European Journal of Human Genetics (2008) 16, 841–847

www.nature.com/ejhg

ARTICLE

Pathogenic expansions of the SCA6 locus are associated with a common *CACNA1A* haplotype across the globe: founder effect or predisposing chromosome?

Kate Craig[1], Yoshihisa Takiyama[2], Bing-Wen Soong[3], Laura B Jardim[4], Maria Luiza Saraiva-Pereira[4], Kieren Lythgow[1], Hiroyuki Morino[5], Hirofumi Maruyama[5], Hideshi Kawakami[5] and Patrick F Chinnery*[,1,6]

[1]*Neurology, Medical School,University of Newcastle upon Tyne, Newcastle upon Tyne, UK;* [2]*Department of Neurology, Jichi Medical School, Shimotsuke-shi, Japan;* [3]*School of Medicine, National Yang-Ming University, Taipei, Taiwan;* [4]*Medical Genetics Service, Hospital de Clinicas de Porto Allegre, Porto Allegre, Brazil;* [5]*Department of Epidemiology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima, Japan;* [6]*Institute of Human Genetics, University of Newcastle upon Tyne, Newcastle upon Tyne, UK*

Spinocerebellar ataxia type 6 (SCA6) is a common cause of dominantly inherited ataxia due to an expansion of the CAG repeat in the *CACNA1A* gene. Affected individuals from the same population share a common haplotype, raising the possibility that most SCA6 cases have descended from a small number of common founders across the globe. To test this hypothesis, we carried out haplotype analysis on SCA6 families from Europe, South America and the Far East, including an established *de novo* SCA6 expansion. A core *CACNA1A* disease haplotype was found in affected individuals across the globe. This was also present in the unaffected father of the *de novo* case, suggesting that the shared chromosome predisposes to the CAG repeat expansion at the SCA6 locus. The SCA6 expansion lies within a CpG island, which could act as a *cis*-acting element predisposing to repeat expansion as for other CAG/CTG repeat diseases. Polymorphic variation in this region may explain the high-risk haplotype found in SCA6 families.
European Journal of Human Genetics (2008) **16,** 841–847; doi:10.1038/ejhg.2008.20; published online 20 February 2008

Keywords: ataxia; spinocerebellar ataxia; trinucleotide repeat; founder effect; SCA6; *CACNA1A*

Introduction

Spinocerebellar ataxia type 6 (SCA6, MIM 183086) is a late-onset, slowly progressive neurodegenerative disorder that characteristically presents with dysarthria, gait and limb ataxia.[1] The disease is caused by an expansion of the CAG repeat in the α1A subunit of the voltage-dependent calcium channel gene *CACNA1A*.[2] SCA6 appears to be a major cause of dominantly inherited ataxia, affecting at least 1.59/100 000 of the UK population[3] and accounting for between 6 and 32% of families with autosomal dominant ataxia.[4] Haplotype analysis in different geographical regions identified shared regions of chromosome 19p13 in affected individuals, suggestive of a common founder chromosome existing in Germany,[4] Japan,[5,6] the Netherlands[7] and the United Kingdom.[3] This raises the possibility that all affected individuals inherited one or a few common founder chromosomes, as has been described

*Correspondence: Professor PF Chinnery, M41014, Neurology, Medical School, University of Newcastle upon Tyne, Framlington Place, Newcastle upon Tyne, NE2 4HH, UK.
Tel: +44 191 222 8334; Fax: +44 191 222 8553;
E-mail: p.f.chinnery@ncl.ac.uk
Received 25 October 2007; revised 18 December 2007; accepted 20 December 2007; published online 20 February 2008

for nucleotide repeat disorders.[8–11] Against this hypothesis, a few proven *de novo* SCA6 CAG expansions have been described,[12,13] pointing toward a predisposing chromosome, rather than a founder effect. We previously identified a common *CACNA1A* haplotype present in 16 pedigrees with SCA6 from the northeast of England.[3] To determine whether this haplotype was due to a founder effect, or was predisposing to CAG repeat expansion in *CACNA1A*, we carried out microsatellite analysis on SCA6 families throughout the world, including a *de novo* case.

