Published in final edited form as:

Neurobiol Aging. 2014 April ; 35(4): . doi:10.1016/j.neurobiolaging.2013.09.024.

Evaluating non-coding nucleotide repeat expansions in amyotrophic lateral sclerosis

Matthew D. Figley^{1,2}, Anna Thomas^{1,3}, and Aaron D. Gitler^{1,4}

¹Department of Genetics, Stanford University School of Medicine

²Stanford Neuroscience Graduate Program, Stanford University School of Medicine, Stanford, CA 94305

³Presentation High School, San Jose, CA 95125

Abstract

Intermediate-length polyglutamine expansions in ataxin 2 are a risk factor for ALS. The polyglutamine tract is encoded by a trinucleotide repeat in a coding region of the ataxin 2 gene (*ATXN2*). Non-coding nucleotide repeat expansions in several genes are also associated with neurodegenerative and neuromuscular diseases. For example, hexanucleotide repeat expansions located in a non-coding region of *C9ORF72* are the most common cause of ALS. We sought to assess a potential larger role of non-coding nucleotide repeat expansions in ALS. We analyzed the nucleotide repeat lengths of six genes (*ATXN8*, *ATXN10*, *PPP2R2B*, *NOP56*, *DMPK* and *JPH3*), which have been previously associated with neurological or neuromuscular disorders, in several hundred sporadic ALS patients and healthy controls. We report no association between ALS and repeat length in any of these genes, suggesting that variation in the non-coding repetitive regions in these genes does not contribute to ALS.

Keywords

ataxin; polyQ; non-coding; nucleotide repeat expansion; ALS

Introduction

Intermediate-length polyglutamine repeat expansions in ataxin 2 are a risk factor for ALS (Elden et al. 2010) and longer expansions cause spinocerebellar ataxia type 2 (SCA2) (Fischbeck and Pulst 2011). Ataxin 2 is a member of a family of polyglutamine (polyQ) disease proteins, which includes huntingtin. The genes encoding polyQ proteins all harbor trinucleotide repeat tracts that, when expanded past a certain threshold, cause neurodegenerative disease (Orr and Zoghbi 2007). We had previously surveyed other polyQ disease genes in ALS patients and found no significant association between polyQ length and ALS in any of the genes tested beyond ataxin 2 (Lee et al. 2011).

© 2013 Elsevier Inc. All rights reserved.

⁴Correspondence should be addressed to: Aaron D. Gitler, 300 Pasteur Drive, M322 Alway Building, Stanford, CA 94305, 650-725-6991 (phone), 650-725-1534 (fax), agitler@stanford.edu.

Disclosure Statement: The authors state no actual or potential conflicts of interest exist

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Mutations in the *C9ORF72* gene were recently identified as the most common cause of ALS and frontotemporal dementia (DeJesus-Hernandez et al. 2011; Renton et al. 2011). The pathogenic mechanism in *C9ORF72*-linked ALS involves the expansion of a non-coding hexanucleotide repeat, GGGGCC, located in an intron of the *C9ORF72* gene, from a few repeats in unaffected individuals to hundreds or even thousands of copies in affected individuals (DeJesus-Hernandez et al. 2011; Renton et al. 2011). The mechanism by which these expansions cause disease is unclear and may involve loss of function of *C9ORF72*, gain of RNA toxicity from transcribed GGGGCC sequences that accumulate in nuclear and cytoplasmic foci, or even proteotoxicity from the non-conventional translation of GGGGCC into dipeptides in multiple reading frames (Ling et al. 2013). The discovery of *C9ORF72* mutations in ALS indicates that, in principle, non-coding repeat expansions in other genes could also contribute to the disease. Indeed, besides *C9ORF72*-ALS, there are several other neurodegenerative and neuromuscular diseases that are caused by expansions of repetitive DNA in noncoding regions, including spinocerebellar ataxias types 8, 10, 12, 31, and 36, Fragile X-associated tremor/ataxia syndrome (FXTAS), Huntington disease-like 2 (HDL2), and myotonic dystrophy types I and II (Cooper et al. 2009; Li and Bonini 2010).

In this report, we expand the analysis of nucleotide repeats in ALS beyond those found in coding regions (e.g., polyQ proteins) and assess the potential role of non-coding repeats. We analyzed the disease-linked nucleotide repeat sequences in the following genes: *ATXN8* – spinocerebellar ataxia type 8 (SCA8), *ATXN10* – SCA10, *PPP2R2B* – SCA12, *NOP56* – SCA36, *JPH3* – Huntington disease-like 2 (HDL2), and *DMPK* – myotonic dystrophy type I in sporadic ALS patients and healthy controls. This analysis revealed no significant association between nucleotide repeat length and ALS in any of the genes tested, suggesting that variation in the noncoding repetitive regions in these genes does not contribute to ALS.

