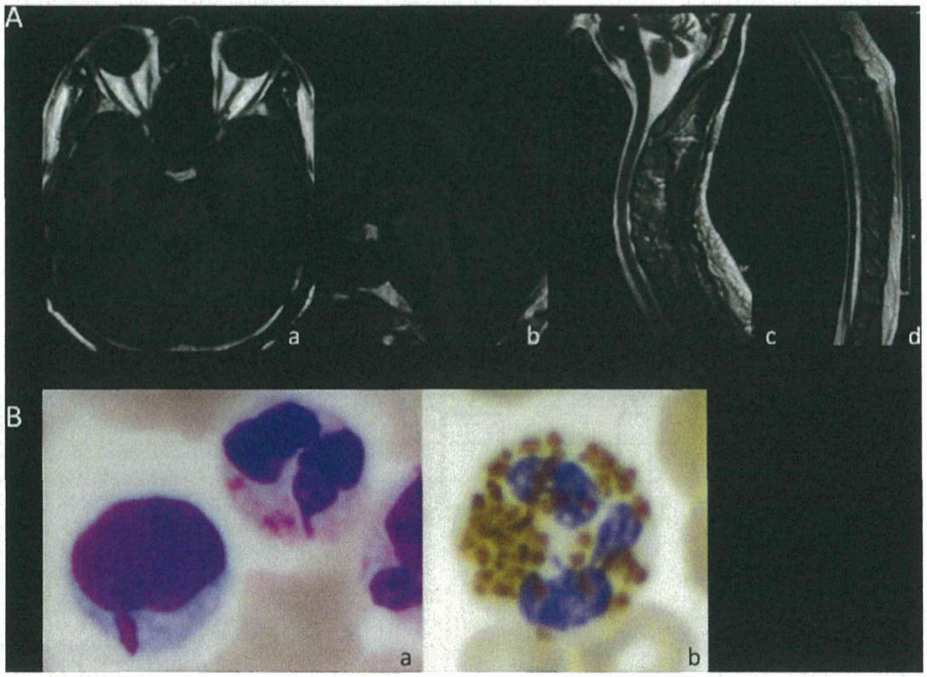

Figure 2 MRI and peripheral blood findings in the proband (II-2). (A) Brain and spinal MRI of the proband. Brain T1-weighted MRI (a, b) showed mild cerebellar atrophy, and spinal T2-weighted MRI (c, d) disclosed mild thoracic cord atrophy. (B) Peripheral blood leucocytes findings in the proband (May–Giemsa stain (a) and peroxidase stain (b)). We could observe large granules in the proband's leucocytes and large peroxidase-positive ones in granulocytes.

c.1045T>C (p.F349L) in *ANGPTL5* in the chromosome 11 candidate area with reference to dbSNP135. These three variations were validated by Sanger sequencing. No such variations were detected in the chromosome 2 and 17 areas. Subsequently, the two candidate gene variants in *FAT3* and *ANGPTL5* could be excluded as polymorphisms because these variants of the two genes were not co-segregated in the normal family members. The another candidate gene variant of *LYST* was predicted to be a functionally deleterious mutation with the prediction programs (PROVEAN, Polyphen-2 and Mutation Taster) and confirmed to be a homozygous missense mutation (c.4189T>G, p.F1397V) on Sanger sequencing in the two patients, II-1 and II-2 (figure 1C). This missense mutation was co-segregated within the family members (figure 1C) and not found in 200 Japanese control genomic DNA samples. This mutation is located at a highly conserved residue (figure 1D) within the concanavalin A (ConA)-like lectin domain (amino acids numbers 1390–1691).[8]

DISCUSSION

The present two patients exhibited spastic paraplegia, peripheral neuropathy and mild cerebellar ataxia with AR transmission. Autosomal recessive hereditary spastic paraplegia (AR-HSP) with cerebellar ataxia and neuropathy is considered to be SPG7 with the *Paraplegin* gene alteration linked to chromosome 16q24.3,[9] SPG21 with the *Maspardin* gene mutation linked to chromosome 15q22.31,[10] SPG27 linked to chromosome 10q22.1–q24.1[11] and SPG30 with the *KIF1A* mutation linked to chromosome 2q37.3.[12 13] However, linkage analysis did not show all reported HSP gene locus linkages, and whole exome sequence analysis did not disclose the *Paraplegin, Maspardin* and *KIF1A* gene mutations. According to these results, we could conclude that the causative gene in this family was not one of the previously reported HSP ones.

Through linkage analysis of the two patients' DNA and whole exome sequencing using one individual's DNA, we could identify a novel homozygous missense mutation in the *LYST* gene. This homozygous mutation was shared by the two patients. We considered the *LYST* gene mutation was causative of the neurological deficits in these two patients because it was co-segregated within their family members, located at a highly conserved amino acid, and not found in the normal controls. Moreover, large granules in leucocytes and reduced natural killer cell activity could support the diagnosis of CHS.

CHS is a rare, AR early-onset disorder characterised by severe immune deficiency, frequent bacterial infections, skin pigmentation or albinism, a bleeding tendency and progressive neurological dysfunction in most cases.[14] It is often complicated by a lymphoproliferative condition called the 'accelerated phase'. A classic diagnostic feature of CHS is enlarged granules in leucocytes, melanocytes, platelets and so forth. Most cases present in early childhood with haematological dysfunction, whereas a small number of cases with the adult form of CHS predominantly exhibit slowly progressive neurological dysfunction without apparent immunodeficiency or a bleeding tendency. Neurological involvement in CHS can include parkinsonism,[15] dementia,[16] cerebellar ataxia, peripheral neuropathy and spastic paraplegia.[17] Although the neurological phenotypes of our cases resembled those previously reported,[17] the main symptom in those patients was cerebellar ataxia that was more severe than that in our cases.

The gene responsible for CHS was identified in 1996, and was called *LYST*.[5 18] The *LYST* gene is a large gene that has 51 coding exons and an open reading frame of 11 403 kb.[6] The LYST protein, which is a large, putative cytosolic protein of 425 kDa (3801 amino acids), is ubiquitously expressed and involved in control of the exocytosis of secretory lysosomes.[5 19] The LYST protein has a BEACH (named after BEige And

Chédiak–Higashi) domain (amino acid numbers 3132–3422),[5] Trp-Asp (WD) 40 repeats (amino acid numbers 3477-3778) and a ConA-like lectin domain (amino acid numbers 1390–1691).[8] The LYST protein has been proposed to act as a scaffold protein in the mediation of fusion or a fission event of vesicles.[20] The mutation in this family (p.F1397V) is located within the ConA-like lectin domain. This domain could be involved in oligosaccharide binding associated with protein trafficking and sorting along the secretory pathway.[8]

Recently, *Drosophila* with a gene mutation of an *LYST* homologue was revealed to exhibit impaired autophagy.[21] The loss of function of some HSP-related proteins, TECPR2[22] and spastizin,[23] caused autophagic dysfunction and induced spastic paraplegia. Therefore, autophagic impairment might have resulted in spastic paraplegia in the CHS patients.

