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

Sophie Tezenas du Montcel,1,2,3 Alexandre 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 Jose´ 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 Brice4,5 and Giovanni Stevanin4,5,27

1 Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Univ Paris 06, UMR_S 1136, Institut Pierre Louis d’Épidémiologie et de Santé Publique, F-75013, Paris, France
2 INSERM, UMR_S 1136, Institut Pierre Louis d’Épidémiologie 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 Turin, 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 Neuroligico ‘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, AOI 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, Brussels, 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 École Pratique des Hautes Études, 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.
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 ATXN2 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.

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 (R²) is the percentage of the variance explained by a given model. An adjusted R² 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

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

<table>
<thead>
<tr>
<th>Alleles</th>
<th>ATXN1</th>
<th>ATXN2</th>
<th>ATXN3</th>
<th>CACNA1A</th>
<th>ATXN7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>&lt; 22</td>
<td>&lt; 16</td>
<td>&lt; 9</td>
<td>&lt; 10</td>
<td></td>
</tr>
<tr>
<td>Medium</td>
<td>22</td>
<td>16-24</td>
<td>-</td>
<td>10-11</td>
<td></td>
</tr>
<tr>
<td>Intermediate short</td>
<td>23-26</td>
<td></td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Intermediate expanded</td>
<td>&gt;=39</td>
<td>&gt;=33</td>
<td>25-35</td>
<td>9-16</td>
<td>12-14</td>
</tr>
<tr>
<td></td>
<td>&gt;=47</td>
<td>&gt;=20</td>
<td>&gt;=36</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.
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%...
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., 1994; Deka 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.

### 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) (R² = 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

---

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

<table>
<thead>
<tr>
<th>Genetic entity (mutated gene)</th>
<th>SCA1 (ATXN1)</th>
<th>SCA2 (ATXN2)</th>
<th>SCA3 (ATXN3)</th>
<th>SCA6 (CACNA1A)</th>
<th>SCA7 (ATXN7)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affected subjects, n</td>
<td>1255</td>
<td>319</td>
<td>309</td>
<td>403</td>
<td>165</td>
<td>59</td>
</tr>
<tr>
<td>Gender, female</td>
<td>595 (48%)</td>
<td>143 (46%)</td>
<td>145 (47%)</td>
<td>201 (50%)</td>
<td>81 (49%)</td>
<td>20 (34%) 0.24</td>
</tr>
<tr>
<td>Transmitting parent, maternal transmission</td>
<td>543 (52%)</td>
<td>133 (48%)</td>
<td>152 (56%)</td>
<td>164 (48%)</td>
<td>70 (61%)</td>
<td>24 (62%) 0.019</td>
</tr>
<tr>
<td>Age at onset Mean (years ± SD)</td>
<td>38 ± 11</td>
<td>36 ± 13</td>
<td>40 ± 12</td>
<td>53 ± 11</td>
<td>30 ± 14</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Range (years)</td>
<td>11–75</td>
<td>7–71</td>
<td>10–78</td>
<td>24–77</td>
<td>8–71</td>
<td></td>
</tr>
<tr>
<td>Expanded CAG repeat length Mean (CAG ± SD)</td>
<td>47 ± 5</td>
<td>39 ± 3</td>
<td>68 ± 4</td>
<td>23 ± 2</td>
<td>45 ± 5</td>
<td></td>
</tr>
<tr>
<td>Non-expanded CAG repeat length Mean (CAG ± SD)</td>
<td>30 ± 2</td>
<td>22 ± 1</td>
<td>21 ± 5</td>
<td>13 ± 1</td>
<td>11 ± 1</td>
<td></td>
</tr>
<tr>
<td>Range (CAG)</td>
<td>26–37</td>
<td>15–29</td>
<td>12–35</td>
<td>7–16</td>
<td>8–14</td>
<td></td>
</tr>
</tbody>
</table>

SD = standard deviation; F = female.
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

