npg

ARTICLE

Haplotype study in Dutch SCA3 and SCA6 families: evidence for common founder mutations

Dineke S Verbeek*,1, Sytse J Piersma[1], Eric FAM Hennekam[1], Elly F Ippel[1], Peter L Pearson[1] and Richard J Sinke[1]

[1]Department of Medical Genetics, University Medical Center, Stratenum, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

This pilot study was initiated to show the existence of founder effects in the Dutch autosomal dominant cerebellar ataxia (ADCA) population. The ADCAs comprise a clinically heterogeneous group of neurodegenerative disorders and the estimated prevalence in the Netherlands is approximately 3:100 000 individuals. Here, we focused on the SCA3 and SCA6 genes because mutations in these genes occur most frequently in the Netherlands. We were able to determine a common origin of the CAG repeat expansions in the majority of Dutch SCA3 and SCA6 families. Haplotype analysis and linkage disequilibrium studies with polymorphic markers revealed shared haplotypes surrounding the SCA3 and SCA6 genes. These results strongly suggest that ADCA families can be traced back to common ancestors in particular parts of the Netherlands.

Introduction

The autosomal dominant cerebellar ataxias (ADCAs) comprise a clinically heterogeneous group of neurodegenerative disorders, characterized by a progressive ataxic syndrome often accompanied by a wide range of additional symptoms including tremor, headache or epilepsy. The disorder is genetically heterogeneous: so far 10 SCA genes (SCA1–3, 6–8, 10, 12, 14 and 17)[1–9] have been isolated and characterized, while an additional nine SCA loci (SCA4, 5, 11, 13, 15, 16, 18, 19, 21 and 24)[10–18] have been identified by linkage analysis. The estimated prevalence of ADCA in the Netherlands is approximately 3:100 000 individuals.[19] The ADCA cohort in Utrecht comprises families in whom in 70% of the cases the clinical diagnosis for SCA1, 2, 3, 6, or 7 could be confirmed by molecular genetic analysis.

Haplotype studies using closely linked microsatellites and intragenic single-nucleotide polymorphisms (SNPs) provided evidence for the existence of founder mutations for the CAG repeat expansion mutations in the SCA2, 3, 6 and 7 genes in different ethnic populations.[20–24] So far, no such extensive haplotype study has been performed in the Dutch ADCA population. We focused on the SCA3 and 6 genes to identify the presence of possible founder effects in the Dutch ADCA population. The SCA3 mutation occurs most frequently (44.1%) in the Dutch ADCA population, followed by mutations in the SCA6 gene (23.4%).[19]

The strategy is based on the fact that seemingly unrelated or distantly related ADCA families will not only share a common mutation but also a conserved region surrounding the disease locus in which marker alleles will be in linkage disequilibrium (LD).

We were able to identify a common origin of the CAG repeat expansions in the majority of Dutch SCA3 and 6 families, and these results imply that independently referred ADCA families may actually be linked to larger families.

Subjects and methods
The patients
Twenty-one apparently independent SCA3 families and 12 SCA6 families, as well as 32 isolated SCA3 patients and 25 isolated SCA6 patients, were selected from the Utrecht ADCA cohort. High molecular DNA was isolated from peripheral blood by a routine salting-out procedure. The clinical diagnosis was confirmed by a positive mutation analysis (data not shown). In order to be able to generate haplotypes, we genotyped at least one affected individual with both parents (when available) to determine the phase. We created simulated haplotypes with unknown phase for the isolated patients. In total, we typed 77 SCA3 patients and 50 SCA6 patients. Control samples from healthy, anonymous nonrelated individuals and parents (total 80 haplotypes) were used to estimate the allele and haplotype frequencies of the markers.

