Genetic and molecular investigation of the spinocerebellar ataxias

Sean I. A. Hayes

Department of Biology
McGill University, Montreal

August, 1999

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements of the degree of Master of Science.

Acknowledgments

I would like to express my sincere appreciation to everyone who helped me make through my Master's studies. Special thanks to my supervisor, Dr. G. A. Rouleau, for his expert advice, patience (and patients), and guidance throughout my time in Montreal. I would also like to thank the members of my supervisory committee, Dr. L. Pinsky and Dr. P. Lasko.

Very special thanks to all of my fellow students, Postdoctoral fellows, technicians, and support staff who helped me, guided me, and taught me along the way. In particular, I would like to thank Iscia Lopes-Cendes, Gustavo Turecki, Ron Lafrenière, Jillian Parboosingh, Patrícia Maciel, Xueping Fan, Zoha Kibar, Heidi Howard, Caroline Fournier, Patrick Dion, Collette Hand, Claude Marineau, Katéri Brisebois, Julie Roussel, Marylène Rousseau, Daniel Rochefort, and Yvonne Gardner. Special thanks to Angie Stocco and Claude Marineau for translating my abstract into French.

For their help, support, and friendship both inside and outside of the lab, I would like to give very special thanks to Albéna Pramatarova and Claudia Gaspar. I would not have made it without you.

Thank you to my family and friends for their support and encouragement throughout, and, of course, the biggest thanks of all to Audrey. Without you I would be lost.

Table of Contents

Contributions of authors

Chapter 2

Hayes, S., Lopes-Cendes, I., Gaspar, C., Jain, S., and Rouleau, G.A. Haplotype analysis in a group of spinocerebellar ataxia type 2 (SCA2) patients.

S. Hayes performed the experiments and wrote the manuscript. I. Lopes-Cendes previously identified all of the SCA2 positive families. S. Jain provided DNA samples. C. Gaspar and G.A. Rouleau contributed to experimental design and analysis.

Chapter 3

Hayes, S., Lopes-Cendes, I., Gaspar, C., Maciel, P., Silviera, I., Rousseau, M., and Rouleau, G. A. Incidence of the SCA6 and SCA7 repeat mutations in a large panel of ataxia patients.

S. Hayes performed the experiments and wrote the manuscript. I. Lopes-Cendes, C. Gaspar, P. Maciel, I. Silviera, and M. Rousseau previously screened for other disease mutations. G. A. Rouleau contributed to experimental design.

Chapter 4

Hayes, S., Turecki, G., Brisebois, K., Lopes-Cendes, I., Gaspar, C., Riess, O., Ranum, L.P.W., Pulst, S.-M., and Rouleau, G.A. CAG repeat length in *hGT1* is associated with age at onset variability in spinocerebellar ataxia type 2 (SCA2).

S. Hayes performed the experiments (with assistance from K. Brisebois) and wrote the manuscript. G. Turecki advised on statistical analysis and helped edit the manuscript. I. Lopes-Cendes identified SCA2 positive samples. O. Riess, L.P.W. Ranum, and S.-M. Pulst contributed DNA samples. C. Gaspar and G.A. Rouleau contributed to experimental design, analysis, and helped edit the manuscript.

Abstract

The spinocerebellar ataxias (SCAs) are a clinically and genetically heterogeneous group of neurodegenerative disorders. To date, ten SCA loci have been described (SCA1-SCA8, SCA10 and SCA11), with six genes having been cloned (SCA1, SCA2, SCA3/MJD, SCA6, SCA7 and SCA8) and shown to contain CAG/CTG repeats.

This study investigated various aspects of the SCA2, SCA6, and SCA7 subtypes. Haplotype analysis in our panel of SCA2 families identified multiple ancestral mutation events to be responsible for disease in this group. Screening for the newly identified SCA6 and SCA7 mutations in our large collection of SCA families and patients revealed that these mutations are rare in our panel, each accounting for less than 1% of our ataxia samples. Finally, the CAG repeat-containing locus *hGT1* was found to be associated with residual age at onset variability in our SCA2 families.

Together, these results add to our growing understanding of the SCAs, and bring us a few steps closer to effective diagnoses of, and treatments for, these devastating diseases.

Résumé

Les ataxies spinocérébelleuses représentent un groupe cliniquement et génétiquement hétérogène de maladies neurodégénératives. À ce jour, dix *loci* SCA ont été décrits (SCA1- SCA8, SCA10, et SCA11) et six gènes ont été clonés (SCA1, SCA2, SCA3/MJD, SCA6, SCA7 et SCA8). Tous ces gènes contiennent des répétitions CAG/CTG.

Ce travail a pour but l'étude de divers aspects des sous-groupes SCA2, SCA6, et SCA7. L'analyse des haplotypes des familles SCA2 a permis d'identifier plusieurs mutations ancestrales responsables de la maladie de ce groupe de patients. En testant nos familles SCA pour les gènes SCA6 et SCA7, récemment identifiés, nous avons constaté que ces mutations sont rares dans notre population. Elles représentent moins de 1% des cas d'ataxies de notre collection. Finalement, nous avons montré que la répétition CAG du locus *hGT1* est associée à la variabilité de l'âge d'apparition de la maladie dans les familles SCA2.

Ensemble, ces résultats nous aident à mieux comprendre les ataxies spinocérébelleuses, ce qui nous permet d'envisager, pour l'avenir, des diagnostics et des traitements plus efficaces.

Chapter 1

Genetic and molecular investigation of the spinocerebellar ataxias

Introduction to the human repeat expansion disorders

Science has long known that human DNA contains repeated sequences, such as microsatellites and variable tandem repeats, interspersed throughout the genome. Despite their usual occurrence in noncoding portions of DNA, these polymorphic repeats have proven to be valuable tools, serving as landmarks in our efforts to map and identify the complex array of human genes. Recently, however, sequence repeats have begun to receive scientific attention for an entirely different reason. A new class of repeat has emerged that, unlike their harmless cousins, in one way or another impact normal gene function, and are thus associated with human disease. This new family of "repeat disorders" is one of the most rapidly growing and intensely investigated areas of genetic research. Case in point, this group of mutations has expanded from nine to 15 members (including two novel repeat types) in the less than three years of this study, to presently include two fragile X loci (FRAXA and FRAXE), myotonic dystrophy (DM), Friedreich's ataxia (FRDA), spinal and bulbar muscular atrophy (SBMA), myoclonus epilepsy type 1 (EPM1), Huntington disease (HD), dentatorubral pallidoluysian atrophy (DRPLA), spinocerebellar ataxia type 1 (SCA1), SCA2, SCA3/Machado-Joseph disease (MJD), SCA6, SCA7, SCA8, and oculopharyngeal muscular dystrophy (OPMD) (1-20) (Fig. 1.1).

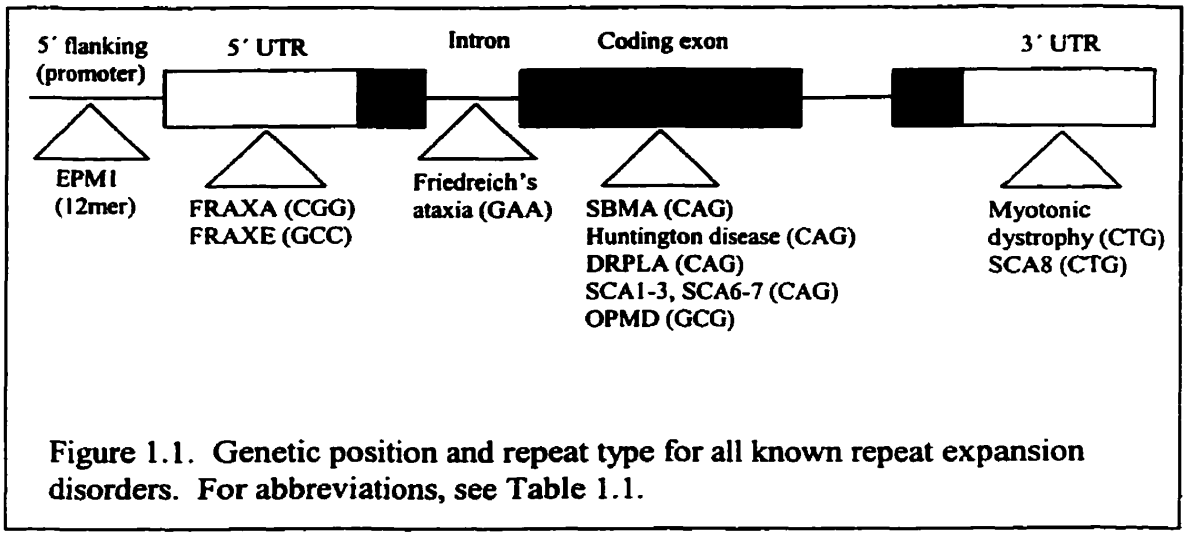

Figure 1.1. Genetic position and repeat type for all known repeat expansion disorders. For abbreviations, see Table 1.1.

Several key features are shared by the repeat disorders. In all cases, normally polymorphic repeat sequences expand beyond a disease-specific threshold, and become associated with disease. Once expanded into the pathological range, the repeats, also known as "dynamic mutations", become unstable, and tend to grow in size as they are passed from generation to generation (21). Interestingly, longer expanded repeats are generally associated with earlier ages of onset and/or more severe phenotypes. This relationship, when combined with the presence of intergenerational instability, has introduced a new wrinkle to human genetics: as repeats expand further, they tend to produce earlier ages of onset and disease severity. This causes disease symptoms to occur at younger ages in successive generations in pedigrees, a phenomenon known as genetic "anticipation". In fact, the discovery of repeat mutations provided a molecular explanation for observed genetic anticipation, which until that point was often said to have resulted from ascertainment bias because it could not be adequately explained by classical Mendelian genetics (22).

Despite the presence of hallmarks such as repeat instability and genetic anticipation, not all repeat mutations are created equal. Several unique qualities and important differences exist which allow the repeat diseases to be divided into subtypes, and these will be discussed in the following section.

Type 1 and type 2 repeat diseases

The identification of new mutations in recent years has made classifying and categorizing repeat disorders increasing more difficult. For instance, the discovery of an expanded dodecamer (12mer) repeat in EPM1 has broken the "rule of three" (23) as the first non-trinucleotide repeat clearly associated with disease, while the novel GCG repeat which causes OPMD (20) is the first repeat mutation to cause no apparent neuronal deficits. Despite these developments, it is still possible to divide the repeat disorders into two main groups based on similarities in repeat type, size, and position.

Until recently, all of the identified expansion mutations were trinucleotide repeats, which were generally divided into two categories, known as type 1 and type 2 trinucleotide repeat diseases (24,25). However, the recent discovery of the EPM1 dodecamer repeat has broadened the definition of repeat diseases, and necessitates dropping the "trinucleotide" from the category titles. As a result, it can now be argued that all known repeat disorders can be divided into two categories known simply as type 1 and type 2 repeat diseases. Table 1.1 lists some of the characteristics of the members of these categories.

Table 1.1. Characteristics of the human repeat expansion disorders.

Disease	Repeat	Location of repeat[a]	Repeat length normal range (premutation)	Repeat length disease range	Gene product
Type 2 repeat diseases					
myoclonus epilepsy type 1 (EPM1)	G/C rich 12mer	promoter	2-3 (12-17)	~30-75	cystatin B
Fragile X syndrome (FRAXA locus)	CGG	5′ UTR	6-52 (~60-200)	~200->2000	FMR1
FRAXE mental retardation	GCC	5′ UTR?	6-25 (~43-200)	>200	FMR2
Myotonic dystrophy (DM)	CTG	3′ UTR	5-37 (~50-180)	~200->2000	DMPK
Spinocerebellar ataxia type 8 (SCA8)	CTG	3′ UTR	16-92	~107-130	unknown
Friedreich's ataxia (FRDA)	GAA	intronic	7-22 (40-60)	200->900	frataxin
Type 1 repeat diseases					
Spinobulbar muscular atrophy (SBMA)	CAG	ORF	9-36	38-65	androgen receptor
Huntington disease (HD)	CAG	ORF	6-34	~36-121	huntingtin
Spinocerebellar ataxia type 1 (SCA1)	CAG	ORF	6-39	39-82	ataxin-1
Spinocerebellar ataxia type 2 (SCA2)	CAG	ORF	14-34	34-77	ataxin-2
Spinocerebellar ataxia type 3 (SCA3)/Machado-Joseph disease (MJD)	CAG	ORF	12-41	55-86	ataxin-3
Spinocerebellar ataxia type 6 (SCA6)	CAG	ORF	4-20	21-33	α_{1A} calcium channel subunit
Spinocerebellar ataxia type 7 (SCA7)	CAG	ORF	4-35	37->300	ataxin-7
Dentatorubralpallido-luysian atrophy (DRPLA)	CAG	ORF	3-36	49-88	atrophin
Oculopharyngeal muscular dystrophy (OPMD)	GCG	ORF	6-7[b]	8-13	PABP2

[a]UTR, untranslated region; ORF, open reading frame; [b]7 repeats causes recessive OPMD when homozygous

Type 1 repeat diseases

The first group, type 1 repeat diseases, currently consists of nine disorders: SBMA, HD, SCA1, SCA2, SCA3/MJD, SCA6, SCA7, DRPLA, and OPMD. All of these diseases are caused by fairly modest expansions of trinucleotide repeats which occur in the coding region of the respective genes (Fig. 1.1). With the exception of OPMD, all of the type 1 diseases are caused by the expansion of a $(CAG)_n$ repeat tract, which codes for an expanded polyglutamine (polyGln) tract in the mature protein. The newest member of this group, OPMD, is caused by a modestly expanded $(GCG)_n$ tract, which produces an expanded polyalanine (polyAla) tract in the PABP2 protein (20). The expansions in OPMD are the smallest of all the type 1 disorders, and are well within the normal range of all other type 1 members (see Table 1.1). OPMD is also unique in that it is the only member of this group that does not show progressive neurodegeneration as a pathological feature, but instead shows progressive atrophy of certain skeletal muscles (20). Despite these differences, type 1 repeat diseases are unified by having progressive phenotypes caused by the existence of modest expansions (8-~100 repeat units) of trinucleotide repeats which occur in the coding regions of their respective genes. It is also generally accepted that all type 1 mutations (with the possible exception of the SCA6 mutation) exert their pathological effects via a toxic gain-of-function of the expanded proteins, not via disruption of normal gene function (24). Members of the type 1 repeat disorders, particularly the spinocerebellar ataxias, will be discussed in greater detail in later sections.

Type 2 repeat diseases

As can be seen in Figure 1.1, several repeat mutations occur outside of the coding region of their respective genes. These mutations make up the type 2 repeat disorders. Mutations of this type have been shown to occur in a promoter region, 5′ untranslated region (UTR), 3′ UTR, and in an intron, and are presently associated with EPM1, FRAXA and FRAXE fragile-X mental retardation, FRDA, DM, and SCA8. Not unlike the type 1 diseases, new additions to the type 2 group have challenged the accepted characteristics of this group somewhat, but in general, type 2 repeat diseases are caused by relatively large (from ~50 to several thousand units), very unstable repeat expansions which occur in non-coding regions of their respective genes, and appear in some way to disrupt normal gene function (25). Unlike type 1 mutations, which tend to lead to progressive, single system disorders, type 2 mutations are often developmental, multisystem disorders. While both type 1 and type 2 mutations are caused by the expansion of unstable repeats into an affected range, type 2 repeats are usually very much larger (see Table 1.1), and tend to show more explosive expansions, with increases of several hundred repeats common during transmission (25). Inheritance patterns in type 2 disorders include autosomal dominant (DM, SCA8), autosomal recessive (EPM1, FRDA), and X-linked (FRAXA, FRAXE) inheritance (2-5,7,19). The occurrence of type 2 mutations in non-coding regions of their respective genes, combined with the fact that in FRDA, EPM1, and FRAXA, point mutations and/or deletions have occasionally been shown to be involved in disease development (5,7,26), suggests that a loss of normal gene function is likely the

underlying cause of most type 2 repeat diseases. While loss-of-function *versus* gain-of-function has, in the past, been a key difference between the type 1 and type 2 diseases, questions raised as a result of the cloning of the SCA6 and SCA8 mutations has begun to blur the line between these two groups somewhat. These two loci, and their proposed pathological etiologies, will be discussed more fully in later sections, as we now direct our attention to the polyGln disorders, the family of diseases which houses the focus of this study: the spinocerebellar ataxias.

The polyglutamine disorders

In 1991, La Spada *et al.* (6) discovered that a normally polymorphic CAG repeat located within the first exon of the androgen receptor (AR) gene was expanded in patients with spinal and bulbar muscular atrophy (SBMA). Since then, seven other diseases have also been shown to be caused by expanded CAG repeats, including HD, DRPLA, SCA1, SCA2, SCA3/MJD, SCA6, and SCA7. These diseases have become collective known as the "polyglutamine disorders" because, in each, the expanded $(CAG)_n$ repeat is translated into an expanded polyGln tract in the mature protein. As with all repeat disorders, a relationship exists between the size of the repeat and the age at disease onset in the polyGln disorders. In general, the repeat size accounts for ~50% of the variability seen in age at disease onset, although the exact magnitude of this effect varies according to the disease and study. Apart from the CAG tracts themselves, the genes show no similarities and, with the exception of the loci involved in SBMA and SCA6

(whose expansion occurs within a calcium channel subunit), code for proteins of unknown function (27). Despite widespread expression of the genes, in each disease a specific subset of neurons degenerate, ranging from motor neurons in SBMA (6), striatal and caudate neurons in HD (28), dentate and globus pallidus neurons in DRPLA (22), and cerebellar, brain stem, and spinal cord neurons in the SCAs (27). With the exception of SBMA, which is X-linked, all known polyGln disorders are inherited as autosomal dominant diseases.

Exactly how and why polyGln tracts cause neurodegeneration is still unclear, but the intense scrutiny these disorders have received in recent years has begun to shed light on many aspects of the polyGln disorders, including clues about their origins, relative prevalence, and mechanisms of pathology. It is generally believed that the presence of a polyglutamine expansion results in disease via a toxic gain-of-function mechanism (with the possible exception of the SCA6 mutation). This notion is supported by several facts: first, dominant inheritance is generally consistent with a gain-of-function (although haploinsufficiency or dominant negative mechanisms are difficult to rule out given the inconsistent reports regarding the phenotypes of individuals homozygous for CAG expansion mutations [14,24,29]). Second, expanded alleles are still expressed and translated into protein, suggesting that altered protein function, not reduced levels, is responsible for disease (24). Third, point mutations in the androgen receptor gene, which causes loss of normal gene function, causes testicular feminization - a disease completely unrelated to SBMA (the disease produced by CAG expansions in the same gene). Finally, several

mouse models have been developed where insertion of expanded transgenes cause neurodegeneration and/or appropriate disease phenotype, suggesting that these transgenes have toxic properties (30,31). Taken together, these factors suggest that polyGln expansions confer some sort of toxic quality to proteins. This idea will be more fully explored in later sections which deal with current research on polyGln pathogenesis. First, however, we will turn our attention to the most fundamental process in CAG repeat pathogenesis: repeat instability.

