Association of Long ATXN2 CAG Repeat Sizes With Increased Risk of Amyotrophic Lateral Sclerosis

Hussein Daoud, PhD; Véronique Belzil, MSc; Sandra Martins, PhD; Mike Sabbagh, MD; Pierre Provencher, MSc; Lucette Lacomblez, MD; Vincent Meiningner, MD; William Camu, MD; Nicolas Dupré, MD, FRCPC; Patrick A. Dion, PhD; Guy A. Rouleau, MD, PhD, FRCPC

Objective: To analyze the ataxin 2 (ATXN2) CAG repeat size in a cohort of patients with amyotrophic lateral sclerosis (ALS) and healthy controls. Large (CAG)n alleles of the ATXN2 gene (27-33 repeats) were recently reported to be associated with an increased risk of ALS.

Design: Case-control study.

Setting: France and Quebec, Canada.

Participants: A total of 556 case patients with ALS and 471 healthy controls; both groups of participants are of French or French-Canadian origin.

Results: We observed a significant association between ATXN2 high-length alleles (∼29 CAG repeats) and ALS in French and French-Canadian ALS populations. Furthermore, we identified spinocerebellar ataxia type 2–pathogenic polyglutamine expansions (∼32 CAG repeats) in both familial and sporadic ALS cases.

Conclusions: Altogether, our findings support ATXN2 high-length repeats as a risk factor for ALS and further indicate a genetic link between spinocerebellar ataxia type 2 and ALS.

Arch Neurol. 2011;68(6):739-742

MYOTROPHIC LATERAL sclerosis (ALS) is a devastating adult-onset neurodegenerative disease characterized by a progressive loss of motor neurons of the cerebral cortex and spinal cord. Clinically, the disease results in spasticity as well as progressive muscle weakness and atrophy, which typically leads to death within 3 to 5 years following symptoms onset. Approximately 5% to 10% of patients with ALS have a family history of ALS (FALS), most frequently with autosomal dominant transmission; the remaining 90% of patients with ALS who are clinically indistinguishable from patients with FALS are referred to as patients with sporadic ALS (SALS).

Despite intensive research efforts, few genes have been unequivocally associated with ALS. Among these, the superoxide dismutase 1 (SOD1) gene is the most frequent causative gene because mutations in SOD1 account for 15% to 20% of all FALS cases and 1% to 2% of all ALS cases. Mutations in TARDBP and FUS, which encode 2 multifunctional DNA/RNA binding proteins, were recently identified in FALS cases and subsequently in SALS cases. Other FALS-causative genes have also been reported but are responsible for only a small number of FALS cases. Similarly, the etiology of most SALS cases remains to be identified. Many genome-wide association studies reported risk alleles for SALS; however, only the 9p21 locus has been replicated in different populations, and none of these associated genes were found to be mutated in patients.

Following a genetic screen in yeast, Elden and colleagues recently reported that the ataxin 2 (ATXN2) protein modulates TDP-43 toxicity, which appears to play a central role in several neurodegenerative diseases including ALS. The 2 proteins interact in an RNA-dependent complex, and both become mislocalized in spinal cord neurons of patients with ALS. Interestingly, Elden et al also found that intermediate-length polyglutamine tracts (range, 24-33 repeats) in ATXN2 confer an increased risk for developing ALS.

