Am. J. Hum. Genet. 54:959-965, 1994

Effect of Trinucleotide Repeat Length and Parental Sex on Phenotypic Variation in Spinocerebellar Ataxia I

Carla Jodice,* Patrizia Malaspina,* Francesca Persichetti,* Andrea Novelletto,* Maria Spadaro,† Paola Giunti,†,§ Cristoforo Morocutti,† Luciano Terrenato,* Anita E. Harding,§ and Marina Frontali‡

*Dipartimento di Biologia, Università "Tor Vergata," †Istituto di Clinica delle Malattie Nervose e Mentali, Università "La Sapienza," and ‡Istituto di Medicina Sperimentale, CNR, Rome; and §Institute of Neurology, London

Summary

Trinucleotide repeat expansion has been found in 64 subjects from 19 families: 57 patients with SCA1 and 7 subjects predicted, by haplotype analysis, to carry the mutation. Comparison with a large set of normal chromosomes shows two distinct distributions, with a much wider variation among expanded chromosomes. The sex of transmitting parent plays a major role in the size distribution of expanded alleles, those with >54 repeats being transmitted by affected fathers exclusively. Our data suggest that alleles with >54 repeats have a reduced chance of survival; these appear to be replaced in each generation by further expansion of alleles in the low- to medium-expanded repeat range, preferentially in male transmissions. Detailed clinical follow-up of a subset of our patients demonstrates significant relationships between increasing repeat number on expanded chromosomes and earlier age at onset, faster progression of the disease, and earlier age at death.

Introduction

The autosomal dominant cerebellar ataxias (ADCAs) constitute a group of late-onset neurodegenerative disorders primarily affecting the cerebellum but often associated with other neurological abnormalities. ADCA type 1 (ADCA1) is characterized by progressive ataxia, often with additional ophthalmoplegia, extrapyramidal features, and peripheral neuropathies (Harding 1982). Genetic heterogeneity within this disorder has been recently shown: at least three genes responsible for this phenotype have been mapped on 6p (spinocerebellar ataxia; SCA1) (Morton et al. 1980), 12q (SCA2) (Gispert et al. 1993), and 14q (SCA3, or Machado-Joseph disease) (Takiyama et al. 1993). By analyzing eight large kindreds in which the disease was tightly linked to the genetic marker D6S89 and was thus classified as SCA1 (Ranum et al. 1991; Zoghbi et al. 1991), we showed that the region that is a candidate to contain the mutation was delimited by the genetic loci D6S89 and D6S274 (Jodice et al. 1993; Kwiatkowski et al. 1993). Within this region Orr et al. (1993) identified a repeated CAG trinucleotide sequence that is selectively expanded in SCA1 patients. They also showed (Chung et al. 1993) that 98% of normal chromosomes contain >22 repeats and one or more CAT trinucleotide(s) interrupting the CAG stretch. However, expanded chromosomes invariably have uninterrupted CAG repeated sequences, a structure found only in normal chromosomes with <22 repeats (2% of chromosomes). Moreover, expanded chromosomes showed marked instability at transmission, with a greater tendency toward increase in size when transmitted by affected fathers (Chung et al. 1993). Earlier age at onset in the affected offspring of males with ADCA1 has been observed elsewhere (Harding 1981). Finally, a strong correlation was found (Orr et al. 1993) between the repeat number on expanded chromosomes and age at onset of SCA1 symptoms. Here we show that the mutational basis of SCA1 in Italian, British, and Malaysian patients is the same as that described by Orr et al. (1993).

Analysis of the largest sample of normal chromosomes (including CEPH families) reported so far con-

Received December 20, 1993; accepted for publication February 1, 1994.

Address for correspondence and reprints: Dr. Carla Jodice, Department of Biology, "Tor Vergata" University, Via della Ricerca Scientifica, 00133 Rome, Italy.