Polymorphic markers
Polymorphic markers flanking the SCA3 and 6 genes were selected from the Marshfield database http://research.marshfieldclinic.org/genetics/ (see Figures 1a and b), with an average spacing of 1–2 cM. If no markers were available in the database to provide optimal spacing, we identified additional polymorphic repeats using a tandem repeats finder program available on the web.[25] The primer sequences of those markers are given in Table 1. The protocol used to amplify the polymorphic markers has been described previously.[17] The different marker alleles were annotated using the CEPH 133101. The alleles that characterize the SCA3 core haplotype are represented by: D14S997 allele 4, 263 base pair (bp); D14S617 allele 6, 161 bp/allele 7, 165 bp; D14S1015 allele 2, 252 bp; D14S1015 allele 6, 138 bp; D14S973 allele 3, 150 bp and D14S977 allele 11, 120 bp. For SCA6, the relevant alleles are: D19S906 allele 3, 158 bp; D19S1165 allele 3, 158 bp; D19S558 (134701) allele 3, 150 bp; D19S1150 (134701) allele 6, 152 bp; D19S840 allele 3, 206 bp.

After amplification, the markers were pooled in equal volumes according to their respective size ranges and the color of the fluorescent label. The markers were run on a 3700 ABI machine 96 capillary automated sequencer (Applied Biosystems, Foster City, CA, USA). The pooled and diluted marker sample was added to a mixture of HiDi and the internal lane standard Rox 500 (ratio 100:1). Analysis, sizing and calling of the fluorescent fragments were performed with Genescan v3.1 and Genotyper v2.1 software. Haplotypes were generated with the program Genehunter v2.1 and checked manually. In total, we used 21 SCA3 and 12 SCA6 haplotypes with known phase and 32 SCA3 and 25 SCA6 haplotypes with unknown phase in our analysis.

Single-nucleotide polymorphisms (SNPs)
We typed three intragenic SNPs, 669A/G, 987C/G and 118A/C, using the SNaP-Shot method (Applied Biosystems, Foster City, CA, USA) in the 21 SCA3 families. Amplification of the SNPs was carried out with a standard PCR and primers as described elsewhere.[21] The templates were purified by a post-PCR shrimp alkaline phosphatase (SAP) and endonuclease 1 enzyme treatment. After inactivation of the enzymes used in the purification, an allele-specific PCR was performed and the products were purified once more. The resulting products were analyzed on the ABI 3700 and Liz 120 (Applied Biosystems, Foster City, CA, USA) was used as an internal standard.

Linkage disequilibrium (LD)
We searched for differences in the overall distribution of alleles between pairs of markers within the SCA3 and 6 regions. LD was calculated with an extension of Fisher's exact probability test using the Arlequin v2.0 program. P-values less than 0.05 were considered statistically significant.

Results
Haplotype analysis of SCA3
In total, 14 markers from the SCA3 locus region were typed in 21 SCA3 families (Figure 1a). We observed a highly conserved core haplotype in 17 of the 21 SCA3 families between markers D14S995 and D14S973, 4-6/7-2-6-3, corresponding to an approximately 1.4 Mb genomic region surrounding the SCA3 gene. In addition, we observed a truncated form of this haplotype, 6/7-2-6-3, in four additional families. The marker D14S617, located in the middle of the core haplotype, showed a 6-allele or a 7-allele (ratio 1:1) on the SCA3 chromosome. This core haplotype was never observed in 80 control chromosomes of unrelated individuals with known phase. Interestingly, extended haplotypes, up to 6.2 Mb in size (see Figure 1a; depicted in white, black and gray boxes), were observed in three groups of families.

Significant LD ($P < 0.05$) was observed, involving a block of markers flanked by markers D14S995 and D14S977 (see Table 2).

We were able to create the core haplotype 4-6/7-2-6-3 in 31% of the constructed alleles of 32 isolated SCA3 patients. Truncated haplotypes derived from this core haplotype could be observed in an additional 33.7% of the chromosomes (data not shown). We used isolated SCA6 patients to control and explore the specificity of this SCA3 haplotype in alleles with unknown phase. The core haplotype 4-6/7-2-6-3 could only be constructed in 2% of the SCA6 chromosomes.