The ATXN2 (12q24.1) polyglutamine tract shows a normal-size range that extends between 14 and 31 repeats, with 22 and 23 repeats being the most frequent ones. Expansions of more than 34 repeats are known to cause spinocerebellar ataxia type 2 (SCA2), an autosomal dominant disorder characterized by progres-
peripheral blood cells using standard methods. We amplified the \textit{ATXN2} sequence at the 5' end (M13-ATXN2F, 5'-TGTTAAACGACGGCGTTCACCATGTCG-3'), and the reverse primer (ATXN2R) was 5'-CGGGTTCGAGACATTGG-3'. The PCR cycles were as follows: 2 minutes at 96°C, 9 minutes at 95°C, 6 cycles (1 minute at 95°C, 30 seconds at 64°C [-0.5°C/ cycle], 1 minute at 72°C), 22 cycles (1 minute at 95°C, 30 seconds at 61°C, 1 minute at 72°C), 8 cycles (1 minute at 95°C, 30 seconds at 53°C, 45 seconds at 72°C), and 10 minutes at 72°C. We determined the CAG repeat sizes with capillary electrophoresis by incorporating a VIC-labeled M13 universal primer (5'-VIC-TGTTAAACGACGGCGTTCAGT-3') into the PCR. The PCR products were then diluted (1:20) and mixed with LIZ-500 size standard (Applied Biosystems, Foster City, California) and processed for size determination on an ABI 3730 DNA analyzer (Applied Biosystems). Repeat sizes were determined using the GeneMapper Software version 4.0 (Applied Biosystems). Sixty-six samples with more than 24 CAG repeats were verified by independent PCRs to further confirm repeat sizes.

**STATISTICAL ANALYSES**

The accuracy of our test to discriminate case patients with ALS from healthy controls was evaluated using receiver operating characteristic curve analysis with GraphPad Prism 5 software (GraphPad Software Inc, La Jolla, California). A 2-tailed Fisher exact test was used to evaluate genetic association between \textit{ATXN2} CAG repeat sizes and ALS (significance was set at \(P < 0.05\)). The correlation between CAG repeats and age at onset was performed by survival analysis using the log-rank Mantel-Cox test with GraphPad Prism 5 software.

**RESULTS**

The \textit{ATXN2} CAG repeat was defined in genomic DNA from 556 patients with ALS and 471 healthy controls from France and Quebec, Canada. The (CAG)\textsubscript{n} size varied from 19 to 32 repeats in controls and from 19 to 37 repeats in case patients. The most abundant alleles were (CAG)\textsubscript{22} (65% of alleles), followed by (CAG)\textsubscript{21} (20%) and (CAG)\textsubscript{23} (7%). We found that 24 of 471 healthy controls (5.1%) harbored 1 intermediate \textit{ATXN2} allele (range, 24-33), whereas 40 of 556 case patients with ALS (7.2%) had 1 allele within that range; this difference is not statistically significant (\(P = .15\)). However, a receiver operating characteristic curve analysis of our data showed that the greatest sensitivity and specificity for discriminating patients with ALS and healthy controls is provided by a cutoff of 27 or more CAG repeats. Nevertheless, when using this cutoff, we found that the difference was still not statistically significant because 19 controls (4.1%) and 35 case patients with ALS (6.3%) have a repeat length of 27 or higher (range, 27-37) (\(P = .09\)). However, we did observe a significant association for the higher CAG repeat sizes (\(\geq 29\) repeats) in our cohort of case patients with ALS (Figure). Indeed, CAG repeats of 29 or higher (range, 29-37) were found in only 4 controls (0.8%), whereas 25 case patients with ALS (4.5%) had such repeats (odds ratio [OR], 5.5 [95% confidence interval {CI}, 1.9-15.9]; \(P = 2.4 \times 10^{-4}\)). Moreover, when 7 of the 95 patients with FALS (7.4%) and 18 of the 461 patients with SALS (3.9%) who carry 1 \textit{ATXN2} allele with 29 CAG repeats or more were compared with the healthy controls, the association with case patients with FALS becomes even stronger (OR, 9.29 [95% CI, 2.66-32.4]; \(P_{\text{FALS}} = 5.2 \times 10^{-3}\) vs OR, 4.74 [95% CI, 1.59-14.13]; \(P_{\text{SALS}} = 7.5 \times 10^{-3}\)). When the SCA2-pathogenic CAG repeat sizes (\(\geq 32\)) are removed, the association remains statistically significant (OR, 4.03 [95% CI, 1.15-14.11]; \(P = .01\)) because 14 case patients with ALS (2.5%) and 3 healthy controls (0.6%) have a repeat length between 29 and 31 repeats.
We then sought to test whether high-length CAG repeats (≥29) could have an effect on the age at which the onset of ALS was reported for each case patient. Hence, we compared 23 case patients with 29 or more CAG repeats vs 489 case patients without 29 or more CAG repeats, but no correlation could be observed (data not shown).