Karim *et al*[24] found apparent genotype–phenotype correlations in CHS, that is, that severe childhood CHS involved a functionally null mutation, whereas missense mutations were seen only in the two later-onset forms. They reported four missense mutations, two of which are located in the ConA-like lectin domain. Our cases correspond to late-onset, slowly progressive neurological CHS with a missense mutation of the *LYST* gene. According to the information on the Japanese cases in the literature,[24] the Japanese adult CHS cases with *LYST* missense mutations (R1563H or V1999D) showed spastic paraplegia, gaze nystagmus and diminished ATRs. Thus, their phenotypes were similar to those of our cases. Moreover, as far as we know, this family had one of the oldest adult CHS cases (onset of 58 years) with a *LYST* gene mutation in the literature. To date, a 57-year-old man has been reported who suffered from sensorimotor polyneuropathy and muscle wasting with a heterozygous *LYST* gene mutation (p.Y2026X).[25]

In this family, the proband showed spastic paraplegia dominantly as well as neuropathy and mild cerebellar ataxia, whereas the brother mainly showed peripheral neuropathy with a positive Babinski sign, cerebellar ataxia and dementia. These two patients did not exhibit parkinsonism. The phenotypic variety in this family might be explained by environmental factors or other modifier gene mutations.

In summary, we could diagnose these patients as having adult CHS presenting spastic paraplegia with neuropathy and cerebellar ataxia. As far as we know, this family includes one of the oldest adult CHS cases in the literature. The clinical spectrum of CHS is broader than previously recognised, and this family shows phenotypic variability. Adult CHS must be considered in the differential diagnosis of AR-HSPs. The linkage analysis and exome sequencing were useful for identifying the causative mutation in this family.

Acknowledgements We thank Professor Miyajima (Department of Neurology, Hamamatsu Medical University, Hamamatsu, Japan) for the Japanese adult CHS patient's information in a reference.[24]

Contributors HS was responsible for conception and design of the work, acquisition, analysis or interpretation of data, drafting the work and revising the work critically for important intellectual content. JH, TN and MN were responsible for acquisition and analysis of data. IN was responsible for analysis or interpretation of data. MY, KN, KY, S-il, HI, YF and YT were responsible for acquisition, analysis or interpretation of data. JG, ST and YT were responsible for conception and design of the work, interpretation of data and revising the work critically for important intellectual content.

Funding This work was supported by a Grant-in-Aid for Scientific Research (C) (23591253 to HS) and MEXT KAKENHI Grant Number 221S0002 from The Ministry of Education, Culture, Sports, Science and Technology in Japan. This work was also supported by a grant from the Research Committee for Ataxic Diseases (YT and HS) of the Ministry of Health, Labor and Welfare, Japan.

Competing interests None.

Patient consent Obtained.

Ethics approval The Institutional Review Boards of the Jichi Medical University, Shinshu University, University of Tokyo, and University of Yamanashi.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement Exome sequencing data are available (DDBJ Sequence Read Archive (DRA) accession number DRA000961).

REFERENCES

1 Harding AE. Classification of the hereditary ataxias and paraplegias. *Lancet* 1983;1:1151–5.
2 Hazan J, Fonknechten N, Mavel D, *et al*. Spastin, a new AAA protein, is altered in the most frequent form of autosomal dominant spastic paraplegia. *Nat Genet* 1999;23:296–303.
3 Stevanin G, Santorelli FM, Azzedine H, *et al*. Mutations in SPG11, encoding spatacsin, are a major cause of spastic paraplegia with thin corpus callosum. *Nat Genet* 2007;39:366–72.
4 Salinas S, Proukakis C, Crosby A, *et al*. Hereditary spastic paraplegia: clinical features and pathogenetic mechanisms. *Lancet Neurol* 2008;7:1127–38.
5 Nagle DL, Karim MA, Woolf EA, *et al*. Identification and mutation analysis of the complete gene for Chediak-Higashi syndrome. *Nat Genet* 1996;14:307–11.
6 Fukuda Y, Nakahara Y, Date H, *et al*. SNP HiTLink: a high-throughput linkage analysis system employing dense SNP data. *BMC Bioinformatics* 2009;10:121.
7 Gudbjartsson DF, Thorvaldsson T, Kong A, *et al*. Allegro version 2. *Nat Genet* 2005;37:1015–16.
8 Burgess A, Mornon JP, de Saint-Basile G, *et al*. A concanavalin A-like lectin domain in the CHS1/LYST protein, shared by members of the BEACH family. *Bioinformatics* 2009;25:1219–22.
9 Casari G, De Fusco M, Ciarmatori S, *et al*. Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* 1998;93:973–83.
10 Simpson MA, Cross H, Proukakis C, *et al*. Maspardin is mutated in mast syndrome, a complicated form of hereditary spastic paraplegia associated with dementia. *Am J Hum Genet* 2003;73:1147–56.
11 Meijer IA, Cossette P, Roussel J, *et al*. A novel locus for pure recessive hereditary spastic paraplegia maps to 10q22.1-10q24.1. *Ann Neurol* 2004;56:579–82.
12 Klebe S, Azzedine H, Durr A, *et al*. Autosomal recessive spastic paraplegia (SPG30) with mild ataxia and sensory neuropathy maps to chromosome 2q37.3. *Brain* 2006;129:1456–62.
13 Erlich Y, Edvardson S, Hodges E, *et al*. Exome sequencing and disease-network analysis of a single family implicate a mutation in KIF1A in hereditary spastic paraparesis. *Genome Res* 2011;21:658–64.
14 Kaplan J, De Domenico I, Ward DM. Chediak-Higashi syndrome. *Curr Opin Hematol* 2008;15:22–9.
15 Pettit RE, Berdal KG. Chediak-Higashi syndrome. Neurologic appearance. *Arch Neurol* 1984;41:1001–2.
16 Uyama E, Hirano T, Ito K, *et al*. Adult Chediak-Higashi syndrome presenting as parkinsonism and dementia. *Acta Neurol Scand* 1994;89:175–83.
17 Sheramata W, Kott HS, Cyr DP. The Chediak-Higashi-Steinbrinck syndrome. Presentation of three cases with features resembling spinocerebellar degeneration. *Arch Neurol* 1971;25:289–94.
18 Barbosa MD, Nguyen QA, Tchernev VT, *et al*. Identification of the homologous beige and Chediak-Higashi syndrome genes. *Nature* 1996;382:262–5.
19 Huynh C, Roth D, Ward DM, *et al*. Defective lysosomal exocytosis and plasma membrane repair in Chediak-Higashi/beige cells. *Proc Natl Acad Sci USA* 2004;101:16795–800.
20 Tchernev VT, Mansfield TA, Giot L, *et al*. The Chediak-Higashi protein interacts with SNARE complex and signal transduction proteins. *Mol Med* 2002;8:56–64.
21 Rahman M, Haberman A, Tracy C, *et al*. Drosophila mauve mutants reveal a role of LYST homologs late in the maturation of phagosomes and autophagosomes. *Traffic* 2012;13:1680–92.
22 Oz-Levi D, Ben-Zeev B, Ruzzo EK, *et al*. Mutation in TECPR2 reveals a role for autophagy in hereditary spastic paraparesis. *Am J Hum Genet* 2012;91:1065–72.
23 Vantaggiato C, Crimella C, Airoldi G, *et al*. Defective autophagy in spastizin mutated patients with hereditary spastic paraparesis type 15. *Brain* 2013;136:3119–39.
24 Karim MA, Suzuki K, Fukai K, *et al*. Apparent genotype-phenotype correlation in childhood, adolescent, and adult Chediak-Higashi syndrome. *Am J Med Genet* 2002;108:16–22.
25 Mottonen M, Lanning M, Baumann P, *et al*. Chediak-Higashi syndrome: four cases from Northern Finland. *Acta Paediatr* 2003;92:1047–51.
26 Robinson JT, Thorvaldsdottir H, Winckler W, *et al*. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24–6.

