos quais o

cromossoma normal poderá afectar a estabilidade intergeracional do repeat expandido, sendo necessários estudos adicionais para esclarecer esta questão.

Estudou-se também, numa população controlo de 125 indivíduos, os polimorfismos $\underline{A}^{669}\text{TG}/\underline{G}^{669}\text{TG}$, $\underline{C}^{987}\text{GG}/\underline{G}^{987}\text{GG}$ e $\text{TAA}^{1118}/\text{TAC}^{1118}$, a fim de determinar possíveis associações entre haplótipos e tamanhos do $(\text{CAG})_n$ que sugerissem uma origem para o cromossoma MJD1 a partir, por exemplo, de um subgrupo de cromossomas normais mais propensos a instabilidade. A observação de uma associação entre os alelos normais de maior tamanho com quatro haplótipos diferentes, sem predominância clara do haplótipo ACA (que é o haplótipo mais frequentemente associado ao $(\text{CAG})_n$ expandido nos doentes), não é no entanto compatível com esta possibilidade. Estes resultados, em combinação com a ausência de qualquer efeito das sequências em *cis* do $(\text{CAG})_n$ expandido na sua instabilidade intergeracional, sugerem que a origem do gene MJD1 mutado não se deva à inserção do *repeat* CAG num haplótipo que predisponha à instabilidade genética.

ARTIGO VI

Study of three intragenic polymorphisms in the Machado-Joseph disease gene (MJD1) in relation to genetic instability of the (CAG)_n tract.

Patrícia Maciel^{1,2}, Claudia Gaspar^{1,2}, Laura Guimarães², Jun Goto³, Iscia Lopes-Cendes¹, Sean Hayes¹, Karin Arvidsson¹, Aureliano Dias¹, Jorge Sequeiros², Alda Sousa⁴ and Guy A. Rouleau¹.

1 Centre for Research in Neuroscience, McGill University and the Montreal General Hospital Research Institute, Montreal, Québec, Canada; 2 Lab. Genética Médica, ICBAS and UnIGENE, IBMC Universidade do Porto, Portugal; 3 Department of Neurology, Tokyo University Medical School, Tokyo, Japan; 4. Genética Aplicada, ICBAS, Universidade do Porto, Portugal.

Running title: Intragenic polymorphisms and instability in MJD

Keywords: ataxia, neurodegenerative disorder, polyglutamine, trinucleotide repeats, dynamic mutations, haplotypes, origin of mutation

Abstract

The intergenerational instability is one of the most important features of the disease-associated trinucleotide expansions, leading to variation in size of the repeat between and within families, which manifests as variable age at onset and severity, and is the basis for the occurrence of anticipation. Several factors are known to affect the degree of instability, namely the type of repeated sequence, its initial size, the presence or absence of interruptions in the repetitive tract and the gender of the transmitting parent. A recent study demonstrated the effect of an intragenic polymorphism in the Machado-Joseph disease causative gene, immediately downstream of the CAG repeat ($C^{987}GG/G^{987}GG$), on the intergenerational instability of the expanded repeat. Surprisingly, there was an effect not only of the polymorphism present in the disease chromosome, but also of the allele of the polymorphism in the normal chromosome, suggesting the existence of an interaction between the normal and expanded alleles that affects the fidelity of replication of the $(CAG)_n$ tract. This effect could be a direct effect of the polymorphism studied, or, alternatively, the polymorphism could be in disequilibrium with some other flanking sequence which affects the instability of the repetitive $(CAG)_n$ tract. In order to confirm the previous results in a different population and to distinguish between a direct and indirect effect of the CGG/GGG polymorphism, we typed 70 pairs parent-progeny for which the variation in the $(CAG)_n$ length in the MJD1 gene was known, for three intragenic polymorphisms, the $C^{987}GG/G^{987}GG$ and two additional, newly described ones, TAA^{1118}/TAC^{1118} and $A^{669}TG/G^{669}TG$. We also typed a control population of 125 individuals for the $C^{987}GG/G^{987}GG$ and TAA^{1118}/TAC^{1118} polymorphisms, in an attempt to identify any association between haplotype and $(CAG)_n$ length in normal chromosomes, suggestive of a instability-predisposing effect of the repeat-flanking sequences, which could have led the origin of the MJD in the human population.

Introduction

Machado-Joseph disease (MJD) is an autosomal dominant spinocerebellar ataxia of adult onset [1], of variable clinical presentation. Its manifestations include cerebellar ataxia and progressive external ophthalmoplegia, associated in a variable degree to pyramidal signs, dystonia, rigidity, amyotrophy, and peripheral neuropathy [1-4]. The mutation associated with this disorder is the expansion of a (CAG)_n tract in the coding region of the MJD1 gene [5], which in normal individuals contains from 12 to 39 repeat units and in expanded chromosomes from 62 to 84 repeat units [5-7]. This mutation is one of the so-called dynamic mutations [8], because the number of trinucleotide units in the expanded repeat segment is not always the same but rather changes along the generations, due to instability of the expanded repeat during mitotic and meiotic cell divisions [5, 6, 9].

Although the mechanisms of trinucleotide repeat instability are still not understood, a variety of factors have been implicated as modulating intergenerational repeat instability in the diseases associated with dynamic mutations, namely the sex of transmitter [6, 7, 10-21], the size of expanded repeat in the transmitting parent [10-15, 21-23], the presence or absence of an interruption of the repeat [19, 24-28] and even the sex of the descendent [29].

We and others have previously shown that in MJD families the sex of the transmitting parent has a significant effect in determination of intergenerational instability [6, 7, 20], with male meiosis associated with larger variations of repeat size, both contractions and expansions. However, there was no association between size of CAG repeat length in the transmitting parent and change in CAG repeat length during transmission [6], nor with the presence of the CAA interruption at the 5' end of the CAG repeat, since both control and expansion-carrying individuals carry this interruption [20, 30].

A single-base substitution polymorphism (C⁹⁸⁷GG/G⁹⁸⁷GG) at the 3' end of the CAG repeat in the MJD1 gene [5] was recently demonstrated to be associated with differential intergenerational instability of (CAG)_n length [20]. Surprisingly, this effect on instability is not observed uniquely for the allele in the disease chromosome (allele in *cis* with the expansion); the combination with the allele in the normal chromosome (allele in *trans*) was shown to be determinant for the degree of intergenerational instability. The biological meaning of these findings is still unclear: while it is possible that this association results from a direct effect of the base substitution at the site examined, it alternatively may reflect the effect of a sequence or gene closely linked to this site.

In order to confirm the previous observations in a different population and to distinguish between a direct and indirect effect of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism, we

analyzed the influence of three intragenic single-base substitution polymorphisms, the $\underline{C}^{987}GG/\underline{G}^{987}GG$ and two additional, newly described ones, TAA^{1118}/TAC^{1118} and $\underline{A}^{669}TG/\underline{G}^{669}TG$, on intergenerational instability of the expanded CAG tract in the MJD1 gene in our set of families. Additionally, we studied a control population for the three intragenic polymorphisms, in an attempt to identify an association between haplotype and $(CAG)_n$ length in normal chromosomes that might give an insight into the origin of the MJD mutation in the human population.

Methods

Subjects

Our study refers to 117 individuals with a previously detected $(CAG)_n$ expansion in the MJD1 gene, including affected individuals and asymptomatic carriers, constituting 70 pairs parent-progeny. These pairs were contributed by 21 different families; one family, B1, contributed 18 pairs (approximately 26%), another contributed 8 pairs (approximately 11%).

Given that the great majority of families in our MJD population was traceable to a Portuguese-Azorean origin, for the study of the association between haplotypes and $(CAG)_n$ length we used a control population of Azorean origin, including 125 normal individuals with no known relationship to an MJD family.

Blood samples were obtained with informed consent from all individuals, and genomic DNA was extracted of lymphocytes as described elsewhere [31].

Size of the CAG repeat and typing of intragenic polymorphisms

Amplification of CAG repeat-containing fragment of the MJD1 gene was performed by PCR using previously described conditions [5], and the size of the PCR products was determined by denaturing polyacrylamide gel electrophoresis, in parallel with an M13 sequence ladder.

The intragenic polymorphism $\underline{C}^{987}GG/\underline{G}^{987}GG$, situated at the 3' end of the CAG repeat, was detected by allele-specific PCR using primers ASP1 (ACTCTGTCCTGATAGGTCCCC) or ASP2 (ACTCTGTCCTGATAGGTCCCG) in combination with MJD52 [5]. The stop codon polymorphism TAA^{1118}/TAC^{1118} [32] was detected by allele-specific PCR using primers ASP3 (GCAAAAATCACATGGAGCTCG) or ASP4 (GCAAAAATCACATGGAGCTCG) in combination with MJD52. In both cases, the PCR reaction was performed using the same conditions used for amplification of the CAG repeat, except for the annealing step, which was performed at 61°C for 30 seconds.

The PCR products were analyzed on a denaturing 6% polyacrylamide gel and visualized by autoradiography.

The $\underline{A}^{669}\text{TG}/\underline{G}^{669}\text{TG}$ polymorphism was detected by SSCP analysis by electrophoresis in a 0.5X MDE gel (FMC BioProducts) with 5% glycerol; the PCR amplification was performed using primers MJDIVSR (TACTAGAGCTTATTTGCCAG) and MJD734R (CAGAGCCCTCTGCAAATCCT), in a solution containing 1X PCR buffer (Perkin Elmer), 0.2 mM dNTP-A, 0.03 mM dATP, 8 ng/ μl each primer, 0.2 μCi ^{35}S dATP, 0.1 $\mu\text{g}/\mu\text{l}$ BSA, 0.12 U AmpliTaq (Perkin Elmer). The amplification conditions were as follows: initial denaturation for 10 minutes at 94°C, followed by 1 minute at 94°C, 1 minute at 56°C and 1 minute at 72°C for 30 cycles and a final elongation period of 5 minutes at 72°C. When necessary, the polymorphisms were typed for both parents and progeny, in order to determine phase.

The relative positions of the three intragenic polymorphisms studied in the MJD1 gene is shown in Figure 1. The results are shown in the tables taking in consideration the order of the polymorphisms, from 5' to 3', in the MJD1 cDNA.

Statistical methods

In order to analyze the associations between intragenic polymorphisms and the size of the CAG repeats in normal chromosomes (of control individuals), we compared the frequencies of the alleles in three classes of normal chromosomes, grouped according to size of (CAG)_n (up to 21, 22 to 26, and more than 27 repeat units), using Fisher's Exact test 2xk and Pearson Chi-square test.

The intergenerational changes in CAG repeat length were analyzed by comparison of absolute change in size of the expanded allele during transmission. Fisher's Exact test (2x2) was used to determine if intergenerational change in repeat number between parent and offspring, expressed as two classes of variation (difference of more than two repeat units between parent and progeny - small variation - and difference of up to two repeat units - large variation), was associated with sex of transmitting parent.

The distribution of the variable intergenerational changes in CAG repeat length as defined is highly skewed, therefore non-parametric methods are appropriate for hypothesis testing; Mann-Whitney U test or Kruskal-Wallis ANOVA were used to assess associations between alleles/haplotypes and instability.

Linkage disequilibrium analysis was performed by determination of the δ value, as described previously [33, 34].

All analyses were performed using procedures of SPSS, release 6.1.3, ©SPSS Inc. (1989-1995) and Exact 2xk- Exact test for a 2xk table, version 2.04, © JH Abramson & PM Gahlinger (1993-1996).

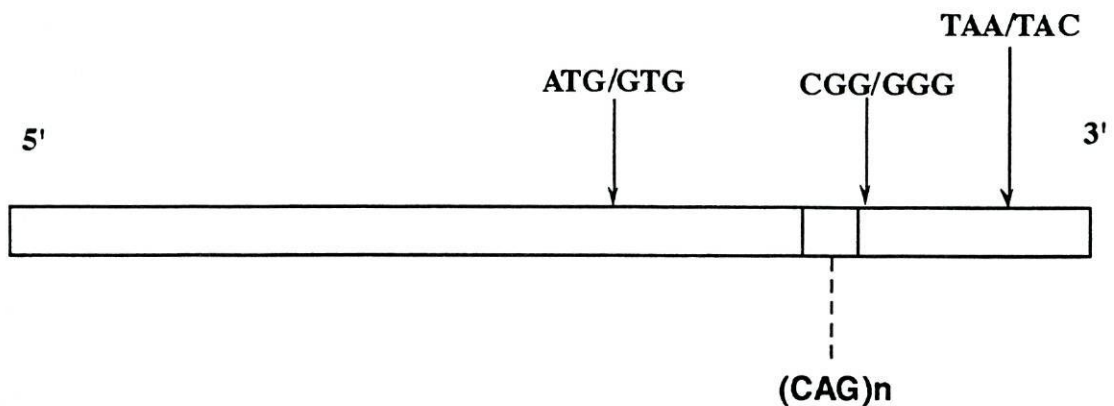


Figure 1. Relative position of the intragenic polymorphisms studied in the MJD1 cDNA.

Results

Larger sizes of the (CAG)_n in MJD1 are not exclusively associated with specific alleles of the intragenic polymorphisms.

The distribution of CAG repeat sizes in the normal chromosomes of our Portuguese-Azorean control population is trimodal, with modes at 14, 24 and 28 CAG repeats. For the analysis of association between intragenic polymorphisms and (CAG)_n length in normal chromosomes, three groups of normal alleles (defined based on the (CAG)_n lengths - from 15 to 21, 22 to 26 and 27 to 36) were compared. The frequencies of the CGG and GGG alleles of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism were significantly different between groups (Fisher's exact test, $p=0.0021$). The detailed analysis of the distribution of these alleles revealed that although there is an increased relative frequency of the CGG allele among the chromosomes with larger CAG repeats when compared with the small and medium CAG repeats, this allele is not exclusively associated with the larger repeat lengths or even the predominant allele in this group of chromosomes (figure 2a).

The same analysis was performed for the TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphism in 231 normal chromosomes, and we observed a difference in distribution ($p=0.0000$), but although there was an enrichment of allele TAA among the larger CAG repeat-containing normal alleles when compared to the small or medium size repeats, no predominance or exclusive association of this allele with larger repeats was observed (figure 2b).

For the $\underline{A}^{669}\text{TG}/\underline{G}^{669}\text{TG}$ polymorphism, in numerous cases it was not possible to determine which allele co-segregated with each $(\text{CAG})_n$ length in normal chromosomes, due to the detection technique used, therefore a limited amount of control cases could be analyzed (only the individuals which were homozygous either for the polymorphism or for the repeat size could be used, $n=154$). Similar results were obtained for this polymorphism (figure 2c, $p=0.051$) as for the $\underline{C}^{987}\text{GG}/\underline{G}^{987}\text{GG}$ and $\text{TAA}^{1118}/\text{TAC}^{1118}$ polymorphisms. A summary of these results is presented in table 1.

Table 1. Comparison of the distribution of CAG repeat length in normal chromosomes carrying different alleles of the three intragenic polymorphisms in the MJD1 gene (Fisher's exact test, considering three groups).

polymorphism	allele	n	significance
$\underline{A}^{669}\text{TG}/\underline{G}^{669}\text{TG}$	a	12	0.0051
	g	142	
$\underline{C}^{987}\text{GG}/\underline{G}^{987}\text{GG}$	c	56	0.0021
	g	176	
$\text{TAA}^{1118}/\text{TAC}^{1118}$	a	49	0.0000
	c	182	

Haplotypes were possible to determine for 133 individuals, and the analysis of the distribution of haplotypes according to $(\text{CAG})_n$ sizes revealed that four different haplotypes are present in the normal chromosomes with larger CAG repeats (figure 3), and that, although we could observe an increased relative frequency of the \underline{ACA} haplotype among these chromosomes when compared to the chromosomes with medium size and small $(\text{CAG})_n$ tracts, there is no exclusive association or even predominance of this haplotype among normal chromosomes carrying larger CAG repeats. The difference in $(\text{CAG})_n$ size distribution among haplotypes was not statistically significant.

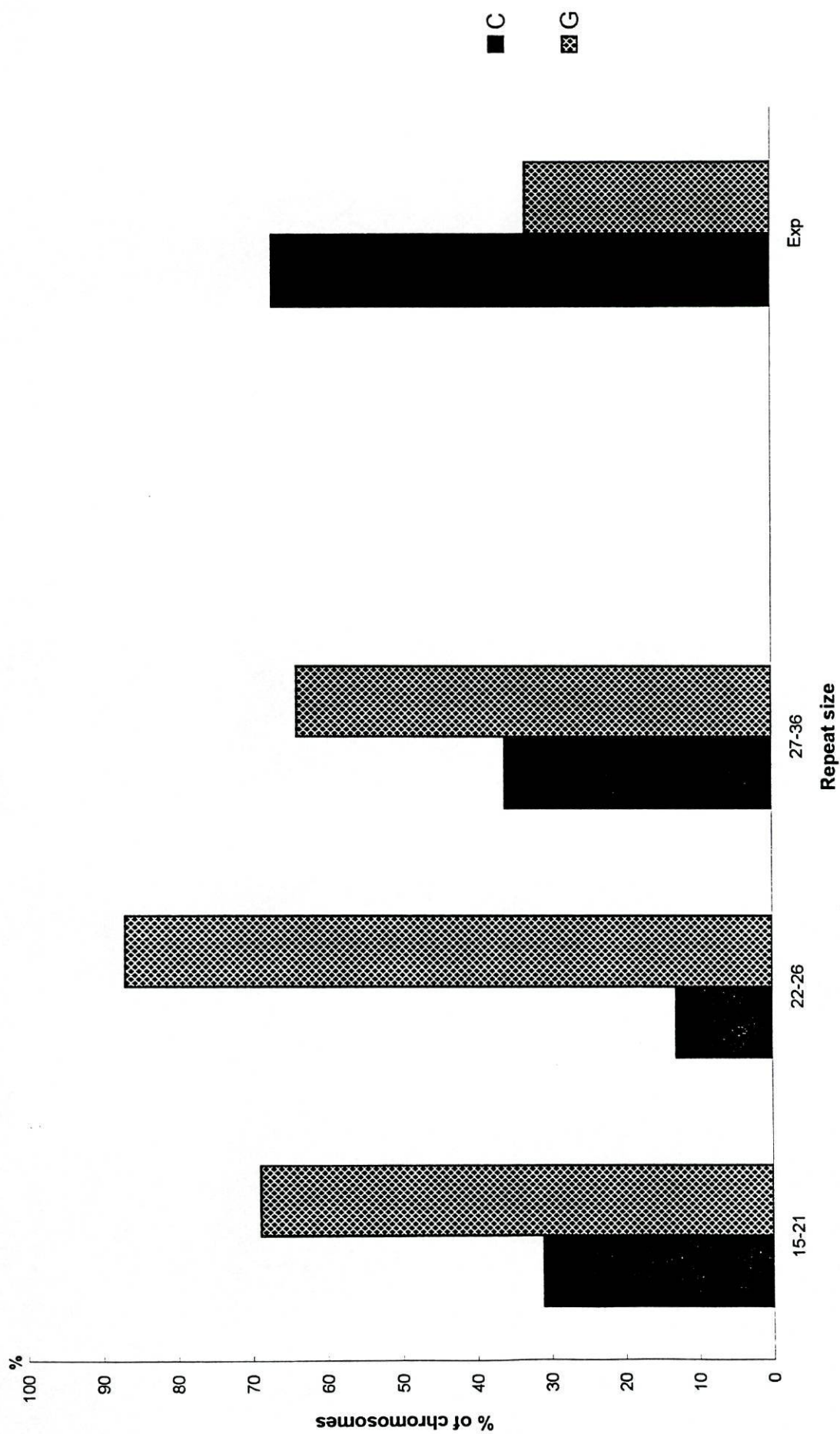


Fig. 2a. Distribution of $(CAG)_n$ sizes among carriers of the two alleles of the $C^{987}GG/G^{987}GG$ polymorphism, in a control population of Portuguese-Azorean origin and in the MJD families of Portuguese descent.

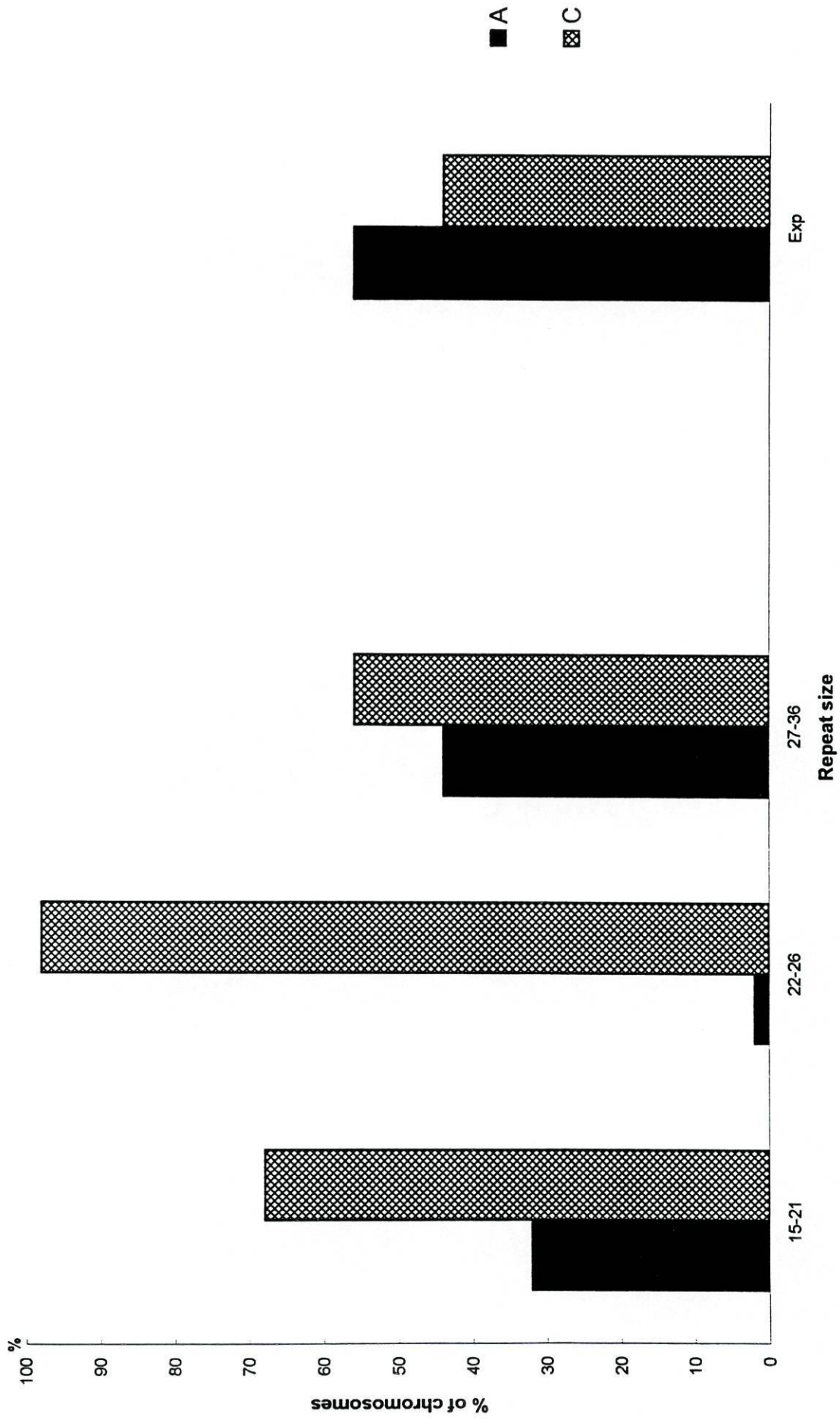


Fig. 2b. Distribution of $(CAG)_n$ sizes among carriers of the TAA^{1118}/TAC^{1118} polymorphism, in a control population of Portuguese-Azorean origin and in the MJD families of Portuguese descent.