Materials and methods

Subjects

Haplotype analysis was carried out on 96 individuals (95 affected and 1 unaffected) from 45 families with a molecular diagnosis of SCA6. Twenty-two families were from the northeast of England ($n=37$ subjects, 16 of these families have been described before[3]), 12 families were Japanese ($n=31$ subjects), 2 families were Brazilian ($n=7$ subjects), 2 families were Finnish ($n=2$ subjects) and 9 families were Taiwanese ($n=18$ subjects). Allele frequencies were determined in control subjects from corresponding geographical regions (northeast of England, $n=50$; Japan, $n=100$; Brazil, $n=50$; Taiwan, $n=56$).

Genetic analysis

All samples were genotyped for the following $(CA)_n$ microsatellite markers in the same laboratory D19S912, D19S906, D19S221, D19S914, D19S1150, D19S840, D19S226, D19S899 and D19S414 (primer sequences and map positions were obtained from the NCBI UniSTS database; Figure 1; Supplementary Table 1, online). Primers were 5′ fluorescent labeled and PCR products for each allele were simultaneously analyzed on a single capillary DNA analyzer (Beckman CEQ 8000). Reaction conditions 0.25 pM of each primer, 1 U of Promega or HotMaster™ *Taq* DNA polymerase with 1 × associated buffer, 2 mM dNTPs and 250 ng of DNA. Amplification was carried out at 94°C for 4 min, (or 94°C for 2 min for HotMaster), followed by 30 cycles at 94°C for 1 min, annealing temperature for specific microsatellite for 1 min and 72°C for 1 min with a final extension of 72°C for 10 min.

Statistical analysis

Haplotypes were constructed manually for the familial samples and inferred using Phase v 2.1.1 for the control subjects.[14] The frequency of individual alleles and haplotypes in cases and controls were compared using Fisher's exact test. Linkage disequilibrium (LD) was estimated by the parameter δ, which is an approximation of the population attributable risk according to the equation, $\delta=(F_d-F_c)/(1-F_c)$, where F_d is the frequency of the allele in carrier chromosomes and F_c is the frequency of the allele in noncarrier chromosomes.[15]

Results

Affected individuals from the 6 additional English families shared microsatellites with the 16 families previously reported (Figure 2a), with highly significant association and LD between the intragenic marker D19S1150 and flanking marker D19S840 (Table 1 and Supplementary Tables 2 and 3, online). A similar haplotype was also found in the Japanese, Brazilian and Finnish SCA6 families (Figures 2b–d). Shared flanking markers between the SCA6 families suggest minor differences in the core haplotype (D19S1150 and D19S840) between these regions that probably arise through single mutation events, which occur between 0 and 7×10^{-3} per locus per gamete per generation.[16] Again, specific alleles and haplotypes in each population were associated and in LD with mutated *CACNA1A* alleles (Tables 1 and 2, Supplementary Tables 2 and 3, online). The same centromeric marker alleles were also found in the Taiwanese families, as were the same telomeric alleles for D19S914/D19S906 (Figure 2e). Given that the same D19S914 alleles were strongly associated, and in LD with mutated SCA6 alleles, in both the Taiwanese and Japanese families (Table 1), this suggests that the different intragenic D19S1150 alleles in Taiwanese families are also due to mutation of the microsatellite, as described for other intragenic diallelic markers defining an ancient founder haplotype, which has spread throughout the world.[17]

Haplotype analysis of a Japanese *de novo* SCA6 patient and his parents was also carried out. Although CAG_{20} alleles have been associated with late-onset mild ataxia,[18]

