Science Translational Medicine AAAS

Submitted Manuscript: Confidential template updated: February 28 2012

Title: Towards allele-specific targeting therapy and pharmacodynamic marker for spinocerebellar ataxia 3

Authors: Mercedes Prudencio[1,2,†], Hector Garcia-Moreno[3,4,†], Karen R. Jansen-West[1,†], Rana Hanna AL-Shaikh[5,†], Tania F. Gendron[1,2], Michael G. Heckman[6], Matthew R. Spiegel[6], Yari Carlomagno[1,2], Lillian M. Daughrity[1], Yuping Song[1], Judith A. Dunmore[1], Natalie Byron[1], Björn Oskarsson[5], Katharine A. Nicholson[7], Nathan P. Staff[8], Sorina Gorcenco[9], Andreas Puschmann[9], João Lemos[10], Cristina Januário[10], Mark S. LeDoux[11], Joseph H. Friedman[12], James Polke[3,4], Robin Labrum[3,4], Vikram Shakkottai[13], Hayley S. McLoughlin[13], Henry L. Paulson[13], Takuya Konno[14], Osamu Onodera[14], Takeshi Ikeuchi[15], Mari Tada[16], Akiyoshi Kakita[16], John D. Fryer [2,17], Ataxia Study Group[18], Zbigniew K. Wszolek[5,*], Paola Giunti[3,4,*], Leonard Petrucelli[1,2,*].

Affiliations:
[1]Department of Neuroscience, Mayo Clinic, Jacksonville, FL, 32224, U.S.A.
[2]Neuroscience Graduate Program, Mayo Clinic Graduate School of Biomedical Sciences, Jacksonville, FL, 32224, USA.
[3]Ataxia Centre, Department of Clinical and Movement Neurosciences, UCL-Queen Square Institute of Neurology, London, WC1N 3BG, United Kingdom.
[4]Ataxia Centre, National Hospital for Neurology and Neurosurgery, University College London Hospitals NHS Trust, London, WC1N 3BG, United Kingdom.
[5]Department of Neurology, Mayo Clinic, Jacksonville, FL, 32224, U.S.A.
[6]Division of Biomedical Statistics and Informatics, Mayo Clinic, Jacksonville, FL, 32224, U.S.A.
[7]Sean M. Healey & AMG Center for ALS, Massachusetts General Hospital (MGH), Boston, MA, 02114, U.S.A.
[8]Department of Neurology, Mayo Clinic, Rochester, MN, 55905, U.S.A.
[9]Lund University, Skåne University Hospital, Department of Clinical Sciences Lund, Neurology, Lund, 22185, Sweden.
[10]Coimbra University Hospital Centre, Coimbra University, Coimbra, 3000-075, Portugal
[11]University of Memphis, and Veracity Neuroscience LLC, Memphis, TN, 38152, U.S.A.
[12]Department of Neurology, Warren Alpert Medical School of Brown University, Providence, RI, 02906, U.S.A.
[13]Department of Neurology, University of Michigan, Ann Arbor, MI, 48109, U.S.A.
[14]Department of Neurology, Brain Research Institute, Niigata University, Niigata, 951-8585, Japan
[15]Department of Molecular Genetics, Brain Research Institute, Niigata University, Niigata, 951-8585, Japan
[16]Department of Pathology, Brain Research Institute, Niigata University, Niigata, 951-8585, Japan
[17]Department of Neuroscience, Mayo Clinic, Scottsdale, AZ, 85259, U.S.A.
[18]A list of members of the Ataxia Study Group is included in Acknowledgements.

*To whom correspondence should be addressed:

Zbigniew K. Wszolek, MD.
Department of Neurology
4500 San Pablo Rd
Jacksonville, FL 32224
Office: +1 904-953-6869
Fax: +1 904-953-0757
e-mail: wszolek.zbigniew@mayo.edu
Paola Giunti, MD, Ph.D.
Department of Clinical and Movement Neurosciences
UCL Queen Square Institute of Neurology
Queen Square London WC1N 3BG
Phone: +44 02034484019
Fax: +44 02034484786
e-mail: p.giunti@ucl.ac.uk

Leonard Petrucelli, Ph.D.
Department of Research, Neuroscience
Mayo Clinic College of Medicine
4500 San Pablo Rd
Jacksonville, FL 32224
Office: +1 904-953-2855
Fax: +1 904-953-6276
e-mail: petrucelli.leonard@mayo.edu

†Equal contributions

Abstract:

Spinocerebellar ataxia 3 (SCA3), caused by a CAG repeat expansion in the ataxin 3 gene (*ATXN3*), is characterized by neuronal polyglutamine (polyQ) ATXN3 protein aggregates. Although there is no cure for SCA3, gene-silencing approaches to reduce toxic polyQ ATXN3 showed promise in preclinical models. However, a major limitation in translating putative treatments for this rare disease to the clinic is the lack of pharmacodynamic markers for use in clinical trials. Here, we developed an immunoassay that readily detects polyQ ATXN3 proteins in human biological fluids, and discriminates SCA3 patients from healthy controls and individuals with other ataxias. We show that polyQ ATXN3 serves as a marker of target engagement in human fibroblasts, which may bode well for its use in clinical trials. Finally, we identified a single nucleotide polymorphism that strongly associates with the expanded allele, thus providing an exciting drug target to abrogate detrimental events initiated by mutant ATXN3. Gene silencing strategies for several repeat diseases are well underway, and our results are expected to improve clinical trial preparedness for SCA3 therapies.

One Sentence Summary: PolyQ ATXN3 is a pharmacodynamic marker for SCA3, and a SNP associated with the expanded allele may serve to specifically target mutant *ATXN3*.

[Main Text:]

Introduction

Spinocerebellar ataxias (SCAs) encompass a group of largely autosomal dominant neurodegenerative disorders associated with progressive degeneration of the cerebellum, spinal cord and other brain regions. This causes gait and limb ataxia, dysarthria, oculomotor dysfunction, as well as variable combinations of pyramidal features (hyperreflexia and/or hyporeflexia), decreased proprioception, and additional types of movement disorders. More than 40 types of SCAs have been described, and many are caused by CAG trinucleotide repeat expansions in the disease-causing gene. SCA3, also known as Machado-Joseph disease, is the most prevalent SCA worldwide. It is characterized by the accumulation of aggregates composed of polyglutamine (polyQ) ataxin 3 (ATXN3) proteins translated from expanded CAG trinucleotide repeats in *ATXN3* (*1-6*). The propensity of polyQ ATXN3 proteins to aggregate into intranuclear and cytosolic neuronal inclusions is thought to confer toxicity through both gain and loss of function mechanisms, affecting a myriad of cellular pathways. However, the exact underlying pathogenic mechanisms of SCA3 remain an active topic of investigation.

At present, there is no treatment to prevent or slow the progression of SCA3, but the main therapeutic strategy being explored is to reduce mutant *ATXN3*. Indeed, in vitro and in vivo studies involving gene-silencing approaches are showing promising results in attenuating polyQ ATXN3-associated toxicity (*7-26*). As potential treatments for SCA3 move towards clinical trials, the need for a reliable method to monitor target engagement and drug efficacy becomes ever more apparent. Current means to assess therapeutic efficacy rely on clinical measures, such as the Scale for the Assessment and Rating of Ataxia (SARA) and others (*27-29*), and magnetic resonance imaging (MRI) to quantify brain atrophy (*3, 30*). Inherent limitations of these measures include examiner bias, the required cooperation by the subject, and the costs associated with MRIs. The identification of molecular markers for SCA3 would thus prove valuable in facilitating the assessment of potential treatments targeting transcripts of the expanded repeat, and in monitoring disease progression. Furthermore, as ATXN3-targeting therapies are developed, it is imperative to identify effective strategies to specifically downregulate mutant *ATXN3*.

Here, we demonstrate that polyQ ATXN3 proteins can be readily detected in human biological fluids by immunoassay, and that polyQ ATXN3 may serve as a pharmacodynamic marker for ATXN3-targeting therapies. Moreover, we found a strong association between a single nucleotide polymorphism (SNP) with the expanded *ATXN3* allele that could be exploited to specifically target mutant *ATXN3* pharmacologically.

Results

PolyQ-ATXN3 proteins accumulate in CSF and plasma from individuals with SCA3, and distinguish SCA3 patients from controls

To quantitatively assess whether polyQ ATXN3 proteins accumulate in human biofluids, we developed an electrochemiluminescent immunoassay employing the Meso Scale Discovery (MSD) system (*31, 32*). Optimized immunoassay conditions were determined by testing various

conditions (different antibody pairs and diluents), and evaluating the limit of blanks, limit of detection and limit of quantitation using a recombinant polyQ ATXN3 protein calibrator as previously described (*33*) (see Methods). Using this immunoassay, we measured polyQ ATXN3 proteins in cerebrospinal fluid (CSF) and/or plasma from: 54 patients with SCA3, 4 asymptomatic *ATXN3* CAG-expansion carriers (pre-symptomatic SCA3), 56 healthy individuals, 21 people with other forms of spinocerebellar ataxia confirmed to be negative for the *ATXN3* CAG-expansion [SCA1 (N=1), SCA2 (N=1), SCA5 (N=2), SCA6 (N=1), cerebellar ataxia cases with a *TRIO* mutation (*34*) (N=4), and other cases negative for any SCA pathogenic mutation (N=12)], and 54 *C9orf72* expansion carriers (an unrelated repeat expansion disease caused by an intronic GGGGCC repeat). For the CSF series, samples from *ATXN3* CAG-expansion carriers were obtained from the Mayo Clinic in Florida (N=24) or Arizona (N=1), the Lund University in Sweden (N=12), the University College of London in the United Kingdom (N=11) and the University of Coimbra in Portugal (N=1). For the main plasma cohort, samples from *ATXN3* CAG-expansion carriers were collected at the Mayo Clinic in Florida (N=31) or Arizona (N=1), the Lund University in Sweden (N=12) and the University of Coimbra in Portugal (N=1). Biofluids from non-SCA3 ataxia patients in the CSF series and the main plasma cohort were collected at Mayo Clinic Florida, whereas samples from healthy individuals and *C9orf72* repeat expansion carriers were collected at the Mayo Clinic Florida, or at the Massachusetts General Hospital and Washington University through their Dominantly Inherited ALS (DIALS) Network study. A detailed description of characteristics of patients and controls in our CSF series and our main plasma cohort is provided in **Table S1**. A second plasma cohort comprised of samples from 30 healthy controls, 40 SCA3 patients and 2 individuals with pre-symptomatic SCA3 was obtained from the University College of London for validation of our findings (**Table S2**).

In our CSF series and our main plasma cohort, polyQ ATXN3 was elevated in patients with SCA3 compared to each control group (no *ATXN3* CAG-expansion) and to all control groups combined both in unadjusted analysis and analysis adjusted for age and sex (median concentration in CSF: 0.13 pg/µl for SCA3 participants, 0.00 pg/µl for controls; median concentration in plasma: 1.28 pg/µl for SCA3 participants, 0.00 pg/µl for controls; **Figs. 1A** and **1B**, **Table S3**). Of interest, even though there were only 4 asymptomatic *ATXN3* CAG-expansion carriers, which lowers the power to detect intergroup differences, CSF polyQ ATXN3 was higher in symptomatic SCA3 patients compared to pre-symptomatic SCA3 individuals (median concentration: 0.07 pg/µl) in unadjusted analysis (P=0.010, **Fig. 1A**), and in analysis adjusted for age and sex (P=0.035, **Table S3**). The latter, however, did not meet our threshold of statistical significance, which was considered P<0.01. However, no difference in plasma polyQ ATXN3 was observed between symptomatic and pre-symptomatic SCA3 cases (**Fig. 1B**, **Table S3**). In our independent validation series of plasma samples, polyQ ATXN3 was also significantly higher in patients with SCA3 (median concentration: 1.08 pg/µl, N=40) compared to healthy controls (median concentration: 0.0 pg/µl, N=30, P<0.001, **Fig. 1B**, **Table S3**).

Next, we assessed the ability of polyQ ATXN3 to discriminate between patients with SCA3 and controls by estimating the area under the receiver operating characteristic curve (AUC). In our CSF series and our main plasma cohort, polyQ ATXN3 perfectly discriminated SCA3 patients from individuals in each control group, and from all control subjects combined, with AUC values equal to 1.00 (**Fig. 1C**, **Table S3**). Similarly, plasma polyQ ATXN3 perfectly discriminated between SCA3 patients and controls in our validation cohort (AUC: 1.00, **Fig. 1C**, **Table S3**). Of note, CSF polyQ ATXN3 also showed good discriminatory ability when comparing SCA3 from asymptomatic *ATXN3* CAG-expansion carriers, with an AUC value of

0.89 (**Fig. 1D**, **Table S3**) whereas plasma polyQ ATXN3 only moderately discriminated between SCA3 patients and asymptomatic *ATXN3* CAG-expansion carriers with an AUC of 0.70 (with 95% confidence intervals ranging from 0.38 to 1.00, **Fig. 1D**, **Table S3**).

Since CSF and plasma polyQ ATXN3 showed similar abilities to differentiate SCA3 patients from controls, we next evaluated whether plasma and CSF polyQ ATXN3 in *ATXN3* CAG-expansion carriers correlate. No such association was observed (Spearman's r: -0.04, P=0.82, N=37). These data, combined with the fact that polyQ ATXN3 in plasma was approximately 10 times higher (median: 1.08-1.28 pg/µl) than in CSF (median: 0.13 pg/µl), suggest that the pool of polyQ ATXN3 in CSF differs from that in blood.

Association of polyQ ATXN3 with clinical features of disease

We subsequently evaluated whether CSF or plasma polyQ ATXN3 associates with clinical features including age of ataxia onset, disease duration (the time elapsed between patient reported symptom onset and sample collection), gait mobility (see Methods) and SARA scores (*27*), and with *ATXN3* CAG-repeat length. In our CSF series, no associations between CSF polyQ ATXN3 and age at onset, disease duration, gait mobility and SARA scores, or *ATXN3* CAG-repeat length were observed (**Table 1**). PolyQ ATXN3 in plasma from SCA3 patients in the main and validation cohorts combined showed only a weak association with an earlier disease onset (N=80, P=0.020), increased gait impairment (N=31, P=0.030), and longer *ATXN3* CAG-repeat length (N=80, P=0.031); however, these associations weakened and were no longer present when correcting for age, sex and disease duration (**Table 1**).

In addition to the above-mentioned clinical features, INAS (Inventory of Non-Ataxia Signs) (*29*), ADL (Activities of Daily Living) (*35*), SCAFI (SCA Functional Index) (*36*) and CCFS (Composite Cerebellar Functional Severity Score) (*37*) data were available for the 40 SCA3 patients in the validation cohort. Nevertheless, no associations were found with any of these test scores and plasma polyQ ATXN3 (**Table 1**).

Neurofilament light also discriminates SCA3 patients from controls

Increased neurofilament light chain (NFL) in CSF and blood is observed in numerous neurological diseases (*38*), including SCA3 (*39, 40*). Thus, we measured NFL in our CSF and plasma series. CSF NFL was significantly elevated in SCA3 patients (median: 3,569 pg/ml) compared to healthy controls (median: 449 pg/ml), and to patients with other forms of ataxia (median: 1,096 pg/ml) in unadjusted analysis (P<0.001, **Fig. 2A**) and adjusted analyses correcting for age and sex (P<0.001, **Table S4**). Notably, CSF NFL was significantly lower in asymptomatic *ATXN3* CAG-expansion carriers (median: 1,352 pg/ml) than in patients with SCA3 in both unadjusted (P<0.001, **Fig. 2A**) and adjusted analyses (P=0.001, **Table S4**). Similar findings were found for plasma NFL both in the main study cohort and the validation cohort (**Fig. 2B**, **Table S4**). We additionally observed that CSF NFL discriminated between SCA3 patients and healthy controls with an AUC value of 1.00 (**Fig. 2C, Table S4**). In a similar fashion, in our main study cohort, plasma NFL discriminated between patients with SCA3 and healthy controls (AUC: 0.89; **Fig. 2C**, **Table S4**) and patients with other forms of ataxia (AUC: 0.94; **Table S4**). In the validation cohort, plasma NFL also discriminated between patients with SCA3 and healthy controls (AUC: 0.97; **Fig. 2C**, **Table S4**). Finally, CSF and plasma NFL discriminated patients with SCA3 from asymptomatic *ATXN3* CAG-expansion carriers, with CSF showing a perfect discriminatory ability (AUC: 1.00) and plasma showing a lower AUC of 0.87 (with 95% confidence intervals ranging from 0.67 to 1.00; **Fig. 2D**, **Table S4**).

As both NFL and polyQ ATXN3 effectively differentiated SCA3 patients from controls, we evaluated, within our SCA3 patients, whether NFL and polyQ ATXN3 associate. No statistically significant correlation between polyQ ATXN3 and NFL was observed in CSF (Spearman's r: 0.07, P=0.67, N=45) or plasma (Spearman's r: -0.004, P=0.97, N=81). We also examined associations between CSF or plasma NFL with clinical features of SCA3 patients. No associations of CSF or plasma NFL with age at onset, disease duration, or SARA, gait mobility, INAS, ADL, SCAFI and CCFS scores were found (**Table S4**).

