Amyotrophic Lateral Sclerosis Risk for Spinocerebellar Ataxia Type 2 ATXN2 CAG Repeat Alleles
A Meta-analysis

Annalese G. Neuenschwander; Khanh K. Thai, MS; Karla P. Figueroa, MS; Stefan M. Pulst, MD

IMPORTANCE Repeats of CAG in the ataxin 2 gene (ATXN2) in the long-normal range (sometimes referred to as intermediate) have been identified as modifiers of amyotrophic lateral sclerosis (ALS) risk. Prior studies have used thresholding considering various cutoffs for ATXN2 repeat length.

OBJECTIVE To calculate association between ATXN2 CAG repeat alleles and increased risk of ALS across multiple ethnic groups.

DATA SOURCES The MEDLINE database was searched for studies published by December 29, 2013, reporting ATXN2 CAG repeat length in patients with ALS and controls.

STUDY SELECTION Studies were included if they reported original data on relative risks or odds ratios (ORs) from ALS and control populations for individual ATXN2 alleles. Review articles that reported no new data were not included in the analysis.

DATA EXTRACTION AND SYNTHESIS Analysis of allele distribution was performed to ensure that all studies followed identical allele sizing. The ORs, 95% confidence intervals, and population attributable risk percentages were calculated according to standard procedures.

MAIN OUTCOMES AND MEASURES Occurrence of ALS associated with ATXN2 repeat alleles, expressed as ORs.

RESULTS Nine studies were analyzed, including 7505 controls and 6151 sporadic ALS cases. The ALS and control cohorts were recruited from different geographical and ethnic regions including the United States, French Canada/Canada, Belgium and the Netherlands, Germany, Italy, mainland China, Turkey, and Flanders-Belgium. The ATXN2 CAG repeat lengths ranged from 13 to 39 in patients with ALS and from 13 to 34 in controls. The ORs were less than 1.00 for alleles with 25 to 28 repeats. The OR was 1.55 for 30 repeats, but this elevation was not statistically significant (95% CI, 0.88-2.73). The ORs were 2.70 (95% CI, 1.47-4.93) for 31 CAG repeats, 11.09 (95% CI, 4.16-29.57) for 32 repeats, and 5.76 (95% CI, 1.79-18.57) for 33 repeats.

CONCLUSIONS AND RELEVANCE In contrast to prior studies with smaller numbers, risk for ALS associated with long-normal alleles is complex. Alleles with 27 and 28 repeats lower ALS risk slightly. The risk for ALS increases beginning with 29 repeats and reaches a maximum at 32 and 33 repeats. Of note, alleles with repeats of these lengths are known to be predisposed to meiotic expansion to full-penetrance mutant alleles. In patients with ALS, alleles with 31 to 33 repeats may have undergone preferential expansion in motor neurons during mitosis or DNA repair. Our meta-analysis provides a framework for counseling individuals with long-normal ATXN2 repeats.
spinocerebellar ataxia type 2 (SCA2) is a rare autosomal dominant neurodegenerative disorder caused by expanded glutamine repeats located in the N-terminal region of the ataxin 2 gene (ATXN2; GenBank BC114546). The ATXN2 polyglutamine allele length, although variable, is most commonly 22 repeats with rarer nonpathological alleles ranging from 14 to 31 repeats. Disease alleles causing ataxia contain 33 or more repeats.\(^5\) Longer pathogenic polyglutamine repeat lengths are inversely correlated with the age at onset.\(^3\) Pathology of SCA2 is characterized by considerable neuronal loss in the cerebellar Purkinje cell layer as well as all 4 deep cerebellar nuclei.\(^4\) In addition to cerebellar ataxia, patients with SCA2 may show other clinical signs such as saccade slowing, early hyporeflexia, severe tremors, and myoclonus.\(^5\)

A recent study by Elden et al\(^6\) showed that ATXN2 was a potent modifier of transactive response DNA-binding protein 43 kDa (TDP-43) through an RNA-dependent mechanism. TDP-43 has been found to be a major component of ubiquitinated cytoplasmic inclusions in neurons of patients with amyotrophic lateral sclerosis (ALS).\(^7,8\) Interestingly, Elden and colleagues went on to show that ATXN2 alleles with 27 to 33 repeats, a class of alleles they designated as intermediate, were significantly associated with ALS. Since then, many studies\(^9-16\) have shown that long-normal ATXN2 alleles contribute to an increased risk of ALS in a variety of ethnic backgrounds. Unfortunately, there seems to be a lack of consistency when defining the boundaries of intermediate-length repeats among the studies.

