Screening for novel hexanucleotide repeat expansions at ALS- and FTD-associated loci

ABSTRACT

Objective: To determine whether GGGGCC (G4C2) repeat expansions at loci other than C9orf72 serve as common causes of amyotrophic lateral sclerosis (ALS).

Methods: We assessed G4C2 repeat number in 28 genes near known ALS and frontotemporal dementia (FTD) loci by repeat-primed PCR coupled with fluorescent fragment analysis in 199 patients with ALS (17 familial, 182 sporadic) and 136 healthy controls. We also obtained blood from patients with ALS4 for evaluation of repeats surrounding the SETX gene locus. C9orf72 expansions were evaluated in parallel.

Results: Expansions of G4C2 repeats in C9orf72 explained 8.8% of sporadic and 47% of familial ALS cases analyzed. Repeat variance was observed at one other locus, RGS14, but no large expansions were observed, and repeat sizes were not different between cases and controls. No G4C2 repeat expansions were identified at other ALS or FTD risk loci or in ALS4 cases.

Conclusions: G4C2 expansions near known ALS and FTD loci other than C9orf72 are not a common cause of ALS.

Neurol Genet 2016;2:e71; doi: 10.1212/NXG.0000000000000071

GLOSSARY

ALS = amyotrophic lateral sclerosis; FTD = frontotemporal dementia; G4C2 = GGGGCC; GWAS = genome-wide association studies; SNP = single-nucleotide polymorphism.

A GGGGCC (G4C2) hexanucleotide repeat expansion in the first intron of C9orf72 is the most common monogenic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD).1,2 There are normally between 2 and 23 G4C2 repeats at this locus. The repeat expands to hundreds in affected individuals,2 although 30 repeats may be sufficient to elicit G4C2-specific pathology.3

A growing body of evidence suggests that the C9orf72 repeat elicits toxicity primarily through gain-of-function mechanisms that are independent of the genetic locus where the repeat resides. Specifically, Drosophila and mouse models of ALS with expression of the repeat outside its normal genomic context suggest that instability and expansion of G4C2 repeats elsewhere in the genome could also cause ALS or FTD. A precedent for this phenomenon is found in spinocerebellar ataxia, in which CAG repeat expansions in a diverse set of genes elicit overlapping clinical phenotypes.4

We therefore hypothesized that cryptic repeat expansions at loci other than C9orf72 could also contribute to ALS and FTD pathogenesis. Using repeat-primed PCR assays, we evaluated whether G4C2 repeats near known ALS and FTD loci identified by linkage analysis or genome-wide association studies (GWAS) exhibited expansions in a cohort of patients with ALS and controls in the University of Michigan ALS Patient Biorepository. Our results confirm that repeat instability and large expansions at C9orf72 are common in sporadic...
ALS in the United States, but expansions at other disease-associated loci are rare in this population and are unlikely to be a common cause of ALS.

**METHODS** Standard protocol approvals, registrations, and patient consents. This study was approved by the Institutional Review Board of the University of Michigan. Individual patients and controls who contributed these DNA samples provided written informed consent via representatives from the Coriell Institute, University of Michigan ALS Patient Biorepository, or individually to a member of the research group.

**ALS and control patient cohorts and DNA isolation.** DNA samples for repeat-primed PCR and genomic PCR were from the following sources: 1 μg of genomic DNA from 199 patients with ALS and 136 healthy controls from the University of Michigan ALS Patient Biorepository. Patients with ALS met the revised El Escorial criteria and were recruited from the University of Michigan ALS Clinic; controls were recruited via the University of Michigan clinical trials Web site (https://umclinicalstudies.org/). Demographic data were analyzed using SAS 9.5 software (SAS Institute Inc., Cary, NC) and summarized in table e-1 at Neurology.org/ng. Two hundred fifty nanograms of genomic DNA from 86 patients with sporadic ALS from Coriell Cell Repository panel #NDPT026 (Coriell Institute) was used for determination of C9orf72 repeat status only. One hundred micrometers of genomic DNA from a patient with ALS and 1 non-ALS sibling from a previously reported large pedigree was extracted from 4 mL of whole blood using a commercial DNA isolation kit (DNeasy Blood & Tissue kit; Qiagen, Netherlands). Genomic DNA from a second published ALS case with a family history was extracted from patient fibroblast cells obtained from a collaborator’s laboratory using the same kit.

