European Journal of Human Genetics (1999) 7, 841–845

http://www.stockton-press.co.uk/ejhg

SHORT REPORT

A common disease haplotype segregating in spinocerebellar ataxia 2 (SCA2) pedigrees of diverse ethnic origin

Joanna Pang[1], Rebecca Allotey[1], Noshir Wadia[2], Hidenao Sasaki[3], Laurence Bindoff[4] and Susan Chamberlain[1]

[1]*Hereditary Ataxia Research Group, Molecular Genetics, Division of Biomedical Sciences, Imperial College of Science, Technology and Medicine, London*
[2]*Department of Neurology, Jaslok Hospital and Research Centre, Bombay, India*
[3]*Department of Neurology, Hokkaido University School of Medicine, Ki ta-ku, Sapporo, Japan*
[4]*Department of Neurology, Middlesbrough General Hospital, Middlesbrough, UK*

The identification of a CAG trinucleotide repeat expansion, located within the coding sequence of the *ataxin-2* gene, as the mutation underlying spinocerebellar ataxia 2 (SCA2) has facilitated direct investigation of pedigrees previously excluded from linkage analysis due to insufficient size or pedigree structure. We have previously described the identification of the ancestral disease haplotype segregating in the Cuban founder population used to assign the disease locus to chromosome 12q23–24.1. We now report evidence for the segregation of the identical core haplotype in pedigrees of diverse ethnic origin from India, Japan and England, established by the analysis of the loci D12S1672 and D12S1333 located 20 kb proximal and 200 kb distal to the triplet repeat motif respectively. Interpretation of this data is suggestive that for these pedigrees at least, the mutation has arisen on a single ancestral or predisposing chromosome.

Keywords: spinocerebellar ataxia 2; common haplotype

Introduction

The autosomal dominant cerebellar ataxias (ADCAs) are a clinically heterogeneous group of neurodegenerative disorders, characterised by a progressive deterioration in balance and co-ordination resulting from premature neuronal loss in the cerebellum and the inferior olivary and pontine nuclei, with degeneration of the spinal cord. Genetic heterogeneity has been demonstrated with the identification of eight disease loci. Five of the disease genes have now been cloned – spinocerebellar ataxia 1 (SCA1),[1] SCA2,[2–4] Machado Joseph disease/SCA3,[5] SCA6[6] and SCA7[7] – and expansion of CAG trinucleotide repeat motifs located within the respective coding sequences identified.

Spinocerebellar ataxia 2 has been reported in most populations and appears to account for a substantial proportion of patients with ADCA worldwide. Since the cloning of the disease gene, *ataxin-2*, direct testing for the mutation has greatly enhanced the investigation of families previously excluded from linkage analysis. These studies have facilitated critical analysis of the

Correspondence: Dr Joanna Pang, Molecular Genetics, Division of Biomedical Science, Imperial College of Science, Technology and Medicine, South Kensington, London, SW7 2AZ, UK. Tel: 0171 594 3235; Fax: 0171 594 3015; E-mail: jpang@hgmp.mrc.ac.uk
Current address: Department of Neurology, Haukeland Sykehus, 5021, Bergen, Norway
Received 24 February 1999; revised 8 June 1999; accepted 14 June 1999

core phenotype and an assessment of the incidence of the diverse clinical features often associated with this disorder,[8,9] including correlation with the size of the expanded allele.

In common with other diseases caused by unstable triplet repeat motifs, efforts are being made to characterise the mechanistic basis and/or genetic factors which permit (or fail to correct) the pathological expansion underlying SCA2. The existence of chromosomes which predispose to instability has been proposed for a number of these disorders, including Huntington's disease[10] and myotonic dystrophy.[11] Haplotype analysis using microsatellite loci in close physical proximity to the respective disease genes has commonly resulted in the identification of one or more haplotypes exhibiting linkage disequilibrium. In the case of Machado Joseph disease, such studies have revealed a founder effect in 64 unrelated families of different geographical origins,[12] raising the possibility that the distribution of the disease chromosome throughout the world could be linked to European migration.