Genealogical research was performed to assist in defining the origin of the core haplotype and was able to link 10 independently referred SCA3 families into four clusters (see
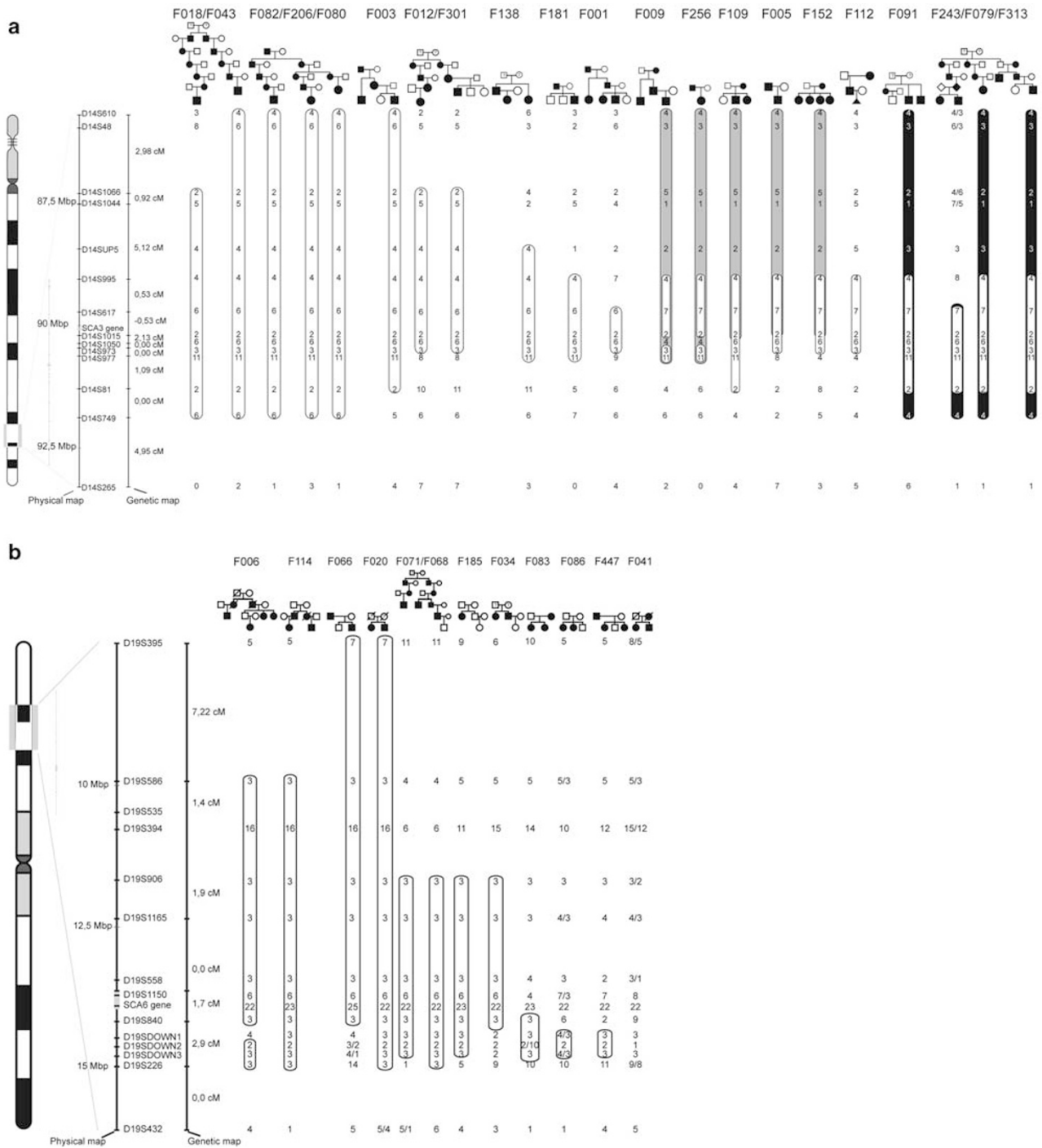
Figure 1 (**a**) Shared haplotype analysis of the SCA3 locus. (**b**) Shared haplotype analysis of the SCA6 locus. The markers were selected from the Marshfield database. The physical distances are based on the Ensembl release 8.0 and the genetic distances (Mb) are based on the Marshfield genetic map (cM). Only one affected chromosome per family is depicted in the figures. The shared haplotypes are boxed and the extended haplotypes are colored by white, gray and black.