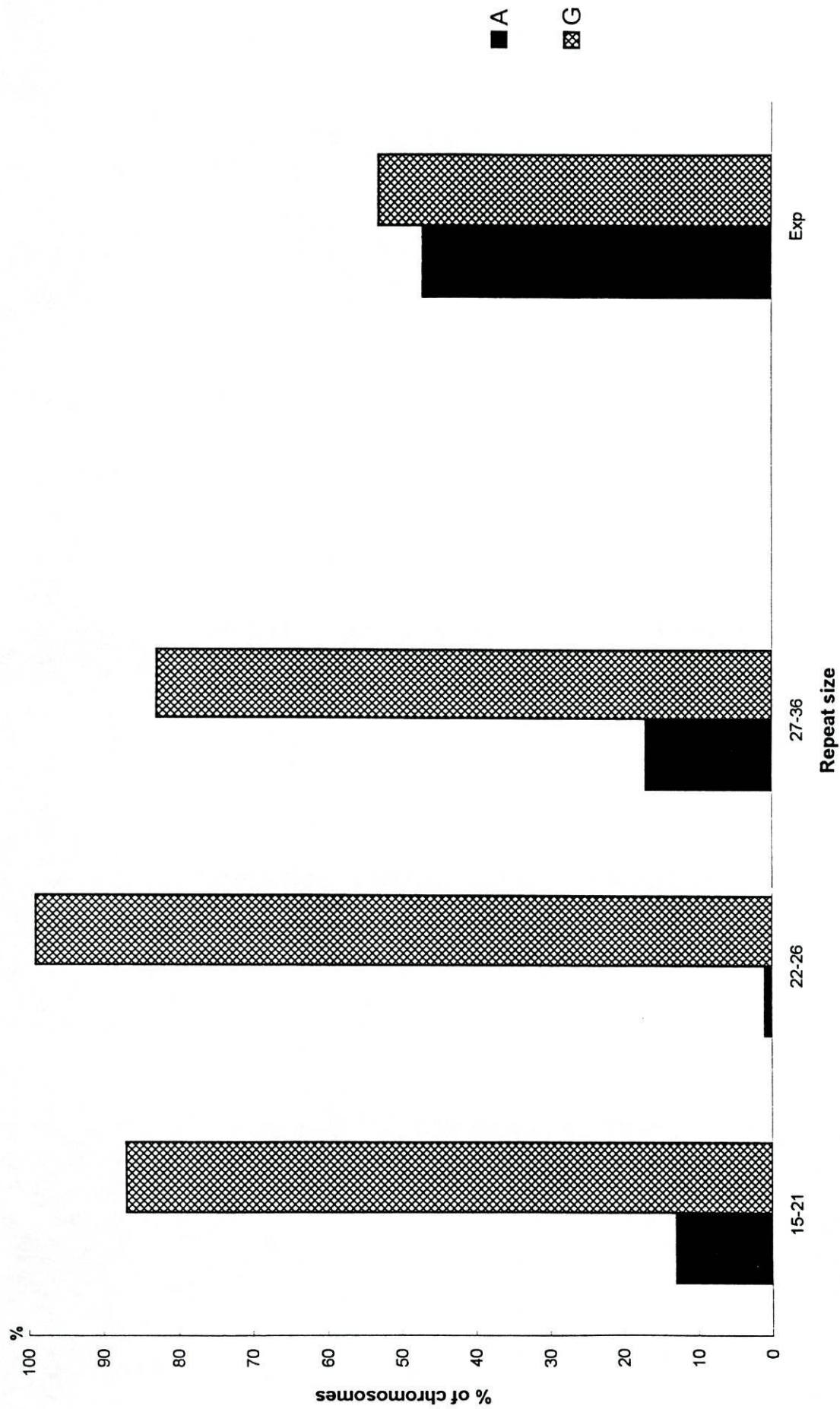


Fig. 2c. Distribution of $(CAG)_n$ sizes among carriers of the two alleles of the $A^{660}TG/G^{660}TG$ polymorphism, in a control population of Portuguese-Azorean origin and in the MJD families of Portuguese descent.

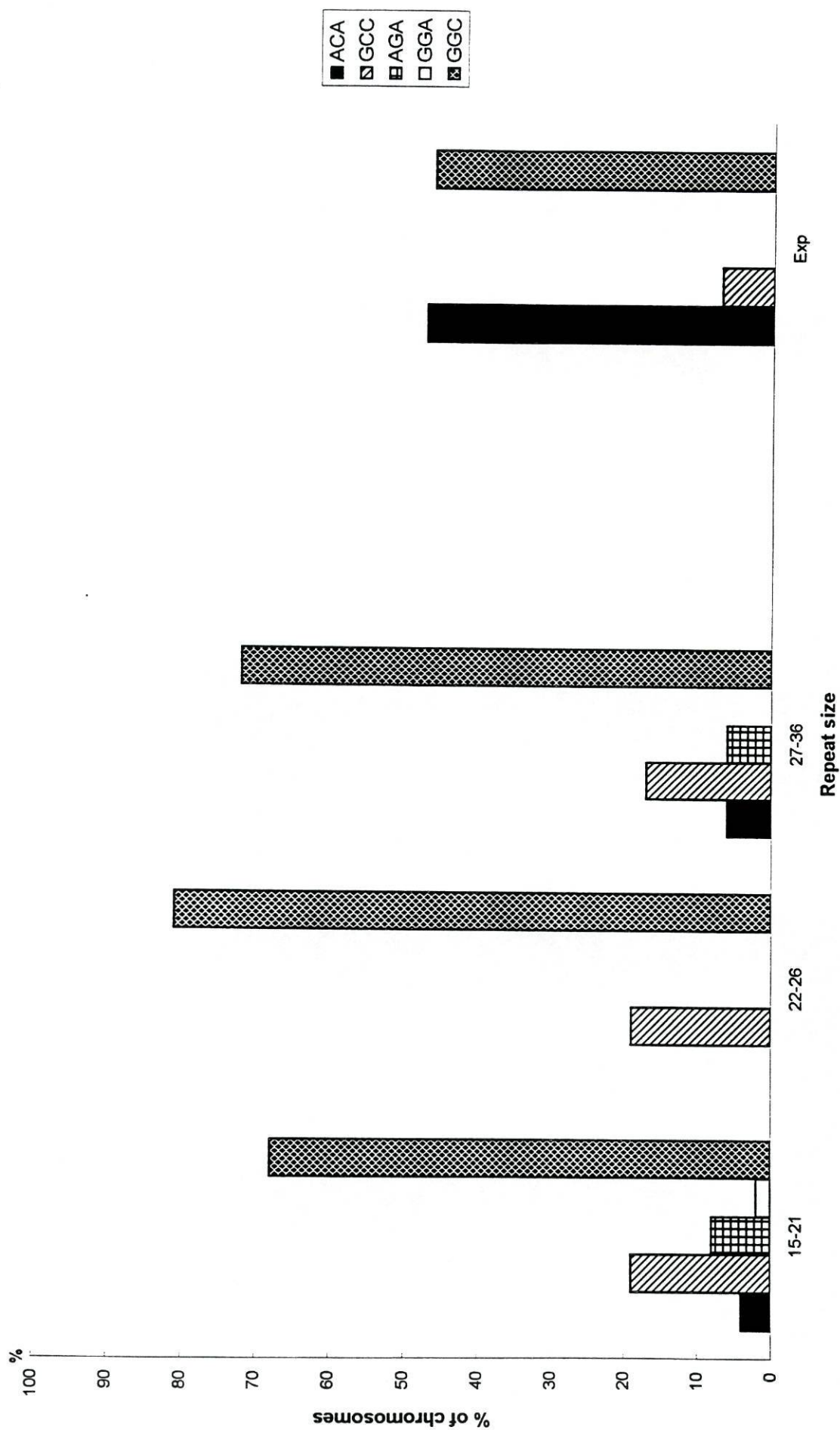


Fig. 3. Distribution of $(CAG)_n$ sizes among carriers of different intragenic haplotypes, defined by the $A^{669}TG/G^{669}TG$, $C^{987}GG/G^{987}GG$ and TAA^{1118}/TAC^{1118} polymorphisms, in a control population of Portuguese-Azorean origin and in the MJD families of Portuguese descent.

Effect of gender of transmitting parent on intergenerational instability of expanded (CAG)_n

We confirmed our previous observation regarding the effect of the gender of transmitting parent on intergenerational instability [6]: when two classes of variation in (CAG)_n size between parent and progeny were considered (small variation and large variation), there was a significant difference between the maternal and paternal transmissions, with increased instability (large variation) associated with male meiosis ($p=0.0017$, Fisher's exact test).

No apparent association between the alleles of intragenic polymorphisms in cis of the (CAG)_n expansion and amplitude of intergenerational variation of expanded (CAG)_n size.

We detected no significant difference in the distribution of absolute variation of (CAG)_n length during transmission between individuals carrying either the C or G alleles of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism in *cis* with the expanded allele (Mann-Whitney U-test, $p=0.150$). Likewise, no significant difference between the CGG or GGG alleles was observed when considering three or five classes of intergenerational variation of repeat size (results not shown). When the gender of the transmitter was taken into account, there was still no significant difference between carriers of CGG or GGG alleles (see table 2).

The comparison of absolute variation in repeat size between individuals carrying either the A or the C variants of the TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphism in *cis* with the expanded CAG also revealed the absence of a significant difference in the median of the absolute variation in repeat size between the two groups (Mann-Whitney U-test, $p=0.450$). The same was observed for the A⁶⁶⁹TG/G⁶⁶⁹TG polymorphism (Mann-Whitney U-test, $p=0.700$).

No association was detected between the presence of any particular intragenic haplotype in *cis* (defined by the three intragenic polymorphisms mentioned above) with increased instability (Kruskal-Wallis test, $p=0.378$). Similar results were obtained for partial haplotypes, defined by combinations of two of these polymorphisms.

Table 2. Association of the alleles of the three intragenic polymorphisms in the *MJD1* gene in *cis* of the expanded (CAG)_n with increased intergenerational instability of the repeat (Mann-Whitney test).

polymorphism	all transmissions	male transmissions	female transmissions
<u>A</u> ⁶⁶⁹ TG/ <u>G</u> ⁶⁶⁹ TG	p=0.700	p=0.692	p=0.616
<u>C</u> ⁹⁸⁷ GG/ <u>G</u> ⁹⁸⁷ GG	p=0.150	p=0.123	p=0.755
TAA ¹¹¹⁸ / <u>TAC</u> ¹¹¹⁸	p=0.450	p=0.287	p=0.613

Influence of the genotype of intragenic polymorphisms on the intergenerational instability of the expanded (CAG)_n.

When the transmission events were divided in groups according to genotype of the transmitting parent for the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism, taking both the alleles in *cis* and in *trans* into account (table 3), there was a significant difference in the median of the absolute variation in repeat length during transmission between the groups, with the parental (CAG)exp-CGG/(CAG)normal-GGG genotype associated with increased instability (Kruskal-Wallis test, p=0.028).

When the genotype of the stop codon polymorphism was considered, the (CAG)exp-TAA/TAC polymorphism was associated with larger intergenerational variation, but the difference was not statistically significant (Kruskal-Wallis test, p=0.080).

The parental genotype for the A⁶⁶⁹TG/G⁶⁶⁹TG polymorphisms did not have a significant effect on instability (Kruskal-Wallis test, p=0.632).

Table 3. Association of genotypes (alleles in *cis+trans* of the expanded repeat) of the three intragenic polymorphisms in the *MJD1* gene with increased instability (Kruskal-Wallis).

polymorphism	all transmissions	male transmissions	female transmissions
<u>A</u> ⁶⁶⁹ TG/ <u>G</u> ⁶⁶⁹ TG	p=0.632	p=0.914	p=0.304
<u>C</u> ⁹⁸⁷ GG/ <u>G</u> ⁹⁸⁷ GG	p=0.028 (CG)	p=0.623	p=0.014 (CG)
TAA ¹¹¹⁸ / <u>TAC</u> ¹¹¹⁸	p=0.080 (AC)	p=0.685	p=0.027 (CC)

Effect of the alleles of intragenic polymorphisms present in the normal chromosome on the amplitude of intergenerational variation of expanded (CAG)_n size in the MJD chromosome.

The previous observations suggest an effect of the genotype for the intragenic polymorphisms C⁹⁸⁷GG/G⁹⁸⁷GG on intergenerational instability but, surprisingly, exclude a noticeable effect of the alleles of this polymorphism in *cis* of the expansion. To further investigate these results, we studied the effect of the alleles of the three intragenic polymorphisms present in the normal chromosome of the transmitting parent (in *trans* of the expansion) on intergenerational instability of the expanded repeat. The results of this analysis are shown in table 4. A highly significant association was observed between the presence of the C variant of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism on the normal chromosome of the transmitting parent and an increased amplitude of intergenerational variation in expanded repeat size (Mann-Whitney U-test, p=0.019)

A similar analysis for both the TAA¹¹¹⁸/TAC¹¹¹⁸ and the A⁶⁶⁹TG/G⁶⁶⁹TG polymorphisms revealed an association of the C variant of the stop codon polymorphism present in the normal chromosome of the transmitter with an increased intergenerational instability of the expanded repeat (Mann-Whitney U-test, p=0.024), but no such association was detected for the A⁶⁶⁹TG/G⁶⁶⁹TG polymorphism (Mann-Whitney U-test, p=0.291).

Table 4. Association of the alleles of the three intragenic polymorphisms in the *MJD1* gene in *trans* of the expanded (CAG)_n with increased intergenerational instability of the repeat (Mann-Whitney test).

polymorphism	all transmissions	male transmissions	female transmissions
<u>A</u> ⁶⁶⁹ TG/ <u>G</u> ⁶⁶⁹ TG	p=0.291	p=0.942	p=0.092 (G)
<u>C</u> ⁹⁸⁷ GG/ <u>G</u> ⁹⁸⁷ GG	p=0.019 (G)	p=0.301	p=0.040 (G)
TAA ¹¹¹⁸ / <u>TAC</u> ¹¹¹⁸	p=0.024 (C)	p=0.650	p=0.003 (C)

We did not detect a significant association of any particular intragenic haplotype (defined by the three intragenic polymorphisms mentioned above) in the normal chromosome of the transmitter with increased intergenerational instability of the expanded repeat (Kruskal-Wallis, p=0.321). However, when partial haplotypes, including only two of the intragenic haplotypes, were analyzed, the haplotype defined by the C⁹⁸⁷GG/G⁹⁸⁷GG and TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphisms seemed to have an effect on instability, with the GC haplotype in *trans* associated with larger variations of the expanded repeat size (Kruskal-Wallis, p=0.042). The partial haplotypes defined by the A⁶⁶⁹TG/G⁶⁶⁹TG and C⁹⁸⁷GG/G⁹⁸⁷GG polymorphisms, and by the A⁶⁶⁹TG/G⁶⁶⁹TG and TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphisms were not associated with differential instability (Kruskal-Wallis, p=0.270 and p=0.197, respectively).

Combined analysis of the effects of intragenic polymorphisms and gender of the transmitting parent.

Knowing that gender of transmitting parent is an important factor determining the degree of intergenerational instability of the expanded (CAG)_n in MJD, we analyzed the influence of intragenic polymorphisms on instability separately for maternal and paternal transmissions.

As is shown in table 3, the significance of the effect of genotype of intragenic polymorphisms on intergenerational instability was increased when we considered maternal transmissions only, but for paternal transmissions no significant effect of the genotype was observed. Similarly, when the analysis of the effect of the polymorphisms in *trans* of the

expanded allele was performed separately for paternal and maternal transmissions, we observed a significant effect only in maternal transmissions (table 4).

The haplotype analysis confirmed these observations: the GC haplotype defined by the C⁹⁸⁷GG/G⁹⁸⁷GG and TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphisms in *trans* was associated with increased instability in maternal transmissions only (Kruskal-Wallis, $p=0.029$).

The direct comparison between haplotypes GA and GC defined by the C⁹⁸⁷GG/G⁹⁸⁷GG and TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphisms in maternal transmissions revealed a borderline significant difference in the degree of instability among carriers of these haplotypes ($p=0.047$, Mann-Whitney U-test), which was not detected when the two genders were analyzed together ($p=0.656$, Mann-Whitney U-test). The comparison of the CC and GC haplotypes revealed no significant difference in the degree of instability ($p=0.555$, Mann-Whitney U-test).

Analysis of linkage disequilibrium between the intragenic polymorphisms in the MJD1 gene, in normal chromosomes.

In order to clarify the meaning of the results obtained with the partial haplotypes defined by the C⁹⁸⁷GG/G⁹⁸⁷GG and TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphisms, in *trans* of the MJD mutation, we tested for the existence of linkage disequilibrium between intragenic polymorphisms in the MJD1 gene, in the normal population. Linkage disequilibrium was detected between the C⁹⁸⁷GG/G⁹⁸⁷GG and TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphisms ($p=0.0000$).

Discussion

Intergenerational instability is one of the most important features of the disease-associated trinucleotide expansions, leading to variation in size of the repeat among and within families, which manifests as variable age of onset and severity of the disease, and is the basis for the occurrence of anticipation.

The molecular basis of the genetic instability of triplet repeats is still unknown. Some speculative models have been proposed [35-39], involving unequal crossing-over, DNA polymerase slippage during replication, mismatch-repair difficulties due to the repetitive nature of the sequences, increased DNA polymerase pausing, and/or formation of stable hairpin loops which would stabilize the sliding strand [40]. Whereas none of these mechanisms separately can account for the observed genetic features of the dynamic mutations associated with human disease, it is likely that a combination of several of them provides the basis for the unique behaviour of these sequences during transmission [38].

The identification of factors such as the length of repetitive tract, presence or absence of interruptions, gender of transmitting parent, chromosomal environment and others that contribute to genetic instability in human disorders caused by expanded trinucleotide repeats provides important clues for the understanding of the biological processes which lead to this instability, and any proposed models for mutation in the trinucleotide repeats should account for these observations. Furthermore, the identification of factors that clearly determine trinucleotide repeat instability in human diseases could possibly allow for some predictability of the behaviour of the expanded repeat during transmission in affected families, which would be of manifest relevance for genetic counseling.

Previous studies of the segregation of the expanded (CAG)_n in Machado-Joseph disease have demonstrated the influence of the gender of the transmitting parent [6, 7, 20] and of an intragenic single-base substitution polymorphism immediately 3' to the (CAG)_n tract on intergenerational instability [20]. In a study by Igarashi and colleagues, the C variant of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism was shown to be preferentially associated both with larger repeats in normal chromosomes and with the expanded repeats in MJD chromosomes, raising the possibility that this substitution confers an increased instability to the (CAG)_n tract. This hypothesis would be in agreement with the fact that most MJD families, even from highly diverse ethnic origins, share a common intragenic haplotype.

Surprisingly, in the same study, the analysis of the effect of this polymorphism on intergenerational instability of the expanded repeat in MJD pedigrees demonstrated that the allele carried by the transmitting parent in the normal chromosome strongly affected intergenerational instability, with the (CAG)_{exp}-CGG/(CAG)_{normal}-GGG genotype associated with increased instability.

The results of our analysis of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism and two additional polymorphisms in a different population confirm some but not all of the previous observations. When the association between (CAG)_n length and the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism was assessed in normal chromosomes, an increased relative frequency of the C allele among larger normal alleles was observed, in agreement with previous studies [20, 41]. Similarly, increased relative frequencies of allele A of the A⁶⁶⁹TG/G⁶⁶⁹TG polymorphism and allele A of the TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphism were detected among longer repeats when compared to small and medium size (CAG)_n. It is interesting however that, in our study as well as the two previous studies [20, 41], there was no exclusive association of the larger alleles with any allele of the polymorphisms in the population of Portuguese origin, in contrast with the observations in populations of other ethnic origins (French and Japanese, but not North-African). In terms of the intragenic haplotypes defined by these polymorphisms, four different haplotypes were found among the group of larger normal alleles in our control population of Azorean origin, and although the ACA haplotype, which was shown in a large population of MJD patients of various ethnic

origins to be present in approximately 70% of the MJD chromosomes [42] was found with increased relative frequency among the larger alleles, the difference in distribution of sizes between haplotypes was not statistically significant, and there was no preferential association of larger sizes with any particular haplotype.

Also, the sequence variations surrounding the expanded (CAG)_n do not seem to have the predicted destabilizing effect during transmission of this repeat in MJD pedigrees: the analysis of the effect of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism on the mutation rate of the expanded (CAG)_n in our population of expansion carriers revealed no significant effect of the allele in *cis* of the expansion on intergenerational instability. This is in agreement with results obtained by Igarashi and colleagues, who could not demonstrate a significant effect of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism in *cis* of the expansion on instability.

These results are not compatible with the suggestion that the expanded alleles in MJD chromosomes might have originated from a pool of expansion-prone normal alleles. On the one hand, although instability-predisposing haplotypes have been demonstrated in other trinucleotide repeat-related human disorders, with specific haplotypes in linkage disequilibrium with repeat copy numbers at the high end of the normal range and with the expanded alleles [21, 26, 43-54], this does not seem to be the case in MJD, at least not in all populations. On the other hand, if there was a instability-predisposing haplotype, its effect should also be detectable on intergenerational instability of the expansion-carrying chromosomes, therefore there should be an effect of the haplotype on instability in *cis*, which we do not detect. In this sense, it will be interesting to determine whether the recently identified intermediate size CAG repeats in the MJD1 gene [55, 56], are preferentially associated with any particular intragenic haplotype.

Both our results and the results of the previous study [20] did, nevertheless, demonstrate the effect of the genotype (combination of the alleles in *cis* and in *trans*) of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism on intergenerational instability. Furthermore, in our study, the analysis of this polymorphism in *trans* of the expanded (CAG)_n separately revealed a highly significant effect, suggesting that the effect observed when the genotype is considered could be mainly contributed by the allele of the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism in *trans* of the expansion. Although there is no mention of the analysis of the allele in *trans* only in the study by Igarashi and colleagues, it is possible that this is the case in their population too.

Additionally, our results of the analysis of two other intragenic polymorphisms in *trans* of the expanded MJD gene suggest that this effect of the normal chromosome could be due specifically to the C/G substitution, and not to a more widespread chromosomal effect. The A⁶⁶⁹TG/G⁶⁶⁹TG polymorphism does not show an effect on instability, whereas the TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphism does. This could be due to the fact that this polymorphism is in linkage disequilibrium with the C⁹⁸⁷GG/G⁹⁸⁷GG polymorphism in the

normal population. Unfortunately, the size of our sample does not allow us to study interactions between haplotype and gender and to develop a model considering the contribution of multiple factors to intergenerational instability of the expanded (CAG)_n in the MJD1 gene. However, when we analyzed the effect of the polymorphisms on the normal chromosome and instability separately in maternal and paternal transmissions, we found that this effect was significant only for maternal transmissions, both when we consider each polymorphism separately or when we analyze the effect of the haplotype defined by the C⁹⁸⁷GG/G⁹⁸⁷GG and the TAA¹¹¹⁸/TAC¹¹¹⁸ polymorphisms. This is in contrast with the observations by Igarashi and colleagues, who observe a more significant effect during paternal transmissions, and was an unexpected result, considering the previously demonstrated (and now reconfirmed) higher instability of the expanded repeat during paternal transmissions.

The main limitation of this type of study, in which size variations were determined between the progenitor's lymphocyte DNA and the descendant's lymphocyte DNA, is that this measure reflects both the instability during gamete formation and the somatic instability during pre and post-natal development of the individual, which can diminish the impact of the factors in study. A more direct approach would be to perform single sperm studies, thus avoiding somatic instability. However, these would allow for study of paternal transmissions only, and our observations suggest a significant effect for maternal transmissions only; studies of the female germline are obviously impractical.

The biological mechanism by which the contribution of the normal chromosome to instability might occur is not clear. One possibility is the occurrence of a direct interaction between chromosomes, such as in gene conversion events [20]. Gene conversion has been shown to occur in the Huntington disease gene, with variation in repeat size but no alteration of flanking markers [57, 58]. Since the variation in size of the expanded allele in MJD pedigrees occurs with no corresponding change in normal allele, the mechanism of unequal crossing-over would have to involve complex interactions, such as hydrogen bonding between hairpins on the homologous alleles during replication, as suggested by McMurray and colleagues (1995). Besides, any changes in repeat number must occur on the 3' side of the CAG tract, because the CAA interruption seems to be in the same position in most chromosomes, or else absent [30]. This polarity of variation was previously demonstrated in gene conversion events leading to mutation of human minisatellites [59, 60]. In this perspective, it is noteworthy that the intragenic polymorphisms which seem to affect the genetic instability of the CAG repeat would be both localized in the 3' flanking region of this repeat.

In conclusion, we believe that both the lack of a significant association of any specific haplotype with larger CAG repeat in normal chromosomes and the absence of an effect of the intragenic haplotype in *cis* on the intergenerational instability of the expanded (CAG)_n in MJD families do not suggest the existence of an instability-predisposing

haplotype leading to recurrent mutation from the normal size range to the disease size range, as the mechanism leading to the origin of the MJD mutation in different ethnic groups throughout the world. We confirmed the previously described effect of the C⁹⁸⁷GG/G⁹⁸⁷GG substitution polymorphism on intergenerational instability when present in the normal chromosome, and, through studies of additional intragenic polymorphisms, demonstrated that the effect is restricted to a small region of the gene, immediately downstream of the CAG repeat, which includes this particular nucleotide substitution and the stop codon of the MJD1 cDNA, and is not a more widespread chromosomal effect. Further studies are needed to clarify the mechanism by which the normal chromosome might contribute to instability of the expanded CAG tract in the MJD causative gene.

References

- 1 Coutinho P and Andrade C: Autosomal dominant system degeneration in Portuguese families of the Azores Islands. *Neurology* 1978;28:703-709.
- 2 Rosenberg RN, Nyhan WL, Bay C and Shore P: Autosomal dominant striatonigral degeneration: a clinical, pathological, and biochemical study of a new genetic disorder. *Neurology* 1976;26:703-714.
- 3 Coutinho P and Sequeiros J: Aspects cliniques, genetiques et pathologiques de la maladie de Machado-Joseph. *Journal de Génétique Humaine* 1981;29:203-209.
- 4 Barbeau A, Roy M, Cunha L, de Vicente AN, Rosemberg RN, Nyhan WL, MacLeod PL, Chazot G, Langston LB, Dawson DM and Coutinho P: The Natural History of Machado-Joseph Disease. *The Canadian Journal of Neurological Sciences* 1984;11:510-525.
- 5 Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, Kimura J, Narumiya S and Kakizuka A: CAG expansions in a novel gene from Machado-Joseph disease at chromosome 14q32.1. *Nature Genetics* 1994;8:221-227.
- 6 Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, Dawson DM, Sudarski L, Guimarães J, Loureiro JEL, Nezarati MM, Corwin LI, Lopes-Cendes I, Rooke K, Rosenberg R, MacLeod P, Farrer LA, Sequeiros J and Rouleau GA: Correlation between CAG repeat length and clinical features in Machado-Joseph Disease. *American Journal of Human Genetics* 1995;57:54-61.
- 7 Maruyama H, Nakamura S, Matsuyama Z, Sakai T, Doyu M, Sobue G, Seto M, Tsujihata M, Oh-i T, Nishio T, Suhonara N, Takahashi R, Hayashi M, Nishino I, Ohtake T, Oda T, Nishimura M, Saida T, Matsumoto H, Baba M, Kawaguchi Y, Kakizuka A and Kawakami H: Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Human Molecular Genetics* 1995;4:807-812.
- 8 Richards RI and Sutherland GR: Dynamic Mutations: A New Class of Mutations Causing Human Disease. *Cell* 1992;70:709-712.
- 9 Takiyama Y, Igarashi S, Rogaeva EA, Endo K, Rogaev EI, Tanaka H, Sherrington R, Sanpei K, Liang Y, Saito M, Tsuda T, Takano H, Ikeda M, Lin C, Chi H, Kennedy JL, Lang AE, Wherrett JR, Segawa M, Nomura Y, Yuasa T, Weissenbach J, Yoshida M, Nishizawa M, Kidd KK, Tsuji S and St George-Hyslop PH: Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. *Human Molecular Genetics* 1995;4:1137-1146.
- 10 Fu Y-H, Kuhi DPA, Pizzuti A, Pierretti M, Sutcliffe JS, Richards S, Verkerk JMH, Holden JJA, Fenwick RGJ, Warren ST, Oostra BA, Nelson DL and Caskey T: Variation of CGG Repeat at the Fragile X Site Results in Genetic Instability: Resolution of the Sherman Paradox. *Cell* 1991;67:1047-1058.