The purpose of our study was to perform a meta-analysis of published data to determine allele-specific risk for ALS for alleles ranging from 24 to 33 repeats.

Methods

Search Strategy

We undertook a MEDLINE database search on PubMed from July 2012 to December 2013 for studies that tracked ATXN2 CAG repeat length in patients with ALS. The Medical Subject Heading terms used were SCA2 OR ataxin-2 OR ATXN2 AND ALS OR amyotrophic lateral sclerosis. The analysis was restricted to articles published in English. The final search was conducted on December 29, 2013. Studies were included if they reported original data, either directly in the publication or provided on request, on relative risks or odds ratios (ORs) from case and control populations, with cases defined as patients with ALS and control samples representative of the general population or controls who were matched for age and/or geographical region to the cases. We excluded review articles that reported no new data and studies that reported on interrupted ATXN2 repeat lengths.

Statistical Analysis

Data were analyzed for genetic association between ATXN2 and ALS using Mantel-Haenszel methods in Stata version 12/12.1 statistical software (StataCorp LP). The ORs were determined for allele counts greater than 23, taking into account the weight of each study. The ORs for alleles with more than 34 repeats could not be computed as these alleles were not found in controls. A repeat of 0.5 was added to the control group so that ORs could be computed in instances in which controls were not reported for the respective repeat length. The addition of the small constant makes the estimate of Mantel-Haenszel pool ORs close to the maximum likelihood estimate.\(^17,18\) Population attributable risk percentages were calculated using standard methods.\(^19\)

Study Populations

Most patients with ALS in previously published studies\(^6,9-16\) were diagnosed according to the El Escorial revised criteria.\(^20\) In all examined studies, control samples were regionally matched to unrelated participants with no reported history of neurological disease.

Results

From the MEDLINE database search, PubMed yielded 28 articles, of which 14 fit the inclusion criteria for the meta-analysis. To be included within the meta-analysis, specific data for individual ATXN2 allele frequencies greater than 23 for both the ALS and control cohorts were required. When sufficient data were found in the publication, the data were extracted from the published information. When adequate data were not published in the article, first and senior authors were contacted via e-mail to request additional information. Of the 14 eligible publications, 9 provided adequate data for analysis. Five of the 14 eligible studies did not respond to e-mails requesting additional data.

From the 9 studies, 7505 controls and 6151 sporadic ALS cases were analyzed. The ALS and control cohorts were recruited from different geographical and ethnic regions including the United States,\(^6\) French Canada/Canada,\(^10\) Belgium and the Netherlands,\(^12\) Germany,\(^11\) Italy,\(^9,16\) mainland China,\(^15\) Turkey,\(^13\) and Flanders-Belgium.\(^14\) The ATXN2 CAG repeat length ranged from 13 to 39 in patients with ALS and from 13 to 34 in controls. Overall ATXN2 CAG repeat length distribution is shown in Figure 1 and in the eTable in the Supplement. Repeat lengths 23 or less represented 97.19% of alleles in the control population and 96.64% of alleles in patients with ALS. This is congruent with prior studies.\(^1,21,22\) As previously reported,\(^25,23\) the third most common repeat allele in controls and patients contained 27 repeats, suggesting that all studies used similar sizing standards, allowing comparisons across studies.

In contrast to prior studies that used a threshold for normal and predisposing alleles, which was often set post hoc, we had a sufficiently large number of alleles for each repeat size that allowed us to determine ORs specific for individual alleles. This meta-analysis revealed ORs of 1.00 or lower for alleles with 25 to 28 repeats. The first study examining ALS risk for CAG repeats between 27 and 33 reported a significantly increased risk for this class of alleles.\(^6\) The Table lists the ORs, 95% confidence intervals, and population attributable risk percentages for each allele summed across all studies. Contrary to previous studies, our analysis showed that ALS risk was decreased with CAG repeat lengths of 27 (OR = 0.77; 95% CI, 0.63-0.95)
The OR was 1.55 for 30 repeats, but this higher OR value did not indicate a statistically significant risk increase for ALS (95% CI, 0.88-2.73). Beginning with 31 repeats, ORs were significantly elevated (Table). The population attributable risk percentage for alleles with more than 30 repeats, however, was relatively low in keeping with the overall rare occurrence of these alleles in the general population (Table).