**Candidate gene selection.** We performed a BLAST search (National Center for Biotechnology Information) against the human genome for G<sub>C2</sub> repeat sequences using a sequence of 5 G<sub>C2</sub>-pure repeats (GGGGGGCCGGGGCCGGGGCCGGGGCCGGGGCC) as a start query. The identified repeat loci were overlaid with published genetic loci associated with ALS or FT (see a recent review and table 1). We constrained our analysis to repeat loci within 2 Mb of either the mapped critical region for an ALS or FTD candidate gene/locus or with single-nucleotide polymorphisms (SNPs) that achieved statistical significance on GWAS in sporadic ALS cohorts. Additional candidate repeats located more than 2 Mb outside of disease-associated loci were identified by requiring at least 3 pure repeats in a gene with abundant neuronal expression in brain based on BioGPS and Proteomic DB database analysis. Three additional candidate genes with G<sub>C2</sub> repeats within the previously identified critical region of ALS but missed by our initial in silico analysis were added after we obtained access to case samples.

**G<sub>C2</sub> repeat determination.** G<sub>C2</sub> repeat numbers in the longer allele were determined by repeat-primed PCR as previously reported, followed by capillary electrophoresis and fragment analysis. The primer sequences are included in table e-2. The individual reverse primers for each candidate gene were designed using Primer 3 tools (http://bioapps.primer3_www.cgi) and were labeled with either 6-FAM or 5-HEX fluorescent dye. The PCR products were diluted in high-deionized formamide (HiDi formamide; ThermoFisher Scientific, Waltham, WA) containing size standard ROX-1000 (BioVenture, TN). Fragment analysis was performed at the University of Michigan Sequencing Core facility on an ABI 3730 Sequencer. The data were analyzed using GeneMarker (SoftGenetics, PA) to determine the maximal repeat number.

**Sanger sequencing.** For samples with fewer than 35 G<sub>C2</sub> repeats in C9orf72, DNA flanking the repeat region was PCR-amplified with the following primers: C9orf72 forward: 5’-CCG CAG CCT GTA GCA AGC-3’ and C9orf72 reverse: 5’-AGT CGC TAG AGG CGA AAG C-3’ using the same thermal cycling program as the repeat-primed PCR. The PCR products were gel purified and subjected to Sanger sequencing at the University of Michigan Sequencing Core facility to determine the exact repeat number.

For samples with a 31-nucleotide insertion in the VAV2 gene, DNA flanking the repeat region was PCR-amplified with the following primers: VAV2 forward: 5’-GCC CAG GAC AGG AGG CCT CAG CA-3’ and VAV2 reverse: 5’-CTC AGG GCC GGG AGG AAG CAC CT-3’ using the same thermal cycling program as the repeat-primed PCR. The PCR products were gel purified and subjected to Sanger sequencing as described above. For repeats at Huwe1 and RGS14, PCR primers flanking the repeat regions were used to determine specific repeat sizes (see supplemental data).

**Southern blot confirmation of repeat expansions at C9orf72.** For Southern blot analysis, 25-μg aliquots of lymphoblast genomic DNA were obtained for 5 selected Coriell samples. Two additional lymphoblast cell lines were obtained from the Coriell Institute and were grown up to 15 million cells, and approximately 100 μg of lymphoblast genomic DNA was extracted using a DNA isolation kit. Ten fibroblast cell lines with G<sub>C2</sub> repeat expansion and 2 control fibroblast lines (1 sample without expansion and 1 control) were obtained from the University of Michigan ALS Patient Biorepository and 100 μg of fibroblast genomic DNA was extracted from 15 million cells.

**Statistical analysis.** Two-tailed Fisher exact tests or χ<sup>2</sup> tests were performed to test for association of repeat length with ALS and to verify genotype frequencies in the VAV2 gene. For RGS14 repeat length, a 2-tailed nonparametric test was performed to compare the median repeat size differences. For the correlation of C9orf72 repeat size and patient age at onset of ALS, the Pearson correlation test was performed.

**RESULTS** We first conducted an in silico experiment to identify all human genes containing at least 2 G<sub>C2</sub> repeats. Our rationale was that transcribed repeats at such loci could become unstable and lead to disease-causing expansions, unless the chromosomal context of the C9orf72 repeat was critical to disease pathogenesis. Our initial analysis revealed that short G<sub>C2</sub> repeats are quite common in the human genome, with 34% identified as intragenic repeats. To narrow our window of potential sites of repeat expansion to a number that we could readily assay, we compared this in silico list of repeat loci with genetic loci implicated in ALS or FT, either through critical region mapping in familial forms or by GWAS. We reasoned that utilization of GWAS conducted in patient populations similar to our cohort as well as known genetic loci in rare families (which have subsequently been shown to contribute
of the causative repeat. Using this approach, we identified 21 candidate genes, including C9orf72, containing at least 2 G4C2 repeats that are located within 2 Mb of known ALS-FTD loci or loci identified in GWAS by at least 2 associated SNPs (figure 1A).