Table 1 Additional SCA6 markers

Marker name	Forward primer	Reverse primer
D19Sdown1	CCAGACTGGTGCTCTGAC	GTGACCAACTTGCCTCTCT
D19Sdown2	CTGTAACCCCAGGAGTTTGT	ACGGAGTTTCGCTTTTGTA
D19Sdown3	GGAGGAGGTAGGAGAGAAGG	CACATTTCCAGGGAGAG

Table 2 Linkage disequilibrium between 12 SCA3 markers

	D14S610	D14S1066	D14S1044	D14Sup5	D14S995	D14S617	D14S1015	D14S1050	D14S973	D14S977	D14S81	D14S749
D14S610	*				+		+					
D14S1066		*	+		+		+					
D14S1044		+	*			+						
D14Sup5				*	+							+
D14S995	+	+			*	+	+	+	+	+	+	
D14S617			+	+	+	*	+	+	+	+		+
D14S1015	+	+			+	+	*	+	+	+	+	
D14S1050					+	+	+	*	+	+		
D14S973					+	+	+	+	*	+		
D14S977					+	+	+	+	+	*		
D14S81	+				+		+				*	
D14S749				+		+						*
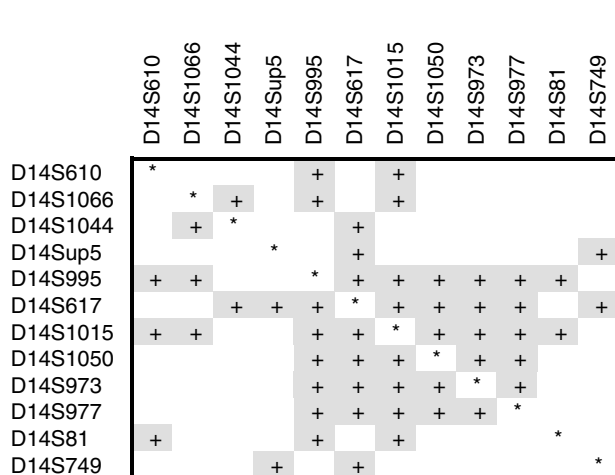
Significant LD (p=0.05) is indicated by + and gray

Table 3 Linkage disequilibrium between 12 SCA markers

	D19S395	D19S586	D19S394	D19S906	D19S1165	D19S1150	D19S840	D19SDOWN1	D19SDOWN2	D19SDOWN3	D19S226	D19S432
D19S395	*	+	+									
D19S586	+	*	+									
D19S394	+	+	*		+						+	
D19S906				*	+							
D19S1165			+	+	*							
D19S1150						*	+					
D19S840						+	*					
D19SDOWN1								*		+		
D19SDOWN2									*			
D19SDOWN3								+		*		
D19S226					+						*	
D19S432												*
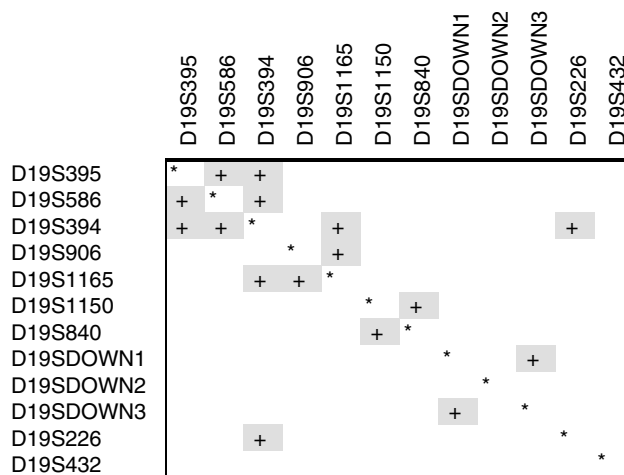
Significant LD (p = 0.05) is indicated by + and gray.