The length of the ATXN3 CAG expansion associates with age of ataxia onset

Since neither polyQ ATXN3 or NFL showed clear associations with clinical features, we evaluated whether CAG-repeat length associates with clinical variables in SCA3. Among SCA3 patients from our series (N=81), repeat lengths ranged from 51 to 75 repeats, with a median repeat length of 69 and 70 in the main study cohort and the validation cohort, respectively (**Tables S1** and **S2**). Consistent with the literature (*41-43*), we found a significant inverse correlation between *ATXN3* CAG-repeat length and age of ataxia onset in both unadjusted and adjusted analyses (P<0.001, **Table 2**). There was a weak association between longer repeat length and a more severe SARA score in adjusted analyses after correcting for multiple testing (P=0.024, where P<0.0071 is considered significant, **Table 2**). Measurements of INAS, ADL, SCAFI and CCFS scores were also available for SCA3 patients in the validation cohort. Despite the smaller sample size, longer repeat lengths trended with worse performance on these tests (**Table 2**).

A single-nucleotide polymorphism frequently associated with ATXN3 CAG-expanded alleles

Reducing ATXN3 may be the most promising therapeutic strategy to treat SCA3, especially if allele-specific approaches are utilized (*7-20*). Although differentiating the mutant from wild-type allele has its challenges, SNPs may be used to target specific alleles that carry the nucleotide change. Nevertheless, although several SNPs have been identified in the *ATXN3* gene (*44-51*), comprehensive studies assessing their association with the expanded allele are lacking. We thus explored the rs7158733 SNP located ~132 nucleotides downstream of the CAG repeat in the *ATXN3* gene. The presence of this particular SNP turns a tyrosine codon in position 349 of the protein sequence (Y349) into a stop codon (TA$\underline{C}^{1118}$ > TA$\underline{A}^{1118}$), leading to the premature termination of the ATXN3 protein (**Fig. 3A**). The resulting truncated ATXN3 protein, when containing an expanded CAG-repeat, is thought to be more aggregation-prone than other known ATXN3 protein isoforms (*52*). Among non-CAG expansion carriers (N=118) in our main study cohort, 56.8% of people had a "TA$\underline{A}^{1118}$" allele (C>A; p.Y349*) (**Fig. 3B**, **Table S6**). The percentage of control individuals with TA$\underline{A}^{1118}$ was lowest in the group of healthy controls (47.7%, N=44), and highest in *C9orf72*-expansion carriers (63%, N=54). Of particular interest, 87.0% of *ATXN3* CAG-expansion carriers had the TA$\underline{A}^{1118}$ allele (N=40). The presence of TA$\underline{A}^{1118}$ was associated with an increased risk of SCA3 when making comparisons to controls (P<0.001), *C9orf72* expansion carriers (P=0.006), other non-SCA3 ataxia patients (P=0.013) and all individuals without an *ATXN3* CAG-expansion combined (P<0.001), with most of these findings remaining significant after correcting for multiple testing (P<0.0125, **Table S6**). We thus assessed on what allele the SNP is located, and found that 90.0% of SCA3 patients with TA$\underline{A}^{1118}$, or 78.3% of all *ATXN3* mutation carriers, also had the *ATXN3* CAG-repeat expansion in the same allele (**Table S6**). In contrast, the percentage of non-*ATXN3* CAG-expansion carriers with TA$\underline{A}^{1118}$on the longer, although non-expanded, allele was lower (52.4% of controls with an

"TAA^{1118}"allele, or 25.0% of all control cases). The occurrence of TAA^{1118} on the longer allele was associated with at least a nominally increased risk of SCA3 compared to the aforementioned non-SCA3 disease groups (all P≤0.023, **Table S6**). Overall, our data indicates the "TAA^{1118} allele strongly associates with the *ATXN3* CAG-repeat expansion in our cohort of *ATXN3* mutation carriers.

PolyQ ATXN3 may serve to assess variability in patients' responses to ATXN3-targeted therapies

In addition to CSF and plasma, fibroblast cell lines generated from skin biopsies were available for 17 SCA3 patients, 2 asymptomatic *ATXN3* CAG-expansion carriers, 13 healthy controls and 14 individuals with other forms of ataxia. As observed with CSF and plasma, polyQ ATXN3 in fibroblast cell lines from SCA3 patients (median concentration: 4.56 pg/µl) were elevated compared to controls (median concentration: 0.38 pg/µl), and perfectly distinguished SCA3 patients from controls (AUC: 1, **Fig. 4A**, **Table S7**). Furthermore, polyQ ATXN3 in fibroblast lines from *ATXN3* CAG-expansion carriers significantly correlated with polyQ ATXN3 in matching CSF (Spearman's r: 0.86, P=0.0012, N=11, **Fig. 4B**) but not plasma (Spearman's r: -0.06, P=0.81, N=18).

To determine whether fibroblast lines could be used to assess patient responses to *ATXN3*-targeting therapies, we grew 4 control and 5 SCA3 lines in parallel, treated them with *ATXN3* or control siRNAs, and measured polyQ ATXN3 post-treatment by immunoassay. As shown in **Fig. 4C**, *ATXN3* siRNA exposure reduced polyQ ATXN3 levels in all 5 SCA3 lines, and in control lines that had predictably low basal polyQ ATXN3. Knockdown efficiency ranged from 37% to 83% among all lines, and from 48% to 72% among SCA3 lines specifically. These data suggest that varying responses to *ATXN3*-targeted treatments exist among individuals, and that measuring polyQ ATXN3 pre- and post-treatment may be used to assess this variability in clinical trials.

Discussion

Despite linking *ATXN3* to SCA3 over 20 years ago, effective therapies for SCA3 are still lacking. Many factors hinder drug development for neurological disorders like SCA3; however, it is becoming increasingly well recognized that diagnostic, prognostic and pharmacodynamic biomarkers will increase the likelihood of successfully developing treatments by improving many aspects of clinical trial design. Herein, we described an immunoassay to measure polyQ ATXN3 proteins in human biospecimens, and discovered that CSF and plasma polyQ ATXN3 discriminate SCA3 patients from healthy controls and individuals with other forms of ataxia. Although clinical genetic testing remains the leading technique to definitively diagnose SCA3, our ability to measure polyQ ATXN3 in biological fluids combined with our findings that polyQ ATXN3 serves as a marker of target engagement in a human preclinical model, support its use as potential pharmacodynamic marker in clinical trials for potential SCA3 therapies. We also identified a SNP that strongly associates with the expanded allele, and thus provides an exciting target to abrogate detrimental downstream events initiated by mutant *ATXN3*.

We show that CSF and plasma polyQ ATXN3 perfectly discriminated patients with SCA3 from controls. Despite a low powered cohort of asymptomatic *ATXN3* CAG-expansion carriers, CSF polyQ ATXN3 additionally differentiated SCA3 patients from asymptomatic *ATXN3* CAG-expansion carriers (AUC: 0.89). This was not observed for plasma polyQ ATXN3, perhaps because polyQ ATXN3 in plasma largely originates from peripheral tissues, which would also

explain the lack of correlation between CSF and plasma polyQ ATXN3. It also bears mentioning that this lack of association between plasma and CSF polyQ ATXN3 suggests that CSF polyQ ATXN3 should be used when evaluating target engagement of ATXN3-targeted therapies in clinical trials in order to accurately monitor drug pharmacodynamics within the central nervous system.

Like CSF and plasma polyQ ATXN3, CSF and plasma NFL effectively discriminated between SCA3 cases and controls, though AUC measures were slightly lower than those involving polyQ ATXN3. In addition, both CSF and plasma NFL distinguished between SCA3 patients and pre-symptomatic SCA3 individuals (AUC $\geq$ 0.87). These findings are congruent with the notion that NFL is released from injured or degenerating neurons, and further validate the use of NFL to estimate neuroaxonal injury in various neurodegenerative diseases (*38, 53-55*).

Our findings highlight the diagnostic ability of CSF and plasma polyQ ATXN3 for SCA3, but we did not observe strong associations between polyQ ATXN3 and clinical features (age of disease onset, time from onset to sample collection, and scores from clinical tests that assess function and ataxia). Similarly to polyQ ATXN3, no association between NFL and clinical features was detected. This was somewhat surprising given that a prior study reported an association between serum NFL and SARA scores in *ATXN3* CAG-repeat expansion carriers (*40*). The apparent discrepancy between the two studies may be explained by the inclusion of SCA3 patients in early stages of disease in the prior report(*40*), whereas our cohorts included only three symptomatic SCA3 patients with a total SARA score of <5.

It is well-established that a longer repeat length associates with an earlier age of ataxia onset and a more severe clinical presentation in SCA3 patients (*41, 42, 56-58*). In our cohort of patients with SCA3, we similarly noted an association between longer repeat length and an earlier age of disease onset (**Table 2**). It should nonetheless be noted that repeat length alone does not completely explain the variability in age of disease onset among all individuals who develop SCA3, indicating that other factors are also at play (*59*). Whereas it did not meet our cut-off for statistical significance (considered to be P<0.0071), repeat length nominally correlated with scores from diverse clinical tests as previously observed (*60, 61*).

Since an efficient method to assess target engagement of potential therapeutics in clinical trials is expected to facilitate and expedite SCA3 drug discovery, we tested whether polyQ ATXN3 could serve to evaluate patients' responses to *ATXN3*-targeting therapies. To this end, we used cultured patient fibroblasts to quantify the reduction of polyQ ATXN3 following treatment with a siRNA against *ATXN3*. We observed that responses differed among fibroblast cultures from different patients. If these results reflect anticipated differences among patients during a clinical trial, not only may CSF polyQ ATXN3 be useful to confirm target engagement, it may also inform the rational selection of dose and schedule. It is thus notable that polyQ ATXN3 in fibroblasts from SCA3 patients correlated strongly with polyQ ATXN3 in matching CSF, further underscoring the use of CSF as potential pharmacodynamics marker. These findings are timely given that different approaches, including small interfering RNAs and antisense oligonucleotides, are currently being tested for their ability to downregulate ATXN3 proteins (*7-24*). Moreover, a similar biomarker assay is currently being used in ongoing clinical trials for Huntington's disease to measure mutant huntingtin protein in human biofluids (*62*), further highlighting the potential utility of the polyQ ATXN3 assay for SCA3.

Most therapeutic approaches for SCA3 not only are able to downregulate mutant ATXN3 protein, but also wild-type ATXN3. Although depleting the mutant protein at the expense of the normal protein may be beneficial, therapies specifically targeting the mutant allele are expected

to be superior. As such, several strategies designed to specifically target the expanded mutant *ATXN3* gene are being investigated (*7, 12, 13, 15-18, 21-24*). In addition, targeting SNPs with a higher frequency in the SCA3 population that are also associated with the expanded allele is an attractive approach. For instance, the A^{669}-C^{987}-A^{1118} haplotype is reportedly present in 72% of SCA3 families but only in about 20% of the normal population (*51*). Whether the A^{669}-C^{987}-A^{1118} haplotype is linked to the mutant allele is unclear, but the C^{987} SNP was observed in the expanded allele in a cohort of 38 Japanese patients (*63*), and the specific downregulation of the C^{987}-containing allele was successfully achieved in animal models of SCA3 (*21*). At present, the frequency of this SNP co-occurring with the expanded allele in the non-Japanese population is unknown. Here, we investigated the A^{1118} SNP and found it to be highly frequent in our main study cohort of *ATXN3* mutation carriers (87%), with the SNP and the CAG repeat expansion co-existing in the same allele for 90% of these individuals (or 78.3% of all *ATXN3* mutation carriers). Previous studies suggest particular polymorphisms exist in the SCA3 population with two main haplotypes, known as Joseph and Machado lineages (*44*). The Joseph lineage, which is represented by the TTACA^{1118}C haplotype, is the most widespread worldwide (*44, 46, 49, 51*). In contrast, the Machado lineage (GTGGC^{1118}A) is more restricted to populations in mainland Portugal and the Azores (*44, 46*). The geographic origin of the individuals carrying the A^{1118} expanded allele in our study cohort was diverse, and included several countries in Europe (the Netherlands, Germany, Sweden, Portugal), as well as China, Mexico, the Caribbean and the United States. Overall, we found a polymorphism with a widespread distribution that strongly associated with the mutant *ATXN3* allele. Of interest, the A^{1118} SNP was detected in the expanded allele of 23 individuals from 10 different Japanese families (*64*). Thus, therapies targeting the A^{1118} SNP are expected to be especially useful in populations where the Joseph lineage is predominant, and represent effective means to specifically target the mutant allele for the majority of SCA3 patients.

Our study has some limitations. First, as this was a retrospective study, the size of some groups, in particular the number of pre-symptomatic *ATXN3* CAG-expansion carriers, was relatively small, which decreased the power of some analyses. In addition, there was no cohort to validate our CSF polyQ ATXN3 findings, and the clinical data available from patients in the main plasma cohort and the validation plasma cohort did not perfectly overlap (INAS, ADL, SCAFI and CCFS data were only available for the SCA3 patients in the validation cohort, whereas gait mobility scores were only available for the main study cohort). Finally, ours was a cross-sectional study, which may have limited our ability to detect associations of interest. For instance, a longitudinal study may have permitted us to examine correlations between the rate of change in polyQ ATXN3 and the rate of change in gait mobility defects and SARA scores.

In summary, polyQ ATXN3 may serve as means to assess target engagement for ATXN3-based therapies, and the identification of the A^{1118} SNP, which strongly associates with the expanded allele in the majority of SCA3 patients, offers a therapeutic opportunity to specifically target the mutant *ATXN3* allele. Whereas progression and prognostic biomarkers for SCA3 are still needed, our studies are expected to accelerate SCA3 therapy development and their testing in clinical trials.

Materials and Methods

Study design

The goals of this study were to: (i) investigate polyQ ATXN3 as a pharmacodynamic marker using CSF and plasma from *ATXN3* mutation carriers and noncarriers and (ii) perform

validations on an independent cohort of plasma samples from *ATXN3* mutation carriers and controls, (iii) evaluate whether polyQ ATXN3 proteins associate with clinical features of SCA3, (iv) determine the value of NFL as a marker for SCA3, (v) validate reported associations between *ATXN3* CAG repeat length and clinical features in our study cohort, (vi) determine the association of a particular SNP with *ATXN3* CAG-expanded allele, and (vii) assess whether polyQ ATXN3 may serve to determine patients' varying responses to ATXN3-targeted therapies using human fibroblasts. Several statistical analyses were performed, as detailed in the "Statistical analysis" section below. Measurements of polyQ ATXN3 were performed by individuals blinded to genotype and disease status (symptomatic vs. asymptomatic) of the samples. Sample size was determined based on availability at the Mayo Clinic and University College of London. Written informed consent was obtained from all participants, and biological samples were obtained with ethics committee approval. For our fibroblast studies, we examined whether *ATXN3* siRNA induced decreases in polyQ ATXN3, and performed these experiments in triplicate.

Statistical analysis

Initial statistical analyses were performed separately for the main study cohort and the plasma validation cohort. Comparisons of polyQ ATXN3 and NFL between SCA3 patients and the separate groups of healthy controls, *C9orf72* carriers, other ataxia patients, pre-symptomatic SCA3 participants, and the combined group of non-SCA3 subjects (the latter group was examined only for comparisons involving polyQ ATXN3 and did not include the pre-symptomatic SCA3 cases) were made in unadjusted analysis using a Wilcoxon rank sum test, and in adjusted analysis using a stratified van Elteren Wilcoxon rank sum test (*65*), where the test was stratified by both age (≤median or >median) and sex. To further examine the ability of polyQ ATXN3 and NFL to discriminate between SCA3 patients and other patient groups, area under the receiver operating characteristic curve (AUC) was estimated along with a 95% confidence interval (CI); an AUC of 0.5 corresponds to predictive ability equal to that of chance, while an AUC of 1.0 represents perfect predictive ability. Separately for each outcome measure (polyQ ATXN3 or NFL in plasma, CSF, and fibroblasts), we applied a Bonferroni correction to adjust for the number of pair-wise comparisons between groups that were made for analysis of the main cohort.

A number of different analyses were performed in SCA3 subjects only. Correlations between polyQ ATXN3 as measured in CSF, plasma, and fibroblasts (and also between NFL as measured in CSF and plasma) were assessed using Spearman's test of correlation. Associations of clinical features of SCA3 with CSF and plasma polyQ ATXN3 and NFL, and also with *ATXN3* CAG-repeat length, were assessed using Spearman's test of correlation in unadjusted analysis and also using multivariable linear regression models. When data was available in both the main study and validation cohorts, the two cohorts were combined, and multivariable models were initially adjusted only for cohort. Subsequently, full multivariable models that also adjusted for age at CSF/plasma collection, sex, and disease duration at CSF/plasma collection were utilized, with the exception of the models for age at onset which were adjusted for sex and *ATXN3* CAG-repeat length. Regression coefficients and 95% CIs were estimated and are interpreted as the change in the mean outcome (polyQ ATXN3 or NFL; *ATXN3* CAG-repeat length) corresponding to a specified increase in the given clinical feature. Plasma NFL was

examined on the natural logarithm scale in all linear regression analysis involving the plasma validation cohort due to its skewed distribution in that cohort. A Bonferroni correction was applied separately for each outcome measure to adjust for multiple testing in linear regression analysis.