We developed a modified fluorescent fragment analysis after repeat-primed PCR to evaluate instability at each of these loci, as previously described for C9orf72 repeat expansions.11 We also assessed repeat size at C9orf72. These assays were tested on control samples, and the actual repeat sizes were confirmed by standard PCR and Sanger sequencing. All 21 assays demonstrated accurate detection of repeat size, based on comparison with Sanger sequencing (data not shown). This assay readily detected G4C2 expansions at C9orf72 in samples from ALS cases known to harbor the repeat.

We applied these assays to a sample collection from the University of Michigan ALS Patient Repository containing 199 ALS cases, of which 17 had a family history of ALS, and 136 healthy controls (University of Michigan). The demographic characteristics of this collection are shown in table e-1. Of note, there were no differences in the age at sample collection, sex composition, or racial/ethnic makeup between ALS cases and controls, and both cases and controls had similar frequencies of a family history of a non-ALS neurodegenerative condition (table e-1). In addition, we used samples from Coriell Cell Repository panel NDPT026, containing 86 sporadic ALS cases, for additional screening for repeat expansions in C9orf72 (Coriell Institute). Consistent with published findings, we observed a wide variation of G4C2 repeat length at C9orf72 in both cases and controls (table 1, table e-3, figure 1B). Pathogenic repeat expansions were observed in 47% of the familial cases and 8.8% (17/182) of sporadic ALS cases in the University of Michigan ALS Patient Repository (figure 1B and table e-3). C9orf72 repeat expansions were observed at a similar frequency (6/86; 7.0%) in ALS samples from the Coriell Institute. To validate these findings and determine the actual repeat expansion size in cases with low repeat numbers, genomic PCR and selective Southern blotting analysis were performed. One sample contained exactly 32 repeats in the longer allele, which was confirmed by Sanger sequencing. Other samples exhibited a single band by genomic PCR reflecting the shorter, PCR-amplifiable allele. Southern blotting on genomic DNA extracted from patient-derived fibroblasts or lymphocytes in a subset of cases demonstrated a wide range of pathologic repeat expansion sizes from 200 to more than 1,500 repeats (figure e-1). Repeat expansion size from peripheral blood samples or cultured cells did not correlate with clinical age at onset (R = 0.173 in

Table 1 G4C2 repeat loci evaluated in this study

<table>
<thead>
<tr>
<th>Gene</th>
<th>Location</th>
<th>Repeat no.</th>
<th>ALS/FTD gene loci or SNP</th>
<th>Different in ALS?</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARMC2</td>
<td>Intron 2</td>
<td>2</td>
<td>F IG4/ALS11</td>
<td>No</td>
</tr>
<tr>
<td>ATXN2</td>
<td>Coding 2</td>
<td>2</td>
<td>ATXN2</td>
<td>No</td>
</tr>
<tr>
<td>C9orf76</td>
<td>Intron 2</td>
<td>2</td>
<td>rs4581057, rs10106208</td>
<td>No</td>
</tr>
<tr>
<td>C9orf72</td>
<td>Intron 2-35</td>
<td>C9orf72</td>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>C16orf72</td>
<td>Intron 2</td>
<td>2</td>
<td>rs1551960, rs7185240</td>
<td>No</td>
</tr>
<tr>
<td>CACNA1G</td>
<td>Coding 2</td>
<td>2</td>
<td>rs8068533, rs1061947</td>
<td>No</td>
</tr>
<tr>
<td>DOCK1</td>
<td>Intron 2</td>
<td>2</td>
<td>rs7082776, rs4363506</td>
<td>No</td>
</tr>
<tr>
<td>GPR123</td>
<td>Intron 2-4</td>
<td>OPTN/ALS12</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>HUWE1</td>
<td>Intron 12-14</td>
<td>UBQLN2/ALS15</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>ITPR3</td>
<td>Coding 3-4</td>
<td>3</td>
<td>rs9380343, rs963733</td>
<td>No</td>
</tr>
<tr>
<td>KMT2C</td>
<td>Coding 2</td>
<td>2</td>
<td>rs4725431, rs10260404</td>
<td>No</td>
</tr>
<tr>
<td>MBD2</td>
<td>Coding 3</td>
<td>3</td>
<td>ALS3</td>
<td>No</td>
</tr>
<tr>
<td>NEURL1B</td>
<td>Intron 2</td>
<td>2</td>
<td>rs871503, rs4868146</td>
<td>No</td>
</tr>
<tr>
<td>PAPD5</td>
<td>Coding 3</td>
<td>3</td>
<td>rs12929266, rs1075875</td>
<td>No</td>
</tr>
<tr>
<td>PRRC2B</td>
<td>Coding 2</td>
<td>2</td>
<td>SETX/ALS4</td>
<td>No</td>
</tr>
<tr>
<td>SUP5H</td>
<td>Coding 2</td>
<td>2</td>
<td>rs12327672, rs116969124, rs2297735</td>
<td>No</td>
</tr>
<tr>
<td>SYNM</td>
<td>Intron 3</td>
<td>3</td>
<td>rs931982, rs380347B</td>
<td>No</td>
</tr>
<tr>
<td>TAF4</td>
<td>Coding 6</td>
<td>6</td>
<td>VAPB/ALS8</td>
<td>No</td>
</tr>
<tr>
<td>TIP3</td>
<td>Coding 2</td>
<td>2</td>
<td>rs2063082, rs16917433</td>
<td>No</td>
</tr>
<tr>
<td>TTC28</td>
<td>Coding 2</td>
<td>2</td>
<td>rs6005863, rs5762919</td>
<td>No</td>
</tr>
<tr>
<td>ZNF423</td>
<td>Intron 4</td>
<td>4</td>
<td>rs1075875, rs1505112</td>
<td>No</td>
</tr>
<tr>
<td>BSN</td>
<td>Coding 5</td>
<td>5</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>GAS7</td>
<td>Intron 3</td>
<td>3</td>
<td>NA</td>
<td>No</td>
</tr>
<tr>
<td>RGS14</td>
<td>Intron 4-15</td>
<td>NA</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>STK39</td>
<td>Coding 3-5</td>
<td>3</td>
<td>rs13015447</td>
<td>No</td>
</tr>
<tr>
<td>PRRC2B</td>
<td>Coding 2</td>
<td>2</td>
<td>SETX/ALS4</td>
<td>No</td>
</tr>
<tr>
<td>PRRX2</td>
<td>Intron 3-5</td>
<td>SETX/ALS, rs395119</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>SURF4</td>
<td>Coding 2</td>
<td>2</td>
<td>SETX/ALS4</td>
<td>No</td>
</tr>
<tr>
<td>VAV2</td>
<td>Coding 2</td>
<td>2</td>
<td>SETX/ALS4</td>
<td>No</td>
</tr>
</tbody>
</table>