Prior to the identification of *ataxin-2*, segregation of an ancestral disease haplotype in pedigrees comprising the Cuban SCA2 founder population was described by Allotey *et al*,[13] incorporating loci known to flank the disease gene at considerable genetic distances. With the advent of markers known to lie in close physical proximity to SCA2, it is now feasible to undertake a comparative study of the disease haplotypes segregating in pedigrees of diverse geographical and ethnic origin. Characterisation of these haplotypes may contribute to the clarification of the phenotypic diversity seen for the disorder. Further, identification of a common disease haplotype segregating in these families could also provide the basis for determining the mutational mechanism giving rise to the expansion of the repeat motif in *ataxin-2*, which otherwise demonstrates a remarkably low level of polymorphism on non-disease chromosomes.

Materials and Methods

Pedigree Resource

Haplotype analysis was undertaken in key affected individuals from pedigrees of diverse ethnic origin, for which expansion of the CAG repeat motif within *ataxin-2*[2] was confirmed in the course of this study. These individuals are derived from seven Indian (B I-VII), 16 Cuban (HC 1-16), one English (DA23) and three Japanese (J29, J35, J36) families.

Briefly, the SCA2 phenotype segregating in each of the pedigrees included in this study is characterised by cerebellar ataxia and slow saccadic eye movements. Detailed clinical descriptions of the Indian[14] and Japanese[15] patients have been reported. In each of these populations, the pedigrees are apparently unrelated and are of diverse geographical and religious origin. The Cuban pedigrees constitute part of the collective of over 500 patients, originally described by Orozco *et al*,[16] used to assign the disease locus to chromosome 12.[17] The English pedigree, originating from the north-east, comprises a two generation pedigree segregating ADCA Type I. Recent neurological evaluation was available for two affected sibs within the second generation, who presented with ataxia of gait in their late twenties. Following a disease duration of > 10 years, the brothers continue to exhibit a comparatively mild cerebellar ataxia and remain ambulatory. No atrophy of the optic discs and no retinal pigmentation has been detected in either case. Pursuit eye movement is normal in speed and extent, but the affected sibs both show markedly slow saccades. Investigation of the mutation at the SCA2 locus showed that the expansion comprises 38 and 37 repeat units respectively. Similarly, their father carries an allele representing 37 repeats and their affected paternal aunt, 35 repeats units.

Haplotype Analysis

We have previously reported the construction of extended haplotypes in the Cuban cohort and identified a common disease chromosome[13] segregating in each of the 16 pedigrees analysed: cen – *D12S353* [allele 3; 103 bp] – (0.00) – *D12S330* [allele 13; 166 bp] – (0.02) – *D12S84* [allele 3; 219 bp] – (0.00) – *D12S105* [allele 4; 147 bp] – (0.00) – *D12S1328* [allele 4; 272 bp] – (0.03) – *D12S1332* [allele 9; 208 bp] – (0.00) – *D12S1330* [allele 3; 213 bp] – (0.01) – *D12S354* [allele 2; 197 bp] – (0.02) – *D12S79* [allele 9; 159 bp] – qter. The SCA2 locus was subsequently localised to the interval defined by *D12S1328* – *D12S1333*, the latter being placed immediately distal to *D12S1332* by the detection of a recombination event in a single Cuban pedigree (Allotey, unpublished data). No recombination between SCA2 and *D12S1332* was observed.

An affected member of one of the pedigrees described in the Allotey *et al*[13] study, pedigree HC-5 individual IV57, was therefore adopted as the reference individual in the comparative study incorporating the microsatellite loci *D12S1328* – (0.03) – *D12S1332* (0.00) – *D12S1672* – (0.00) – *D12S1333* – (0.01) – *D12S1329*. The phase of the disease chromosome for the previously untyped loci was determined in the pedigree and confirmed in the Cuban cohort. The order and location of these loci were confirmed from the physical map of the SCA2 region. *D12S1332* is located approximately 350 kb proximal to *D12S1672*. Following the characterisation of the *ataxin-2* gene, the loci *D12S1672* and *D121333* were shown to lie 20 kb proximal and 200 kb distal to the triplet repeat motif respectively.[9]

Segregation analysis was undertaken in each of the Indian, English and Japanese pedigrees, and genotypes compared with the Cuban reference individual for whom the phase of the disease chromosome had been established. PCR amplification of the five polymorphic microsatellite loci was carried out according to conditions described on the Genome Database (GDB). Products were resolved on 6% polyacrylamide gels and visualised by autoradiography.

