

SHORT REPORT

# The (CAG)<sub>n</sub> tract of Machado–Joseph Disease gene (*ATXN3*): a comparison between DNA and mRNA in patients and controls

Conceição Bettencourt<sup>\*,1,2</sup>, Cristina Santos<sup>1,3</sup>, Rafael Montiel<sup>4</sup>, Teresa Kay<sup>5</sup>, João Vasconcelos<sup>6</sup>, Patrícia Maciel<sup>7</sup> and Manuela Lima<sup>1,2</sup>

Machado–Joseph disease (MJD) is an autosomal dominant neurodegenerative disorder of late onset (occurring at a mean age of 40.2 years). The clinical manifestation of MJD is dependent on the presence of an expansion of the (CAG)<sub>n</sub> motif within exon 10 of the *ATXN3* gene, located at 14q32.1. The variance in onset of MJD is only partially correlated (~50–80%) with the extension of the CAG tract in genomic DNA (gDNA). The main aim of this work was to determine whether there are discrepancies in the size of the (CAG)<sub>n</sub> tract between gDNA and mRNA, and to establish whether there is a better association between age at onset and repeat size at the mRNA level. We typed gDNA and cDNA samples for the (CAG)<sub>n</sub> tract totaling 108 wild-type and 52 expanded *ATXN3* alleles. In wild-type alleles no differences were found between gDNA and cDNA. In expanded alleles, the CAG repeat size in gDNA was not always directly transcribed into the mRNA; on average there were differences of +1 repeat at the cDNA level. The slight discrepancies obtained were insufficient to cause significant differences in the distribution of the expanded alleles, and therefore no improvement in onset variance explanation was obtained with mRNA.

*European Journal of Human Genetics* (2010) 18, 621–623; doi:10.1038/ejhg.2009.215; published online 25 November 2009

**Keywords:** MJD; SCA3; CAG repeats; transcript variation; polyglutamine disorders

## INTRODUCTION

Overall, spinocerebellar ataxias (SCAs) are considered to be rare disorders (prevalence ~3:100 000). Machado–Joseph disease (MJD) (MIM 109150), also known as SCA type 3 (SCA3), is one of the most common SCAs worldwide,<sup>1</sup> reaching its highest prevalence value in Azores Islands (Portugal) (~1:3472).<sup>2</sup> MJD is a clinically heterogeneous autosomal dominant neurodegenerative disorder of late onset (occurring at a mean age of 40.2 years), involving the cerebellar, ocular motor, pyramidal, extrapyramidal and peripheral motor systems.<sup>3</sup>

MJD is caused by the expansion of a CAG repeat at exon 10 of the *ATXN3* gene (14q32.1), which encodes for ataxin-3.<sup>4,5</sup> Wild-type alleles present 12–44 CAG units, whereas expanded alleles contain between 53 and 87 CAG repeats.<sup>6,7</sup> Because in genomic DNA (gDNA) the (CAG)<sub>n</sub> number in the expanded alleles is only partially correlated with the age at onset, accounting for 50–80% of its variability,<sup>8,9</sup> precise predictions of the onset based on repeat size are unfeasible.

As in other polyglutamine disorders, both germlinal<sup>8</sup> and somatic<sup>10</sup> instability have been described for expanded MJD alleles at the gDNA level. The existence of such molecular instability, which is related to DNA replication and repair processes, raises the hypothesis that changes in the CAG repeat number could also occur during transcription. Previous studies on yeast models have highlighted the possibility that transcription-related mechanisms could account for

the generation of longer transcripts.<sup>11</sup> However, this hypothesis is still much unexplored for polyglutamine disorders.

The main aim of this work was to determine whether there are discrepancies in the size of the (CAG)<sub>n</sub> tract at the *ATXN3* gene between gDNA and mRNA from MJD patients and controls, and to establish whether there is a better association between age at onset and repeat size at the mRNA level.

## MATERIALS AND METHODS

### Subjects

After obtaining informed consent, blood samples were collected from 52 clinically confirmed MJD patients of Azorean ancestry, for whom age at onset was available (mean age 37.92 years). Twenty-eight apparently healthy individuals, of the same background and with no familial history for MJD, were used as controls.

### gDNA and cDNA samples

DNA and total RNA were extracted from peripheral blood leukocytes, using a salting-out procedure and Trizol reagent (Invitrogen, Ltd, Paisley, UK), respectively. Total RNA (weight 2 µg) was subjected to direct cDNA synthesis using the ThermoScript reverse transcription (RT)-PCR System (Invitrogen) with Oligo(dT)<sub>20</sub>, at 60°C for 50 min.