doi:10.1093/brain/awu174 Brain 2014: 137; 2444–2455 | 2444

BRAIN
A JOURNAL OF NEUROLOGY

Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes

Sophie Tezenas du Montcel,[1,2,3] Alexandra Durr,[4,5] Peter Bauer,[6] Karla P. Figueroa,[7] Yaeko Ichikawa,[8] Alessandro Brussino,[9] Sylvie Forlani,[5] Maria Rakowicz,[10] Ludger Schöls,[11,12] Caterina Mariotti,[13] Bart P. C. van de Warrenburg,[14] Laura Orsi,[15] Paola Giunti,[16] Alessandro Filla,[17] Sandra Szymanski,[18] Thomas Klockgether,[19] José Berciano,[20] Massimo Pandolfo,[21] Sylvia Boesch,[22] Bela Melegh,[23] Dagmar Timmann,[24] Paola Mandich,[25] Agnès Camuzat,[5] Clinical Research Consortium for Spinocerebellar Ataxia (CRC-SCA),* the EUROSCA network,* Jun Goto,[8] Tetsuo Ashizawa,[26] Cécile Cazeneuve,[4] Shoji Tsuji,[8] Stefan-M. Pulst,[7] Alfredo Brusco,[9] Olaf Riess,[6] Alexis Brice[4,5] and Giovanni Stevanin[4,5,27]

1 Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Univ Paris 06, UMR_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France
2 INSERM, UMR_S 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France
3 AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Biostatistics Unit, Paris, F-75013, France
4 AP-HP, Groupe Hospitalier Pitié-Salpêtrière, Department of Genetics and Cytogenetics, F-75013, Paris, France
5 Inserm U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, F-75013, Paris, France
6 Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany
7 Department of Neurology, University of Utah, Salt Lake City, USA
8 Department of Neurology, University of Tokyo, Graduate School of Medicine, Tokyo, Japan
9 University of Torino, Department of Medical Sciences, and Medical Genetics Unit, Az. Osp. 'Città della Salute e della Scienza', Torino, Italy
10 Institute of Psychiatry and Neurology Warsaw, Sobieskiego 9, 02-957 Warsaw, Poland
11 Department of Neurology and Hertie-Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany
12 German Centre of Neurodegenerative Diseases (DZNE), Tübingen, Germany
13 SOSD Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS, Istituto Neurologico 'Carlo Besta', Milan, Italy
14 Department of Neurology, Donders Institute for Brain, Cognition, and Behaviour, Radbound University Medical Centre, Nijmegen, The Netherlands
15 Neurologic Division I, Department of Neuroscience and Mental Health, AOU Città della Salute e della Scienza, Torino, Italy
16 Institute of Neurology, Department of Molecular Neuroscience, UCL, Queen Square, London, UK
17 Department of Neurological Sciences, Federico II University, Naples, Italy
18 Department of Neurology, St. Josef Hospital, University Hospital of Bochum, Bochum, Germany
19 Department of Neurology, University Hospital of Bonn, Bonn, Germany
20 Department of Neurology, University Hospital 'Marqués de Valdecilla', UC, IDIVAL and CIBERNED, 39008 Santander, Spain
21 Department of Neurology, ULB-Hôpital Erasme, Université Libre de Bruxelles, CP 231, Campus Plaine, ULB, Brusssels, Belgium
22 Department of Neurology, Medical University Innsbruck, Innsbruck, Austria
23 Department of Medical Genetics, and Szentagothai Research Centre, University Pécs, Hungary
24 Department of Neurology, University Clinic Essen, University of Duisburg-Essen, Essen, Germany
25 Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal Child Health, University of Genova, and U.O. Medical Genetics of IRCCS AOU S. Martino Institute, Genova, Italy
26 Department of Neurology and McKnight Brain Institute, University of Florida, Gainesville, Florida, USA
27 Ecole Pratique des Hautes Etudes, heSam Université, laboratoire de neurogénétique, ICM, Groupe Hospitalier Pitié-Salpêtrière, F-75013 Paris, France

*See Appendix 1 for details of the Clinical Research Consortium for Spinocerebellar Ataxia and EUROSCA network.

Received March 17, 2014. Revised May 20, 2014. Accepted May 23, 2014. Advance Access publication June 26, 2014

Correspondence to: Sophie Tezenas du Montcel,
Unité de Biostatistiques et Information Médicale,
Département de Santé Publique - Bâtiment Mazarin,
Groupe Hospitalier Pitié-Salpêtrière - 47-83 Bd de l'Hôpital,
75651 Paris Cedex 13, France
E-mail: sophie.tezenas@psl.aphp.fr

Correspondence may also be addressed to: Alexis Brice, Institut du Cerveau et de la Moelle épinière, Groupe Hospitalier Pitié-Salpêtrière - 47-83 Bd de l'Hôpital, 75651 Paris Cedex 13, France E-mail: alexis.brice@upmc.fr

Polyglutamine-coding (CAG)n repeat expansions in seven different genes cause spinocerebellar ataxias. Although the size of the expansion is negatively correlated with age at onset, it accounts for only 50–70% of its variability. To find other factors involved in this variability, we performed a regression analysis in 1255 affected individuals with identified expansions (spinocerebellar ataxia types 1, 2, 3, 6 and 7), recruited through the European Consortium on Spinocerebellar Ataxias, to determine whether age at onset is influenced by the size of the normal allele in eight causal (CAG)n-containing genes (*ATXN1–3, 6–7, 17, ATN1* and *HTT*). We confirmed the negative effect of the expanded allele and detected threshold effects reflected by a quadratic association between age at onset and CAG size in spinocerebellar ataxia types 1, 3 and 6. We also evidenced an interaction between the expanded and normal alleles in *trans* in individuals with spinocerebellar ataxia types 1, 6 and 7. Except for individuals with spinocerebellar ataxia type 1, age at onset was also influenced by other (CAG)n-containing genes: *ATXN7* in spinocerebellar ataxia type 2; *ATXN2, ATN1* and *HTT* in spinocerebellar ataxia type 3; *ATXN1* and *ATXN3* in spinocerebellar ataxia type 6; and *ATXN3* and *TBP* in spinocerebellar ataxia type 7. This suggests that there are biological relationships among these genes. The results were partially replicated in four independent populations representing 460 Caucasians and 216 Asian samples; the differences are possibly explained by ethnic or geographical differences. As the variability in age at onset is not completely explained by the effects of the causative and modifier sister genes, other genetic or environmental factors must also play a role in these diseases.