Finally, we observed a significant difference in the number of SCA2-pathogenic CAG repeats (size, ≥32) between the case patients with ALS and the healthy controls (11 patients with ALS vs 1 control; P = .001). Of the 11 patients with such pathogenic CAG repeats, 7 were found to have 32 repeats, and 4 were found to have repeat sizes between 35 and 37 (Table). Two of these case patients with ALS had FALS, whereas 9 had SALS and all presented with classical ALS signs and mean age at onset of 68.7 years. In all case patients, the disease started in the limbs. None of these patients presented with features of SCA2 such as cerebellar or brainstem atrophy.

### Table. Clinical Characteristics of Patients with ATXN2 Alleles With CAG Repeat Sizes of 32 or Higher

<table>
<thead>
<tr>
<th>Patient No./Sex/Age at Onset</th>
<th>Type of ALS</th>
<th>CAG Repeat Size</th>
<th>Site of Onset</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/F/not available</td>
<td>FALS</td>
<td>32</td>
<td>Not available</td>
</tr>
<tr>
<td>2/M/75</td>
<td>FALS</td>
<td>32</td>
<td>Not available</td>
</tr>
<tr>
<td>3/M/75</td>
<td>SALS</td>
<td>32</td>
<td>Spinal</td>
</tr>
<tr>
<td>4/F/76</td>
<td>SALS</td>
<td>32</td>
<td>Spinal</td>
</tr>
<tr>
<td>5/M/67</td>
<td>SALS</td>
<td>32</td>
<td>Spinal</td>
</tr>
<tr>
<td>6/M/58</td>
<td>SALS</td>
<td>37</td>
<td>Not available</td>
</tr>
<tr>
<td>7/M/60</td>
<td>SALS</td>
<td>32</td>
<td>Spinal</td>
</tr>
<tr>
<td>8/M/52</td>
<td>SALS</td>
<td>35</td>
<td>Spinal</td>
</tr>
<tr>
<td>9/F/76</td>
<td>SALS</td>
<td>35</td>
<td>Spinal</td>
</tr>
<tr>
<td>10/M/72</td>
<td>SALS</td>
<td>36</td>
<td>Spinal</td>
</tr>
<tr>
<td>11/F/76</td>
<td>SALS</td>
<td>32</td>
<td>Spinal</td>
</tr>
</tbody>
</table>

Abbreviations: ALS, amyotrophic lateral sclerosis; FALS, family history of ALS; SAL, sporadic ALS.

Recently, Elden and colleagues elegantly showed that intermediate-length polyglutamine tracts in the ATXN2 protein are associated with increased risk of ALS. In our study, we have assessed the ATXN2 polyglutamine repeat length in our cohort of case patients with ALS and healthy controls. Although intermediate-length CAG tracts were not significantly associated with ALS, we did find a significant association between higher lengths (≥29 repeats) and ALS in our cohort of French and French-Canadian patients. This association was also significant in our patients with FALS and in our patients with SALS, separately, with FALS presenting a stronger association.