0002-9297/94/5406-0004$02.00

firms that the distributions of repeat numbers on normal and affected chromosomes are clearly nonoverlapping and that normal chromosomes are stable, even with a high normal number of repeats, as long as an interrupted CAG stretch is present. The sex of the transmitting parent plays a major role in the size distribution of expanded alleles, those with >54 repeats being transmitted by affected fathers exclusively. Finally, an analysis of the relationship between the length of trinucleotide expansion and phenotypic aspects of SCA1 shows that repeat length appears to be related not only to age at onset of the disease but also to rate of disease progression. Some of these correlations depend dramatically on whether the affected chromosome is inherited from the father or the mother.

Subjects and Methods

Subjects

The present study includes 19 SCA1 families. Six of these (Ns, Ps, Pu, Sa, Mr, and Lz) were already defined as SCA1 on the basis of linkage analysis (Jodice et al. 1993).

Ten further families—two Italian, seven British, and one Malaysian—showing the clinical features of ADCA1 are also reported here. Only one was large enough for linkage analysis (Shrimpton et al. 1993); the others were classified as SCA1 on the basis of the presence of an expanded number of repeats in affected members.

Overall, 172 family members were typed, including 57 affected, 7 subjects at high risk as determined by linkage analysis, and 108 unaffected relatives. Subjects <18 years old were not sampled unless already affected. Thirty-four additional normal families (456 individuals), derived from the CEPH reference panel, were also studied.

Methods

As a measure of a patient's disability, a modified Kurtzke's Disability Status Scale (DSS) adapted for SCA1 clinical features was used (Spadaro et al. 1992). The score attained on the DSS was used to generate two different indexes. The first index was the ratio of DSS recorded at the last neurological examination to the disease duration (in years) at that time. This index was calculated for patients seen one or more times. The second index was obtained only for patients followed up for more than 2 years, by calculating the difference in DSS score between the first and last examination, divided by the number of years elapsed between the two evaluations.

Analysis of the CAG-containing segment was performed by PCR as described elsewhere (Orr et al. 1993). PCR products were sized by comparison with M13 sequencing ladders and a subject with known genotype (provided by H. Orr). Affected subjects from the same pedigrees were run on the same gels.

The presence of one or more CAT trinucleotides within the repeated CAG sequence was detected by *Sfa*NI restriction of the unlabeled PCR product and separation of fragments on 2.5% agarose gels (Chung et al. 1993). Genomic sequencing was performed by the method of Murray (1989), after purification of the relevant band from agarose gels (Magic PCR Prep Promega).

Results

Distribution and Structure of Repeat Units

The repeat-size distribution for the larger allele observed in the 57 affected patients and in 7 asymptomatic subjects known to carry SCA1 by linkage analysis is shown in figure 1*a*. Of 64 SCA1 chromosomes, 49 (77%) were already considered to carry the SCA1 mutation, on the basis of linkage studies. Families with different geographic origin, in which SCA1 is transmitted along with different haplotypes, most likely reflecting independent mutational events, appear to have the same degree of expansion, on average. The mean repeat value of all 64 expanded chromosomes was 52.4 ± 4.5, with a range of 46–66. Figure 1*b* shows the distribution of repeat sizes on 133 independent normal chromosomes from the families that we studied. This sample includes the shorter (normal) alleles of the affected or at-high-risk subjects. The mean is 30.5 ± 1.4, with a range of 27–37. This distribution closely matches the distribution of repeat number in the larger sample, of 232 parental chromosomes derived from the CEPH reference panel (fig. 1*c*). This latter has a mean of 30.6 ± 1.6 with a range of 26–36. Genotype distribution in this group fits the Hardy-Weinberg equilibrium.

In order to investigate the internal structure of the repeated CAG segment, we analyzed *Sfa*NI digestion of the PCR product obtained from one affected subject of each family. When compared with undigested products, in all cases the larger allele was not altered in length, while the smaller allele was cut into two smaller bands, of ~130 and ~100 bp. This indicates that the expanded alleles always have the structure of an uninterrupted $(CAG)_n$ sequence but that the normal alleles contain one or more CAT trinucleotides.