Keywords: spinocerebellar ataxia; age at onset; trinucleotide repeat; modifier
Abbreviation: SCA = spinocerebellar ataxia

Introduction

Autosomal dominant cerebellar ataxias, also known as spinocerebellar ataxias (SCA), are clinically and genetically heterogeneous neurodegenerative diseases. Major advances have been made since the 1990s in our understanding of their causes. So far, mutations in 20 genes have been identified as responsible for the diseases. They comprise conventional mutations, non-coding nucleotide expansions and coding (CAG)n expansions (Schols *et al.*, 2004; Durr, 2010; Matilla-Dueñas *et al.*, 2014). SCA1, SCA2, SCA3, SCA6, SCA7, SCA12, SCA17 and dentato-rubro-pallidoluysian atrophy (DRPLA) are caused by (CAG)n repeat expansions in the *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *PPP2R2B*, *TBP* and *ATN1* genes, respectively. All lead to the expansion of a polyglutamine tract in the corresponding proteins (Stevanin *et al.*, 2000). A polyglutamine tract also contributes to the disease process in SCA8 expansion carriers (Ikeda *et al.*, 2008).

The so-called polyglutamine ataxias share many features with each other, as well as with Huntington's disease and Kennedy syndrome (spinobulbar muscular atrophy): a negative relation between age at onset and the number of repeats in the expansion; in general, a more severe disease with larger expansions; and a phenotype that is variable in affected individuals with the same genotype due, to some extent, to the size of the expansion (Stevanin *et al.*, 2000; Orr and Zoghbi, 2007). However, the repeat length only explains 50–80% of the variability of age at onset, suggesting that other genetic factors contribute to the variability, as shown recently in Huntington's disease (van Dellen and Hannan, 2004).

The involvement of other 'familial,' thus possibly genetic, factors was suggested early on (DeStefano *et al.*, 1996), and was confirmed more recently in a large Dutch and French cohort (van de Warrenburg *et al.*, 2005) and in a large Cuban SCA2 population (Pulst *et al.*, 2005). Normal polymorphic stretches on the unaffected allele in *trans* have been shown to affect SCA1, SCA3, and SCA6 (Durr *et al.*, 1996; van de Warrenburg *et al.*, 2005) and Huntington disease (Aziz *et al.*, 2009). The effects of repeat alleles in other, non-causal, SCA genes have also been examined in a few studies. One found that disease onset in cases with SCA2 with long normal CAG repeats in the *CACNA1A* gene was earlier than would be expected from the size of the CAG expansion in *ATXN2* (Pulst *et al.*, 2005). In a Brazilian cohort, the age at onset of SCA2 was earlier in affected subjects with longer normal CAG repeats in *ATXN3* (de Castilhos *et al.*, 2014).

To find other factors involved in the variability of the age at onset and analyse the functional relationships among SCA genes, we performed a regression analysis in 1255 affected subjects with known types of SCA to determine the influence of the size of the normal alleles in eight polymorphic (CAG)n-containing genes

(*ATN1*, *HTT*, *TBP*, *CACNA1A*, *ATXN1*, *2*, *3* and *7*) on the age at onset. This SCA cohort, recruited through the integrated European project on the spinocerebellar ataxias (EUROSCA) consortium, is the largest to have been studied so far. The study was replicated in 676 subjects originating from four independent cohorts from the USA, Japan, France and Italy.

Materials and methods

Subjects

Affected subjects (n = 1255), at least 78% of which were of European Caucasian ancestry (ancestry unknown in 15%), were recruited by the EUROSCA study group (http:www.eurosca.org) from 10 countries (Austria, Belgium, France, Germany, Hungary, Italy, Netherlands, Poland, Spain and UK). All subjects with genetically determined CAG repeat expansions in a causal gene (e.g. *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*) were invited to enter the cohort. To recruit the largest pedigrees possible, participating subjects were asked to inform their relatives about the study. Only living affected subjects were included. Disease onset was defined by the appearance of gait disorders (Globas *et al.*, 2008). Affected subjects were included in the database with age at onset as indicated by themselves during their examination by the neurologist, as indicated in their medical records. Blood samples were collected with informed consent according to ethical committees in each country.

The four independent cohorts recruited to replicate the results included a series of 216 Japanese subjects and three Caucasian groups of 291, 93 and 76 subjects from the USA, France and Italy, respectively.

Genotype analysis and classification

The genotypes of individuals in the EUROSCA cohort were first determined at the centre where the individual was recruited. To homogenize sizing of the CAG repeats, all EUROSCA subjects were re-genotyped in a central laboratory (Tübingen, Germany). Only subjects in whom the second genotype matched the reported CAG expansion (±2 CAG repeats) were included in the current study. The participating subjects were also genotyped in the central laboratory for eight other polymorphic (CAG)n-containing genes (*ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *ATN1* and *HTT*).

Genotyping of the EUROSCA and French subjects was performed by multiplex PCR amplification of the CAG tracts (primers and conditions available upon request to the authors), and the genotypes were resolved by capillary electrophoresis in a CEQ8000 automated sequencer (Beckman Coulter) followed by analysis with CEQ 8.0 software, or on an ABI3730 sequencer followed by analysis with GeneMapper software (Applied Biosystems). Repeats in the Italian, US, and Japanese cohorts were sized by independent PCR amplifications resolved in an automated sequencer using classical procedures.

An allele in the pathological range was designated the 'expanded' allele according to the threshold indicated in Table 1. At loci without expansions, the allele containing the larger repeat was designated as the 'longer' allele, the other was termed the 'shorter' allele. In the statistical models, the shorter and the longer alleles were considered separately. With respect to the multimodal or skewed distribution shown in Fig. 1, non-expanded normal repeats in the *ATXN2*, *ATXN3*, *ATXN7* and *CACNA1A* genes were classified as short, medium, short intermediate or intermediate (Table 1). The *ATXN2* genotypes were divided into four classes as proposed previously (Elden *et al.*, 2010): (i) at least one short allele; (ii) homozygous medium 22 CAG alleles; (iii) at least one intermediate allele; and (iv) at least one short intermediate allele with or without a medium 22 CAG allele (Table 1). The same strategy was used for *ATXN3* and *ATXN7* genotypes that were divided into three classes: (i) at least one short allele; (ii) homozygous medium alleles; and (iii) at least one intermediate allele with a medium allele or homozygous intermediate alleles (Table 1). The *CACNA1A* genotypes were divided into three classes: (i) homozygous short alleles; (ii) heterozygous intermediate alleles; and (iii) homozygous intermediate alleles (Table 1). For the other genes (*ATXN1*, *TBP*, *ATN1* and *HTT*), the sizes of the shorter and the longer alleles were considered separately. The statistical analysis also took into account the interaction between the two alleles, the mean of the length of the two alleles and the difference between the lengths of the two alleles.

Table 1 Classification of SCA gene alleles according to the number of CAG repeats

Alleles	*ATXN1*	*ATXN2*	*ATXN3*	*CACNA1A*	*ATXN7*
Short		<22	<16	<9	<10
Medium		22	16-24	-	10–11
Intermediate short		23–26	-	-	-
Intermediate		27–29	25–35	9–16	12–14
Expanded	⩾39	⩾33	⩾47	⩾20	⩾36

An allele in the pathological range was designated the 'expanded' allele if it contained at least the number of CAG repeats in the table.
Non-expanded normal repeats in the *ATXN2*, *ATXN3*, *ATXN7* and *CACNA1A* genes were classified as short, medium, short intermediate or intermediate according to the threshold in the table. The normal allele of the *ATXN1* gene was considered to be a quantitative variable, therefore no thresholds are given. *TBP*, *ATN1* and *HTT* repeats were also considered to be linear values; no cases had mutations in these genes in our cohorts.