CAG repeat instability

Like most repeat sequences in the genome, CAG repeats are inherently unstable. Rare changes in repeat size during countless historical meioses has produced a polymorphic normal repeat range in the general population for most of the CAG repeat sequences studied (see Table 1.1). Despite this polymorphic nature, normal sized repeats are almost always stably transmitted at the level of individual families (24). However, once repeats expand past a pathological threshold, which is approximately 37-40 repeats for most of the polyGln disorders, they become much more unstable, undergoing further expansions (as well as occasional contractions) during transmission. As was already discussed, this tends to lead to longer repeat sizes, and corresponding earlier disease onset, in successive generations.

In addition to this meiotic instability, expanded alleles also appear to be unstable during mitotic cell division, causing any given population of cells to exhibit a range of expanded allele sizes, a phenomenon known as "somatic

mosaicism". As a result, polymerase chain reaction (PCR) amplification of expanded repeats for almost all of the polyGln disorders produces a characteristic, multi-banded pattern containing a prominent, central band (the inherited size), with progressively fainter bands above and below. There is also some evidence that within individuals, different allele ranges may be seen in different cell populations. For instance, cerebellar cortex tissue in SCA1 and SCA3/MJD shows slightly smaller expanded repeat sizes than do other brain regions (32,33), a phenomenon also reported for DRPLA (34).

This propensity for CAG repeats to further mutate during meiosis and mitosis is likely also responsible for a parent-of-origin effect often seen in the polyGln diseases. Typically, paternal transmissions of expanded alleles are associated with greater transmission instability (10,35-37). It has been proposed that this phenomenon occurs because more cell divisions, and therefore more opportunities for further mutational events, are required in the production of sperm as compared to oocytes (38). This instability has been confirmed in several studies where single sperm PCR, performed on sperm from affected males, produces a wide range of expanded alleles (36,39-41), although this effect was not seen in affected SCA6 males (42).

One element that appears to influence repeat stability, and seems to be crucial for CAG repeats to make the jump from the normal range to the expanded range, is the loss of interruptions. In many cases, normal alleles are not pure CAG tracts, but are instead punctuated by other trinucleotides. For example, 98% of SCA1 normal alleles are interrupted by 1-3 CAT trinucleotides, while all SCA1

expanded alleles are composed of pure CAG tracts (35). SCA2 alleles only expand into the pathological range after losing the 1 or 2 CAA repeats present in almost all normal alleles (13-15). In this way, the presence of interruptions appears to afford a degree of stability to CAG tracts, with the loss of these interruptions being an important initiating step in the transition from the normal to expanded range.

Exactly how and why CAG repeats become unstable is not fully understood, but it is commonly believed that hairpin formation plays an important role in the process. Figure 1.2 outlines the ways in which hairpin formation during DNA replication might lead to expansion and/or contraction of a CAG repeat tract. Hairpins form when CAG and/or CTG tracts loop back on themselves such that cytosine residues can form hydrogen bonds with guanine residues, adopting the following configuration:

```
.....C A G C A G C A G C A G C
     |   | |   | |   | |   | A
.....G A C G A C G A C G A C G
```

As can be seen in Figure 1.2, if a hairpin forms in the newly synthesized daughter strand during replication, the result will be an expansion of the repeat, while a hairpin in the parent template strand will lead to a contraction (43).

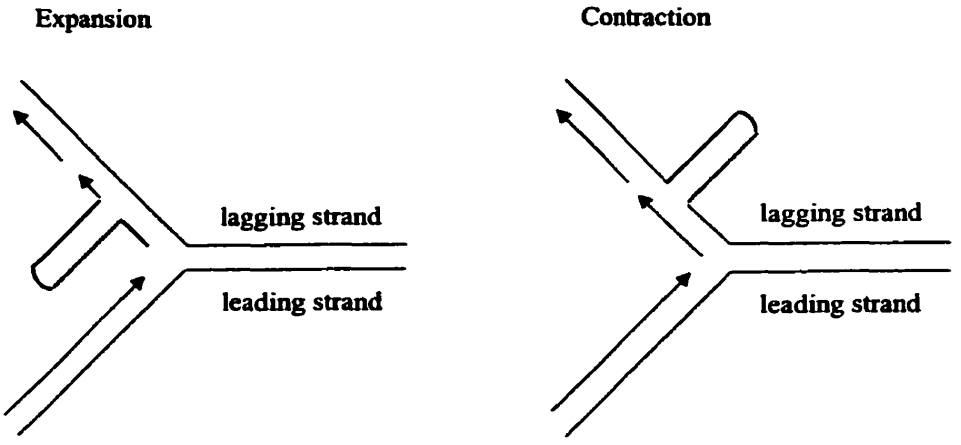

Figure 1.2. Role of hairpins in CAG repeat tract expansion/contraction during DNA synthesis. Hairpin formation in the newly formed strand will lead to expansions; hairpin formation in the template strand will lead to contractions.

Hairpin stability, and thus the likelihood that they will form, can be influenced by the length and composition of repeat sequences. As a result, longer repeats are more unstable than short repeats because longer repeats more readily form stable hairpins (43,44). Similarly, the loss of sequence interruptions (such as a CAA repeat) to produce a pure repeat may increase repeat instability by increasing the stability of hairpins (44). Interruptions may also increase repeat stability by serving as reference points, or "anchors," within the repeat tract, decreasing slippage by DNA replication machinery (a mechanism also proposed to contribute to repeat mutation)(45). Regardless of the mechanism, it is obvious that interruptions are crucial for repeat stability, and that the loss of these interruptions, along with increasing repeat length, are central to the mutability of CAG repeat tracts.

One final factor has recently become implicated as possibly contributing to CAG repeat tract instability. Three separate studies have reported that a CGG/GGG polymorphism immediately downstream of the CAG repeat in the *MJD1* gene effects intergenerational instability (40,46,47). Surprisingly, the effect appears to act both in *cis* and in *trans*, as it is specifically the [expanded CAG-CGG/normal CAG-GGG] genotype which is associated with the greatest instability. This suggests that some sort of inter-allelic interaction may be involved in intergenerational instability of the *MJD1* repeat tract. Exactly how or why this effect occurs, or if a similar phenomenon occurs in other polyGln disorders, remains to be determined. Nevertheless, these results indicate that intragenic regions other than the CAG repeat itself may play a role in conferring instability to CAG repeat tracts.

To summarize, instability of CAG repeat tracts, an integral component of the pathology of the polyGln diseases, appears to be influenced by several factors, including repeat length, presence or absence of interruptions, gender of transmitting parent, and perhaps the presence of intragenic polymorphisms. As has been discussed, these factors not only contribute to instability within the expanded range, but may also play a role in the expansion of normal alleles into the expanded range, with obvious implications for disease origins and prevalence.

The above examination of some of the general characteristics of the CAG repeat disorders provides a framework for the introduction of the spinocerebellar ataxias, which, in terms of numbers of identified genes, constitute the majority of the polyglutamine disorders. Besides being the focus of this investigation, many

of the identified characteristics of the SCA subtypes have provided important clues regarding aspects of polyglutamine-induced pathology in general, making it important to discuss these qualities before finally addressing the most recent theories regarding polyglutamine pathogenesis.

The spinocerebellar ataxias

Until very recently, it would have been quite logical to include the spinocerebellar ataxias (SCAs) as a group within the polyGln disorders, but the recent cloning of the SCA8 gene has challenged the notion that all SCAs are caused by polyglutamine-coding CAG repeats, a fact which will be discussed later. The SCAs are, however, unified to some extent by overlapping clinical features, which usually include gait and limb ataxia, dysarthria, and other associated signs, as well as by common pathology of the brain stem, cerebellum, and spinal cord (27).

Prior to the identification of any SCA mutations, attempts were made to group and classify these disorders based solely on clinical observations. The most widely cited example of this was Harding's classification system (48) which recognized three cerebellar ataxia groups: autosomal dominant cerebellar ataxia type I (ADCA I), which is characterized by cerebellar ataxia in conjunction with various associated features; autosomal dominant cerebellar ataxia type II (ADCA II), which consists of cerebellar ataxia with retinal degeneration; and autosomal dominant cerebellar ataxia type III (ADCA III), which consists of pure cerebellar ataxia without associated features. While this system is still used, it is really

genetic classification, which has resulted from gene cloning, which is now most useful. To date, six SCA genes have been cloned (SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and SCA8), while four other loci have been mapped in SCA families (SCA4, SCA5, SCA10, and SCA11) (49-52).

The first SCA gene discovered, SCA1, was cloned in 1993. The identification of an expanded CAG repeat as the causal mutation in SCA1 gave direction to the search for similar mutations in other forms of SCA, and opened the door for the cloning of SCA3/MJD in 1994, SCA2 in 1996, SCA6 in 1997, and SCA7, later in 1997. Each gene was found to contain an expanded, polyglutamine-coding CAG tract as the causal mutation. Four other loci have been linked in families suffering from SCA; SCA4 on chromosome 16q22.1, SCA5 on the centromeric region of chromosome 11, and recently SCA10 on chromosome 22q13 and SCA11 on chromosome 15q14-21.3. Efforts to identify mutations in these candidate regions have thus far failed. As has already been mentioned, very recently a sixth SCA gene, spinocerebellar ataxia type 8 (SCA8), was cloned (19), and shown to contain an untranslated CTG expansion in a subset of SCA patients. This finding obviously challenges the notion that all SCAs are caused by polyglutamine-encoding CAG expansions, and suggests that SCA8 may be related to the type 2 repeat disease myotonic dystrophy, which is also caused by an untranslated CTG repeat in the 3′ UTR. SCA8 also appears unique in that it is found to segregate in autosomal dominant, autosomal recessive, and isolated SCA families/individuals. All other known SCA genes generally affect autosomal dominant families, although occasional sporadic cases have been reported (53-60).

Three of the more recently cloned forms of SCA were specifically examined in this study: SCA2, SCA6, and SCA7. In order to help establish the rationale for the approaches of this study, a brief description of each of these loci, including some of the unique characteristics of each, follows.

Spinocerebellar ataxia type 2 (SCA2)

The gene for SCA2 was cloned independently in 1996 by three different groups (13-15) having already been mapped and shown to be associated with polyglutamine expansions using an antibody specific to expanded polyglutamine motifs (61). Clinically, SCA2 typically features ataxia, dysarthria, and characteristic slow saccades (27). At the time of the cloning, SCA2 expanded alleles were the smallest seen in the CAG repeat disorders, with the smallest pathological allele (35 repeats) well below the pathological threshold of the other CAG repeat diseases identified at that time. Unlike most of the other CAG repeat disorders, which are highly polymorphic in the normal population, the vast majority (>90%) of SCA2 normal alleles consist of 22 repeats, with the remaining rare alleles ranging between 14 and 34 repeats (13,14,39,58). No gap exists between the largest normal allele and smallest expanded allele in SCA2, with 34 repeat being reported as both a normal allele and an "at risk" allele by different investigators (39,58).

While the relative prevalence of each SCA subtype varies depending on the ethnic and geographic group being studied, most surveys show that SCA2 is the second most common cause of spinocerebellar ataxia (behind SCA3/MJD),

generally accounting for between 10-20% of all ataxia patients (39,55,62-64). In Britain, however, SCA2 is reported to be the most common cause of SCA at 40% (65), while in Japan it is generally a rare cause of SCA (~5%) behind SCA3/MJD, DRPLA, and SCA6 in terms of prevalence (53,56,66).

Spinocerebellar ataxia type 6 (SCA6)

In contrast to SCA2, which was already defined prior to its cloning by virtue of linkage to chromosome 12, SCA6 became a distinct genetic entity only after the SCA6 repeat was discovered. Using a $(GCT)_7$ probe to screen a human brain cDNA library, Zhuchenko *et al.* (17) identified SCA6 as a short CAG repeat which ranged from 4-16 CAG repeats in normal controls, but was expanded to 21-27 repeats in a subset of tested SCA patients. This repeat is located in the 3′ end of the well characterized α_{1A}-voltage-dependent calcium channel gene, now known as CACNA1A (67), and was predicted to code for polyglutamine in three of the six isoforms discovered. Initial screening results suggested that SCA6 was typified by late onset, relatively pure cerebellar ataxia (ADCA III) (17), but more recent studies have found some SCA6 patients with early onset and associated signs (59,66,68).

Many characteristics of the SCA6 subtype make it a unique member of the spinocerebellar ataxia family. First, the size of the expanded alleles are within the normal range of all other CAG repeat diseases. At present, normal alleles have been found to range from 4-20 repeats, with expanded alleles ranging from 21-33 repeats (29,69,70). As with all the CAG repeat diseases, a negative correlation

exists between the size of the expanded repeat and the age of disease onset, but in SCA6 this association is particularly pronounced, with approximately 65-75% of the variability in age at onset explained by repeat size in some panels (29). Thus, the addition of even a single glutamine residue can have a profound effect on disease onset.

Several factors, including the small repeat size and strong negative correlation between disease onset and repeat size, have prompted some researchers to suggest that SCA6 pathogenesis results from a loss of normal gene function, as opposed to the gain-of-function mechanism proposed for most other polyGln disorders (17,71). This characteristic, so far unique to SCA6, is intriguing as it is difficult to imagine how two CAG tracts, differing in size by as few as two repeat units (in the case of the largest SCA6 repeat and the smallest SCA2 repeat), could produce nearly identical pathological and clinical pictures via completely different pathological mechanisms. Nevertheless, there does appear to be evidence that disruption of normal gene function may be at the root of SCA6 pathogenesis. For instance, recessive (loss of function) mutations in the mouse homologue of the α_{1A}-voltage-dependent calcium channel gene are associated with the mouse tottering (*tg*) and leaner (tg^{la}) phenotypes, which are characterized by cerebellar ataxia and seizures (17,71). Also intriguing are recent findings linking SCA6 with two other diseases also caused by mutations in the CACNA1A locus. Missense mutations in this gene have been found to cause familial hemiplegic migraine (FHM), while mutations causing protein truncation cause episodic ataxia type-2 (EA-2) (72). All three of these "allelic disorders" can be

associated with a degree of ataxia and cerebellar atrophy, in addition to the features specific to each disease (71). While FHM and EA-2 are also inherited in an autosomal dominant fashion, it is generally accepted that the mutations that cause these disorders disrupt normal channel function, causing disease via haploinsufficiency or dominant negative mechanisms (71). With this in mind, it is very interesting to note that families with episodic ataxia-like features have been found to harbour SCA6 expansions (67,68), while individuals with progressive ataxia have been found with point mutations in the CACNA1A locus (73). In fact, Jodice *et al.* (67) suggest that EA-2 and SCA6 should be considered to be the same disorder. These findings do indeed seem to suggest that the SCA6 expansion causes a disruption of normal channel function, perhaps also through haploinsufficiency or dominant negative mechanisms, as opposed to conferring a gain of toxic protein function, as is accepted for the other polyGln disorders. Having said this, very recent evidence of inappropriate cellular localization of the α_{1A}-voltage-dependent calcium channel protein (74), which will be discussed later, may in fact suggest that SCA6 has more in common pathologically with the other polyGln disorders than originally suggested.

A third unique characteristic that appears to separate SCA6 from the other polyGln diseases is the relative stability of the expanded repeat tracts. Expanded repeats in HD, DRPLA, SBMA, and the other SCAs are rarely transmitted faithfully, but rather undergo meiotic expansions or contractions during transmission. SCA6, on the other hand, is almost always stably transmitted, even within the expanded range (17,56,70). Perhaps this should not be surprising

considering the small expansion sizes seen in SCA6, which are similar to the stable, normal alleles of other SCA subtypes. At least two studies have reported, however, SCA6 expanded alleles changing size during transmission (53,75), pointing out that a measure of transmission instability can be seen in SCA6.

Once again, as in SCA2, relative prevalence of SCA6 depends greatly on the geographic and/or ethnic group being examined, even within a country. For instance, in some areas of Japan, SCA6 accounts for in excess of 30% of ataxia patients (56,75), while in other regions it is found in only 6% of ataxia patients (66). In general, however, SCA6 appears to be responsible for approximately 10-20% of all SCA cases (54,55,59,68,76), usually ranking it third behind SCA3/MJD and SCA2 in terms of relative prevalence, although in some areas like France, parts of North America, and Portugal, it appears to be a very rare cause of ataxia (less than 5%) (62,64,77). Several authors have found SCA6 in sporadic or isolated patients (53-57), although Schols *et al.* (59) have noted that often this may be the result of misdiagnosis due to the missing of very late manifesting ataxia symptoms in the parents, as opposed to genuine *de novo* SCA6 mutations, like that documented by Shizuka *et al.* (60).

One final characteristic which appears to be more common in SCA6 than in the other polyGln disorders is the occurrence of patients homozygous for the SCA6 expansion. To date, at least six SCA6 homozygotes (also known as compound heterozygotes) have been reported (53,56,68,69,75). High local prevalence and founder effects, particularly in regions of Japan, may be playing a role in the generation of these homozygotes. Interestingly, authors appear divided

as to the effects of a homozygous mutation, with some studies reporting earlier onset and more severe disease than heterozygotes (53,56,68), while others report no effect (69,75). This uncertainty mirrors the ongoing argument of the effects of homozygous mutations in other polyGln diseases such as DRPLA, MJD, SCA2, and HD (14,29,78-80). While clear evidence of increased disease severity in homozygotes would provide clues about the pathological mechanisms in diseases such as SCA6, the lack of a consensus opinion in this area unfortunately makes these data ambiguous at best.

Spinocerebellar ataxia type 7 (SCA7)

The long search for the gene mutated in SCA7 ended in 1997 when David *et al.* (18) cloned a highly unstable CAG repeat on chromosome 3p12-p13. Aided once again by the 1C2 antibody which recognizes expanded polyglutamine motifs, the cloning of the SCA7 gene was guided by the knowledge that an expanded polyglutamine repeat was responsible for this disease (61).