Moreover, we found SCA2-pathogenic polyglutamine expansions (32 and higher) in approximately 2% of our cohort of patients with ALS (2 patients with FALS and 9 patients with SALS), whereas only 1 control participant had an allele with 32 expansions. Although the presence of a 32-repeat allele in 1 control participant is intriguing, the effect of ATXN2 alleles with 32 to 33 repeats on pathogenesis of SCA2 has been a matter of speculation ever since asymptomatic individuals carrying these intermediate alleles have been reported.16,17

Altogether, the association of high-length ATXN2 alleles (≥29 CAG repeats) in our cohort of patients with FALS or SALS, as well as the presence of SCA2-pathogenic polyglutamine expansions in some patients both with familial and sporadic ALS, strongly suggests that large repeats of ATXN2 confer an increased risk of ALS and support a genetic link between ALS and SCA2. Interestingly, this genetic link is strengthened by the previous association of motor neuron disease in some patients with genetically confirmed SCA2.19,20

In summary, our findings support ATXN2 high-length repeats as a risk factor for ALS and further indicate a genetic link between SCA2 and ALS. Given that long polyglutamine tracts in ATXN2 have been proposed to enhance its interaction with TDP-43, and that this interaction makes TDP-43 more prone to aggregation, it is tempting to speculate that other polyglutamine proteins involved in neurodegenerative disorders are good candidates for ALS. For now, whether these interactions contribute to the pathogenesis of ALS is still unknown. This will lead to exciting research to identify novel proteins that cause ALS.

Accepted for Publication: January 27, 2011.

**Author Affiliations:** Centre of Excellence in Neuroinformatics, CHUM Research Centre (Drs Daoud, Martins, Sabbag, Dion, and Rouleau and Ms Belzil), Faculty of Medicine, Department of Pathology and Cell Biology (Dr Dion), Research Center, CHU Sainte-Justine, and Department of Pediatrics and Biochemistry (Dr Rouleau), Université de Montréal, and Faculty of Medicine, Laval University, Centre Hospitalier Affilié Universitaire de Québec, Enfant-Jesus Hospital, Ville de Québec (Mr Provancher and Dr Dupré), Canada; Institute of Molecular Pathology and Immunology of the University of Porto, Portugal (Dr Martins); and Fédération des maladies du système nerveux, Centre de référence maladies rares sclérose latérale amyotrophique, Hôpital Pitié-Salpêtrière, Paris (Dr Lacomblez and Meininger), and Unité de Neurologie Compportementale et Dégénérative, Institute of Biology, Montpellier (Dr Camu), France.

**Corresponding Author:** Guy A. Rouleau, MD, PhD, Centre for Excellence in Neuroinformatics, CHUM Research Center, Université de Montréal, 2009 Alexandre De-Seve St, Rm Y-3633, Montreal, QC H2L 2W5, Canada (guy.rouleau@umontreal.ca).

**Author Contributions:** Study concept and design: Daoud and Martins. Acquisition of data: Daoud. Analysis and interpretation of data: Daoud, Belzil, Martins, Sabbag, Provancher, Lacomblez, Meininger, Camu, Dupré, Dion, and Rouleau. Drafting of the manuscript: Daoud. Critical revision of the manuscript for important intellectual content: Belzil, Martins, Sabbag, Provancher, Lacomblez, Meininger, Camu, Dupré, Dion, and Rouleau. Obtained funding: Rouleau. Administrative, technical, and material support: Daoud, Belzil, Martins, Sabbag, Provancher, Lacomblez, Meininger, Camu, and Dupré. Study supervision: Dion and Rouleau.

**Financial Disclosure:** None reported.
**Funding/Support:** Dr Daoud and Ms Belzil are supported by a postdoctoral and a doctoral fellowship from the Canadian Institutes of Health Research, respectively. Dr Martins is supported by the postdoctoral fellowship SFRH/BPD/29225/2006 from Fundação para a Ciência e a Tecnologia (Portugal). Dr Rouleau holds a Canada Research Chair in Genetics of the Nervous System and the Jeanne-et-J-Louis-Lévesque Chair in Genetics of Brain Diseases.

**Role of the Sponsor:** The funders did not participate in the design and conduct of the study; in the collection, analysis, and interpretation of the data; or in the preparation, review, or approval of the manuscript.

**Additional Contributions:** We thank all the patients and their families involved in our study. We thank Annie Levert, DEC, and Sylvia Dobrzeniecka, MSc, for their technical assistance.

**REFERENCES**