Associations of presence of the "TAA^{1118}" allele for *ATXN3* rs7158733 (regardless of allele and when considering the allele with the longer *ATXN3* CAG-repeat length) with risk of SCA3 (vs. the separate groups of healthy controls, *C9orf72* carriers, other ataxia patients, and the combined group non-SCA3 subjects) were made using logistic regression models that were adjusted for age and sex. Odds ratios (ORs) and 95% CIs were estimated in comparison to the reference SCA3 group. A Bonferroni correction was made to account for each pair-wise comparison vs. SCA3 subjects that was made for the two separate *ATXN3* rs7158733 variables that were considered.

All statistical tests were two-sided, and performed using R Statistical Software (version 3.6.1; R Foundation for Statistical Computing, Vienna, Austria) and GraphPad Prism (version 8.3.0). GraphPad Prism was also utilized to originate all graphs. The statistical significance P-value threshold for a given analysis is displayed in the footnote of the table displaying that given analysis.

Supplementary Materials

Materials and Methods

Fig. S1. Purified rATXN3 protein containing 80 glutamine repeats (Q80).

Table S1. Subject characteristics for the CSF series and the main plasma cohort.

Table S2. Subject characteristics for the plasma validation cohort.

Table S3. Comparisons of polyQ ATXN3 between SCA3 patients and healthy controls, carriers of the C9orf72 repeat expansion, other ataxia patients, and pre-symptomatic SCA3 individuals.

Table S4. Comparisons of NFL between SCA3 patients and healthy controls, other ataxia patients, and pre-symptomatic SCA3 individuals.

Table S5. Associations of NFL with clinical features of SCA3.

Table S6. Association of presence of the "TAA^{1118}" allele for ATXN3 rs7158733 with SCA3 or at risk of SCA3 (pre-symptomatic SCA3) in comparison to healthy controls, *C9orf72* expansion carriers, and other ataxia patients).

Table S7. Comparisons of polyQ ATXN3 in fibroblasts between SCA3 patients and healthy controls, other ataxia patients, and pre-symptomatic SCA3 individuals.

Table S8. Data for each individual data point presented in Figure 4A and B.

Table S9. Data for each individual data point presented in Figure 4C.

References and Notes:

1. G. Haberhausen, M. S. Damian, F. Leweke, U. Muller, Spinocerebellar ataxia, type 3 (SCA3) is genetically identical to Machado-Joseph disease (MJD). *J Neurol Sci* **132**, 71-75 (1995); published online EpubSep (
2. Y. Kawaguchi, T. Okamoto, M. Taniwaki, M. Aizawa, M. Inoue, S. Katayama, H. Kawakami, S. Nakamura, M. Nishimura, I. Akiguchi, et al., CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* **8**, 221-228 (1994); published online EpubNov (10.1038/ng1194-221).
3. N. R. Whaley, S. Fujioka, Z. K. Wszolek, Autosomal dominant cerebellar ataxia type I: a review of the phenotypic and genotypic characteristics. *Orphanet J Rare Dis* **6**, 33 (2011); published online EpubMay 28 (10.1186/1750-1172-6-33).
4. M. M. Evers, L. J. Toonen, W. M. van Roon-Mom, Ataxin-3 protein and RNA toxicity in spinocerebellar ataxia type 3: current insights and emerging therapeutic strategies. *Mol Neurobiol* **49**, 1513-1531 (2014); published online EpubJun (10.1007/s12035-013-8596-2).
5. S. Fujioka, C. Sundal, Z. K. Wszolek, Autosomal dominant cerebellar ataxia type III: a review of the phenotypic and genotypic characteristics. *Orphanet J Rare Dis* **8**, 14 (2013); published online EpubJan 18 (10.1186/1750-1172-8-14).
6. H. S. McLoughlin, L. R. Moore, H. L. Paulson, Pathogenesis of SCA3 and implications for other polyglutamine diseases. *Neurobiol Dis* **134**, 104635 (2020); published online EpubFeb (10.1016/j.nbd.2019.104635).
7. J. Hu, M. Matsui, K. T. Gagnon, J. C. Schwartz, S. Gabillet, K. Arar, J. Wu, I. Bezprozvanny, D. R. Corey, Allele-specific silencing of mutant huntingtin and ataxin-3 genes by targeting expanded CAG repeats in mRNAs. *Nat Biotechnol* **27**, 478-484 (2009); published online EpubMay (nbt.1539 [pii]
10.1038/nbt.1539).
8. L. R. Moore, G. Rajpal, I. T. Dillingham, M. Qutob, K. G. Blumenstein, D. Gattis, G. Hung, H. B. Kordasiewicz, H. L. Paulson, H. S. McLoughlin, Evaluation of Antisense Oligonucleotides Targeting ATXN3 in SCA3 Mouse Models. *Mol Ther Nucleic Acids* **7**, 200-210 (2017); published online EpubJun 16 (10.1016/j.omtn.2017.04.005).
9. E. Rodriguez-Lebron, C. Costa Mdo, K. Luna-Cancalon, T. M. Peron, S. Fischer, R. L. Boudreau, B. L. Davidson, H. L. Paulson, Silencing mutant ATXN3 expression resolves molecular phenotypes in SCA3 transgenic mice. *Mol Ther* **21**, 1909-1918 (2013); published online EpubOct (10.1038/mt.2013.152).
10. F. Huang, L. Zhang, Z. Long, Z. Chen, X. Hou, C. Wang, H. Peng, J. Wang, J. Li, R. Duan, K. Xia, D. M. Chuang, B. Tang, H. Jiang, miR-25 alleviates polyQ-mediated cytotoxicity by silencing ATXN3. *FEBS Lett* **588**, 4791-4798 (2014); published online EpubDec 20 (10.1016/j.febslet.2014.11.013).
11. C. Costa Mdo, K. Luna-Cancalon, S. Fischer, N. S. Ashraf, M. Ouyang, R. M. Dharia, L. Martin-Fishman, Y. Yang, V. G. Shakkottai, B. L. Davidson, E. Rodriguez-Lebron, H. L. Paulson, Toward RNAi therapy for the polyglutamine disease Machado-Joseph disease. *Mol Ther* **21**, 1898-1908 (2013); published online EpubOct (10.1038/mt.2013.144).
12. V. M. Miller, H. Xia, G. L. Marrs, C. M. Gouvion, G. Lee, B. L. Davidson, H. L. Paulson, Allele-specific silencing of dominant disease genes. *Proc Natl Acad Sci U S A* **100**, 7195-7200 (2003); published online EpubJun 10 (10.1073/pnas.1231012100).

13. C. Nobrega, I. Nascimento-Ferreira, I. Onofre, D. Albuquerque, H. Hirai, N. Deglon, L. P. de Almeida, Silencing mutant ataxin-3 rescues motor deficits and neuropathology in Machado-Joseph disease transgenic mice. *PLoS One* **8**, e52396 (2013)10.1371/journal.pone.0052396).
14. J. Liu, H. Pendergraff, K. J. Narayanannair, J. G. Lackey, S. Kuchimanchi, K. G. Rajeev, M. Manoharan, J. Hu, D. R. Corey, RNA duplexes with abasic substitutions are potent and allele-selective inhibitors of huntingtin and ataxin-3 expression. *Nucleic Acids Res* **41**, 8788-8801 (2013); published online EpubOct (10.1093/nar/gkt594).
15. S. Alves, I. Nascimento-Ferreira, G. Auregan, R. Hassig, N. Dufour, E. Brouillet, M. C. Pedroso de Lima, P. Hantraye, L. Pereira de Almeida, N. Deglon, Allele-specific RNA silencing of mutant ataxin-3 mediates neuroprotection in a rat model of Machado-Joseph disease. *PLoS One* **3**, e3341 (2008); published online EpubOct 8 (10.1371/journal.pone.0003341).
16. M. M. Evers, H. D. Tran, I. Zalachoras, B. A. Pepers, O. C. Meijer, J. T. den Dunnen, G. J. van Ommen, A. Aartsma-Rus, W. M. van Roon-Mom, Ataxin-3 protein modification as a treatment strategy for spinocerebellar ataxia type 3: removal of the CAG containing exon. *Neurobiol Dis* **58**, 49-56 (2013); published online EpubOct (10.1016/j.nbd.2013.04.019).
17. Y. Aiba, J. Hu, J. Liu, Q. Xiang, C. Martinez, D. R. Corey, Allele-selective inhibition of expression of huntingtin and ataxin-3 by RNA duplexes containing unlocked nucleic acid substitutions. *Biochemistry* **52**, 9329-9338 (2013); published online EpubDec 23 (10.1021/bi4014209).
18. J. Liu, D. Yu, Y. Aiba, H. Pendergraff, E. E. Swayze, W. F. Lima, J. Hu, T. P. Prakash, D. R. Corey, ss-siRNAs allele selectively inhibit ataxin-3 expression: multiple mechanisms for an alternative gene silencing strategy. *Nucleic Acids Res* **41**, 9570-9583 (2013); published online EpubNov (10.1093/nar/gkt693).
19. M. M. Evers, B. A. Pepers, J. C. van Deutekom, S. A. Mulders, J. T. den Dunnen, A. Aartsma-Rus, G. J. van Ommen, W. M. van Roon-Mom, Targeting several CAG expansion diseases by a single antisense oligonucleotide. *PLoS One* **6**, e24308 (2011)10.1371/journal.pone.0024308).
20. L. J. Toonen, I. Schmidt, M. S. Luijsterburg, H. van Attikum, W. M. van Roon-Mom, Antisense oligonucleotide-mediated exon skipping as a strategy to reduce proteolytic cleavage of ataxin-3. *Sci Rep* **6**, 35200 (2016); published online EpubOct 12 (10.1038/srep35200).
21. Y. Li, T. Yokota, R. Matsumura, K. Taira, H. Mizusawa, Sequence-dependent and independent inhibition specific for mutant ataxin-3 by small interfering RNA. *Ann Neurol* **56**, 124-129 (2004); published online EpubJul (10.1002/ana.20141).
22. C. Nobrega, I. Nascimento-Ferreira, I. Onofre, D. Albuquerque, N. Deglon, L. P. de Almeida, RNA interference mitigates motor and neuropathological deficits in a cerebellar mouse model of Machado-Joseph disease. *PLoS One* **9**, e100086 (2014)10.1371/journal.pone.0100086).
23. J. Hu, K. T. Gagnon, J. Liu, J. K. Watts, J. Syeda-Nawaz, C. F. Bennett, E. E. Swayze, J. Randolph, J. Chattopadhyaya, D. R. Corey, Allele-selective inhibition of ataxin-3 (ATX3) expression by antisense oligomers and duplex RNAs. *Biol Chem* **392**, 315-325 (2011); published online EpubApr (10.1515/BC.2011.045).

24. S. Ouyang, Y. Xie, Z. Xiong, Y. Yang, Y. Xian, Z. Ou, B. Song, Y. Chen, Y. Xie, H. Li, X. Sun, CRISPR/Cas9-Targeted Deletion of Polyglutamine in Spinocerebellar Ataxia Type 3-Derived Induced Pluripotent Stem Cells. *Stem Cells Dev* **27**, 756-770 (2018); published online EpubJun 1 (10.1089/scd.2017.0209).
25. H. S. McLoughlin, L. R. Moore, R. Chopra, R. Komlo, M. McKenzie, K. G. Blumenstein, H. Zhao, H. B. Kordasiewicz, V. G. Shakkottai, H. L. Paulson, Oligonucleotide therapy mitigates disease in spinocerebellar ataxia type 3 mice. *Ann Neurol* **84**, 64-77 (2018); published online EpubJul (10.1002/ana.25264).
26. S. Alves, I. Nascimento-Ferreira, N. Dufour, R. Hassig, G. Auregan, C. Nobrega, E. Brouillet, P. Hantraye, M. C. Pedroso de Lima, N. Deglon, L. P. de Almeida, Silencing ataxin-3 mitigates degeneration in a rat model of Machado-Joseph disease: no role for wild-type ataxin-3? *Hum Mol Genet* **19**, 2380-2394 (2010); published online EpubJun 15 (10.1093/hmg/ddq111).
27. T. Schmitz-Hubsch, S. T. du Montcel, L. Baliko, J. Berciano, S. Boesch, C. Depondt, P. Giunti, C. Globas, J. Infante, J. S. Kang, B. Kremer, C. Mariotti, B. Melegh, M. Pandolfo, M. Rakowicz, P. Ribai, R. Rola, L. Schols, S. Szymanski, B. P. van de Warrenburg, A. Durr, T. Klockgether, R. Fancellu, Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* **66**, 1717-1720 (2006); published online EpubJun 13 (10.1212/01.wnl.0000219042.60538.92).
28. T. Schmitz-Hubsch, S. Tezenas du Montcel, L. Baliko, S. Boesch, S. Bonato, R. Fancellu, P. Giunti, C. Globas, J. S. Kang, B. Kremer, C. Mariotti, B. Melegh, M. Rakowicz, R. Rola, S. Romano, L. Schols, S. Szymanski, B. P. van de Warrenburg, E. Zdzienicka, A. Durr, T. Klockgether, Reliability and validity of the International Cooperative Ataxia Rating Scale: a study in 156 spinocerebellar ataxia patients. *Mov Disord* **21**, 699-704 (2006); published online EpubMay (10.1002/mds.20781).
29. H. Jacobi, M. Rakowicz, R. Rola, R. Fancellu, C. Mariotti, P. Charles, A. Durr, M. Kuper, D. Timmann, C. Linnemann, L. Schols, O. Kaut, C. Schaub, A. Filla, L. Baliko, B. Melegh, J. S. Kang, P. Giunti, B. P. van de Warrenburg, R. Fimmers, T. Klockgether, Inventory of Non-Ataxia Signs (INAS): validation of a new clinical assessment instrument. *Cerebellum* **12**, 418-428 (2013); published online EpubJun (10.1007/s12311-012-0421-3).
30. J. B. Schulz, J. Borkert, S. Wolf, T. Schmitz-Hubsch, M. Rakowicz, C. Mariotti, L. Schols, D. Timmann, B. van de Warrenburg, A. Durr, M. Pandolfo, J. S. Kang, A. G. Mandly, T. Nagele, M. Grisoli, R. Boguslawska, P. Bauer, T. Klockgether, T. K. Hauser, Visualization, quantification and correlation of brain atrophy with clinical symptoms in spinocerebellar ataxia types 1, 3 and 6. *Neuroimage* **49**, 158-168 (2010); published online EpubJan 1 (10.1016/j.neuroimage.2009.07.027).
31. T. F. Gendron, J. Chew, J. N. Stankowski, L. R. Hayes, Y. J. Zhang, M. Prudencio, Y. Carlomagno, L. M. Daughrity, K. Jansen-West, E. A. Perkerson, A. O'Raw, C. Cook, L. Pregent, V. Belzil, M. van Blitterswijk, L. J. Tabassian, C. W. Lee, M. Yue, J. Tong, Y. Song, M. Castanedes-Casey, L. Rousseau, V. Phillips, D. W. Dickson, R. Rademakers, J. D. Fryer, B. K. Rush, O. Pedraza, A. M. Caputo, P. Desaro, C. Palmucci, A. Robertson, M. G. Heckman, N. N. Diehl, E. Wiggs, M. Tierney, L. Braun, J. Farren, D. Lacomis, S. Ladha, C. N. Fournier, L. F. McCluskey, L. B. Elman, J. B. Toledo, J. D. McBride, C. Tiloca, C. Morelli, B. Poletti, F. Solca, A. Prelle, J. Wuu, J. Jockel-Balsarotti, F. Rigo, C. Ambrose, A. Datta, W. Yang, D. Raitcheva, G. Antognetti, A. McCampbell, J. C. Van

Swieten, B. L. Miller, A. L. Boxer, R. H. Brown, R. Bowser, T. M. Miller, J. Q. Trojanowski, M. Grossman, J. D. Berry, W. T. Hu, A. Ratti, B. J. Traynor, M. D. Disney, M. Benatar, V. Silani, J. D. Glass, M. K. Floeter, J. D. Rothstein, K. B. Boylan, L. Petrucelli, Poly(GP) proteins are a useful pharmacodynamic marker for C9ORF72-associated amyotrophic lateral sclerosis. *Sci Transl Med* **9**, (2017); published online EpubMar 29 (10.1126/scitranslmed.aai7866).