SCA7 is typified by cerebellar ataxia and retinal degeneration, and corresponds to ADCA II in Harding's classification system (48). Although SCA7 screening reports have generally supported the notion that ADCA II is genetically homogeneous (36,55,76,81-83), the very recent identification of a SCA7 negative ADCA II family suggests this phenotype can be associated with a degree of genetic heterogeneity (84). It should also be mentioned that SCA7 expansions have been found in patients without retinal degeneration, at least for the initial

phase of disease development (36,81), indicating that the SCA7 phenotype may not be as well defined as originally suggested.

Normal alleles at the SCA7 locus range from 4 to 35 repeats, with 10 CAG repeats accounting for 65-75% of all normal alleles (36,76,81,83). Expanded alleles range from 37 to >300 repeats, and are extremely unstable upon transmission, particularly when transmitted paternally (83,84), although maternal transmissions are most common (36,83). The largest SCA7 repeats (>200 CAG units), associated with infantile disease onset and early death, are the largest repeats found thus far in any of the polyGln diseases (76,81). This propensity for large expansions is likely a reflection of the fact that the SCA7 repeat is the most unstable of all known disease-causing CAG repeats (36).

Assessing the prevalence of SCA7 by examining the literature is a difficult task, as some authors discuss ADCA II completely separately from other forms of ADCA, and therefore never compare SCA7 to other SCAs (36,81-85). Studies which do integrate ADCA groups, however, generally report that SCA7 is a rare to moderately rare cause of cerebellar ataxia, accounting for approximately 5-10% of all SCA cases (55,76).

Having introduced the key concepts and players in the CAG repeat diseases, we will next briefly discuss the most recent theories about the cellular pathology of the polyglutamine disorders, before finally establishing the specific goals and aims of this study.

Protein accumulation in the polyglutamine disorders

While there is now very little doubt that expanded polyglutamine motifs are in some way toxic to neurons, very little is known about the actual mechanisms that may be responsible for cellular pathogenesis, including what role expanded polyglutamine tracts play in cell death, and why specific populations of neurons are susceptible in different diseases. While answers to these questions are still possibly years away, several exciting new studies have begun to shed some light in this area by investigating the possibility that improper protein localization and aggregation may be key contributors to polyglutamine-induced neurodegeneration.

While evidence for protein accumulation in affected neurons in the polyGln disorders has only recently been generated, theories proposing such accumulation and/or aggregation have been established for several years. Green (86) speculated in 1993 that expanded polyglutamine motifs may act as substrates for covalent cross-linking by transglutaminases, while Perutz *et al.* (87) suggested that these repeats may act as "polar zippers," gradually causing disease protein to aggregate and accumulate through non-covalent hydrogen bonding. In both cases, it was assumed that aggregates would somehow interfere with normal cell function, and ultimately contribute to cell death. These models also introduced the possibility that expanded polyglutamine-containing proteins might interact with cell-specific factors, rendering specific populations of neurons at risk for each disease.

Direct evidence for protein aggregation in the polyGln diseases first appeared in 1997 when Davies *et al.* (88) noticed small, neuronal intranuclear inclusions (NIs) in brain specimens of HD transgenic mice. These inclusions, which labeled with antibodies to huntingtin and ubiquitin (Ub), were present in brain regions typically affected in HD, and only occurred in animals carrying an expanded transgene. These results appeared to suggest that the presence of a polyglutamine expansion was causing inappropriate protein accumulation and/or nuclear translocation, perhaps with degenerative consequences.

Since this initial report, NIs have been clearly seen in transgenic mouse and *Drosophila* models, cellular models, and patient brain material for SCA1, SCA3/MJD, DRPLA, HD, SBMA and SCA7 (89-99), as well as with cellular and mouse models using CAG repeats alone, or inserted into non-disease genes (31,100). While the complete make-up of these aggregate bodies is not yet known, other cellular proteins (in addition to Ub), such as components of the proteasome, as well as chaperones, have also been found to be associated with aggregates in some cases (101,102).

Recently, protein aggregation has been reported in the polyGln diseases with the shortest pathological alleles, SCA2 and SCA6. Two recent studies have demonstrated the presence of the SCA2 and SCA6 disease proteins in cytoplasmic accumulations and aggregates (74,103). These results are somewhat surprising given the generally accepted notion that SCA6 pathogenesis occurs via different mechanisms than in the other polyGln disorders. Having said this, it is still possible that protein aggregation in SCA6 may lead to loss of proper calcium

channel function, and not necessarily to a gain of toxic qualities. Further data is needed to adequately address this issue.

The presence of NIs in disease models, as well as patient material, originally lead researchers to suggest that these structures must be responsible for cell death in some way. New evidence has surfaced, however, which suggests that NIs are not necessarily required for cell death - indeed that these structures may actually be beneficial to cell survival (98,104-106). Basically, these data indicate that cell death can occur in the absence of discernible inclusions (104), while other cells can survive despite the presence of NIs (98,106). Additionally, blocking the formation of inclusion bodies actually increased the rate of cell death in one study (105). It is therefore possible, based on these findings, to imagine a model whereby expanded polyglutamine-containing protein is toxic to cells in its unbound form, and that the formation of NIs is a protective strategy utilized by the cell in an attempt to neutralize and/or degrade this protein (thus explaining the presence of degradation machinery such as Ub and chaperones) (105). It may be possible that some populations of cells have a lesser capacity to handle toxic protein in this way, eventually rendering them vulnerable to toxic protein build-up. Having said this, it is important to note that this model is based on preliminary reports only, and does not preclude the possibility that protein accumulation and aggregation may play an important role in the neurodegenerative process. Suffice it to say that further exploration is needed before we can adequately address the role protein aggregation plays in polyGln-induced pathogenesis.

Objectives of this study

Several aspects of the genetics of the spinocerebellar ataxias were independently investigated in this study. While these studies are able to stand alone as separate bodies of research, together they contribute to our increasing understanding of the SCAs. Using techniques such as haplotype analysis, mutation screening, and candidate gene analysis, we investigated aspects of mutation origin, prevalence, and genetic modification of disease onset in certain SCA subtypes.

This study consists of three independent projects. The specific objectives of these projects are:

Investigating the origins and spread of the spinocerebellar ataxia type 2 (SCA2) mutation.

1. to use haplotype analysis to examine whether the SCA2 mutation arose independently in our SCA2 positive families, or whether these families are linked by common ancestral mutations.

Investigating the role of the SCA6 and SCA7 repeat mutations in a large panel of ataxia patients.

2. to determine the frequency of the SCA6 and SCA7 mutations in our large panel of SCA patients, and to compare these frequencies to those found for the previously tested forms of SCA.

Examination of CAG repeat-containing loci as possible genetic modifiers of age at onset in spinocerebellar ataxia type 2 (SCA2).

3. to determine whether the size of CAG repeat tracts found in various CAG repeat-containing loci can account for any of the variability in SCA2 age at disease onset not explained by the size of the expanded SCA2 repeat tract itself.

Each of these studies attempts to provide another small piece to the bigger puzzle that is the spinocerebellar ataxias, as well as to contribute to our knowledge of the polyglutamine disorders in general.

Chapter 2

Investigating the origins and spread of the spinocerebellar ataxia type 2 (SCA2) mutation

This first study addressed the possibility that all, or a subset, of cases of SCA2 identified in our lab may have resulted from a common ancestral mutation, reflecting a founder effect. Similar studies in other SCA subtypes and polyglutamine disorders have often found multiple mutational events responsible for disease prevalence, even within restricted ethnic or geographic groups. Work of this type can reveal a great deal about the origins and spread of a particular mutation, as well as give valuable information regarding the relative commonness or rarity of mutational events.

This work will be submitted for publication as:

Hayes, S., Lopes-Cendes, I., Gaspar, C., Jain, S., and Rouleau, G.A. Haplotype analysis in a group of spinocerebellar ataxia type 2 (SCA2) patients.

Haplotype analysis in a group of spinocerebellar ataxia type 2 (SCA2) patients

Introduction

The autosomal dominant spinocerebellar ataxias (SCAs) are a genetically diverse group of neurodegenerative disorders typically affecting the cerebellum, brain stem, and spinal cord. The growing list of loci identified as causing SCA presently contains six genes (SCA1, SCA2, SCA3/Machado-Joseph disease (MJD), SCA6, SCA7, and SCA8) (12-19) as well as four linked loci (SCA4, SCA5, SCA10, and SCA11) (49-52). Five genes identified so far have an expanded $(CAG)_n$ repeat as their pathological mutation. These repeats are located in the coding region of their respective genes and code for polyglutamine tracts, making these forms of SCA members of a larger group of "polyglutamine (polyGln) disorders" which also includes spinal and bulbar muscular atrophy (SBMA) (6), Huntington disease (HD) (9), and dentatorubral-pallidoluysian atrophy (DRPLA) (10,11). Current evidence suggests that SCA8 is not a polyGln disorder, as its pathological expansion is an untranslated CTG, making this form of SCA similar to the expansion disease myotonic dystrophy (DM) (19). It is therefore possible that the pathological mechanism at work in SCA8 is different than that of the other polyGln disorders, and similar to that of DM.

The mechanism by which polyGln expansions cause the characteristic patterns of neurodegeneration seen in each disorder is unknown, as is how and why these expansions arise. It has been proposed that certain factors, including repeat size (107), as well as *cis* and/or *trans* acting elements (40,46,47), predispose certain normal alleles to expand to become pathological alleles. It has

also been suggested that present day distributions of these diseases may, in some cases, be due to founder effects, with ancient mutations responsible for most or all current cases (108). While it is true that some studies have reported apparent founder effects in certain geographic regions (109), the vast majority of evidence seems to indicate that the unstable nature of CAG repeats has lead to numerous mutational events for most of the polyGln diseases, even within apparently homogeneous populations. For example, haplotype analysis in South African SCA1 families revealed the existence of at least three ancient mutations (110), while at least two founder mutations appear to be responsible for modern cases of MJD in the Azore Islands (C. Gaspar, in preparation) (111). Multiple founders have likewise been reported for SCA2, SCA6, SCA7, DRPLA, and HD (64,65,82,112-114). Several authors have also reported *de novo* mutations occurring in polyGln families (60,82,84,115), further supporting the notion that expansions from normal to pathological ranges are not limited to rare, ancient events.

In order to further investigate the possible origins of CAG expansions in the polyGln diseases, we used haplotype analysis to identify potential founder chromosomes in SCA2. Unlike most of the other polyGln disorders, SCA2 normal alleles are relatively non-polymorphic, with 22 repeats accounting for in excess of 80% of all normal alleles (14,65,112). Given this apparent stability of normal length SCA2 alleles, one might speculate that expansion events will occur less frequently in this form of SCA. However, even within geographic and ethnic groups, we were able to identify multiple founder chromosomes, suggesting that

mutational events have occurred frequently in SCA2. Our data also allowed us to make some assumptions about the spread of the SCA2 mutation within our families.

Materials and Methods

Subjects

Twenty-four seemingly unrelated SCA2 families, including nine isolated patients, were included in this experiment. The families examined were from three main geographic/ethnic groups. Samples from seven of the families were collected in and around Delhi, India, and practiced the Hindu faith. Three families had ties to Canada, with one family from the Gaspé region of Québec, one from Newfoundland, and one with branches in Saskatchewan and Austria. Samples from five families, and one isolated individual, were collected in either Portugal or Brazil. An isolated Greek-Canadian individual, as well as seven isolated pathological brain specimens of unknown ethnic origin, were also included.

Genomic DNA was isolated from peripheral lymphocytes or autopsy materials using standard procedures (116). All families and isolated cases were identified as having the SCA2 mutation using previously reported methods (117). At least one affected individual from each family was genotyped, as well as parents and/or sibs (when available) in order to determine chromosome phase. Control data was gathered from unaffected spouses of affected individuals, as well as from patients' non-segregating normal chromosomes.

PCR analysis of microsatellite markers

Three chromosome 12 microsatellite CA repeat markers were used to establish haplotypes in this study. Markers D12S1333 and D12S1332 flank the SCA2 gene, with D12S1333 located approximately 200 kb telomeric to the gene, and D12S1332 approximately 350 kb centromeric (13). The third marker, D12S1672, is located in the first intron of the SCA2 gene (118). Marker repeat sizes were determined by polymerase chain reaction (PCR) in a total volume of 13.0 µl, containing approximately 100 ng of genomic DNA, 10 mM Tris-HCl (pH 8.8), 1.5 mM $MgCl_2$, 50 mM KCl, 10% dimethylsulfoxide (DMSO), 250 mM each of dCTP, dGTP and dTTP, 25 mM dATP, 1.5 µCi α^{35}S-dATP, 100 ng of each primer, and 1.5 U *Taq* polymerase (Perkin-Elmer). The following primer pairs and annealing temperatures were used: D12S1332a (5'-GCC AGG TAC AGT GGC TC-3') and D12S1332b (5'-CTG GGA CCA CAG GTG TAG-3') at 60°C; D12S1333a (5'-TTC AGG TGG TAC AGC CGT-3') and D12S1333b (5'-CAT CAG AAG GCT TCA TAG GAA T-3') at 50°C; D12S1672a (5'-CAG AGG GAG ATT CCA TCC AA-3') and D12S1672b (5'-CGG TTT GAC AAG TTT CGA GA-3') at 49°C. Products were resolved on 6% polyacrylamide gels, and allele sizes were determined by comparison to an M13 DNA sequencing ladder.

Results

Table 2.1 contains the haplotypes and/or genotypes of all of the SCA2 families and isolated patients examined in this study. Haplotypes are included

where phase could be determined, or where phase was unequivocal due to homozygosity at a given locus. Genotypes are included where phase could not be determined.

Table 2.1. SCA2 disease chromosome haplotypes[a]

SCA2 family/individual	D12S1333[b]	D12S1672[c]	D12S1332[d]
Families			
Indian 1	8	3	4
Indian 6	8	3	4
Indian 5	8	2	7
Indian 7	8	2	7
Indian 2	8	2	8
Indian 3	8	1	7
Indian 4	9	3	10
Gaspé	7	3	6
Austrian-Canadian	4	5	7
Newfoundland	4	3	7
Brazilian 2	6	8	5
Brazilian 4	6	5/8	10
Brazilian 3	4	5	7
Portuguese 1	4	5	7
Portuguese 2	7	2	7
Isolated cases			
Brazilian 1	6/8	3/5	4/7
Greek	4/7	3/4	2/7
Brain 94-546	4/6	5	1/7
Brain 92-403	4/6	5	2/7
Brain 91-289	4	3/5	3/5
Brain 94-542	4/6	5	7
Brain 94-517	4/8	3	5
Brain 86-84	4/8	4/5	7
Brain 95-669	6/8	3/5	5/7

[a]double entries denote phase unknown genotypes
[b]allele 4=240 bp, 6=236 bp, 7=228 bp, 8=226 bp, 9=230 bp; [c]allele 1=301 bp, 2=299 bp, 3=297 bp, 4=295 bp, 5=293 bp, 8=287 bp; [d]allele 1=217 bp, 2=215 bp, 3=213 bp, 4=211 bp, 5=209 bp, 6=207 bp, 7=205 bp, 8=203 bp, 10=197 bp

Complete haplotypes, which included all three loci, were generated for 14 families, including all seven Indian families, all three families with Canadian members, and four of the five Portuguese/Brazilian families. Ten different haplotypes were found in these 14 families, with no single haplotype present in more than three families. The most common haplotype, 4-5-7 (D12S1333-D12S1672-D12S1332), segregated with the SCA2 disease chromosome in the Austrian-Canadian family, one of the Brazilian families, and one of the Portuguese families. Two other haplotypes, 8-3-4 and 8-2-7, were both found in two of the Indian families. All other haplotypes were present only once. Partial haplotypes (consisting of one or two phase-determined markers) were seen in one Brazilian family, and in five of the isolated brain samples. Normal control allele frequencies for each marker are shown in Fig. 2.1. In general, increasing allele number for each marker corresponds to decreasing CA-repeat size, with the exception of allele "9" at D12S1333, which is intermediate in size between alleles "6" and "7" (see Table 2.1).

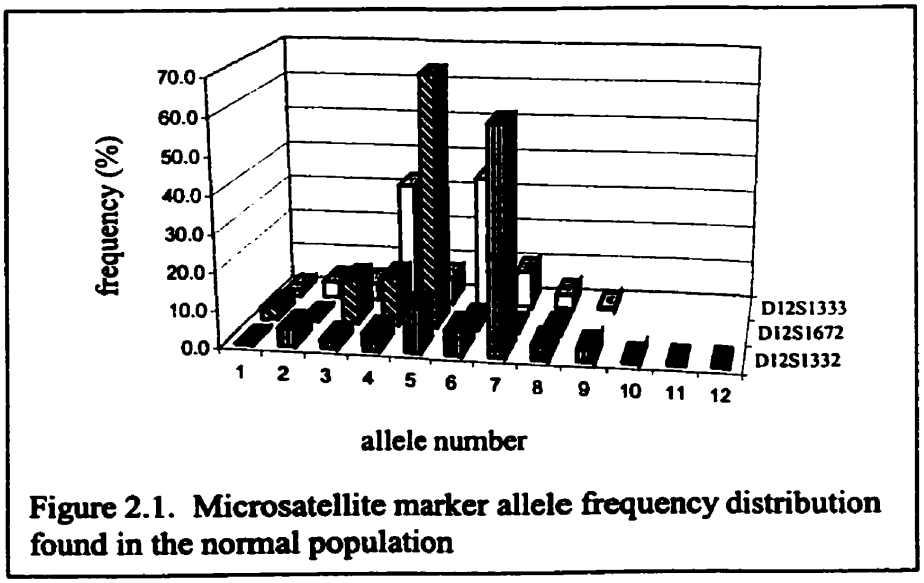

Figure 2.1. Microsatellite marker allele frequency distribution found in the normal population

Discussion

Consistent with results obtained in studies dealing with other CAG repeat disorders, haplotype analysis in our SCA2 families revealed that multiple mutational events appear to be responsible for the cases of SCA2 seen in our panel. Several distinct haplotypes exist in these families, both between and within geographic and ethnic groups.