### Allele size determination

For gDNA, fragments containing the (CAG)<sub>n</sub> tract of the *ATXN3* gene (198 bp+(CAG)<sub>n</sub>) were amplified using previously described conditions.<sup>12</sup>

<sup>1</sup>Center of Research in Natural Resources (CIRN) and Department of Biology, University of the Azores, Ponta Delgada, Portugal; <sup>2</sup>Molecular and Cellular Biology Institute (IBMC), University of Porto, Porto, Portugal; <sup>3</sup>Unitat Antropologia Biològica, Dep. Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona, Bellaterra (Barcelona), Spain; <sup>4</sup>Laboratorio Nacional de Genómica para la Biodiversidad, CINVESTAV – IPN, Irapuato, Mexico; <sup>5</sup>Department of Clinical Genetics, Hospital of D. Estefania, Lisbon, Portugal; <sup>6</sup>Department of Neurology, Hospital of Divino Espirito Santo, Ponta Delgada, Portugal; <sup>7</sup>Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal

\*Correspondence: Dr C Bettencourt, Center of Research in Natural Resources (CIRN) and Department of Biology, University of the Azores, Rua Mãe de Deus – Apartado 1422, Ponta Delgada 9501-801, Azores, Portugal. Tel: +351 296650477; Fax: +351 296650100; E-mail: mcbettencourt@uac.pt

Received 3 July 2009; revised 19 October 2009; accepted 27 October 2009; published online 25 November 2009

For cDNA, the same methodology was applied, after substituting the forward primer by MJD814F (5'-TATTCAGCTAAGTATGCAAG-3'). The amplification products obtained also contained the (CAG)<sub>n</sub> tract plus 198 bp. Allele size determination was made according to Bettencourt *et al.*<sup>12</sup>

### Statistical analysis

Wilcoxon's test was performed using SPSS v.15.0<sup>13</sup> to detect discrepancies between the sizes of the expanded *ATXN3* alleles in gDNA and mRNA. An exact test was conducted in GENEPOP v.1.2 (<http://genepop.curtin.edu.au/>)<sup>14</sup> to test for differences in the expanded allele size distribution between gDNA and mRNA. Correlation analysis was performed to determine the relationship between the CAG repeat number in the expanded alleles and the age at onset, using SPSS v.15.0.<sup>13</sup>

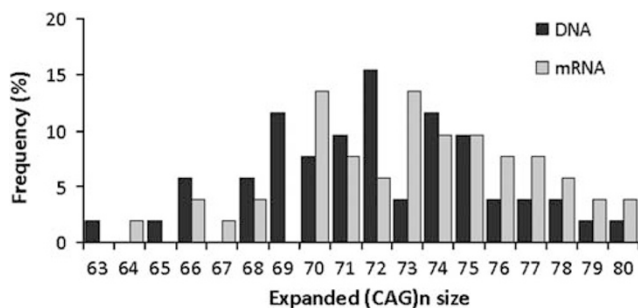
### RESULTS

With regard to wild-type alleles, no discrepancies were observed between the CAG repeat number in gDNA and in mRNA (Table 1). In the expanded alleles (Table 1), however, differences were observed in 48 cases (92.3%), with statistically significant (Wilcoxon's test,  $Z=-6.239$ ,  $P<0.000$ ) increases of +1 (alleles 63–78) or +2 CAG repeats (alleles 66–78) in mRNA (inferred from cDNA). Nevertheless, when comparing the expanded allele size distributions (gDNA *versus* mRNA), no significant differences were found (exact test,  $P=0.286$ ) (Figure 1), indicating that, globally, the allelic distribution is similar in both types of samples.

Negative correlations between (CAG)<sub>n</sub> size and age at onset were observed at both gDNA and mRNA levels ( $r=-0.825$  and  $r=-0.819$ , respectively), with similar percentages of explanation of onset variance (68% gDNA and 67% mRNA).

**Table 1 Summary statistics of *ATXN3* allelic profiles in MJD patients and controls**

| Statistics     | Controls      |               | MJD patients  |               |
|----------------|---------------|---------------|---------------|---------------|
|                | Wild type     | Wild type     | Expanded DNA  | Expanded mRNA |
| No. of alleles | 56            | 52            | 52            | 52            |
| Mean size (SE) | 21.79 (±0.59) | 21.46 (±0.61) | 71.94 (±0.51) | 73.25 (±0.52) |
| Mode           | 23            | 23            | 72            | 70, 73        |
| Skewness (SE)  | -0.63 (±0.32) | -0.66 (±0.33) | -0.06 (±0.33) | -0.29 (±0.33) |
| Minimum        | 14            | 14            | 63            | 64            |
| Maximum        | 30            | 29            | 80            | 80            |