To establish the significance of association between the disease and genetic variants at the polymorphic loci segregating in these families, the frequency of each allele was determined by the genotyping of unrelated individuals derived from the respective ethnic populations. Within the Cuban and Japanese pedigrees, 37 and 8 spouses respectively were available for analysis, and the Indian population was represented by 36 randomly selected individuals, unrelated to the disease pedigrees and of mixed religious and geographical derivation. These 81 individuals were genotyped using the markers *D12S1672* and *D121333*. Genotypes for the CEPH reference pedigrees at these loci are available on GDB.

Results

The disease haplotype segregating in each of the 12 pedigrees is shown in Table 1. Alleles are assigned according to the CEPH designation, sized in relation to CEPH individuals 134702. For those alleles not described previously, variants are designated using the next available consecutive number. Individual allele sizes are included in the legend for clarity. Full phase for the individual disease chromosomes could be ascertained at most loci; where this was not possible, the complete genotype is given.

Formal statistical analysis was not undertaken in view of the small number of disease chromosomes studied. However, marked association was observed between the disease locus and *D121333* [allele 2; 225 bp] in each of the ethnic populations, with at least 9, and possibly 11, of the 12 disease chromosomes represented by this allele. In contrast, allele 2 was infrequent in the 72 Indian (0.11), 74 Cuban (0.04), 16 Japanese (0.06) and 54 CEPH (0.11) reference chromosomes. A lesser degree of association was also observed between SCA2 and allele 2 (283 bp) at the *D12S1672* locus, represented on at least seven (and possibly nine) of the 12 disease chromosomes included in this study. However, this allele was significantly less frequent in the Cuban (0.20), Indian (0.18), Japanese (0.13) and CEPH (0.09) control chromosomes. Variance from the associated allele at *D12S1672* was observed in three of the Indian families, where SCA2 was found to segregate with allele 8 (corresponding to the presence of a single extra dinucleotide) in the pedigrees BIV, BVI and BVII. This particular allele was only represented on one of the 72 normal Indian chromosomes.

No opinion on the degree of association between the disease locus and *D12S1329* could be drawn from this study, as allele 1 is equally infrequent in both the disease and reference chromosome populations. Finally, no overall evidence of association between SCA2 and the loci *D12S1328* or *D12S1332* was apparent.

Construction of the disease haplotypes for these families indicates that SCA2 preferentially segregates with the core haplotype *D12S1672* (allele 2) – *D121333* (allele 2), definitively established in six of the 12 pedigrees, including the Cuban (HC-5) and English (DA23) families, and potentially present in a further two (BI and BV). Diversification of the disease haplotype with

Table 1 Haplotype analysis of 12 unrelated SCA2 pedigrees of diverse origin with microsatellite loci known to span the ataxin-2 locus

Ethnic origin	*Pedigree*	*D12S1328*	*D12S1332*	*D12S1672*	*D12S1333*	*D12S1329*
Cuban	HC-5	4	9	2	2	6
English	DA23	1	1/3	2	2	1
Indian	BI	1	1	1/2	2/5	1
	BII	3	9	2	2	1
	BIII	4	2	2	2	1
	BIV	1	3	8	2	1
	BV	1	11/9	2/8	2/5	1/2
	BVI	3	11	8	2	5
	BVII	3	3	8	2	1
Japanese	J29	4	11/3	2	2	2
	J35	4	11	2	2	2
	J36	1	11	2	6	2

Alleles are assigned with reference to CEPH individual 134702.
D12S1328: 1=270bp; 3=268bp; 4=272bp. *D12S1332:* 1=206bp; 2=204bp; 3=202bp; 9=208bp; 11=198bp. *D12S1672:* 1=279; 2=283bp; 8=285bp. *D12S1333:* 2=225bp; 5=239bp; 6=237bp. *D12S1329:* 1=145bp; 2=147bp; 5=149bp; 6=133bp. Where phase could not be established, the genotype is given. Shading indicates the conserved haplotype.

markers flanking these loci was apparent, although in the individual ethnic populations, a degree of association with an extended haplotype could be observed. In the case of the Japanese families, the core haplotype could be extended to include the locus *D12S1332* (allele 11), providing further evidence that these three families are related. In contrast, the *D12S1672* (allele 2) – *D12S1333* (allele 2) haplotype was extremely rare in the control samples, accounting for just six (three Indian and three Cuban) of the 162 non-SCA2 chromosomes (haplotype data is unavailable for the CEPH reference pedigrees). CAG repeat length analysis on these six individuals showed them to be homozygous for the 22 repeat allele.