We observed *Sfa*NI restriction of the longest normal

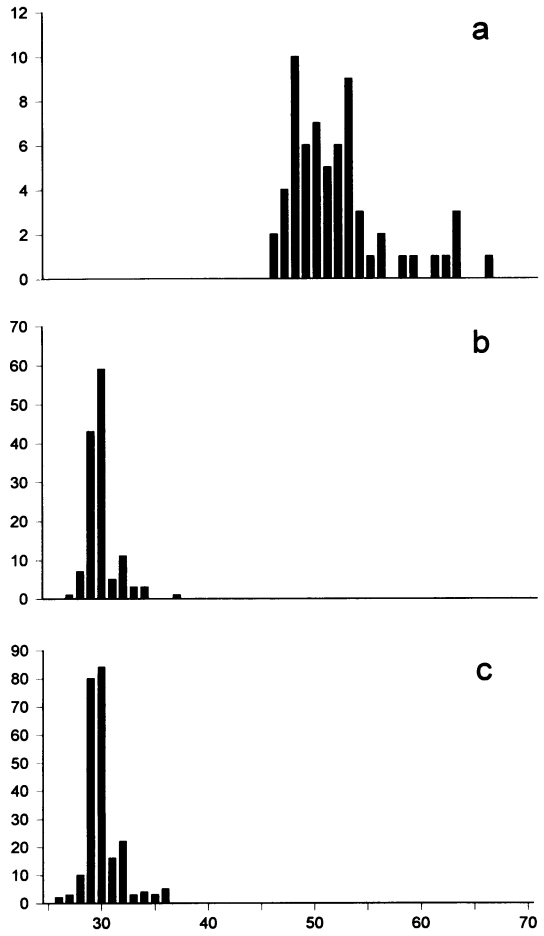

Figure 1 Histograms of repeat numbers found on different sets of chromosomes: chromosomes carrying expansion in 57 SCA1 patients and in 7 high-risk subjects (*a*), 133 independent normal chromosomes from families segregating SCA1 (*b*), and 232 independent normal chromosomes from the parental generation of 34 CEPH reference families (*c*).

allele so far described, which carries 37 repeat units. On genomic sequencing, the structure of this allele was found to be $(CAG)_{17}CAT\ CAG\ CAT(CAG)_{17}$. This allele was found in seven healthy subjects from the Pu pedigree, presently aged 26–93 years and separated by 12 meiotic events, always associated with a haplotype not carrying SCA1.

Parental Sex Effect at Transmission

The sex of the affected parent was known for all affected and at-risk subjects. When expanded alleles are partitioned according to repeat number and sex of the transmitting parent (table 1, row 1), a striking result emerges, alleles with >54 repeats (17% of the sample) being transmitted only by affected fathers. On the other hand, alleles with 46–54 repeats are transmitted by affected fathers and mothers in equal proportions. The unbalanced distribution of repeat number and parental origin is statistically significant (Fisher test, $P = .001$).

Table 1 (rows 2 and 3) also shows that within each class a normal sex ratio of offspring is observed, i.e., affected and high-risk subjects with repeat numbers either ≤54 or >54 are males and females in a proportion approaching the expected 50:50. If a steady-state distribution of repeat numbers is assumed to persist through the generations, this raises the question as to why affected females transmitting alleles with >54 repeats are lacking, while females receiving >54 repeats do exist. In view of the earlier onset of the disease in subjects with greater expansions (see below), impaired biological fitness is likely. Subjects of both sexes who have >54 repeats, when compared with those carrying ≤54 repeats, have fewer children, on average (*t*-test, $P < .05$) (table 1, row 4). This is unlikely to be accounted for by their younger age, since they already exhibit the same (if not greater) disease severity (table 1, rows 5 and 6). We could directly examine 12 transmissions of expanded chromosomes from affected mothers (8 transmissions) and fathers (4 transmissions). Among transmitting mothers (range 49–53 repeats), increases (two cases), null changes (two cases), and decreases (four cases) were observed, with an average variation of −0.5 units. On the other hand, among transmitting fathers (range 47–53 repeats), only a single null change and three increases were observed, with an average variation of +1.75 units. We also examined the instability of expanded alleles by taking into account the range observed among affected sibs. Three sibships born to affected mothers show ranges of 0, 1, and 3 repeat units, whereas six sibships born to affected fathers show values that have a range of 1–11 repeat units.