- 11 Snow K, Doud LK, Hagerman R, Pergolizi RG, Erster SH and Thibodeau SN: Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. *American Journal of Human Genetics* 1993;53:1217-1228.
- 12 Fisch GS, Snow K, Thibodeau SN, Chalifaux M, Holden JJA, Nelson DL, Howard-Peebles PN and Maddalena A: The fragile X premutation in carriers and its effect on mutation size in offspring. *American Journal of Human Genetics* 1995;56:1147-1155.
- 13 Brunner HG, Nelen M, Breakefield O, Ropers HH and van Oost BA: Abnormal Behavior Associated with a Point Mutation in the Structural Gene for Monoamine Oxidase A. *Science* 1993;262:578-580.
- 14 Lavedan C, Radvanyi HH, Shelbourne P, Rabes JP, Duros C, Savoy D, Dehaupas I, Luce S, Johnson K and Junien C: Myotonic Dystrophy: Size-and Sex-Dependent Dynamics of CTG Meiotic Instability, and Somatic Mosaicism. *American Journal of Human Genetics* 1993;52:875-883.
- 15 Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Fontali M, Folstein S, Ross C, Franz M, Abbot M, Gray J, Conneally P, Young A, Penney J, Hollingsworth Z, Shoulson I, Lazzarini A, Falek A, Koroshetz W, Sax D, Bird E, Vonsattel J, Bonilla E, Alvir J, Conde JB, Cha J-H, Dure L, Gomez F, Ramos M, Ramos JS, Snodgrass S, Young M, Wexler N, Moscovitz C, Penschaszadeh G, MacFarlane H, Anderson M, Jenkins B, Srinidhi J, Barnes G, Gusella J and MacDonald M: Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics* 1993;4:387-392.
- 16 Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, Telenius H, Adam S, Sajoo A, Starr E, Heiberg A, Wolff G and Hayden MR: Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. *Nature Genetics* 1993;5:174-179.
- 17 De Rooij KE, De Kooning Gans PAM, Skraastad MI, Belfrois RDM, Vegter-Van der Vlis M, Roos RAC, Van Ommen G-JB, Den Dunnen JT and Losekoot M: Dynamic mutation in Dutch Huntington's disease patients: increased paternal repeat instability extending to within the normal size range. *Journal of Medical Genetics* 1993;30:996-1002.
- 18 Ranen NG, Colin Stine O, Abbott MH, Sherr M, Codori A-M, Franz ML, Chao NI, Chung AS, Pleasant N, Callahan C, Kasch LM, Ghaffari M, Chase GA, Kazazian HH, Brandt J, Folstein SE and Ross CA: Anticipation and instability of IT-15 (CAG)_n repeats in parent-offspring pairs with Huntington disease. *American Journal of Human Genetics* 1995;57:593-602.
- 19 Chung M, Ranum LPW, Duvick LA, Servadio A, Zoghbi HY and Orr HT: Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nature Genetics* 1993;5:254-258.
- 20 Igarashi S, Takiyama Y, Cancel G, Rogaeva EA, Sasaki H, Wakisaka A, Zhou Y-X, Takano H, Endo K, Sanpei K, Oyake M, Tanaka H, Stevanin G, Abbas N, Durr A, Rogaev EI, Sherrington R, Tsuda T, Ikeda M, Cassa E, Nishizawa M, Benomar A, Julien J, Weissenbach J, Wang G-X, Agid Y, George-Hyslop PHS, Brice A and Tsuji S: 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. *Human Molecular Genetics* 1996;5:923-932.
- 21 Pianese L, Cavalcanti F, De Michele G, Filla A, Campanella G, Calabrese O, Castaldo I, Monticelli A and Coccozza S: The effect of parental gender on the GAA dynamic mutation in the FRDA gene. *American Journal of Human Genetics* 1997;60:460-463.
- 22 Yu S, Mulley J, Loesch D, Turner G, Donnelly A, Gedeon A, Hillen D, Kremer E, Lynch M, Pritchard M, Sutherland GR and Richards RI: Fragile-X syndrome: unique genetics of the heritable unstable element. *American Journal of Human Genetics* 1992;50:968-980.
- 23 Hunter A, Tsilfidis C, Mettler G, Jacob P, Mahadevan M, Surh L and Korneluk R: The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy. *Journal of Medical Genetics* 1992;29:774-779.
- 24 Eichler E, Holden J, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA and Nelson DL: Length of uninterrupted CGG repeats determines instability in the FMR-1 gene. *Nature Genetics* 1994;8:88-94.
- 25 Kunst CB and Warren ST: Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* 1994;77:853-861.
- 26 Snow K, Tester DJ, Kruckeberg KE and Thibodeau SN: Sequence analysis of the Fragile X trinucleotide repeat: correlations with stability and haplotype, and implications for the origin of Fragile X alleles. *American Journal of Human Genetics* 1994;55:A243.
- 27 Nancarrow JK, Holman K, Mangelsdorf M, Hori T, Denton M, Sutherland GR and Richards RI: Molecular basis of p(CGG)_n repeat instability at the FRA16A fragile site locus. *Human Molecular Genetics* 1995;4:367-372.
- 28 Leeflang EP and Arnheim N: A novel repeat structure at the myotonic dystrophy locus in a 37 repeat allele with unexpectedly high stability. *Human Molecular Genetics* 1995;4:135-136.
- 29 Loesch D, Huggins R, Petrovic V and Slater H: Expansion of the CGG repeat in Fragile x in the FMR1 gene depends on the sex of the offspring. *American Journal of Human Genetics* 1995;57:
- 30 Limprasert P, Nouri N, Heyman R, Nopparatana C, Kamonsilp M, Deininger PL and Keats BJB: 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. *Human Molecular Genetics* 1996;5:207-213.
- 31 Sambrook J, Fritsch EF and Maniatis T: *Molecular cloning: a laboratory manual*, 2nd. Cold Spring Harbor, N.Y., Cold Spring Harbor Laboratory Press, 1989.
- 32 Goto J, Watanabe M, Ichikawa Y, Yee S-B, Ihara N, Misumi Y, Igarashi S, Tsuji S, Nomura Y, Segawa M, Gaspar C, Maciel P, Rouleau GA and Kanazawa I: The usage of different stop codons by an alternative splicing and polymorphism in the Machado-Joseph disease gene.
- 33 Bowcock AM, Tomfohrde J, Weissenbach J, Bonne-Tamir B, St George-Hyslop P, Giagheddu M, Cavalli-Sforza LL and Farrer LA: Refining the position of Wilson disease by linkage disequilibrium with polymorphic microsatellites. *American Journal of Human Genetics* 1994;54:79-87.

- 34 Gaspar C, Lopes-Cendes I, DeStefano AL, Maciel P, Silveira I, Coutinho P, MacLeod P, Sequeiros J, Farrer LA and Rouleau GA: Linkage disequilibrium analysis in Machado-Joseph disease patients of different origins. *Human Genetics* 1996;98:620-624.
- 35 Strand M, Prolla TA, Liskay RM and Petes TD: Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 1993;365:274-276.
- 36 Kunkel TA: Slippery DNA and Diseases. *Nature* 1993;365:207-208.
- 37 Richards RA and Sutherland GR: Simple repeat DNA is not replicated simply. *Nature Genetics* 1994;6:114-116.
- 38 McMurray CT: Mechanisms of DNA expansion. *Chromosoma* 1995;104:2-13.
- 39 Wells RD: Molecular basis of genetic instability of triplet repeats. *Journal of Biological Chemistry* 1996;271:2875-2878.
- 40 Gacy AM, Goellner G, Juranic N, Macura S and McMurray CT: Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* 1995;81:533-540.
- 41 Stevanin G, Lebre A-S, Mathieux C, Cancel G, Abbas N, Didierjean O, Durr A, Trottier Y, Agid Y and Brice A: Linkage disequilibrium between the spinocerebellar ataxia 3/Machado-Joseph disease mutation and two intragenic polymorphisms, one of which, (X359Y), affects the stop codon. *American Journal of Human Genetics* 1997;60:1548-1552.
- 42 Gaspar C, Goto J, Lopes-Cendes I, Hayes S, Arvidsson K, Maciel P, Silveira I, Coutinho P, Stevanin G, DeStefano A, Reiss O, Sasaki H, Giunti P, Wood N, Brunt E, Tranebjaerg L, Nicholson G, Higgins JJ, Lauritzen M, Watanabe M, Volpini V, Hsieh M, Soong B-W, Wang G-X, Ranum L, Tsuji S, Brice A, Farrer LA, Sequeiros J and Rouleau GA: Different origins of mutation in Machado-Joseph disease patients. 1997,
- 43 Richards RI, Holman K, Friend K, Kremer E, Hillen D, Staples A, Brown WT, Goonewardena P, Tarleton J, Schwartz C and Sutherland GR: Evidence of founder chromosomes in fragile X syndrome. *Nature Genetics* 1992;1:257-260.
- 44 MacPherson JN, Bullman H, Youngs SA and Jacobs PA: Insert size and flanking haplotype in fragile X and normal populations: possible multiple origins for the fragile X mutation. *Human Molecular Genetics* 1994;3:399-405.
- 45 Imbert G, Kretz C, Johnson K and Mandel J-L: Origin of the expansion mutation in myotonic dystrophy. *Nature Genetics* 1993;4:72-76.
- 46 Neville C, Mahadevan MS, Barceló JM and Korneluk R: High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of myotonic dystrophy mutation. *Human Molecular Genetics* 1994;3:45-51.
- 47 Rubinsztein DC, Leggo J, Amos W, Barton DE and Ferguson-Smith MA: Myotonic dystrophy CTG repeats and the associated insertion/deletion polymorphisms in human and primate populations. *Human Molecular Genetics* 1994;3:2031-2035.
- 48 Squitieri F, Andrew SE, Goldberg YP, Kremer B, Spence N, Zeisler J, Nichol K, Theilmann J, Greenberg J, Goto J, Kanazawa I, Vesa J, Peltonen L, Almqvist E, Anvret M, Telenius H, Lin B, Napolitano G, Morgan K and Hayden MR: DNA haplotype analysis of Huntington disease reveals

- clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. *Human Molecular Genetics* 1994;3:2103-2114.
- 49 Rubinsztein DC, Leggo J, Goodburn S, Barton DE and Ferguson-Smith MA: Haplotype analysis of the delta-2642 and (CAG)_n polymorphisms in the Huntington's disease (HD) gene provides an explanation for the apparent "founder" effect. *Human Molecular Genetics* 1995;4:203-206.
- 50 Almqvist E, Chong S, Telenius H, La Tray L, Nichol K, Bourdelat-Parks B, Goldberg YP, Richards F, Sillence D, Greenberg C, Ives E, Van den Engh G, Hughes M and Hayden MR: Contribution of DNA sequence and (CAG)_n size to mutational frequencies of intermediate alleles (IAs) in Huntington disease (HD): evidence from single sperm analyses. *American journal of Human Genetics* 1996;59:A48.
- 51 Almqvist E, Spence N, Nichol K, Andrew SE, Vesa J, Peltonen L, Anvret M, Goto J, Kanazawa I, Goldberg YP and Hayden MR: Ancestral differences in the distribution of the delta-2642 glutamic acid polymorphism is associated with varying CAG repeat lengths in normal chromosomes: insights into the genetic evolution of Huntington disease. *Human Molecular Genetics* 1995;4:207-214.
- 52 Igarashi S, Takiyama Y, Takano H, K S, Oyake M, Tanaka H, Sasaki H, Wakisaka A, Nishizawa M, Horai S, Brice A, Roganova E, St. George-Hyslop P and Tsuji S: Single base polymorphism at the 3' end of the CAG repeat in MJD1 gene is associated with the size of CAG repeat and the intergenerational instability. *American Journal of Human Genetics* 1996;5:923-932.
- 53 Murray A, MacPherson JN, Pound MC, Sharrock A, Youings SA, Dennis NR, McKechnie N, Linehan P, Morton NE and Jacobs PA: The role of the size, sequence and haplotype in the stability of FRAXA and FRAXE alleles during transmission. *Human Molecular Genetics* 1997;6:173-184.
- 54 Yanagisawa H, Fujii K, Nagafuchi S, Nakahori Y, Nakagome Y, Akane A, Nakamura M, Sano A, Komure O, Kondo I, Jin DK, Sorensen SV, Potter NT, Young SR, Nakamura K, Nukina N, Nagao Y, Tadokoro K, Okuyama T, Miyashita T, Inoue T, Kanazawa I and Yamada M: A unique origin and multistep process for the generation of expanded DRPLA triplet repeats. *Human Molecular Genetics* 1996;5:373-379.
- 55 Takiyama Y, Sakoe K, Nakano I and Nishizawa M: Machado-Joseph disease: Cerebellar ataxia and autonomic dysfunction in a patient with the shortest known expanded allele (56 CAG repeat units) of the MJD1 gene. *Neurology* 1997;49:604-606.
- 56 Maciel P, Rousseau M, Gaspar C, Coutinho P, Sequeiros J and Rouleau GA: An intermediate allele of 51 repeats found at the Machado-Joseph disease (MJD) locus: implications for predictive testing and the study of the origin of the MJD mutation. 1997,
- 57 Warner JP, Barron LH, Fitzpatrick DR and Brock DJH: A gene conversion event at the Huntington's CAG repeat. *American Journal of Human Genetics* 1996;59:1697.
- 58 Nicol K, Goldberg YP, Almqvist E, Sillence D, Richards F and M.R. H: Interallelic conversion events associated with trinucleotide instability in the Huntington disease gene. *American Journal of Human Genetics* 1996;59:1590.

- 59 Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL and Armour JAL: Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics* 1994;6:136-145.
- 60 Monckton DG, Neumann R, Guram T, Fretwell N, Tamaki K, MacLeod A and Jeffreys AJ: Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nature Genetics* 1994;8:162-170.

III. DISCUSSÃO GLOBAL

III. DISCUSSÃO GLOBAL

A identificação da mutação causadora da doença de Machado-Joseph, a expansão de uma sequência repetitiva (CAG)_n no gene MJD1, que na proteína (ataxina-3) é traduzida como um segmento poliglutamina, veio permitir o diagnóstico molecular desta doença por detecção directa da mutação, aplicável na confirmação do diagnóstico clínico ou patológico da doença e ao teste preditivo e aconselhamento genético em famílias portadoras desta mutação.

ESTABELECIMENTO DE UM TESTE MOLECULAR PARA A DMJ

Tendo-se confirmado a presença de uma expansão do *repeat* CAG no gene MJD1 na grande maioria das famílias DMJ descritas, de origem portuguesa (artigo I) e de outras origens [Maciel *et al.*, 1995; Maruyama *et al.*, 1995; Burt *et al.*, 1996; Gadoth *et al.*, 1996; Lopes-Cendes *et al.*, 1996c; Watanabe *et al.*, 1996], uma das questões de maior relevância para o estabelecimento de um teste genético para diagnóstico da DMJ era a caracterização da mutação associada à DMJ, incluindo a definição dos tamanhos do *repeat* CAG, em indivíduos normais e em indivíduos afectados. O nosso estudo (artigo I) demonstrou que os indivíduos normais eram portadores de 12 a 37 tripletos CAG, enquanto os indivíduos afectados ou portadores assintomáticos tinham entre 62 e 84 unidades.

Os resultados combinados dos vários estudos (ver tabela 7), realizados em diferentes populações [Kawaguchi *et al.*, 1994; Cancel *et al.*, 1995; Haberhausen *et al.*, 1995; Maciel *et al.*, 1995; Maruyama *et al.*, 1995; Matilla *et al.*, 1995; Ranum *et al.*, 1995; Schols *et al.*, 1995; Stevanin *et al.*, 1995b; Takiyama *et al.*, 1995; Silveira *et al.*, 1996; Watanabe *et al.*, 1996; Soong *et al.*, 1997; Zhou *et al.*, 1997; Silveira *et al.*, submetido a publicação] confirmam em geral estas distribuições (tendo em conta possíveis diferenças de uma a duas unidades devidas a diferenças metodológicas na

Tabela 7. Definição dos tamanhos do *repeat* CAG na população normal e em doentes DMJ/SCA3 nos diversos estudos realizados até à data, em populações de origens étnicas diversas.

Referência	Origem da população	Controlos	Cromossomas DMJ
Kawaguchi, 1994	Japonesa	13-36	68-79
Maciel, 1995	Portuguesa	12-37	62-85
Takiyama, 1995	Japonesa	14-37	68-84
Maruyama, 1995	Japonesa	14-34	61-84
Ranum, 1995	Norte-americana (diversas origens)	14-40	68-82
Matilla, 1995	Norte-americana (diversas origens)	14-31	66-79
Cancel, 1995	Francesa	14-34	66-82
Stevanin, 1995	Brasileira	14-28	66-77
Haberhausen, 1995	Alemã	14-34	67-75
Silveira, 1996	diversas origens	-	66-81
Watanabe, 1996	Japonesa	13-33	65-78
Schöls, 1996	Alemã	14-30	67-80
Soong, 1997	Chinesa	14-39	63-81
Zhou, 1997	Chinesa	14-40	72-86

determinação dos tamanhos dos *repeats*): os cromossomas normais apresentam de 12 a 41 *repeats* CAG no gene MJD1, enquanto todos os indivíduos afectados ou portadores estudados têm entre 61 e 86 *repeats*.

Notória, em todos os estudos acima referidos, é a ausência de alelos "intermédios". Na DMJ não se verifica uma sobreposição de tamanhos entre os alelos normais e os alelos expandidos, nem foram descritas até à data transições de alelos normais de maior tamanho para alelos expandidos, levando a casos isolados da doença, como é o caso na doença de Huntington. Esta distância entre os tamanhos normais e expandidos facilita muito o diagnóstico molecular da DMJ. No entanto, foi muito recentemente identificado por nós um alelo de 51 unidades CAG a segregar (independentemente da mutação causadora da doença) numa pequena família DMJ [Maciel *et al.*, 1997]. Não é ainda claro se este alelo está ou não associado à doença, qual a sua frequência e se este achado poderá vir a levantar problemas sérios no diagnóstico preditivo e aconselhamento genético.

BASES MOLECULARES DA VARIABILIDADE CLÍNICA

Efeito do tamanho do *repeat* CAG expandido na idade de início da DMJ

Uma vez que todos os doentes com DMJ apresentam um mesmo tipo de mutação, a expansão da sequência repetitiva (CAG)_n no gene MJD1, levantava-se a questão de saber qual era a base molecular da variabilidade fenotípica característica desta doença. A resposta (parcial) está na diversidade de tamanhos do *repeat* CAG expandido, uma vez que se observou uma correlação inversa entre o tamanho deste e a idade de início da doença (artigo I). Esta correlação, observada também na doença de Huntington, SCA1, ADRPL e SCA2, indica que, qualquer que seja o mecanismo patogénico da expansão do *repeat* CAG/poli glutamina, o efeito tóxico desta sequência aumenta com o seu comprimento.

Tabela 8. Resultados obtidos nos diversos estudos de correlação entre o tamanho do *repeat* CAG expandido e a idade de início da DMJ/SCA3 realizados até à data, em populações de origem étnica diversa.

Referência	Origem da população	Número de casos estudados	Correlação c/ idade de início
Kawaguchi, 1994	Japonesa	11	$r^2=0.45$
Maciel, 1995	Portuguesa	156	$r^2=0.46$
Takiyama, 1995	Japonesa	116	$r^2=0.54$
Maruyama, 1995	Japonesa	83	$r^2=0.76$
Ranum, 1995	Norte-americana (diversas origens)	57	$r^2=0.70$
Matilla, 1995	Norte-americana (diversas origens)	16	$r^2=0.58$
Cancel, 1995	Francesa	32	$r^2=0.58$
Stevanin, 1995	Brasileira/ Portuguesa	28	$r^2=0.62$
Watanabe, 1996	Japonesa	20	$r^2=0.85$
Schols, 1996	Alemã	42	$r^2=0.67$
Soong, 1997	Chinesa	15	$r^2=0.79$
Zhou, 1997	Chinesa	25	$r^2=0.59$

Resultados similares foram obtidos por diferentes grupos (ver tabela 8), em populações de doentes DMJ de origens étnicas diversas [Kawaguchi *et al.*, 1994; Cancel *et al.*, 1995; Maciel *et al.*, 1995; Maruyama *et al.*, 1995; Matilla *et al.*, 1995; Ranum *et al.*, 1995; Stevanin *et al.*, 1995b; Takiyama *et al.*, 1995; Schols *et al.*, 1996; Silveira *et al.*, 1996; Watanabe *et al.*, 1996; Soong *et al.*, 1997; Zhou *et al.*, 1997; Silveira *et al.*, submetido a publicação]. Os coeficientes de correlação observados entre o tamanho do *repeat* e a idade de início da doença variavam entre -0.67 e -0.92, o que significa que apenas parte (45 a 85%) da variabilidade em termos de idade de início poderá ser explicada pela variação de tamanho do (CAG)_n expandido. Esta correlação incompleta pode ser devida a vários factores, quer metodológicos - tais como a imprecisão da medida do "tamanho do (CAG)_n expandido" (detectam-se múltiplas bandas, devido ao mosaicismo somático existente nos linfócitos dos portadores de expansões) e dificuldade de definição de "idade de início" (embora os seus critérios tenham sido claramente estabelecidos, o início da doença é muito insidioso e a informação, geralmente fornecida pelo doente ou familiares, tem um certo grau de subjectividade) - quer biológicos - o *repeat* é medido em linfócitos circulantes e a doença afecta sobretudo o sistema nervoso central, onde este pode ter tamanhos diferentes) [Maciel *et al.*, 1995] - mas pode significar também que outros factores contribuem para determinar a idade de início.

O efeito do tamanho do *repeat* CAG no gene MJD1 do cromossoma normal na variação da idade de início foi excluído na nossa população [Maciel *et al.*, 1995] e na população japonesa [Takiyama *et al.*, 1995]. O sexo do doente, avançado como possível factor determinante da idade de início por Kawaguchi e colaboradores [1995b] foi também excluído pelos nossos resultados (artigo II), bem como pelos de outros grupos [Durr *et al.*, 1995].

Outros factores moduladores da idade de início na DMJ

A etapa seguinte do estudo foi, assim, a procura de outros factores, além do tamanho do *repeat* CAG expandido no gene MJD1, que influenciassem a idade de início

da DMJ. Em particular, uma vez que havia sido identificado em famílias com ataxia espinocerebelosa tipo 1 um efeito familiar que explicava cerca de 5% da variabilidade na idade de início da doença [Ranum *et al.*, 1994], pensou-se que um efeito semelhante poderia existir na DMJ. De facto, as correlações positivas da idade de início da DMJ observadas entre pares de irmãos, primos e tios, após ajustamento para o tamanho do *repeat* CAG (artigo III) revelaram que um factor familiar, independente do *repeat* expandido, influenciava a idade de início da doença, e que este factor poderia ser genético. Embora a correlação negativa das idades de início entre pais e filhos, pudesse pôr em causa a possibilidade de se tratar de um factor genético, é possível que esta fosse apenas o resultado de um enviesamento (quase inevitável) da amostragem: o facto de se dispôr de dados relativos ao tamanho do *repeat* e à idade de início para um dado conjunto de pares progenitor-descendente significa, em geral, que se trata de progenitores com início mais tardio da doença (e, como tal, vivos até uma idade mais avançada) e descendentes com início mais precoce da doença (doentes ao mesmo tempo que os pais, ou quase), ou seja, pares em que a correlação de idades de início é previsivelmente baixa ou mesmo inversa.

O facto de os resultados obtidos com dois métodos estatísticos diversos, em que a variável dependente é tratada de forma completamente diferente, serem concordantes, parece mostrar que não se trata de uma artefacto da análise. Por outro lado, estes resultados não são de todo inesperados, uma vez que, antes mesmo da identificação do gene alterado na DMJ, a existência de modificadores genéticos ou ambientais que influenciassem a expressão da mutação principal fora já proposta por vários autores como explicação para a heterogeneidade de fenótipo e idade de início [Woods e Schaumburg, 1972; Bharucha *et al.*, 1986; Sequeiros, 1989], embora não se previsse a variabilidade possível da mutação do gene principal, por serem desconhecidas à data as expansões de trinucleotídeos como causa de doenças genéticas.

Homozigotia e idade de início

Embora não seja necessária a homozigotia para explicar idades de início muito precoces da DMJ, uma vez que a presença de um *repeat* expandido de tamanho extremo pode originar um início juvenil da doença (artigo I), observou-se que os poucos casos de homozigotos para a mutação DMJ confirmados por teste molecular [Takiyama *et al.*, 1995; Gadoth *et al.*, 1996], tinham de facto idades de início mais precoces e formas particularmente graves da doença, sugerindo que a mutação DMJ, ao contrário da mutação associada à doença de Huntington, não tem um efeito dominante puro. Esta diferença entre as duas doenças é difícil de explicar, assumindo um mecanismo patogénico comum.