Within the largest and most homogeneous group of families, the seven Indian families, five different haplotypes were seen. This is somewhat surprising considering that these families all share similar ethnic, religious, and geographic parameters. It is, however, interesting to note that despite the presence of five different haplotypes, six of the seven Indian families share allele "8" for D12S1333. While it is possible that the four haplotypes containing this "8" allele arose independently, the rarity of this allele in our control population (3.8%; see Fig. 2.1) as well as in 35 Indian control chromosomes (2.9%, data not shown), makes this less likely. Examination of the haplotypes containing this "8" allele reveals that ancestral relationships between some of these haplotypes may have been masked by marker mutation events. For instance, three of the haplotypes which contain the rare "8" allele (8-2-7, 8-2-8, and 8-1-7) might have evolved from a common ancestral mutation as a result of simple marker mutations. It is possible that the 8-2-7 haplotype, which is seen in two families, could have produced the 8-2-8 haplotype after the loss of a single CA repeat in D12S1332, as well as the 8-1-7 haplotype as a result of the addition of a single CA repeat in D12S1672. This scenario is even more likely considering the rarity of the

D12S1672 "1" and "2" alleles in the control population (3.4% and 0.6% respectively, see Fig. 2.1) and in 35 Indian control chromosomes (2.9% for each, data not shown). Indeed, the chances of an "8-2-X" haplotype arising twice independently are extremely remote ($[0.038 \times 0.006]^2$). As for the remaining "8" allele-containing haplotype (8-3-4), more complicated scenarios, perhaps including a recombination event, would be needed to link these chromosomes to the other 8-X-X haplotypes. Nonetheless, it is possible that our Indian families may have arisen from three ancestral mutations, not five, as the data might suggest. Conversely, it is possible that the presence of this "8" allele marks chromosomes in this population as having SCA2 repeats which are predisposed to undergoing expansion, causing the prevalence of this allele on SCA2 disease chromosomes in our Indian families. Examination of more markers, perhaps including intragenic single nucleotide polymorphisms, would be necessary to further investigate this question.

Three of the SCA2 families examined segregated the 4-5-7 haplotype: the Austrian-Canadian kindred, a Brazilian kindred, and a Portuguese kindred. While a case can certainly be made linking the Brazilian and Portuguese to a common ancestral mutation, it is also possible that these disease chromosomes, as well as the Austrian-Canadian chromosome, arose independently. As Fig. 2.1 shows, these alleles are relatively common in the control population (32.8%, 67.6%, and 59.9% for allele "4" at D12S1333, allele "5" at D12S1672, and allele "7" at D12S1332, respectively). The chances of an expanded allele occurring randomly on a chromosome with this haplotype is roughly 13% given these numbers, and

supposing complete linkage equilibrium between the markers. In fact, four of the isolated cases for whom it was possible to construct partial haplotypes could also potentially be carrying the 4-5-7 haplotype (see Table 2.1).

As with the Indian families, multiple haplotypes are seen within the Canadian pedigrees, the Brazilian pedigrees, and the Portuguese pedigrees, further supporting the notion that the formation of SCA2 mutations has not been a historically rare occurrence, even within geographic and ethnic groups . It is interesting to note, however, that one of the Portuguese families, the Portuguese 2 kindred, is reported to have ancestry in Goa, Western India, a region formerly known as Portuguese India. This family's ancestral link to India is interesting in light of the fact that the Portuguese 2 kindred segregates a 7-2-7 haplotype, reminiscent of the 8-2-7 seen in two of the Indian kindreds. The extreme rarity of the "2" allele at D12S1672, seen only once in 179 control chromosomes, makes a link between this Portuguese family and the Indian pedigrees an intriguing possibility. Unfortunately, sufficiently detailed genealogical data was not available on this Portuguese family to investigate this potential relationship further.

Very recently, Didierjean *et al.* (119) published findings similar to those found in this study. Using the same three markers, the authors reported several distinct haplotypes among the 23 families they investigated, with common haplotypes shared by small clusters of families, likely as a result of founder effects. Although the calculated allele sizes are slightly different between these published results and those found in this study, it is possible to compensate for

these differences by comparing the sizes of the predominant alleles for each marker in the control populations. Comparison of haplotypes reveals our Austrian-Canadian family as having the same haplotype as the three Austrian families in the Didierjean *et al.* study. Also interesting were haplotypes shared between our Portuguese 2 family and the Didierjean *et al.* Portuguese family, and between our Gaspé family and the Didierjean *et al.* French and German families. These shared haplotypes, combined with the potential ancestral connections between these families, make it quite possible that these families share common ancestral SCA2 mutations.

The data generated in these studies seem to support the increasingly apparent notion that the formation of CAG expansion mutations has not been restricted to a few, rare, ancestral events, but instead has occurred several times for each CAG disease locus. While several studies have generated convincing haplotype data in support of this concept (110,114,119), perhaps the most compelling evidence comes from groups reporting identification of *de novo* mutations in polyGln diseases. *De novo* mutations have thus far been reported in HD (115), SCA1 (113), SCA7 (82,84), SCA2 (58), and SCA6 (60), and have long been observed in other trinucleotide disorders such as fragile X syndrome (24). Interestingly, our data suggest that one of the isolated cases examined in this study may have resulted from a *de novo* mutation. The Greek patient listed in Table 2.1 is an affected male with an SCA2 expanded allele consisting of 42 repeats. His parents, both in their 60s, presently show no signs of neurological dysfunction. Molecular examination revealed SCA2 alleles of 22 and 28 repeats in the father,

and 22 and 34 repeats in the mother. It remains to be seen if the mother's 34 repeat allele will eventually be associated with disease, as this size allele appears to lie on the border between normal and expanded alleles for SCA2 (39,58). Regardless, it appears that this allele further expanded during transmission to the affected son, potentially representing a *de novo* expansion. It should be noted that subsequent testing of the proband's two sisters revealed alleles of 35 repeats in his 40 year old sister, and 37 repeats in his 32 year old sister. Neither sister currently shows signs of ataxia, although both should now be considered at risk (the proband was considered an isolated case for this study as other family members were not made available until after haplotype analysis was completed). Results such as these further demonstrate the unstable nature of CAG repeats, illustrate the way in which new SCA families must have been historically generated, and help explain the multiple haplotypes seen in the SCA2 families examined in this study.

In summary, we have used haplotype analysis to determine that SCA2 expanded alleles appear to have arisen through multiple, independent events in our SCA2 families. It was, however, possible to make some speculative links between families based on haplotype information, with a partial founder effect in the Indian population a distinct possibility. We were also able to make some assumptions about the spread of the SCA2 mutation between our families based on historical relationships and shared haplotype information. Nevertheless, our data underscore the unstable nature of the SCA2 repeat, and support the notion

that expansion events in trinucleotide repeat disorders may not be as rare as was originally thought.

Chapter 3

Investigating the role of the SCA6 and SCA7 repeat mutations in a large panel of ataxia patients

Haplotype analyses, like that discussed in chapter 2, point out that expansion from the normal range to the expanded range in many of the CAG repeat diseases is not a historically rare event. The relative commonness or rarity of such expansion events can also be investigated by screening for expansion mutations in a given population of ataxia patients. While general trends do exist, examination of the literature reveals that each panel of ataxia patients has a unique frequency distribution for each SCA subtype, depending on the geographic and ethnic make-up of the samples. This study investigated the frequency of the newly cloned SCA6 and SCA7 mutations in our large panel of ataxia patients.

This work will be submitted for publication as:

Hayes, S., Lopes-Cendes, I., Gaspar, C., Maciel, P., Silviera, I., Rousseau, M., and Rouleau, G. A. Incidence of the SCA6 and SCA7 repeat mutations in a large panel of ataxia patients.

Incidence of the SCA6 and SCA7 repeat mutations in a large panel of ataxia patients

Introduction

The autosomal dominant spinocerebellar ataxias are a heterogeneous group of often fatal disorders which typically involve neurodegeneration of the cerebellum, brain stem, and spinal cord. To date, ten spinocerebellar ataxia (SCA) loci have been identified (SCA1-SCA8, SCA10, and SCA11) (49-52), with six genes having already been cloned (SCA1, SCA2, SCA3/Machado-Joseph disease, SCA6, SCA7, and SCA8) (12-19). With the exception of SCA8, all of the genes identified so far have an expanded, polyglutamine encoding $(CAG)_n$ tract as the disease causing mutation.

The identification of the various genes involved in SCA has not only allowed scientists to examine the relative prevalence of the SCAs, it has enabled physicians to begin establishing a clinical phenotype for each SCA subtype. Early reports have suggested that two of the most recently cloned genes, SCA6 and SCA7, are associated with characteristic and identifiable clinical patterns. For instance, authors have suggested that SCA6 comprises a relatively "pure" cerebellar syndrome with late onset and few or no associated signs (17), while SCA7 is reported to be the only SCA to include retinal degeneration as a clinical feature (36). New evidence, however, suggests that the phenotypes of these forms of SCA might be more complicated. For example, SCA6 expanded alleles have been found in patients with associated signs, such as peripheral neuropathy (59),

in patients with early onset (66), and in patients suffering with episodic ataxia (67,68). Perhaps more surprisingly, SCA7 expansions have been found in patients where retinal degeneration occurs years after initial symptoms, as well as in patients with no history of visual disturbances at all (36,81). Such findings warn against ruling out (or in) these forms of SCA based on clinical observations, and emphasize the requirement that multiple studies be performed in multiple populations before a clear clinical picture of any of the SCA subtypes can be established. Caution should also be used when deciding which patients to include in a screen for SCA mutations based on mode of inheritance. Several studies have reported finding SCA expansions in patients with no family history (53-60), while a recent study by Moseley *et al.* (55) found SCA gene CAG expansions in autosomal dominant families, autosomal recessive families, and in isolated cases. As a result, it may be wise to include all cases which fit a broad SCA phenotype, regardless of inheritance pattern, in a screen for any of the SCA mutations.

In order to assess the relative frequencies of the recently cloned SCA6 and SCA7 subtypes in comparison to the other forms of SCA already screened by our laboratory, we performed a screen for these mutations in a large group of ataxia patients collected without regard to inheritance pattern or associated clinical features. This screen also allowed us to further investigate the clinical boundaries which have been established for these mutations.

Methods and Materials

We collected a total of 198 blood samples from unrelated individuals for SCA6 and SCA7 testing. These had previously been shown to be negative for the SCA1, SCA2, SCA3/MJD, and DRPLA mutations. All samples were either collected by us, or were sent to us for SCA testing by physicians from various parts of the world, with robust contributions from Portugal, Brazil, and North America. Individuals from families with autosomal dominant inheritance, autosomal recessive inheritance, as well as isolated cases, were included. The majority of the samples collected came from patients with progressive ataxia as their main clinical feature, although other clinical findings, such as episodic ataxia, peripheral neuropathy, and multiple system atrophy, were occasionally reported. All samples coming from patients with ataxia and retinal degeneration were tested for SCA7, as were 118 additional samples from patients with no history of visual difficulties. All 198 samples were examined for SCA6 expansions.

Blood from at least one affected individual per family was collected and used to establish lymphoblastoid cell lines, or directly for DNA extraction. Genomic DNA was obtained from lymphocytes or lymphoblastoid cell lines using standard protocols (116).

SCA6 and SCA7 repeat size were determined by polymerase chain reaction (PCR) in a total volume of 13.0 μl containing approximately 100 ng of genomic DNA, 10 mM Tris-HCl (pH 8.8), 1.5 mM $MgCl_2$, 50 mM KCl, 10% dimethylsulfoxide (DMSO), 250 mM each of dCTP, dGTP and dTTP, 25 mM

dATP, 1.5 μCi α^{35}S-dATP, 100 ng of each primer, and 1.5 U *Taq* polymerase (Perkin-Elmer). The SCA6 repeat was amplified using S-5-F1 (17), and SCA6-D (5'-CCG GGC CCC GCC GTG TCG-3'). SCA6-D is a primer of our design and was effective at reducing shadow bands obtained with S-5-R1 (17). The SCA7 repeat was amplified using the primer pair of 7ALT and 4U716 (83). PCR products were run on 6% denaturing polyacrylamide gels and sized by comparison with an M13 sequencing ladder.

Results

Of the 198 samples collected for this study, 57 samples (28.8%) came from families with a clear autosomal dominant pattern of inheritance. Forty-five samples (22.7%) were from families with apparent autosomal recessive inheritance, having at least one affected sibling in most cases, and phenotypically normal parents. Eighty-seven isolated cases (43.9%) were also included. These patients had no recorded family history of ataxia, although in some cases complete parental information was not available due to early parental death or other factors. Finally, nine samples (4.5%) were tested for which we had no known mode of inheritance.

SCA6 and SCA7 testing in our panel revealed only four individuals (2.0%) with an expanded repeat in either the SCA6 or SCA7 gene. The mutation responsible for SCA6 was found in a single individual, while three samples were positive for the SCA7 expansion.

The SCA6 positive sample we identified came from a French-Canadian family which showed a clear autosomal dominant pattern of inheritance, and whose affected individuals had previously been diagnosed as having Holmes ataxia. We tested nine additional affected individuals from this family, from three generations, and found an expanded allele of 26 repeats in all cases (Figure 3.1). All of the affected individuals in this pedigree were reported to suffer from severe ataxia, with dysarthria and nystagmus occurring in some patients. Ages of onset ranged from 22 years of age to 35 years of age, with the majority of patients reporting initial symptoms in their mid 20's.

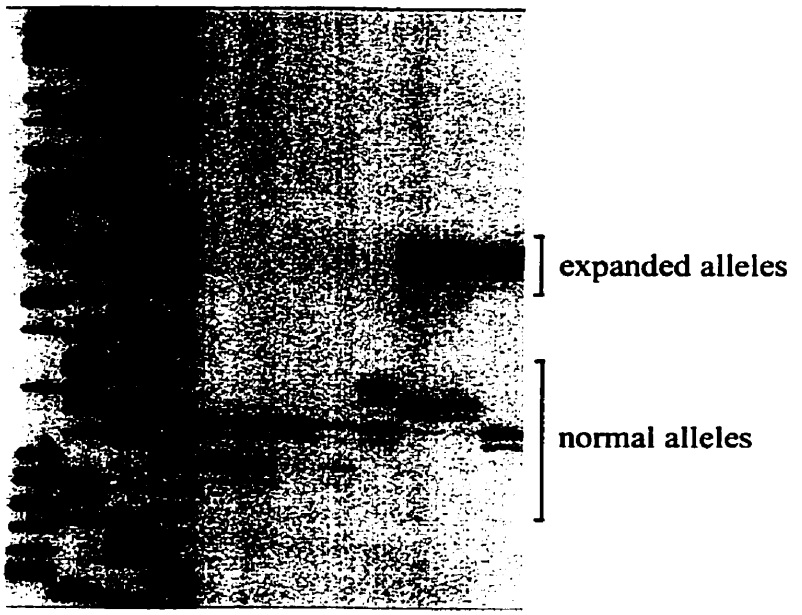

Figure 3.1. SCA6 normal and expanded alleles. The three lanes on the right contain three affected individuals from the same family.

Figure 3.2 shows the distribution of SCA6 alleles found in 236 individuals for whom we amplified this repeat. These data include results obtained during SCA6 screening for this study, as well as from SCA2 patients and normal controls typed for another experiment (S. Hayes, in preparation). SCA6 normal alleles range from 4 repeats to 18 repeats, with 11-13 repeats accounting for 82.1% of the alleles found.

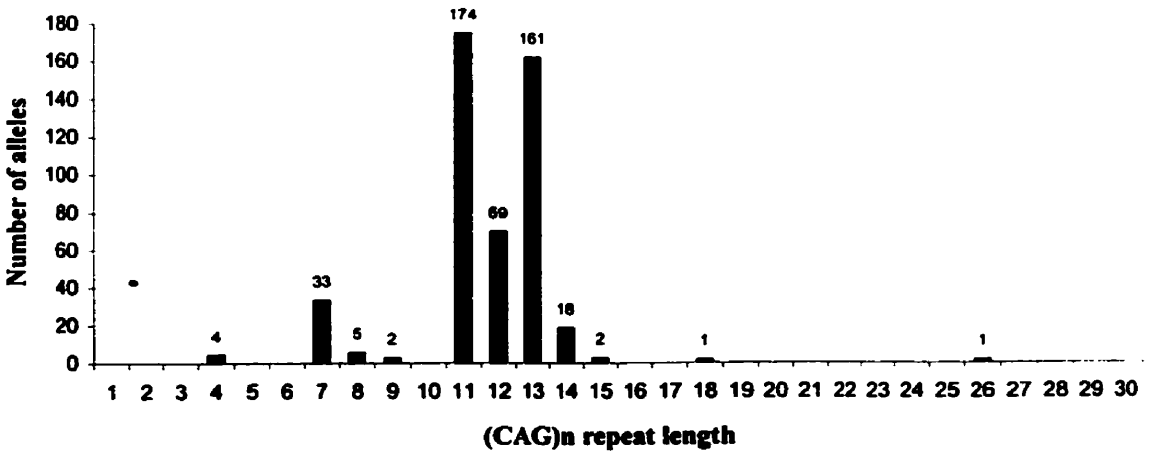

Figure 3.2. Distribution of normal and expanded $(CAG)_n$ repeats in the SCA6 gene.

Figure 3.3 shows normal and expanded allele ranges for the SCA7 repeat. Three unrelated individuals were found to carry SCA7 repeat expansions, with expanded allele sizes of 44, 51, and 54 repeats. An affected son of the patient with 44 repeats was also tested, and found to carry an expanded allele of 52 repeats, reflecting a further expansion of 8 repeats during this mother/son transmission. All three SCA7 positive samples were from autosomal dominant

families, and were associated with clinical features including limb and gait ataxia, dysarthria, hyperreflexia, and blindness. Age of onset ranged from early twenties to 40 years of age, with the maternal increase of 8 repeats associated with an 18 year decrease in the age at onset in this mother/son pair. No SCA7 repeat expansions were found in samples from patients without retinal degeneration. SCA7 normal alleles ranged from 7 repeats to 35 repeats, with 10 repeats accounting for 76% of the normal alleles typed. No alleles were found between 13 repeats and 35 repeats, with 35 repeats equaling the largest normal SCA7 allele yet reported (36,82).