32. T. F. Gendron, C. O. N. S. Group, L. M. Daughrity, M. G. Heckman, N. N. Diehl, J. Wuu, T. M. Miller, P. Pastor, J. Q. Trojanowski, M. Grossman, J. D. Berry, W. T. Hu, A. Ratti, M. Benatar, V. Silani, J. D. Glass, M. K. Floeter, A. Jeromin, K. B. Boylan, L. Petrucelli, Phosphorylated neurofilament heavy chain: A biomarker of survival for C9ORF72-associated amyotrophic lateral sclerosis. *Ann Neurol* **82**, 139-146 (2017); published online EpubJul (10.1002/ana.24980).
33. D. A. Armbruster, T. Pry, Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev* **29 Suppl 1**, S49-52 (2008); published online EpubAug (
34. R. Hanna Al Shaikh, T. Caulfield, A. J. Strongosky, M. Matthew, K. R. Jansen-West, M. Prudencio, J. D. Fryer, L. Petrucelli, R. J. Uitti, Z. K. Wszolek, TRIO gene segregation in a family with cerebellar ataxia. *Neurol Neurochir Pol* **52**, 743-749 (2018); published online EpubNov - Dec (10.1016/j.pjnns.2018.09.006).
35. M. E. Mlinac, M. C. Feng, Assessment of Activities of Daily Living, Self-Care, and Independence. *Arch Clin Neuropsychol* **31**, 506-516 (2016); published online EpubSep (10.1093/arclin/acw049).
36. T. Schmitz-Hubsch, P. Giunti, D. A. Stephenson, C. Globas, L. Baliko, F. Sacca, C. Mariotti, M. Rakowicz, S. Szymanski, J. Infante, B. P. van de Warrenburg, D. Timmann, R. Fancellu, R. Rola, C. Depondt, L. Schols, E. Zdzienicka, J. S. Kang, S. Dohlinger, B. Kremer, B. Melegh, A. Filla, T. Klockgether, SCA Functional Index: a useful compound performance measure for spinocerebellar ataxia. *Neurology* **71**, 486-492 (2008); published online EpubAug 12 (10.1212/01.wnl.0000324863.76290.19).
37. S. T. du Montcel, P. Charles, P. Ribai, C. Goizet, A. Le Bayon, P. Labauge, L. Guyant-Marechal, S. Forlani, C. Jauffret, N. Vandenberghe, K. N'Guyen, I. Le Ber, D. Devos, C. M. Vincitorio, M. U. Manto, F. Tison, D. Hannequin, M. Ruberg, A. Brice, A. Durr, Composite cerebellar functional severity score: validation of a quantitative score of cerebellar impairment. *Brain* **131**, 1352-1361 (2008); published online EpubMay (10.1093/brain/awn059).
38. J. Gaiottino, N. Norgren, R. Dobson, J. Topping, A. Nissim, A. Malaspina, J. P. Bestwick, A. U. Monsch, A. Regeniter, R. L. Lindberg, L. Kappos, D. Leppert, A. Petzold, G. Giovannoni, J. Kuhle, Increased neurofilament light chain blood levels in neurodegenerative neurological diseases. *PLoS One* **8**, e75091 (2013)10.1371/journal.pone.0075091).
39. C. Wilke, F. Bender, S. N. Hayer, K. Brockmann, L. Schols, J. Kuhle, M. Synofzik, Serum neurofilament light is increased in multiple system atrophy of cerebellar type and in repeat-expansion spinocerebellar ataxias: a pilot study. *J Neurol* **265**, 1618-1624 (2018); published online EpubJul (10.1007/s00415-018-8893-9).
40. Q. F. Li, Y. Dong, L. Yang, J. J. Xie, Y. Ma, Y. C. Du, H. L. Cheng, W. Ni, Z. Y. Wu, Neurofilament light chain is a promising serum biomarker in spinocerebellar ataxia type 3. *Mol Neurodegener* **14**, 39 (2019); published online EpubNov 4 (10.1186/s13024-019-0338-0).

41. P. Maciel, C. Gaspar, A. L. DeStefano, I. Silveira, P. Coutinho, J. Radvany, D. M. Dawson, L. Sudarsky, J. Guimaraes, J. E. Loureiro, et al., Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am J Hum Genet* **57**, 54-61 (1995); published online EpubJul (
42. H. Maruyama, S. Nakamura, Z. Matsuyama, T. Sakai, M. Doyu, G. Sobue, M. Seto, M. Tsujihata, T. Oh-i, T. Nishio, et al., Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Hum Mol Genet* **4**, 807-812 (1995); published online EpubMay (10.1093/hmg/4.5.807).
43. A. Durr, G. Stevanin, G. Cancel, C. Duyckaerts, N. Abbas, O. Didierjean, H. Chneiweiss, A. Benomar, O. Lyon-Caen, J. Julien, M. Serdaru, C. Penet, Y. Agid, A. Brice, Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. *Ann Neurol* **39**, 490-499 (1996); published online EpubApr (10.1002/ana.410390411).
44. S. Martins, J. Sequeiros, Origins and Spread of Machado-Joseph Disease Ancestral Mutations Events. *Adv Exp Med Biol* **1049**, 243-254 (2018)10.1007/978-3-319-71779-1_12).
45. I. P. D. Costa, B. C. Almeida, J. Sequeiros, A. Amorim, S. Martins, A Pipeline to Assess Disease-Associated Haplotypes in Repeat Expansion Disorders: The Example of MJD/SCA3 Locus. *Front Genet* **10**, 38 (2019)10.3389/fgene.2019.00038).
46. T. Li, S. Martins, Y. Peng, P. Wang, X. Hou, Z. Chen, C. Wang, Z. Tang, R. Qiu, C. Chen, Z. Hu, K. Xia, B. Tang, J. Sequeiros, H. Jiang, Is the High Frequency of Machado-Joseph Disease in China Due to New Mutational Origins? *Front Genet* **9**, 740 (2018)10.3389/fgene.2018.00740).
47. S. Martins, F. Calafell, V. C. Wong, J. Sequeiros, A. Amorim, A multistep mutation mechanism drives the evolution of the CAG repeat at MJD/SCA3 locus. *Eur J Hum Genet* **14**, 932-940 (2006); published online EpubAug (10.1038/sj.ejhg.5201643).
48. S. Martins, C. E. Pearson, P. Coutinho, S. Provost, A. Amorim, M. P. Dube, J. Sequeiros, G. A. Rouleau, Modifiers of (CAG)(n) instability in Machado-Joseph disease (MJD/SCA3) transmissions: an association study with DNA replication, repair and recombination genes. *Hum Genet* **133**, 1311-1318 (2014); published online EpubOct (10.1007/s00439-014-1467-8).
49. R. Sharony, S. Martins, I. P. D. Costa, R. Zaltzman, A. Amorim, J. Sequeiros, C. R. Gordon, Yemenite-Jewish families with Machado-Joseph disease (MJD/SCA3) share a recent common ancestor. *Eur J Hum Genet* **27**, 1731-1737 (2019); published online EpubNov (10.1038/s41431-019-0449-7).
50. G. Stevanin, A. S. Lebre, C. Mathieux, G. Cancel, N. Abbas, O. Didierjean, A. Durr, Y. Trottier, Y. Agid, A. Brice, Linkage disequilibrium between the spinocerebellar ataxia 3/Machado-Joseph disease mutation and two intragenic polymorphisms, one of which, X359Y, affects the stop codon. *Am J Hum Genet* **60**, 1548-1552 (1997); published online EpubJun (10.1016/S0002-9297(07)64251-7).
51. C. Gaspar, I. Lopes-Cendes, S. Hayes, J. Goto, K. Arvidsson, A. Dias, I. Silveira, P. Maciel, P. Coutinho, M. Lima, Y. X. Zhou, B. W. Soong, M. Watanabe, P. Giunti, G. Stevanin, O. Riess, H. Sasaki, M. Hsieh, G. A. Nicholson, E. Brunt, J. J. Higgins, M. Lauritzen, L. Tranebjaerg, V. Volpini, N. Wood, L. Ranum, S. Tsuji, A. Brice, J. Sequeiros, G. A. Rouleau, Ancestral origins of the Machado-Joseph disease mutation: a

worldwide haplotype study. *Am J Hum Genet* **68**, 523-528 (2001); published online EpubFeb (10.1086/318184).
52. D. Weishaupl, J. Schneider, B. Peixoto Pinheiro, C. Ruess, S. M. Dold, F. von Zweydorf, C. J. Gloeckner, J. Schmidt, O. Riess, T. Schmidt, Physiological and pathophysiological characteristics of ataxin-3 isoforms. *J Biol Chem* **294**, 644-661 (2019); published online EpubJan 11 (10.1074/jbc.RA118.005801).
53. L. M. Byrne, F. B. Rodrigues, K. Blennow, A. Durr, B. R. Leavitt, R. A. C. Roos, R. I. Scahill, S. J. Tabrizi, H. Zetterberg, D. Langbehn, E. J. Wild, Neurofilament light protein in blood as a potential biomarker of neurodegeneration in Huntington's disease: a retrospective cohort analysis. *Lancet Neurol* **16**, 601-609 (2017); published online EpubAug (10.1016/S1474-4422(17)30124-2).
54. N. Siller, J. Kuhle, M. Muthuraman, C. Barro, T. Uphaus, S. Groppa, L. Kappos, F. Zipp, S. Bittner, Serum neurofilament light chain is a biomarker of acute and chronic neuronal damage in early multiple sclerosis. *Mult Scler* **25**, 678-686 (2019); published online EpubApr (10.1177/1352458518765666).
55. A. M. Zeitlberger, G. Thomas-Black, H. Garcia-Moreno, M. Foiani, A. J. Heslegrave, H. Zetterberg, P. Giunti, Plasma Markers of Neurodegeneration Are Raised in Friedreich's Ataxia. *Front Cell Neurosci* **12**, 366 (2018)10.3389/fncel.2018.00366).
56. L. B. Jardim, M. L. Pereira, I. Silveira, A. Ferro, J. Sequeiros, R. Giugliani, Neurologic findings in Machado-Joseph disease: relation with disease duration, subtypes, and (CAG)n. *Arch Neurol* **58**, 899-904 (2001); published online EpubJun (10.1001/archneur.58.6.899).
57. Y. Takiyama, S. Igarashi, E. A. Rogaeva, K. Endo, E. I. Rogaev, H. Tanaka, R. Sherrington, K. Sanpei, Y. Liang, M. Saito, et al., Evidence for inter-generational instability in the CAG repeat in the MJD1 gene and for conserved haplotypes at flanking markers amongst Japanese and Caucasian subjects with Machado-Joseph disease. *Hum Mol Genet* **4**, 1137-1146 (1995); published online EpubJul (10.1093/hmg/4.7.1137).
58. P. Giunti, M. G. Sweeney, A. E. Harding, Detection of the Machado-Joseph disease/spinocerebellar ataxia three trinucleotide repeat expansion in families with autosomal dominant motor disorders, including the Drew family of Walworth. *Brain* **118 (Pt 5)**, 1077-1085 (1995); published online EpubOct (10.1093/brain/118.5.1077).
59. C. Globas, S. T. du Montcel, L. Baliko, S. Boesch, C. Depondt, S. DiDonato, A. Durr, A. Filla, T. Klockgether, C. Mariotti, B. Melegh, M. Rakowicz, P. Ribai, R. Rola, T. Schmitz-Hubsch, S. Szymanski, D. Timmann, B. P. Van de Warrenburg, P. Bauer, L. Schols, Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. *Mov Disord* **23**, 2232-2238 (2008); published online EpubNov 15 (10.1002/mds.22288).
60. J. Zhou, L. Lei, X. Liao, J. Wang, H. Jiang, B. Tang, L. Shen, Related factors of ICARS and SARA scores on spinocerebellar ataxia type 3/Machado-Joseph disease. *Zhong Nan Da Xue Xue Bao Yi Xue Ban* **36**, 498-503 (2011); published online EpubJun (10.3969/j.issn.1672-7347.2011.06.005).
61. T. Schmitz-Hubsch, M. Coudert, P. Bauer, P. Giunti, C. Globas, L. Baliko, A. Filla, C. Mariotti, M. Rakowicz, P. Charles, P. Ribai, S. Szymanski, J. Infante, B. P. van de Warrenburg, A. Durr, D. Timmann, S. Boesch, R. Fancellu, R. Rola, C. Depondt, L. Schols, E. Zdienicka, J. S. Kang, S. Dohlinger, B. Kremer, D. A. Stephenson, B. Melegh, M. Pandolfo, S. di Donato, S. T. du Montcel, T. Klockgether, Spinocerebellar ataxia

types 1, 2, 3, and 6: disease severity and nonataxia symptoms. *Neurology* **71**, 982-989 (2008); published online EpubSep 23 (10.1212/01.wnl.0000325057.33666.72).
62. S. J. Tabrizi, B. R. Leavitt, G. B. Landwehrmeyer, E. J. Wild, C. Saft, R. A. Barker, N. F. Blair, D. Craufurd, J. Priller, H. Rickards, A. Rosser, H. B. Kordasiewicz, C. Czech, E. E. Swayze, D. A. Norris, T. Baumann, I. Gerlach, S. A. Schobel, E. Paz, A. V. Smith, C. F. Bennett, R. M. Lane, I.-H. S. S. T. Phase 1-2a, Targeting Huntingtin Expression in Patients with Huntington's Disease. *N Engl J Med* **380**, 2307-2316 (2019); published online EpubJun 13 (10.1056/NEJMoa1900907).
63. R. Matsumura, T. Takayanagi, K. Murata, N. Futamura, M. Hirano, S. Ueno, Relationship of (CAG)nC configuration to repeat instability of the Machado-Joseph disease gene. *Hum Genet* **98**, 643-645 (1996); published online EpubDec (10.1007/s004390050276).
64. J. Goto, M. Watanabe, Y. Ichikawa, S. B. Yee, N. Ihara, K. Endo, S. Igarashi, Y. Takiyama, C. Gaspar, P. Maciel, S. Tsuji, G. A. Rouleau, I. Kanazawa, Machado-Joseph disease gene products carrying different carboxyl termini. *Neurosci Res* **28**, 373-377 (1997); published online EpubAug (10.1016/s0168-0102(97)00056-4).
65. P. H. van Elteren, On the combination of independent two-sample tests of Wilcoxon. *Bulletin of the International Statistical Institute*, 351-361 (1960).
66. K. Schon, N. J. H. van Os, N. Oscroft, H. Baxendale, D. Scoffings, J. Ray, M. Suri, W. P. Whitehouse, P. R. Mehta, N. Everett, L. Bottolo, B. P. van de Warrenburg, P. J. Byrd, C. Weemaes, M. A. Willemsen, M. Tischkowitz, A. M. Taylor, A. E. Hensiek, Genotype, extrapyramidal features, and severity of variant ataxia-telangiectasia. *Ann Neurol* **85**, 170-180 (2019); published online EpubFeb (10.1002/ana.25394).
67. T. Lee, Y. R. Li, A. Chesi, M. P. Hart, D. Ramos, N. Jethava, D. Hosangadi, J. Epstein, B. Hodges, N. M. Bonini, A. D. Gitler, Evaluating the prevalence of polyglutamine repeat expansions in amyotrophic lateral sclerosis. *Neurology* **76**, 2062-2065 (2011); published online EpubJun 14 (10.1212/WNL.0b013e31821f4447).

Acknowledgments: We thank the patients and their families for their contributions. We especially thank the Donald G. and Jodi P. Heeringa Family for their generous gift towards the support of this study. We recognize Ms. Maria del Carmen Labrador from Mayo Clinic International Services in Mexico City, Mexico. We also would like to thank Ms. Kathleen M. Hebenstreit and Ms. Lauren T. Daniel from Mayo Clinic Florida, and Ms. Tammy Y. Finalley from Baptist MD Anderson Cancer Center, for their contribution through CSF collection. We also want to thank the Mayo Clinic Neurodegeneration Lab within the Center for Regenerative Medicine for providing fibroblast lines. We thank Ms. Nita Solanky for her contributions with patients' recruitment and organization. The Ataxia Study Group is composed of: Christin Karremo (Lund University, Lund, Sweden), Inês Gomes (University of Coimbra, Coimbra, Portugal), John N. Caviness and Mark R. Pittelkow (Mayo Clinic, Scottsdale, AZ, U.S.A.), Jan Aasly (Norwegian University of Science and Technology, Trondheim, Norway), Ronald F. Pfeiffer and Venka Veerappan (Oregon Health & Science University, Portland, OR, U.S.A.), Eric R. Eggenberger, William D. Freeman, Josephine F. Huang, Ryan J. Uitti, Klaas J. Wierenga, Iris V. Marin Collazo and Philip W. Tipton (Department of Neurology, Mayo Clinic, Jacksonville, FL, U.S.A.), Jay A. van Gerpen (University of Alabama at Birmingham, AL, U.S.A.), Marka van Blitterswijk and Guojun Bu (Department of Neuroscience, Mayo Clinic, Jacksonville, FL, U.S.A.).

Funding: This work was supported by the National Institutes of Health/National Institute of Neurological Disorder and Stroke [R35NS097273 (L.P.); P01NS084974 (L.P.); P01NS099114 (T.F.G., L.P.); R01NS088689 (L.P.); R35NS097263 (L.P.); R21NS084528 (L.P.)]; U01NS10667 (H.S.M., H.L.P.), National Institute of Environmental Health Services [R01ES20395 (L.P.)]; Department of Defense [ALSRP AL130125 (L.P.)]; Mayo Clinic Foundation (L.P.); Amyotrophic Lateral Sclerosis Association (L.P., M.P., K.A.N.), Robert Packard Center for ALS Research at Johns Hopkins (L.P.), Target ALS Foundation (T.G., L.P., K.A.N.), Mayo Clinic Center for Regenerative Medicine (Z.K.W.), gifts from The Sol Goldman Charitable Trust (Z.K.W.) and the Donald G. and Jodi P. Heeringa Family (Z.K.W., L.P.), the Haworth Family Professorship in Neurodegenerative Diseases fund (Z.K.W.), The Albertson Parkinson's Research Foundation (Z.K.W.), SCA Network Sweden (A.P., S.G.), ALF, Sweden (S.G.), Region Skåne, Sweden (A.P.)., Muscular Dystrophy Association (K.A.N), ALS Finding A Cure (K.A.N.), North Thames Clinical Research Network, CRN (P.G.), MRC (P.G.), JPND (P.G.), ESMI (P.G.), ESMI grant (H.G.), Cure SCA3 and Fathers Foundation (H.G.). P.G. works at University College London Hospitals/University College London, which receives a proportion of funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding scheme.