Abbreviations: ALS = amyotrophic lateral sclerosis; FTD = frontotemporal dementia; G4C2 = GGGGCC; GWAS = genome-wide association study; NA = not applicable; SNP = single-nucleotide polymorphism.

a G4C2 repeats found within 2 Mb of GWAS SNPs from references 12-14, 28-36. b MBD2 has 2 separate G4C2 repeats that are 2 and 3 repeats long. Data shown for the 3-repeat region. The 2-repeat region was invariant. c TAF4 has 6 imperfect G4C2 repeats (C2G4)3CCGGG(C2G4)3. d ZNF423 has 2 separate G4C2 repeats that are 2 and 4 repeats long. Data are shown for the 4-repeat region. The 2-repeat region was invariant. e BSN has 2 separate G4C2 repeats. One is 3 repeats long and the other has 5 imperfect repeats (C0GGCCCGCCGGCCCGGGG). Data are shown for the 5-repeat region. The 3-repeat region was invariant. f HUWE1 has 12 imperfect G4C2 repeats (C2G4)2CCAGG(C2G4)6CCAGGGCCGGTG (C2G4).  

to a percentage of sporadic cases) would identify most repeat expansions that confer a substantial contribution to the genetic burden of ALS in the United States. Consistent with this idea, the C9orf72 locus was identified on multiple GWAS surveys in ALS and FTD in sporadic populations prior to identification

Neurology: Genetics  3

© 2016 American Academy of Neurology. Unauthorized reproduction of this article is prohibited.
Pearson correlation test, $p = 0.572$, figure e-2), consistent with previous reports.\textsuperscript{15}

In contrast to $C9orf72$, $G_{4}C_{2}$ repeats assayed at other loci exhibited little or no variance in repeat number in the University of Michigan ALS Patient Biorepository samples. Of the 20 novel repeats analyzed, 16 showed no variance at all across the cohort (80%). Four genes displayed small variations in repeat number (table 1), but none of these cases contained expansions of more than 2 repeats greater than the

Figure 1  $G_{4}C_{2}$ repeat numbers from controls and patients with ALS

(A) Flow chart of candidate genes selection for this study. (B–H) Graphs show the distribution of repeat sizes from ALS cases and controls for 7 genes. *Significant ($p < 0.05$) difference in the distribution of repeats larger than 33 times between cases and controls by Fisher exact test. ALS = amyotrophic lateral sclerosis; $G_{4}C_{2} = GGGGCC$. 