No caso, já referido, de um indivíduo portador de um alelo expandido (com 71 unidades CAG) e de um alelo "intermédio" com 51 unidades CAG, a idade de início não era no entanto especialmente precoce (32 anos) e a apresentação clínica era típica do tipo 2, sem a espasticidade e a distonia marcadas que caracterizam os casos juvenis, o que poderia interpretar-se como evidência de que o alelo de 51 *repeats* é de facto um alelo intermédio, sem o efeito patológico de um alelo expandido. No entanto, será necessário confirmar por sequenciação que este alelo não tenha uma interrupção do *repeat* CAG ou outra variação na sua sequência que lhe retire o potencial patogénico.

Ritmo de progressão da DMJ e tamanho do *repeat* CAG expandido

Estudos quantitativos da história natural da SCA1, SCA2 e DMJ (designada SCA3 pelos autores) [Klockgether *et al.*, 1997], revelaram que, nestas ataxias, o tamanho do *repeat* tem um impacto considerável não só na idade de início dos sintomas, mas também no ritmo de progressão da doença e na sobrevivência. O impacto era variável entre as três doenças, observando-se que na SCA2 o aumento do ritmo de progressão por cada unidade CAG adicional no *repeat* expandido era maior que na DMJ e na SCA1.

Tamanho do *repeat* CAG expandido e subfenótipo

Como foi já referido, observa-se na DMJ, além da diversidade em termos de idade de início da doença, uma grande heterogeneidade da apresentação clínica, com quadros muito variáveis, em que predominam diferentes tipos de sintomas, e a que podem corresponder uma gravidade e ritmo de progressão muito diferentes. Esta variabilidade foi classificada em 3 subtipos [Coutinho e Andrade, 1978; Coutinho, 1992], descritos atrás (ver Introdução). A identificação do subtipo clínico é extremamente útil para estudos da diversidade de expressão da doença, uma vez que é definida uma variável que integra a idade de início, os sintomas principais e secundários da doença e o ritmo de progressão desta.

No nosso estudo de correlações entre o tamanho do *repeat* e a apresentação clínica (artigo I) observou-se uma correlação entre o tamanho do (CAG)_n expandido e o subtipo, com diferenças significativas entre o subtipo 1 e os subtipos 2 e 3. Embora se observasse uma grande sobreposição entre os subtipos em termos de tamanhos do *repeat*, os doentes com tipo 2 e 3 apresentavam tamanhos médios do *repeat* menores que os doentes com o tipo 1.

Resultados semelhantes foram obtidos por outros grupos [Maruyama *et al.*, 1995; Schols *et al.*, 1996; Watanabe *et al.*, 1996]: observaram-se diferenças em termos de tamanhos do *repeat* entre os subtipos e, no estudo por Schöls e colaboradores [1996], em que se fez uma análise quantitativa, observaram-se diferenças significativas no tamanho médio do *repeat* entre o subtipo 1 e os subtipos 2 e 3, confirmando assim os nossos resultados. Num estudo algo diferente [Takiyama *et al.*, 1995], em que se compararam as frequências de determinados sintomas entre grupos de doentes DMJ, definidos em termos de tamanho do alelo expandido, observou-se que, dos três sintomas mais frequentes na DMJ, a ataxia e a oftalmoplegia externa progressiva (comum aos três tipos) se distribuíam igualmente por todos os grupos, assim como o nistagmo, enquanto os sinais piramidais (mais pronunciados no tipo 1) e a pseudoexoftalmia estavam mais presentes nos portadores de alelos maiores ($p < 0.001$ e $p < 0.05$, respectivamente). Detectavam-se ainda

tendências, estatisticamente não significativas, para o predomínio de fasciculações da face e de distonia (predominantes no tipo 1) nos portadores de alelos maiores e de atrofia muscular (característica do tipo 3) nos portadores de alelos menores. Estas distribuições de sintomas correspondem pois exactamente às observadas entre os subtipos [Coutinho, 1994].

Além dos três subtipos principais definidos por Coutinho e Andrade, foi definido um quarto subtipo, com predominância de um quadro parkinsoniano, em doentes geralmente mais velhos [Rosenberg e Fowler, 1981]; embora alguns doentes com este quadro tenham sido confirmados molecularmente como portadores da mutação DMJ [Tuite *et al.*, 1995], o tipo 4 é muito raro e a sua existência real permanece controversa.

Uma questão que se levanta, face a este resultados, é a de saber de que forma o tamanho do *repeat* CAG expandido no gene MJD1 determina as manifestações clínicas da DMJ. O efeito do tamanho do *repeat* no envolvimento patológico de diferentes vias neuronais, com as correspondentes manifestações clínicas, pode ser apenas indirecto, mediado pelo seu efeito na idade de início da doença: conforme sugerido por Coutinho [1992], a correlação entre a idade de início e o subtipo parece reflectir a susceptibilidade diferencial das estruturas do SNC ao efeito tóxico da mutação ao longo da vida do indivíduo; assim, o tamanho do *repeat*, ao determinar a idade de início da doença (entendendo-se, neste caso, doença como patologia celular, antes ainda das manifestações clínicas), iria determinar quais as estruturas do SNC afectadas. Alternativamente, o tipo de células afectadas poderia depender directamente do tamanho do *repeat* expandido (devido a diferentes limiares de tolerância em diferentes populações celulares ao efeito tóxico da mutação).

Redefinição nosológica e fenotípica das ataxias espinocerebelosas

A identificação dos genes associados a formas diversas de ataxia espinocerebelosa (ver tabela 9) veio abrir a possibilidade de redefinição da história natural e de reavaliação dos critérios de diagnóstico clínico destas doenças, com base na identificação segura das

diferentes mutações através dos testes moleculares, agora disponíveis. Os estudos de rastreio molecular de ataxias levados a cabo em vários países são de particular utilidade para este fim [Cancel *et al.*, 1995; Matilla *et al.*, 1995; Ranum *et al.*, 1995; Schols *et al.*, 1995; Silveira *et al.*, 1996; Soong *et al.*, 1997], ao permitirem o estudo de famílias que não seriam identificadas de outra forma. Torna-se necessário, no entanto, um trabalho de análise aprofundada dos fenótipos (a nível clínico e patológico), uma vez determinada a mutação causadora da doença em cada família, e a comparação do espectro fenotípico completo das diferentes mutações. Só assim se poderão tirar conclusões sobre as diferenças existentes entre estas, com aplicação clínica a nível do diagnóstico e do aconselhamento genético, bem como na compreensão das alterações fisiológicas que levam à doença. Embora não existam ainda muitos trabalhos publicados nesse campo e a maioria dos estudos esteja ainda a decorrer, há já dados que apontam para algumas conclusões interessantes.

Um dos pontos a realçar é que a avaliação clínica e/ou patológica não é sempre suficiente para o diagnóstico quer da DMJ quer de outras formas de ataxia, dada a sobreposição fenotípica entre estas, sendo por isso fundamental a realização do teste molecular para um diagnóstico de certeza [Lopes-Cendes *et al.*, 1996b]. Uma das formas de ataxia mais frequentemente confundidas com a DMJ a nível clínico é a SCA2, por ser uma mutação bastante frequente e de apresentação clínica variável (semelhante em alguns casos à da DMJ), tendo sido descritos doentes, portadores da expansão no gene SCA2, que apresentam, além da ataxia e oftalmoparésia, fasciculações faciais e retracção palpebral [Geschwind *et al.*, 1997], características que são mais frequentemente associadas à DMJ.

Como exemplo da redefinição das características patológicas da DMJ, estudou-se recentemente uma pequena família originária das Caraíbas, na qual o diagnóstico clínico da DMJ não fora possível, já que os três doentes da família apresentavam quadros clínicos inespecíficos, dificultando o diagnóstico diferencial com os outros tipos de ataxia espinocerebelosa [Lopes-Cendes *et al.*, 1996b]. O teste molecular mostrou a presença da mutação DMJ nos três doentes, um dos quais faleceu após 35 anos de doença. A autópsia

Tabela 9. Localização cromossômica e identificação dos genes associados às diversas formas de ataxia espinocerebelosa autossômica dominante (classificação genética das ataxias); n.i. = não identificado.

Designação	Localização cromossômica	Gene mutado	Mutação
SCA1	6p	SCA1	CAG expandido
SCA2	12q	SCA2	CAG expandido
MJD/SCA3	14q	MJD1	CAG expandido
SCA4	16q	n.i.	n.i.
SCA5	11q	n.i.	n.i.
SCA6	19p	CACNL1A4	CAG expandido (pequena expansão)
SCA7	3p	SCA7	CAG expandido

deste indivíduo mostrou atrofia marcada dos núcleos pônticos, pedúnculos cerebelosos médios, núcleos dos nervos cranianos motores, cornos anteriores da medula e colunas de Clarke e atrofia moderada dos núcleos subtalâmicos e do *locus niger*, além de atrofia moderada do córtex cerebeloso e das olivas bulbares, geralmente consideradas como não envolvidas na DMJ.

No âmbito da reavaliação dos fenótipos associados às diferentes mutações, será também interessante referir alguns (ainda poucos) estudos quantitativos de diversos aspectos do fenótipo, efectuados após a identificação molecular das famílias como DMJ ou outra forma de ataxia, em particular: a) o estudo dos potenciais motores evocados em doentes com ataxia espinocerebelosa tipo 1 ou 3 (DMJ) por Schöls e colaboradores [1997], em que se demonstrou que os estudos electrofisiológicos constituem um método rápido, não doloroso e não invasivo, com utilidade na diferenciação das diversas formas de ataxia espinocerebelosa, b) os estudos do metabolismo cerebral por tomografia de emissão de positrões em famílias chinesas com a mutação DMJ [Soong *et al.*, 1997] e c) os estudos comparativos entre SCA1 e SCA3/DMJ, por posturografia e electronistagmografia [Subramony e King, 1997].

BASES MOLECULARES DA ESPECIFICIDADE DE MORTE NEURONAL NA DMJ

Exclusão do mosaicismo somático do CAG expandido no gene MJD1 como causa da especificidade da degenerescência na DMJ

Conforme foi já discutido (ver Introdução), não é ainda compreendido o mecanismo pelo qual a expansão de um segmento $(CAG)_n$, acima de um determinado limiar (geralmente cerca de 40 unidades CAG), leva à morte neuronal precoce nas diversas doenças associadas a este tipo de mutação. Embora se assuma que um mecanismo patológico comum está na base da degenerescência neuronal em todas estas

doenças, é evidente que os grupos de células afectadas divergem de doença para doença. Torna-se então necessário esclarecer a diferente susceptibilidade das estruturas do SNC a um mesmo tipo de mutação, quando presente em genes diferentes. Uma vez que a expressão do gene MJD1 é ubíqua a nível do ARN_m [Nishiyama *et al.*, 1996] e da proteína [Paulson *et al.*, 1997a], tal como acontece na maioria dos genes das doenças associadas à expansão de uma sequência repetitiva (CAG)_n, há que procurar explicações alternativas à especificidade de expressão, como base molecular da especificidade da morte celular no SNC.

Os nossos resultados (artigo IV) permitiram excluir o papel do mosaïcismo somático, em termos de tamanho do *repeat* CAG, como base da morte neuronal diferencial no SNC de indivíduos portadores de expansões no gene MJD1 (assim como na SCA1): a ausência de correlação entre o mosaïcismo somático e o padrão de degenerescência neuronal, assim como a semelhança entre os padrões de mosaïcismo observados entre a DMJ e a SCA1, apesar das diferenças a nível patológico entre estas duas doenças, sugere que o padrão de mosaïcismo somático depende apenas da composição celular das diferentes estruturas e não é a causa da especificidade de morte celular. Resultados semelhantes foram encontrados por outros investigadores [Hashida *et al.*, 1997] em doentes com DMJ.

Assim, deverão existir factores celulares específicos, que conferem às estruturas do SNC uma susceptibilidade diferencial ao efeito tóxico da mutação DMJ. Uma possibilidade seria a expressão variável das formas alternativas de *splicing* do gene MJD1, a nível das diversas estruturas do SNC, que poderia estar na base desta susceptibilidade diferencial (se a uma das formas de *splicing* estivesse associada uma maior toxicidade para a célula). Por outro lado, existem provavelmente outras proteínas que, interagindo com a ataxina-3 (numa situação normal ou apenas se a proteína contiver a sequência poliglutamina expandida), poderão desempenhar um papel na toxicidade da proteína mutante. Entre estas proteínas poderão existir algumas que, sendo de expressão restrita a certos tipos celulares, iriam definir a especificidade da degenerescência neuronal.

Interações proteicas como mecanismo determinante da morte celular na DMJ

A identificação das proteínas que interactivam com a ataxina-3 será pois da maior relevância, quer para o esclarecimento das funções normais da proteína na célula, quer para conhecer o mecanismo molecular que aumenta a susceptibilidade de grupos particulares de neurónios à presença de uma expansão poliglutamina nesta proteína. A GAPDH e as diversas isoformas da transglutaminase são candidatas possíveis, que ocorrem de imediato (ver Introdução - Modelos propostos para o mecanismo patogénico dos *repeats* CAG/poliglutamina expandidos), embora a GAPDH seja, dado o seu papel central no metabolismo energético, de expressão generalizada. Em outras doenças associadas à expansão de poliglutaminas, foram já identificados alguns "parceiros" moleculares das proteínas principais: a ataxina-1 interactiva consigo própria [Burrigh et al., 1997] e com a GAPDH [Koshy et al., 1996]; a huntingtina interactiva com uma proteína designada HAP1 (*huntingtin associated protein 1*) [Li et al., 1995], com a GAPDH [Burke et al., 1996] e com a calmodulina [Bao et al., 1996]; a atrofina com a GAPDH [Burke et al., 1996]; no entanto, a expressão destas proteínas associadas não parece, em nenhum dos casos, localizar-se especificamente nos tecidos mais afectados em cada uma das doenças (SCA1, doença de Huntington e atrofia dentatorubro-palidoluysiana, respectivamente).

A possibilidade de outros factores celulares específicos contribuírem para o mecanismo patológico das poliglutaminas expandidas está também de acordo com as observações a nível genético, já discutidas, que sugerem a existência de genes modificadores do fenótipo causado pelo gene mutante principal, com influência na idade de início da doença e na sua apresentação clínica: é possível que polimorfismos nos genes que codificam para as proteínas que interactivam com a ataxina-3 tenham efeito modulador da toxicidade da proteína mutada, portadora de uma poliglutamina expandida, determinando assim parcialmente a idade de início e a apresentação clínica da doença.

Caso o efeito das expansões da sequência (CAG)_n ocorresse, conforme sugerido por McLaughlin e colaboradores [1996], a nível do ARN codificado pelo gene MJD1, por interacção deste com proteínas de ligação ao ARN, a susceptibilidade diferencial dos diferentes tipos celulares poderia depender quer dos níveis expressão das proteínas em questão nas células, quer da sua interacção diferencial com os genes *downstream*, ou seja, com os genes cuja expressão fosse regulada por estas proteínas.

Mosaicismo somático do *repeat* CAG expandido: contribuições dos diferentes tipos celulares e progressão ao longo da vida

O mosaicismo somático do *repeat* CAG expandido no gene MJD1 [Maciel *et al.*, 1995; Hashida *et al.*, 1997] reflecte-se, por um lado, na observação de múltiplas bandas de tamanhos diferentes para o alelo expandido (e não para o alelo normal), quando se leva a cabo a amplificação do *repeat* num dado tecido, correspondentes às sub-populações celulares portadoras de *repeats* expandidos de tamanhos diferentes presentes nesse tecido, e, por outro lado, nas diferenças de tamanho observadas entre os vários tecidos. É interessante notar que o padrão de mosaicismo no sistema nervoso central nos *specimens* de doentes DMJ estudados, que, como se referiu, não correspondia ao padrão de degenerescência neuronal, era semelhante ao observado em doentes com SCA1 ou ADRPL [Maciel *et al.*, 1995; Lopes-Cendes *et al.*, 1996a; Hashida *et al.*, 1997], embora a amplitude de tamanhos fosse maior na ADRPL e SCA1 do que na DMJ (reflectindo maior instabilidade mitótica). Nas três doenças, tal como fora observado para a doença de Huntington [Telenius *et al.*, 1994], o cortex cerebeloso evidencia um tamanho médio da expansão menor do que as outras estruturas do SNC e uma amplitude de tamanhos menor (menor grau de mosaicismo somático).

A hipótese avançada para explicar esta observação [Chong *et al.*, 1995] foi que o cortex cerebeloso é proporcionalmente mais rico em neurónios (células pós-mitóticas) do que outras estruturas e que a principal contribuição para o mosaicismo somático no SNC viria, não dos neurónios, mas das células da glia, que continuam a dividir-se ao longo da

vida do indivíduo: a instabilidade mitótica do *repeat* expandido originaria então células com diferentes tamanhos do (CAG)_n, levando a um maior grau de mosaicismos (e a *repeats*, em média, maiores, dada a maior tendência para a expansão observada para a maior parte dos *repeats* CAG) nas estruturas com maior actividade mitótica.

Um corolário desta teoria seria que, nessas estruturas, o grau de mosaicismos aumentaria ao longo da vida do indivíduo. Esta progressão do grau de mosaicismos com a idade foi observada, de facto, para a ADRPL, no estudo de Takano e colaboradores [1996], em que se detectou que os indivíduos portadores de expansões do (CAG)_n cuja morte ocorrera mais precocemente apresentavam um grau de mosaicismos menor do que os indivíduos falecidos em idade mais tardia. Esta correlação inversa entre a idade de morte e o grau de mosaicismos não foi no entanto replicada na DMJ, nem na SCA1 (artigo V), sugerindo que o *repeat* CAG expandido poderá ter um comportamento algo diferente consoante a sequência génica em que se integra.

Por outro lado, estudos de análise mais directa da contribuição dos diferentes tipos celulares para o mosaicismos somático na ADRPL, por dissecação de estruturas do SNC [Takano *et al.*, 1996; Hashida *et al.*, 1997], revelaram diferenças significativas entre o cortex e a substância branca, quer a nível do cérebro quer do cerebelo, o que leva a crer que são de facto as células não-neuronais que contribuem maioritariamente para o mosaicismos no SNC nesta doença. Num estudo semelhante realizado por nós em *specimens* de cérebro de doentes com DMJ e SCA1 (artigo V) não foram no entanto detectadas diferenças significativas entre o cortex cerebral e a matéria branca, parecendo confirmar-se, assim, o comportamento diverso do *repeat* CAG nestas doenças.

Em ratinhos transgénicos portadores de expansões do (CAG)_n no gene da doença de Huntington [Mangiarini *et al.*, 1996; Mangiarini *et al.*, 1997] observou-se também mosaicismos somático, em termos de tamanho do *repeat* expandido, com aumento nítido do grau de mosaicismos ao longo do desenvolvimento pós-natal na generalidade dos tecidos, com excepção do cerebelo, coração e baço, que apresentavam um grau de mosaicismos constante ao longo da vida do ratinho, resultando em *repeats* em média menores nas fases mais tardias.

Na ataxia de Friedreich, associada à expansão de um outro tipo de *repeat* - (GAA)_n, observa-se, em geral, um grau de mosaïcismo somático mais acentado do que nas doenças associadas a *repeats* CAG; também em portadores desta mutação é observada uma diferença significativa entre o cortex cerebeloso e outras estruturas [Montermini *et al.*, 1997]. Neste caso, ao contrário do que se observa nas outras doenças, parece haver uma tendência predominante para a contracção do *repeat* GAA durante as divisões celulares, com maior número de bandas de menor tamanho nas estruturas mitoticamente mais activas. Em conformidade, o cerebelo apresenta um maior tamanho médio do *repeat* do que outras estruturas, bem como um menor grau de mosaïcismo.

BASES MOLECULARES DO FENÓMENO DE ANTECIPAÇÃO: INSTABILIDADE INTERGERACIONAL DO REPEAT CAG EXPANDIDO

Instabilidade genética do *repeat* CAG expandido no gene MJD1: correlações com a antecipação

O estudo molecular da variação intergeracional do tamanho do (CAG)_n expandido na nossa população de portadores da mutação DMJ (artigo I) demonstrou que existia instabilidade em cerca de 55% das transmissões do gene mutado, na sua maior parte expansões (80%). A variação média de tamanho era de +0.84 unidades CAG, indicando uma tendência geral do *repeat* expandido para aumentar de tamanho na população, que se reflectiria (dada a correlação inversa entre o tamanho do *repeat* e a idade de início) num efeito global de antecipação. No entanto, seriam de esperar também (cerca de 20 %) casos com antecipação negativa, dado que também se observam contracções do *repeat*. Seria também de esperar uma correlação entre a variação do tamanho do *repeat*, entre progenitores e descendentes afectados, com a antecipação; no entanto, esta correlação não era significativa na nossa população (artigo I), nem em outras populações [Brunt *et al.*,

1996], o que pode ser devido apenas ao pequeno tamanho das amostras e aos inevitáveis enviesamentos de amostragem [Sousa *et al.*, 1997].

São pois necessários estudos adicionais para se chegar a uma compreensão mais aprofundada das bases moleculares da antecipação na DMJ, em particular em modelos animais (ver Introdução), em que se poderão possivelmente ultrapassar muitas das dificuldades inerentes ao estudo de doenças de início tardio em humanos. Por outro lado, é necessário aprofundar o conhecimento dos mecanismos moleculares que levam à instabilidade genética dos *repeats* de trinucleotídeos.

Factores que contribuem para a instabilidade genética do (CAG)_n expandido na DMJ

Dos resultados obtidos na nossa população de portadores de uma expansão no gene DMJ (artigos I e VI) e noutras populações, surgem como principais factores contribuintes para a instabilidade intergeracional do *repeat* CAG (1) o sexo do progenitor [Maciel *et al.*, 1995; Maruyama *et al.*, 1995; Igarashi *et al.*, 1996a], com as transmissões paternas associadas a variações mais amplas (quer expansões quer contracções) do *repeat* e (2) um polimorfismo intragénico situado imediatamente após o *repeat* CAG, cuja variante no cromossoma normal parece interferir com a instabilidade do *repeat* expandido no cromossoma DMJ [Igarashi *et al.*, 1996a], por mecanismos ainda não compreendidos.

Este intrigante efeito do polimorfismo no cromossoma normal sobre a instabilidade do *repeat* expandido foi confirmado na nossa população (artigo VI) e em estudos do esperma de portadores de expansões no gene MJD1 [Takiyama *et al.*, 1997]; os estudos de esperma têm a vantagem de permitir analisar milhões de meioses, em contraste com o número limitado de meioses acessível em estudos familiares.

De acordo com os nossos resultados (artigo VI), o efeito detectado restringe-se aparentemente aos polimorfismos $\underline{C}^{987}\underline{GG}/\underline{G}^{987}\underline{G}$ e $\underline{TAA}^{1118}/\underline{TAC}^{1118}$, situados na região proximal a 3' do *repeat*, não se tratando, pois, de um efeito cromossómico mais alargado, uma vez que o outro polimorfismo intragénico estudado, a substituição

$\underline{A}^{669}\text{TG}/\underline{G}^{669}\text{TG}$, situada a montante (5') do *repeat* CAG, não parece ter efeito na instabilidade intergeracional deste.

Por outro lado, as variações de tamanho observadas durante a transmissão do CAG, normal ou expandido, ocorrem provavelmente no extremo 3' do *repeat*, uma vez que a interrupção do *repeat* CAG pela sequência CAAAAG (ver figuras 2.a e 2.b), se situa na mesma posição relativa, no extremo 5' do *repeat*, em todos os indivíduos [Rubinsztein *et al.*, 1995a; Limprasert *et al.*, 1996].

Estas observações são de certa forma paralelas às descritas para um minisatélite localizado no cromossoma 1q42-43 [Jeffreys *et al.*, 1994; Monckton *et al.*, 1994], em que, por análise da estrutura interna dos diversos alelos do minisatélite, se identificou uma forte tendência para a variabilidade se concentrar numa das extremidades da sequência repetitiva, a extremidade 5'. O tipo de alterações observadas na sequência do minisatélite indicava o envolvimento de trocas interalélicas no processo de mutação deste, com transferência de blocos de sequência repetitiva de um alelo para o outro durante a meiose, mas a ausência de alterações nas sequências flanqueantes do minisatélite implicava que um mecanismo de conversão génica complexo (e não uma troca desigual entre alelos) estaria na base da variação de sequência no seio do minisatélite [Jeffreys *et al.*, 1994]. Curiosamente, foi identificado um polimorfismo flanqueante (do lado 5') do minisatélite, uma transversão de G para C, a que estava associado um efeito de supressão das mutações [Monckton *et al.*, 1994], cujo efeito se fazia sentir em *cis*, mas não em *trans*. Por outro lado, foi descrita a ocorrência de conversões génicas no gene da doença de Huntington [Nicol *et al.*, 1996; Warner *et al.*, 1996], sugerindo que este mecanismo pode também estar envolvido na instabilidade dos *repeats* CAG expandidos. A comprovar-se, esse mecanismo poderia explicar o efeito de sequências presentes no cromossoma normal na instabilidade genética do cromossoma mutante na DMJ.