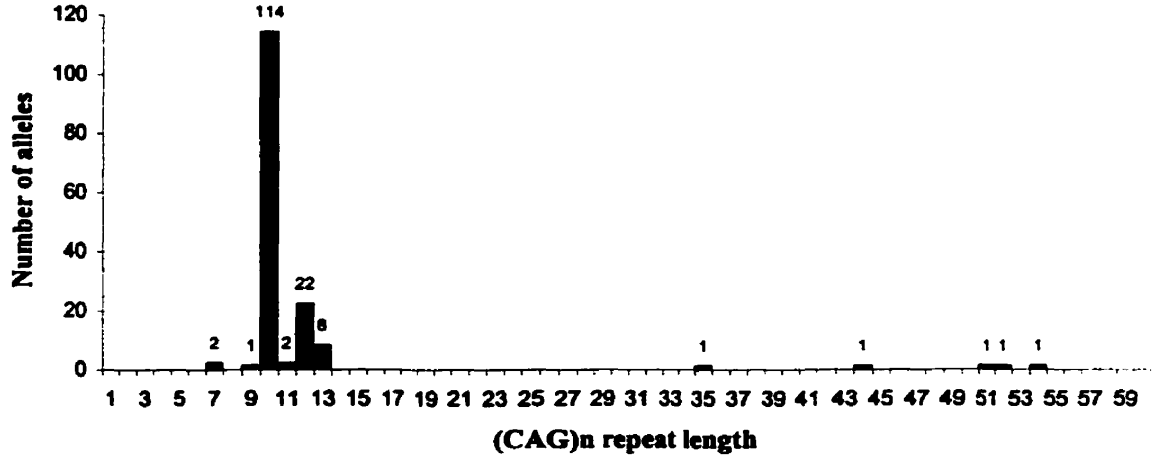

Figure 3.3. Distribution of normal and expanded $(CAG)_n$ repeats in the SCA7 gene.

SCA6 or SCA7 expanded repeats were not found in any patients from autosomal recessive families, or in any of the isolated cases tested. The four

samples identified with either SCA6 or SCA7 expansions accounted for 7.0% of the 57 autosomal dominant families tested in this study.

As was mentioned earlier, the panel tested in this study contained samples which had previously been shown to be negative for the SCA1, SCA2, SCA3/MJD, and DRPLA mutations. Table 3.1 shows the frequencies of all of the SCA loci tested, with data combined from previous studies as well as from this study.

Table 3.1. Frequencies of SCA1, 2, 3, 6, 7, and DRPLA among 351 ataxia samples.

Locus	Overall panel[a] (n=351)	Dominant families (n=206)	Recessive families (n=45)	Isolated cases (n=91)
SCA1	9 (2.6%)	9 (4.4%)	0 (0.0%)	0 (0.0%)
SCA2	20 (5.7%)	18 (8.7%)	0 (0.0%)	2 (2.2%)
SCA3/MJD	121 (34.5%)	119 (57.8%)	0 (0.0%)	2 (2.2%)
DRPLA	3 (0.9%)	3 (1.5%)	0 (0.0%)	0 (0.0%)
SCA6	1 (0.3%)	1 (0.5%)	0 (0.0%)	0 (0.0%)
SCA7	3 (0.9%)	3 (1.5%)	0 (0.0%)	0 (0.0%)
Unknown[b]	194 (55.3%)	53 (25.7%)	45 (100%)	87 (95.6%)

[a] includes 9 samples with undetermined inheritance

[b] families/samples which remain negative for all tested loci

Overall, we examined 351 unrelated samples for expansions, including 206 dominant samples, 45 recessive samples, 91 isolated cases, and 9 samples with unknown inheritance.

Discussion

We screened 198 ataxia patients for the SCA6 mutation, and 121 of these patients for the SCA7 mutation, and found these loci to be rare in our panel. Only 2.0% of the samples tested in this study were positive for either SCA6 or SCA7, which represented 7.0% of the autosomal dominant families examined.

Unlike the other SCAs, and consistent with previous reports (29), SCA6 expanded alleles appear to be stably transmitted from generation to generation in the single SCA6 positive family we identified. All ten affected individuals tested showed an expanded allele of 26 repeats, with no changes in repeat size in seven meioses over three generations (data not shown). Despite carrying the same sized expansion in all cases, affected individuals in this pedigree showed a 13 year range in age at onset, raising the possibility that other genetic or environmental factors may influence disease onset. This is particularly intriguing in light of the fact that later generations in this pedigree tend to show later ages of onset (data not shown), therefore lessening the likelihood that age of onset variability was the result of ascertainment bias. In contrast to reports labeling SCA6 as a late onset ataxia (17,68), this pedigree showed relatively early disease onset (22 to 35 years of age), perhaps reflecting the large SCA6 repeat carried by this family. Clinical symptoms were consistent with relatively pure cerebellar ataxia.

It is worth noting that point mutations in the gene which contains the SCA6 CAG repeat (CACNA1A) have been found to be responsible for episodic ataxia type 2 (72), as well as rare cases of progressive ataxia (73). As a result, we cannot rule out the possibility that some of our samples may contain point

mutations in the CACNA1A gene, especially those with episodic clinical episodes.

All four SCA7 positive samples identified in this study suffered from ataxia as well as retinal degeneration, consistent with published reports (36). The single SCA7 mother to son expansion seen in our data also seems to support published data which characterizes SCA7 as an extremely unstably transmitted expanded repeat (36,83). The maternally transmitted +8 repeat expansion seen in our sample is consistent with the average range of maternally transmitted SCA7 increases in the literature (36,76,83), and was also associated with clinical anticipation. Although we did not find sufficient SCA7 positive samples to formally regress repeat size against age at onset, it does appear that there is a general trend in our sample of increasing repeat size corresponding to earlier ages of onset.

It should be mentioned that the extremely unstable nature of the SCA7 repeat can lead to expansions too large to accurately identify by standard PCR techniques (81), and it is therefore possible that we have missed such expansions in our screen. However, our panel does not contain any patients with retinal degeneration and very early ages of onset, as would be expected from large SCA7 alleles, making the likelihood of missing such expansions quite small.

Combining the data generated in this study with screening results for other SCAs allows us to examine the true impact the SCA6 and SCA7 loci have in our panel. In 351 SCA samples, the most commonly found mutation was SCA3/MJD, which accounted for 34.5% of all SCA cases tested, followed by SCA2 at 5.7%,

SCA1 at 2.6%, SCA7 and DRPLA at 0.9%, and SCA6 at 0.3% (see Table 3.1). In our 206 autosomal dominant families, SCA3/MJD was the most frequent (57.8%), followed by SCA2 (8.7%), SCA1 (4.4%), SCA7 and DRPLA (1.5%) and SCA6 (0.5%). Overall, 194 (55.3%) of the samples in our panel remain genetically undefined, including 53 (25.7%) of the autosomal dominant families.

While it is interesting to compare the distribution of SCA expansions in our samples with other published reports, it should be noted that our panel is somewhat biased by a preponderance of Portuguese and Brazilian patients, likely contributing to an inflated occurrence of SCA3/MJD expansions. In fact, in our panel, all other SCA loci combined (other than SCA3/MJD) account for only 10.3% of our total SCA samples, and 16.5% of the autosomal dominant families. By comparison, SCA6 alone accounts for over 20% of SCA cases in Germany (63), and over 30% of autosomal dominant cases in Japan (56). It is becoming evident with each published SCA genotyping report that SCA distributions may vary greatly between geographic and ethnic groups, with local founder effects likely playing a major role in influencing relative proportions.

In summary, it appears that SCA6 and SCA7 are rare causes of SCA in our panel. While our SCA6 and SCA7 positive samples tended to fit the established clinical criteria for these loci, we did find relatively early disease onset in our SCA6 pedigree. We did not find SCA6 or SCA7 to be associated with any of our autosomal recessive SCA families or isolated cases. Despite screening for five known SCA loci, as well as DRPLA, 53 dominant (25.7%), 45 recessive (100%), and 87 sporadic samples (95.6%) in our panel remain genetically undefined.

Recent findings of homozygous Friedreich's ataxia (FRDA) mutations in isolated and autosomal recessive SCA patients (54,55) raises the possibility that FRDA expansions may be present in our panel as well. It will also be interesting to see whether the newly identified SCA8 will be a common cause of SCA in our panel, especially considering this locus' apparent association with recessive and sporadic SCA (19). Whether the mutations responsible for SCA4, SCA5, SCA10 and SCA11 will play a significant role in our panel awaits the discovery of these genes, but the clinical diversity of our samples, combined with the rather well-defined phenotypes associated with these SCA subtypes, seem to suggest that these loci will not be major contributors. Given the fact that the six loci examined thus far are responsible for less than half of the SCA in our panel, even with a substantial contribution from SCA3/MJD, it appears likely that several dominant and/or recessive SCA loci are still awaiting discovery.

Chapter 4

Examination of CAG repeat-containing loci as possible genetic modifiers of age at onset in spinocerebellar ataxia type 2 (SCA2)

For all of the SCAs, a significant proportion (~50%) of the age at onset variability seen between patients is not accounted for by the size of the expanded CAG repeat. This final study tested the hypothesis that other, non-expanded CAG repeats may act as genetic modifiers of age at onset variability in the SCAs. This study was inspired by the recent discovery of inappropriate protein aggregation in the polyglutamine disorders, and is based on the assumption that a) polyglutamine tracts have the ability to interact with one another within the cell, perhaps influencing protein aggregation, and b) variations in the repeat length of interacting CAG repeat-containing loci may influence this rate of aggregation, and thus disease progression (age at onset).

This work will be submitted for publication as:

Hayes, S., Turecki, G., Brisebois, K., Lopes-Cendes, I., Gaspar, C., Riess, O., Ranum, L.P.W., Pulst, S.-M., and Rouleau, G.A. CAG repeat length in *hGT1* is associated with age at onset variability in spinocerebellar ataxia type 2 (SCA2).

CAG repeat length in *hGT1* is associated with age at onset variability in spinocerebellar ataxia type 2 (SCA2)

Introduction

The autosomal dominant cerebellar ataxias (ADCAs) are a heterogeneous group of neurodegenerative disorders typified by gait, speech, and limb incoordination. The diseases are progressive in nature and generally feature degeneration of the cerebellum, brainstem, and spinocerebellar tracts (48). To date, ten spinocerebellar ataxia (SCA) loci have been implicated in the ADCAs (SCA1-SCA8, SCA10, and SCA11) (49-52), while six genes have been cloned (SCA1, SCA2, SCA3/Machado-Joseph disease, SCA6, SCA7, SCA8) (12-19) and shown to contain polymorphic $(CAG/CTG)_n$ repeats. With the exception of SCA8, these repeats are translated into polyglutamine tracts which appear to be toxic when expanded into a pathological range. For this group of "polyglutamine disorders", which also includes Huntington disease (HD) (9), dentatorubral-pallidoluysian atrophy (DRPLA) (10,11), and spinobulbar muscular atrophy (SBMA) (6), a negative correlation exists between the size of the repeat tract and the age at disease onset (27), with variations in repeat size generally accounting for ~25-80% of the variability in the age at onset (24,63,75), depending on the disease being examined. Several studies have suggested that the variability in age at onset not attributable to the repeat size is in part due to the existence of modifying genetic factors (24,120-122). How these modifiers affect the disease process is not known, but it is an attractive possibility that these factors may

interact with the disease proteins and, by virtue of their distribution, help explain why specific subsets of neurons appear vulnerable in each disease despite widespread expression of the disease proteins themselves (123). In fact, models have been proposed where polyglutamine-containing disease proteins interact with cell-specific factors via non-covalent "polar-zippers" (87), or via transglutaminase mediated covalent cross-linking (86), producing complexes which perhaps play a role in the characteristic patterns of cell death seen in each disease.

Recently, insoluble intranuclear inclusions (NIs) have been reported in both affected patient brain material, as well as transfected cells, for many of the polyglutamine disorders (123,124). While some evidence exists indicating that NIs may not be necessary for initiation of SCA1 and HD pathogenesis (104,105), the existence of expanded polyglutamine tracts in these structures, as well as the finding that NIs are commonly found in affected brain regions, suggest that NIs are, at some level, disease-related structures (125). Much attention is now being focused on determining the protein make-up of NIs, and initial results suggest that some proteins may enter these complexes in a disease-specific manner. For instance, a protein expressed predominantly in cerebellar Purkinje cells, the leucine-rich acidic nuclear protein (LANP), colocalizes to NIs in SCA1 (126) but is not found in NIs in SCA7 (98). Identification of the full protein complement of NIs is important as it is possible that the spectrum of interacting proteins found in each disorder may influence the process of cell death, while rates at which these proteins enter aggregates may impact on the rate of neurodegeneration.

While the list of possible interacting candidate proteins is indeed large, the ability of polyglutamine tracts to interact with one another makes proteins with even non-pathological polyglutamine stretches excellent candidates for recruitment into nuclear aggregates. In addition, the polymorphic nature of these repeats could potentially impact on the rate of aggregate formation. This in turn may influence age of disease onset, which makes polymorphic polyglutamine-coding genes strong candidates for roles as genetic modifiers of age at onset in the SCAs. Support for this hypothesis can be seen in Perez *et al.*'s (125) recent demonstration that ataxin-3 (the disease protein in SCA3/MJD) can interact with both ataxin-1 (the disease protein in SCA1) and the polyglutamine-containing TATA-binding protein (TBP), in nuclear aggregates. There is also evidence that an expanded repeat-containing protein can recruit the same protein with a non-expanded repeat into aggregates (91,95). However, the pathological consequences of these interactions, or their relationship to disease progression, remain unclear.

SCA2 is a typical polyglutamine disorder, with expansions of 34-59 repeats associated with disease (123). While no clear evidence of NIs has yet been reported in SCA2, the appearance of punctate, perinuclear ataxin-2 staining in patient Purkinje cells suggests that protein aggregation and sequestering may also play a role in this disease (103). Although expanded repeat sizes are relatively small in SCA2, expanded ataxin-2 still has the potential to form aggregates and interact with other polyglutamine-containing proteins, with potential pathological consequences. Therefore, as a way to begin addressing the possible role of polyglutamine-containing proteins as genetic modifiers in the

SCAs, we have investigated several CAG-containing loci as potential modifiers of the age at disease onset in SCA2.

Materials and methods

Patients

A panel consisting of 46 individuals taken from 10 SCA2 pedigrees was assembled for an initial screen with potential modifier loci. Genomic DNA was extracted from peripheral lymphocytes or lymphoblastoid cell lines using standard protocols (116). SCA2 repeat size was determined as reported elsewhere (77).

A second panel was assembled which included 26 individuals chosen at random from SCA2 pedigrees such that a single individual was tested per pedigree (in the case of small pedigrees) or per nuclear family (in the case of large pedigrees). This panel was used to re-test any loci which yielded suggestive results in the initial screen in order to address any biases that may have been introduced by testing multiple affected individuals from single families. Any loci which still produced significant results were also tested in 47 additional unrelated SCA2 patients obtained through collaboration, yielding a total of 73 unrelated SCA2 patients for this final screen.

Finally, a panel of 68 unrelated SCA3/MJD patients was assembled and used to test any loci which produced suggestive results with the SCA2 panels.

Molecular analysis

A total of 10 loci known to contain polymorphic $(CAG)_n$ tracts were examined as potential modifiers of SCA2 age of onset. These included six genes known to be involved in neurological disease (SCA1, SCA3/MJD, SCA6, SCA7, DRPLA, and HD), as well as four loci not known to be directly responsible for a disease phenotype (TBP, hSKCa3 [accession number AF031815], STS accession number G09710, and EST accession number U80757). These latter loci were chosen from a panel of 50 CAG repeat containing loci previously identified through database searches and used in an unrelated study (127). They were included based on their high heterozygosity and/or long repeat length. Repeats were amplified by polymerase chain reaction (PCR) using conditions reported elsewhere (77) and the primer pairs and annealing temperatures listed in Table 4.1. SCZ41 and SCZ42 amplify both CAG tracts in hSKCa3, which were added together to give total CAG length for this locus. Ten percent DMSO was included in all reactions except for SCA3/MJD, which used reaction conditions reported elsewhere (37). PCR products were run on 6% denaturing polyacrylamide gels and sized by comparison with an M13 sequencing ladder. For each locus, the number of CAG repeats in both alleles was recorded.

Table 4.1. PCR annealing temperatures, primers, and allelic results obtained for tested CAG-containing loci

Locus	Primer Pairs[a]	Annealing Temp.(°C)	Number of Alleles	Range of Alleles[b]
SCA1	Rep1 and Rep2 (12)	54	9	24 - 33
SCA3/MJD	MJD52 and MJD25a (37)	60	13	15 - 41
SCA6	S-5-F1 (17) and SCA6-D (5′-CCGGGCCCCGCCGTGTCG-3′)	68	8	4 - 15
SCA7	7ALT and 4U716 (83)	56	5	10 - 35
DRPLA	B37 primer pair (128)	62	16	11 - 27
hSKCa3	SCZ41 (5′-GGGCTGGGGGACTTGGATAA-3′) SCZ42 (5′-GGGCGAACTGAGACAGGGGATGC-3′)	61	12	24 - 35
HD	HD2 (5′-ATGGCGACCCTGGAAAAGCTGATGAA-3′) HD1 (5′-GGCGGCTGAGGAAGCTGAGGA-3′)	59	17	14 - 38
TBP	TBP1 (5′-CTGTCTATTTTGGAAGAGCAACAAAGG-3′) TBP2 (5′-CTGCTGGGACGTTGACTGCTGAAC-3′)	62	11	27 - 40
G09710	SCZ15 (5′-GGGGCAGCGGGTCCAGAATCTTC-3′) SCZ16 (5′-CTGGCCTTGCTGCCCGTAGTGCT-3′)	62	5	10 - 18
U80757	SCZ91 (5′-TGGTCGGGAAGCCAGTTTGTTTG-3′) SCZ92 (5′-CCACCGCATTCGGGGCAGAG-3′)	62	6	13 - 28

[a]primer references shown in brackets when available
[b]minimum and maximum sizes of CAG tracts

Statistical analysis

Statistical analysis was carried out using step-wise multiple linear regression with age at onset as the dependent variable (Systat V5.05). This analysis was considered suitable to study the potential effect of the investigated loci as age at onset modifiers because it controls for the SCA2 expansion. Number of CAG repeats were treated quantitatively and regressions were performed using the smaller allele at each locus, the larger allele at each locus, and the total of the two alleles (smaller plus larger). Overall "CAG load" was also tested by summing the large allele of each locus together (total-large), as well as both alleles at each locus together (total-all).