Figure 1a). Interestingly, the SCA3 families that showed the 6-allele for marker D14S617 were clustered in the eastern part of the Netherlands (province of Drenthe) and the families with the 7-allele were clustered in the western part (province of South Holland, data not shown).

SNP haplotype analysis in SCA3
In order to confirm the existence of the worldwide intragenic SNP haplotype, described by Gaspar *et al* (2001), in the Dutch SCA3 population, we typed the SNPs 669A/G, 987C/G and 118A/C in 22 SCA3 families. The A-C-A haplotype is highly conserved throughout all Dutch SCA3 chromosomes (data not shown).

Haplotype analysis in SCA6
Haplotypes were constructed for 12 SCA6 families with 13 markers (see Figure 1b) surrounding the SCA6 gene. Eight families showed a shared haplotype between the markers D19S1165 and D19S840, 3-3-6-22-3, including the SCA6 gene. This core haplotype was not observed in 80 control chromosomes. Three additional families did not show the core haplotype. Two of the three families had an identical

extended haplotype and one did not show any similarity at all.

Genealogical research showed only that families F068 and F071 were related and that most of the SCA6 families are concentrated in the province of North Holland (data not shown). Significant LD results were observed around the SCA6 gene for intragenic marker D19S1150 with marker D19S840 ($P = 0.024$) (Table 3).

The core haplotype could be created in 36% of the putative disease chromosomes of 25 isolated SCA6 patients. Truncated haplotypes derived from the initial shared haplotype were also detected in 44% of the chromosomes (data not shown). The core haplotype 3-3-6-3 was never observed in the sporadic SCA3 patients that were used as controls.

Discussion
Founder mutations have been described previously in different ethnic ADCA populations. However, no such extensive shared haplotype study had been performed in the Dutch SCA3 and SCA6 families. We were able to demonstrate the presence of one major founder SCA3

mutation in the Dutch ADCA population. In addition, geographical clustering of most of these ADCA families in a similar province supports our findings and suggests that the ADCA families can be traced back to common ancestors in particular parts of the Netherlands.

Interestingly, we noticed three groups of families with more extended haplotypes in the SCA3 region. This indicates that families within one haplotype group are more closely related than families from different groups, or than between families with less extended haplotypes. This was further confirmed by genealogical research; families from an extended haplotype group could be traced to a common link, whereas no link could be identified between families exhibiting haplotypes of different lengths.

In addition, we were also able to detect the presence of the worldwide intragenic SNP haplotype in the Dutch SCA3 families. Not surprisingly, we observed significant LD between markers surrounding both the SCA3 and 6 genes. However, the LD data were less convincing for SCA6, as a result of the small number of families, marker informativity and the more heterogenic background of the families. LD mapping within candidate regions can be a useful tool in localizing disease genes if it can be found, but in the absence of significant LD one can never be sure whether the disease locus has been missed.

We observed multiple origins of the mutation for SCA6 in the Dutch population. However, one major founder haplotype accounted for approximately 70% of the SCA6 families. This finding was strengthened by regional clustering of most of the SCA6 families in the province of North Holland.

Intergenerational stability of the CAG repeat number has been considered to be a specific molecular feature of SCA6. To date, a limited number of reports have described meiotic instability of the SCA6 repeat.[26,27] Nevertheless, we observed CAG repeat numbers ranging from 22, 23 and 25 repeats in families which belonged to the same haplotypic group. This might indicate that evidence for meiotic instability of the repeat can be found when there are sufficient numbers of generations between families. Genealogical research is an important tool in reconstructing the extended families needed to study meiotic instability.

In conclusion, we have identified the major common haplotypes for SCA3 and 6 families in the Dutch ADCA population. The combined use of both genealogical and genotype data allows linking of small pedigrees into extended families. This will be particularly useful for the identification of additional SCA loci in the small ADCA families in which no mutation has been found yet.

References

1 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.

2 Kawaguchi Y, Okamoto T, Taniwaki M *et al*: CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* 1994; **8**: 221–228.