<table>
<thead>
<tr>
<th>Genotypes</th>
<th>SCA1</th>
<th>SCA2</th>
<th>SCA3</th>
<th>SCA6</th>
<th>SCA7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model with the expanded allele only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Linear model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expanded allele</td>
<td>(-0.021 \pm 0.001)</td>
<td>(-0.045 \pm 0.002)</td>
<td>(-0.024 \pm 0.001)</td>
<td>(-0.0351 \pm 0.0040)</td>
<td>(-0.037 \pm 0.002)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>(&lt;0.0001)</td>
<td>(&lt;0.0001)</td>
<td>(&lt;0.0001)</td>
<td>(&lt;0.0001)</td>
<td>(&lt;0.0001)</td>
</tr>
<tr>
<td>Adjusted (R^2)</td>
<td>0.652</td>
<td>0.620</td>
<td>0.504</td>
<td>0.326</td>
<td>0.807</td>
</tr>
<tr>
<td>Quadratic model</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expanded allele (\text{(quadratic term)})</td>
<td>(+2.9 \times 10^{-5} \pm 12.5 \times 10^{-5})</td>
<td>(+1.4 \times 10^{-3} \pm 0.4 \times 10^{-3})</td>
<td>(-1.4 \times 10^{-3} \pm 0.1 \times 10^{-3})</td>
<td>(+8.0 \times 10^{-3} \pm 2.4 \times 10^{-3})</td>
<td>(+4.3 \times 10^{-4} \pm 3.3 \times 10^{-4})</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.652</td>
<td>0.637</td>
<td>0.594</td>
<td>0.369</td>
<td>0.813</td>
</tr>
<tr>
<td>Adjusted (R^2)</td>
<td>0.650</td>
<td>0.634</td>
<td>0.592</td>
<td>0.362</td>
<td>0.806</td>
</tr>
<tr>
<td>Model with allelic interaction at the causative gene</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expanded allele</td>
<td>(0.027 \pm 0.018) ((0.14))</td>
<td>(&lt;0.0001)</td>
<td>(&lt;0.0001)</td>
<td>(0.028 \pm 0.024)</td>
<td>(-0.040 \pm 0.003)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>(0.079 \pm 0.028) ((0.0056))</td>
<td>(0.52)</td>
<td>(0.4367)</td>
<td>(0.122 \pm 0.048)</td>
<td>(-0.507 \pm 0.255)</td>
</tr>
<tr>
<td>Interaction</td>
<td>(-0.0016 \pm 0.0006) ((0.0088))</td>
<td>(0.48)</td>
<td>(0.3814)</td>
<td>(-0.0056 \pm 0.0021)</td>
<td>(+0.011 \pm 0.005)</td>
</tr>
<tr>
<td>(R^2)</td>
<td>0.664</td>
<td>0.628</td>
<td>0.599</td>
<td>0.374</td>
<td>0.812</td>
</tr>
<tr>
<td>Adjusted (R^2)</td>
<td>0.660</td>
<td>0.621</td>
<td>0.590</td>
<td>0.362</td>
<td>0.822</td>
</tr>
<tr>
<td>Model with causative gene and non-causal polymorphic (CAG)n-containing genes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gene involved, (\text{R}^2), adjusted (\text{R}^2)</td>
<td>(-)</td>
<td>ATXN7 0.636, 0.634</td>
<td>ATXN1 0.602, 0.597</td>
<td>ATXN1, 0.346, 0.333</td>
<td>ATXN3, 0.878, 0.884</td>
</tr>
<tr>
<td></td>
<td></td>
<td>HTT, 0.605, 0.600</td>
<td>ATXN3, 0.354, 0.346</td>
<td>TBP, 0.859, 0.850</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>ATXN2, 0.616, 0.613</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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 \(\pm S D\) (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 \((-0.0319 \pm 0.0120, P = 0.0087, R^2 = 35.4\%)\). 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
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 R² 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
this point. Secondly, a larger size of the normal alleles in trans in the causative genes lowered the age at onset in SCA1 and SCA6 affected subjects, as reported previously in smaller cohorts (van de Warrenburg et al., 2005). The additional increase in $R^2$ was 1.2% for SCA1 and 4.8% for SCA6. Although the allelic distributions were similar in the Japanese, French and American subjects, these results were not replicated. The number of subjects with SCA6 was too small in these cohorts to identify an additional $R^2$ of 4.8%; a sample size of at least 129 would be needed to obtain an additional $R^2$ of 4.8% with a power of 80%. At the SCA7 locus, a small but significant tendency towards an earlier age at onset was due to effects of the smaller alleles. Normal ATXN2 and ATXN3 alleles had no effects, probably because of the low degree of polymorphism of their CAG repeats. Effects of normal alleles at the SCA7 and SCA6 loci have been reported in a simplified model (Durr et al., 1996; van de Warrenburg et al., 2005).