Additional funding sources to disclose but not pertinent to this study include grants from Abbvie, Inc. (M15-562 and M15-563 to Z.K.W.), Biogen, Inc. (228PD201 to Z.K.W.),) and Biohaven Pharmaceuticals, Inc. (BHV4157-206 and BHV3241-301 to Z.K.W.).

Author contributions: M.P, K.R.J., and L.P. designed the study. M.P., H.G., K.R.J., T.F.G., Y.C., L.M.D., N.B. performed experiments, R.H.A., L.M.D., Y.S., J.A.D, B.O., K.A.N., N.P.S., S.G., A.P., J.L., C.J., M.S.L, J.H.F., J.P., R.L., V.G.S., H.S.M., H.L.P., T.K., O.O., T.I., M.T., A.K., J.D.F., members of the Ataxia Study Group, Z.K.W., P.G., and L.P. contributed with patient referrals, patient samples, clinical information, sample collection and processing. M.G.H and M.R.S performed the statistical analysis. M.P, H.G., K.J.W., R.H.A., T.F.G, M.G.H., M.S.L., Z.K.W., P.G. and L.P. wrote the manuscript. All authors read and approved the manuscript.

Competing interests: B.O. has consulted for Biogen, Medicinova, Mitsubishi, Amylyx and Tsumura; K.A.N. has performed consulting for Alector, AI Therapeutics, Biogen, MT Pharma, Avanir Pharmaceuticals, and Biohaven, A.P. receives reimbursement from Elsevier for serving as Associate Editor for Parkinsonism and Related Disorders; H.L.P. has consulted for Exicure, and collaborated with Biogen and Ionis; M.S.L., has served as a consultant for US WorldMeds, and speaker for US WorldMeds, Acadia Pharmaceuticals, Teva Pharmaceutical Industries, Kyowa Kirin, Amneal Pharmaceuticals, and Acorda Therapeutics; J.H.F. serves as consultant for Acorda and Concert Pharmaceuticals; R.F.P receives honoraria from Acadia and Acorda, has a research grant from Acorda, and receives royalties for book editing from CRC Press and Humana Press; P.G. is a consultant for Reata Pharmaceutical, Triplet Therapeutics, and Vico Therapeutics; Z.K.W. serves as PI of the Mayo Clinic American Parkinson Disease Association (APDA) Information and Referral Center; L.P. is a consultant for Expansion Therapeutics.**Data and material availability:** all the data used for this study are included in the main text or in the supplementary material.

Figure captions:

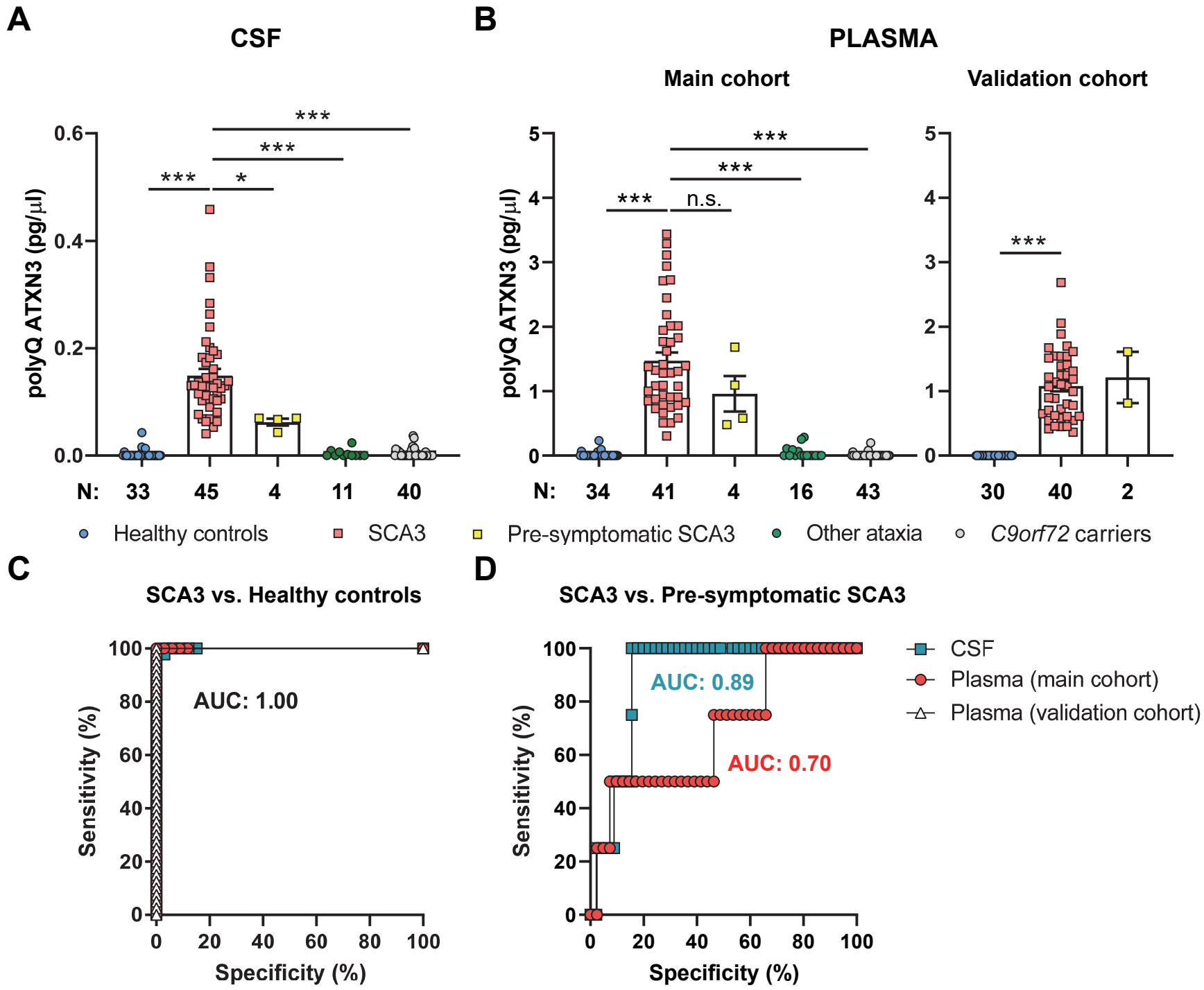
Fig. 1. **PolyQ-ATXN3 proteins accumulate in CSF and plasma from individuals with SCA3, and distinguish SCA3 patients from controls. A-B**) PolyQ ATXN3in CSF (**A**) and plasma (**B**) was measured using our immunoassay as described (see Methods). The number of cases per study group is included below the graphs. Graphs represent mean±SEM. Statistical differences represent unadjusted and adjusted analyses. Complete statistical analyses can be found in **Table S3.** *P < 0.05, ***P < 0.001, n.s.: non-significant differences. **C-D**) Area under the receiver operating characteristic curve (AUC) for polyQ ATXN3 between SCA3 and healthy controls (**C**) or pre-symptomatic SCA3 carriers (**D**) in both CSF and plasma. Additional AUC values with other controls are included in **Table S3.**

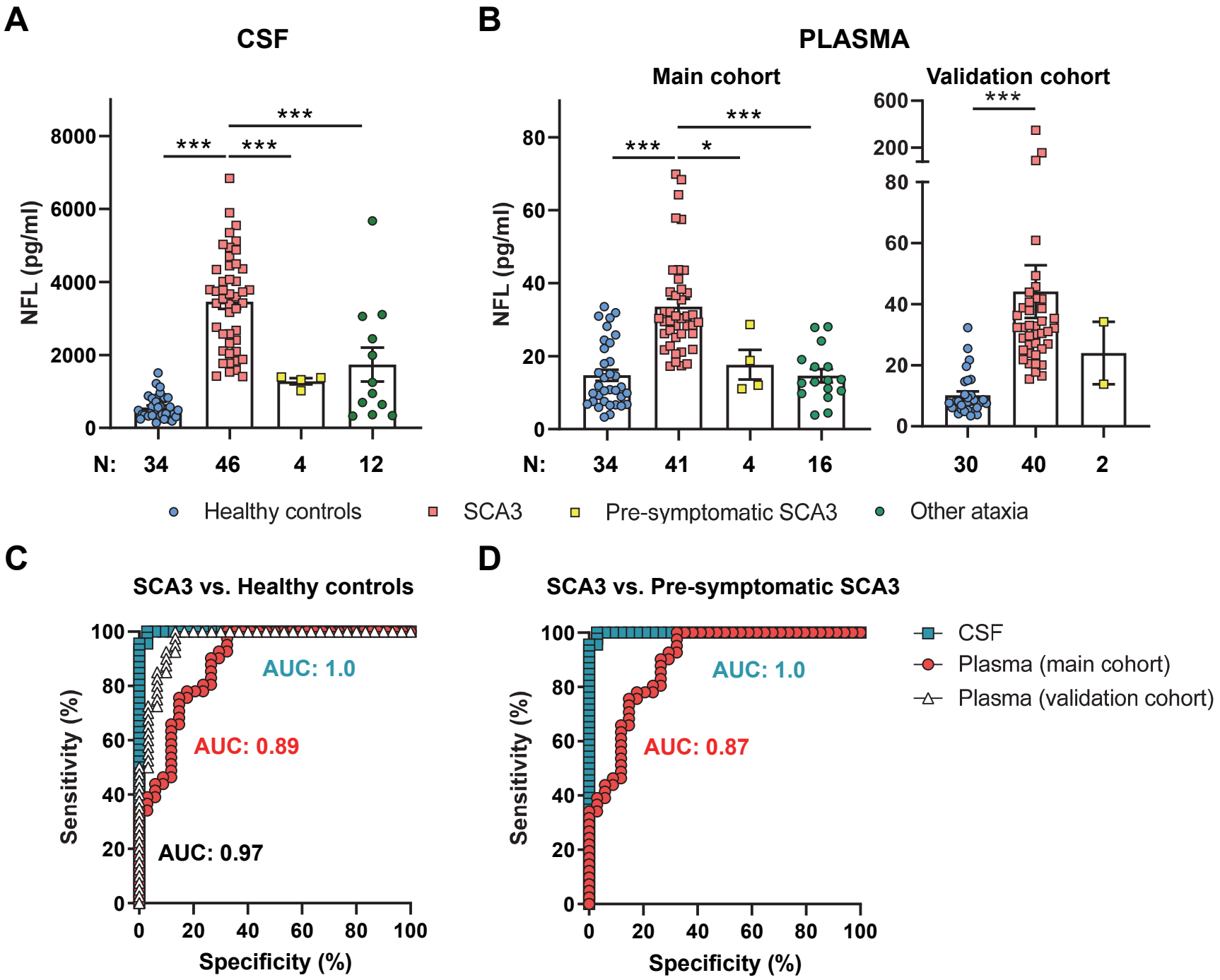
Fig. 2. Neurofilament light can also discriminate SCA3 patients from controls. A-B) NFL in CSF (**A**) and plasma (**B**) was measured using the commercial NF-light kit on the Simoa HD-1 analyzer (Quanterix) (see Methods). The number of cases per study group is included below the graphs. Graphs represent mean±SEM. Statistical differences represent unadjusted and adjusted analyses. Complete statistical analyses can be found in **Table S4**. *P < 0.05, ***P < 0.001, n.s.: non-significant differences. **C-D**) Area under the receiver operating characteristic curve (AUC) for NFL between SCA3 and healthy controls (**C**) or pre-symptomatic SCA3 carriers (**D**) in both CSF and plasma. Additional AUC values with other controls are included in **Table S4**.

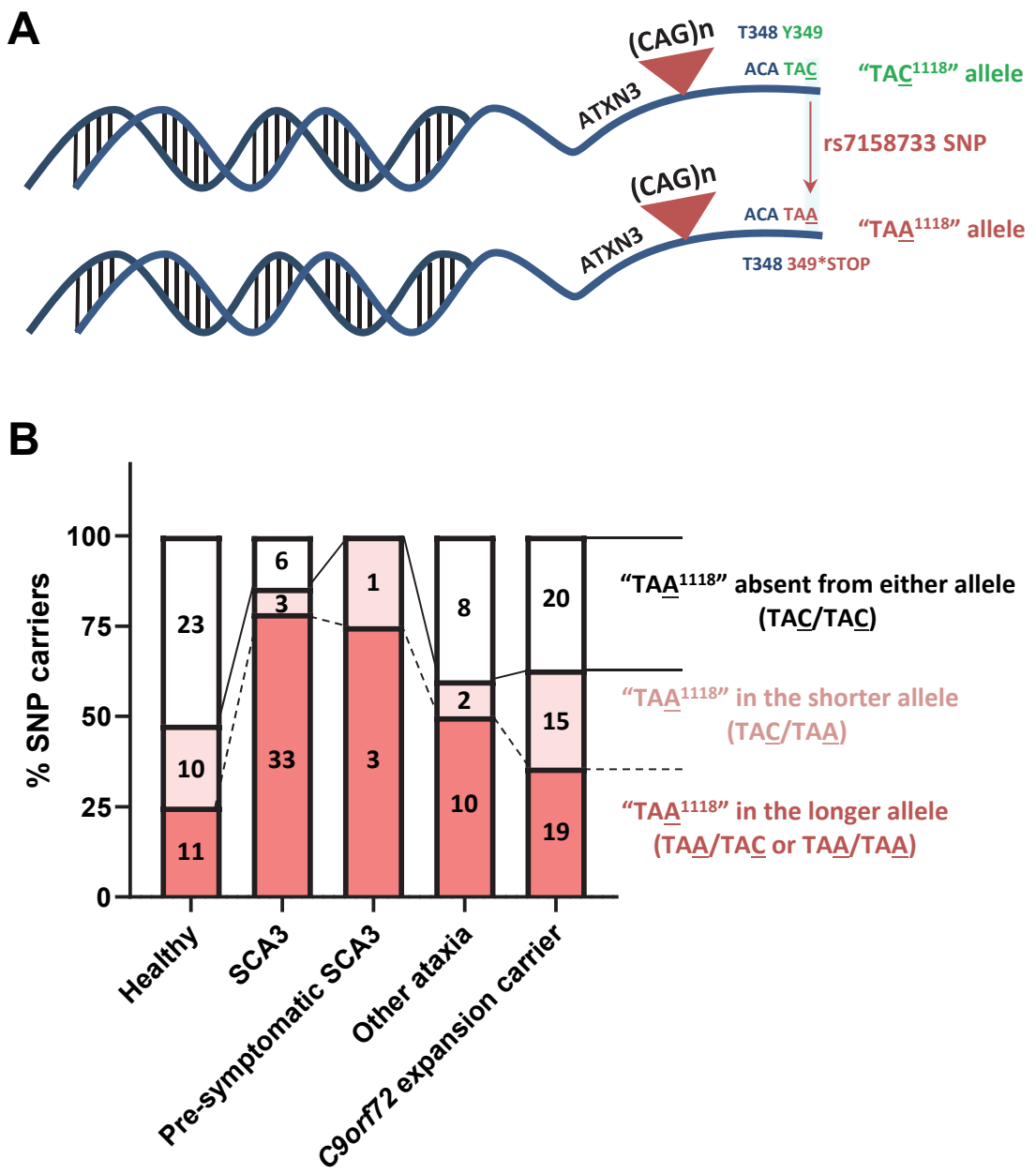
Fig. 3. The rs7158733 SNP is highly prevalent in our SCA3 patient cohort and associates with the expanded allele. A) Schematic depicting the reference "TA$\underline{C}^{1118}$" allele (top) in which the presence of C^{1118} allows corresponding codon (TA$\underline{C}^{1118}$) to encode for tyrosine at position 349 of the protein sequence (Y349). The "TA$\underline{A}^{1118}$" allele (bottom) is characterized by the single nucleotide change to A^{1118}, encoding for a TA$\underline{A}^{1118}$ stop codon at position 349 and leading to generation of a truncated variant of ATXN3. **B**) Bar graph indicating the % of cases with either the "TA$\underline{A}^{1118}$" (salmon) or "TA$\underline{C}^{1118}$" allele (white). The exact number of cases is shown within each section of the bar graph. The darker salmon color depicts the proportion of cases that have the "TA$\underline{A}^{1118}$" on the expanded allele.

Fig. 4. PolyQ ATXN3 proteins accumulate in human fibroblasts derived from *ATXN3* CAG-expansion carriers and treatment with *ATXN3* siRNA leads to varying degrees

of polyQ ATXN3 downregulation. A) Fibroblast lines were grown from 17 SCA3 patients, 13 healthy controls, 14 non-SCA3 ataxia patients and 2 asymptomatic *ATXN3* CAG-expansion carriers (pre-symptomatic SCA3). PolyQ ATXN3 proteins were measured using our developed immunoassay. Bars represent mean±SEM. ****P<0.001, n.s.: not-significant differences. **B**) Spearman correlation analyses of polyQ ATXN3 in fibroblasts and matching CSF from *ATXN3* CAG-expansion carriers (N=11). The solid line represents the estimated regression line. **C**) Four healthy control lines and 5 SCA3 lines were grown in parallel and the same amount of cells were treated with control or *ATXN3* siRNA for 5 days. Lysates were extracted and polyQ ATXN3 was measured using our immunoassay. Bars represent mean±SEM of at least three replicates per line. Information on *ATXN3* CAG-repeat length and the presence of the rs7158733 SNP on each allele from matching blood samples, as well as the percentage of polyQ ATXN3 knockdown from the different fibroblast lines are indicated underneath the graph. Statistical analyses were performed using a 2-way ANOVA followed by Sidak's multiple comparisons test. *P<0.05, **P<0.005, ****P<0.0001. Individual-level data is shown in **Tables S8** and **S9**.