In summary, therefore, comparison of the disease haplotypes segregating in each of the 12 families would support the existence of a common ancestral or predisposing chromosome in pedigrees of Cuban, Indian, English and Japanese origin.

Discussion

Allelic distribution at a specific locus may vary according to the ethnic origin of the pedigrees studied and this has certainly been evident in the detailed analysis of the Cuban pedigrees, for example, where previously unreported or rare alleles have been commonly observed. We have therefore analysed unrelated ethnically matched individuals to ascertain the frequency of alleles comprising the core haplotype in non-disease chromosomes. To reflect the diversity of the Indian SCA2 pedigrees, the 72 normal Indian chromosomes analysed are also of diverse geographical and religious origin. Regarding the Cuban population, we have been able to examine 74 chromosomes in unrelated family members (spouses). Unfortunately, only a limited number of individuals from the Japanese population (eight spouses) were available in this analysis. However, in a recent study, Mizushima *et al*[18] reported a comparison of the distribution of chromosome 12 markers in SCA2 patients and 60 unrelated normal controls. Assuming the adoption of CEPH allele nomenclature, the reported frequencies can therefore be incorporated into this study.

The argument for a common disease haplotype segregating in ethnically diverse pedigrees is strengthened by the frequency of the individual associated alleles and hence the *D12S1672 – D12S1333* haplotype observed in these reference populations. The 225 bp variant (allele 2) at *D121333*, which shows the highest degree of association with SCA2 in our study (segregating in potentially 11/12 pedigrees), has frequencies of only 0.10 (8/72), 0.04 (3/74), 0.07 (9/136) and 0.11 (6/54) in the Indian, Cuban, Japanese and CEPH reference chromosomes, respectively. Similarly, the 283 bp allele at *D12S1672* is also comparatively rare, with respective frequencies in the non-SCA2 chromosomes of just 0.18, 0.20, 0.15 and 0.09.

More definitively, the *D12S1672* (allele 2) – *D12S1333* (allele 2) haplotype, which accounts for at least 50% of the SCA2 chromosomes, is present in only 3.7% (6/162) of the control chromosomes. Consequently, as the 12 pedigrees originate from four ethnically and geographically distinct populations, the association observed between SCA2 and the *D12S1672–D12S1333* haplotype would therefore appear even more remarkable. The sizes of the SCA2 CAG repeats in the six control individuals with the *D12S1672* (allele 2) – *D12S1333* (allele 2) haplotype were determined to investigate the potential of a predisposing SCA2 allele of an intermediate size. *De novo* expansions of intermediate alleles giving rise to expansions within the pathological range have so far only been demonstrated in HD[10,19–21] and SCA7.[22] However, our analysis of the control individuals revealed no evidence for a premutation allele as all of the chromosomes possessed the 22 repeat allele.

There is not complete association of the *D12S1672–D12S1333* haplotype with SCA2 in our families, reflecting that microsatellite loci are prone to mutation as the direct result of slippage.[23] In the case of *D12S1672*, the alternative allele definitively segregating in three of the Indian families represents an increase of the variant size by just 2 bp, consistent with this hypothesis. In reality, the slippage event may have occurred once on the ancestral chromosome and the consequence of this subsequently propagated into each of three pedigrees, thus reducing the degree of association.

With an increasing number of neurodegenerative disorders arising as the result of unstable triplet repeat motifs, the quest to understand the genetic basis of predisposition to such events continues to pose a major research challenge. The identification of a common disease chromosome for spinocerebellar ataxia 2 should facilitate the characterisation of events preceding the hyperexpansion of the repeat motif underlying this disorder.

Acknowledgements

This work has been supported by ATAXIA, UK. We thank the families who have provided continuing support by their willingness to donate samples.