3 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.

4 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.

5 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.

6 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.

7 Chen DH, Brkanac Z, Verlinde CL *et al*: Missense mutations in the regulatory domain of PKC gamma: a new mechanism for dominant nonepisodic cerebellar ataxia. *Am J Hum Genet* 2003; **72**: 839–849.

8 Koide R, Kobayashi S, Shimohata T *et al*: A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Hum Mol Genet* 1999; **8**: 2047–2053.

9 Nakamura K, Jeong SY, Uchihara T *et al*: SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet* 2001; **10**: 1441–1448.

10 Flanigan K, Gardner K, Alderson K *et al*: Autosomal dominant spinocerebellar ataxia with sensory axonal neuropathy (SCA4): clinical description and genetic localization to chromosome 16q22.1. *Am J Hum Genet* 1996; **59**: 392–399.

11 Ranum LP, Schut LJ, Lundgren JK *et al*: Spinocerebellar ataxia type 5 in a family descended from the grandparents of President Lincoln maps to chromosome 11. *Nat Genet* 1994; **8**: 280–284.

12 Herman-Bert A, Stevanin G, Netter JC *et al*: Mapping of spinocerebellar ataxia 13 to chromosome 19q13.3–q13.4 in a family with autosomal dominant cerebellar ataxia and mental retardation. *Am J Hum Genet* 2000; **67**: 229–235.

13 Knight MA, Kennerson ML, Anney RJ *et al*: Spinocerebellar ataxia type 15 (sca15) maps to 3p24.2–3pter: exclusion of the ITPR1 gene, the human orthologue of an ataxic mouse mutant. *Neurobiol Dis* 2003; **13**: 147–157.

14 Miyoshi Y, Yamada T, Tanimura M *et al*: A novel autosomal dominant spinocerebellar ataxia (SCA16) linked to chromosome 8q22.1–24.1. *Neurology* 2001; **57**: 96–100.

15 Worth PF, Giunti P, Gardner-Thorpe C *et al*: 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* 1999; **65**: 420–426.

16 Brkanac Z, Fernandez M, Matsushita M *et al*: Autosomal dominant sensory/motor neuropathy with ataxia (SMNA): linkage to chromosome 7q22–q32. *Am J Med Genet* 2002; **114**: 450–457.

17 Verbeek DS, Schelhaas JH, Ippel EF *et al*: Identification of a novel SCA locus (SCA19) in a Dutch autosomal dominant cerebellar ataxia family on chromosome region 1p21–q21. *Hum Genet* 2002; **111**: 388–393.

18 Swartz BE, Burmeister M, Somers JT *et al*: A form of inherited cerebellar ataxia with saccadic intrusions, increased saccadic speed, sensory neuropathy, and myoclonus. *Ann N Y Acad Sci* 2002; **956**: 441–444.

19 van de Warrenburg BP, Sinke RJ, Verschuuren-Bemelmans CC *et al*: Spinocerebellar ataxias in the Netherlands: prevalence and age at onset variance analysis. *Neurology* 2002; **58**: 702–708.

20 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.

21 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.

22 Dichgans M, Schols L, Herzog J *et al*: Spinocerebellar ataxia type 6: evidence for a strong founder effect among German families. *Neurology* 1999; **52**: 849–851.

23 Mori M, Adachi Y, Kusumi M *et al*: Spinocerebellar ataxia type 6: founder effect in Western Japan. *J Neurol Sci* 2001; **185**: 43–47.

24 Jonasson J, Juvonen V, Sistonen P *et al*: Evidence for a common Spinocerebellar ataxia type 7 (SCA7) founder mutation in Scandinavia. *Eur J Hum Genet* 2000; **8**: 918–922.

25 Benson G: Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999; **27**: 573–580.

26 Shizuka M, Watanabe M, Ikeda Y *et al*: Molecular analysis of a *de novo* mutation for spinocerebellar ataxia type 6 and (CAG)n repeat units in normal elder controls. *J Neurol Sci* 1998; **161**: 85–87.

27 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.