A
CSF
B
PLASMA
Main cohort
Validation cohort
polyQ ATXN3 (pg/µl)
***
***
***
*
n.s.
N: 33 45 4 11 40
N: 34 41 4 16 43
30 40 2
Healthy controls
SCA3
Pre-symptomatic SCA3
Other ataxia
C9orf72 carriers
C
SCA3 vs. Healthy controls
AUC: 1.00
Sensitivity (%)
Specificity (%)
D
SCA3 vs. Pre-symptomatic SCA3
AUC: 0.89
AUC: 0.70
CSF
Plasma (main cohort)
Plasma (validation cohort)


A
CSF
NFL (pg/ml)
8000
6000
4000
2000
0
***
***
***
N: 34 46 4 12
B
PLASMA
Main cohort
Validation cohort
NFL (pg/ml)
80
60
40
20
0
***
*
***
N: 34 41 4 16
600
200
80
60
40
20
0
***
30 40 2
Healthy controls
SCA3
Pre-symptomatic SCA3
Other ataxia
C
SCA3 vs. Healthy controls
Sensitivity (%)
100
80
60
40
20
0
AUC: 1.0
AUC: 0.89
AUC: 0.97
0 20 40 60 80 100
Specificity (%)
D
SCA3 vs. Pre-symptomatic SCA3
Sensitivity (%)
100
80
60
40
20
0
AUC: 1.0
AUC: 0.87
0 20 40 60 80 100
Specificity (%)
CSF
Plasma (main cohort)
Plasma (validation cohort)


A
(CAG)n
ATXN3
T348 Y349
ACA TAC
"TAC1118" allele
rs7158733 SNP
(CAG)n
ATXN3
ACA TAA
"TAA1118" allele
T348 349*STOP
B
% SNP carriers
100
75
50
25
0
23
10
11
6
3
33
1
3
8
2
10
20
15
19
"TAA1118" absent from either allele
(TAC/TAC)
"TAA1118" in the shorter allele
(TAC/TAA)
"TAA1118" in the longer allele
(TAA/TAC or TAA/TAA)
Healthy
SCA3
Pre-symptomatic SCA3
Other ataxia
C9orf72 expansion carrier

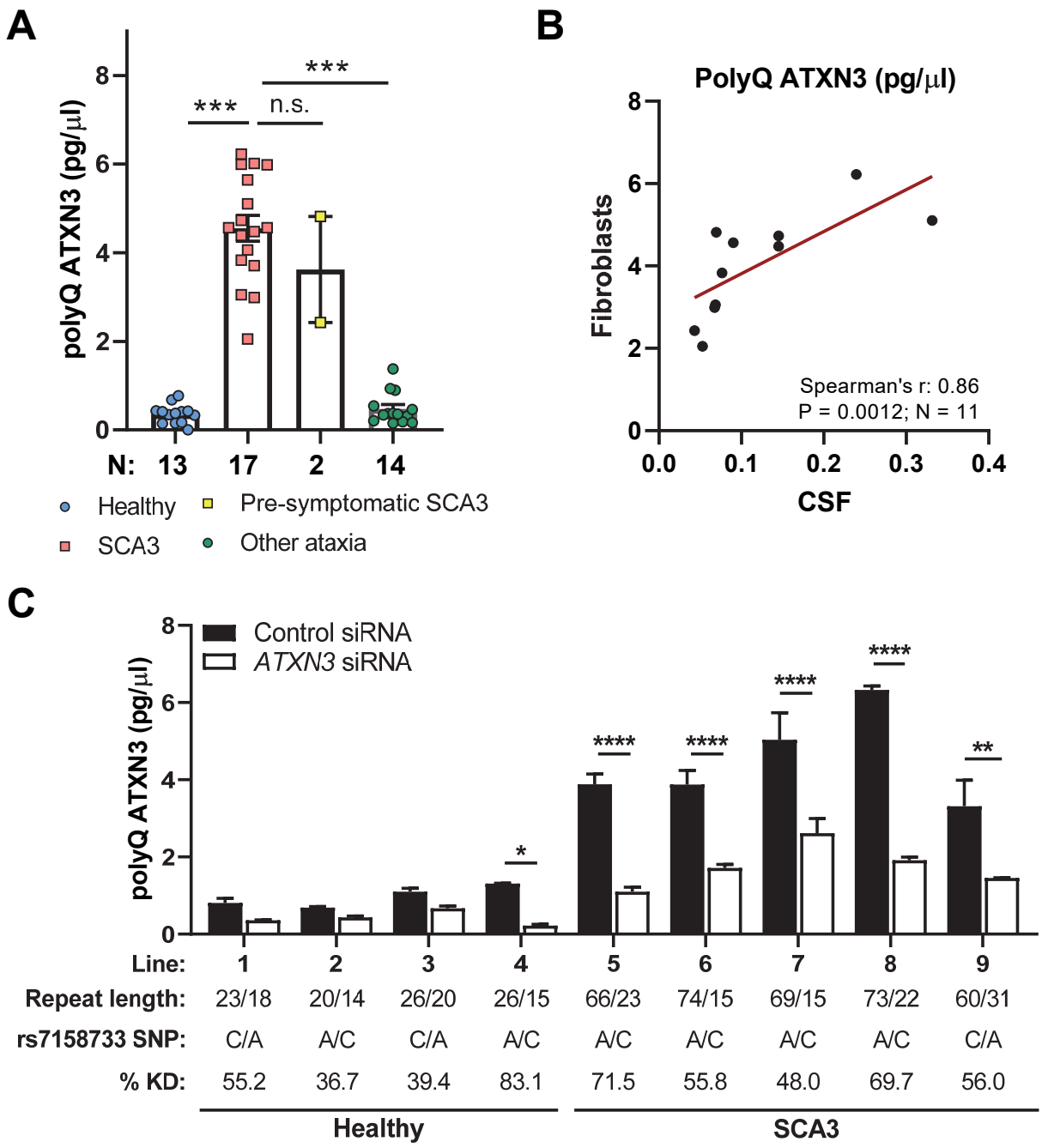

A
polyQ ATXN3 (pg/µl)
***
***
n.s.
N: 13 17 2 14
Healthy
SCA3
Pre-symptomatic SCA3
Other ataxia
B
PolyQ ATXN3 (pg/µl)
Fibroblasts
CSF
Spearman's r: 0.86
P = 0.0012; N = 11
C
Control siRNA
ATXN3 siRNA
polyQ ATXN3 (pg/µl)
Line: 1 2 3 4 5 6 7 8 9
Repeat length: 23/18 20/14 26/20 26/15 66/23 74/15 69/15 73/22 60/31
rs7158733 SNP: C/A A/C C/A A/C A/C A/C A/C A/C C/A
% KD: 55.2 36.7 39.4 83.1 71.5 55.8 48.0 69.7 56.0
Healthy
SCA3

Table 1. Associations of CSF and plasma polyQ ATXN3 with clinical features of SCA3.

		Unadjusted analysis		Adjusting only for cohort		Full multivariable analysis	
Clinical feature	**N**	**Spearman's *r* (95% CI)**	**P-value**	**Regression coefficient (95% CI)**	**P-value**	**Regression coefficient (95% CI)**	**P-value**
Association with CSF polyQ ATXN3							
Age of onset (10 year increase)	45	0.21 (-0.12, 0.51)	0.16	N/A	N/A	0.02 (-0.01, 0.06)	0.18
Disease duration (5 year increase)	45	-0.28 (-0.53, 0.01)	0.063	N/A	N/A	-0.02 (-0.04, 0.00)	0.053
ATXN3 CAG-repeat length (5 unit increase)	45	-0.11 (-0.40, 0.20)	0.47	N/A	N/A	0.00 (-0.05, 0.05)	0.91
SARA total (5 unit increase)	44	-0.22 (-0.50, 0.09)	0.15	N/A	N/A	-0.01 (-0.02, 0.01)	0.54
Gait mobility score (1 unit increase)	24	-0.09 (-0.51, 0.33)	0.66	N/A	N/A	0.01 (-0.03, 0.05)	0.55
Association with plasma polyQ ATXN3 (combined cohort)							
Age of onset (10 year increase)	80	-0.26 (-0.46, -0.02)	0.020	-0.16 (-0.28, -0.04)	0.011	-0.13 (-0.30, 0.03)	0.11
Disease duration (5 year increase)	80	0.13 (-0.10, 0.33)	0.27	0.04 (-0.06, 0.14)	0.45	0.08 (-0.03, 0.18)	0.15
ATXN3 CAG-repeat length (5 unit increase)	80	0.24 (0.01, 0.46)	0.031	0.19 (0.03, 0.36)	0.024	0.06 (-0.17, 0.29)	0.61
SARA total (5 unit increase)	80	0.03 (-0.17, 0.23)	0.82	0.00 (-0.09, 0.10)	0.96	-0.02 (-0.14, 0.10)	0.72
Gait mobility score (1 unit increase)	31	0.39 (0.04, 0.68)	0.030	N/A	N/A	0.19 (-0.05, 0.44)	0.11
INAS total (1 unit increase)	40	0.10 (-0.20, 0.37)	0.54	N/A	N/A	-0.03 (-0.11, 0.06)	0.53
ADL total (5 unit increase)	40	0.06 (-0.27, 0.37))	0.73	N/A	N/A	-0.08 (-0.24, 0.09)	0.37
SCAFI (1 unit increase)	34	-0.22 (-0.50, 0.11)	0.21	N/A	N/A	-0.11 (-0.33, 0.10)	0.28
CCFS (0.5 unit increase)	32	0.07 (-0.30, 0.42)	0.70	N/A	N/A	0.31 (-0.83, 1.45)	0.58

CI = confidence interval. In unadjusted analysis, P-values result from Spearman's test of correlation. In multivariable analysis, regression coefficients, 95% CIs, and P-values result from linear regression models. Regression coefficients are interpreted as the change in mean polyQ ATXN3 corresponding to the increase given in parenthesis. In the full multivariable analysis, models were adjusted for cohort (main or validation, only if the given feature was assessed in both cohorts), age at plasma/CSF collection, sex, and disease duration with the exceptions of the model for age at onset which was adjusted for cohort, sex, and *ATXN3* CAG-repeat length, and the model for CCFS which was adjusted for sex and disease duration. For clinical features that were assessed in only the main or the validation cohorts, but not both, analysis adjusting only for cohort was not performed. After applying a Bonferroni adjustment for multiple testing separately for each polyQ ATXN3 outcome measure, P-values <0.010 (associations with CSF polyQ ATXN3) and <0.0056 (associations with plasma polyQ ATXN3) were considered as statistically significant.

Table 2. Associations of *ATXN3* CAG-repeat length with clinical features of SCA3 in the combined cohort.

		Unadjusted analysis		Adjusting only for cohort[1]		Full multivariable analysis	
Clinical feature	**N**	**Spearman's *r* (95% CI)**	**P-value**	**Regression coefficient (95% CI)**	**P-value**	**Regression coefficient (95% CI)**	**P-value**
Age of onset (10 year increase)	80	-0.68 (-0.80, -0.51)	<0.001	-0.50 (-0.63, -0.37)	<0.001	-0.50 (-0.62, -0.37)	<0.001
SARA total (5 unit increase)	80	0.20 (-0.03, 0.41)	0.075	0.09 (-0.03, 0.22)	0.14	0.14 (0.02, 0.25)	0.024
Gait mobility score (1 unit increase)	32	0.34 (-0.03, 0.65)	0.057	N/A	N/A	0.04 (-1.18, 1.26)	0.95
INAS total (1 unit increase)	39	0.24 (-0.09, 0.51)	0.14	N/A	N/A	0.09 (-0.01, 0.20)	0.081
ADL total (5 unit increase)	39	0.22 (-0.10, 0.51)	0.18	N/A	N/A	0.26 (0.05, 0.46)	0.015
SCAFI (1 unit increase)	34	-0.16 (-0.53, 0.19)	0.35	N/A	N/A	-0.28 (-0.57, 0.00)	0.051
CCFS (0.5 unit increase)	31	-0.05 (-0.41, 0.28)	0.77	N/A	N/A	1.72 (0.15, 3.30)	0.034

CI = confidence interval. In unadjusted analysis, P-values result from Spearman's test of correlation. In multivariable analysis, regression coefficients, 95% CIs, and P-values result from linear regression models. Regression coefficients are interpreted as the change in mean outcome (allele length) corresponding to the increase given in parenthesis (continuous variables). In the full multivariable analysis, models were adjusted for cohort (main or validation), age at plasma/CSF collection, sex, and disease duration with the exception of the model for age at onset which was adjusted for cohort and sex, and the model for CCFS which was adjusted for cohort (main or validation), sex, and disease duration. [1]For clinical features that were assessed in only the main cohort or the validation cohort, but not both, analysis adjusting only for cohort was not performed. After applying a Bonferroni adjustment for multiple testing, P-values <0.0071 were considered as statistically significant.

Materials and Methods

Standard protocol approval and patient consent

All individuals agreed to be in the study and biological samples were retrieved after obtaining informed consent from all participants. All protocols were approved by the Mayo Clinic Institution Review Board and Ethics Committee, and the Ethical review boards responsible for the sites where patients were included. All collaborative sites were also approved by the Mayo Clinic IRB Committee and material transfer agreements were developed.

Clinical assessment of SCA3 participants

Diagnosis of SCA3 was independently ascertained by trained movement disorder specialists upon examinations, and neurological evaluations were recorded. In particular, neurological evaluations were comprised of the Scale for the Assessment and Rating of Ataxia (SARA) (*27*), the Inventory of Non-Ataxia Signs (INAS) (*29*), Activities of Daily Living (ADL) (*35*), the SCA Functional Index (SCAFI) (*36*), the Composite Cerebellar Functional Severity Score (CCFS) (*37*), and a gait mobility score. Of note, the gait mobility score was determined by R.H., based on the Schon et al scale (*66*) to provide a classification for the current level of mobility for each participant. This scale is comprised of five levels of motor functionality: 0-asymptomatic, 1-impaired gait (no assistance required), 2-impaired gait (utilizes cane when ambulating), 3-requires walker, 4-wheelchair bound, 5-bedridden. In addition, study participants were evaluated by family history of neurological conditions, history of falls, vision changes, swallowing difficulties, loss of sensation, presence/absence of paresthesia, and urine and stool incontinence. On semi-quantitative examinations, patients were observed for signs of eye movement impairment, nystagmus, frontal lobe release signs, cognitive and behavioral dysfunction, pyramidal signs and lower motor dysfunction, ataxia, peripheral neuropathy, and parkinsonism. Commercial genetic test results were also obtained from SCA3 participants. Finally, we also measured *ATXN3* CAG-repeat length to evaluate associations with age of onset and other disease metrics widely described in the literature. A summary of patient characteristics is described in **Tables S1** and S**2**.

Sample collection

Patient biofluids (blood, plasma and CSF) and/or skin biopsies from patients with SCA3, other forms of ataxia, *C9orf72* carriers, and healthy control individuals were obtained following standardized protocols at Mayo Clinic in Florida and Arizona, University College of London (United Kingdom), Lund University (Sweden), University of Coimbra (Portugal), and the Massachusetts General Hospital and Washington University through their Dominantly Inherited ALS (DIALS) Network study. Blood, plasma and CSF were placed on ice upon collection from the patients and transferred to the laboratory for processing. Within 30 minutes following CSF and plasma collection, samples were spun at 4°C for 15 minutes at either 2,465 x g (plasma) or 453 x g (CSF), and supernatant stored in 200 µl aliquots. Blood was aliquoted in 2 ml aliquots for DNA extraction. All skin biopsies consisted of a 3 mm punch from the non-dominant upper extremity (forearm), and were collected by the same clinician (except 1 case collected at Mayo Clinic Arizona) at the same patient visit in which CSF and blood was collected. Skin biopsies were also transferred immediately to the research lab for immediate generation of fibroblast lines, as described below.

CSF series, main plasma cohort, and fibroblast cell lines

This cohort consisted of 54 SCA3 patients, 21 other ataxia patients, 54 *C9orf72* carriers, 56 healthy controls, and 4 pre-symptomatic SCA3 participants. Information was collected regarding polyQ ATXN3 (plasma, CSF, and fibroblasts), NFL (plasma and CSF), age at onset, sex, race, age at sample collection, disease duration at sample collection, *ATXN3* CAG-repeat length in both alleles, *ATXN3* rs7158733 long/short allele genotype, gait mobility score, and total SARA score. Note that CSF from 11 patients from the UCL validation cohort was also included in the CSF series since no CSF control samples were available from the independent sample collection.

Plasma validation cohort

The plasma validation cohort was comprised of 40 SCA3 patients, 2 pre-symptomatic SCA3 participants, and 30 healthy controls collected at University College of London (UCL) at the United Kingdom. Information was collected regarding age at onset, sex, polyQ ATXN3, NFL, age at plasma collection, disease duration at plasma collection, *ATXN3* CAG-repeat length, total SARA score, total INAS score, total ADL score, SCAFI, and CCFS.