\textsuperscript{15}Neurology: Genetics  \textcopyright 2016 American Academy of Neurology. Unauthorized reproduction of this article is prohibited.
prevalent alleles. The degree of repeat length variation was not significantly different between control and ALS groups, demonstrating a lack of association between repeat number variation and ALS/FTD in our cohort (figure 1, B–E, \( p < 0.001 \) for \( \text{C9orf72} \), \( p = 0.76 \) for \( \text{GPR123} \), \( p = 0.69 \) for \( \text{HUWE1} \), \( p = 0.374 \) for \( \text{ITPR3} \); \( \chi^2 \) test for \( \text{C9orf72} \) and \( \text{ITPR3} \), and Fisher exact test for \( \text{GPR123} \) and \( \text{HUWE1} \) used for the rest for different allele frequencies).

To expand the initial screening we selected 4 additional candidate genes that are highly expressed in brain and contain at least 3 \( \text{G}_4\text{C}_2 \) repeats in controls. In these candidate genes, no large repeat expansions were found in ALS or control groups (table 1, figure 1, F and G). However, an intronic repeat in \( \text{RGS14} \) displayed a much larger variation than those in other loci studied, with between 4 and 15 repeats present in both controls and ALS cases. The distribution of repeat lengths in \( \text{RGS14} \) was independent of clinical ALS (figure 1F) or \( \text{C9orf72} \) expansion status (control \( 8.49 \pm 1.74 \) vs ALS \( 8.1 \pm 1.93 \), \( p = 0.082 \), 2-tailed student \( t \) test).

As a third approach, we examined additional loci harboring \( \text{G}_4\text{C}_2 \) repeats located near the \( \text{ALS4} \) locus in patients with this disorder. \( \text{ALS4} \) is a form of juvenile ALS with very slow progression that is thought to result from dominant mutations in SETX.\(^6\) The mechanism of these dominant mutations is unclear, because recessive mutations in \( \text{SETX} \) cause a distinct neurodegenerative disorder, ataxia with oculomotor apraxia type 2, which does not include motor neuron degeneration.\(^6\) Within the critical region defined by linkage analysis,\(^6\) there are 4 genes containing short \( \text{G}_4\text{C}_2 \) repeats: \( \text{PRRX2} \), \( \text{PCCR2B} \), \( \text{SURF4} \), and \( \text{VAV2} \) (figure 2A). To investigate whether a cryptic repeat expansion might provide an alternative explanation for the \( \text{ALS4} \) phenotype, we analyzed repeat numbers in 2 patients with \( \text{ALS4} \) and an unaffected control from 2 previously described \( \text{ALS4} \) families.\(^6,7\) No repeat expansions were observed (figure 2B). Of these 4 loci, only 1 (\( \text{PRRC2B} \)) was included in our initial repeat screen in the University of Michigan ALS Patient Repository. Assessment of the additional 3 loci in this larger cohort identified no repeat expansions, only a minor repeat length variance in gene \( \text{PRRX2} \) (table 1, figure 1H). We did observe an amplicon length variant of 31 bp in \( \text{VAV2} \) by fragment analysis in the \( \text{ALS4} \) family (figure 2B). Sanger sequencing identified it as a 31-bp intronic insertion in the longer allele that contains an imperfect duplication of intronic sequence (GCCGGGACGGCCGTGCTGCCTACGGCAGTGGACC). This insertion was common in controls and ALS cases in the University of Michigan ALS Patient Biorepository panel (figure 2C), and appears to be nonpathogenic (\( p = 0.398 \), \( \chi^2 \) test).

**DISCUSSION** \( \text{G}_4\text{C}_2 \) repeat expansions in \( \text{C9orf72} \) are the most common known cause of ALS and FTD. Here we tested whether \( \text{G}_4\text{C}_2 \) repeat instability at other loci could be a common cause of ALS. Our results confirmed that \( \text{G}_4\text{C}_2 \) repeat expansions and instability are common at the \( \text{C9orf72} \) locus in both familial and sporadic ALS cases in 2 patient cohorts (University of Michigan ALS Patient Repository and the NDPT026 ALS collection from Coriell). However, for 27 other genomic loci that harbor \( \text{G}_4\text{C}_2 \) repeats, no pathogenic \( \text{G}_4\text{C}_2 \) expansions were identified. Thus, repeat instability at these alternative loci are unlikely to be major contributors to the genetic burden in ALS in the United States.