As a possible variable influencing instability of repeat number on expanded chromosomes, we considered the age of the transmitting parent at the time of birth of the examined patients. Among 53 subjects with available data, there was a weakly but significantly positive correlation ($r = .28$, $P < .05$). In order to see whether an advanced paternal age could be responsible for the production of SCA1 alleles >54 repeats, we divided the three groups of patients reported in table 1, according to the affected parent's age at the patient's birth (table 1, rows 7 and 8). The median ages of transmitting fathers and mothers at birth of carriers were 32 and 28

Table 1

Parental Sex Effect on Repeat Number and Associated Characteristics

	≤54 Repeats		>54 Repeats	
	Paternal Inheritance	Maternal Inheritance	Paternal Inheritance	Maternal Inheritance
Total no. of cases	26	27	11	0
Sex of affected subjects:				
M	14 (54±10)	11 (41±9)	6 (55±15)	...
F	12 (46±10)	16 (59±9)	5 (45±15)	...
Average no. of children	2.12 ± 2.31	1.52 ± 1.45	0.70 ± 1.34	...
Average age (years)[a]	41.7 ± 10.6	43.0 ± 11.2	31.9 ± 9.3	...
Average DSS	5.85 ± 2.41	4.80 ± 2.46	6.37 ± 2.20	...
Parental age (years) at birth of proband:				
Above median[b]	10 (50±11)	12 (52±10)	6 (60±15)	...
Below median[b]	10 (50±11)	11 (48±10)	4 (40±15)	...

NOTE.—Data in parentheses are percent ± standard error.

[a] Age at time of study or age at death.

[b] Median ages of affected parents at birth of probands were 32 and 28 years for fathers and mothers, respectively.

years, respectively. When these values are used as thresholds, the subgroups thus obtained are expected to show a 1:1 proportion. Rows 7 and 8 of table 1 show no evidence that the patients receiving SCA1 alleles with >54 repeats have affected parents who are older than expected. Therefore, advanced parental age does not seem to play a major role in large increases in repeat length.

Finally, instability of the CAG repeated segment in normal chromosomes was also examined. No instances of variation were found either in families segregating SCA1 or in the much larger sample of 576 meioses of 34 CEPH reference families.

Correlations between Repeat Number and Clinical Parameters

Table 2 shows the data illustrating the linear relationships between the number of repeat units on expanded chromosomes and different parameters of the SCA1 phenotype. The data were obtained from different subsets of affected subjects and are divided according to the parental origin of the SCA1 chromosome.

Age at onset (range 15–51 years) shows a striking correlation with repeat number (fig. 2). The linear and exponential (not shown) regression models fit the data similarly, and this result seems to be qualitatively different from that reported by Orr et al. (1993), mainly because of the lack, in our series, of subjects with extremely early and late ages at onset. The 95% confidence limits of the regression line define an interval of ±11.8 years for the expected age at onset, on the basis of repeat number (fig. 2). We also investigated the effect that repeat length of normal chromosomes carried by affected subjects had on age at onset, and we found no correlation, either in patients receiving SCA1 from their mothers or in patients receiving SCA1 from their fathers ($r = -.15$ and $r = .13$, respectively).

Thirty-six patients were scored on the DSS at least once. The mean score was 6.03 ± 2.37, with a median of 6. Two different indexes were generated from the DSS score to quantify disease progression (see Subjects and Methods). The first one is of more general applicability, since patients undergoing a single clinical examination can be scored. This index appears to be highly correlated with the repeat number on the affected subjects' expanded chromosomes (table 2). When the entire sample is divided according to the origin of the SCA1 chromosome, only the group of patients inheriting the disease from the father shows such a high correlation coefficient, because of the subgroup of subjects (six) with >59 repeats, whose progression in DSS was >0.7 points/year (mean 0.60±0.18) in all cases.