Establishment of fibroblasts cell lines and culturing

Human primary dermal fibroblasts were cultured from skin biopsies in standard culture media comprised of Dulbecco's modified Eagle's medium (DMEM, Fisher Scientific) with 10% fetal bovine serum (Sigma), 1% non-essential amino acids (Fisher), 1% penicillin-streptomycin (Fisher) and 1% amphotericin B (Gemini Bio). Skin biopsies were dissected into 4 pieces and cultured in 6 well plates until outgrowth covered approximately 20% of each well, then the biopsy tissue was removed and fibroblasts were passaged and combined into one T25 flask. At confluency, cells were passaged 1:3 to larger flasks, to obtain three T225s flasks (~$1x10^6$ cells), at which time the cells were harvested and either frozen in FBS with 10% DMSO, or washed and frozen as a cell pellet. Cultures were screened for mycoplasma contamination using the MycoAlert PLUS Mycoplasma Detection Kit (Lonza) when confluent in the T25 flask, and again immediately before freezing.

ATXN3 knockdown in human fibroblasts

From three T225s flasks per line, cells were washed with 1x PBS, combined and counted. The same number of cells was seeded into 10 cm^2 dishes ($3x10^6$ cells/dish). A total of 6 dishes per line were established, with exception of fibroblasts line #9 for which only 4 dishes were generated. Twenty-four hours following seeding into 10cm^2 dishes, cells were treated with 20nM control (siGENOME Non-Targeting siRNA Pool #1, Cat#D-001206-13-20, Dharmacon) or *ATXN3* (SMARTpool: ON-TARGETplus ATXN3 siRNA, Cat#L-012013-00-0005, Dharmacon) siRNA, in triplicates (except line #9 which only allowed duplicates), for a total of 5 days. Cells were then harvested for polyQ ATXN3 analyses using our immunoassay, as described above.

Generation of fibroblast lysates

Fibroblast cell pellets were lysed in Co-IP buffer (50 mm Tris–HCl, pH 7.4, 300 mM NaCl, 1% Triton-X-100, 5 mM EDTA) with 2% sodium dodecyl sulfate (SDS), phenylmethylsulfonyl fluoride (PMSF), and both a protease and phosphatase inhibitor mixture (1:100), and sonicated. After centrifugation at 16,000-20,000 xg for 20 min at 4-15°C, the supernatant was collected and protein concentration determined by BCA assay. Samples were diluted in 1x TBS prior to loading onto the MSD plate.

PolyQ ATXN3 protein immunoassay development

To detect and quantify ATXN3 proteins containing the CAG-repeat expansion characteristic of patients with SCA3, we established a sandwich immunoassay using the Meso Scale Discovery (MSD) electrochemiluminescence detection technology.

To determine the best antibody pairs and conditions that delivered the best signal-to-noise ratio, we first coated a MSD Multi-array 96-well plate (Cat#L15XA-6, MSD) with different anti-ATXN3 or anti-polyQ antibodies, at different concentrations (2, 1, 0.5, and 0.25 μg/ml diluted in TBS: Tris buffered saline) overnight at 4°C. Then, wells were washed with TBS-T (0.2% Tween 20 in TBS) and blocked with 3% milk in TBS-T at room temperature for 2 hours. A total of 1 μg of protein from untransfected HEK293T cells or HEK293T cells overexpressing the human ATXN3 coding sequence with an 84 poly-glutamine expansion were added to the wells as negative and positive controls, respectively. Lysates were incubated on the plate for 2 hours at room temperature at 600 rpm, and washed 3 times with TBS-T. For detection, the SULFO-TAGGED anti-polyQ antibody (clone 3B5H10, Cat#P1874 from Sigma) was added at diverse testing concentrations (2, 1, 0.5, and 0.25 μg/ml), and the plate was incubated another 2 hours at room temperature and 600 rpm, after which time the plate was washed again with TBS-T, and read with 1X MSD Read Buffer T with Surfactant (Cat#R92TC-2, MSD) on the MSD Meso QuikPlex SQ 120 Plate Reader. The best 2 antibody pair combinations were then repeated using lysates of HEK293T cells overexpressing the human ATXN3 coding sequence with a polyQ expansion of either 28 or 84 glutamines. Cell lysates were used at 4x dilutions starting at 1 μg/well. Furthermore, several blocking reagents were also tested: 3% milk, 5% MSD Blocker A (Cat#R93AA-1, MSD), or 0.25% casein. The assay that gave the highest signal to noise ratio, and which was used in subsequent assays, used: 1) a mono-clonal anti-ATXN3 antibody (clone 1H9, Cat# MAB5360 from Millipore) at 0.25 μg/ml as capture antibody, 2) an anti-polyQ detection antibody (clone 3B5H10, Cat#P1874, Sigma) at 0.5 μg/ml and conjugated to a SULFO-TAG NHS-Ester group according to the manufacturer's recommendation (Cat# R91AO-1, MSD), and 3) 5% MSD Blocker A. To increase sensitivity of the assay when testing biofluids, the concentration of the capture antibody was increased to 2 μg/ml and the concentration of the detection antibody was increased to 1.25 μg/ml. Each new lot of capture and labeled detection antibodies was re-optimized to give the best signal to noise ratio.

Generation of recombinant ATXN3

To further optimize immunoassay conditions for the detection of polyQ ATXN3, as well as to later serve as standards to quantify ATXN3 in biological fluids, we generated recombinant ATXN3 (rATXN3) with 80 glutamine repeats (Q80), and purified from Inclusion Body (IB) as a fusion protein with a C-terminal His tag. In brief, to generate rATXN3-Q80, the human polyQ ATXN3 gene (Addgene) was PCR amplified, sequenced, repeat length verified, and cloned into the pET30a bacterial expression vector (Novagen) containing a C-terminal His tag. ATXN3-Q80-His containing plasmids were overexpressed and grown in Rosetta 2 (DE3)pLysS cells (EMD Millipore) at 37°C and 220rpm until OD600 reached 0.5. Then, IPTG was added at final concentration of 0.5 mM and cells were left to grow for 4 hours.

Cell pellets were resuspended in lysis buffer (10 mM HEPES, 0.5 M NaCl, 1 mM $MgCl_2$, 1 mM TCEP, 1 mM PMSF, 1X PIC, 21 U/mL DNase, 2% Triton X-100), and sonicated. Inclusion bodies were pelleted by centrifugation at 16,000 x g for 10 minutes, and resuspended in denaturing buffer (10 mM HEPES, 300 mM NaCl, 1 mM TCEP, 20 mM Imidazole, 8 M urea)

and filtered through a 0.45 µm membrane. Then, rATXN3 proteins were bound to a pre-equilibrated 5 ml His-trap HP column (GE Healthcare). The column was washed with x10 volume of washing buffer (denaturing buffer with 40 mM Imidazole), and the protein was eluted with elution buffer (denaturing buffer with 250 mM Imidazole). rATXN3 containing fractions were pooled, flash frozen and stored at -80°C. Before use, the frozen rATXN3 samples were thawed on ice and desalted with PBS using a 0.5 ml Zeba Spin Desalting column (Thermo Scientific) according to the manufacturer's instructions. Protein concentration was calculated using the bicinchoninic acid assay (BCA). To confirm the purity and specificity of the fractions, we performed coomassie brilliant blue stain, and western blotting against ATXN3 (Cat# MAB5360, clone 1H9, Millipore Sigma) and the His tag (Cat#2366, clone 27E8, Cell Signaling) (**Fig. S1**). Finally, we used the recombinant polyQ ATXN3 protein to evaluate the limit of blanks (0.0088 pg/ul), limit of detection (0.0134 pg/ul) and limits of quantitation (upper limit: >200 pg/ul, lower limit: 0.05 pg/ul) of our immunoassay, as previously described (*33*).

PolyQ ATXN3 immunoassay

To detect and quantify ATXN3 proteins containing the CAG-repeat expansion characteristic of patients with SCA3, we established a sandwich immunoassay using the Meso Scale Discovery (MSD) electrochemiluminescence detection technology. Additional details on assay development can be found in a previous section entitled: PolyQ ATXN3 protein immunoassay analyses development.

To quantify polyQ ATXN3 proteins in human samples, we first coated a MSD Multi-array 96-well plate (Cat#L15XA-6, MSD) with a mono-clonal anti-ATXN3 antibody (clone 1H9, Cat# MAB5360 from Millipore) at 2 µg/ml as capture antibody (diluted in TBS: Tris buffered saline) overnight at 4°C. Then, wells were washed with TBS-T (0.2% Tween 20 in TBS) and blocked with 3% MSD Blocker A (Cat#R93AA-1, MSD) in TBS-T at room temperature for 2 hours. A total 90 µl of CSF, 10 µl of plasma (total 50 µl/well, 1:5 dilution in TBS-T) or 20 µg of protein from human fibroblast lysates were added to each well (each sample run in duplicates) and incubated on the plate for 2 hours at room temperature at 600 rpm, and washed 3 times with TBS-T. For detection, an anti-polyQ detection antibody (clone 3B5H10, Cat#P1874, Sigma) at 1.25 µg/ml and conjugated to a SULFO-TAG NHS-Ester group was used according to the manufacturer's recommendation (Cat# R91AO-1, MSD). The plate was incubated 2 hours at room temperature and 600 rpm, after which time the plate was washed again with TBS-T, and read with 1X MSD Read Buffer T with Surfactant (Cat#R92TC-2, MSD) on the MSD Meso QuikPlex SQ 120 Plate Reader.

NFL immunoassay

NFL measurements were acquired using the Simoa HD-1 analyzer and the commercial NF-light kit from Quanterix as per the manufacturer's protocol. CSF and plasma samples were run in Simoa 96-well plates by trained personnel, with kits of the same lot. The Simoa platform performs 2 repeated measures and provides the mean of both readings as well as the coefficient of variation (CV).

Assessment of repeat length from blood genomic DNA

For analysis of the CAG repeat length in patient samples and controls, PCR was performed on genomic DNA extracted from blood, as described (Cat#158389, Qiagen). Sequencing primers were as previously reported (*67*): a 6-FAM fluorophore-labeled 5' primer (5'

CCAGTGACTACTTTGATTCG 3') and an unlabeled 3' primer (5' CTTACCTAGATCACTCCCAA 3'). A touchdown PCR program (3 min 95°C; 10 cycles of 15s 94°C, 15s 65°C decreasing 1°C per cycle, 30s 72°C; 30 cycles 15s 94°C, 15s 55°C, 30s 72°C; and 5 min 72°C) was used with an Apex Taq DNA Polymerase (Genesee). Products were mixed with the GeneScan 500 LIZ dye Size Standard (Fisher), analyzed for size determination on an ABI3730 sequencer, and repeat sizes were determined with GeneMapper™ 4.0 software (Applied Biosystems).

Determination of rs7158733 SNP and location on non-expanded/expanded allele

For analysis of the rs7158733 SNP-specific CAG repeat length in patient samples and controls, PCR was performed on gDNA using a 6-FAM fluorophore-labeled 5' primer (5' CCAGTGACTACTTTGATTCG 3'), and unlabeled 3' primers specific for either the "TAC^{1118}" allele (5' AAAAATCACATGGAGCTCG 3') or the "TAA^{1118}" allele (5' AAAAATCACATGGAGCTCT 3'), and a touchdown PCR program (3 min 95°C; 10 cycles of 15s 94°C, 15s 70°C decreasing 1°C per cycle, 30s 72°C; 30 cycles 15s 94°C, 15s 60°C, 30s 72°C; and 5 min 72°C) using a Apex Taq DNA Polymerase (Genesee). Products were mixed with the GeneScan 500 LIZ dye Size Standard (Fisher), analyzed for size determination on an ABI3730 sequencer, and repeat sizes were determined with GeneMapper™ 4.0 software (Applied Biosystems).

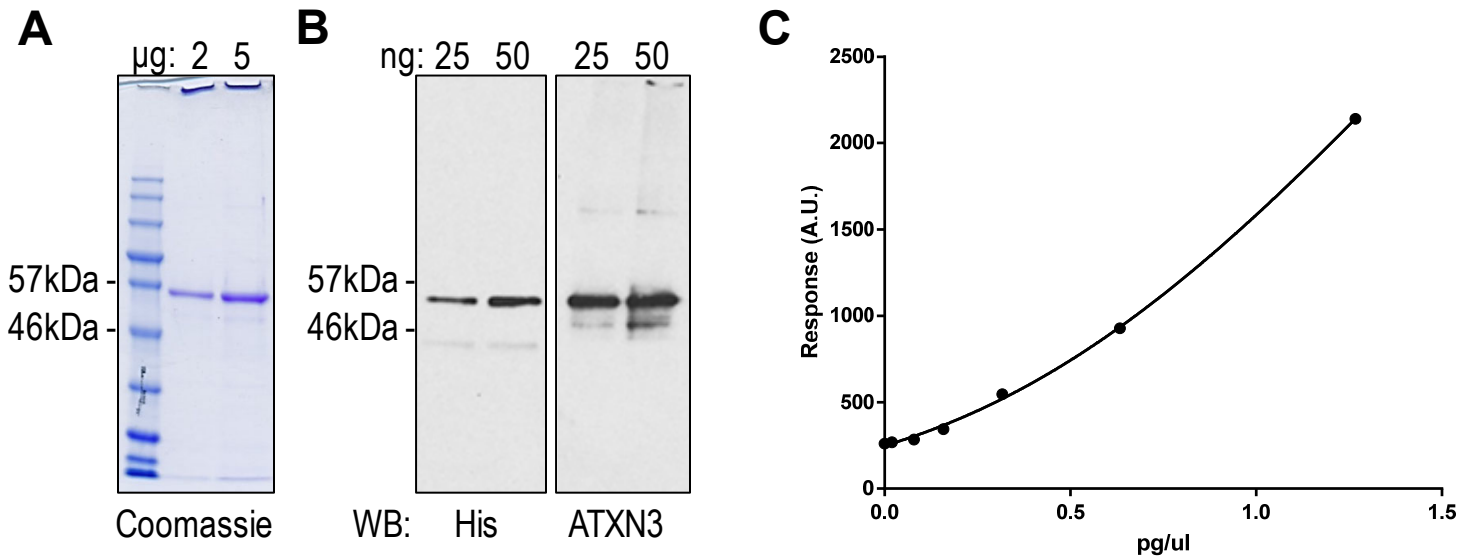


Fig. S1. Purified rATXN3 protein containing 80 glutamine repeats (Q80). A) Coomassie brilliant blue staining on gels containing 2 and 5 µg of rATXN3-Q80-His protein. **B**) Western blots (WB) of 25 and 50 ng total protein shows the specificity of rATXN3-Q80-His fractions by probing with antibodies against the His tag (left) and ATXN3 (right). **C**) Standard curve using different concentrations of rATXN3 protein.

Table S1. Subject characteristics for the CSF series and the main plasma cohort.

	SCA3 patients (N=54)		Other ataxia patients (N=21)		*C9orf72* expansion carriers (N=54)		Healthy controls (N=56)		Pre-symptomatic SCA3 (N=4)	
Variable	**N**	**Summary[1]**	**N**	**Summary[1]**	**N**	**Summary[1]**	**N**	**Summary[1]**	**N**	**Summary[1]**
Age at ataxia onset (years)	54	39 (18, 65)	21	36 (3, 72)	0	N/A	0	N/A	0	N/A
Sex (Male)	54	19 (35.2%)	21	9 (42.9%)	54	26 (48.1%)	56	23 (41.1%)	4	1 (25.0%)
Race (White)	43	33 (76.6%)	20	15 (75.0%)	52	51 (98.1%)	50	41 (82.0%)	4	4 (100.0%)
Age at plasma collection (years)	41	53 (27, 75)	16	54 (20, 80)	43	57 (24, 73)	34	55 (26, 79)	4	32 (31, 67)
Age at CSF collection (years)	46	51 (24, 75)	12	54 (20, 78)	40	58 (23, 77)	30	45 (27, 73)	4	32 (31, 67)
Disease duration of ataxia at plasma collection (years)	41	11 (4, 38)	16	14 (1, 37)	0	N/A	0	N/A	0	N/A
Disease duration of ataxia at CSF collection (years)	46	11 (1, 38)	12	9 (2, 40)	0	N/A	0	N/A	0	N/A
ATXN3 CAG-repeat length	53	69 (51, 75)	20	25 (14, 37)	54	23 (14, 33)	44	23 (14, 33)	4	65 (59, 69)
Gait mobility score	32		20		0		30		2	
0=Asymptomatic		0 (0.0%)		0 (0.0%)		N/A		30 (100.0%)		2 (100.0%)
1=Impaired gait (no assistance)		10 (31.2%)		9 (45.0%)		N/A		0 (0.0%)		0 (0.0%)
2=Impaired gait (require cane)		3 (9.4%)		2 (10.0%)		N/A		0 (0.0%)		0 (0.0%)
3=Requires a walker		4 (12.5%)		1 (5.0%)		N/A		0 (0.0%)		0 (0.0%)
4=Wheelchair bound		14 (43.8%)		8 (40.0%)		N/A		0 (0.0%)		0 (0.0%)
5=Bedridden		1 (3.1%)		0 (0.0%)		N/A		0 (0.0%)		0 (0.0%)
SARA total	52	12 (1, 34)	20	12 (5, 39)	0	N/A	30	0 (0, 3)	4	0 (0, 0)

[1]The sample median (minimum, maximum) is given for continuous variables, and No. (%) of subjects is given for categorical variables.