This study is not exhaustive in the number of repeat loci assayed or the number of patient and familial cohorts examined. Therefore, we cannot rule out \( \text{G}_4\text{C}_2 \) repeat expansions as rare causes of ALS. For practical reasons, we assayed fewer than 10% of the possible genomic loci that harbor \( \text{G}_4\text{C}_2 \) repeats located near protein-coding genes and a much smaller fraction of the GC-rich repeats in the human genome. Specifically, we focused our analysis on repeats located adjacent to loci already implicated in ALS or FTD. Our rationale for this choice was that utilization of GWAS conducted in similar patient populations as well as known genetic loci in rare families (which have subsequently been shown to contribute to a percentage of sporadic cases) would colocalize with repeat expansions that confer a substantial contribution to the genetic burden of ALS in the United States. Consistent with this idea, the \( \text{C9orf72} \) locus was identified on multiple GWAS surveys in ALS and FTD in sporadic populations prior to identification of the causative repeat. Moreover, we reasoned that cryptic repeat expansions might have escaped identification during the initial analyses of these loci, several of which lack an identified causative gene or a clear pathogenic mechanism for the currently implicated gene. As such, our findings that none of these alternative loci harbor repeat expansions in 199 ALS cases suggest that \( \text{G}_4\text{C}_2 \) repeat expansions at non-\( \text{C9orf72} \) loci are unlikely to be a common cause of ALS in our patient population, given that most common highly penetrant mutations would be detected by this approach. However, it is important to acknowledge that for highly penetrant loci identified in familial ALS cases, it might be necessary to examine the original published cohorts, which might carry rare or private mutations. We took this step for \( \text{ALS4} \) and found no \( \text{G}_4\text{C}_2 \) repeat instability at 4 neighboring loci. Application of this approach to other isolated pedigrees could more definitively rule out expansions at these genetically implicated loci.

The lack of \( \text{G}_4\text{C}_2 \) repeat expansions at alternative gene loci suggests that there could be something about the \( \text{C9orf72} \) locus that promotes instability. The vast majority (75%) of the other hexanucleotide repeats studied here exhibited no variation among
335 individuals, and variance in repeat size was largely absent from the 1,000-genome database for the 321 intragenic G4C2 repeat loci not assayed (data not shown). With one exception, the repeat size variants observed were limited to 1 or 2 repeats. Studies in vitro suggest that stretches of G4C2 as short as 4 repeats are capable of forming intramolecular G quadruplexes, which might contribute to repeat instability through activation of DNA mismatch repair pathways and mistemplating.\(^1\) In general, our data are consistent with this threshold, since only genes with at least 4 repeats had variation of more than 1 repeat. However, some of the G4C2 repeats we studied were larger than this threshold but were invariant, suggesting that the length of the repeat is insufficient to explain instability. While future work will be needed to determine which factors drive selective instability, it may reflect specific aspects of the GC-rich sequence surrounding the repeat at the C9orf72 loci. This concept is supported by the finding that all pathologic repeat expansions at C9orf72 observed to date have occurred on a single haplotype background.\(^1\)

Our approach here focused solely on the role of G4C2 repeat expansions in ALS. However, G4C2 repeat expansions in genes other than C9orf72 may be enriched in neurologic disorders other than ALS. Phenotypes of polyglutamine disorders associated with CAG repeat expansions, for example, are not all the same.\(^1\) In this context, it is interesting to note that the C9orf72 repeat expansion exhibits neurologic symptoms in addition to motor neuron disease and frontotemporal cognitive dysfunction, which are still being explored.\(^1,2\)

We observed noteworthy repeat instability in both cases and controls in one other gene: a G4C2 intronic repeat in RGS14 with a mean repeat length of 9 and a range of 4–14 repeats. RGS14 encodes a member of the regulator of G-protein signaling family of proteins. It is highly expressed in the developing brain and has known functions in hippocampal learning and synaptic plasticity.\(^2\) Although we observed no association between the size of this repeat and clinical ALS, our study cannot rule out a role for expansion in rare ALS families or other racial/ethnic populations. As such, reevaluation of this locus may be worthy of further study in familial ALS, FTD, and other neurodegenerative disorders. Other repeat expansions are implicated in causing or modifying motor neuron disease phenotypes. A GGCCCTG repeat expansion in NOP56 causes spino-cerebellar ataxia type 36 and has motor neuron disease as a clinical feature. Moderate CAG repeat expansions (encoding polyglutamine) in ATXN2 and GCG repeat expansions (encoding polyalanine) in NIP1A are associated with an increased risk of ALS. However, other known pathogenic CAG and CGG repeat expansions do not influence the risk of ALS or FTD, suggesting that the observed effects may be gene specific rather than repeat sequence specific.\(^2,5–7\)

We screened 28 loci harboring G4C2 repeats near known ALS/FTD loci in 199 patients with ALS and 335 controls. Our approach here focused solely on the role of G4C2 repeat expansions in ALS. However, G4C2 repeat expansions in genes other than C9orf72 may be enriched in neurologic disorders other than ALS. Phenotypes of polyglutamine disorders associated with CAG repeat expansions, for example, are not all the same.\(^1\) In this context, it is interesting to note that the C9orf72 repeat expansion exhibits neurologic symptoms in addition to motor neuron disease and frontotemporal cognitive dysfunction, which are still being explored.\(^1,2\)