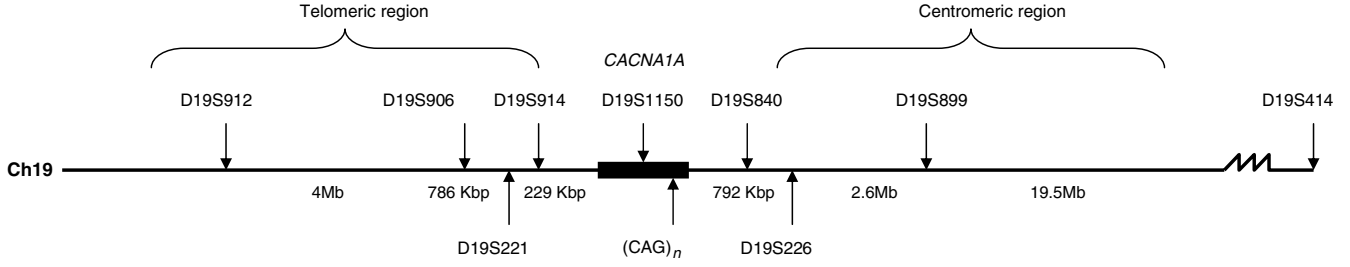

Figure 1 Schematic representation of microsatellite marker positions on chromosome 19. Black box = *CACNA1A*. (Mb = megabase, kbp = kilobase pairs).