Instabilidade genética do *repeat* CAG: pistas possíveis para a origem da mutação DMJ

O estudo dos três polimorfismos intragênicos do gene MJD1, $\underline{C}^{987}\text{GG}/\underline{G}^{987}\text{GG}$, $\text{TAA}^{1118}/\text{TAC}^{1118}$ e $\underline{A}^{669}\text{TG}/\underline{G}^{669}\text{TG}$, numa população controle de origem açoriana (artigo VI), revelou que, em geral, os alelos dos polimorfismos mais frequentemente associados aos $(\text{CAG})_n$ expandidos neste gene apresentavam também uma frequência relativa algo aumentada nos alelos normais com maior tamanho do $(\text{CAG})_n$, o que poderia ser compatível com a hipótese de um efeito desestabilizador do *repeat* por parte das sequências flanqueantes deste. No entanto, não se observava uma associação exclusiva, nem mesmo uma predominância, de nenhum alelo dos polimorfismos entre os cromossomas com $(\text{CAG})_n$ de maior tamanho, em contraste com observações prévias, em populações de outras origens étnicas [Igarashi *et al.*, 1996a; Stevanin *et al.*, 1997]. A comparação dos tamanhos do $(\text{CAG})_n$ do gene MJD1 associados aos diferentes haplótipos intragênicos definidos pelos polimorfismos (na população normal) não revelou uma diferença significativa entre estes, não existindo nenhum haplótipo preferencialmente associado aos *repeats* de maior tamanho nos cromossomas normais.

Estes resultados estão em contraste com os descritos para o *repeat* CAG no gene da doença de Huntington [Squitieri *et al.*, 1994; Almqvist *et al.*, 1995; Rubinsztein *et al.*, 1995b], em que se verificaram diferenças na distribuição de tamanhos do $(\text{CAG})_n$ relativamente às variantes do polimorfismo de inserção/delecção $\Delta 2642$ ácido glutâmico, com os cromossomas portadores de uma delecção associados a maiores *repeats*, sendo a delecção claramente sobre-representada também nos cromossomas da doença.

Tomados em conjunto com a ausência de um efeito do haplótipo em *cis* na instabilidade do *repeat* expandido (artigo VI), os resultados obtidos parecem excluir a possibilidade de um efeito desestabilizador do *repeat* por parte das sequências gênicas flanqueantes como componente determinante para a origem da mutação DMJ nas diversas populações em que foi descrita esta entidade clínica.

DESENVOLVIMENTO DO TESTE PREDITIVO DA DMJ

Como implicação destes trabalhos, resta discutir a aplicação prática do conhecimento das bases moleculares das características clínicas, patológicas e genéticas da DMJ, em particular no que respeita à única via de intervenção contra esta doença a que é possível recorrer desde já: os testes preditivo e pré-natal para o aconselhamento genético dos indivíduos em risco.

Em particular, interessa aqui abordar a relevância para o diagnóstico molecular de algumas das conclusões deste estudo e discutir brevemente possíveis aplicações de resultados desta investigação em situações concretas, como a resolução de resultados ambíguos ou outras questões complexas no aconselhamento genético da DMJ e outras ataxias hereditárias.

Interpretação do resultado do teste molecular: tamanhos dos (CAG)_n expandidos

A grande separação existente entre as amplitudes normal e de expansão do (CAG)_n é em geral reconfortante em termos da segurança dos resultados da detecção directa da mutação. No entanto, conforme se referiu acima, foram recentemente identificados três indivíduos de uma família portuguesa com um alelo de tamanho intermédio (51 unidades CAG) [Maciel *et al.*, 1997], o que levanta novos problemas em termos de aconselhamento, porque se desconhece ainda se a este tamanho de *repeat* estará ou não associado o fenótipo da doença. Estão já a decorrer estudos mais detalhados desta família, incluindo a análise de um maior número de indivíduos portadores deste alelo, a sequenciação da sequência repetitiva para determinar a sua estrutura exacta e o seguimento clínico desses indivíduos, para se tentar esclarecer esta questão.

Por outro lado, o tamanho do alelo expandido não permite por si só um prognóstico preciso em termos da idade de início ou gravidade da doença; como exemplo, apenas com base num modelo de regressão (usando o coeficiente de correlação obtido) e

para um intervalo de confiança de 95%, poderá prever-se, para um indivíduo portador de um alelo com 67 unidades CAG, o início da doença entre os 35 e os 68 anos de idade. Espera-se que, uma vez identificados os factores (genéticos ou ambientais) modificadores da idade de início e da apresentação clínica, estes possam ser incorporados num modelo preditivo mais adequado, que permita o estabelecimento de um prognóstico quanto à idade de aparecimento e à gravidade das manifestações clínicas.

Homoalelismo para o alelo normal: possibilidades de eliminação da incerteza do resultado do teste molecular

A obtenção de um resultado de "homozigotia" para o alelo normal, ou seja, a detecção de dois alelos normais do mesmo tamanho (mais correctamente chamado homoalelismo), levanta algumas incertezas no caso dos *repeats* de trinucleotídeos expandidos, uma vez que é sempre possível não se ter conseguido amplificar um *repeat* expandido devido ao seu tamanho excessivamente grande (dada a competição entre os dois alelos na mistura de reacção de PCR, em que os produtos de menor tamanho são cineticamente favorecidos), ou mesmo, como foi já descrito na doença de Huntington, pela existência de polimorfismos na região de emparelhamento dos *primers* na reacção de PCR.

É fundamental excluir esta possibilidade, nos limites possíveis da tecnologia disponível, a fim de reduzir a incerteza do resultado a transmitir ao consultando. A primeira estratégia a utilizar para este fim é, evidentemente, o estudo de outros membros da família, em particular a determinação dos tamanhos do (CAG)_n nos progenitores, para se determinar se o homoalelismo é compatível com esses resultados. Em segundo lugar, pode proceder-se ao estudo de polimorfismos intragénicos no indivíduo: se os alelos apresentarem o mesmo tamanho do *repeat* CAG, mas estiverem associados a diferentes haplótipos intragénicos, é assim possível distinguir os dois alelos do gene.

Finalmente, estão em desenvolvimento métodos alternativos de detecção da mutação DMJ por *Southern blotting*, que, embora mais trabalhosos e necessitando uma

maior quantidade de ADN, permitiriam ultrapassar as dificuldades da amplificação por PCR em situações como esta. Embora seja pouco provável que na DMJ haja alelos mutados de tamanho suficientemente grande para não serem detectados por PCR, só fazendo esta análise por *Southern blotting* caso a caso se poderia eliminar por completo esta possibilidade. Este tipo de método teria também grande aplicabilidade em outras situações e seria ainda mais necessário nas ataxias dominantes em que a variabilidade populacional dos alelos normais é muito menor (ou seja, o homoalelismo é mais frequente), de que são exemplos a SCA2 e a SCA6.

Comportamento não-mendeliano da mutação DMJ: relevância para o aconselhamento genético

Além das dificuldades já discutidas (estabelecimento de um prognóstico para os portadores de expansões e situações de homoalelismo), há outras dificuldades inerentes ao aconselhamento genético nas doenças associadas a *repeats* expandidos, dado que a mutação é instável e o seu comportamento durante a transmissão de uma geração para a seguinte não é previsível.

A compreensão dos mecanismos de instabilidade genética e dos factores que a regulam poderá também vir a ser utilizável no aconselhamento genético. Um exemplo desta aplicação vem da doença de Huntington, em que há muito se conhece a existência de alelos intermédios (alelos cujo tamanho se situa na gama de sobreposição entre o normal e o expandido, não estando eles próprios associados à doença, mas cuja instabilidade genética leva à expansão do *repeat* para tamanhos maiores na descendência). Muitos destes alelos surgem em famílias em que se tinham identificado casos isolados de doença de Huntington. Foi identificado um haplótipo que, quando presente em *cis* de *repeats* intermédios, originava uma instabilidade aumentada destes, com maior frequência das transições para CAGs expandidos [Chong *et al.*, 1997]. Este efeito foi identificado quer a nível da descendência dos indivíduos portadores, quer em estudos de esperma destes indivíduos por *single-cell PCR*, em que se observou que existia uma gama de

tamanhos de alelos mais alargada nos portadores do haplótipo desestabilizador, bem como uma maior frequência de transições para a expansão completa. Esta informação passou então a ser incorporada no aconselhamento das famílias com portadores de *repeats* intermédios, calculando-se o risco de surgirem alelos completamente expandidos na sua descendência.

Da maior importância para o aconselhamento genético é também o efeito, recentemente identificado, de desvio meiótico associado à transmissão do gene DMJ mutado: os descendentes de portadores de expansões no $(CAG)_n$ do gene MJD1 transmitiriam mais frequentemente o gene mutado do que o gene alelo normal [Igarashi *et al.*, 1994; Igarashi *et al.*, 1996b; Takiyama *et al.*, 1997].

O efeito de desvio meiótico, previamente descrito em famílias com distrofia miotónica [Carey *et al.*, 1994] e em famílias com ADRPL [Ikeuchi *et al.*, 1996], foi detectado em famílias DMJ [Igarashi *et al.*, 1996b], e confirmado em estudos de esperma por *single-cell PCR* [Takiyama *et al.*, 1997], em que se analisaram milhares de transmissões do gene: cerca de 60% dos gâmetas de indivíduos portadores de uma expansão no gene MJD1 (heterozigotos) apresentavam uma expansão (de tamanho igual ou ligeiramente diferente) nesse gene. São necessários estudos adicionais para confirmar estas observações numa população mais alargada e livre de enviesamentos de amostragem, para se determinar se o desvio meiótico, a existir, estará associado unicamente a transmissões paternas ou também a transmissões maternas e para se esclarecerem os mecanismos pelos quais os gâmetas portadores da expansão do $(CAG)_n$ adquiririam uma vantagem selectiva que os torne mais frequentes. A confirmarem-se, estes resultados são, evidentemente, muito relevantes para a correcção da determinação do risco para a descendência de um portador da mutação DMJ.

Necessidade de envolvimento da componente de investigação no aconselhamento genético

O aconselhamento genético inclui o fornecimento aos consultandos das informações sobre a doença e sobre os mecanismos genéticos relevantes para a compreensão do significado do teste preditivo e dos seus resultados possíveis. Em doenças como a DMJ, as ataxias espinocerebelosas ou a doença de Huntington, ainda em fase de estudo, isto implica um envolvimento constante com a componente de investigação, que permita incorporar os novos conhecimentos progressivamente disponíveis sobre a doença, devendo pois considerar-se que o próprio processo de aconselhamento está ainda em fase experimental. São exemplos disso os casos, acima referidos, da identificação de alelos de tamanho intermédio (que não se sabe se serão ou não causadores de doença) da existência ou não de desvio meiótico e da tentativa de distinção entre alelos normais com o mesmo tamanho do $(CAG)_n$ por estudo de polimorfismos intragénicos ou por *Southern blotting*, em indivíduos aparentemente homocigotos para o alelo normal. É pois muito importante que o teste genético seja feito, pelo menos nesta fase inicial, em laboratórios com experiência de investigação com esta doença e em que se esteja ao corrente das dificuldades possíveis, já encontradas ou que ainda venham a surgir.

Por outro lado, uma vez que no caso da DMJ se trata de uma doença ainda sem cura, deve ser feita uma avaliação e um acompanhamento psicológico e social adequado, que permita aos indivíduos em risco que optem pela realização do teste lidar melhor com o seu resultado. Ou seja, o teste molecular deve ser feito apenas no contexto de um programa bem estabelecido de aconselhamento genético, seguido por uma equipa multidisciplinar e seguindo os critérios e normas internacionalmente estabelecidos [Sequeiros e Nance, 1997], e não como uma mera análise laboratorial.

V. CONCLUSÕES
E
DIRECÇÕES FUTURAS DE INVESTIGAÇÃO

V. CONCLUSÕES E DIRECÇÕES FUTURAS DE INVESTIGAÇÃO

CONCLUSÕES

- O gene causador da DMJ localiza-se no cromossoma 14q32.1, sendo a mutação causadora da doença a expansão de uma sequência repetitiva (CAG)_n no gene MJD1.

- Esta mutação está presente na grande maioria das famílias previamente identificadas como DMJ, de origem portuguesa e de outras origens, bem como em outras famílias, identificadas clinicamente como ataxia espinocerebelosa indiferenciada, de origens geográficas muito diversas. A análise destas famílias permitirá a redefinição fenotípica da DMJ.

- O *repeat* CAG no gene MJD1 é altamente polimórfico, varia habitualmente de tamanho entre 12 e 41 unidades CAG na população normal; nos cromossomas DMJ dos doentes e portadores assintomáticos estudados, os *repeats* contêm de 51 a 86 tripletos CAG.

- O tamanho do *repeat* expandido no gene MJD1 correlaciona-se inversamente com a idade de início da DMJ: a *repeats* maiores está associado um início mais precoce da doença.

- A apresentação clínica da DMJ, classificada em 3 subtipos, é também influenciada pelo tamanho do *repeat* CAG expandido. As formas mais graves da doença, caracterizadas por um início mais precoce e progressão mais rápida da doença, e pela presença de espasticidade e distonia - tipo 1- estão associadas a *repeats* maiores do que as formas mais benignas, os tipos 2 e 3.

- A correlação entre o tamanho do *repeat* e a idade de início da DMJ é, no entanto, incompleta, sugerindo que existem outros factores moduladores da idade de início desta

doença. Isto impede, para já, a utilização do tamanho do *repeat* como indicador do prognóstico no teste preditivo e aconselhamento genético.

- Identificou-se a existência de um factor familiar adicional, independente do *repeat* CAG expandido, com influência na idade de início da DMJ, não sendo no entanto ainda possível distinguir entre um efeito genético e um efeito ambiental.

- O *repeat* CAG expandido associado à DMJ apresenta instabilidade somática e intergeracional; observam-se contracções e expansões do $(CAG)_n$ durante a transmissão do gene MJD1 mutado de progenitor para descendente.

- O mosaicismo somático dos genes associados à DMJ e à SCA1, resultante da instabilidade mitótica dos *repeats* expandidos, apresenta uma distribuição tecidual semelhante nestas duas doenças e não se relaciona com o grau de envolvimento patológico das estruturas do SNC observado nos doentes; parece depender apenas do tipo de tecido, possivelmente da sua composição celular, actividade mitótica e/ou nível de expressão do gene.

- O grau de mosaicismo somático dos *repeats* CAG expandidos associados à DMJ e SCA1 não aumenta significativamente ao longo da vida adulta de indivíduos portadores de expansões.

- O grau de mosaicismo somático dos *repeats* CAG expandidos associados à DMJ e SCA1 não difere significativamente entre o cortex e a substância branca cerebral, apesar da diferente composição celular destes.

- A instabilidade intergeracional do $(CAG)_n$ expandido no gene MJD1 leva quer a expansões, quer a contracções do *repeat*, durante a transmissão do gene mutado de progenitor a descendente. A variação média de tamanho na nossa população é de + 0.84 tripletos CAG, o que significa que a tendência geral na população é para um aumento do tamanho do *repeat*, de quase um triplete CAG por geração.

- A variação de tamanho, devida à instabilidade intergeracional, do *repeat* CAG expandido no gene MJD1, apesar de ser a base do fenómeno de antecipação, não pôde ser directamente correlacionável com a variação observada na idade de início nos pares

progenitor-descendente testados na nossa população, possivelmente apenas devido ao enviesamento (inevitável) da amostragem.

- A instabilidade intergeracional do *repeat* CAG expandido no gene MJD1 é modulada por pelo menos dois factores: o sexo do progenitor que transmite a mutação, estando as transmissões paternas associadas a uma maior instabilidade, e as variantes de polimorfismos intragénicos (flanqueantes, a juzante, do *repeat*) presentes no cromossoma normal.

- Não há evidência de um efeito de predisposição à instabilidade pelo haplótipo flanqueante do *repeat* CAG no gene MJD1, nem na população normal, nem nos cromossomas DMJ, reforçando a hipótese de um efeito fundador, pelo menos em algumas populações.

DIRECÇÕES FUTURAS DE INVESTIGAÇÃO

As etapas seguintes dos estudos de genética molecular na DMJ serão, por um lado, o aprofundar dos conhecimentos sobre o gene MJD1 e a proteína (ataxina-3), em particular a identificação da sua função normal na célula, e, por outro, a identificação do mecanismo de toxicidade neuronal da mutação associada à doença, com vista a uma possível abordagem terapêutica.

A determinação da estrutura genómica do gene MJD1 é uma questão já em estudo, mas ainda a completar, da mesma forma que os estudos detalhados da expressão deste gene, a nível de ARN_m e proteína. Os estudos de expressão e de localização subcelular podem, como se discutiu, contribuir para a identificação da função da ataxina-3, uma vez que a localização tecidual e intracelular desta proteína poderá fornecer pistas sobre a sua função. Outros indícios sobre a função da ataxina-3 poderão provir dos estudos de homologia desta proteína com proteínas de outras espécies: foi já identificado um gene homólogo do gene MJD1 na *Caenorhabditis elegans*, de função ainda desconhecida, mas seria sem dúvida interessante procurar outras homologias, em espécies variadas; estas podem contribuir para a identificação de domínios da proteína associados a funções

específicas. A criação de *knock-outs* para a ataxina-3 em ratinhos, na levedura e em outras espécies poderá também permitir tirar conclusões sobre o papel da proteína a nível celular e fisiológico. A determinação da estrutura tridimensional da ataxina-3 será também importante neste aspecto: a comparação das estruturas tridimensionais das proteínas normal e mutada poderá contribuir para a elucidação, quer da função da proteína, quer do mecanismo molecular de patogénese da mutação DMJ, dado que as alterações conformacionais induzidas pela expansão da sequência poliglutamina podem estar relacionadas com a toxicidade das proteínas com expansões.

Outra das vias de investigação a prosseguir é a identificação de factores celulares modificadores do fenótipo, por métodos genéticos, de biologia molecular ou bioquímica, bem como de possíveis interações com o ambiente que tenham relevância para o prognóstico da doença.

Os estudos, já iniciados, com vista à compreensão dos diversos aspectos do comportamento não-mendeliano da mutação DMJ, nomeadamente a instabilidade intergeracional do *repeat* CAG expandido e o desvio meiótico, serão relevantes quer para a compreensão das bases moleculares destes fenómenos (na DMJ e noutras doenças genéticas), quer na sua aplicação prática ao teste preditivo e aconselhamento genético.

Finalmente, outro objectivo fundamental será o desenvolvimento de modelos animais da DMJ, que permitam estudos mais aprofundados do mecanismo patológico (alterações fisiológicas e celulares) e o ensaio de potenciais estratégias terapêuticas para esta doença. O único modelo disponível até à data seria o ratinho transgénico [Ikeda *et al.*, 1996] contendo um fragmento do ADN_C do gene MJD1 com um *repeat* CAG expandido, sob regulação do promotor da *Purkinje cell protein 2* (*pcp-2*), que não apresenta no entanto características patológicas semelhantes à DMJ, uma vez que se observa (como seria de esperar dada a escolha do promotor) apenas a degenerescência das células de Purkinje, normalmente não afectadas na DMJ. Torna-se pois necessário desenvolver um modelo animal que se aproxime o mais possível do fenótipo observado nos doentes; para isto será possivelmente mais apropriado recorrer a *constructs* contendo fragmentos genómicos incluindo a totalidade do gene MJD1, incluindo intrões, que

podem conter elementos reguladores da expressão importantes (donde a importância de se esclarecer a estrutura genómica deste gene), bem como o uso de promotores que dirijam esta expressão para as células apropriadas.

Bibliografia

Almqvist E, Spence N, Nichol K, Andrew SE, Vesa J, Peltonen L, Anvret M, Goto J, Kanazawa I, Goldberg YP, Hayden MR (1995) Ancestral differences in the distribution of the delta-2642 glutamic acid polymorphism is associated with varying CAG repeat lengths in normal chromosomes: insights into the genetic evolution of Huntington disease. *Human Molecular Genetics* 4:207-214

Amorim A, Prata MJ, Veiga I, Almeida VM, Santos MT, Loureiro JL, Coutinho P, Sequeiros J (1995) Machado-Joseph disease: linkage relationships. *American Journal of Human Genetics* 49:355

Auburger G, Diaz GO, Capote RF, Sanchez SG, Perez MP, Estrada del Cueto M, Meneses MG, Farrall M, Williamson R, Chamberlain S, Baute LH (1990) Autosomal dominant ataxia: genetic evidence for locus heterogeneity from a Cuban founder-effect population. *American Journal of Human Genetics* 46:1163-1177

Babcock M, deSilva D, Oaks R, Davis-Kaplan S, Jiralerspong S, Montermini L, Pandolfo M, Kaplan J (1997) Regulation of mitochondrial iron accumulation by Yfh1p, a putative homolog of frataxin. *Science* 276:1709-1712

Banfi S, Servadio A, Chung M-Y, Kwiatkowski TJ, McCall AE, Duvick LA, Shen Y, Roth EJ, Orr HT, Zoghbi HY (1994) Identification and characterization of the gene causing type 1 spinocerebellar ataxia. *Nature Genetics* 7:513-519

Bao J, Sharp AH, Wagster MW, Becher M, Schilling G, Ross CA, Dawson VL, Dawson TM (1996) Expansion of polyglutamine repeat in huntingtin leads to abnormal protein interactions involving calmodulin. *Proceedings of the National Academy of Sciences USA* 93:5037-5042

Barbeau A, Roy M, Cunha L, de Vicente AN, Rosemberg RN, Nyhan WL, MacLeod PL, Chazot G, Langston LB, Dawson DM, Coutinho P (1984) The Natural History of Machado-Joseph Disease. *The Canadian Journal of Neurological Sciences* 11:510-525

Bassett AS, Honer WG (1994) Evidence for Anticipation in Schizophrenia. *American Journal of Human Genetics* 54:864-870

Bharucha NE, Bharucha EP, Bhabha SK (1986) Machado-Joseph in India. *Archives of Neurology* 43:142-144

Bingham PM, Scott MO, Wang S, McPhaul MJ, Wilson EM, Garbern JY, Merry DE, Fischbeck KH (1995) Stability of an expanded trinucleotide repeat in the androgen receptor gene in transgenic mice. *Nature Genetics* 9:191-196

Blass J, Sheu R, Cedarbaum J (1988) Energy metabolism in disorders of the nervous system. *Rev Neurol(Paris)* 144:543-563

Boutté MI (1987) The stumbling disease: a case study of stigma among Azorean-Portuguese. *Social Sciences & Medicine* 24:209-217

Brais B, Xie Y-G, Sanson M, Morgan K, Weissenbach J, Korczyn AD, Blumen SC, Fardeau M, Tomé FMS, Bouchard J-P, Rouleau GA (1995) The oculopharyngeal muscular dystrophy locus maps to the region of the cardiac α and β myosin heavy chain genes on chromosome 14q11.2-q13. *Human Molecular Genetics* 4:429-434

Brook JD, McCurrach ME, Harley HG, Buckler AJ, Church D, Aburatani H, Hunter K (1992) Molecular basis of myotonic dystrophy: expansion of a trinucleotide (CTG) repeat at the 3' end of a transcript encoding a protein kinase family member. *Cell* 68:799-808

Brunner HG, Nelen M, Breakefield O, Ropers HH, van Oost BA (1993) Abnormal Behavior Associated with a Point Mutation in the Structural Gene for Monoamine Oxidase A. *Science* 262:578-580

Brunt ERP, Verschuuren CC, Mensink AJ, Stolte I, Scheffer H (1996) CAG repeat extension correlates with age of onset but does not explain anticipation in Dutch SCA3/MJD family. *Neurology* 46:A197

Burke JR, Enghild JJ, Martin ME, Jou Y-S, Myers RM, Roses AD, Vance JM, Strittmatter W (1996) Huntingtin and DRPLA proteins selectively interact with the enzyme GAPDH. *Nature Medicine* 2:347-349

Burright EN, Davidson JD, Duvick LA, Koshy B, Zoghbi HY, Orr HT (1997) Identification of a self-association region within the SCA1 gene product, ataxin-1. *Human Molecular Genetics* 6:513-518

Burt T, Currie B, Kilburn C, Lethlean AK, Dempsey K, Blair I, Cohen A, Nicholson G (1996) Machado-Joseph disease in east Arnhem Land, Australia: chromosome 14q32.1 expanded repeat confirmed in four families. *Neurology* 46:1118-1122

Cancel G, Abbas N, Stevanin G, Durr A, Chneiweiss H, Néri C, Duyckaerts C, Penet C, Cann HM, Agid Y, Brice A (1995) Marked phenotypic heterogeneity associated with expansion of a CAG repeat sequence at the spinocerebellar ataxia 3 / Machado-Joseph disease locus. *American Journal of Human Genetics* 57:809-816

Carey N, Johnson K, Nokelainen P, Peltonen L, Savontaus M-L, Juvonen V, Anvret M, Grandell U, Robertson E, Middleton-Price H, Malcom S (1994) Meiotic drive at the myotonic dystrophy locus? *Nature Genetics* 6:117-118

Carson WJ, Radvany J, Farrer LA, Vincent D, Rosemberg RN, MacLeod PM, Rouleau GA (1992) The Machado-Joseph Disease Locus is Different from the Spinocerebellar Ataxia Locus (SCA1). *Genomics* 13:852-855

Cha J-HJ, Dure LSI (1994) Trinucleotide repeats in neurologic diseases: a hypothesis concerning the pathogenesis of Huntington's disease, Kennedy's disease and spinocerebellar ataxia type 1. *Life Sciences* 54:1459-1464

Chen B, Kubelik AR, Mohr S, Breitenberger CA (1996) Cloning and characterization of the *Neurospora crassa* cyt-5 gene. A nuclear-coded mitochondrial RNA polymerase with a polyglutamine repeat. *The Journal of Biological Chemistry* 271:6537-6544

Chen X, Mariappan SVS, Catasti P, Ratliff R, Moyzis RK, Laayoun A, Smith SS, Bradbury EM, Gupta G (1995) Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications. *Proceedings of the National Academy of Sciences U.S.A.* 92:5199-5203