Statistical analyses

The logarithmically (decimal) transformed ages at onset were treated as dependent variables. Univariate linear regression analyses were first performed to determine the effect on age at onset of (i) the expanded allele (linear and quadratic effect); (ii) the normal allele in *trans* in addition to the expanded allele; and (iii) the interaction between both alleles of the causative SCA gene. In addition, effects of the normal (CAG)n repeats in seven other polymorphic in non-causal genes were added to this model. To take into account a possible influence of the genotype at a given gene, the interaction between the two alleles and a combination of both alleles were considered. For each allele or combination of alleles, three models were tested: (i) a model taking into account the polymorphic (CAG)n tract in each of the seven additional genes only; (ii) a model combining the (CAG)n tracts at each gene in addition to the repeat in the causative SCA gene; and (iii) a model testing the interaction between the (CAG)n tracts in each gene and the repeats in the causative gene to determine whether the effects of the additional genes differed as a function of the pathological (CAG)n tract. The study was replicated in the independent cohorts. The determinant coefficient (R^2) is the percentage of the variance explained by a given model. An adjusted R^2 was computed to take into account the number of parameters included in the model. All the final models were tested for a family effect. Because the results were similar (P = not significant), only models without familial effects are reported. Because most of the EUROSCA patients were of

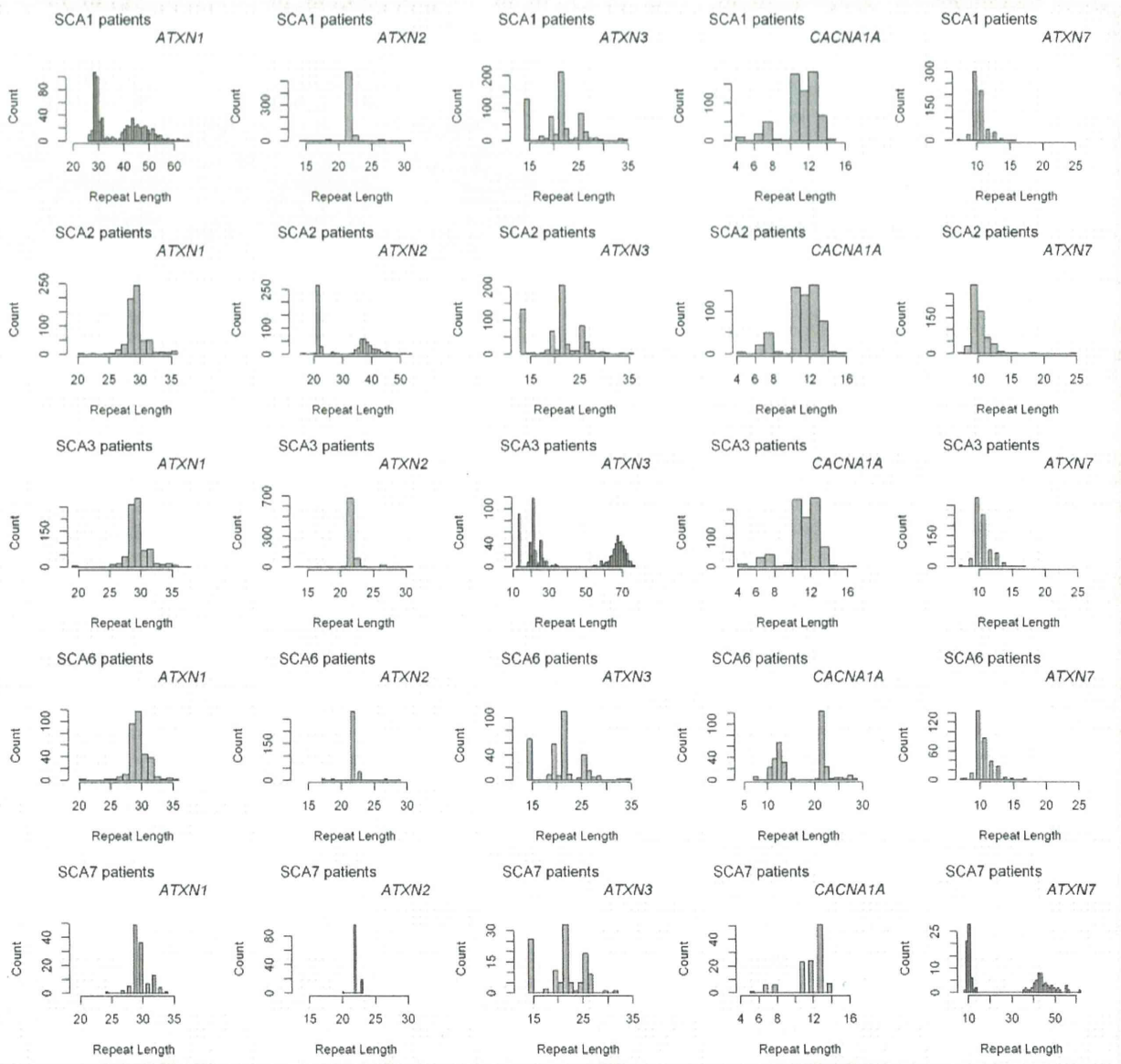

Figure 1 Histograms of the normal and expanded allele repeat lengths for *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A* and *ATXN7* in affected individuals of the EUROSCA kindreds.

Caucasian ancestry and as only 8% were clearly of non-Caucasian origin, ethnicity was not considered as a parameter in this study. The replication cohorts, which diverged in ethnicities and geographical origins, were considered independently. Similarly, the geographical origin had no influence on the results obtained in this study (data not shown).

To verify the validity of the model obtained, residuals were inspected and three extreme outliers were eliminated. They corresponded to (i) a subject with SCA2 who had onset when he was 1-year-old and the following genotypes: *ATXN2* gene, 22/36 repeats; *ATXN7* gene, 10/12 repeats; (ii) a subject with SCA6 who had onset when he was 16 and the following genotypes: *CACNA1A* gene, 11/25 repeats; *ATXN1*, 29/32 repeats; *ATXN3*, 23/25 repeats; and (iii) a subject with SCA7 who had onset when he was aged 5 and the following genotypes: *ATXN7* gene, 15/92 repeats; *ATXN3* gene, 14/26; *TBP* gene, 36/37. The residual plots were reconsidered after these exclusions.

All reported *P*-values are two-tailed. A type I error rate of 5% was used. Analyses were performed with the SAS 9.2 statistical package (SAS Institute Inc.).