The ORs of each study for alleles of 30, 31, 32, and 33 repeats are shown in Figure 2. The ORs were 2.70 (95% CI, 1.47-4.93) for 31 CAG repeats, 11.09 (95% CI, 4.16-29.57) for 32 repeats, and 5.76 (95% CI, 1.79-18.57) for 33 repeats. Not all studies were informative for each allele class, when the respective allele was not present in cases and in controls. This is particularly true for alleles of more than 34 repeats that were found in patients with ALS but not in controls (Figure 1).

Table. Summary of Odds Ratios, 95% Confidence Intervals, and Population Attributable Risk Percentages for Ataxin 2 Gene Alleles With 24 to 33 CAG Repeats

<table>
<thead>
<tr>
<th>CAG Repeats, No.</th>
<th>Odds Ratio (95% CI)</th>
<th>Population Attributable Risk %</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>1.10 (0.74-1.64)</td>
<td>0.0415</td>
</tr>
<tr>
<td>25</td>
<td>0.70 (0.39-1.25)</td>
<td>0.0113</td>
</tr>
<tr>
<td>26</td>
<td>0.80 (0.23-2.75)</td>
<td>−0.0003</td>
</tr>
<tr>
<td>27</td>
<td>0.77 (0.63-0.95)</td>
<td>−0.2434</td>
</tr>
<tr>
<td>28</td>
<td>0.44 (0.18-1.04)</td>
<td>−0.0306</td>
</tr>
<tr>
<td>29</td>
<td>1.15 (0.73-1.82)</td>
<td>0.0133</td>
</tr>
<tr>
<td>30</td>
<td>1.55 (0.88-2.73)</td>
<td>0.0616</td>
</tr>
<tr>
<td>31</td>
<td>2.70 (1.47-4.93)</td>
<td>0.1182</td>
</tr>
<tr>
<td>32</td>
<td>11.09 (4.16-29.57)</td>
<td>0.1948</td>
</tr>
<tr>
<td>33</td>
<td>5.76 (1.79-18.57)</td>
<td>0.0694</td>
</tr>
</tbody>
</table>

Sizing of CAG repeats can be challenging. Because none of the studies reported a common sizing standard, we assessed the quality of sizing by examining the frequency distribution of alleles. All studies showed the expected distribution with 22 repeats as the most common allele, followed by alleles with 23 and 27 repeats.\(^1\)\(^,\)\(^2\)\(^,\)\(^12\)\(^,\)\(^22\) The frequency of the 27-repeat allele in individual studies, however, showed significant variability ranging from 0.2%\(^13\) to 3.6%.\(^15\) Similar observations hold true for the 28- and 29-repeat alleles. The observed variation in frequency likely represents unrecognized population substructure and can, in retrospect, explain the spurious association of increased ALS risk for these alleles. It is also possible that these alleles have an increased risk in specific populations only.

Although the stated objectives of the studies included in our meta-analysis were similar, there were some differences in the definition and composition of ALS and control populations. The ALS and control cohorts were recruited from different geographical and ethnic regions including the United States,\(^6\) French Canada/Canada,\(^10\) Belgium and the Netherlands,\(^12\) Germany,\(^11\) Italy,\(^9\)\(^,\)\(^16\) mainland China,\(^15\) Turkey,\(^13\) and Flanders-Belgium.\(^14\) Additionally, the source and type of...
controls differed from one study to the next; controls were defined as neurologically normal,6,11,12 neurologically normal without any known history of neurological disorders,13,15 neurologically normal matched for age,10 neurologically normal matched for age and sex,11,12,19,20 or neurologically normal with no known history of neurological disorders matched for age,11,12,19,20 with a few exceptions.16