Results

Table 4.1 lists the size ranges and the number of alleles found for each of the loci tested. In our initial screen of 46 SCA2 patients, regression with the small allele of each locus tested, or the total of the two alleles, did not yield significant results (data not shown). When the large alleles were considered, the size of the SCA2 expanded repeat accounted for 48% of the variability seen in the age at onset ($R = -0.693$; $R^2 = 0.480$; $P < 0.0001$). Of the remaining loci that were tested, only G09710 produced a significant improvement of fit ($R^2 = 0.543$; $P = 0.021$). A log transformation of age at onset slightly increased the fit of the regression equation for the SCA2 repeat alone ($R^2 = 0.516$; $P < 0.0001$), as well as for G09710 ($R^2 = 0.576$; $P = 0.019$), and allowed DRPLA to enter the equation. After accounting for the effect of the SCA2 repeat and G09710, the large allele of

the DRPLA repeat improved the fit to $R^2 = 0.615$ ($p = 0.050$). At no time did overall CAG-load, as represented by summing both alleles of all loci, or the large allele of all loci, have an effect on age at onset.

In order to address the possibility that the presence of multiple affected individuals from single nuclear families may have been producing false positive results in the initial screen, we decided to perform a second screen using a single SCA2 positive individual per nuclear family. This new panel of 26 samples was used to retest SCA2, G09710, and DRPLA. Using the large alleles of each of these loci in step-wise multiple regressions versus age at onset produced an increase in the fit only for G09710 ($R^2 = 0.572$; $P = 0.003$ up from $R^2 = 0.362$; $P = 0.001$ for SCA2 repeat alone).

To further investigate this possible association, we obtained an additional and independent sample of 47 SCA2 subjects to produce a panel of 73 unrelated SCA2 patients. This panel was used to test G09710 in a final screen. Once again, the large allele of the G09710 locus significantly increased the fit from $R^2 = 0.447$ ($P < 0.0001$) for the SCA2 repeat alone, to $R^2 = 0.488$ ($P = 0.021$) for the two loci together. This represents an increase of 4.1% in the total explained SCA2 age at onset variability.

Finally, the possible role this locus may play in modifying age at onset in other SCAs was examined by assembling a panel of 68 unrelated SCA3/MJD patients and testing this panel for the effect of the SCA3/MJD repeat, as well as G09710. While the SCA3/MJD repeat accounted for 37.3% of age at onset

variability ($R = -0.610$; $P < 0.0001$), G09710 did not produce a significant effect after step-wise multiple linear regression (data not shown).

Discussion

We have identified the polymorphic $(CAG)_n$ repeat within the STS G09710 as a potential modifier of age at disease onset for SCA2. In our SCA2 patients, the large allele of this locus accounts for an additional 4.1% of the age at onset variability after accounting for the effect of the SCA2 expanded repeat.

Interestingly, the magnitude of the effect attributable to the repeat in G09710 is similar to that predicted by Ranum *et al.* (120), who found that approximately 5% of age at onset variation in SCA1 may be due to unidentified genetic factors. A similar percentage of the total variation in age of onset in HD (4.1%) was reported by Rubinsztein *et al.* (121) to be associated with a polymorphic TAA repeat in GluR6 kainate receptor locus. Although the identification of potential modifiers for the polyglutamine disorders is a significant finding, it should be noted that the modest size of these effects still makes predictive testing in these disorders impossible, as too much age of onset variation remains unexplained. As a result, the role of loci such as those reported here may be more important for the clues they can provide about the pathways and mechanisms that may be involved in the pathogenesis of the polyglutamine disorders.

As noted by Joober *et al.* (127) G09710 appears to be the human homologue of the mouse gene *GT1*, a 196 kDa protein of unknown function which

is expressed during retinoic acid-induced differentiation of mouse embryonic carcinoma cells into neurons (129). This neuronal expression reinforces the possibility that *GT1* might be involved in a neurodegenerative disease. If human *GT1* (*hGT1*) is also expressed in neurons, it may be available for interaction with ataxin-2. In our study, larger *hGT1* alleles were associated with earlier ages of onset (after accounting for SCA2 repeat size). While it is impossible to establish the true nature of this association at this time, it is tempting to speculate that hGT1 protein might interact with ataxin-2 more readily as the hGT1 polyglutamine tract increases in size. This may in turn influence the rate of protein aggregation and therefore impact disease progression, as reflected by the age at onset association. Further attempts to identify discernible protein aggregates in SCA2, as well as their constituents, will be necessary to test the validity of this model. If protein aggregation is occurring in SCA2, it would be interesting to see if the hGT1 gene product, or other polyglutamine-containing proteins, are involved in their formation. Such interactions may simply influence the rate of aggregate formation and growth, or serve to sequester important cellular components and disrupt cellular pathways. Many groups are presently attempting to address these questions.

Recently, Perez *et al.* (125) reported that ataxin-3 can interact with both ataxin-1 and TBP in cellular aggregates. It is interesting to note that we found no association between the TBP CAG repeat and age of onset for SCA2 patients, and that similar testing with SCA3/MJD patients did not produce significant results (data not shown). This is not to say that TBP is not involved in aggregate

formation in these diseases, but simply that it does not appear to influence age at onset in our panels. In addition to those noted by Perez *et al.*, ataxin-3 may interact with other polyglutamine-containing proteins, including the hGT1 gene product, although we did not find *hGT1* to be associated with variation in SCA3/MJD disease onset. This is perhaps not surprising considering that it is possible that each polyglutamine disorder will have a unique spectrum of interacting proteins as well as genetic modifiers. Nevertheless, Perez and colleagues' findings further support the possibility that "normal length" polyglutamine-containing proteins may be involved in the pathogenesis of the polyglutamine disorders. To this end, we had hypothesized that overall "CAG load", that is the total of all the polyglutamine proteins found in the affected neurons, might have an influence on the disease process. While the data we obtained for the total of the 10 loci tested did not support this idea, many variables exist which were not controllable. For example, the cells which are vulnerable in SCA2 potentially only express a subset of the 10 loci tested, while at the same time expressing other polyglutamine-containing proteins which were not tested. Expression levels and compartmentalization of the various proteins will also influence cellular polyglutamine levels. As a result, our "CAG load" is most likely not an accurate reflection of the actual levels found within the affected cells. Nevertheless, it appears that certain CAG tract-containing genes, such as *hGT1*, may play a role in modifying the age of disease onset for at least some of the SCAs.

It should be noted that the precise magnitude of the effect any locus has on age at onset in the polyglutamine disorders is subject to the accuracy of several variables. A certain degree of error can be expected when patients or their families are asked to recall an accurate age of disease onset. Likewise, exact sizing of polymorphic CAG repeats can be difficult, especially when somatic mosaicism produces several bands, as is seen in the expanded alleles of most SCAs. Indeed, the source tissue for DNA extraction (lymphocytes, lymphoblastoid cell lines, brain material) used in the analysis may also affect the repeat size results obtained (39). As a result, replication of results such as those obtained in this study will be important before the absolute validity of any associations can be established.

As Rubinsztein *et al.* (121) points out, the boundaries between monogenic and multifactorial diseases have become blurred somewhat in recent years by the identification of genetic modifiers. Diseases such as the polyglutamine disorders appear to be more genetically complex than originally imagined. Discoveries of even modest modifier effects are important as they may serve to increase our overall understanding of the pathological mechanisms of these diseases. Hopefully, this increased understanding will ultimately lead to our ability to effectively treat, or even cure, these devastating disorders.

Chapter 5

Conclusion

Summary and implications of thesis results

This work comprising this thesis attempted to address several interesting aspects of the genetics of the spinocerebellar ataxias. The first part of this study was designed to investigate the origins of the SCA2 mutation in our SCA2 positive families. Using haplotype analysis, we were able to determine that multiple mutational events were responsible for the cases of SCA2 found in our panel. Having said this, these data also indicated potential ancestral relationships between several of the families examined.

Results of this type can reveal a great deal about the origins and spread of CAG expansion mutations, and are also useful indicators of the historical commonness or rarity of mutational events. Our results add to the growing cache of data which seem to indicate that the transition from the normal range to the expanded range in the CAG repeat diseases is not as uncommon as it was once believed to be. Even in fairly geographically and/or ethnically homogeneous groups, logically perfect candidates for founder effects, it is now evident that multiple mutational events can be responsible for disease. In fact, the discovery of *de novo* mutations seen in several polyGln disease families (60,82,84,115) points out that the generation of new expansion mutations is an ongoing process.

There is some evidence that large normal CAG alleles in any given population can serve as a reservoir for the introduction of new expanded

mutations into that population (107), similar to the role premutations play in type 2 repeat disorders like fragile X syndrome. This continual reintroduction of disease alleles into a population is not surprising when it is considered that such a mechanism is necessary to maintain disease prevalence over time. As anticipation drives disease onset below the reproductive years, new mutations arise, thus maintaining disease prevalence. Therefore, while it is certainly true that in some populations a founder effect is responsible for high disease prevalence, haplotype analysis results such as those seen in this study seem to suggest that high disease prevalence may, in some cases, result from the existence of a large pool of normal alleles predisposed to expansion within that population. If this is indeed the case, it will hopefully prompt researchers to begin identifying what factors and mechanisms are responsible for CAG repeat instability, ultimately impacting future disease monitoring and treatment.

The factors which contribute to the occurrence of disease mutations in a particular population also directly impact the second part of this study, which examined the prevalence of the SCA6 and SCA7 mutations in our collection of SCA patients. While the identification of each new SCA mutation provides many research opportunities, groups inevitably begin by determining the relative prevalence of new mutations in their SCA panels. Not only does this allow researchers to determine which SCA subtype(s) to focus future research energy upon, it can also have an important clinical impact - allowing physicians to inform patients of their genetic status, and begin implementing more effective genetic counseling, which may include genetic testing of at-risk family members.

Examination of the literature quickly reveals that, while general trends exist as to the relative prevalence of each SCA subtype, each group's testing produces a unique subtype frequency distribution. This distribution is, of course, a reflection of the specific geographic and ethnic background of the patients that make up any given panel. We were able to determine that the SCA6 and SCA7 mutations were relatively rare causes of SCA in our panel, accounting for only 0.3% and 0.9%, respectively, of all SCA cases examined. Our overall testing, including the SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and DRPLA loci, has identified the causal mutation for less than half of our SCA patient panel. These data would seem to suggest that several more SCA loci remain undiscovered. While it is a distinct possibility that the recently cloned SCA8 locus, as well as the presently uncloned SCA4, SCA5, SCA10 and SCA11 loci will account for a significant proportion of our genetically undefined samples, it would seem logical that other loci will also be involved.

The increasing genetic diversity of the spinocerebellar ataxias is a reflection of the increasing importance of dynamic repeat mutations in the modern study of human genetics. It is truly amazing how a mutation type which was unknown even a decade ago has grown to garner so much attention. With the potential to have an impact upon many other diseases, perhaps including psychiatric conditions such as schizophrenia and bipolar disorder, the scrutiny dynamic mutations receive will likely continue to expand.

As was mentioned earlier, genetic testing for identified SCA mutations has become a useful clinical tool, but the value of such testing on genetic counseling

is tempered by only a moderate level of predicative validity between repeat size and age of disease onset. Ideally, genetic counselors would like to be able to make concrete predictions about disease onset and severity based on mutant repeat size, but at present too much variability in age of onset data remains unexplained by repeat size. Identification of genetic modifiers which can account for some of the ~50% variability in age of onset not explained by repeat size would not only prove clinically valuable, it could potentially provide insights into the molecular mechanisms of neurodegeneration by highlighting putative molecular interactions and pathways. It is also quite possible that each CAG repeat disease will have a unique spectrum of genetic modifiers and interacting proteins, which may help explain the population-specific susceptibility seen in each disease.

The third component of this study attempted to address some of these issues by investigating potential contributions of candidate loci to age of onset variability in SCA2. Our study was based on the hypothesis that gene products from CAG-containing loci are ideal candidates for interaction with polyglutamine-containing disease proteins, perhaps through transglutaminase-mediated crosslinking, or non-covalent "polar-zippering". By screening a panel of SCA2 patients with CAG-containing candidate loci, we were able to identify a locus, *hGT1*, whose CAG repeat tract accounted for 4.1% of SCA2 age of onset variability not explained by the SCA2 expanded repeat itself. While this effect is certainly modest, it does suggest that CAG repeat-containing loci may be involved in modifying the pathological effects of polyglutamine disease proteins. These data are particularly interesting in light of the recent findings of neuronal

intranuclear inclusions (NIs) in several of the polyglutamine disorders (123,130). If these protein aggregates are involved in disease pathogenesis (a topic currently under debate), it is distinctly possible that other proteins involved in these complexes may modify the neurodegenerative process. Efforts to identify the components of nuclear aggregates are underway, and thus far have identified several proteins, including: ubiquitin in most disease subtypes (123), chaperones and proteasome elements in SCA1 and SCA3/MJD (101,102), the Purkinje cell-specific leucine-rich acidic nuclear protein (LANP) in SCA1 (126), and the polyglutamine-containing TATA-binding protein in SCA3/MJD (125). This last result once again illustrates the potential for polyglutamine-containing proteins to interact, as was implied by the SCA2 modifying effects of *hGT1*. Interestingly, we did not find the *hGT1* repeat to be associated with age of onset variability in SCA3/MJD, perhaps supporting the notion that each polyglutamine disease has a unique spectrum of interacting proteins.

As more loci and/or proteins are identified as being associated with CAG-repeat disease pathology, we inch one step closer to understanding the complex factors which influence disease pathogenesis in the polyglutamine diseases. Although the three components of this study deal with specific SCA subtypes, the findings and conclusions are in many cases transferable to other subtypes, and contribute to our growing knowledge and understanding of the spinocerebellar ataxias, as well as the polyglutamine disorders in general. This new found knowledge, combined with the almost daily contributions made by other researchers in this area, will hopefully soon lead to effective treatment strategies

for disorders such as Huntington disease and the spinocerebellar ataxias, at long last giving some hope to individuals who suffer from these devastating diseases.

References

1. Kremer, E.J., Pritchard, M., Lynch, M., Yu, S., Holman, K., Baker, E., Warren, S.T., Schlessinger, D., Sutherland, G.R. and Richards, R.I. (1991) Mapping of DNA instability at the fragile X to a trinucleotide repeat sequence p(CCG)n. *Science,* **252,** 1711-1714.
2. Verkerk, A.J., Pieretti, M., Sutcliffe, J.S., Fu, Y.H., Kuhl, D.P., Pizzuti, A., Reiner, O., Richards, S., Victoria, M.F., Zhang, F.P., Eussen, B.E., van Ommen, G.J.B., Blonden, L.A.J., Riggins, G.J., Chastain, J.L., Kunst, C.B., Galjaard, H., Caskey, C.T., Nelson, D.L., Oostra, B.A. and Warren, S.T. (1991) Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell,* **65,** 905-914.
3. Knight, S.J., Flannery, A.V., Hirst, M.C., Campbell, L., Christodoulou, Z., Phelps, S.R., Pointon, J., Middleton-Price, H.R., Barnicoat, A., Pembrey, M.E., Holland, J., Oostra, B.A., Bobrow, M. and Davies, K.E. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell,* **74,** 127-134.
4. Brook, J.D., McCurrach, M.E., Harley, H.G., Buckler, A.J., Church, D., Aburatani, H., Hunter, K., Stanton, V.P., Thirion, J.P., Hudson, T., Sohn, R., Zemelman, B., Snell, R., Rundle, S., Crow, S., Davies, J., Shelbourne, P., Buxton, J., Jones, C., Juvonen, V., Johnson, K., Harper, P., Shaw, D. and Housman, D. (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,* **69,** 385.
5. Campuzano, V., Montermini, L., Molto, M.D., Pianese, L., Cossee, M., Cavalcanti, F., Monros, E., Rodius, F., Duclos, F., Monticelli, A., Zara, F., Canizares, J., Koutnikova, H., Bidichandani, S.I., Gellera, C., Brice, A., Trouillas, P., De Michele, G., Filla, A., De Frutos, R., Palau, F., Patel, P.I., Di Donata, S., Mandel, J.-L., Cocozza, S., Koenig, M. and Pandolfo, M.

(1996) Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. *Science,* **271**, 1423-1427.

6. La Spada, A.R., Wilson, E.M., Lubahn, D.B., Harding, A.E. and Fischbeck, K.H. (1991) Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature,* **352**, 77-79.
7. Lafreniere, R.G., Rochefort, D.L., Chretien, N., Rommens, J.M., Cochius, J.I., Kalviainen, R., Nousiainen, U., Patry, G., Farrell, K., Soderfeldt, B., Federico, A., Hale, B.R., Cossio, O.H., Sorensen, T., Pouliot, M.A., Kmiec, T., Uldall, P., Janszky, J., Pranzatelli, M.R., Andermann, F., Andermann, E. and Rouleau, G.A. (1997) Unstable insertion in the 5' flanking region of the cystatin B gene is the most common mutation in progressive myoclonus epilepsy type 1, EPM1. *Nat. Genet.,* **15**, 298-302.
8. Lalioti, M.D., Scott, H.S., Buresi, C., Rossier, C., Bottani, A., Morris, M.A., Malafosse, A. and Antonarakis, S.E. (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. *Nature,* **386**, 847-851.
9. The 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.
10. 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. and Tsuji, S. (1994) Unstable expansion of CAG repeat in hereditary dentatorubral-pallidoluysian atrophy (DRPLA). *Nat. Genet.,* **6**, 9-13.
11. Nagafuchi, S., Yanagisawa, H., Sato, K., Shirayama, T., Ohsaki, E., Bundo, M., Takeda, T., Tadokoro, K., Kondo, I., Murayama, N., Tanaka, Y., Kikushima, H., Umino, K., Kurosawa, H., Furukawa, T., Nihei, K., Inoue, T., Sano, A., Komure, O., Takahaski, M., Yoshizawa, T., Kanazawa, I. and Yamada, M. (1994) Dentatorubral and pallidoluysian

atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat. Genet.*, **6**, 14-18.

12. Orr, H.T., Chung, M.Y., Banfi, S., Kwiatkowski, T.J., Jr., Servadio, A., Beaudet, A.L., McCall, A.E., Duvick, L.A., Ranum, L.P. and Zoghbi, H.Y. (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat. Genet.*, **4**, 221-226.
13. Pulst, S.M., Nechiporuk, A., Nechiporuk, T., Gispert, S., Chen, X.N., Lopes-Cendes, I., Pearlman, S., Starkman, S., Orozco-Diaz, G., Lunkes, A., DeJong, P., Rouleau, G.A., Auburger, G., Korenberg, J.R., Figueroa, C. and Sahba, S. (1996) Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat. Genet.*, **14**, 269-276.
14. 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., Takahashi, H. and Tsuji, S. (1996) Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat. Genet.*, **14**, 277-284.
15. 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. and Brice, A. (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.
16. 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. (1994) CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat. Genet.*, **8**, 221-228.