Results

Cohorts and repeat length distribution in affected subjects

The EUROSCA cohort comprised 1255 affected subjects from 775 families with a definite diagnosis (25% SCA1, 23% SCA2, 32%

Table 2 Description of the EUROSCA cohort analysed in the modifier study

Genetic entity (mutated gene)		SCA1 (*ATXN1*)	SCA2 (*ATXN2*)	SCA3 (*ATXN3*)	SCA6 (*CACNA1A*)	SCA7 (*ATXN7*)	P-value
Affected subjects, *n*	1255	319	309	403	165	59	
Gender, female	595 (48%)	148 (46%)	145 (47%)	201 (50%)	81 (49%)	20 (34%)	0.24
Transmitting parent, maternal transmission	543 (52%)	133 (48%)	152 (56%)	164 (48%)	70 (61%)	24 (62%)	0.019
Age at onset	Mean (years ± SD)	38 ± 11	36 ± 13	40 ± 12	53 ± 11	30 ± 14	<0.0001
	Range (years)	11–75	7–71	10–78	24–77	8–71	
Expanded CAG repeat length	Mean (CAG ± SD)	47 ± 5	39 ± 3	68 ± 4	23 ± 2	45 ± 5	–
	Range (CAG)	39–66	33–54	47–77	21–29	36–62	
Non-expanded CAG repeat length	Mean (CAG ± SD)	30 ± 2	22 ± 1	21 ± 5	13 ± 1	11 ± 1	–
	Rang (CAG)	26–37	15–29	12–35	7–16	8–14	

SD = standard deviation; F = female.

SCA3, 16% SCA6, 4% SCA7). The mean number of affected subjects per family was 1.6 ± 1.5 (min = 1; max = 17); there were no statistical differences among the SCAs. The sex ratio was close to 1 and similar in all the SCAs (Table 2). In each SCA group, the lengths of the longer and shorter CAG repeat alleles were distributed as expected (Table 1 and Figs 1 and 2) from previous reports (Kremer *et al.*, 1994; Deka *et al.*, 1995; Takano *et al.*, 1998; Giunti *et al.*, 1999; Fujigasaki *et al.*, 2001; Silveira *et al.*, 2002).

The replication cohorts (Supplementary Table 1) were composed of 291 North-American cases ($n = 51$ SCA1, $n = 6$ SCA2, $n = 110$ SCA3, $n = 67$ SCA6), 216 Japanese cases ($n = 126$ SCA3, $n = 90$ SCA6), 76 Italian cases ($n = 24$ SCA1, $n = 52$ SCA2) and 93 French cases ($n = 25$ SCA1, $n = 24$ SCA2, $n = 44$ SCA3). Repeat lengths in the replication series were similar to those of the EUROSCA cohort (P = not significant).

For the following statistics, as they did not influence the results (see 'Materials and methods' section), we considered all cases regardless of their familial relation or their ethnicity or country of origin. Various models were tested, but only the most significant and relevant ones are shown here.

Effect of the causative spinocerebellar ataxia gene on age at onset in EUROSCA individuals

The age at onset was similar in the different SCAs (SCA1: 38 ± 11, SCA2: 36 ± 13, SCA3: 40 ± 12), except for SCA7 in which onset was earlier (30 ± 14 years) and SCA6 in which onset was later (53 ± 11 years) ($P < 0.0001$) (Table 2). Age at onset did not differ according to the gender of the subject or of the transmitting parent, except for SCA3 in which onset was earlier in subjects with a maternally inherited expansion (maternal: 37 ± 11 years, paternal: 41 ± 12, $P = 0.004$).

The log age at onset was determined by the size of the expanded allele in all of the SCAs (Table 3); the determinant coefficients varied from 0.32 (SCA6) to 0.80 (SCA7), as reported previously in a subset of this series (van de Warrenburg *et al.*, 2005). Significant quadratic effects of the expanded alleles were found for *ATXN2* (positive effect, $P = 0.0002$), *ATXN3* (negative effect $P < 0.0001$) (Fig. 3) and *CACNA1A* (positive effect, $P = 0.001$) (Table 3). Similar results were obtained in the SCA3 cases of the USA (beta = −0.0019 ± 0.00044, $P < 0.0001$), Japan (beta = −0.0014 ± 0.00024, $P < 0.0001$) and France (beta = −0.0016 ± 0.00041, $P = 0.0003$) cohorts (Fig. 3).

A significant interaction between the non-expanded (wild-type) allele in *trans* and age at onset was also evidenced in subjects with SCA1, SCA6 and SCA7 (Table 3). Intermediate normal alleles interacting with the expanded allele had stronger effects (decreased age at onset) than small normal alleles in SCA1 and SCA6 (interaction term between normal and expanded alleles: SCA1 subjects −0.0016 ± 0.0006, $P = 0.0088$; SCA6 cases −0.0056 ± 0.0021, $P = 0.0069$). In SCA7, the opposite was observed; short or medium (<12 repeats) normal alleles interacting with the expanded allele had a stronger effect (decreased age at onset) than large normal alleles (12 repeats or more) ($P = 0.044$). Only nine cases with SCA7 had intermediate alleles. Significant effects of the normal alleles in *trans* were not replicated in the other cohorts, which were smaller in size.

Additional effects on age at onset of non-causative polymorphic (CAG)n-containing genes

The effects of seven polymorphic (CAG)n repeats were tested in combination with the effect of the causative genes in EUROSCA subjects. None of the non-causative trinucleotide repeats tested influenced the age at onset in SCA1 (Supplementary Table 2).

In contrast, other genes affected the age at onset in SCA2, 3, 6 and 7 subjects in addition to the linear, quadratic, linear, and linear effects of the expanded alleles, respectively.

Spinocerebellar ataxia type 2

Age at onset in SCA2 subjects was influenced by the number of CAG repeats in the *ATXN7* gene (Supplementary Table 3 and Fig. 4) ($R^2 = 63.6\%$). Subjects with an *ATXN7* allele containing >12 repeats (33%) had a log age at onset of −0.0425 ± 0.013 ($P = 0.0014$), which was earlier than subjects with ⩽12 repeats. This effect was independent of the size of the *ATXN2* expansion. These results were not replicated in the other cohorts, although a