Thirdly, the negative association between age at onset and the size of the CAG repeat expansion was exacerbated when the normal repeat in a modifier gene at another genetic locus was large in size: ATXN3 repeats in SCA7 and SCA6 patients, ATXN2 and ATN1 in SCA3, TBP in SCA7, ATXN7 in SCA2 (Fig. 5). The additional increase in $R^2$ ranged from 0.8% (ATN1 for SCA3 subjects) to 7.1% (ATXN3 for SCA7 subjects). A similar tendency, although not significant, was observed for the ATXN7 allele in American and French SCA2 subjects. The effect of ATXN2 alleles on age at onset in SCA3 was not validated in the replication cohorts, probably due to the low prevalence of intermediate alleles at this locus. This was already illustrated in a homogeneous SCA3 Brazilian cohort (Jardim et al., 2003). The effects of TBP and ATXN3 alleles on subjects with SCA7 could not be tested in the replication populations because they were too few. Surprisingly, the size of the normal repeat in the Htt gene had the opposite effect; an increase in the size of the normal allele delayed the age at onset in SCA3 subjects. As we had no information on the genotypes of the replication cohorts at the Htt locus, we could not validate these results. A large difference in the number of CAG

Figure 4 The presence of at least one ATXN7 allele ≥ 12 CAG results in an earlier age at onset in SCA2 cases. Open circles and solid lines: no ATXN7 alleles with 12 CAG repeats or more; Crosses and dotted lines: at least one ATXN7 alleles with 12 CAG repeats or more. Model parameters: (A) EUROSCA: Cases with no alleles with both ATXN7 alleles over 12: $LOA = 3.318 - 0.046 \ Exp$; Cases with at least one allele ≥ 12 repeats: $LOA = 3.275 - 0.046 \ Exp$. (B) USA: Cases with no alleles with both ATXN7 alleles over 12: $LOA = 3.361 - 0.045 \ Exp$; Cases with at least one allele ≥ 12 repeats: $LOA = 3.331 - 0.045 \ Exp$. (C) Italy: Cases with no alleles with both ATXN7 alleles over 12: $LOA = 3.120 - 0.040 \ Exp$; Cases with at least one allele ≥ 12 repeats: $LOA = 3.166 - 0.040 \ Exp$. (D) France: Cases with no alleles with both ATXN7 alleles over 12: $LOA = 2.949 - 0.035 \ Exp$; Cases with at least one allele ≥ 12 repeats: $LOA = 2.914 - 0.035 \ Exp$. With $LOA = \log(\text{Age at onset})$, $\Exp = \text{ATXN2 expanded allele.}$
Modifier genes of age at onset in SCAs

Figure 5 Summary of interactions among (CAG)n-containing genes in the EUROSCA cohort. For the SCA type, the black labels indicate the gene, the blue label, the age at onset of individuals with an expansion of the gene. Black arrows are for effect in interaction with the major gene. Nb > 12: = alleles with more than 12 repeats; Interm = intermediate allele; + = positive effect; – = negative effect; Shorter = shorter allele; Longer = longer allele; Longer > 25 = longer allele with > 25 repeats; Difference = difference between the longer and the shorter allele.

repeats in normal ATXN1 alleles delayed the age at onset in SCA6 subjects, but this observation was not replicated in the American and Japanese SCA6 populations. The only modifier effect observed in SCA1 was that of the normal ATXN1 allele.