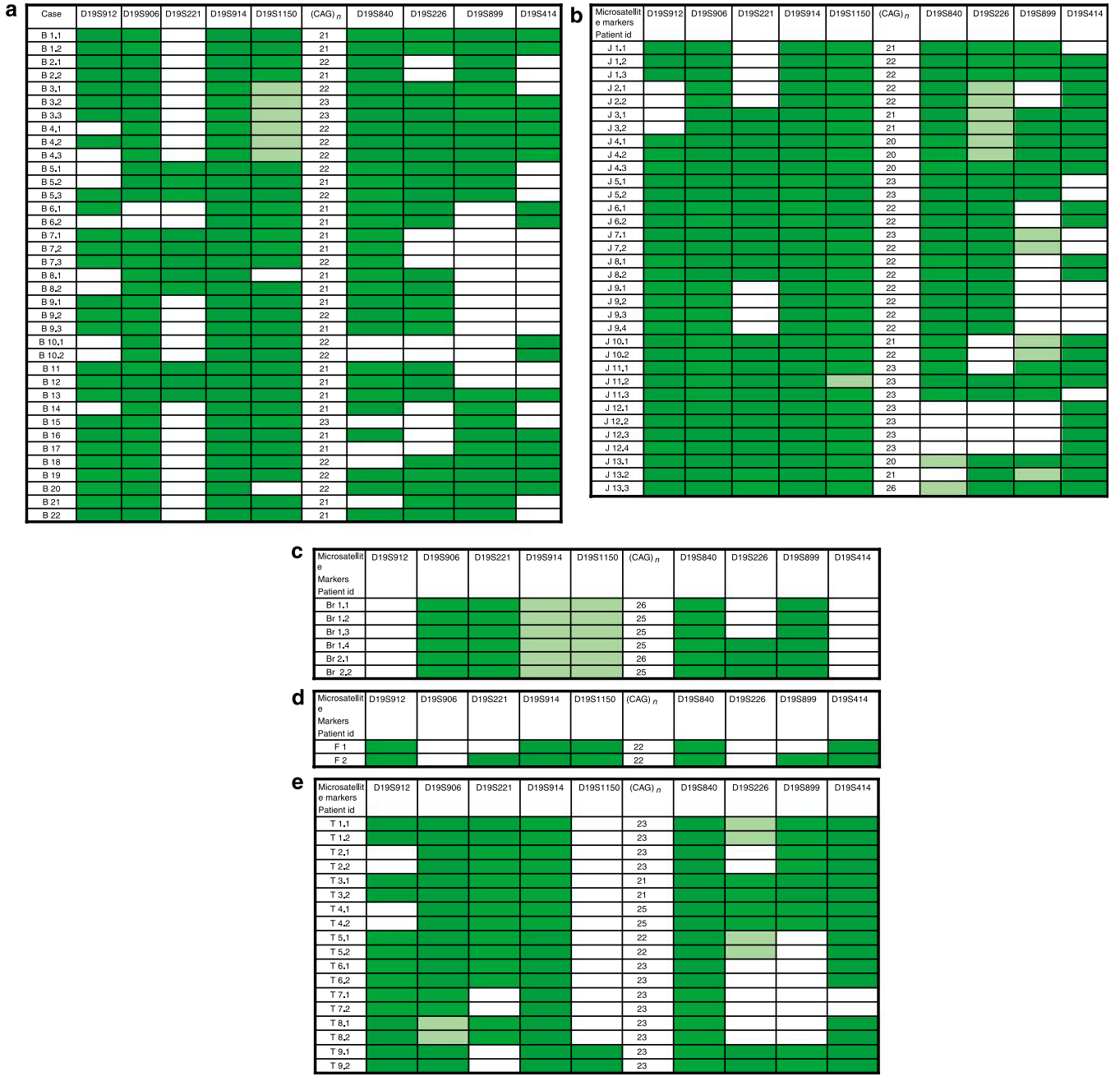

Figure 2 CACNA1A haplotypes for the SCA6 subjects. (**a**) British, (**b**) Japanese, (**c**) Brazilian, (**d**) Finnish and (**e**) Taiwanese SCA6 patients. Numbers after the decimal place represent individuals within a family. Haplotypes were determined manually. Both alleles were considered where it was not possible to determine phase. The most common (core) haplotype is shown in dark green and the allele sizes are as follows: D19S912(178), D19S906(158), D19S221(198), D19S914(90), D19S1150(160), D19S840(204), D19S226(245), D19S899(112) and D19S414(163). With paler shading representing slightly different haplotypes that are closely related to the core haplotype (differing by one or two dinucleotide repeats) and probably differ due to microsatellite instability. Unshaded regions are regions that differ through recombination and are not related to the core haplotype.

the patient was previously reported as being *de novo* by Shimazaki *et al*,[12] given that the parents were neurologically and radiologically normal at the time of the original study, and on follow-up prior to this study (Figure 3). The extended haplotype suggested that the CAG expansion occurred during paternal transmission of the CAG_{20} allele, which is characteristic of CAG repeat disorders.[19] Although the core haplotype carrying the mutated SCA6 allele in this

Table 1 Frequency of microsatellite marker alleles in SCA6 families and control subjects from the same geographic region

Ethnic origin	*Marker/allele*		*Frequency*			
			Controls	*SCA6 families*	*Fisher's exact (P)*	δ
British	D19S1150	154	0.34	0.13	0.0021	0.42
	D19S1150	160	0.19	0.41	0.0018	0.16
	D19S1150	164	0.02	0.13	0.0082	0.32
	D19S840	216	0.01	0.16	0.0002	0.153
Japanese	D19S914	88	0.1	0.22	0.0114	0.138
	D19S914	90	0.69	0.46	0.0008	−0.738
	D19S914	92	0.12	0.25	0.0183	0.147
	D19S1150	158	0.17	0.47	0.0001	0.359
	D19S1150	160	0.28	0.15	0.0332	−0.186
	D19S1150	166	0.22	0.03	0.0001	−0.252
	D19S1150	168	0.02	0.09	0.0207	0.069
	D19S840	202	0.03	0.12	0.0099	0.09
Brazilian	D19S914	94	0.07	0.5	0.0005	0.462
	D19S840	206	0.25	0.58	0.0363	0.444
	D19S840	218	0	0.25	0.001	0.25
Taiwanese	D19S914	90	0.6	0.36	0.019	−0.597
	D19S914	92	0.11	0.39	0.0007	0.313
	D19S1150	166	0.32	0	0.0001	−0.464
	D19S1150	168	0.05	0.56	0.0001	0.533

P-values were calculated using Fisher's exact test. The parameter δ is an approximation of the population attributable risk according to the equation, $\delta=(F_d-F_c)/(1-F_c)$, where F_d is the frequency of the allele in carrier chromosomes and F_c is the frequency of the allele in noncarrier chromosomes. Control subjects were from the same geographic region (see text). Only significant associations are shown in this table. The entire data set is shown in Supplementary Table 2, which is available online.