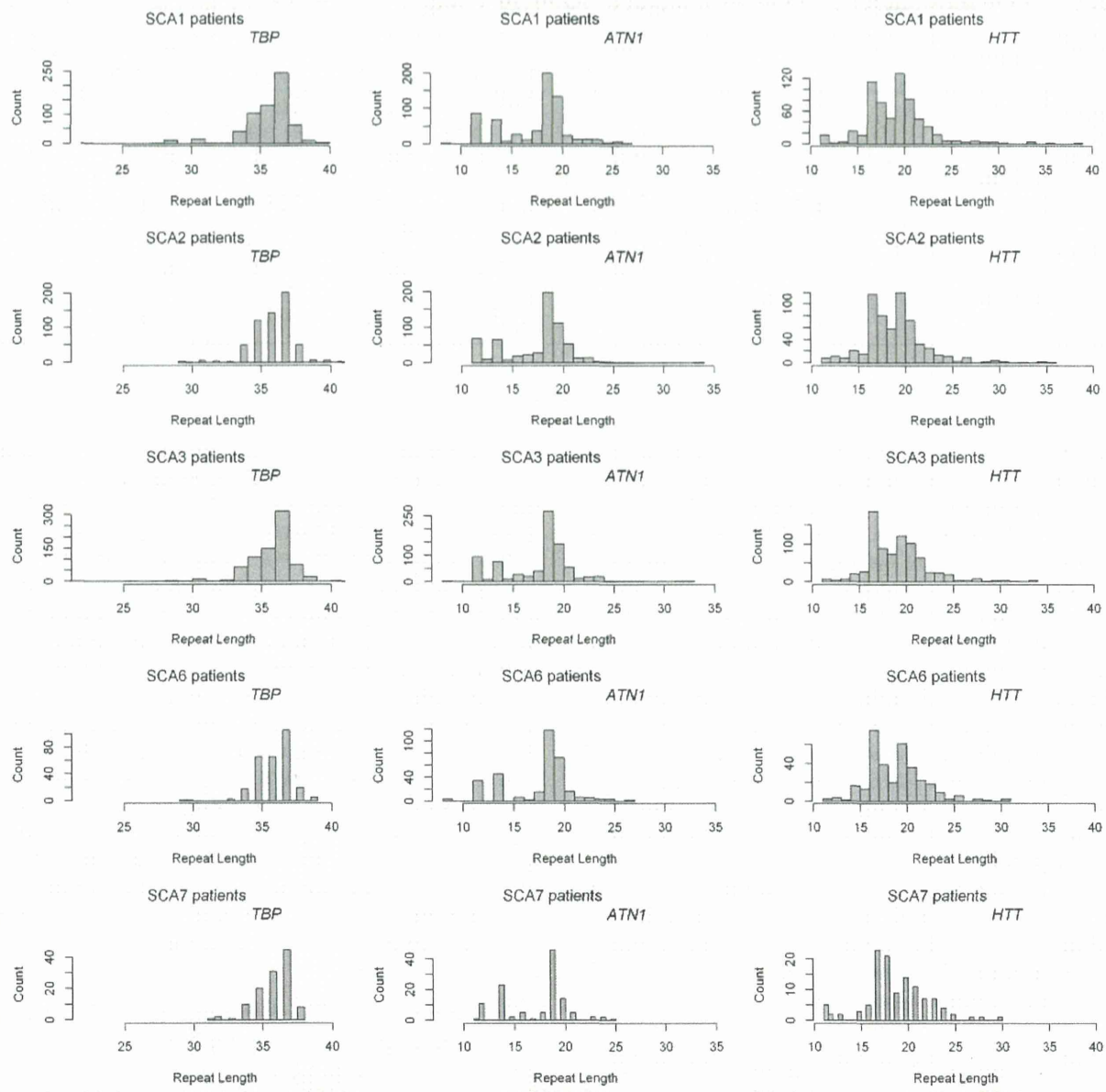

Figure 2 Histograms of the normal and expanded allele repeat lengths for *TBP*, *ATN1* and *HTT* in affected individuals of the EUROSCA kindreds.

non-significant tendency was observed in the American and French cases with SCA2 (Fig. 4).

Spinocerebellar ataxia type 3

In the SCA3 subgroup, age at onset was influenced by the number of repeats in the *ATXN2* (longer intermediate allele), *ATN1* (longer wild-type allele) and *HTT* (shorter allele) genes (Supplementary Table 4). SCA3 subjects with an intermediate *ATXN2* allele (27–29) (7% of the subjects) had an earlier age at onset than subjects with shorter *ATXN2* alleles (0.073 ± 0.017 earlier log age at onset, $P < 0.0001$, $R^2 = 61.6\%$). The larger *ATN1* allele interacting with the *ATXN3* expansion (that correlates negatively with age at onset in SCA3), also decreased the age at onset in SCA3 cases ($P = 0.036$, $R^2 = 60.2\%$). On the contrary, the shorter *HTT* allele ($P = 0.038$, $R^2 = 60.5\%$), interacting with the *ATXN3* expansion, increased the age at onset in subjects with SCA3. The effects of *ATN1*, *HTT* and *ATXN2* were not replicated. This might be due, in the case of *ATXN2*, to the rarity of intermediate *ATXN2* alleles in the subjects with SCA3 in the replication cohorts: Japan (0%), France (2%, $n = 1$), USA (5%, $n = 6$) compared to the EUROSCA group (7%, $n = 28$) ($P = 0.0034$). Similarly, in the EUROSCA population, intermediate *ATXN2* alleles (Table 1) were

Table 3 Effect of expanded and normal alleles of causative genes on log age at onset

Genotypes	SCA1	SCA2	SCA3	SCA6	SCA7
MODEL with the expanded allele only					
Linear model					
Expanded allele	-0.021 ± 0.001 (<0.0001)	-0.045 ± 0.002 (<0.0001)	-0.024 ± 0.001 (<0.0001)	-0.0351 ± 0.0040 (<0.0001)	-0.037 ± 0.002 (<0.0001)
R^2	0.652	0.620	0.504	0.326	0.807
Adjusted R^2	0.651	0.619	0.503	0.322	0.804
Quadratic model					
Expanded allele (quadratic term)	$+2.9 \times 10^{-5} \pm 12.5 \times 10^{-5}$ (0.81)	$+1.4 \times 10^{-3} \pm 0.4 \times 10^{-3}$ (0.0002)	$-1.4 \times 10^{-3} \pm 0.1 \times 10^{-3}$ (<0.0001)	$+8.0 \times 10^{-3} \pm 2.4 \times 10^{-3}$ (0.001)	$4.3 \times 10^{-4} \pm 3.3 \times 10^{-4}$ (0.19)
R^2	0.652	0.637	0.594	0.369	0.813
Adjusted R^2	0.650	0.634	0.592	0.362	0.806
MODEL with allelic interaction at the causative gene					
Expanded allele	0.027 ± 0.018 (0.14)	(<0.0001)	(<0.0001)	0.028 ± 0.024 (0.25)	-0.040 ± 0.003 (0.0028)
Non-expanded allele	0.079 ± 0.028 (0.0056)	(0.52)	(0.4367	0.122 ± 0.048 (0.01)	-0.507 ± 0.255 (0.051)
Interaction	-0.0016 ± 0.0006 (0.0088)	(0.48)	(0.3814)	-0.0056 ± 0.0021 (0.007)	$+0.011 \pm 0.005$ (0.044)
R^2	0.664	0.628	0.599	0.374	0.812
Adjusted R^2	0.660	0.621	0.590	0.362	0.822
MODEL with causative gene and non-causal polymorphic (CAG)n-containing genes					
Gene involved, R^2, adjusted R^2	–	*ATXN7* 0.636, 0.634	*ATN1*, 0.602, 0.597 *HTT*, 0.605, 0.600 *ATXN2*, 0.616, 0.613	*ATXN1*, 0.346, 0.333 *ATXN3*, 0.354, 0.346	*ATXN3*, 0.878, 0.864 *TBP*, 0.859, 0.850

Expanded alleles were analysed as quantitative variables. Non-expanded alleles in the*ATXN1* and *CACNA1A* genes were analysed as quantitative variables, whereas non-expanded alleles of the other genes were analysed using subclasses of repeat sizes (see 'Materials and methods' section). Data are expressed as beta ± SD (*P*-value), except when the *P*-values alone of the non-expanded allele and the interaction between both alleles were not significant.
R^2: Determinant coefficient, i.e. percentage at the variance explained by the model; the adjusted R^2 takes into account the number of parameters included in the model used.

more frequent in SCA3 (7%, $n = 28$ from 24 different families) than in other SCAs (SCA1: 4%, $n = 11$; SCA6: 2%, $n = 4$; SCA7: 0%) ($P = 0.015$).