Materials and Methods

Genomic DNA from human patients with ALS and healthy controls was obtained from the Coriell Institute for Medical Research (Coriell). These genomic DNA samples were from DNA panels from the National Institute of Neurological Disorders and Stroke Human Genetics Resource Center DNA and Cell Line Repository (<http://ccr.coriell.org/ninds>). The submitters that contributed samples are acknowledged in detailed descriptions of each panel: ALS (NDPT025, NDPT026, NDPT103, and NDPT106) and control (NDPT084, NDPT090, NDPT093, NDPT094). The Coriell non-ALS control samples represent unrelated North American Caucasian individuals (ages 36–48 years) who themselves were never diagnosed with a neurologic disorder nor had a first-degree relative with one. We used polymerase chain reaction (PCR)-based fragment analysis to determine the repeat lengths of each gene analyzed, using a similar protocol as in (Lee et al. 2011). PCR primers and cycling conditions are available upon request. It remains possible that our analysis method (PCR fragment analysis) could have missed exceptionally long repeat expansions that are refractory to PCR amplification but we think that this is unlikely since we did not observe an increase frequency of apparently homozygous repeat alleles in ALS cases compared to controls, except for *PPP2R2B*. For that gene there was a statistically significant increase in the number of ALS samples with homozygous alleles compared to controls ($p=0.01$ (169 out of 358 ALS and 127 out of 338 controls)). However, our analysis method for this gene would have detected the presence of any pathogenic expansion (55–78 repeats). We also note that both *ATXN8* and *NOP56* contain another repetitive motif just next to the disease-causing repetitive motif. As with sizing done in previous studies, fragment analysis can only capture the combined number of both repetitive regions (Table 1).

Results

To evaluate the potential contribution of non-coding nucleotide repeat genes to ALS, we defined the nucleotide repeat length in six non-coding repeat genes in ALS patients and healthy controls (Table 1). We selected the following genes: *ATXN8* – spinocerebellar ataxia type 8 (SCA8), *ATXN10* – SCA10, *PPP2R2B* – SCA12, *NOP56* – SCA36, *JPH3* – Huntington disease-like 2 (HDL2), and *DMPK* – myotonic dystrophy type I. For each gene, we used the polymerase chain reaction (PCR) to amplify the nucleotide repeat region, incorporating the fluorescent dye 6-FAM into the 5' PCR primer. We determined the repeat length by resolving PCR amplicons by capillary electrophoresis, followed by size determination with fragment analysis, compared to known size standards. Figure 1A and Table 1 show the genes we analyzed and the normal range of repeat lengths. Figure 1B-G show the distribution of nucleotide repeats in each of the genes analyzed in both cases and controls. We did not observe significant differences in the repeat lengths between ALS cases and healthy controls (*ATXN8* (>23 repeats or >32 repeats), $p=0.49$ and $p=0.08$, respectively; *ATXN10* (>14 repeats), $p=0.41$; *PPP2R2B* (>10 repeats), $p=0.11$; *NOP56* (>9 repeats), $p=0.19$; *JPH3* (>14 repeats), $p=0.11$; *DMPK* (>5 repeats), $p=0.61$). Of note, two ALS patients had *ATXN8* CAG repeat lengths of 70 and 85 (Figure 1B, which border the SCA8-causing repeat range, though *ATXN8* expanded repeats show incomplete penetrance (Day et al. 2000).

Discussion

The discoveries of nucleotide repeat expansions in *ATXN2* and *C9ORF72* as contributors to ALS (Ling et al. 2013) raise the possibility that repeat expansions in other genes might also contribute to the disease. In this report, we evaluated six genes harboring non-coding nucleotide repeats that are expanded in neurological and neuromuscular diseases, and did not find an association with ALS. In a previous report, we assessed seven polyQ-encoding genes and also did not find an association with ALS (Lee et al. 2011). Other groups have also assessed additional nucleotide repeat expansion genes in ALS. Groen et al. (2012) reported no association of the non-coding CGG-repeat expansions in *FMR1* (Groen et al. 2012). Blauw et al. (2012) identified an association between polyalanine-encoding GCG repeats in *NIPA1* and ALS (Blauw et al. 2012). In contrast to our previous findings (Lee et al. 2011), Conforti et al. (2012) reported increased frequency of trinucleotide repeat expansions in the ataxin 1 gene in ALS patients (Conforti et al. 2012).

Our findings suggest two possibilities. First, the effects of repeat expansions in *ATXN2* and *C9ORF72* that contribute to ALS are specific to those particular genes and thus further understanding the normal functions of these genes will provide insight into their role in disease. Second, our analysis of selected candidate nucleotide repeat genes here and previously (Lee et al. 2011) could have certainly missed other repeat-containing genes. Thus, we suggest that a comprehensive genomewide assessment of nucleotide repeat expansions is warranted. Such an analysis will likely require the development and implementation of novel methods to analyze genome sequencing data, since repetitive sequences are typically refractory to standard next generation sequencing alignment approaches (DeJesus-Hernandez et al. 2011; Renton et al. 2011). Maybe the difficulty in mapping long repeat expansions could be turned into an advantage and used to scour genome sequence data for mapping discrepancies that may in fact be hidden repeat expansions.