Table S2. Subject characteristics for the plasma validation cohort.

	SCA3 patients (N=40)		Healthy controls (N=30)		Pre-symptomatic SCA3 (N=2)	
Variable	**N**	**Summary[1]**	**N**	**Summary[1]**	**N**	**Summary[1]**
Age at ataxia onset (years)	39	39 (16, 69)	0	N/A	0	N/A
Sex (Male)	40	13 (32.5%)	30	12 (40.0%)	2	0 (0%)
Age at plasma collection (years)	40	53 (25, 77)	30	43 (22, 72)	2	38 (34, 41)
Disease duration of ataxia at plasma collection (years)	39	11 (1, 41)	0	N/A	0	N/A
ATXN3 CAG-repeat length	39	70 (56, 75)	25	24 (14, 30)	2	67 (62, 71)
SARA total	40	17 (2, 35)	0	N/A	2	2 (2, 2)
INAS total	40	7 (2, 9)	0	N/A	2	2 (1, 3)
ADL total	40	17 (3, 34)	0	N/A	2	2 (0, 4)
SCAFI	34	-0.6 (-3.1, 1.8)	0	N/A	2	1.0 (1.0, 1.0)
CCFS	32	1.1 (-0.9, 1.5)	0	N/A	2	1.0 (0.9, 1.0)

[1]The sample median (minimum, maximum) is given for continuous variables, and No. (%) of subjects is given for categorical variables.

Table S3. Comparisons of polyQ ATXN3 between SCA3 patients and healthy controls, carriers of the *C9orf72* repeat expansion, other ataxia patients, and pre-symptomatic SCA3 individuals.

			P-value in comparison to SCA3 patients		
Outcome/group	**N**	**Median (pg/µl) (minimum, maximum)**	**Unadjusted analysis**	**Adjusting for age and sex**	**AUC (95% CI) vs. SCA3 patients)**
CSF polyQ ATXN3					
SCA3 patients	45	0.13 (0.04, 0.46)	N/A	N/A	N/A
Healthy controls, *C9orf72* carriers, and other ataxia patients	84	0.00 (0.00, 0.04)	<0.001	<0.001	1.00 (1.00, 1.00)
Healthy controls	33	0.00 (0.00, 0.04)	<0.001	<0.001	1.00 (1.00, 1.00)
C9orf72 carriers	40	0.00 (0.00, 0.04)	<0.001	<0.001	1.00 (1.00, 1.00)
Other ataxia patients	11	0.00 (0.00, 0.02)	<0.001	<0.001	1.00 (1.00, 1.00)
Pre-symptomatic SCA3	4	0.07 (0.04, 0.07)	0.010	0.035	0.89 (0.80, 0.99)
Plasma polyQ ATXN3 – Main cohort					
SCA3 patients	41	1.28 (0.30, 3.44)	N/A	N/A	N/A
Healthy controls, *C9orf72* carriers, and other ataxia patients	93	0.00 (0.00, 0.28)	<0.001	<0.001	1.00 (1.00, 1.00)
Healthy controls	34	0.00 (0.00, 0.23)	<0.001	<0.001	1.00 (1.00, 1.00)
C9orf72 carriers	43	0.00 (0.00, 0.20)	<0.001	<0.001	1.00 (1.00, 1.00)
Other ataxia patients	16	0.00 (0.00, 0.28)	<0.001	<0.001	1.00 (1.00, 1.00)
Pre-symptomatic SCA3	4	0.84 (0.48, 1.68)	0.22	0.21	0.70 (0.38, 1.00)
Plasma polyQ ATXN3 – Validation cohort					
SCA3 patients	40	1.08 (0.36, 2.68)	N/A	N/A	N/A
Healthy controls	30	0.00 (0.00, 0.00)	<0.001	<0.001	1.00 (1.00, 1.00)
Pre-symptomatic SCA3	2	1.21 (0.81, 1.61)	N/A	N/A	N/A

AUC = area under the ROC curve. In unadjusted analysis P-values result from a Wilcoxon rank sum test. In analysis adjusted for age and sex, P-values result from a van Elteren stratified Wilcoxon rank sum test, where the tests were stratified by a 4-category variable based on combination of age (≤ median in the given group [main or validation cohort], >median in the given group [main or validation cohort]) and sex. After applying a Bonferroni correction for multiple testing in the main cohort, P-values <0.010 were considered as statistically significant. P-values <0.05 were considered as significant in analysis of the plasma validation cohort. A statistical test was not performed for the pre-symptomatic SCA3 patients in the validation cohort since only 2 patients had a measurement.

Table S4. Comparisons of NFL between SCA3 patients and healthy controls, other ataxia patients, and pre-symptomatic SCA3 individuals.

			P-value in comparison to SCA3 patients		
Outcome/group	**N**	**Median (pg/ml) (minimum, maximum)**	**Unadjusted analysis**	**Adjusting for age and sex**	**AUC (95% CI) vs. SCA3 patients)**
CSF NFL					
SCA3 patients	46	3569 (1413, 6837)	N/A	N/A	N/A
Healthy controls	34	449 (137, 1512)	<0.001	<0.001	1.00 (1.00, 1.00)
Other ataxia patients	12	1096 (327, 5672)	<0.001	<0.001	0.82 (0.65, 0.99)
Pre-symptomatic SCA3	4	1352 (1019, 1398)	<0.001	0.001	1.00 (1.00, 1.00)
Plasma NFL – Main cohort					
SCA3 patients	41	30.31 (17.24, 69.90)	N/A	N/A	N/A
Healthy controls	34	11.20 (3.31, 33.56)	<0.001	<0.001	0.89 (0.81, 0.96)
Other ataxia patients	16	13.14 (3.90, 28.04)	<0.001	<0.001	0.94 (0.87, 1.00)
Pre-symptomatic SCA3	4	15.42 (11.05, 28.68)	0.011	0.022	0.87 (0.67, 1.00)
Plasma NFL – Validation cohort					
SCA3 patients	40	32.31 (15.50, 348.08)	N/A	N/A	N/A
Healthy controls	30	7.98 (3.47, 32.35)	<0.001	<0.001	0.97 (0.93, 1.00)
Pre-symptomatic SCA3	2	24.04 (13.85, 34.23)	N/A	N/A	N/A

AUC = area under the ROC curve. In unadjusted analysis p-values result from a Wilcoxon rank sum test. In analysis adjusted for age and sex, P-values result from a van Elteren stratified Wilcoxon rank sum test, where the tests were stratified by a 4-category variable based on combination of age (≤ median in the given group [main or validation cohort], >median in the given group [main or validation cohort]) and sex. After applying a Bonferroni correction for multiple testing in the main cohort, P-values <0.0167 were considered as statistically significant. P-values <0.05 were considered as significant in analysis of the plasma validation cohort. A statistical test was not performed for the pre-symptomatic SCA3 patients in the validation cohort since only 2 patients had a measurement.

Table S5. Associations of NFL with clinical features of SCA3.

		Unadjusted analysis		Adjusting only for cohort		Multivariable analysis	
Clinical feature	**N**	**Spearman's *r* (95% CI)**	**P-value**	**Regression coefficient (95% CI)**	**P-value**	**Regression coefficient (95% CI)**	**P-value**
Association with CSF NFL (combined cohort)							
Age of onset (10 year increase)	46	0.16 (-0.14, 0.44)	0.30	N/A	N/A	293.8 (-181.6, 769.1)	0.22
Disease duration (5 year increase)	46	0.05 (-0.24, 0.34)	0.73	N/A	N/A	-4.3 (-278.0, 271.5)	0.98
ATXN3 CAG-repeat length (5 unit increase)	46	0.03 (-0.25, 0.32)	0.83	N/A	N/A	375.4 (-285.5, 1036.3)	0.26
SARA total (5 unit increase)	45	0.01 (-0.28, 0.30)	0.95	N/A	N/A	-9.5 (-300.1, 281.1)	0.95
Gait mobility score (1 unit increase)	25	0.03 (-0.46, 0.48)	0.89	N/A	N/A	-51.9 (-459.2, 355.4)	0.79
Association with plasma NFL (combined cohort)							
Age of onset (10 year increase)	80	-0.03 (-0.26, 0.18)	0.77	0.16 (-7.19, 7.52)	0.96	1.19 (-8.79, 11.17)	0.81
Disease duration (5 year increase)	80	0.14 (-0.09, 0.34)	0.20	3.61 (-2.21, 9.42)	0.22	3.47 (-2.69, 9.62)	0.27
ATXN3 CAG-repeat length (5 unit increase)	80	0.00 (-0.24, 0.23)	0.99	0.94 (-8.90, 10.77)	0.85	5.70 (-8.14, 19.53)	0.41
SARA total (5 unit increase)	80	0.19 (-0.01, 0.39)	0.099	5.54 (0.09, 11.00)	0.047	5.71 (-1.37, 12.80)	0.11
Gait mobility score (1 unit increase)	31	-0.02 (-0.38, 0.32)	0.93	N/A	N/A	0.77 (-4.06, 5.61)	0.75
INAS total (1 unit increase)	40	0.06 (-0.27, 0.39)	0.72	N/A	N/A	0.05 (-0.04, 0.15)	0.27
ADL total (5 unit increase)	40	0.11 (-0.19, 0.40)	0.50	N/A	N/A	0.08 (-0.11, 0.26)	0.42
SCAFI (1 unit increase)	34	-0.15 (-0.48, 0.21)	0.41	N/A	N/A	-0.15 (-0.42, 0.12)	0.26
CCFS (0.5 unit increase)	32	0.05 (-0.33, 0.42)	0.79	N/A	N/A	-0.17 (-1.10, 0.75)	0.71

CI = confidence interval. In unadjusted analysis, p-values result from Spearman's test of correlation. In multivariable analysis, regression coefficients, 95% CIs, and P-values result from linear regression models. Regression coefficients are interpreted as the change in mean NFL corresponding to the increase given in parenthesis (continuous variables). Multivariable models were adjusted for age at plasma/CSF collection, sex, and disease duration with the exception of the model for age at onset which was adjusted for sex and *ATXN3* CAG-repeat length. For clinical features that were assessed in only the main cohort or the validation cohort, but not both, analysis adjusting only for cohort was not performed. After applying a Bonferroni adjustment for multiple testing separately for each NFL outcome measure, P-values <0.010 (associations with CSF NFL) and <0.0056 (associations with plasma NFL) were considered as statistically significant.

Table S6. Association of presence of the "TAA^{1118}" allele for *ATXN3* rs7158733 with SCA3 or at risk of SCA3 (pre-symptomatic SCA3) in comparison to healthy controls, *C9orf72* expansion carriers, and other ataxia patients).

ATXN3 rs7158733 measure	Fraction (%) of carriers of the "TA$\underline{A}^{1118}$" allele	Association of the "TA$\underline{A}^{1118}$" allele with risk of SCA3	
		OR[1] (95% CI)	**P-value**
Presence of the "TA$\underline{A}^{1118}$" allele on either the long or short allele			
SCA3 and pre-symptomatic SCA3	40/46 (87.0%)	1.00 (reference)	N/A
Controls	21/44 (47.7%)	0.13 (0.04, 0.36)	<0.001
C9orf72 expansion carriers	34/54 (63.0%)	0.23 (0.07, 0.62)	0.006
Other ataxia patients	12/20 (60.0%)	0.19 (0.05, 0.68)	0.013
Controls, *C9orf72* carriers, and other ataxia patients	67/118 (56.8%)	0.18 (0.06, 0.43)	<0.001
Presence of the "TA$\underline{A}^{1118}$" allele on the longer allele			
SCA3 and pre-symptomatic SCA3	36/46 (78.3%)	1.00 (reference)	N/A
Controls	11/44 (25.0%)	0.09 (0.03, 0.23)	<0.001
C9orf72 expansion carriers[2]	19/53 (35.8%)	0.14 (0.05, 0.34)	<0.001
Other ataxia patients	10/20 (50.0%)	0.27 (0.08, 0.83)	0.023
Controls, *C9orf72* carriers, and other ataxia patients	40/117 (34.2%)	0.13 (0.06, 0.29)	<0.001

OR = odds ratio; CI = confidence interval. [1]ORs, 95% CIs, and P-values result from logistic regression models that were adjusted for age and sex. In these logistic regression models, disease status was the dependent variable and *ATXN3* rs7158733 was the independent variable; correspondingly, a separate logistic regression model was examined for each pair-wise comparison vs. SCA3 patients. ORs less than one (with SCA3 as the reference group) indicate a negative correlation between the "TA$\underline{A}^{1118}$" allele and the risk of SCA3. [2]The long and short allele were the same length for one of the *C9orf72* carriers; this *C9orf72* carrier was excluded from analysis involving presence of the "TA$\underline{A}^{1118}$" allele on the long allele as there was no long allele for this patient. After applying a Bonferroni adjustment for multiple testing separately for each *ATXN3* rs7158733 measure, P-values <0.0125 were considered as statistically significant.

Table S7. Comparisons of polyQ ATXN3 in fibroblasts between SCA3 patients and healthy controls, other ataxia patients, and pre-symptomatic SCA3 individuals.

Outcome/group	N	Median (pg/µl) (minimum, maximum)	P-value in comparison to SCA3 patients		
			Unadjusted analysis	Adjusting for age and sex	AUC (95% CI) vs. SCA3 patients)
Fibroblasts polyQ ATXN3 – Main cohort					
SCA3 patients	17	4.56 (2.05, 6.22)	N/A	N/A	N/A
Healthy controls and other ataxia patients	27	0.37 (0.00, 1.38)	<0.001	<0.001	1.00 (1.00, 1.00)
Healthy controls	13	0.38 (0.00, 0.77)	<0.001	<0.001	1.00 (1.00, 1.00)
Other ataxia patients	14	0.36 (0.15, 1.38)	<0.001	<0.001	1.00 (1.00, 1.00)
Pre-symptomatic SCA3	2	3.62 (2.43, 4.82)	N/A	N/A	N/A

AUC = area under the ROC curve. In unadjusted analysis P-values result from a Wilcoxon rank sum test. In analysis adjusted for age and sex, P-values result from a van Elteren stratified Wilcoxon rank sum test, where the tests were stratified by a 4-category variable based on combination of age (≤ median) and sex. After applying a Bonferroni correction for multiple testing, P-values <0.0167 were considered as statistically significant. A statistical test was not performed for fibroblasts polyQ ATXN3 in the pre-symptomatic SCA3 patients since only 2 patients had a measurement.

Table S8. Data for each individual data point presented in Figure 4A and B.

Sample ID	Study group	Fibroblasts_polyQ_ATXN3_(pg/ul)	CSF_polyQ_ATXN3_(pg/ul)
1	Healthy control	0.42	0.00
2	Healthy control	0.17	N/A
10	Healthy control	0.00	N/A
11	Healthy control	0.41	N/A
12	Healthy control	0.15	0.00
13	Healthy control	0.35	N/A
14	Healthy control	0.15	N/A
15	Healthy control	0.43	0.00
16	Healthy control	0.68	0.00
17	Healthy control	0.38	0.01
18	Healthy control	0.33	0.00
19	Healthy control	0.41	N/A
20	Healthy control	0.77	N/A
21	Other ataxia	0.31	N/A
22	Other ataxia	0.37	N/A
23	Other ataxia	0.15	0.00
24	Other ataxia	0.89	N/A
25	Other ataxia	1.38	N/A
26	Other ataxia	0.46	N/A
27	Other ataxia	0.93	N/A
28	Other ataxia	0.18	N/A
29	Other ataxia	0.37	0.00
30	Other ataxia	0.17	0.00
31	Other ataxia	0.34	0.00
32	Other ataxia	0.21	N/A
33	Other ataxia	0.36	0.00
34	Other ataxia	0.54	N/A
35	Pre-symptomatic SCA3	2.43	0.04
36	Pre-symptomatic SCA3	4.82	0.07
7	SCA3	4.48	0.15
8	SCA3	5.64	N/A
9	SCA3	5.10	0.33
37	SCA3	2.99	0.07
38	SCA3	6.01	N/A
39	SCA3	3.71	N/A
40	SCA3	4.57	N/A
41	SCA3	6.00	N/A
42	SCA3	6.22	0.24
43	SCA3	5.98	N/A
44	SCA3	4.39	N/A
45	SCA3	4.06	N/A
46	SCA3	3.06	0.07
47	SCA3	3.83	0.08
48	SCA3	2.05	0.05
49	SCA3	4.56	0.09
50	SCA3	4.73	0.15

N/A: not available

Table S9. Data for each individual data point presented in Figure 4C.

Sample ID	Study group	Control siRNA			*ATXN3* siRNA		
1	Healthy control	0.98	0.56	0.88	0.38	0.35	0.36
2	Healthy control	0.64	0.74	0.69	0.39	0.50	0.42
3	Healthy control	0.92	1.20	1.18	0.63	0.78	0.59
4	Healthy control	1.33	1.30	1.30	0.28	0.23	0.16
5	SCA3	4.41	3.67	3.58	1.03	1.32	0.97
6	SCA3	4.61	3.46	3.57	1.84	1.77	1.53
7	SCA3	3.66	5.44	6.00	1.98	2.57	3.30
8	SCA3	6.22	6.54	6.23	1.80	1.89	2.07
9	SCA3	2.64	3.99	N/A	1.46	1.46	N/A

N/A: not available