Table 2 Frequency of *CACNA1A* haplotypes in SCA6 families and controls defined by microsatellite markers D19S914, D19S1150 and D19S840

Ethnic origin	*CACNA1A haplotype*	*Frequency*			
		Controls	*SCA6 families*	*Fisher's exact (P)*	δ
British	90/154/204	0.1	0	0.0054	−0.111
	90/162/206	0.03	0.42	0.0001	0.441
	92/160/206	0	0.20	0.0001	0.203
	92/160/214	0	0.05	0.0312	0.054
	92/162/206	0.01	0.09	0.0110	0.085
	92/164/204	0	0.05	0.0312	0.054
Japanese	88/154/214	0	0.04	0.0158	0.044
	90/158/204	0	0.14	0.0001	0.147
	90/158/214	0	0.15	0.0001	0.147
	90/158/206	0	0.09	0.0001	0.103
	90/160/204	0.56	0.01	0.0001	−1.214
	90/160/206	0.07	0	0.0240	−0.0752
	90/166/204	0.17	0	0.0001	−0.2048
	92/156/204	0	0.04	0.0158	0.044
	92/168/204	0	0.04	0.0158	0.044
Brazilian	90/154/218	0	0.25	0.001	0.25
	92/160/204	0.47	0	0.0012	−0.887
	94/162/206	0	0.5	0.0001	0.5
	96/160/204	0.49	0	0.001	−0.961
Taiwanese	88/152/210	0.44	0	0.0001	−0.778
	90/154/208	0.49	0	0.0001	−0.965
	90/160/214	0	0.14	0.0007	0.139
	92/168/204	0	0.31	0.0001	0.305

P-values were calculated using Fisher's exact test. The parameter δ is an approximation of the population attributable risk according to the equation, $\delta=(F_d-F_c)/(1-F_c)$, where F_d is the frequency of the allele in carrier chromosomes and F_c is the frequency of the allele in noncarrier chromosomes. Control subjects were from the same geographic region (see text). Only significant associations are shown in this table. The entire data set is shown in Supplementary Table 3, which is available online.

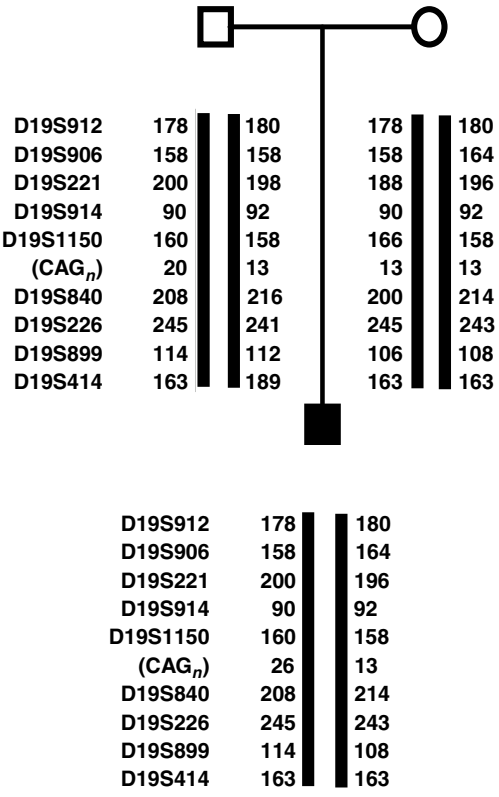

Figure 3 Pedigree for the *de novo* SCA6 patient and their parents showing the haplotypes and the possible inherited haplotype from parent to offspring.