Acknowledgments

Fragment sizing was performed by the Nucleic Acid and Protein Core facility at the Children's Hospital of Philadelphia. We thank Katherine Carr for helping with PCR setup and fragment analysis. This work was supported by NIH Director's New Innovator Award DP2OD004417 (A.D.G.), and NIH grants R01NS065317 and R01NS073660 (A.D.G.). M.D.F. is supported by the Stanford Genome Training Program.

References

- Blauw HM, et al. NIPA1 polyalanine repeat expansions are associated with amyotrophic lateral sclerosis. *Hum Mol Genet.* 2012; 21:2497–2502. [PubMed: 22378146]
- Conforti FL, et al. Ataxin-1 and ataxin-2 intermediate-length PolyQ expansions in amyotrophic lateral sclerosis. *Neurology.* 2012; 79:2315–2320. [PubMed: 23197749]
- Cooper TA, et al. RNA and disease. *Cell.* 2009; 136:777–793. [PubMed: 19239895]
- Day JW, et al. Spinocerebellar ataxia type 8: clinical features in a large family. *Neurology.* 2000; 55:649–657. [PubMed: 10980728]
- DeJesus-Hernandez M, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron.* 2011; 72:245–256. [PubMed: 21944778]
- Elden AC, et al. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature.* 2010; 466:1069–1075. [PubMed: 20740007]
- Fischbeck KH, Pulst SM. Amyotrophic lateral sclerosis and spinocerebellar ataxia 2. *Neurology.* 2011; 76:2050–2051. [PubMed: 21562249]
- Garcia-Murias M, et al. 'Costa da Morte' ataxia is spinocerebellar ataxia 36: clinical and genetic characterization. *Brain.* 2012; 135:1423–1435. [PubMed: 22492559]
- Groen EJ, et al. CGG-repeat expansion in FMR1 is not associated with amyotrophic lateral sclerosis. *Neurobiol Aging.* 2012; 33:1852 e1851–1853. [PubMed: 22507827]
- Lee T, et al. Evaluating the prevalence of polyglutamine repeat expansions in amyotrophic lateral sclerosis. *Neurology.* 2011; 76:2062–2065. [PubMed: 21562248]
- Li LB, Bonini NM. Roles of trinucleotide-repeat RNA in neurological disease and degeneration. *Trends Neurosci.* 2010; 33:292–298. [PubMed: 20398949]
- Ling SC, et al. Converging Mechanisms in ALS and FTD: Disrupted RNA and Protein Homeostasis. *Neuron.* 2013; 79:416–438. [PubMed: 23931993]
- Orr HT, Zoghbi HY. Trinucleotide repeat disorders. *Annu Rev Neurosci.* 2007; 30:575–621. [PubMed: 17417937]
- Renton AE, et al. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron.* 2011; 72:257–268. [PubMed: 21944779]
- Todd PK, Paulson HL. RNA-mediated neurodegeneration in repeat expansion disorders. *Ann Neurol.* 2010; 67:291–300. [PubMed: 20373340]

Figure 1. Analysis of non-coding nucleotide repeat expansions in ALS and healthy controls. **A)** Schematic of nucleotide repeat containing genes analyzed in this study, including the location and composition of each repeat. **B–G)** Distribution of repeat lengths in nucleotide repeat disease genes in ALS patients and healthy controls. **B)** *ATAXIN8*, **C)** *ATAXIN10*, **D)** *PPP2R2B*, **E)** *NOP56*, **F)** *JPH3*, **G)** *DMPK*. The distribution of repeat lengths for the genes assessed was not different in ALS cases and controls.

Table 1

Non-coding repeat genes analyzed in ALS and controls

Gene	Associated Disease	Repeat Sequence	# ALS patients analyzed	# healthy controls analyzed	observed repeat size range (max. detectable)	Previously reported normal repeat size range*	Disease repeat expansion range*
ATXN8	SCA8	(CTA) _n -(CTG) _n	365	350	15-85 (100)	15-50	~71-1400 (Inc. penetrance)
ATXN10	SCA10	(ATTCT) _n	356	355	9-22 (32)	10-29	400-4500
PPP2R2B	SCA12	(CAG) _n	358	338	9-25 (126)	7-32	51-78
NOP56	SCA36	(GGCCTG) _n -(CGCCCTG) _n	352	342	6-14 (43)	3-14	~650-2500
JPH3	HDL2	(CTG) _n	360	352	5-29 (142)	6-28	>41
DMPK	DM1	(CTG) _n	333	329	5-39 (100)	5-38	>50

* (Todd and Paulson 2010; Garcia-Murias et al. 2012)