It should be noted that the results concerning SCA6 and SCA7 were obtained in small numbers of cases. Mutations in these genes are rare. Thus, despite sample sizes among the largest available for these diseases, we were not able to perform a multivariate analysis for the SCA7 gene. For the SCA6 causative gene, we found an effect on age at onset of two non-causal genes, ATXN1 and ATXN3, in addition to SCA6. This might be due to over-fitting, however, although results based on the univariate analysis are not normally prone to this bias. However, we did not reproduce the effects of CACNA1A repeats (Pulst et al., 2005) or long ATXN3 alleles (de Castilhos et al., 2014) previously reported in Cuban and Brazilian subjects with SCA2. Differences in CAG repeat size or phenotypic expression in affected subjects of different ethnic or geographic origins (Subramony et al., 2002) might explain why these results were not found in our large SCA series.

Are the genetic effects observed of biological relevance?

The exact functions of the genes involved in SCAs are often not known. Recent studies suggest that most of the mutated proteins contribute to pathogenesis in relation to their ‘normal’ cellular functions. The ATXN1 protein forms dimers (de Chiara et al., 2013); a dimer between the wild-type protein and pathological ATXN1 might underline the significant effect in trans of the ATXN1 expansion. Normal repeats in ATXN2 and ATXN3 vary little in size. This might explain why we did not detect interactions of these genes in trans with the expanded CAG repeat. In a previous study by Lim et al. (2006), biological interactions between some of the proteins involved in ataxias were detected in a 2-hybrid screen. In particular, an indirect interaction between the ATXN1 and ATN1 proteins, might be at the origin of the effects of the corresponding genes on age at onset we observed in SCA3 cases. Another study, in Drosophila, showed that ATXN2 modifies the SCA7 phenotype (Latouche et al., 2007). This effect was not found in our small SCA7 cohort, but the biological interaction between their protein products, observed in flies, might account for the effect of ATXN7 repeats observed in multiple SCA2 cohorts in our study.

Long normal HTT alleles were protective in our subjects with SCA3, as in subjects with Huntington’s disease reported in a previous study (Djoussé et al., 2003; Aziz et al., 2009). Both beneficial (mitigation of mutant protein toxicity) and detrimental (loss of normal HTT function) effects of HTT have been described, depending on the size of the expansion (Aziz et al., 2009). Larger stretches of polyglutamine in normal HTT might promote stronger associations with mutant protein fragments, facilitating their co-aggregation, which would prevent them from interfering with other proteins. A reported interaction between the ATXN3 protein and HTT associated protein 1 (HAP1) (Takeshita et al., 2011) might explain the negative effect of HTT on age at onset we observed in SCA3.

The propensity of the polyglutamine stretch, encoded by the polymorphic CAG repeat, to aggregate increases with its size (Perutz, 1995). It has been hypothesized that the formation of inclusions in neurons is promoted by long repeats that trap partners of the aggregated proteins in the inclusions. The sequestration of the ATXN3 protein and the normal ATXN7 isoform in inclusions in subjects with SCA7 (Zander et al., 2001; Takeshita et al., 2011) depends on the size of the pathological repeat, but also on the size of the normal alleles. It is therefore tempting to postulate that this is illustrated by the genetic interaction we observed between ATXN3 and ATXN7 alleles in SCA7. It should be noted that this explanation assumes that the aggregate formation is deleterious to cells, a hypothesis that is still debated (Michalik and Van Broeckhoven, 2003; Ross and Poirier,
Furthermore, it is probably dependent on disease duration.

Finally, the functional relationship between ATXN7, a component of SAGA, and TBP (Martinez et al., 1998; Mohibullah and Hahn, 2008) would explain the genetic interaction between wild-type TBP and the ATXN7 expansion.