Data are shown for 30 (A), 31 (B), 32 (C), and 33 (D) CAG repeats relative to the baseline of 23 or fewer CAG repeats. Horizontal lines indicate 95% CI; arrow (D) indicates an off-scale 95% CI. Dashed line indicates the pooled OR, with the width of the diamond representing the 95% CI. Square size is representative of the size or weight of the individual study.
matched for age and ethnicity,9,10,15 and neurologically normal with exclusion of incipient Parkinson disease through midbrain sonography.11

As shown in both Figure 1 and the eTable in the Supplement, several studies identified pathogenic SCA2 alleles with 34 or more repeats in patients with ALS. In total, 15 patients with ALS and full-length pathologic SCA2 repeats (≥34) presented with a motor neuron syndrome as opposed to a cerebellar syndrome. A recent study described a family with coexisting SCA2 and ALS.24 The proband with ATXN2 CAG repeats of 40 and 22 presented with degenerative ataxia, while her uncle had an upper and lower motor neuron syndrome with CAG repeats of 39 and 22. Familial ALS, caused by mutant ATXN2 alleles, was also observed in a Cuban population.25 This study, in addition to our findings, further supports the notion that expanded ATXN2 CAG repeats can give rise to several neurological disorders with a wide range of phenotypes.

It is not known what causes the shift from a typical cerebellar phenotype to a rare motor phenotype in SCA2. In principle, genetic, environmental, and stochastic factors could play a role. The occurrence of individuals with the typical cerebellar phenotype and the ALS phenotype in the same pedigree argues against importance of cis-acting factors. Nonallelic modifiers (genetic background) cannot be excluded, but the presence of the SCA2 ALS phenotype in different ethnic groups may make this mechanism less likely. The relative rarity of the ALS phenotype suggests that we should stochastic events such as expansion of long-normal ATXN2 alleles in motor neurons during mitosis or DNA repair, both of which have genetic, epigenetic, or environmental underpinnings, may explain the exclusive motor neuron phenotype in some individuals. The possibility that only ataxin 2 molecules with a specific and narrow window of polyglutamine repeat lengths have an abnormal interaction with TDP-43 seems less likely. At this point, however, experimental data are lacking both in model systems and in human patients.

The ATXN2 alleles with 32 and 33 repeats are considered extremely rare and have been associated with very late-onset cerebellar ataxia for 32 repeats in the homozygous state in one publication26 (and by personal observation by one of us [S.M.P.]) and 33 repeats in the heterozygous state in another.2 Consistent with that notion, only 1 allele with 32 repeats and no alleles with 33 repeats were found among approximately 15 000 control alleles. Among approximately 12 000 ALS alleles, there were 43 alleles with 32 repeats and 15 alleles with 33 repeats.

In ALS, it is highly probable that both known and unknown susceptibility genes interact with environmental exposures to modulate disease risk.27 In addition to long-normal ATXN2 repeats, several other genes and environmental factors have been shown to have an association with ALS. Carers such as farming (OR = 2.0; 95% CI, 1.1-3.5)28 and some electoral occupations (OR = 2.3; 95% CI, 1.29-4.09)29 may be related to an increased risk of ALS.

Additionally, having ever smoked a cigarette (OR = 2.0; 95% CI, 1.3-3.2)30 or having a high intake of dietary fat (OR = 2.78; 95% CI, 0.9-8.0)31 have shown association with ALS; however, dietary fiber (OR = 0.3; 95% CI, 0.1-0.7)32 may have a preventive effect. Thus, ORs of 2.70, 11.09, and 5.76 for 31, 32, and 33 ATXN2 repeats, respectively, represent significant risk factors for ALS.

Conclusions

This meta-analysis provides a new framework for interpretation and counseling of patients with regard to ATXN2 alleles. Alleles with up to 30 CAG repeats can be considered normal with no apparent increased risk for neurodegenerative disease. In fact, alleles of 27 and 28 repeats had a slight but significantly reduced risk of ALS. Our analysis confirms that there is an increased risk for ALS with alleles of 31, 32, and 33 repeats. It is not known whether repeat interruptions modify this risk. Counseling also needs to include information that long-normal alleles show increased meiotic instability and can expand to full-mutation alleles in offspring. For alleles of 34 or more repeats, most individuals will develop a cerebellar degeneration but some may present with motor neuron disease (Figure 1). Of note, many patients with SCA2 presenting with cerebellar dysfunction will develop symptoms and signs of motor neuron disease late in the course of their illness.