The second index is more precise, relying on an objective evaluation of both the initial DSS score and the interval between scores. This index also shows a high correlation with repeat number on expanded chromosomes, a correlation that once again is mainly contributed by cases with paternal inheritance.

Table 2

Relationship between Repeat Number on SCA1 Chromosomes and Different Parameters Describing SCA1 Phenotype

Variable (*n*)	Regression Coefficient (Standard error)	*t*	Correlation coefficient	*P*
Age at onset:				
Paternal inheritance (31)	−1.22 (.18)	−6.63	−.78	⩽.001
Maternal inheritance (24)	−2.31 (.56)	−4.12	−.66	<.001
Overall (55)	−1.37 (.17)	−8.08	−.74	⩽.001
Progression (DSS/disease duration [years]):				
Paternal inheritance (21)	.026 (.005)	5.05	+.76	<.001
Maternal inheritance (15)	−.001 (.022)	−.01	−.02	n.s.
Overall (36)	.023 (.005)	4.56	+.62	<.001
Progression (DSS/years between first and last examination):				
Paternal inheritance (7)	.031 (.011)	2.89	+.79	<.05
Maternal inheritance (7)	.035 (.027)	1.29	+.50	n.s.
Overall (14)	.031 (.009)	3.45	+.71	<.01
Total disease duration (years):				
Paternal inheritance (7)	−.480 (.173)	−2.77	−.78	<.05
Maternal inheritance (2)	n.a. (n.a.)	n.a.	n.a.	n.a.
Overall (9)	−.406 (.209)	−1.94	−.59	n.s.

NOTE.—n.s. = not significant; and n.a. = not applicable.

A further confirmation of the relationship between disease progression and repeat expansion has been obtained in nine patients who died during the course of the study. The overall duration of the disease (table 2) appears to be fairly well correlated ($P < .05$) with the number of repeats in the seven subjects inheriting their SCA1 chromosomes from their fathers. The negative correlation is not significant in the whole sample.

Discussion

The molecular basis of SCA1, an expansion of a tandemly repeated CAG trinucleotide (Orr et al. 1993), together with the typical features of an autosomal dominant neurodegenerative disorder of the central nervous system, establishes remarkable similarities among SCA1, spinal and bulbar muscular atrophy (SBMA) (La Spada et al. 1991), and Huntington disease (HD) (Huntington's Disease Collaborative Research Group 1993). We have confirmed the mutational basis of SCA1 in six large Italian pedigrees already shown, by linkage analysis, to carry the disease (Jodice et al. 1993). The same expansion was found in 13 other families who had the typical clinical features of the disease but who were not suitable for linkage studies.

By typing the largest number of independent normal chromosomes so far, we have shown that the distributions of repeat number on normal and SCA1 chromosomes are widely separated, with a gap of 9 units. The assay for SCA1 in the presence of an ADCA phenotype can therefore be considered reliable whenever a PCR product corresponding to >46 repeats is found. On the other hand, we cannot rule out SCA1 in other ADCA patients who do not have expansion, since mutations of other type may affect the same gene.

At the upper end of the normal distribution, we observed a chromosome with 37 repeats, a size not in the bimodal distribution reported by Orr et al. (1993). As do most normal alleles, this has an interrupted structure. The ages of the carriers of this allele (all of whom were healthy), as well as the associated haplotype, exclude its involvement in the disorder. Moreover, this chromosome proves to be stable over the transmission events linking the carrier subjects.