In conclusion, we have demonstrated, in a large cohort, that the polyglutamine genes interact with each other in SCA diseases to modify age at onset even when they contain a number of repeats considered to be normal. The influence of these genes on the severity and progression of the diseases should also be investigated as done in previous studies in smaller series of patients (Jardim et al., 2003; de Castilhos et al., 2014). In addition, the variability in age at onset is still not completely explained by the effects of the major and additional genes examined here, suggesting that other genetic or environmental factors play a role in these diseases. Non-polyglutamine genes have also been found to affect age at onset in SCA2 [GLUR6 (now known as GRIK2), CA150 (now known as TCERG1), RA11, mitochondrial complex 1 (Hayes et al., 2000; Holbert et al., 2001; Chattopadhyay et al., 2003)] and might contribute partially to the variability of phenotypic expression. In our EUROSCA cohort, 34%, 34%, 38%, 65% and 12% of the variability in age at onset for SCA1, SCA2, SCA3, SCA6 and SCA7 cohorts, respectively, is still unexplained by known genetic factors and modifiers. High-throughput sequencing in large cohorts should facilitate the identification of more modifier variants, which would help design therapeutic strategies aimed at slowing disease progression or delaying onset. This study, which enrolled the largest number of SCA cases so far, did not confirm some previously reported genetic interactions. This suggests that ethnic differences are likely to exist and question the role of replication studies when the same stringent and uniform methods are not used and when only limited number of patients are available when searching for weak contributors of the phenotypic variability. Finally, even if some functional observations are compatible with several of the genetic interactions highlighted here, additional studies are needed to confirm these effects at the cellular level.

Acknowledgements

The authors would like to thank the affected individuals and their relatives for participating in the study. They are also grateful to the DNA and cell bank of the Institut du Cerveau et de la Moelle épinière (technical coordinator: Sylvie Forlani) and the ICM genotyping/sequencing facilities (technical coordinator: Yannick Marie). The excellent technical assistance of Philippe Martin-Hardy, Michèle Viémont, Claudia Bauer, Daniela Lenz, Vivi Zuri and N’ama Lanxner was much appreciated. We also thank Merle Ruberg for the critical reading of the manuscript.

Funding

This work was supported in part by the European Union 6th and 7th framework programs, EUROSCA/LSHM-CT-2004-503304 (to Alexis Brice, Ludger Schöls, Thomas Klockgether, Olaf Riess, Peter Bauer, Alessandro Filla) and NEUROMICS/FP7-2012-305121 (to Alexis Brice, Ludger Schöls, Thomas Klockgether, Olaf Riess, Peter Bauer), the Programme Hospitalier de Recherche Clinique (PHRC, to Alexandra Durr), the Association Connaître les Syndromes Cérébelleux (to Giovanni Stevanin, Alexandra Durr and the SPATAX network), the VERUM Foundation (to Alexis Brice) and the Fondation Roger the Spoelberch (to Alexis Brice). This study also benefited from funding from the program ‘Investissements d’avenir’ ANR-10-IAIHU-06 (to the Institut du Cerveau et de la Moelle épinière, Paris), PRIN 2010/2011 (to Alfredo Brusco) and the grant 3 PO5B 019 24 (2003 - 2006) from the Polish Ministry of Science and Higher Education (to Maria Rakowicz) and the grant POIG 01.01.02&8211;05/09/01 (2010-2014) from the Polish Ministry of Regional Development Operating Programme Innovative Economy (to Maria Rakowicz).

Paola Giunti works at University College London Hospitals/University College London which receives a proportion of its funding from the Department of Health’s National Institute for Health Research Biomedical Research Centres funding scheme. The CRCSCA was funded by a NIH grant (NS068897) and a NAF (Pioneer grant to Tetsuo Ashizawa).

Supplementary material

Supplementary material is available at Brain online.

References


Appendix 1
Clinical Research Consortium for Spinocerebellar Ataxia
EUROSCA network
Sophie Tezenas du Montcel, Alexandra Durr, Peter Bauer, Sylvie Forlani, Maria Rakowicz, Anna Sulek, Ludger Schöls, Caterina Mariotti, Bart P.C. van de Warrenburg, Paola Giunti, Alessandro Filla, Sandra Szymanski, Thomas Klockgether, José Berciano, Massimo Pandolfo, Sylvia Boesch, Melegh Bela, Laszlo Baliko, Kinga Hadzsiev, Dagmar Timmann, Olaf Riess, Alexis Brice, Giovanni Stevanin.