References

1 Orr HT, Chung MY, Banfi S *et al*: Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 1993; **4**: 221–226.
2 Pulst SM, Nechiporuk A, Nechiporuk T *et al*: Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* 1996; **14**: 269–276.
3 Sanpei K, Takano H, Igarashi S *et al*: Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* 1996; **14**: 277–284.
4 Imbert G, Saudou F, Yvert G *et al*: Cloning of the gene for spinocerebellar ataxia 2 reveals a locus with high sensitivity to expanded CAG/glutamine repeats. *Nat Genet* 1996; **14**: 285–291.
5 Kawaguchi Y, Okamoto T, Taniwaki M *et al*: CAG expansion in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* 1994; **8**: 221–227.
6 Zhuchenko O, Bailey J, Bonnen P *et al*: Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the α1A-voltage-dependent calcium channel. *Nat Genet* 1997; **15**: 62–68.
7 David G, Abbas N, Stevanin G *et al*: Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* 1997; **17**: 65–70.
8 Cancel G, Durr A, Didierjean O *et al*: Molecular and clinical correlations in spinocerebellar ataxia 2: a study of 32 families. *Hum Mol Genet* 1997; **6**: 709–715.
9 Geschwind DH, Perlman S, Figueroa CP, Treiman LJ, Pulst SM: 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* 1997; **60**: 842–850.
10 Squitieri F, Andrew SE, Goldberg YP *et al*: DNA haplotype analysis of Huntington disease reveals clues to the origins and mechanisms of CAG expansion and reasons for geographic variations of prevalence. *Hum Mol Genet* 1994; **3**: 2103–2114.
11 Neville CE, Mahadevan MS, Barcelo JM, Korneluk RG: High resolution genetic analysis suggests on ancestral haplotype for the origin of the myotonic dystrophy mutation. *Hum Mol Genet* 1994; **3**: 45–51.
12 Gaspar C, Lopes-Cendes I, DeStefano AL *et al*: Linkage disequilibrium analysis in Machado-Joseph diseae patients of different ethnic origins. *Hum Genet* 1996; **98**: 620–624.
13 Allotey R, Twells R, Cemal C *et al*: The spinocerebellar ataxia 2 locus is located within a 3-cM interval on chromosome 12q23–24.1. *Am J Hum Genet* 1995; **57**: 185–189.
14 Wadia N: Autosomal dominant cerebellar ataxia with slow saccades and peripheral neuropathy – a variety of olivopontocerebellar degeneration. In: de Jong JMBV (ed.). *Handbook of Clinical Neurology: Hereditary Neuropathies and Spinocerebellar Atrophies*. Elsevier Science Publishers BV: Oxford, 1991, vol. 16 (60), pp 491–504.
15 Sasaki H, Fukazawa T *et al*: Central phenotype and related varieties of spinocerebellar ataxia 2 (SCA2): a clinical and genetic study with a pedigree in the Japanese. *J Neurol Sci* 1996; **144**: 176–181.
16 Orozco-Diaz G, Fleites AN, Sagaz RC, Auburger G: Autosomal dominant cerebellar ataxia: clinical analysis of 263 patients from a homogeneous in Holguin, Cuba. *Neurology* 1990; **40**: 1369–1375.
17 Gispert S, Twells R, Orozco G *et al*: Chromosomal assignment of a second locus for autosomal dominant cerebellar ataxia (SCA2) to chromosome 12q23–24.1. *Nat Genet* 1993; **4**: 295–299.
18 Mizushima K, Watanabe M, Kondo I *et al*: Analysis of spinocerebellar ataxia type 2 gene and haplotype analysis: $(CCG)_{1-2}$ polymorphism and contribution to founder effect. *J Med Genet* 1999; **36**: 112–114.
19 Myers RH, MacDonald ME, Koroshetz WJ *et al*: *De novo* expansion of a $(CAG)_n$ repeat in sporadic Huntington's disease. *Nat Genet* 1993; **5**: 168–173.
20 Goldberg YP, Kremer B, Andrew SE *et al*: Molecular analysis of new mutations for Huntington's disease: intermediate alleles and sex of origin effects. *Nat Genet* 1993; **5**: 174–179.
21 Goldberg YP, McMurray CT, Ziesler J *et al*: Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. *Hum Mol Genet* 1995; **4**: 1911–1918.
22 Stevanin G, Giunti P, David G *et al*: *De novo* expansion of intermediate alleles in spinocerebellar ataxia 7. *Hum Mol Genet* 1998; **7**: 1809–1813.
23 Weber JL, Wong C: Mutation of short tandem repeats. *Hum Mol Genet* 1993; **2**: 1123–1128.