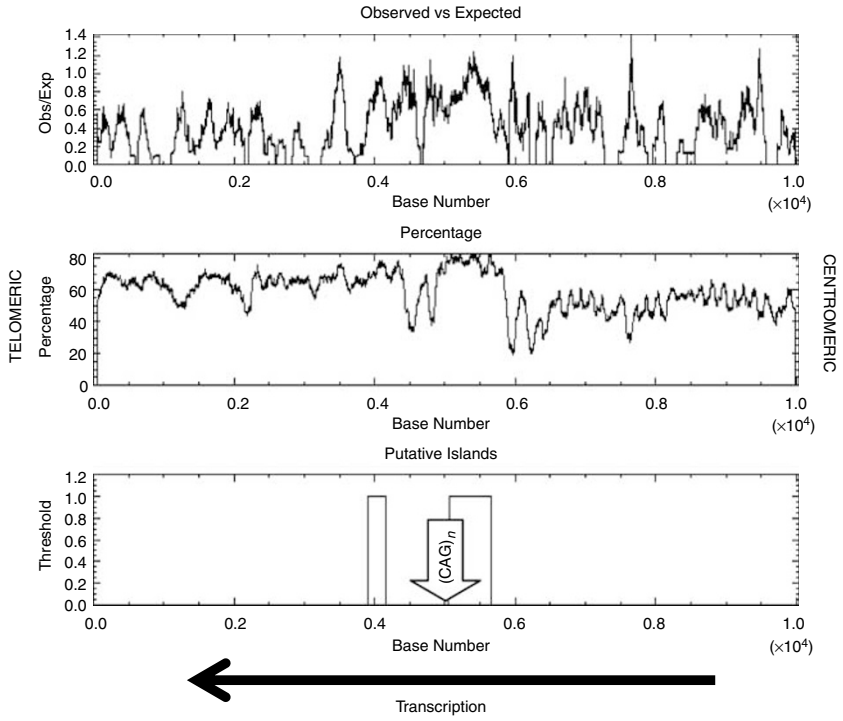

Figure 4 Bioinformatic analysis of the 5000-bp of DNA sequence flanking either side of the *CACNA1A* $(CAG)_n$ repeat, based on the algorithm of Gardiner-Garden and Frommer (1987).[27] Upper panel: CpG prediction plots (observed/expected). Middle panel: percentage of CG residues (% CG). Lower panel: predicted CpG islands. The $(CAG)_n$ repeat is indicated by an open arrow. Note that the numbering from left to right corresponds to the direction of the NCBI chromosomal sequence. The gene is transcribed from right to left (solid arrow).

family is rare (D19S914(90)-D19S1150(160)-D19S840(208)) (Supplementary Table 3, online), both telomeric (D19S914(90) and D19S1150(160)) centromeric to D19S840 markers that define the SCA6 chromosome in this family were also found in other Japanese families (Supplementary Table 2, online). Thus, although it is possible that the father will go on to develop a mild form of SCA6 late in life (and thus have a pathogenic allele), the haplotype analysis described here indicates that the unstable paternal chromosome is the likely origin of the *de novo* expansion, and that this occurred on a haplotype found in other Japanese SCA families.

Discussion

The identification of a common *CACNA1A* haplotype in affected individuals with SCA6 from Europe, Brazil and Japan supports the hypothesis that all SCA6 patients descend from a small pool of founder individuals. However, the demonstration a *de novo* SCA6 expansion on a similar genetic background raises the possibility of a predisposing haplotype leading to new mutation events in different populations across the globe. Recent evidence from SCA7 transgenic mice has shown that *cis*-acting elements 3′ to the repeat drive instability of the $(CAG)_n$ at the SCA7 locus[20] and a similar mechanism may operate for SCA6. Mice generally require a larger $(CAG)_n$ tract than humans to show instability of repeat lengths[21–23] but introducing large regions of flanking human sequence allows instability for moderate repeat lengths in some mouse models.[20,21]

CpG methylation may also influence trinucleotide repeat tract stability through an effect on chromosomal structure.[24,25] We therefore used a bioinformatic approach using a previously described method to study GC content flanking other trinucleotide repeat genes.[26,27] Genomic sequence for 5000 bp upstream and 5000 bp downstream of the *CACNA1A* $(CAG)_n$ repeat sequence were downloaded from NCBI and analyzed using cpgplot (available through EMBOSS) in the 5′–3′ orientation, with a moving window of 500 bp and a step of 100 bp.[27] This revealed a large CpG island immediately upstream of the *CACNA1A* $(CAG)_n$ repeat sequence (Figure 4, 56 bp upstream of the CAG repeat spanning 611 bp). These sequence characteristics are associated with repeat instability in other disorders.[26] At 77.94%, the percentage GC content of the chromosomal region immediately upstream of the *CACNA1A* $(CAG)_n$ repeat is greater than that found for other unstable

pathogenic repeat sequences (including *HD*, *SCA1* and *SCA3*) and just less than the region flanking the highly unstable SCA7 at 83.5%.[26]