We observed noteworthy repeat instability in both cases and controls in one other gene: a G4C2 intronic repeat in RGS14 with a mean repeat length of 9 and a range of 4–14 repeats. RGS14 encodes a member of the regulator of G-protein signaling family of proteins. It is highly expressed in the developing brain and has known functions in hippocampal learning and synaptic plasticity.\(^2\) Although we observed no association between the size of this repeat and clinical ALS, our study cannot rule out a role for expansion in rare ALS families or other racial/ethnic populations. As such, reevaluation of this locus may be worthy of further study in familial ALS, FTD, and other neurodegenerative disorders. Other repeat expansions are implicated in causing or modifying motor neuron disease phenotypes. A GGCCCTG repeat expansion in NOP56 causes spino-cerebellar ataxia type 36 and has motor neuron disease as a clinical feature. Moderate CAG repeat expansions (encoding polyglutamine) in ATXN2 and GCG repeat expansions (encoding polyalanine) in NIP1A are associated with an increased risk of ALS. However, other known pathogenic CAG and CGG repeat expansions do not influence the risk of ALS or FTD, suggesting that the observed effects may be gene specific rather than repeat sequence specific.\(^2,5–7\)

We screened 28 loci harboring G4C2 repeats near known ALS/FTD loci in 199 patients with ALS and 335 controls. Our approach here focused solely on the role of G4C2 repeat expansions in ALS. However, G4C2 repeat expansions in genes other than C9orf72 may be enriched in neurologic disorders other than ALS. Phenotypes of polyglutamine disorders associated with CAG repeat expansions, for example, are not all the same.\(^1\) In this context, it is interesting to note that the C9orf72 repeat expansion exhibits neurologic symptoms in addition to motor neuron disease and frontotemporal cognitive dysfunction, which are still being explored.\(^1,2\)

We observed noteworthy repeat instability in both cases and controls in one other gene: a G4C2 intronic repeat in RGS14 with a mean repeat length of 9 and a range of 4–14 repeats. RGS14 encodes a member of the regulator of G-protein signaling family of proteins. It is highly expressed in the developing brain and has known functions in hippocampal learning and synaptic plasticity.\(^2\) Although we observed no association between the size of this repeat and clinical ALS, our study cannot rule out a role for expansion in rare ALS families or other racial/ethnic populations. As such, reevaluation of this locus may be worthy of further study in familial ALS, FTD, and other neurodegenerative disorders. Other repeat expansions are implicated in causing or modifying motor neuron disease phenotypes. A GGCCCTG repeat expansion in NOP56 causes spino-cerebellar ataxia type 36 and has motor neuron disease as a clinical feature. Moderate CAG repeat expansions (encoding polyglutamine) in ATXN2 and GCG repeat expansions (encoding polyalanine) in NIP1A are associated with an increased risk of ALS. However, other known pathogenic CAG and CGG repeat expansions do not influence the risk of ALS or FTD, suggesting that the observed effects may be gene specific rather than repeat sequence specific.\(^2,5–7\)

We screened 28 loci harboring G4C2 repeats near known ALS/FTD loci in 199 patients with ALS and 335 controls. Our approach here focused solely on the role of G4C2 repeat expansions in ALS. However, G4C2 repeat expansions in genes other than C9orf72 may be enriched in neurologic disorders other than ALS. Phenotypes of polyglutamine disorders associated with CAG repeat expansions, for example, are not all the same.\(^1\) In this context, it is interesting to note that the C9orf72 repeat expansion exhibits neurologic symptoms in addition to motor neuron disease and frontotemporal cognitive dysfunction, which are still being explored.\(^1,2\)

We observed noteworthy repeat instability in both cases and controls in one other gene: a G4C2 intronic repeat in RGS14 with a mean repeat length of 9 and a range of 4–14 repeats. RGS14 encodes a member of the regulator of G-protein signaling family of proteins. It is highly expressed in the developing brain and has known functions in hippocampal learning and synaptic plasticity.\(^2\) Although we observed no association between the size of this repeat and clinical ALS, our study cannot rule out a role for expansion in rare ALS families or other racial/ethnic populations. As such, reevaluation of this locus may be worthy of further study in familial ALS, FTD, and other neurodegenerative disorders. Other repeat expansions are implicated in causing or modifying motor neuron disease phenotypes. A GGCCCTG repeat expansion in NOP56 causes spino-cerebellar ataxia type 36 and has motor neuron disease as a clinical feature. Moderate CAG repeat expansions (encoding polyglutamine) in ATXN2 and GCG repeat expansions (encoding polyalanine) in NIP1A are associated with an increased risk of ALS. However, other known pathogenic CAG and CGG repeat expansions do not influence the risk of ALS or FTD, suggesting that the observed effects may be gene specific rather than repeat sequence specific.\(^2,5–7\)
did not find pathogenic repeat expansions other than that previously reported in C9orf72. While future studies in larger cohorts and select families are still needed, our results suggest that G4C2 repeat expansions outside the C9orf72 gene are not a common cause of ALS, supporting the importance of identifying cis elements that explain the genetic instability at the C9orf72 locus.