Chong SS, Almqvist E, Telenius H, LaTray L, Nichol K, Bourdelat-Parks B, Goldberg YP, Haddad BR, Richards F, Sillence D, Greenberg CR, Ives E, Van den Engh G, Hayden MR (1997) Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. *Human Molecular Genetics* 6:301-309

Chong SS, McCall AE, Cota J, Subramony SH, Orr HT, Hughes MR, Zoghbi H (1995) Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics* 10:344-349

Chung M, Ranum LPW, Duvick LA, Servadio A, Zoghbi HY, Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nature Genetics* 5:254-258

Coutinho P (1992) Doença de Machado-Joseph. Tentativa de definição. Tese - Universidade do Porto

Coutinho P, Andrade C (1978) Autosomal dominant system degeneration in Portuguese families of the Azores Islands. *Neurology* 28:703-709

Coutinho P, Guimarães A, Scaravilli F (1982) The pathology of Machado-Joseph disease: report of a possible homozygous case. *Acta Neuropathologica* 58:48-54

Cross G, Pitt T, Sharif A, Bates G, Lehrach H (1994) False-negative result for Huntington's disease mutation. *New England Journal of Medicine* 343:1232

David G, Abbas N, Stevanin G, Durr A, Yvert G, Cancel G, Weber C, Imbert G, Saudou F, Antoniou E, Drabkin H, Gemmill R, Giunti P, Benomar A, Wood N, Ruberg M, Agid Y, Mandel J-L, Brice A (1997) Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nature Genetics* 17:65-70

Dayn A, Samadashwily GM, Mirkin SM (1992) Intramolecular DNA triplexes: unusual sequence requirements and influence on DNA polymerization. *Proceedings of the National Academy of Sciences U.S.A.* 89:11406-11410

De Rooij KE, De Koning Gans PAM, Skraastad MI, Belfrois RDM, Vegter-Van der Vlis M, Roos RAC, Van Ommen G-JB, Den Dunnen JT, Losekoot M (1993) Dynamic mutation in Dutch Huntington's disease patients: increased paternal repeat instability extending to within the normal size range. *Journal of Medical Genetics* 30:996-1002

DiFiglia M, Sapp E, Chase K, Schwarz C, Meloni A, Young C, Martin E, Vonsattel J-P, Carraway R, Reeves SA, Boyce FM, Aronin N (1995) Huntingtin is a cytoplasmic protein associated with vesicles in human and rat brain neurons. *Neuron* 14:1075-1081

Drwinga HL, Toji LH, Kim CH, Greene AE, Mullivor RA (1993) NIGMS human/rodent somatic cell hybrid panels 1 and 2. *Genomics* 16:311-314

Durr A, Stevanin G, Cancel G, Abbas N, Chneiweiss H, Agid Y, Feingold J, Brice A (1995) Gender equality in Machado-Joseph disease. *Nature Genetics* 11:118

Duvoisin R, Chokroverty S, Lepore F, Nicklas W (1983) Glutamate dehydrogenase deficiency in patients with olivopontocerebellar atrophy. *Neurology* 33:1322-1326

Duyao M, Ambrose C, Myers R, Novelletto A, Persichetti F, Fontali M, Folstein S, Ross C, Franz M, Abbot M, Gray J, Conneally P, Young A, Penney J, Hollingsworth Z, Shoulson I, Lazzarini A, Falek A, Koroshetz W, Sax D, Bird E, Vonsattel J, Bonilla E, Alvir J, Conde JB, Cha J-H, Dure L, Gomez F, Ramos M, Ramos JS, Snodgrass S, Young M, Wexler N, Moscovitz C, Penschaszadeh G, MacFarlane H, Anderson M, Jenkins B, Srinidhi J, Barnes G, Gusella J, MacDonald M (1993) Trinucleotide repeat length instability and age of onset in Huntington's disease. *Nature Genetics* 4:387-392

Eichler E, Holden J, Popovich BW, Reiss AL, Snow K, Thibodeau SN, Richards CS, Ward PA, Nelson DL (1994) Length of uninterrupted CGG repeats determines instability in the FMR-1 gene. *Nature Genetics* 8:88-94

Fisch GS, Snow K, Thibodeau SN, Chalifaux M, Holden JJA, Nelson DL, Howard-Peebles PN, Maddalena A (1995) The fragile X premutation in carriers and its effect on mutation size in offspring. *American Journal of Human Genetics* 56:1147-1155

Fleischer B (1918) Uber myotonische dystrophie mit katarakt: eine hereditare, familiare degeneration. *Archiv Ophthalmol* 96:91-133

Floriijn RJ, Bonden LAJ, Vrolijk J, Wiegant J, Vaandrager JW, Baas F, Dunnen JT, Tanke HJ, Ommen GJB, Raap AK (1995) High-resolution DNA fiber-FISH for genomic DNA mapping and colour bar-coding of large genes. *Human Molecular Genetics* 5:831-836

Forse RA, MacLeod PM, Holden GGA, White BN (1989) DNA marker studies show that MJD is not an allele of the Huntington disease locus. *Journal of Neurogenetics* 5:155-158

Fu Y-H, Kuhl DPA, Pizzuti A, Pierretti M, Sutcliffe JS, Richards S, Verkerk JMH, Holden JJA, Fenwick RGJ, Warren ST, Oostra BA, Nelson DL, Caskey T (1991) Variation of CGG repeat at the fragile X site results in genetic instability: resolution of the Sherman paradox. *Cell* 67:1047-1058

Gacy AM, Goellner G, Juranic N, Macura S, McMurray CT (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell* 81:533-540

Gadoth N, Slogotora J, Merims D (1996) Correlation between clinical and molecular genetic analysis in Machado-Joseph disease. *Neurology* 46:A329

Gaspar C, Goto J, Lopes-Cendes I, Hayes S, Arvidsson K, Maciel P, Silveira I, Coutinho P, Stevanin G, DeStefano A, Reiss O, Sasaki H, Giunti P, Wood N, Brunt E, Tranebjaerg L, Nicholson G, Higgins JJ, Lauritzen M, Watanabe M, Volpini V, Hsieh M, Soong B-W, Wang G-X, Ranum L, Tsuji S, Brice A, Farrer LA, Sequeiros J, Rouleau GA (1997) Different origins of mutation in Machado-Joseph disease patients. *1st International Symposium on Inherited Ataxias, Montreal*.

Gaspar C, Lopes-Cendes I, DeStefano AL, Maciel P, Silveira I, Coutinho P, MacLeod P, Sequeiros J, Farrer LA, Rouleau GA (1996) Linkage disequilibrium analysis in Machado-Joseph disease patients of different origins. *Human Genetics* 98:620-624

Gavrieli Y, Sherman Y, Ben-Sasson A (1992) Identification of programmed cell death in situ via specific labelling of nuclear DNA fragmentation. *Journal of Cell Biology* 119:493-501

Geschwind DH, Perlman S, Figueroa CP, Treiman LJ, Pulst SM (1997) The prevalence and wide clinical spectrum of the spinocerebellar ataxia type 2 trinucleotide repeat in patients with autosomal dominant cerebellar ataxia. *American Journal of Human Genetics* 60:842-850

Ghosh S, Palmer SM, Rodrigues NR, Cordell HJ, Hearne CM, Cornall RJ, Prins J-B, McShane P, Lathrop GM, Peterson LB, Wicker LS, Todd JS (1993) Polygenic control of autoimmune diabetes in nonobese diabetic mice. *Nature Genetics* 4:404-409

Giunti P, Sweeney MG, Harding AE (1995) Detection of the Machado-Joseph disease/spinocerebellar ataxia three trinucleotide repeat expansion in families with autosomal dominant motor disorders, including the Drew family of Walworth. *Brain* 118:1077-1085

Goellner GM, Tester D, Thibodeau S, Almqvist E, Goldberg YP, Hayden MR, McMurray CT (1997) Different mechanisms underlie DNA instability in Huntington disease and colorectal cancer. *American Journal of Human Genetics* 60:879-890

Goldberg YP, Kalchman MA, Metzler M, Nasir J, Zeisler J, Graham R, Koide HB, O'Kusky J, Sharp AH, Ross CA, Jirik F, Hayden MR (1996) Absence of disease phenotype and intergenerational stability of the CAG repeat in transgenic mice expressing the human Huntington disease transcript. *Human Molecular Genetics* 5:177-185

Goldberg YP, Kremer B, Andrew SE, Theilmann J, Graham RK, Squitieri F, Telenius H, Adam S, Sajoo A, Starr E, Heiberg A, Wolff G, Hayden MR (1993) Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. *Nature Genetics* 5:174-179

Goldberg YP, Nicholson DW, Rasper DM, Kalchman MA, Koide HB, Graham RK, Bromm M, Kazemi-Esfarjani P, Thornberry NA, Vaillancourt JP, Hayden MR (1996) Cleavage of huntingtin by apopain, a proapoptotic cysteine protease, is modulated by the polyglutamine tract. *Nature Genetics* 13:442-449

Gostout B, Q. L, Sommer S (1993) "Cryptic" repeating triplets of purines and pyrimidines (cRRY(i)) are frequent and polymorphic: analysis of coding cRRY(i) in the Propiomelanocortin (POMC) and TATA-binding protein (TBP) genes. *American Journal of Human Genetics* 52:1182-1190

Gourdon G, Radvanyi F, Lia A-S, Duros C, Blanche M, Abitbol M, Junien C, Hoffman-Radvanyi H (1997) Moderate intergenerational and somatic instability of a 55-CTG repeat in transgenic mice. *Nature Genetics* 15:190-192

Gouw LG, Kaplan CD, Haines JH, Digre KB, Rutledge SL, Matilla A, Leppert M, Zoghbi HY, Ptáček LJ (1995) Retinal degeneration characterizes a spinocerebellar ataxia mapping to chromosome 3p. *Nature Genetics* 10:89-93

Green H (1993) Human genetic diseases due to codon reiteration: relationship to an evolutionary mechanism. *Cell* 74:955-956

Haberhausen G, Damian MS, Leweke F, Muller U (1995) Spinocerebellar ataxia, type 3 (SCA3) is genetically identical to Machado-Joseph disease. *Journal of the Neurological Sciences* 132:71-75

Hannan AJ (1996) Trinucleotide-repeat expansions and neurodegenerative disease: a mechanism of pathogenesis. *Clinical and Experimental Pharmacology and Physiology* 23:1015-1020

Harper PS, Harley HG, Reardon W, Shaw DJ (1992) Anticipation in myotonic dystrophy: new light on an old problem. *American Journal of Human Genetics* 51:10-16

Hashida H, Goto J, Kurisaki H, Mizusawa H, Kanazawa I (1997) Brain regional differences in the expansion of a CAG repeat in the spinocerebellar ataxias: dentatorubral-pallidoluysian atrophy, Machado-Joseph disease and spinocerebellar ataxia type 1. *Annals of Neurology* 41:505-511

Higgins JJ, Nee LE, Vasconcelos O, Ide SE, Lavedan C, Goldfarb LG, Ploymeropoulos MH (1996) Mutations in American families with spinocerebellar ataxia (SCA) type 3: SCA3 is allelic to Machado-Joseph disease. *Neurology* 46:208-213

Housman D (1995) Gain of glutamines, gain of function? *Nature Genetics* 10:3-4

Howeler CJ, Busch HFM, Geraedts JPM, Niermeijer MF, Staal A (1989) Anticipation in myotonic dystrophy: fact or fiction? *Brain* 112:779-797

Hudson TH, Engelstein M, Lee MK, Ho EC, Rubenfield MJ, Adams CP, Housman DE, Dracopoli NC (1992) Isolation and chromosomal assignment of 100 highly informative human simple sequence repeat polymorphisms. *Genomics* 13:622-629

Hunter A, Tsilfidis C, Mettler G, Jacob P, Mahadevan M, Surh L, Korneluk R (1992) The correlation of age of onset with CTG trinucleotide repeat amplification in myotonic dystrophy. *Journal of Medical Genetics* 29:774-779

Igarashi S, Onodera O, Tanaka H, Weissenbach J, Tsuji S (1994) Isolation and characterization of human cerebellum cDNAs containing polymorphic CAG trinucleotide repeats. *American Journal of Human Genetics* 55:A189

Igarashi S, Takiyama Y, Cancel G, Rogaeva EA, Sasaki H, Wakisaka A, Zhou Y-X, Takano H, Endo K, Sanpei K, Oyake M, Tanaka H, Stevanin G, Abbas N, Durr A, Rogaev EI, Sherrington R, Tsuda T, Ikeda M, Cassa E, Nishizawa M, Benomar A, Julien J, Weissenbach J, Wang G-X, Agid Y, George-Hyslop PHS, Brice A, Tsuji S (1996a) 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. *Human Molecular Genetics* 5:923-932

Igarashi S, Takiyama Y, Takano H, K S, Oyake M, Tanaka H, Sasaki H, Wakisaka A, Nishizawa M, Horai S, Brice A, Roganova E, St. George-Hyslop P, Tsuji S (1996b) Single base polymorphism at the 3' end of the CAG repeat in MJD1 gene is associated with the size of CAG repeat and the intergenerational instability. *American Journal of Human Genetics* 5:923-932

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

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

Imbert G, Kretz C, Johnson K, Mandel J-L (1993) Origin of the expansion mutation in myotonic dystrophy. *Nature Genetics* 4:72-76

Imbert G, Saudou F, Yvert G, Devys D, Trottier Y, Garnier J-M, Weber C, Mandel J-L, Cancel G, Abbas N, Durr A, Didierjean O, Stevanin G, Agid Y, Brice A (1996) Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nature Genetics* 14:285-291

Imbert G, Trottier G, Beckman J, Mandel JL (1994) The gene for the TATA binding protein (TBP) that contains a highly polymorphic protein coding CAG repeat maps to 6q27. *Genomics* 21:667-668

Jackson J, Currier M, Terasaki P, Morton N (1977) Spinocerebellar ataxia and HLA linkage: risk prediction by HLA typing. *New England Journal of Medicine* 296:1138-1141

Jeffreys AJ, Tamaki K, MacLeod A, Monckton DG, Neil DL, Armour JAL (1994) Complex gene conversion events in germline mutation at human minisatellites. *Nature Genetics* 6:136-145

Kahlem P, Terré C, Green H, Djian P (1996) Peptides containing glutamine repeats as substrates for transglutaminase-catalyzed cross-linking: relevance to diseases of the nervous system. *PNAS* 93:14580-14585

Kanda T, Isozaki E, Kato S, Tanabe H, Oda M (1989) Type III Machado-Joseph disease in a Japanese family: a clinicopathological study with special reference to the peripheral nervous system. *Clinical Neuropathology* 8:134-141

Kark RAP, Blass JP, Spence A (1977) Physostigmine in familial ataxias. *Neurology* 27:70-72

Kawaguchi Y, Okamoto T, Taniwaki M, Aizawa M, Inoue M, Katayama S, Kawakami H, Nakamura S, Nishimura M, Akiguchi I, Kimura J, Narumiya S, Kakizuka A (1994) CAG expansions in a novel gene from Machado-Joseph disease at chromosome 14q32.1. *Nature Genetics* 8:221-227

Kawakami H, Maruyama H, Nakamura S, Kawaguchi Y, Kakizuka A, Doyu M, Sobue G (1995) Unique features of the CAG repeats in Machado-Joseph disease. *Nature Genetics* 9:344-345

Kish S, Currier R, Schut L, Perry T, Morito L (1987) Brain cholineacetyltransferase reduction in dominantly inherited olivopontocerebellar atrophy. *Annals of Neurology* 22:272-275

Klockgether T, Evert B, Ludtke R, Kramer B, Abele M, Burk K, Schols L, Riess O, Laccone F, Lopes-Cendes I, Brice A, Dichgans J (1997) The natural history of degenerative ataxia: a retrospective study in 420 patients. *1st International Symposium on Inherited Ataxias, Montreal.*

Koide R, Ikeuchi T, Onodera O, Tanaka h, Igarashi S, Endo K, Takahashi H, Kondo R, Ishikawa A, Hayashi T, Saito M, Tomoda A, Miike T, Naito H, Ikuta F, Tsuji S (1994) Unstable

expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nature Genetics* 6:9-13

Korneluk RG, Narang MA (1997) Anticipating anticipation. *Nature Genetics* 15:119-120

Koshy B, Matilla T, Burchright EN, Merry DE, Fischbeck KH, Orr HT, Zoghbi HY (1996) Spinocerebellar ataxia type 1 and spinobulbar muscular atrophy gene products interact with glyceraldehyde-3-phosphate dehydrogenase. *Human Molecular Genetics* 5:1311-1318

Kremer EJ, Pritchard M, Lynch M, Yu S, Holman K, Baker E, Warren ST, Schlessinger D, Sutherland GR, Richards RI (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CGG)_n. *Science* 252:1711-1714

Kunst CB, Warren ST (1994) Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* 77:853-861

La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 352:77-79

Landwehrmeyer GB, McNeil SM, Dure LS, Ge P, Aizawa H, Huang Q, Ambrose CM, Duyao MP, Bird ED, Bonilla E, de Young M, Avila-Gonzales AJ, Wexler NS, DiFiglia M, Gusella JF, MacDonald ME, Penney JB, Young AB, Vonsattel JP (1995) Huntington's disease gene: regional and cellular expression in brain of normal and affected individuals. *Annals of Neurology* 37:218-230

Lang AE, Rogaeva EA, Tsuda T, Hutterer J, St. George-Hyslop P (1994) Homozygous inheritance of the Machado-Joseph Disease gene. *Annals of Neurology* 36:443-447

Lathrop GM, Lalouel JM (1984) Easy calculations of lod scores and genetic risks on small computers. *American Journal of Human Genetics* 36:460-465

Lavedan C, Radvanyi HH, Shelbourne P, Rabes JP, Duros C, Savoy D, Dehaupas I, Luce S, Johnson K, Junien C (1993) Myotonic Dystrophy: Size-and Sex-Dependent Dynamics of CTG Meiotic Instability, and Somatic Mosaicism. *American Journal of Human Genetics* 52:875-883

Leeflang EP, Arnheim N (1995) A novel repeat structure at the myotonic dystrophy locus in a 37 repeat allele with unexpectedly high stability. *Human Molecular Genetics* 4:135-136

Li X-J, Li S-H, Sharp AH, Nucifora Jr FC, Schilling G, Lanahan A, Worley P, Snyder SH, Ross CA (1995) A huntingtin-associated protein enriched in brain with implications for pathology. *Nature* 378:398-402

Lima L, Coutinho P (1980) Clinical criteria for diagnosis of Machado-Joseph disease: Report of a non-Azorean Portuguese family. *Neurology* 30:319-322

Limprasert P, Nouri N, Heyman R, Nopparatana C, Kamonsilp M, Deininger PL, Keats BJB (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. *Human Molecular Genetics* 5:207-213

Lindblad K, Zander C, Nylander P-O, De Bruyn A, Sourey D, Engstrom C, Holmgren G, Hudson T, Chotai J, Mendlewicz J, Van Broeckhoven C, Adolfsson R, Carrero-valenzuela R, Payami H, Johnson W, Stenroos ES, Shattue S, Nutt J, Brice A, Litt M, Auburger G, Ratzlaff T, Schalling M (1995) Detection of CAG repeat expansions in neuropsychiatric disorders using the repeat expansion detection (RED) method. *Cold Spring Harbor*

Lopes-Cendes I, Maciel P, Kish S, Gaspar C, Robitaille Y, Clark HB, Koeppen AH, Nance M, Schut L, Silveira I, Coutinho P, Sequeiros J, Rouleau GA (1996a) Somatic mosaicism in CNS

specimens of spinocerebellar ataxia type1 (SCA1) and Machado-Joseph disease (MJD) patients. *Annals of Neurology* 40:199-206

Lopes-Cendes I, Silveira I, Maciel P, Gaspar C, Radvany J, D. C, Babul R, Stewart J, Dolliver M, Robitaille Y, Rouleau GA, Sequeiros J (1996b) Limits of clinical assessment in the accurate diagnosis of Machado-Joseph disease. *Archives of Neurology* 53:1168-1174

Lopes-Cendes I, Steiner C, Silveira I, Pinto-Júnior W, Maciel E, Rouleau GA (1996c) Clinical and molecular characteristics of a Brazilian family with spinocerebellar ataxia type 1. *Arquivos de Neuropsiquiatria* 54:412-418

Lopes-Cendes I, Teive HAG, Silveira I, Maciel JA, Radvany J, Arruda WO, Bettencourt BC, Calcagnotto ME, Da Costa JC, Steiner CE, Pinto-Júnior W, Santos AJ, Rouleau GA (1996d) Frequency of the three different mutations causing spinocerebellar ataxia (SCA1, MJD/SCA3, DRPLA) in Brazilian patients. *9th International Congress of Human Genetics, Rio de Janeiro.*

Luty JA, Litt M (1991) Dinucleotide repeat polymorphism at the D14S45 locus. *Nucleic Acids Research* 19:4308

MacDonald ME, Barnes G, Srinidhi J, Duyao MP, Ambrose CM, Myers RH, Gray J, Conneally M, Young A, Penney J, Shoulson I, Hollingsworth Z, Koroshetz W, Bird E, Vonsattel JP, Bonilla E, Moscowitz C, Penchaszadeh G, Brzustowicz L, Alvir J, Bickham Conde J, Cha J-H, Dure L, Gomez F, Ramos-Arroyo M, Sanchez-Ramos J, Snodgrass SR, de Young M, Wexler NS, MacFarlane H, Anderson MA, Jenkins B, Gusella JF (1993) Gametic but not somatic instability of CAG repeat length in Huntington's disease. *Journal of Medical Genetics* 30:982-986

Maciel P, Gaspar C, DeStefano AL, Silveira I, Coutinho P, Radvany J, Dawson DM, Sudarski L, Guimarães J, Loureiro JEL, Nezarati MM, Corwin LI, Lopes-Cendes I, Rooke K, Rosenberg R,

MacLeod P, Farrer LA, Sequeiros J, Rouleau GA (1995) Correlation between CAG repeat length and clinical features in Machado-Joseph Disease. *American Journal of Human Genetics* 57:54-61

Maciel P, Rousseau M, Gaspar C, Coutinho P, Sequeiros J, Rouleau GA (1997) An intermediate allele of 51 repeats found at the Machado-Joseph disease (MJD) locus: implications for predictive testing and the study of the origin of the MJD mutation. *47th Annual Meeting of the American Society of Human Genetics, Baltimore.*

MacPherson JN, Bullman H, Youngs SA, Jacobs PA (1994) Insert size and flanking haplotype in fragile X and normal populations: possible multiple origins for the fragile X mutation. *Human Molecular Genetics* 3:399-405

Malter HE, Iber JC, Willemsen R, de Graaf E, Tarleton JC, Leisti J, Warren ST, Oostra BA (1997) Characterization of the full fragile X syndrome mutation in fetal gametes. *Nature Genetics* 15:165-169

Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, Bates GP (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nature Genetics* 15:197-200

Mangiarini L, Sathasivam K, Seller M, Cozens B, Harper A, Hetherington C, Lawton M, Trotter Y, Lehrach H, Davies SW, Bates GP (1996) Exon 1 of the HD gene with an expanded CAG repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 87:493-506

Maruyama H, Nakamura S, Matsuyama Z, Sakai T, Doyu M, Sobue G, Seto M, Tsujihata M, Oh-i T, Nishio T, Suhonara N, Takahashi R, Hayashi M, Nishino I, Ohtake T, Oda T, Nishimura M, Saida T, Matsumoto H, Baba M, Kawaguchi Y, Kakizuka A, Kawakami H (1995) Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Human Molecular Genetics* 4:807-812

Matesanz F, Alcina A (1996) Glutamine and tetrapeptide repeat variations affect the biological activity of different mouse interleukin-2 alleles. *European Journal of Immunology* 26:1675-1682

Matilla T, McCall A, Subramony SH, Zoghbi HY (1995) Molecular and clinical correlations in spinocerebellar ataxia type 3 and Machado-Joseph disease. *Annals of Neurology* 38:68-72

McInnis M (1996) Anticipation: an old idea in new genes. *American Journal of Human Genetics* 59:973-979

McLaughlin A, Spencer C, Eberwinne J (1996) CAG trinucleotide repeats interact with RNA-binding proteins. *American Journal of Human Genetics* 59:561-569

McMurray CT (1995) Mechanisms of DNA expansion. *Chromosoma* 104:2-13

Mello FA, Abbot BP (1988) Effect of sulfamethoxazole and trimethoprim on neurologic dysfunction in a patient with Joseph's disease. *Archives of Neurology* 45:210-213

Mhatre AN, Trifiro M, Kaufman M, Kazemi-Esfarjani P, Figlewicz D, Rouleau GA, Pinsky L (1993) Reduced transcriptional regulatory competence of the androgen receptor in X-linked spinal and bulbar muscular atrophy. *Nature Genetics* 5:184-187

Monckton DG, Coolbaugh MI, Ashizawa KT, Siciliano MJ, Caskey CT (1997) Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nature Genetics* 15:193-196

Monckton DG, Neumann R, Guram T, Fretwell N, Tamaki K, MacLeod A, Jeffreys AJ (1994) Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nature Genetics* 8:162-170

Montermini L, Kish SJ, Jiralerspong S, Lamarche JB, Pandolfo M (1997) Somatic mosaicism for the Friedreich ataxia GAA triplet repeat expansions in central nervous system. *1st International Symposium on Inherited Ataxias, Montreal.*