Chung et al. (1993) showed that, in contrast with SBMA and HD, the great majority of normal alleles at the SCA1 locus have one or more CAT trinucleotides interrupting the CAG sequence, whereas SCA1 expanded alleles have a continuous $(CAG)_n$ sequence. We have detected CAT trinucleotides in the sequence of all tested normal alleles (21) but none of the expanded ones. It is notable that we did not observe the short normal alleles reported, by Chung et al. (1993), with an uninterrupted $(CAG)_n$ structure of 19–21 repeats, in either SCA1 or CEPH families (lowest limit was 26). These authors hypothesized that interrupted CAG se-

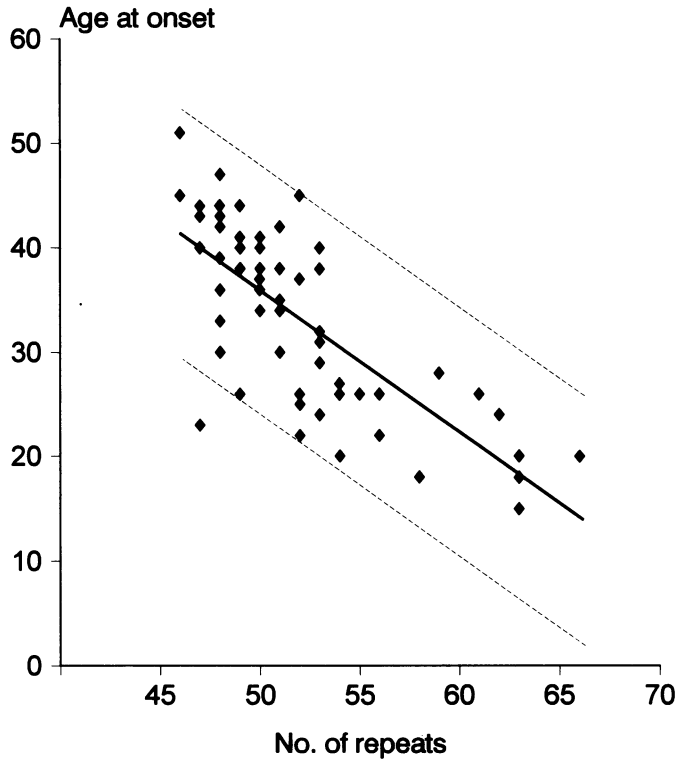

Figure 2 Scattergram of age at onset of SCA1, vs. repeat number on expanded chromosomes. The solid line denotes the interpolated regression line (for parameters, see table 2), and the dotted lines denote 95% confidence limits of the regression.

quences might be intrinsically more stable than continuous ones. This can be supported by the low variance in the distribution of normal alleles in SCA1. Moreover, no instances of instability of normal alleles at the SCA1 locus were detected, despite the screening of >600 transmission events.

Our data show a highly unbalanced transmission of expanded alleles, according to the sex of the affected parent. Alleles with >54 repeat units (17% of the sample) were only transmitted by males, but they were received by males and females in equal proportions. Our results point toward reduced biological fitness of patients carrying such chromosomes. Therefore, we suggest a model in which the selective loss of the largest alleles is counterbalanced by large increases in size, preferentially occurring in male transmissions, envisaging a mutation/selection balance for this mutation. This is supported both by our data and by the larger series of Chung et al. (1993), since both studies show a tendency toward an increase of the length of the expanded alleles, in paternal transmissions. However, a small proportion of females in both series transmitted alleles that contracted slightly, by ≤6 repeats.

Several quantifiable disease parameters were correlated with the repeat number on expanded alleles. An effect of the sex of the transmitting parent was apparent, with higher correlation coefficients among subjects inheriting their SCA1 allele from their fathers. The underlying difference between subjects with maternal versus paternal inheritance is that SCA1 alleles had a 46–54-repeat range in the former and a 47–66-repeat range in the latter. This is paralleled by the finding that patients with the more extreme presentation of the disease were present in the second group only. This subset of data pairs enhances correlation figures, and the biological basis of the differences in correlations is thus to be considered the greater increases of the repeat size occurring in paternal transmissions.