Spinocerebellar ataxia type 6

In subjects with SCA6, age at onset was influenced by the number of repeats in *ATXN1* (difference between the two alleles) and *ATXN3* (longer allele with 25 repeats or more) (Supplementary Table 5). Cases with at least one intermediate *ATXN3* repeat (41%) had earlier onset than those with short or medium *ATXN3* alleles (-0.0319 ± 0.0120, $P = 0.0087$, $R^2 = 35.4\%$). The difference between the number of repeats on the two *ATXN1* alleles decreased the effect of the *CACNA1A* expansions: the greater the difference, the smaller the effect of *CACNA1A* (interaction term between *CACNA1A* and *ATXN1*: $+0.0055 \pm 0.0027$ for one additional repeat, $P = 0.0405$, $R^2 = 34.6\%$). Among the seven cases with differences of more than five repeats between the alleles, all but one had an intermediate *ATXN1* allele (allele with 34 repeats or more). The effects of *ATXN1* and *ATXN3* were not found in the American and Japanese SCA6 cohorts. Intermediate *ATXN3* alleles were present, however, in 51% of the American subjects with SCA6 and 62% of the Japanese subjects with SCA6 suggesting that the effect of *ATXN3* is mediated by another factor that depends on the genetic background.

Spinocerebellar ataxia type 7

The observation of an interaction between the short and expanded *ATXN7* alleles is likely due to the small number of subjects with intermediate *ATXN7* alleles. To limit the sample bias, effects of non-causal polymorphic (CAG)n were tested only in subjects with SCA7 with short or medium normal *ATXN7* alleles (<12 repeats).

Age at onset in SCA7 subjects was influenced by the number of repeats in *ATXN3* (shorter allele) and *TBP* (shorter allele) interacting with the causative gene (Supplementary Table 6). The effect of the *ATXN7* expansion on age at onset increased with the number of normal repeats in the *ATXN3* gene: the larger the number of repeats in the *ATXN3* gene, the earlier the onset ($R^2 = 87.8\%$). A large wild-type *TBP* allele decreased more the age at onset in SCA7, in conjunction with the expanded *ATXN7* allele, than a shorter *TBP* allele (interaction term between the shorter *TBP* allele and the expanded *ATXN7* allele: -0.0063 ± 0.0018, $P = 0.0011$, $R^2 = 85.9\%$). In summary, onset was increasingly earlier in SCA7 cases (i) with increasing numbers of CAG repeats in the expanded *ATXN7* allele; (ii) with normal

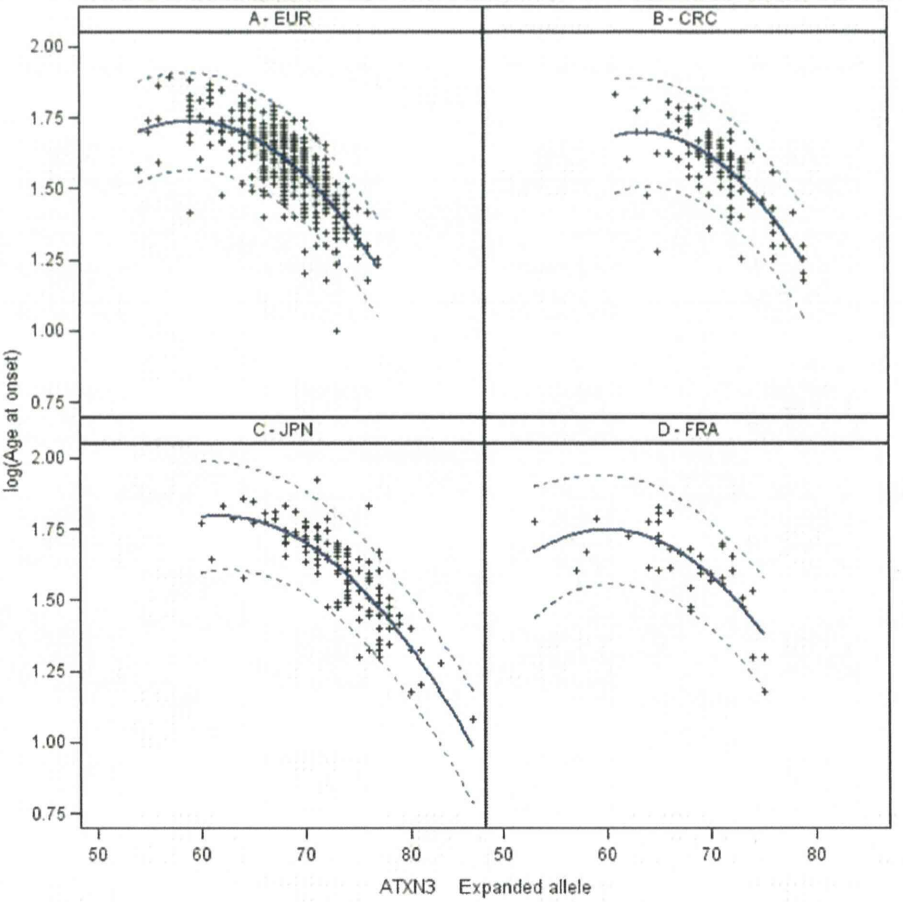

Figure 3 Modification of the age at onset due to *ATXN3* genotypes (expanded allele) in SCA3 cases from the EUROSCA population and the three replication populations. (*Top left*) EUR: Model parameters: Log(age at onset) (LOA) = −3.8098 + 0.1880 Exp−0.00159 Exp^2. (*Top right*) USA: Model parameters: Log(age at onset) = −5.9960 + 0.2423 Exp–0.00191 Exp^2. (*Bottom left*) Japan: Model parameters: Log(age at onset) = −3.3923 + 0.1684 Exp–0.00137 Exp^2. (*Bottom right*) France: Model parameters: Log(age at onset) = −4.1042 + 0.1951 Exp–0.00162 Exp^2. With Exp = Expanded *ATXN3* allele.

ATXN7 alleles of intermediate size; and (iii) with larger wild-type alleles in *TBP*. As SCA7 cases were rare in the replication cohorts, we could not test these effects in other populations.

Discussion

This study was performed in a large series of affected SCA subjects (the largest ever examined for SCA1, SCA2 and SCA3) from 10 European countries. They were analysed with shared clinical methods and scales (Schmitz-Hubsch *et al.*, 2006), and CAG repeat size was systematically re-analysed in a single laboratory. This enabled us to identify new genetic interactions among SCA genes and to confirm others. Some of the results were validated in additional populations, although, despite international recruitment, the replication samples were small, limiting their power and then replication.

Discovery of a polygenic effect on age at onset

First, we identified a quadratic effect of CAG repeat size on age at onset in our subjects with SCA2, SCA3 and SCA6, which was replicated in the American, French and Japanese SCA3 populations. The increase in R^2 ranged from 1.7% for SCA2 to 9% for SCA3. This quadratic effect reflects the existence of two or more different slopes in the association curves relating age at onset to the number of CAG repeats. In subjects with SCA3, the slope of the curve is relatively stable in the 55–65 repeat class, suggesting that the number of repeats and age at onset are relatively independent in this class. Above 65 repeats, CAG size has a stronger effect on age at onset as the number of repeats increases. More than 65 repeats in the ATXN3 protein might induce a conformational change that confers greater toxicity, but this has not been documented to our knowledge and requires functional studies at