17. Zhuchenko, O., Bailey, J., Bonnen, P., Ashizawa, T., Stockton, D.W., Amos, C., Dobyns, W.B., Subramony, S.H., Zoghbi, H.Y. and Lee, C.C. (1997) Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat. Genet.*, **15**, 62-69.

18. 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. and Brice, A. (1997) Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat. Genet.*, **17**, 65-70.

19. Koob, M.D., Moseley, M.L., Schut, L.J., Benzow, K.A., Bird, T.D., Day, J.W. and Ranum, L.P. (1999) An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat. Genet.*, **21**, 379-384.

20. Brais, B., Bouchard, J.P., Xie, Y.G., Rochefort, D.L., Chretien, N., Tome, F.M., Lafreniere, R.G., Rommens, J.M., Uyama, E., Nohira, O., Blumen, S., Korczyn, A.D., Heutink, P., Mathieu, J., Duranceau, A., Codere, F., Fardeau, M., Rouleau, G.A. and Korczyn, A.D. (1998) Short GCG expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. *Nat. Genet.*, **18**, 164-167.

21. Richards, R.I. and Sutherland, G.R. (1997) Dynamic mutation: possible mechanisms and significance in human disease. *Trends Biochem. Sci.*, **22**, 432-436.

22. Ashley, C.T., Jr. and Warren, S.T. (1995) Trinucleotide repeat expansion and human disease. *Annu. Rev. Genet.*, **29**, 703-728.

23. Mandel, J.-L. (1997) Breaking the rule of three. *Nature*, **386**, 767-769.

24. Paulson, H.L. and Fischbeck, K.H. (1996) Trinucleotide repeats in neurogenetic disorders. *Annu. Rev. Neurosci.*, **19**, 79-107.

25. Reddy, P.S. and Housman, D.E. (1997) The complex pathology of trinucleotide repeats. *Curr. Opin. Cell Biol.*, **9**, 364-372.

26. Meijer, H., de Graaff, E., Merckx, D.M., Jongbloed, R.J., de Die-Smulders, C.E., Engelen, J.J., Fryns, J.P., Curfs, P.M. and Oostra, B.A. (1994) A deletion of 1.6 kb proximal to the CGG repeat of the FMR1 gene causes the clinical phenotype of the fragile X syndrome. *Hum. Mol. Genet.*, **3**, 615-620.

27. Koshy, B.T. and Zoghbi, H.Y. (1997) The CAG/polyglutamine tract diseases: gene products and molecular pathogenesis. *Brain Pathol.*, **7**, 927-942.

28. Martin, J.B. and Gusella, J.F. (1986) Huntington's disease: pathogenesis and management. *N. Engl. J. Med.*, **315**, 1267-1276.

29. Zoghbi, H.Y. (1997) CAG repeats in SCA6. Anticipating new clues. *Neurology*, **49**, 1196-1199.

30. Burright, E.N., Orr, H.T. and Clark, H.B. (1997) Mouse models of human CAG repeat disorders. *Brain Pathol.*, **7**, 965-977.

31. Ordway, J.M., Tallaksen-Greene, S., Gutekunst, C.A., Bernstein, E.M., Cearley, J.A., Wiener, H.W., Dure, L.S.I., Lindsey, R., Hersch, S.M., Jope, R.S., Albin, R.L. and Detloff, P.J. (1997) Ectopically expressed CAG repeats cause intranuclear inclusions and a progressive late onset neurological phenotype in the mouse. *Cell*, **91**, 753-763.

32. Lopes-Cendes, I., Maciel, P., Kish, S., Gaspar, C., Robitaille, Y., Clark, H.B., Koeppen, A.H., Nance, M., Schut, L., Silveira, I., Coutinho, P., Sequeiros, J. and Rouleau, G.A. (1996) Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado-Joseph disease. *Ann. Neurol.*, **40**, 199-206.

33. Cancel, G., Gourfinkel-An, I., Stevanin, G., Didierjean, O., Abbas, N., Hirsch, E., Agid, Y. and Brice, A. (1998) Somatic mosaicism of the CAG repeat expansion in spinocerebellar ataxia type 3/Machado-Joseph disease. *Hum. Mutat.*, **11**, 23-27.

34. Hashida, H., Goto, J., Kurisaki, H., Mizusawa, H. and 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. *Ann. Neurol.*, **41**, 505-511.

35. Chung, M.Y., Ranum, L.P., Duvick, L.A., Servadio, A., Zoghbi, H.Y. and Orr, H.T. (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat. Genet.*, **5**, 254-258.

36. David, G., Durr, A., Stevanin, G., Cancel, G., Abbas, N., Benomar, A., Belal, S., Lebre, A.S., Abada-Bendib, M., Grid, D., Holmberg, M., Yahyaoui, M., Hentati, F., Chkili, T., Agid, Y. and Brice, A. (1998) Molecular and clinical correlations in autosomal dominant cerebellar ataxia with progressive macular dystrophy (SCA7). *Hum. Mol. Genet.*, **7**, 165-170.

37. Maciel, P., Gaspar, C., DeStefano, A.L., Silveira, I., Coutinho, P., Radvany, J., Dawson, D.M., Sudarsky, L., Guimaraes, J., Loureiro, J.E.L., Nezarati, M.M., Corwin, L.I., Lopes-Cendes, I., Rooke, K., Rosenberg, R., MacLeod, P., Farrer, L.A., Sequeiros, J. and Rouleau, G.A. (1995) Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am. J. Hum. Genet.*, **57**, 54-61.

38. Hannan, A.J. (1996) Trinucleotide-repeat expansions and neurodegenerative disease: a mechanism of pathogenesis. *Clin. & Exper. Pharm. & Physiol.*, **23**, 1015-1020.

39. Cancel, G., Durr, A., Didierjean, O., Imbert, G., Burk, K., Lezin, A., Belal, S., Benomar, A., Abada-Bendib, M., Vial, C., Guimaraes, J., Chneiweiss, H., Stevanin, G., Yvert, G., Abbas, N., Saudou, F., Lebre, A.S., Yahyaoui, M., Hentati, F., Vernant, J.C., Klockgether, T., Mandel, J.L., Agid, Y. and Brice, A. (1997) Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. *Hum. Mol. Genet.*, **6**, 709-715.

40. Takiyama, Y., Sakoe, K., Soutome, M., Namekawa, M., Ogawa, T., Nakano, I., Igarashi, S., Oyake, M., Tanaka, H., Tsuji, S. and 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. *Hum. Mol. Genet.*, **6**, 1063-1068.

41. Watanabe, M., Abe, K., Aoki, M., Yasuo, K., Itoyama, Y., Shoji, M., Ikeda, Y., Iizuka, T., Ikeda, M., Shizuka, M., Mizushima, K. and Hirai, S. (1996) Mitotic and meiotic stability of the CAG repeat in the X-linked spinal and bulbar muscular atrophy gene. *Clin. Genet.*, **50**, 133-137.
42. Shizuka, M., Watanabe, M., Ikeda, Y., Mizushima, K., Kanai, M., Tsuda, T., Abe, K., Okamoto, K. and Shoji, M. (1998) Spinocerebellar ataxia type 6 - CAG trinucleotide expansion, clinical characteristics and sperm analysis. *Eur. J. of Neurol.*, **5**, 381-387.
43. Wells, R.D. (1996) Molecular basis of genetic instability if triplet repeats. *J. Biol. Chem.*, **271**, 2875-2878.
44. Gacy, A.M., Goellner, G., Juranic, N., Macura, S. and McMurray, C.T. (1995) Trinucleotide repeats that expand in human disease form hairpin structures in vitro. *Cell*, **81**, 533-540.
45. Djian, P. (1998) Evolution of simple repeats in DNA and their relation to human disease. *Cell*, **94**, 155-160.
46. Igarashi, S., Takiyama, Y., Cancel, G., Rogaeva, E.A., Sasaki, H., Wakisaka, A., Zhou, Y.X., Takano, H., Endo, K., Sanpei, K., Oyake, M., Tanaka, H., Stevanin, G., Abbas, N., Durr, A., Rogaev, E.I., Sherrington, R., Tsuda, T., Ikeda, M., Cassa, E., Nishizawa, M., Benomar, A., Julien, J., Weissenbach, J., Wang, G.-X., Agid, Y., St. George-Hyslop, P.H., Brice, A. and Tsuji, S. (1996) Intergenerational instability of the CAG repeat of the gene for Machado- Joseph disease (*MJD1*) is affected by the genotype of the normal chromosome: implications for the molecular mechanisms of the instability of the CAG repeat. *Hum. Mol. Genet.*, **5**, 923-932.

47. Maciel, P., Gaspar, C., Guimaraes, L., Goto, J., Lopes-Cendes, I., Hayes, S., Arvidsson, K., Dias, A., Sequeiros, J., Sousa, A. and Rouleau, G.A. (1999) Study of three intragenic polymorphisms in the Machado-Joseph disease gene (*MJD1*) in relation to genetic instability of the (CAG)n tract. *Eur. J. Hum. Genet.*, **7**, 147-156.

48. Harding, A.E. (1993) Clinical features and classification of inherited ataxias. *Adv. Neurol.*, **61**, 1-14.

49. Flanigan, K., Gardner, K., Alderson, K., Galster, B., Otterud, B., Leppert, M.F., Kaplan, C. and Ptacek, L.J. (1996) Autosomal dominant spinocerebellar ataxia with sensory axonal neuropathy (SCA4): clinical description and genetic localization to chromosome 16q22.1. *Am. J. Hum. Genet.*, **59**, 392-399.

50. Ranum, L.P., Schut, L.J., Lundgren, J.K., Orr, H.T. and Livingston, D.M. (1994) Spinocerebellar ataxia type 5 in a family descended from the grandparents of President Lincoln maps to chromosome 11. *Nat. Genet.*, **8**, 280-284.

51. Zu, L., Figueroa, K.P., Grewal, R. and Pulst, S.M. (1999) Mapping of a new autosomal dominant spinocerebellar ataxia to chromosome 22. *Am. J. Hum. Genet.*, **64**, 594-599.

52. Worth, P.F., Giunti, P., Gardner-Thorpe, C., Dixon, P.H., Davis, M.B. and Wood, N.W. (1999) Autosomal dominant cerebellar ataxia type III: Linkage in a large British family to a 7.6-cM region on chromosome 15q14-21.3. *Am. J. Hum. Genet.*, **65**, 420-426.

53. Ikeuchi, T., Takano, H., Koide, R., Horikawa, Y., Honma, Y., Onishi, Y., Igarashi, S., Tanaka, H., Nakao, N., Sahashi, K., Tsukagoshi, H., Inoue, K., Takahashi, H. and Tsuji, S. (1997) Spinocerebellar ataxia type 6: CAG repeat expansion in alpha1A voltage-dependent calcium channel gene and clinical variations in Japanese population. *Ann. Neurol.*, **42**, 879-884.

54. Leggo, J., Dalton, A., Morrison, P.J., Dodge, A., Connarty, M., Kotze, M.J. and Rubinsztein, D.C. (1997) Analysis of spinocerebellar ataxia types

1, 2, 3, and 6, dentatorubral-pallidoluysian atrophy, and Friedreich's ataxia genes in spinocerebellar ataxia patients in the UK. *J. Med. Genet.*, **34**, 982-985.

55. Moseley, M.L., Benzow, K.A., Schut, L.J., Bird, T.D., Gomez, C.M., Barkhaus, P.E., Blindauer, K.A., Labuda, M., Pandolfo, M., Koob, M.D. and Ranum, L.P.W. (1998) Incidence of dominant spinocerebellar and Friedreich triplet repeats among 361 ataxia families. *Neurology*, **51**, 1666-1671.

56. Matsumura, R., Futamura, N., Fujimoto, Y., Yanagimoto, S., Horikawa, H., Suzumura, A. and Takayanagi, T. (1997) Spinocerebellar ataxia type 6. Molecular and clinical features of 35 Japanese patients including one homozygous for the CAG repeat expansion. *Neurology*, **49**, 1238-1243.

57. Riess, O., Schols, L., Bottger, H., Nolte, D., Vieira-Saecker, A.M., Schimming, C., Kreuz, F., Macek, M., Jr., Krebsova, A., Macek, M.S., Klockgether, T., Zuhlke, C. and Laccone, F.A. (1997) SCA6 is caused by moderate CAG expansion in the alpha1A-voltage-dependent calcium channel gene. *Hum. Mol. Genet.*, **6**, 1289-1293.

58. Schols, L., Gispert, S., Vorgerd, M., Menezes Vieira-Saecker, A.M., Blanke, P., Auburger, G., Amoiridis, G., Meves, S., Epplen, J.T., Przuntek, H., Pulst, S.M. and Riess, O. (1997) Spinocerebellar ataxia type 2. Genotype and phenotype in German kindreds. *Arch. Neurol.*, **54**, 1073-1080.

59. Schols, L., Kruger, R., Amoiridis, G., Przuntek, H., Epplen, J.T. and Riess, O. (1998) Spinocerebellar ataxia type 6: genotype and phenotype in German kindreds. *J. of Neurol. Neurosurg. Psychiatry*, **64**, 67-73.

60. Shizuka, M., Watanabe, M., Ikeda, Y., Mizushima, K., Okamoto, K. and Shoji, M. (1998) Molecular analysis of a *de novo* mutation for spinocerebellar ataxia type 6 and (CAG)n repeat units in normal elder controls. *J. Neurol. Sci.*, **161**, 85-87.

61. Trottier, Y., Lutz, Y., Stevanin, G., Imbert, G., Devys, D., Cancel, G., Saudou, F., Weber, C., David, G., Tora, L., Agid, Y., Brice, A. and Mandel, J.-L. (1995) Polyglutamine expansion as a pathological epitope in Huntington's disease and four dominant cerebellar ataxias. *Nature*, **378**, 403-406.

62. Lorenzetti, D., Bohlega, S. and Zoghbi, H.Y. (1997) The expansion of the CAG repeat in ataxin-2 is a frequent cause of autosomal dominant spinocerebellar ataxia. *Neurology*, **49**, 1009-1013.

63. Schols, L., Amoiridis, G., Buttner, T., Przuntek, H., Epplen, J.T. and Riess, O. (1997) Autosomal dominant cerebellar ataxia: phenotypic differences in genetically defined subtypes? *Ann. Neurol.*, **42**, 924-932.

64. Stevanin, G., Durr, A., David, G., Didierjean, O., Cancel, G., Rivaud, S., Tourbah, A., Warter, J.M., Agid, Y. and Brice, A. (1997) Clinical and molecular features of spinocerebellar ataxia type 6. *Neurology*, **49**, 1243-1246.

65. Giunti, P., Sabbadini, G., Sweeney, M.G., Davis, M.B., Veneziano, L., Mantuano, E., Federico, A., Plasmati, R., Frontali, M. and Wood, N.W. (1998) The role of the SCA2 trinucleotide repeat expansion in 89 autosomal dominant cerebellar ataxia families. Frequency, clinical and genetic correlates. *Brain*, **121**, 459-467.

66. Watanabe, H., Tanaka, F., Matsumoto, M., Doyu, M., Ando, T., Mitsuma, T. and Sobue, G. (1998) Frequency analysis of autosomal dominant cerebellar ataxias in Japanese patients and clinical characterization of spinocerebellar ataxia type 6. *Clin. Genet.*, **53**, 13-19.

67. Jodice, C., Mantuano, E., Veneziano, L., Trettel, F., Sabbadini, G., Calandriello, L., Francia, A., Spadaro, M., Pierelli, F., Salvi, F., Ophoff, R.A., Frants, R.R. and Frontali, M. (1997) Episodic ataxia type 2 (EA2) and spinocerebellar ataxia type 6 (SCA6) due to CAG repeat expansion in the *CACNA1A* gene on chromosome 19p. *Hum. Mol. Genet.*, **6**, 1973-1978.

68. Geschwind, D.H., Perlman, S., Figueroa, K.P., Karrim, J., Baloh, R.W. and Pulst, S.M. (1997) Spinocerebellar ataxia type 6. Frequency of the mutation and genotype-phenotype correlations. *Neurology,* **49**, 1247-1251.

69. Ishikawa, K., Tanaka, H., Saito, M., Ohkoshi, N., Fujita, T., Yoshizawa, K., Ikeuchi, T., Watanabe, M., Hayashi, A., Takiyama, Y., Nishizawa, M., Nakano, I., Matsubayashi, K., Miwa, M., Shoji, S., Kanazawa, I., Tsuji, S. and Mizusawa, H. (1997) Japanese families with autosomal dominant pure cerebellar ataxia map to chromosome 19p13.1-p13.2 and are strongly associated with mild CAG expansions in the spinocerebellar ataxia type 6 gene in chromosome 19p13.1. *Am. J. Hum. Genet.,* **61**, 336-346.

70. Yabe, I., Sasaki, H., Matsuura, T., Takada, A., Wakisaka, A., Suzuki, Y., Fukazawa, T., Hamada, T., Oda, T., Ohnishi, A. and Tashiro, K. (1998) SCA6 mutation analysis in a large cohort of the Japanese patients with late-onset pure cerebellar ataxia. *J. Neurol. Sci.,* **156**, 89-95.

71. Greenberg, D.A. (1997) Calcium channels in neurological disease. *Ann. Neurol.,* **42**, 275-282.

72. Ophoff, R.A., Terwindt, G.M., Vergouwe, M.N., van Eijk, R., Oefner, P.J., Hoffman, S.M., Lamerdin, J.E., Mohrenweiser, H.W., Bulman, D.E., Ferrari, M., Haan, J., Lindhout, D., van Ommen, G.J., Hofker, M.H., Ferrari, M.D. and Frants, R.R. (1996) Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene *CACNL1A4. Cell,* **87**, 543-552.

73. Yue, Q., Jen, J.C., Nelson, S.F. and Baloh, R.W. (1997) Progressive ataxia due to a missense mutation in a calcium-channel gene. *Am. J. Hum. Genet.,* **61**, 1078-1087.

74. Ishikawa, K., Fujigasaki, H., Saegusa, H., Ohwada, K., Fujita, T., Iwamoto, H., Komatsuzaki, Y., Toru, S., Toriyama, H., Watanabe, M., Ohkoski, N., Shoji, S., Kanazawa, I., Tanabe, T. and Mizusawa, H. (1999) Abundant expression and cytoplasmic aggregations of alpha 1A-voltage-

dependent calcium channel protein associated with neurodegeneration in spinocerebellar ataxia type 6. *Hum. Mol. Genet.*, **8**, 1185-1193.