AUTHOR CONTRIBUTIONS
F.H., M.H.M., D.Z., B.C.C., and P.K.T. conceptualized and designed the study. F.H., J.M.J., and C.F.-R. performed the experiments. F.H., J.M.J., M.H.M., B.C.C., and P.K.T. analyzed the data. M.H.M., E.L.F., S.A.G., and B.C.C. provided patient samples. F.H. and P.K.T. wrote the manuscript. All authors revised the manuscript.

ACKNOWLEDGMENT
The authors thank all patients who provided the samples for this study and Blake Swihart and Jayna Duell, RN, for assistance in study coordination. The authors thank Suzanne Genik for technical assistance in fragment analysis and Zixing Wang for assistance on data processing. They thank the University of Michigan ALS Patient Biorepository for access to patient fibroblasts. The authors thank Michio Hirano for providing patient DNA and Phillip Chance and David Cornblath for helpful discussions. This repository is funded by the Center for Disease Control (CDC) Agency for Toxic Substances and Disease Registry (contract 200-2013-56856), the Program for Neurology Research & Discovery, and the A. Alfred Taubman Medical Research Institute. The authors also thank colleagues in P.K.T.’s laboratory for help in experiments and data interpretation and proofreading of the manuscript.

STUDY FUNDING
This project was funded by NIH R01GM24872 to M.H.M.; grants from the Centers for Disease Control and Prevention (200-2013-56856) and the A. Alfred Taubman Medical Research Institute to E.L.F.; and grants from the Michigan Alzheimer’s Disease Center at the University of Michigan, NIH R01NS08681001, and the Department of Veterans Affairs BLRD 1121BX001841 to P.K.T. The funders played no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

DISCLOSURE
Drs. He and Jones report no disclosures. Dr. Figueroa-Romero has received research support from the Katherine Rayner Fund and the A. Alfred Taubman Medical Research Institute. Dr. Zhang receives intramural funding support from the NIH. Dr. Feldman receives funding support from Novo Nordisk, the NIH, the Juvenile Diabetes Research Foundation, the US-Israel Binational Science Foundation, the American Diabetes Association, and the A. Alfred Taubman Medical Research Institute. Dr. Gourtman receives research funding support from Neuralyst, Inc; the Centers for Disease Control and Prevention/Agency for Toxic Substances and Disease Registry; and the Agency for Toxic Substances and Disease Registry (contract #200-2013-56856). Dr. Meisler receives funding support from the NIH. Dr. Callaghan receives funding support from Impeto Medical Inc and the NIH, has received travel/speaker honoraria from the American Academy of Neurology and the World Federation of Neurology, performs medical consultations for Advance Medical, consults for a PCORI grant, has received honoraria from the British Medical Journal, has worked with the ALS Association, and has acted as a consultant for medical legal cases. Dr. Todd receives funding support from the NIH, the Veterans Administration, the National Ataxia Foundation, the National Fragile X Foundation, and the Muscular Dystrophy Association; serves as a paid consultant for Denali Therapeutics and for an NIH grant; serves as an unpaid consultant for the National Fragile X Foundation; has received travel/speaker honoraria from the American Academy of Neurology, the National Fragile X Foundation, and the Japanese Society of Neurochemistry; has received publishing royalties from UpToDate; and has received license fee payments for antibodies developed by his group from Millipore for distribution to scientists and commercial groups, and for vectors and antibodies developed by his group from Denali Therapeutics. Go to Neurology.org/ng for full disclosure forms.

Received December 10, 2015. Accepted in final form March 1, 2016.

REFERENCES
Screening for novel hexanucleotide repeat expansions at ALS- and FTD-associated loci
Fang He, Julie M. Jones, Claudia Figueroa-Romero, et al.

Neurol Genet 2016;2;
DOI 10.1212/NXG.0000000000000071

This information is current as of May 11, 2016

Updated Information & Services
including high resolution figures, can be found at:
http://ng.neurology.org/content/2/3/e71.full.html

Supplementary Material
Supplementary material can be found at:
http://ng.neurology.org/content/suppl/2016/05/11/2.3.e71.DC1.html

References
This article cites 36 articles, 8 of which you can access for free at:
http://ng.neurology.org/content/2/3/e71.full.html#ref-list-1

Permissions & Licensing
Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
http://ng.neurology.org/misc/about.xhtml#permissions

Reprints
Information about ordering reprints can be found online:
http://ng.neurology.org/misc/addir.xhtml#reprintsus