Mott FW (1911) A lecture on heredity and insanity. *Lancet* 1:1251-1259

Murray A, MacPherson JN, Pound MC, Sharrock A, Youings SA, Dennis NR, McKechnie N, Linehan P, Morton NE, Jacobs PA (1997) The role of the size, sequence and haplotype in the stability of FRAXA and FRAXE alleles during transmission. *Human Molecular Genetics* 6:173-184

Myers RH, MacDonald ME, Koroshetz WJ, Duyao MP, Ambrose CM, Taylor SAM, Barnes G, Srinidhi J, Lin CS, Whaley WL, Lazzarini AM, Schwarz M, Wolff G, Bird ED, Vonsattel J-PG, Gusella JF (1993) De novo expansion of a (CAG) repeat in sporadic Huntington's disease. *Nature Genetics* 5:168-173

Nagafuchi S, Yanagisawa H, Ohsaki E, Shirayama T, Tadokoro K, Inoue T, Yamada M (1994) Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA). *Nature Genetics* 8:177-182

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

Nancarrow JK, Holman K, Mangelsdorf M, Hori T, Denton M, Sutherland GR, Richards RI (1995) Molecular basis of p(CGG)_n repeat instability at the FRA16A fragile site locus. *Human Molecular Genetics* 4:367-372

Neville C, Mahadevan MS, Barceló JM, Komeluk R (1994) High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of myotonic dystrophy mutation. *Human Molecular Genetics* 3:45-51

Nicol K, Goldberg YP, Almqvist E, Sillence D, Richards F, M.R. H (1996) Interallelic conversion events associated with trinucleotide instability in the Huntington disease gene. *American Journal of Human Genetics* 59:1590

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

Nygaard TG, Wilhelmsen KC, Risch NJ, Brown DL, Trugman JM, Gilliam TC, Fahn S, Weeks D (1993) Linkage mapping dopa-responsive dystonia (DRD) to chromosome 14q. *Nature Genetics* 5:386-391

O'Donovan MC, Guy C, Craddock N, Murphy KC, Cardno AG, Jones LA, Owen MJ, McGuffin P (1995) Expanded CAG repeats in schizophrenia and bipolar disorder. *Nature Genetics* 10:380-381

Ohshima K, Wells RD (1997) GAA.TTC structure and Friedreich's ataxia. *1st International Symposium on Inherited Ataxias, Montreal.*

Oliveira J, Teixeira CM, Coutinho P, Sequeiros J, Sousa A (1994) Analysis of anticipation of age at onset in Machado-Joseph disease. *3rd International Workshop on Machado-Joseph disease, Furnas.*

Onodera O, Roses AD, Tsuji S, Vance JM, Stritmatter WJ, Burke JR (1996) Toxicity of expanded polyglutamine-domain proteins in *Escherichia coli*. *FEBS Journal* 399:135-139

Ophoff RA, Terwindt GM, Vergouve MN, van Eijk R, Oefner PJ, Hoffman SMG, Lamerdin JE, Mohrenwiser HW, Bulman DE, Ferrari M, Haan J, Lindhout D, Ommen G-JB, Hofker MH, Ferrari MD, Frants RR (1996) Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca²⁺ channel gene CACNL1A4. *Cell* 87:543-552

Orr HT, Chung M-y, Banfi S, Kwiatkowski TJJ, Servadio A, Beaudet AL, McCall AE, Duvick LA, Ranum LPW, Zoghbi HY (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature Genetics* 4:221-226

Oruc L, Lindblad K, Verheyen GR, Ahlberg S, Jakovljevic M, Ivezic S, Raeymaekers P, Van Broeckhoven S M. (1997) CAG repeat expansions in bipolar and unipolar disorders. *American Journal of Human Genetics* 60:730-732

Paul C (1996) Reações esperadas aos resultados do teste preditivo - A experiência dos programas da doença de Huntington. In: J. Sequeiros (ed) *O teste preditivo da Doença de Machado-Joseph*, UnIGENE, IBMC, Porto, pp 79-93

Paulson HL, Das SS, Crino PB, Perez MK, Patel SC, Gotsdiner D, K.H. F, Pittman RN (1997a) Machado-Joseph disease gene product is a cytoplasmic protein widely expressed in brain. *Annals of Neurology* 41:453-462

Paulson HL, Subramony SH, Das SS, Perez MK, Pittman RN (1997b) Characterization of the Machado-Joseph disease gene product, ataxin-3, in normal and disease brain. *1st International Symposium on Inherited Ataxias, Montreal.*

Penrose LS (1948) The problem of anticipation in the pedigrees of dystrophia myotonica. *Annals of Eugenics* 14:125-132

Perutz MF, Johnson T, Suzuki M, Finch JT (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proceedings of the National Academy of Science, U.S.A.* 91:5355-5358

Pianese L, Cavalcanti F, De Michele G, Filla A, Campanella G, Calabrese O, Castaldo I, Monticelli A, Coccozza S (1997) The effect of parental gender on the GAA dynamic mutation in the FRDA gene. *American Journal of Human Genetics* 60:460-463

Portera-Cailliau C, Hedreen JC, Price DL, Koliatsos VE (1995) Evidence for apoptotic cell death in Huntington disease and excitotoxic animal models. *The Journal of Neuroscience* 15:3775-3787

Pulst S-M, Nechiporuk A, Nechiporuk T, Gispert S, Chen X-N, Lopes-Cendes I, Pearlman S, Starkman S, Orozco-Diaz G, Lunke A, DeJong P, Rouleau GA, Auburger G, Korenberg JR, Figueroa C, Sahba S (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nature Genetics* 14:269-276

Ranen NG, Colin Stine O, Abbott MH, Sherr M, Codori A-M, Franz ML, Chao NI, Chung AS, Pleasant N, Callahan C, Kasch LM, Ghaffari M, Chase GA, Kazazian HH, Brandt J, Folstein SE, Ross CA (1995) Anticipation and instability of IT-15 (CAG)_n repeats in parent-offspring pairs with Huntington disease. *American Journal of Human Genetics* 57:593-602

Ranum LPW, Chung M, Banfi S, Bryer A, Schut LJ, Ramesar R, Duvick LA, McCall A, Subramony SH, Goldfarb L, Gomez C, Sandkuijl LA, Orr HT, Zogby HY (1994) Molecular and clinical correlations in spinocerebellar ataxia type 1: evidence for familial effects on the age at onset. *American Journal of Human Genetics* 55:244-252

Ranum LPW, Lundgren JK, Schut LJ, Ahrens MJ, Perlman S, Aita J, Bird TD, Gomez C, Orr HT (1995) Spinocerebellar ataxia type 1 and Machado-Joseph disease: incidence of CAG expansion among adult-onset ataxia patients from 311 families with dominant, recessive or sporadic ataxia. *American Journal of Human Genetics* 57:603-608

Richards RI, Holman K, Friend K, Kremer E, Hillen D, Staples A, Brown WT, Goonewardena P, Tarleton J, Schwartz C, Sutherland GR (1992) Evidence of founder chromosomes in fragile X syndrome. *Nature Genetics* 1:257-260

Richards RI, Sutherland GR (1992) Dynamic Mutations: A New Class of Mutations Causing Human Disease. *Cell* 70:709-712

Rosen A (1996) Huntingtin: new marker along the road to death? *Nature Genetics* 13:380-382

Rosenberg RN (1977) Azorean disease of the nervous system. *New England Journal of Medicine* 297:729

Rosenberg RN, Fowler HL (1981) Autosomal dominant motor system disease of the Portuguese: a review. *Neurology* 31:1124-1126

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

Rubinsztein DC, Leggo J, Amos W, Barton DE, Ferguson-Smith MA (1994) Myotonic dystrophy CTG repeats and the associated insertion/deletion polymorphisms in human and primate populations. *Human Molecular Genetics* 3:2031-2035

Rubinsztein DC, Leggo J, Coetze GA, Irvine RA, Buckley M, Ferguson-Smith MA (1995a) Sequence variation and size ranges of CAG repeats in the Machado-Joseph disease, spinocerebellar ataxia type 1 and androgen receptor genes. *Human Molecular Genetics* 4:1585-1590

Rubinsztein DC, Leggo J, Goodburn S, Barton DE, Ferguson-Smith MA (1995b) Haplotype analysis of the delta-2642 and (CAG)_n polymorphisms in the Huntington's disease (HD) gene provides an explanation for the apparent "founder" effect. *Human Molecular Genetics* 4:203-206

Sakai T, Ohta M, Ishino H (1983) Joseph disease in a non-Portuguese family. *Neurology* 33:74-80

Sangla S, De Broucker TH, Cheron F, Cambier J, Dehen H (1990) Amélioration d'une maladie de Joseph para la sulfaméthoxazole-triméthoprime. *Revue de Neurologie* 146:213-214

Sanpei K, Takano H, Igarashi S, Sato T, Oyake M, Sasaki H, Wakisaka A, Tashiro K, Ishida Y, Ikeuchi T, Koide R, Saito M, Sato A, Tanaka T, Hanyu S, Takiyama Y, Nishizawa M, Shimizu N, Nomura Y, Segawa M, Iwabuchi K, Eguchi I, Tanaka H, Tsuji S (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nature Genetics* 14:277-291

Schaefer FV, Calikoglu AS, Whetsell LH (1994) Characterization of conservative somatic instability of the CAG repeat region in Huntington's disease. *American Journal of Human Genetics* 55:A241

Schalling M, Hudson TJ, Buetow KH, Housman DE (1993) Direct detection of novel expanded trinucleotide repeats in the human genome. *Nature Genetics* 4:135-139

Schlotterer C, Tautz D (1992) Slippage synthesis of simple sequence DNA. *Nucleic Acids Research* 20:211-215

Schols L, Amoiridis G, Epplen JT, Langkafel M, Przuntek H, Riess O (1996) Relations between genotype and phenotype in German patients with the Machado-Joseph disease mutation. *Journal of Neurology, Neurosurgery and Psychiatry* 61:466-470

Schols L, Vieira-Saecker AMM, Schols S, Przuntek H, Epplen JT, Riess O (1995) Trinucleotide expansion within the MJD1 gene presents clinically as spinocerebellar ataxia and occurs most frequently in German SCA patients. *Human Molecular Genetics* 4:1001-1005

Schweitzer JK, Livingston DM (1997) Destabilization of CAG trinucleotide repeat tracts by mismatch repair mutations in yeast. *Human Molecular Genetics* 6:349-355

Seipel K, Georgiev O, Gerber H-P, Schaffner W (1994) Basal components of the transcription apparatus (RNA polymerase II, TATA binding protein) contain activation domains: is the repetitive C-terminal domain (CTD) of RNA polymerase II a "portable enhancer domain"? *Molecular Reproduction and Development* 39:215-225

Sequeiros J (1989) Análise genética da variação fenotípica na doença de Machado-Joseph. Tese - Universidade do Porto

Sequeiros J (1996) Protocolo geral do programa nacional de teste preditivo e aconselhamento genético na doença de Machado-Joseph. In: J. Sequeiros (ed) *O teste preditivo da doença de Machado-Joseph*, UnIGENE, IBMC, Porto, pp 123-149

Sequeiros J, Coutinho P (1981) Genetic Aspects of Machado-Joseph Disease. *Brotéria-Genética* 2:137-147

Sequeiros J, Coutinho P (1993) Epidemiology and Clinical Aspects of Machado-Joseph Disease. In: A. Harding, T. Deufel and S. Chamberlain (ed) *Advances in Neurology*, Raven Press, New York, pp 139-153

Sequeiros J, Maciel P, Taborda F, Ledo S, Rocha JC, Lopes A, Reto F, Fortuna AM, Gaspar C, Fleming M, Coutinho P, Rouleau GA, Santos Jorge C (In press) Prenatal diagnosis of Machado-Joseph disease by direct mutation analysis: a protocol for dominant ataxias and other late-onset disorders.

Sequeiros J, Nance MA (1997) Offer and utilization of genetic testing in the dominant ataxias: towards guidelines. *47th Annual Meeting of the American Society of Human Genetics, Baltimore.*

Sequeiros J, Oliveira J, Teixeira CM, Coutinho P, Sousa A (1994a) Quantitative genetic studies of anticipation of age-at-onset in Machado-Joseph disease. *International Workshop on Friedreich's ataxia and dominant ataxias, Tunis.*

Sequeiros J, Silveira I, Maciel P, Coutinho P, Manaia A, Gaspar C, Bulet P, Loureiro L, Guimarães J, Tanaka H, Takiyama Y, Sakamoto H, Nishizawa M, Nomura Y, Segawa M, Tsuji S, Melki J, Munnich A (1994b) Genetic linkage studies of Machado-Joseph disease with chromosome 14q STRPs in 16 Portuguese-Azorean kindreds. *Genomics* 21:645-648

Sharp AH, Loev SJ, Schilling G, Li S-H, Li X-J, Bao J, Wagster MV, Kotzuk JA, Steiner JP, Lo A, Hedreen J, Sisodia S, Snyder SH, Dawson TM, Ryugo DK, Ross CA (1995) Widespread expression of Huntington's disease gene (IT15) protein product. *Neuron* 14:1065-1074

Silveira I, Coutinho P, Maciel P, Gaspar C, Hayes S, Dias A, Guimarães J, Loureiro L, Maciel P, Sequeiros J, Rouleau GA (submetido a publicação) Analysis of SCA1, SCA2, MJD and SCA6 CAG repeats in 48 Portuguese ataxia families.

Silveira I, Lopes-Cendes I, Kish S, Maciel P, Gaspar C, Coutinho P, Botez MI, Teive H, Arruda W, Steiner CE, Pinto-Junior W, Maciel JA, Jain S, Sack G, Andermann E, Sudarsky L, Rosenberg RN, MacLeod P, Sequeiros J, Rouleau GA (1996) Frequency of spinocerebellar ataxia type 1, dentatorubropallidoluysian atrophy and Machado-Joseph disease mutations in a large group of spinocerebellar ataxia patients. *Neurology* 46:214-218

Silveira I, Manaia A, Melki J, Magariño C, Lunkes A, Hernandez A, Gispert S, Bulet P, Rozet J-M, Coutinho P, Leal Loureiro JE, Guimarães J, Auburger G, Munnich A, Sequeiros J (1993)

Machado-Joseph disease is genetically different from Holguin dominant ataxia (SCA2). *Genomics* 17:556-559

Snow K, Doud LK, Hagerman R, Pergolizi RG, Erster SH, Thibodeau SN (1993) Analysis of a CGG sequence at the FMR-1 locus in fragile X families and in the general population. *American Journal of Human Genetics* 53:1217-1228

Snow K, Tester DJ, Kruckeberg KE, Thibodeau SN (1994) Sequence analysis of the Fragile X trinucleotide repeat: correlations with stability and haplotype, and implications for the origin of Fragile X alleles. *American Journal of Human Genetics* 55:A243

Soong B-w, Cheng C-h, Liu R-s, Shan D-e (1997) Machado-Joseph disease: clinical, molecular and metabolic characterization in chinese kindreds. *Annals of Neurology* 41:446-452

Sorbi S, Piacenti S, Marini P (1988) Lack of efficacy of phosphatidylcholine in ataxias. *Neurology* 38:649-650

Sousa A, Coutinho P, Oliveira J, Teixeira CM, Silva RM, Sudarsky LA, Dawson DM, Rosenberg RN, MacLeod PM, Sequeiros J (1997) Anticipation of age-at-onset in Machado-Joseph disease-a quantitative study. *1st International Symposium on Inherited Ataxias, Montreal*.

Squitieri F, Andrew SE, Goldberg YP, Kremer B, Spence N, Zeisler J, Nichol K, Theilmann J, Greenberg J, Goto J, Kanazawa I, Vesa J, Peltonen L, Almqvist E, Anvret M, Telenius H, Lin B, Napolitano G, Morgan K, Hayden MR (1994) DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. *Human Molecular Genetics* 3:2103-2114

Stevanin G, Cancel G, Didierjean O, Durr A, Abbas N, Cassa E, Feingold J, Agid Y, Brice A (1995a) Linkage disequilibrium at the Machado-Joseph disease/spinal cerebellar ataxia 3 locus: evidence

for a common founder effect in French and Portuguese-Brazilian families as well as a second ancestral Portuguese-Azorean mutation. *American Journal of Human Genetics* 57:1247-1250

Stevanin G, Cassa E, Cancel G, Abbas N, Durr A, Jardim E, Agid Y, Sousa PS, Brice A (1995b) Characterization of the unstable expanded CAG repeat in the MJD1 gene in four Brazilian families of Portuguese descent with Machado-Joseph disease. *Journal of Medical Genetics* 32:827-830

Stevanin G, Guern EL, Ravisé N, Chneiweiss H, Durr A, Cancel G, Vignal A, Boch A-L, Ruberg M, Penet C, Pothin Y, Lagroua I, Haguenu M, Rancurel G, Weissenbach J, Agid Y, Brice A (1994) A Third Locus for Autosomal Dominant Cerebellar Ataxia Maps to Chromosome 14q24.3-qter: Evidence for the Existence of a Fourth Locus. *American Journal of Human Genetics* 54:11-20

Stevanin G, Lebre A-S, Mathieux C, Cancel G, Abbas N, Didierjean O, Durr A, Trottier Y, Agid Y, Brice A (1997) Linkage disequilibrium between the spinocerebellar ataxia 3/Machado-Joseph disease mutation and two intragenic polymorphisms, one of which, (X359Y), affects the stop codon. *American Journal of Human Genetics* 60:1548-1552

Stevanin G, Trottier Y, Cancel G, Durr A, David G, Didierjean O, Burk K, Imbert G, Saudou F, Abada-Bendib M, Gourfinkel-An I, Benomar A, Abbas N, Klockgether T, Grid D, Agid Y, Mandel J-L, Brice A (1996) Screening for proteins with polyglutamine expansions in autosomal dominant cerebellar ataxias. *Human Molecular Genetics* 5:1887-1892

Stine OC, Li S-H, Pleasant N, Wagster MV, Hedreen JC, Ross CA (1995) Expression of the mutant allele of IT-15 (the HD gene) in striatum and cortex of Huntington's disease patients. *Human Molecular Genetics* 4:15-18

Stine OC, Pleasant N, Wagster MV, Kasch L, Hedreen J, Ross CA (1994) Somatic instability of the expanded allele of IT-15 from patients with Huntington disease. *American Journal of Human Genetics* 55:A244

Strand M, Prolla TA, Liskay RM, Petes TD (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. *Nature* 365:274-276

Strong T, Tagle DA, Valdes JM, Elmer LW, Boehm K, Swaroop M, Kaatz KW, Collins FS, Albin RL (1993) Widespread expression of the human and rat Huntington's disease gene in brain and nonneural tissues. *Nature Genetics* 5:259-265

Subramony SH, Currier RD, Jackson JF, Collins GM, Jackson MS (1990) Presumed Joseph disease in three new kindreds and response to anti-Parkinson drugs. *Annals of Neurology* 28:233

Subramony SH, King WM (1997) Phenotypic differences between spinocerebellar ataxia 1 and spinocerebellar ataxia 3 detected by posturography and electronystagmography. *1st International Symposium on Inherited Ataxias, Montreal.*

Takano H, Onodera O, Takahashi H, Igarashi S, Yamada M, Oyake M, Ikeuchi T, Reiji K, Tanaka H, Iwabuchi K, Tsuji S (1996) Somatic mosaicism of expanded CAG repeats in brains of dentatorubral-pallidoluysian atrophy (DRPLA): cellular population-dependent dynamics of mitotic instability. *American Journal of Human Genetics* 58:1212-1222

Takiyama Y, Igarashi S, Rogaeva EA, Endo K, Rogaev EI, Tanaka H, Sherrington R, Sanpei K, Liang Y, Saito M, Tsuda T, Takano H, Ikeda M, Lin C, Chi H, Kennedy JL, Lang AE, Wherrett JR, Segawa M, Nomura Y, Yuasa T, Weissenbach J, Yoshida M, Nishizawa M, Kidd KK, Tsuji S, St George-Hyslop PH (1995) Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. *Human Molecular Genetics* 4:1137-1146

Takiyama Y, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, Shimazaki H, Soutome M, Endo K, Ohta S, Kagawa Y, Kanazawa I, Mizuno Y, Yoshida M, Yuasa T, Horikawa

y, Oyanagi K, Nagai H, Kondo T, Inuzuka T, Onodera O, Tsuji S (1993) The gene for Machado-Joseph disease is mapped to chromosome 14q. *Nature Genetics* 4:300-304

Takiyama Y, Sakoe K, Soutome M, Namekawa M, Ogawa T, Nakano I, Igarashi S, Oyake M, Tanaka H, Tsuji S, Nishizawa M (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. *Human Molecular Genetics* 6:1063-1068

Telenius H, Kremer B, Goldberg YP, Jane Theilmann J, Andrew SE, Zeisler J, Adam S, Greenberg C, Ives EJ, Clarke LA, Hayden M (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nature Genetics* 6:409-414

The Huntington Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72:971-983

Trottier Y, Cancel G, Zeder-lutz G, Devys D, Imbert G, Lunkes A, Yvert G, Tora L, Lutz Y, Hirsch E, Brice A, Mandel J-L (1997) Effect of polyglutamine expansion in mutated proteins causing Huntington's disease and spinocerebellar ataxia 3. *1st International Symposium on Inherited Ataxias, Montreal.*

Trouillas P, Brudon F, Adeleine P (1988) Improvement of cerebellar ataxia with levorotatory form of 5-hydroxytryptophan. A double-blind study with double data processing. *Archives of Neurology* 45:1217-1222

Tuite P, Rogaeva E, St George-Hyslop P, Lang A (1995) Dopa-responsive Parkinsonism phenotype of Machado-Joseph disease: confirmation of 14q CAG expansion. *Annals of Neurology* 38:684-687

Twist EC, Causaubon LK, Rutledge M, Rao VS, MacLeod PM, Radvany J, Zhao Z, Rosenberg RN, Farrer LA, Rouleau GA (1995) Machado-Joseph disease maps to the same region of chromosome 14 as the spinocerebellar ataxia type 3 locus. *Journal of Medical Genetics* 31:823-829

Twist EC, Farrer LA, MacLeod PM, Radvany J, Chamberlain S, Rosenberg RN, Rouleau GA (1994) Machado-Joseph disease is not an allele of the spinocerebellar ataxia 2 locus. *Human Genetics* 93:335-338

Verkerk AJMH, Pieretti M, Sutcliffe JS, Fu Y-H, Kuhl DPA, Pizzuti A, Reiner O, Richards S, Victoria MF, Zhang F, Eussen BE, Ommen G-JB, Blonden LAJ, Riggins GJ, Chastain JL, Kunst CB, Galjaard H, Caskey CT, Nelson DL, Oostra BA, Warren ST (1991) Identification of a Gene (FMR-1) Containing a CGG Repeat Coincident with a Breakpoint Cluster Region Exhibiting Variation in Fragile X Syndrome. *Cell* 65:905-914

Wang Z, Weber JL (1992) Continuous Linkage Map of Human Chromosome 14 Short Tandem Repeat Polymorphisms. *Genomics* 13:532-536

Warner JP, Barron LH, Fitzpatrick DR, Brock DJH (1996) A gene conversion event at the Huntington's CAG repeat. *American Journal of Human Genetics* 59:1697

Watanabe M, Abe K, Aoki M, Kameya T, Kaneko J, Shoji M, Ikeda M, Shizuka M, Ikeda Y, Iizuka T, Hirai S, Itoyama Y (1996) Analysis of CAG trinucleotide expansion associated with Machado-Joseph disease. *Journal of the Neurological Sciences* 136:101-107

Weber J, May PE (1989) Abundant Class of Human DNA Polymorphisms which can be typed using the Polymerase Chain Reaction. *American Journal of Human Genetics* 44:388-396

Wells RD (1996) Molecular basis of genetic instability of triplet repeats. *Journal of Biological Chemistry* 271:2875-2878

Wiegant J, Kalle W, Mullenders L, Brookes S, Hoovers JMN, Dauwerse JG, Ommen GJB, Raap AK (1992) High-resolution in situ hybridization using DNA halo preparations. *Human Molecular Genetics* 1:587-591

Woods BT, Schaumburg HH (1972) Nigro-spinal-dentatal degeneration with nuclear ophthalmoplegia: a unique and partially treatable clinico-pathological entity. *Journal of Neurological Sciences* 17:149-166

Yanagisawa H, Fujii K, Nagafuchi S, Nakahori Y, Nakagome Y, Akane A, Nakamura M, Sano A, Komure O, Kondo I, Jin DK, Sorensen SV, Potter NT, Young SR, Nakamura K, Nukina N, Nagao Y, Tadokoro K, Okuyama T, Miyashita T, Inoue T, Kanazawa I, Yamada M (1996) A unique origin and multistep process for the generation of expanded DRPLA triplet repeats. *Human Molecular Genetics* 5:373-379

Yu S, Mulley J, Loesch D, Turner G, Donnelly A, Gedeon A, Hillen D, Kremer E, Lynch M, Pritchard M, Sutherland GR, Richards RI (1992) Fragile-X syndrome: unique genetics of the heritable unstable element. *American Journal of Human Genetics* 50:968-980

Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Baker E, Holman K, Mulley JC, Warren ST, Schlessinger D, Sutherland GR, Richards RI (1991) Fragile X genotype characterized by an unstable region of DNA. *Science* 252:1179-1181

Yuasa T, Miyatake T, Ohama E, Coutinho P, Sequeiros J, Leite I, Andrade C (1990) A comparative study of Machado-Joseph disease in Portuguese and Japanese. *Shinkei Kenkyu no Shinpo* 34:102-112

Zhou YX, Takiyama Y, Igarashi S, Li YF, Zhou BY, Gui DC, Endo K, Tanaka H, Chen ZH, Zhou LS, Fan MZ, Yang BX, Weissenbach J, Wang GX, Tsuji S (1997) Machado-Joseph disease in

four Chinese pedigrees: molecular analysis of 15 patients including two juvenile cases and clinical correlations. *Neurology* 48:482-485

Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Stockton DW, Amos C, Dobyns WB, Subramony SH, Zoghbi HY, Lee CC (1996) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha1A-voltage-dependent calcium channel. *Nature Genetics* 15:62-69

Zuhlke C, Riess O, Bockel B, Lange H, Thies U (1993) Mitotic stability and meiotic variability of the (CAG)_n repeat in the Huntington disease gene. *Human Molecular Genetics* 2:2063-2067

Abreviaturas utilizadas

ADN- Ácido desoxirribonucleico

ADRPL- Atrofia dentatorubropalidoluisiana

ARN-Ácido ribonucleico

DMJ-Doença de Machado-Joseph

GAPDH- *Glyceraldehyde 3-phosphate dehydrogenase* (Gliceraldeído 3-fosfato desidrogenase)

PCR-*Polymerase chain reaction*

RT-PCR-*Reverse transcription+ Polymerase chain reaction*

SCA 1-*Spinocerebellar ataxia type 1* (ataxia espinocerebelosa tipo 1)

SCA 2-*Spinocerebellar ataxia type 2* (ataxia espinocerebelosa tipo 2)

SCA 3-*Spinocerebellar ataxia type 3* (ataxia espinocerebelosa tipo 3)

SCA 4-*Spinocerebellar ataxia type 4* (ataxia espinocerebelosa tipo 4)

SCA 5- *Spinocerebellar ataxia type 5* (ataxia espinocerebelosa tipo 5)

SCA 6-*Spinocerebellar ataxia type 6* (ataxia espinocerebelosa tipo 6)

SCA 7-*Spinocerebellar ataxia type 7* (ataxia espinocerebelosa tipo 7)

SNC- Sistema nervoso central

TBP- *TATA binding protein* (proteína de ligação à *TATA box*)

ANEXO

**Descrição das famílias e
generalidades sobre os métodos utilizados**

ANEXO: Descrição das famílias e generalidades sobre os métodos utilizados

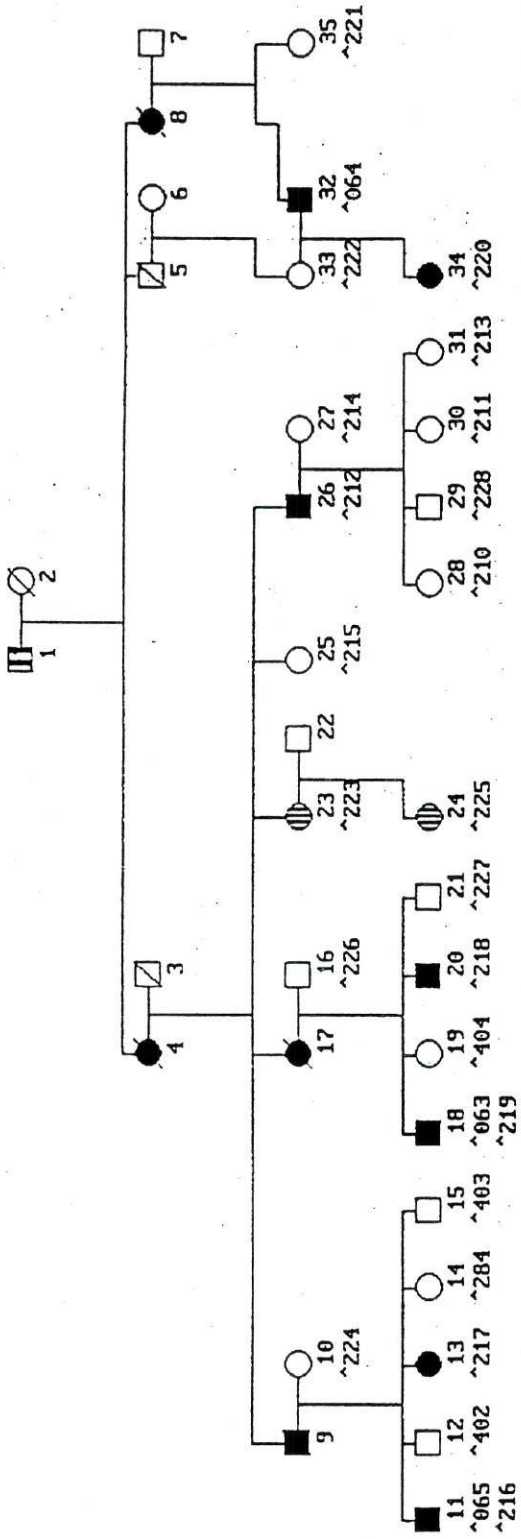
Descrição das famílias

Para os estudos genéticos da DMJ, foram colhidas amostras de sangue de 551 indivíduos, incluindo 212 afectados, pertencentes a 26 famílias açorianas e de Portugal continental. As árvores familiares das 16 famílias seleccionadas para os estudos de ligação a marcadores do cromossoma 14 podem ver-se na figura 5.

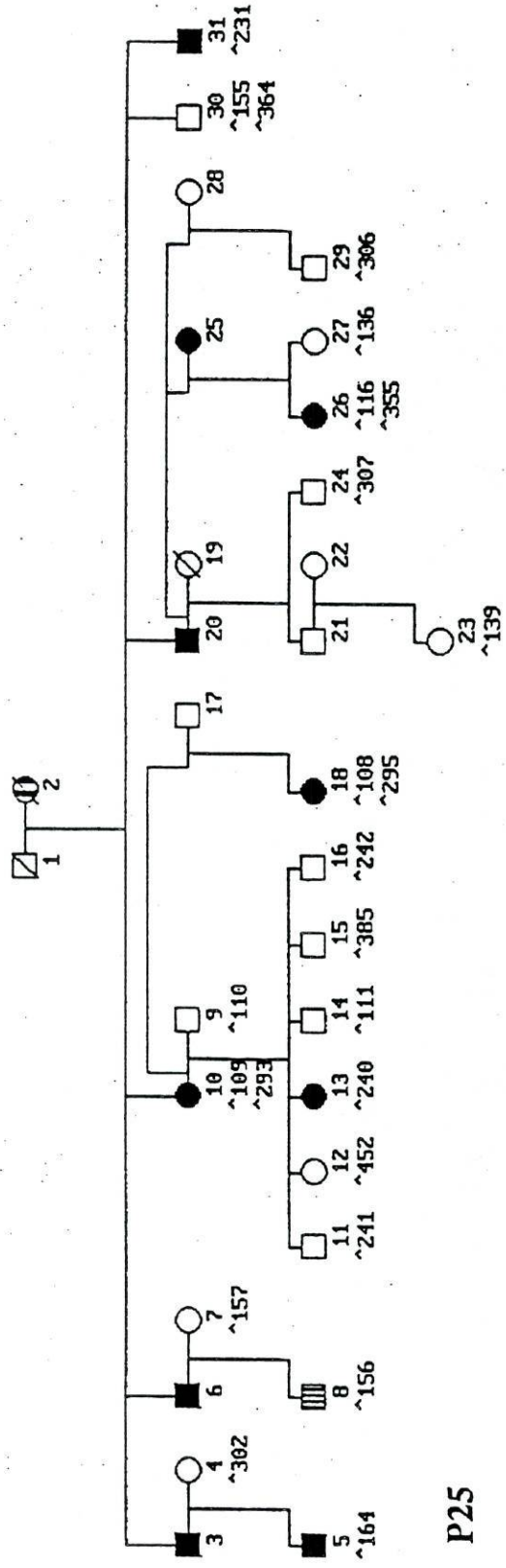
Nos estudos de correlação genótipo-fenótipo, incluíram-se 33 famílias, 15 portuguesas (açorianas), 16 norte-americanas de origem açoriana e 2 brasileiras de origem presumidamente portuguesa, uma das quais a grande família de Santa Catarina (B1), previamente descrita na literatura [Radvany, 1993 #297].

Os estudos de correlação entre a patologia e o mosaicismo foram efectuados em *specimens post-mortem*, tendo sido obtidas amostras de 20 regiões do SNC, de 3 doentes DMJ e 3 doentes SCA1 (com confirmação molecular da mutação), do banco de cérebros de Stephen Kish, Toronto. Foi obtida informação detalhada sobre a patologia nestes doentes, assim como informação clínica, sempre que possível, e a idade de morte do doente. Para a análise da correlação entre a idade de morte e o grau de mosaicismo estudaram-se *specimens* de cortex e substância branca cerebral de 11 doentes com DMJ e 7 doentes com SCA1.

Nos estudos da instabilidade genética foram utilizados todos pares transmissor-descendente para os quais se dispunha dos tamanhos dos (CAG)_n expandidos, incluindo indivíduos pertencentes às famílias portuguesas P1, P8, P14, P16, P25, P32, às famílias brasileiras B1, H e B3, às famílias norte-americanas (de origem açoriana) N2, N3a, N3b, N4, N5, N6, N10 e N13 e a duas famílias de origem chinesa, C1 e C2 (que contribuem com apenas 4 pares).



P17



P25

Estudo clínico

O estudo clínico das famílias foi efectuado por neurologistas com experiência na área da DMJ ou das ataxias em geral. As famílias portuguesas foram na sua totalidade observadas por Paula Coutinho e/ou Jorge Sequeiros, com a vantagem de se conseguir uma grande homogeneidade nos critérios de estudo, tais como a definição de idades de início e subtipos. As famílias americanas foram estudadas por diversos clínicos, em particular David Dawson, Lewis Sudarsky, Lee Corwin e Patrick MacLeod. Na família brasileira B1, dado que a descrição da apresentação clínica era feita em detalhe na literatura [Radvany, 1993 #297], fez-se uma conversão dessa informação para a classificação em subtipos (Isclia Lopes-Cendes).

Estudo patológico

Os estudos patológicos de quantificação da degenerescência neuronal nas diversas regiões do cérebro em doentes com DMJ ou SCA1 foram efectuados por patologistas com alguma experiência prévia com ataxias e doença de Huntington (Dr Stephen Kish, de Toronto, Dr Robitaille, de Montreal, e colaboradores).

Métodos de extracção do ADN

O método utilizado na extracção do ADN genómico a partir de linfócitos foi o método de extracção pelo fenol [Sambrook, 1989 #317], incluindo as seguintes etapas: 1. separação das células do plasma, 2. lise dos eritrócitos, 3. centrifugação e recolha dos glóbulos brancos, 4. digestão pela protease, 5. extracção por fenol, 6. obtenção do ADN genómico.

Quando se extraiu ADN genómico de material patológico congelado (*specimens* de SNC), utilizou-se uma variante do método, em que na fase inicial se tritura o material a baixas temperaturas na presença de uma solução tampão apropriada, que inibe reacções enzimáticas indesejáveis (solução de Warner); deixou-se também a protease actuar mais tempo do que o habitual.

Estudos de ligação genética (linkage)

A distância genética entre dois *loci* é um parâmetro determinado pela frequência das recombinações meióticas ou "*crossing-over*". As recombinações são eventos absolutamente aleatórios e imprevisíveis, que obedecem no entanto a uma lei estatística: a probabilidade de ocorrer uma recombinação entre dois *loci* depende da distância física que os separa. Quanto maior esta distância, mais possibilidades existem de ocorrer uma recombinação entre os dois *loci*. Inversamente, quanto menor esta distância, mais diminui a possibilidade de ocorrer uma recombinação - diz-se então que os *loci* estão geneticamente *ligados*. Esta ligação genética só é mensurável por uma observação empírica: a co-segregação dos dois *loci* ao longo de gerações sucessivas sem mudança de fase. Para se determinar a distância genética entre dois *loci* (e a existência ou não de ligação genética) é necessário proceder a uma análise familiar em que se examina o acoplamento de marcadores num dado cromossoma. É assim determinada a taxa de recombinação entre os dois *loci*, θ , que é proporcional à distância física entre eles. No caso em que se procura a localização cromossômica do gene causador de uma doença, um dos *loci* seria o da doença, e o outro o marcador genético em estudo.

Um primeiro requisito para este tipo de estudo é que os *loci* alélicos devem ser distinguíveis, por um marcador genotípico (como um RFLP, um VNTR ou um microsatélite) ou um marcador fenotípico como um "traço" patológico. O segundo requisito é que o número de acontecimentos que podem dar lugar a uma recombinação, ou seja, o número de meioses, seja suficientemente grande para permitir que a análise estatística seja significativa.

Actualmente, existem programas de computador que permitem calcular a probabilidade de dois *loci* serem ligados e de se situarem a uma dada distância genética. A análise é feita pelo método dos MLE (*Maximum Likelihood Estimates*), ou o cálculo da verosimilhança máxima. O logaritmo ("l") do quociente das probabilidades ("*odds*") - quociente entre o valor da probabilidade de o *locus* da doença se encontrar a uma distância θ do marcador e a probabilidade de não-ligação ($\theta=0.50$) - é estimado para cada marcador estudado. A utilização destes índices de ligação, designados *lod scores* ou *lods*

(Z), como logaritmos das probabilidades permite que estes sejam aditivos, e não multiplicáveis; assim a análise é sequencial, família a família, até à obtenção dos valores necessários à afirmação de uma ligação ($Z \geq +3.00$) ou à sua exclusão ($Z \leq -2.00$), para um dado valor de θ . O valor de θ dá-nos a proximidade entre os *loci* testados: $\theta=0.00$ indica que o *locus* da doença e o do marcador são o mesmo ou tão próximos que não existe recombinação entre eles; $\theta=0.5$ significa que os dois *loci* não estão ligados. Valores intermédios indicam a proximidade entre os *loci*.

Se na análise se consideram apenas dois *loci* (dois marcadores genéticos ou o *locus* da doença e um marcador genético), trata-se de uma análise *pairwise*; é no entanto possível incorporar informação sobre vários marcadores, numa análise de *multipoint*, em que se obtém uma estimativa das taxas de recombinação entre múltiplos *loci*.

A análise de ligação genética pode ser complicada pela existência de heterogeneidade não-alélica na população em estudo (dois ou mais *loci* independentes causando um mesmo fenótipo). Assim, é frequente levantarem-se as seguintes questões: (1) dado um teste de ligação positivo na população total, existe evidência de que uma proporção das famílias segregue o gene ligado e outra segregue um gene não-ligado? (2) apesar de não existirem provas da existência de ligação genética assumindo homogeneidade entre as famílias, poderia demonstrar-se ligação genética caso se permitisse uma percentagem de heterogeneidade? O programa HOMOG pode ser utilizado para testar ambas estas hipóteses [Ott, 1991 #688].

A utilização dos programas do pacote "Linkage", desenvolvidos por Ott e colaboradores [Ott, 1991], apresenta numerosas vantagens, uma vez que permite considerar parâmetros como a frequência dos genes e a taxa de mutação para o *locus* da doença, o fenótipo (afectado ou não afectado), a penetrância do gene dependente da idade (classificação em classes de risco segundo a idade actual das pessoas em risco assintomáticas, particularmente relevante nesta doença de início tardio e muito variável) e o tratamento das ansas de consanguinidade e de casamento em cada árvore genealógica.

Finalmente, é importante não confundir a noção de ligação genética, que exprime a proximidade de dois *loci* no ADN sem pressupostos sobre os alelos presentes, e a

noção de desequilíbrio de ligação que é uma associação preferencial entre certos alelos de dois *loci*. O desequilíbrio de ligação pode permitir a datação e a demonstração da origem geográfica de certas doenças monogénicas (monoalélicas), e saber se o evento mutacional ocorreu uma única vez num único indivíduo e se difundiu por migração - mutação fundadora - ou se as mutações foram múltiplas e independentes, em regiões separadas.

PCR (polymerase chain reaction)

Esta técnica permite amplificar selectivamente *in vitro* um dado segmento de ADN, de forma a que, partindo de amostras contendo quantidades ínfimas dessa sequência, se consegue originar ADN suficiente para permitir a sua análise em algumas horas. Para isso, é necessário conhecer-se a sequência das regiões flanqueantes do segmento a amplificar e escolher 2 fragmentos nucleotídicos (*primers*) capazes de emparelhar com elas e assim iniciar a replicação do ADN pela *ADN polimerase*. Misturando os *primers* com o ADN genómico, em condições de hibridação, estes posicionam-se face às suas sequências complementares respectivas. Em seguida, por acção da ADN polimerase, cada fragmento é alongado, originando uma sequência exactamente complementar à da cadeia original. Resulta então uma duplicação da sequência considerada, uma vez que cada cadeia é copiada de novo. A operação é em seguida recomeçada, consistindo cada ciclo em três etapas: (a) um período de desnaturação, a 94°C, em que as duas cadeias de ADN são separadas, permitindo o acesso dos *primers*; (b) um período de hibridação entre os *primers* e as cadeias de ADN genómico (*annealing*), em que a temperatura é seleccionada conforme a sequência destes, de forma a tornar o emparelhamento específico; e (c) um período de extensão das cadeias complementares.

Cada etapa dura em geral cerca de um minuto. Como cada ciclo produz uma duplicação da sequência compreendida entre os dois primers, a amplificação é exponencial; ao fim de n ciclos obtêm-se teoricamente 2^n exemplares do segmento de ADN desejado. Este método é muito simples porque não é necessária outra manipulação além da variação de temperatura, sendo todos os ciclos realizados num mesmo tubo. A

introdução de uma ADN polimerase termo-resistente, a *Taq polimerase*, evita ter que se adicionar enzima a cada ciclo e aumenta a especificidade de hibridação dos *primers* porque se podem utilizar temperaturas mais elevadas na etapa de *annealing*. Pela sua simplicidade, esta técnica presta-se bem à automatização, existindo aparelhos capazes de efectuar as variações térmicas de forma automática e programável.

A PCR pode ser utilizada na detecção directa de mutações ou de polimorfismos de formas variadas. (1) A amplificação por PCR pode ser seguida de digestão com enzimas de restrição apropriadas à detecção do polimorfismo. (2) Na PCR específica de alelo, escolhem-se os *primers* de forma a que um deles tenha na extremidade 3' (determinante para a especificidade do oligonucleotídeo) o nucleotídeo variante e desenvolvem-se as condições de amplificação de forma a que esta seja específica apenas para um alelo (o de sequência exactamente complementar à do *primer* utilizado); se o indivíduo tiver a outra variante do polimorfismo, não se dá amplificação (é importante ter controlos positivos neste tipo de teste, pelo que normalmente se usa um *primer* para cada variante).

Na RT-PCR, realiza-se uma etapa prévia de "transcrição reversa", em que se obtêm cadeias simples de DNA a partir do RNA mensageiro (utilizando uma enzima chamada transcriptase reversa, presente em certos vírus de ARN), seguida de uma amplificação por PCR do ADNc assim obtido. Isto permite uma análise a nível do ARNm (determinação da sequência, presença/ausência de expressão, presença de formas truncadas ou formas de *splicing* alternativas), em que se ultrapassa o problema da quantidade de material presente na célula: mesmo genes cuja expressão a nível do ARN é muito baixa podem ser analisados por este método.

Na análise de *SSCP* (*single strand conformation polymorphism*) para determinação de mutações ou de polimorfismos, após amplificação por PCR do fragmento de ADN que contém a mutação ou o polimorfismo em estudo, desnaturam-se os produtos de PCR, que se separam em seguida por electroforese em gel de poli-acrilamida, em condições que permitem a auto-renaturação das cadeias, o que vai

originar diferentes conformações terciárias do ADN (*hairpins* e outros), dependentes da sequência primária deste; as diferentes conformações podem apresentar velocidades de migração diferentes, permitindo a distinção das formas variantes do gene.

Southern Blot

O objectivo da técnica de Southern *blot* é visualizar uma sequência única de ADN. Para isso, o ADN é fragmentado por uma ou várias enzimas de restrição apropriadas, os fragmentos de tamanhos diferentes são separados por electroforese em gel de agarose, desnaturados *in situ*, transferidos por capilaridade para um suporte sólido - filtro de *nylon* ou de nitrocelulose e hibridados com uma sonda marcada radioactivamente. Os fragmentos reconhecidos pela sonda formam com esta um *duplex* resistente à lavagem exaustiva e revelam-se por um sinal persistente a nível do filtro.

Clonagem

A clonagem é uma técnica *in vitro* para obter múltiplas cópias idênticas de um dado fragmento de ADN. O primeiro passo é isolar o fragmento de interesse, que em seguida é introduzido num vector de clonagem (plasmídeo, bacteriófago, cosmídeo, YAC, BAC). Estes vectores replicam-se autonomamente numa célula bacteriana hospedeira (*Escherichia coli*), ou levedura (*Saccaromyces cerevisiae*) no caso de YACs, o que permite multiplicar indefinidamente o número de cópias desse fragmento de ADN e ultrapassar assim as limitações quantitativas à sua análise.

No trabalho aqui apresentado recorreu-se à clonagem de fragmentos de ADN_c, amplificados por RT-PCR a partir do ADN genómico de indivíduos, num vector com características especiais que permitem a clonagem imediata de um produto de PCR obtido com a Taq polimerase (*TA vector, Invitrogen*). Esta enzima acrescenta um nucleotídeo extra (A) à extremidade 5' dos fragmentos de ADN por ela produzidos; o vector tem uma extremidade coesiva T, aonde os produtos de PCR se ligam, circularizando-se assim o ADN recombinante. O plasmídeo confere resistência à ampicilina, e actividade de Beta-galactosidase, que é eliminada por inserção do fragmento de ADN, permitindo assim a

selecção dos *clones* de interesse através do crescimento em meio com ampicilina e a eliminação de clones contendo plasmídeo recircularizado (sem *insert*) por simples observação da cor das colónias. Um outro vector utilizado na clonagem dos ADNc do gene MJD1 (sequências apresentadas nas figuras 3.a e 3.b) foi o plasmídeo pBluescriptII SK (+).

Utilizaram-se também os cosmídeos sCos1 [Evans *et al.*, 1989] do banco de ADN genómico humano do cromossoma 14. Este vector confere resistência à ampicilina e à kanamicina.

Uma vez introduzidas as moléculas de ADN recombinante em bactérias, leva-se a cabo a sua multiplicação em meio de cultura a 37°C, e extrai-se o ADN do vector pelo processo designado *Miniprep* (de cosmídeos ou de plasmídeos), que consiste basicamente na lise celular bacteriana, seguida da precipitação das proteínas e ADN genómico bacteriano, que se separa do ADN do vector e do ARN nesta fase, digestão do ARN e, finalmente, a separação e lavagem do ADN do vector (protocolo adaptado de Sambrook, 1989).

Sequenciação do ADN

No trabalho aqui apresentado utilizou-se uma variante do método de sequenciação de Sanger, em que se leva a cabo a síntese da cadeia complementar à sequência que se deseja conhecer, na presença de um nucleotídeo marcado radioactivamente e de análogos dideoxi de nucleotídeos em baixa concentração. Estes últimos, dada a sua baixa concentração, incorporam-se raramente e de modo aleatório na cadeia complementar, e, por carecerem do grupo 3' hidroxilo, interrompem a síntese desta. Esta interrupção aleatória origina fragmentos de tamanhos diversos, que correspondem à distância entre o ponto de interrupção e a extremidade inicial da cadeia, e que são analisados em gel de poliacrilamida desnaturante. Como a resolução destes géis é de uma base, o encadeamento de fragmentos na autoradiografia corresponde ao encadeamento de bases no fragmento de ADN. Este método exige que a sequência de interesse seja de cadeia simples e purificada. Para se obter uma amostra nestas condições,

pode proceder-se à subclonagem do fragmento a sequenciar num vector apropriado, de cadeia simples, o que permite também obter as quantidades de ADN necessárias às reacções envolvidas na sequenciação. Alternativamente, podem utilizar-se várias combinações das técnicas de amplificação por PCR com o método de Sanger, com a vantagem de se necessitar apenas de uma pequena quantidade inicial de DNA, e de permitir ultrapassar a etapa de clonagem.

Agradecimentos

Aos meus pais, à Rita e à Ana Paula, por esperarem sempre o melhor de mim, por todo o encorajamento e apoio, em tudo e desde sempre.

Aos avós, tios e tias, primos e primas, por todo o carinho. Obrigada pelos telefonemas, postais, bolinhos, amêndoas, todo o mimo da família, que me ajudou sempre a aguentar os momentos mais difíceis e com quem gosto de partilhar os momentos mais alegres.

Aos amigos de sempre: à Dora, à Teresa, ao Maurício, à Claudia, ao Zé e à Marta.

To my new old-friends-to-be in Montreal: thanks Albéna, Ana Luisa and Elaine for a lot of happy moments. When will we do our Thelma and Louise thing? Thanks to Chris for lending me his CDs, his tape recorder, and, most of all, for the beautiful fluorescent green wooly hat!

À Dra. Alda, pelo acompanhamento e encorajamento na parte final da escrita da tese.

Ao Prof. Pedro Pinho Costa, que ainda antes da minha entrada na Universidade contribuiu grandemente para o meu entusiasmo pela Ciência e para a minha decisão de seguir a via da investigação, por toda a simpatia e apoio que me deu e por demonstrar que, embora talvez com maior dificuldade do que em outros sítios, é possível fazer-se Ciência a sério em Portugal.

To all the students and technicians at the Rouleau lab, for all the help and useful interaction. Thanks in particular to Ron, Iscia and Jaime for all the advice.

Às Profs. Maria João Saraiva e Paula Coutinho, que desde sempre estiveram disponíveis com a maior paciência e simpatia, para discutir os projectos e artigos e me aconselharem no rumo a seguir em vários momentos decisivos.

To Guy, for giving me the opportunity to work and learn in an excellent lab.

Ao Jorge, pela oferta de uma excelente oportunidade de aprendizagem, e pelo óptimo ambiente de trabalho.