75. Matsuyama, Z., Kawakami, H., Maruyama, H., Izumi, Y., Komure, O., Udaka, F., Kameyama, M., Nishio, T., Kuroda, Y., Nishimura, M. and Nakamura, S. (1997) Molecular features of the CAG repeats of spinocerebellar ataxia 6 (SCA6). *Hum. Mol. Genet.*, **6**, 1283-1287.

76. Benton, C.S., Desilva, R., Rutledge, S.L., Bohlega, S., Ashizawa, T. and Zoghbi, H.Y. (1998) Molecular and clinical studies in SCA-7 define a broad clinical spectrum and the infantile phenotype. *Neurology*, **51**, 1081-1086.

77. Silveira, I., Coutinho, P., Maciel, P., Gaspar, C., Hayes, S., Dias, A., Guimaraes, J., Loureiro, L., Sequeiros, J. and Rouleau, G.A. (1998) Analysis of SCA1, DRPLA, MJD, SCA2, and SCA6 CAG repeats in 48 Portuguese ataxia families. *Am. J. Med. Genet.*, **81**, 134-138.

78. Sato, K., Kashihara, K., Okada, S., Ikeuchi, T., Tsuji, S., Shomori, T., Morimoto, K. and Hayabara, T. (1995) Does homozygosity advance the onset of dentatorubral-pallidoluysian atrophy? *Neurology*, **45**, 1934-1936.

79. Kremer, B., Goldberg, P., Andrew, S.E., Theilmann, J., Telenius, H., Zeisler, J., Squitieri, F., Lin, B., Bassett, A. and Almqvist, E. (1994) A worldwide study of the Huntington's disease mutation. The sensitivity and specificity of measuring CAG repeats. *N. Engl. J. Med.*, **330**, 1401-1406.

80. Takiyama, Y., Igarashi, S., Rogaeva, E.A., Endo, K., Rogaev, E.I., Tanaka, H., Sherrington, R., Sanpei, K., Liang, Y., Saito, M., Tsuda, T., Takano, H., Ikeda, M., Lin, C., Chi, H., Kennedy, J.L., Lang, A.E., Wherrett, J.R., Segawa, M., Nomura, Y., Yuasa, T., Weissenbach, J., Yoshida, M., Nishizawa, M., Kidd, K.K., Tsuji, S. and St. George-Hyslop, P.H. (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. *Hum. Mol. Genet.*, **4**, 1137-1146.

81. Johansson, J., Forsgren, L., Sandgren, O., Brice, A., Holmgren, G. and Holmberg, M. (1998) Expanded CAG repeats in Swedish spinocerebellar ataxia type 7 (SCA7) patients: effect of CAG repeat length on the clinical manifestation. *Hum. Mol. Genet.*, **7**, 171-176.

82. Stevanin, G., Giunti, P., David, G., Belal, S., Durr, A., Ruberg, M., Wood, N. and Brice, A. (1998) *De novo* expansion of intermediate alleles in spinocerebellar ataxia 7. *Hum. Mol. Genet.*, **7**, 1809-1813.

83. Gouw, L.G., Castaneda, M.A., McKenna, C.K., Digre, K.B., Pulst, S.M., Perlman, S., Lee, M.S., Gomez, C., Fischbeck, K., Gagnon, D., Storey, E., Bird, T., Jeri, F.R. and Ptacek, L.J. (1998) Analysis of the dynamic mutation in the SCA7 gene shows marked parental effects on CAG repeat transmission. *Hum. Mol. Genet.*, **7**, 525-532.

84. Giunti, P., Stevanin, G., Worth, P.F., David, G., Brice, A. and Wood, N.W. (1999) Molecular and clinical study of 18 families with ADCA type II: Evidence for genetic heterogeneity and De Novo mutation. *Am. J. Hum. Genet.*, **64**, 1594-1603.

85. Del-Favero, J., Krols, L., Michalik, A., Theuns, J., Lofgren, A., Goossens, D., Wehnert, A., Van den Bossche, D., Van Zand, K., Backhovens, H., van Regenmorter, N., Martin, J.J. and Van Broeckhoven, C. (1998) Molecular genetic analysis of autosomal dominant cerebellar ataxia with retinal degeneration (ADCA type II) caused by CAG triplet repeat expansion. *Hum. Mol. Genet.*, **7**, 177-186.

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

87. Perutz, M.F., Johnson, T., Suzuki, M. and Finch, J.T. (1994) Glutamine repeats as polar zippers: their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. USA*, **91**, 5355-5358.

88. Davies, S.W., Turmaine, M., Cozens, B.A., DiFiglia, M., Sharp, A.H., Ross, C.A., Scherzinger, E., Wanker, E.E., Mangiarini, L. and Bates, G.P. (1997) Formation of neuronal intranuclear inclusions underlies the

neurological dysfunction in mice transgenic for the HD mutation. *Cell*, **90**, 537-548.

89. Becher, M.W., Kotzuk, J.A., Sharp, A.H., Davies, S.W., Bates, G.P., Price, D.L. and Ross, C.A. (1998) Intranuclear neuronal inclusions in Huntington's disease and dentatorubral and pallidoluysian atrophy: correlation between the density of inclusions and *IT15* CAG triplet repeat length. *Neurobiol. Dis.*, **4**, 387-397.

90. Cooper, J.K., Schilling, G., Peters, M.F., Herring, W.J., Sharp, A.H., Kaminsky, Z., Masone, J., Khan, F.A., Delanoy, M., Borchelt, D.R., Dawson, V.L., Dawson, T.M. and Ross, C.A. (1998) Truncated N-terminal fragments of huntingtin with expanded glutamine repeats form nuclear and cytoplasmic aggregates in cell culture. *Hum. Mol. Genet.*, **7**, 783-790.

91. Igarashi, S., Koide, R., Shimohata, T., Yamada, M., Hayashi, Y., Takano, H., Date, H., Oyake, M., Sato, T., Sato, A., Egawa, S., Ikeuchi, T., Tanaka, H., Nakano, R., Tanaka, K., Hozumi, I., Inuzuka, T., Takahashi, H. and Tsuji, S. (1998) Suppression of aggregate formation and apoptosis by transglutaminase inhibitors in cells expressing truncated DRPLA protein with an expanded polyglutamine stretch. *Nat. Genet.*, **18**, 111-117.

92. Li, S.H. and Li, X.J. (1998) Aggregation of N-terminal huntingtin is dependent on the length of its glutamine repeats. *Hum. Mol. Genet.*, **7**, 777-782.

93. Li, M., Miwa, S., Kobayashi, Y., Merry, D.E., Yamamoto, M., Tanaka, F., Doyu, M., Hashizume, Y., Fischbeck, K.H. and Sobue, G. (1998) Nuclear inclusions of the androgen receptor protein in spinal and bulbar muscular atrophy. *Ann. Neurol.*, **44**, 249-254.

94. Lunkes, A. and Mandel, J.L. (1998) A cellular model that recapitulates major pathogenic steps of Huntington's disease. *Hum. Mol. Genet.*, **7**, 1355-1361.

95. Paulson, H.L., Perez, M.K., Trottier, Y., Trojanowski, J.Q., Subramony, S.H., Das, S.S., Vig, P., Mandel, J.L., Fischbeck, K.H. and Pittman, R.N.

(1997) Intranuclear inclusions of expanded polyglutamine protein in spinocerebellar ataxia type 3. *Neuron,* **19,** 333-344.

96. Schilling, G., Becher, M.W., Sharp, A.H., Jinnah, H.A., Duan, K., Kotzuk, J.A., Slunt, H.H., Ratovitski, T., Cooper, J.K., Jenkins, N.A., Copeland, N.G., Price, D.L., Ross, C.A. and Borchelt, D.R. (1999) Intranuclear inclusions and neuritic aggregates in transgenic mice expressing a mutant N-terminal fragment of huntingtin. *Hum. Mol. Genet.,* **8,** 397-407.

97. Duyckaerts, C., Durr, A., Cancel, G. and Brice, A. (1999) Nuclear inclusions in spinocerebellar ataxia type 1. *Acta Neuropathol. (Berl),* **97,** 201-207.

98. Holmberg, M., Duyckaerts, C., Durr, A., Cancel, G., Gourfinkel-An, I., Damier, P., Faucheux, B., Trottier, Y., Hirsch, E.C., Agid, Y. and Brice, A. (1998) Spinocerebellar ataxia type 7 (SCA7): a neurodegenerative disorder with neuronal intranuclear inclusions. *Hum. Mol. Genet.,* **7,** 913-918.

99. Merry, D.E., Kobayashi, Y., Bailey, C.K., Taye, A.A. and Fischbeck, K.H. (1998) Cleavage, aggregation and toxicity of the expanded androgen receptor in spinal and bulbar muscular atrophy. *Hum. Mol. Genet.,* **7,** 693-701.

100. Moulder, K.L., Onodera, O., Burke, J.R., Strittmatter, W.J. and Johnson, E.M., Jr. (1999) Generation of neuronal intranuclear inclusions by polyglutamine-GFP: analysis of inclusion clearance and toxicity as a function of polyglutamine length. *J. Neurosci.,* **19,** 705-715.

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

102. Chai, Y., Koppenhafer, S.L., Shoesmith, S.J., Perez, M.K. and Paulson, H.L. (1999) Evidence for proteasome involvement in polyglutamine

disease: localization to nuclear inclusions in SCA3/MJD and suppression of polyglutamine aggregation in vitro. *Hum. Mol. Genet.*, **8**, 673-682.

103. Huynh, D.P., Del Bigio, M.R., Ho, D.H. and Pulst, S.M. (1999) Expression of ataxin-2 in brains from normal individuals and patients with Alzheimer's disease and spinocerebellar ataxia 2. *Ann. Neurol.*, **45**, 232-241.

104. Klement, I.A., Skinner, P.J., Kaytor, M.D., Yi, H., Hersch, S.M., Clark, H.B., Zoghbi, H.Y. and Orr, H.T. (1998) Ataxin-1 nuclear localization and aggregation - role in polyglutamine-induced disease in SCA1 transgenic mice. *Cell*, **95**, 41-53.

105. Saudou, F., Finkbeiner, S., Devys, D. and Greenberg, M.E. (1998) Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions. *Cell*, **95**, 55-66.

106. Warrick, J.M., Paulson, H.L., Gray-Board, G.L., Bui, Q.T., Fischbeck, K.H., Pittman, R.N. and Bonini, N.M. (1998) Expanded polyglutamine protein forms nuclear inclusions and causes neural degeneration in *Drosophila*. *Cell*, **93**, 939-949.

107. Takano, H., Cancel, G., Ikeuchi, T., Lorenzetti, D., Mawad, R., Stevanin, G., Didierjean, O., Durr, A., Oyake, M., Shimohata, T., Sasaki, R., Koide, R., Igarashi, S., Hayashi, S., Takiyama, Y., Nishizawa, M., Tanaka, H., Zoghbi, H., Brice, A. and Tsuji, S. (1998) Close associations between prevalences of dominantly inherited spinocerebellar ataxias with CAG-repeat expansions and frequencies of large normal CAG alleles in Japanese and Caucasian populations. *Am. J. Hum. Genet.*, **63**, 1060-1066.

108. Gaspar, C., Lopes-Cendes, I., DeStefano, A.L., Maciel, P., Silveira, I., Coutinho, P., MacLeod, P., Sequeiros, J., Farrer, L.A. and Rouleau, G.A. (1996) Linkage disequilibrium analysis in Machado-Joseph disease patients of different ethnic origins. *Hum. Genet.*, **98**, 620-624.

109. Dichgans, M., Schols, L., Herzog, J., Stevanin, G., Weirich-Schwaiger, H., Rouleau, G., Burk, K., Klockgether, T., Zuhlke, C., Laccone, F., Riess, O.

and Gasser, T. (1999) Spinocerebellar ataxia type 6: evidence for a strong founder effect among German families. *Neurology*, **52**, 849-851.

110. Ramesar, R.S., Bardien, S., Beighton, P. and Bryer, A. (1997) Expanded CAG repeats in spinocerebellar ataxia (SCA1) segregate with distinct haplotypes in South african families. *Hum. Genet.*, **100**, 131-137.

111. Stevanin, G., Lebre, A.S., Mathieux, C., Cancel, G., Abbas, N., Didierjean, O., Durr, A., Trottier, Y., Agid, Y. and 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. *Am. J. Hum. Genet.*, **60**, 1548-1552.

112. Geschwind, D.H., Perlman, S., Figueroa, C.P., Treiman, L.J. and Pulst, S.M. (1997) The prevalence and wide clinical spectrum of the spinocerebellar ataxia type 2 trinucleotide repeat in patients with autosomal dominant cerebellar ataxia. *Am. J. Hum. Genet.*, **60**, 842-850.

113. Frontali, M., Sabbadini, G., Novelletto, A., Jodice, C., Naso, F., Spadaro, M., Giunti, P., Jacopini, A.G., Veneziano, L., Mantuano, E., Malaspina, P., Ulizzi, L., Brice, A., Durr, A. and Terrenato, L. (1996) Genetic fitness in Huntington's disease and spinocerebellar ataxia 1: a population genetics model for CAG repeat expansions. *Ann. Hum. Genet.*, **60**, 423-435.

114. Yanagisawa, H., Fujii, K., Nagafuchi, S., Nakahori, Y., Nakagome, Y., Akane, A., Nakamura, M., Sano, A., Komure, O., Kondo, I., Jin, D.K., Sorensen, S.A., Potter, N.T., Young, S.R., Nakamura, K., Nukina, N., Nagao, Y., Tadokoro, K., Okuyama, T., Miyashita, T., Inoue, T., Kanazawa, I. and Yamada, M. (1996) A unique origin and multistep process for the generation of expanded DRPLA triplet repeats. *Hum. Mol. Genet.*, **5**, 373-379.

115. Myers, R.H., MacDonald, M.E., Koroshetz, W.J., Duyao, M.P., Ambrose, C.M., Taylor, S.A., Barnes, G., Srinidhi, J., Lin, C.S., Whaley, W.L., Lazzarini, A.M., Schwarz, M., Wolff, G., Bird, E.D., Vonsattel, J.P. and

Gusella, J.F. (1993) *De novo* expansion of a (CAG)n repeat in sporadic Huntington's disease. *Nat. Genet.*, **5**, 168-173.

116. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

117. Lopes-Cendes, I., Teive, H.G., Calcagnotto, M.E., Da Costa, J.C., Cardoso, F., Viana, E., Maciel, J.A., Radvany, J., Arruda, W.O., Trevisol-Bittencourt, P.C., Rosa Neto, P., Silveira, I., Steiner, C.E., Pinto Junior, W., Santos, A.S., Correa Neto, Y., Werneck, L.C., Araujo, A.Q., Carakushansky, G., Mello, L.R., Jardim, L.B. and Rouleau, G.A. (1997) Frequency of the different mutations causing spinocerebellar ataxia (SCA1, SCA2, MJD/SCA3 and DRPLA) in a large group of Brazilian patients. *Arq. Neuropsiquiatr.*, **55**, 519-529.

118. Sahba, S., Nechiporuk, A., Figueroa, K.P., Nechiporuk, T. and Pulst, S.M. (1998) Genomic structure of the human gene for spinocerebellar ataxia type 2 (SCA2) on chromosome 12q24.1. *Genomics*, **47**, 359-364.

119. Didierjean, O., Cancel, G., Stevanin, G., Durr, A., Burk, K., Benomar, A., Lezin, A., Belal, S., Abada-Bendid, M., Klockgether, T. and Brice, A. (1999) Linkage disequilibrium at the SCA2 locus. *J. Med. Genet.*, **36**, 415-417.

120. Ranum, L.P., Chung, M.Y., Banfi, S., Bryer, A., Schut, L.J., Ramesar, R., Duvick, L.A., McCall, A., Subramony, S.H., Goldfarb, L., Gomez, C., Sandkuijl, L.A., Orr, H.T. and Zoghbi, H.Y. (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.

121. Rubinsztein, D.C., Leggo, J., Chiano, M., Dodge, A., Norbury, G., Rosser, E. and Craufurd, D. (1997) Genotypes at the GluR6 kainate receptor locus are associated with variation in the age of onset of Huntington disease. *Proc. Natl. Acad. Sci. USA*, **94**, 3872-3876.

122. DeStefano, A.L., Cupples, L.A., Maciel, P., Gaspar, C., Radvany, J., Dawson, D.M., Sudarsky, L., Corwin, L., Coutinho, P., MacLeod, P., Sequeiros, J., Rouleau, G.A. and Farrer, L.A. (1996) A familial factor independent of CAG repeat length influences age at onset of Machado-Joseph disease. *Am. J. Hum. Genet.*, **59**, 119-127.

123. Paulson, H.L. (1999) Protein fate in neurodegenerative proteinopathies: polyglutamine diseases join the (mis)fold. *Am. J. Hum. Genet.*, **64**, 339-345.

124. Davies, S.W., Beardsall, K., Turmaine, M., DiFiglia, M., Aronin, N. and Bates, G.P. (1998) Are neuronal intranuclear inclusions the common neuropathology of triplet-repeat disorders with polyglutamine-repeat expansions? *Lancet*, **351**, 131-133.

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

126. Matilla, A., Koshy, B.T., Cummings, C.J., Isobe, T., Orr, H.T. and Zoghbi, H.Y. (1997) The cerebellar leucine-rich acidic nuclear protein interacts with ataxin-1. *Nature*, **389**, 974-978.

127. Joober, R., Benkelfat, C., Lafreniere, R., Lal, S., Bloom, D., Labelle, A., Lalonde, P., Turecki, G., Alda, M., Palmour, R. and Rouleau, G.A. (1999) Correlation between *hGT1* gene CAG repeat length and severity of schizophrenia. *Am. J. Med. Genet., in press* .

128. Li, S.H., McInnis, M.G., Margolis, R.L., Antonarakis, S.E. and Ross, C.A. (1993) Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms. *Genomics*, **16**, 572-579.

129. Imai, Y., Suzuki, Y., Matsui, T., Tohyama, M., Wanaka, A. and Takagi, T. (1995) Cloning of a retinoic acid-induced gene, *GT1*, in the embryonal carcinoma cell line P19: neuron-specific expression in the mouse brain. *Brain Res. Mol. Brain Res.*, **31**, 1-9.

130. Rubinsztein, D.C., Wyttenbach, A. and Rankin, J. (1999) Intracellular inclusions, pathological markers in diseases caused by expanded polyglutamine tracts? *J. Med. Genet.*, **36**, 265-270.