In agreement with Orr et al. (1993), we note that repeat number on expanded chromosomes could explain approximately two-thirds of the variation in age at onset in our series. However, repeat length alone does not allow a precise prediction of age at onset for testing purposes. We also analyzed three additional parameters assessing disease progression, and we found them too to be remarkably correlated with repeat number. Our

data show that a higher number of repeats is predictive of a more extreme SCA1 phenotype, with both earlier onset and faster progression toward disability and early death. This suggests that the mechanisms leading from CAG expansion to SCA1 phenotype have cumulative effects after the first appearance of symptoms. One possible explanation for this is provided by the model described by Green (1993), in which the products of poly-CAG translation act as a substrate for unscheduled transglutamination, eventually resulting in cell damage.

Acknowledgments

This work was partially supported by "TELETHON Italia" (no. 358), Progetto Finalizzato Ingegneria Genetica (support to M.F.), C.N.R. contracts 93.00031.PF99 (to A.N.) and 92.03000.CT26 and 93.01093.PF70 (both to L.T.), and grants from Associazone Italiana Sindromi Atassiche and the Friedreich's Ataxia Group (to P.G. and A.E.H.). CEPH reference families were used within the framework of the EUROGEM Project.

References

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

Gispert S, Twells R, Oroczo G, Brice A, Weber J, Heredero L, Scheufler K, et al (1993) Chromosomal assignment of the second locus for autosomal dominant cerebellar ataxia (SCA2) to chromosome 12q23-24.1. Nature Genet 4:295–299

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

Harding AE (1981) Genetic aspects of autosomal dominant late onset cerebellar ataxia. J Med Genet 18:436–441

——— (1982) The clinical features and classification of the late onset autosomal dominant cerebellar ataxias. Brain 105:1–28

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

Jodice C, Frontali M, Persichetti F, Novelletto A, Pandolfo M, Spadaro M, Giunti P, et al (1993) The gene for spinal cerebellar ataxia 1 (SCA1) is flanked by two closely linked highly polymorphic microsatellite loci. Hum Mol Genet 2:1383–1387

Kwiatkowski TJ Jr, Orr HT, Banfi S, McCall AE, Jodice C, Persichetti F, Novelletto A, et al (1993) The gene for autosomal dominant spinocerebellar ataxia (SCA1) maps centromeric to D6S89 and shows no recombination, in nine large kindreds, with a dinucleotide repeat at the AM10 locus. Am J Hum Genet 53:391–400

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

Morton NE, Lalouel JM, Jackson JF, Currier RD, Yee S (1980) Linkage studies in spinocerebellar ataxia (SCA). Am J Med Genet 6:251–257

Murray V (1989) Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res 17:8889

Orr HT, Chung M, Banfi S, Kwiatkowski TJ Jr, Servadio A, Beaudet AL, McCall AE, et al (1993) Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. Nature Genet 4:221–226

Ranum PWL, Duvick LA, Rich SS, Schut LJ, Litt M, Orr HT (1991) Localization of the autosomal dominant HLA-linked spinocerebellar ataxia (SCA1) locus, in two kindreds, within an 8-cM subregion of chromosome 6p. Am J Hum Genet 49:31–41

Shrimpton AE, Davidson R, MacDonald N, Brock DJH (1993) Presymptomatic test for autosomal dominant spinocerebellar ataxia type I. J Med Genet 30:616–617

Spadaro M, Giunti P, Lulli P, Frontali M, Jodice C, Cappellacci S, Morellini M, et al (1992) HLA-linked spinocerebellar ataxia: a clinical and genetic study of large Italian kindreds. Acta Neurol Scand 85:257–265

Takiyama J, Nishizawa M, Tanaka H, Kawashima S, Sakamoto H, Karube Y, Shimazaki H, et al (1993) The gene for Machado-Joseph disease maps to human chromosome 14q. Nature Genet 4:300–304

Zoghbi HY, Jodice C, Sandkuijl LA, Kwiatkowski TJ Jr, McCall AE, Huntoon SA, Lulli P, et al (1991) The gene for autosomal dominant spinocerebellar ataxia (SCA1) maps telomeric to the HLA complex and is closely linked to the D6S89 locus in three large kindreds. Am J Hum Genet 49:23–30