It is intriguing that the region 5′ to the CAG repeat in the *de novo* expansion (Figure 3) was defined by markers strongly associated with other mutated SCA6 alleles (Table 1). Conversely, the region 3′ to the CAG repeat on the newly mutated chromosome contained alleles not associated with $(CAG)_n$ expansions (Table 1). This is in keeping with a common telomeric region predisposing to pathogenic repeat formation. Mammalian CpG islands are known sites for the initiation of transcription and may also act as origins of replication,[28] and modification of repeat instability has been linked to both transcription and replication origin events.[29] Methylation of CpG islands appears to stabilize GCG tracts in fragile X syndrome,[30] although the mechanisms involved are likely to be complex,[19] and may relate to GC content and chromatin structure.[25] It is therefore possible that sequence variation within the CpG island immediately upstream of the $(CAG)_n$ alters the susceptibility of specific *CACNA1A* haplotypes to cause repeat expansion, as had been described for other repeat sequences.[31] Identifying the underlying molecular mechanism will have important implications for our understanding of the onset and genetic anticipation of SCA6 in large dominant families.

Acknowledgements

This study was funded by Ataxia UK. PFC is a Wellcome Trust Senior Clinical Research Fellow who also receives funding from the United Mitochondrial Diseases Foundation, a research grant from the United States Army and the EU FP program EUmitocombat and MITOCIRCLE. We thank Dr Kirsi Huoponen for supplying the SCA6 families from Finland.

Electronic database information

Accession numbers and URLs for data presented herein are as follows: European Bioinformatics Institute (EMBOSS), CpG plot, http://www.ebi.ac.uk/emboss/cpgplot/; National Centre for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/; Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for SCA6 and *CACNA1A*); and UniSTS, http://www.ncbi.nlm.nih.gov.

References

1 Schols L, Kruger R, Amoiriddis G: Spinocerebellar ataxia type 6: genotype and phenotype in German kindreds. *J Neurol Neurosurg Psych* 1998; **64**: 67–73.
2 Zhuchenko O, Bailey J, Bonnen P *et al*: Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat Genet* 1997; **15**: 62–69.
3 Craig K, Keers SM, Archibald K, Curtis A, Chinnery PF: Molecular epidemiology of spinocerebellar ataxia type 6. *Ann Neurol* 2004; **55**: 752–755.
4 Dichgans M, Schols L, Hertzog J, Stevanin G *et al*: Spinocerebellar ataxia type 6: evidence for a strong founder effect among German families. *Neurology* 1999; **52**: 849–851.
5 Mori M, Adachi Y, Kusumi M, Nakashima K: Spinocerebellar ataxia type 6: founder effect in Western Japan. *J Neurol Sci* 2001; **185**: 43–47.
6 Terasawa H, Oda M, Morino H *et al*: A novel haplotype of spinocerebellar ataxia type 6 contributes to the highest prevalence in western Japan. *Neurosci Lett* 2004; **358**: 107–110.
7 Verbeek DS, Piersma SJ, Hennekam EF, Ippel EF, Pearson PL, Sinke RJ: Haplotype study in Dutch SCA3 and SCA6 families: evidence for common founder mutations. *Eur J Hum Genet* 2004; **12**: 441–446.
8 Neville CE, Mahadevan MS, Barcelo JM, Korneluk RG: High resolution genetic analysis suggests one ancestral predisposing haplotype for the origin of the myotonic dystrophy mutation. *Hum Mol Genet* 1994; **3**: 45–51.
9 Bachinski LL, Udd B, Meola G *et al*: Confirmation of the type 2 myotonic dystrophy (CCTG)n expansion mutation in patients with proximal myotonic myopathy/proximal myotonic dystrophy of different European origins: a single shared haplotype indicates an ancestral founder effect. *Am J Hum Genet* 2003; **73**: 835–848.
10 Saleem Q, Choudhry S, Mukerji M *et al*: Molecular analysis of autosomal dominant hereditary ataxias in the Indian population: high frequency of SCA2 and evidence for a common founder mutation. *Hum Genet* 2000; **106**: 179–187.
11 Gaspar C, Lopes-Cendes I, Hayes S *et al*: Ancestral origins of the Machado–Joseph disease mutation: a worldwide haplotype study. *Am J Hum Genet* 2001; **68**: 523–528.
12 Shimazaki H, Takiyama Y, Sakoe K *et al*: Meiotic instability of the CAG repeats in the SCA6/CACNA1A gene in two Japanese SCA6 families. *J Neurol Sci* 2001; **185**: 101–107.
13 Lindquist SG, Norremolle A, Hjermind LE, Hasholt L, Nielsen JE: Meiotic CAG repeat instability in spinocerebellar ataxia type 6: maternally transmitted elongation in a presumed sporadic case. *J Neurol Sci* 2006; **241**: 95–98.
14 Stephens M, Smith NJ, Donnelly P: A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001; **68**: 978–989.
15 Devlin B, Risch N: A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 1995; **29**: 311–322.
16 Brinkman B, Klintschar M, Neuhuber F, Huhne J, Rolf B: Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* 1998; **62**: 1408–1415.
17 Zabetian CP, Hutter CM, Yearout D *et al*: LRRK2 G2019S in families with Parkinson disease who originated from Europe and the Middle East: evidence of two distinct founding events beginning two millennia ago. *Am J Hum Genet* 2006; **79**: 752–758.
18 Yabe I, Sasaki H, Matsuura T *et al*: SCA6 mutation analysis in a large cohort of the Japanese patients with late-onset pure cerebellar ataxia. *J Neurol Sci* 1998; **156**: 89–95.
19 Pearson CE, Nichol Edamura K, Cleary JD: Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet* 2005; **6**: 729–742.
20 Libby RT, Monckton DG, Fu YH *et al*: Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. *Hum Mol Genet* 2003; **12**: 41–50.
21 Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, Bates GP: Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. *Nat Genet* 1997; **15**: 197–200.
22 Monckton DG, Coolbaugh MI, Ashizawa KT, Siciliano MJ, Caskey CT: Hypermutable myotonic dystrophy CTG repeats in transgenic mice. *Nat Genet* 1997; **15**: 193–196.
23 Sato T, Oyake M, Nakamura K *et al*: Transgenic mice harboring a full-length human mutant DRPLA gene exhibit age-dependent intergenerational and somatic instabilities of CAG repeats

comparable with those in DRPLA patients. *Hum Mol Genet* 1999; **8**: 99–106.
24 Carbonell P, Lopez I, Gabarron J *et al*: FRAXE mutation analysis in three Spanish families. *Am J Med Genet* 1996; **64**: 434–440.
25 Otten AD, Tapscott SJ: Triplet repeat expansion in myotonic dystrophy alters the adjacent chromatin structure. *Proc Natl Acad Sci USA* 1995; **92**: 5465–5469.
26 Brock GJ, Anderson NH, Monckton DG: *Cis*-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum Mol Genet* 1999; **8**: 1061–1067.
27 Gardiner-Garden M, Frommer M: CpG islands in vertebrate genomes. *J Mol Biol* 1987; **196**: 261–282.
28 Delgado S, Gomez M, Bird A, Antequera F: Initiation of DNA replication at CpG islands in mammalian chromosomes. *EMBO J* 1998; **17**: 2426–2435.
29 Kang S, Jaworski A, Ohshima K, Wells RD: Expansion and deletion of CTG repeats from human disease genes are determined by the direction of replication in *E. coli*. *Nat Genet* 1995; **10**: 213–218.
30 Nichol Edamura K, Leonard MR, Pearson CE: Role of replication and CpG methylation in fragile X syndrome CGG deletions in primate cells. *Am J Hum Genet* 2005; **76**: 302–311.
31 Monckton DG, Neumann R, Guram T *et al*: Minisatellite mutation rate variation associated with a flanking DNA sequence polymorphism. *Nat Genet* 1994; **8**: 162–170.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (http://www.nature.com/ejhg)