**Figure 1** Distribution of the expanded *ATXN3* alleles at DNA and mRNA levels in samples of MJD patients ( $N=52$ ).

### DISCUSSION

The main aim of this work was to determine whether there are discrepancies in the size of the (CAG)<sub>n</sub> tract at the *ATXN3* gene between gDNA and mRNA in expanded and wild-type alleles. On analysing cDNA, we consistently detected differences from +1 to +2 repeats between gDNA and mRNA in expanded alleles. The possibility that this could result from a methodological artefact produced during *in vitro* RT could be raised. However, and to minimize the occurrence of *in vitro* errors, an enzyme with high thermal stability was used, enabling the increase of RT reaction temperature (in this case to 60°C), which reduces the probability of RNA secondary structure formation and improves priming specificity. On the other hand, the absence of discrepancies for wild-type alleles, and the fact that there was no relationship between increases of +1 or +2 repeats and the size of the expanded allele in gDNA, does not support a systematic error of the technique. Therefore, it is most likely that the size differences between gDNA and mRNA observed for the expanded alleles are in fact produced *in vivo* during transcription. It is known that triplet repeat sequences are able to form alternative structures, such as slipped strand structures.<sup>15</sup> The presence of those structures along the gDNA template may cause transient pausing and backward slippage of the RNA polymerase complex. As it was proposed previously, from the study of yeast models with long CAG or CTG tracts,<sup>11</sup> this could result in the resynthesis of the same RNA sequence, leading to the formation of longer transcripts. The present results support the hypothesis that transcriptional slippage occurs at the MJD locus in a size-dependent way, as it was observed only in expanded *ATXN3* alleles.

To our knowledge, comparisons at the gDNA and mRNA levels with regard to the size of the expanded *ATXN3* alleles were only made in a previous study,<sup>16</sup> which analysed neuronal and non-neuronal post-mortem tissues of a single MJD patient. The referred study additionally included the analysis of the corresponding loci in one patient with dentatorubral-pallidoluyian atrophy (DRPLA) and another with spinal and bulbar muscular atrophy (SBMA). Their results indicated that the tissue-specific size variation of the expanded CAG repeats present in the gDNA was directly transcribed into the mRNAs of the three analysed loci (MJD, DRPLA and SBMA). In the present study, the number of MJD patients was largely increased; results showed that the CAG repeats in gDNA were not always directly transcribed into the mRNAs. The slight discrepancies observed, however, were insufficient to cause significant differences in the expanded alleles' distribution, and no improvement in onset variance explanation was obtained at the mRNA level. Other factors (genetic, environmental or both) should be influencing the onset variance and should be further analyzed.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

This work was supported by projects 'PRI-DMJ' (funded by the Regional Government of the Azores) and 'Transcriptional variation of the *ATXN3* gene as modulator of the clinical heterogeneity in Machado-Joseph disease (MJD)' (PIC/IC/83074/2007, funded by 'Fundação para a Ciência e a Tecnologia' - FCT). CB (SFRH/BD/21875/2005) is the recipient of a PhD grant from FCT.

1 Schols L, Bauer P, Schmidt T, Schulte T, Riess O: Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. *Lancet Neurol* 2004; **3**: 291-304.

2 Bettencourt C, Santos C, Kay T, Vasconcelos J, Lima M: Analysis of segregation patterns in Machado-Joseph disease pedigrees. *J Hum Genet* 2008; **53**: 920-923.

- 3 Coutinho P, Andrade C: Autosomal dominant system degeneration in Portuguese families of the Azores Islands. A new genetic disorder involving cerebellar, pyramidal, extrapyramidal and spinal cord motor functions. *Neurology* 1978; **28**: 703–709.
- 4 Kawaguchi Y, Okamoto T, Taniwaki M *et al*: CAG expansions in a novel gene for Machado-Joseph disease at chromosome 14q32.1. *Nat Genet* 1994; **8**: 221–228.
- 5 Ichikawa Y, Goto J, Hattori M *et al*: The genomic structure and expression of MJD, the Machado-Joseph disease gene. *J Hum Genet* 2001; **46**: 413–422.
- 6 Maciel P, Costa MC, Ferro A *et al*: Improvement in the molecular diagnosis of Machado-Joseph disease. *Arch Neurol* 2001; **58**: 1821–1827.
- 7 van Alfen N, Sinke RJ, Zwarts MJ *et al*: Intermediate CAG repeat lengths (53,54) for MJD/SCA3 are associated with an abnormal phenotype. *Ann Neurol* 2001; **49**: 805–807.
- 8 Maciel P, Gaspar C, DeStefano AL *et al*: Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am J Hum Genet* 1995; **57**: 54–61.
- 9 Maruyama H, Nakamura S, Matsuyama Z *et al*: Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Hum Mol Genet* 1995; **4**: 807–812.
- 10 Lopes-Cendes I, Maciel P, Kish S *et al*: Somatic mosaicism in the central nervous system in spinocerebellar ataxia type 1 and Machado-Joseph disease. *Ann Neurol* 1996; **40**: 199–206.
- 11 Fabre E, Dujon B, Richard G-F: Transcript and nuclear transport of CAG/CTG trinucleotide repeats in yeast. *Nucleic Acids Res* 2002; **30**: 3540–3547.
- 12 Bettencourt C, Fialho RN, Santos C *et al*: Segregation distortion of wild-type alleles at the Machado-Joseph disease locus: a study in normal families from the Azores islands (Portugal). *J Hum Genet* 2008; **53**: 333–339.
- 13 SPSS Inc.: *SPSS for Windows: Release 15*. Chicago: SPSS Inc., 2006.
- 14 Raymond M, Rousset F: GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *J Heredity* 1995; **86**: 248–249.
- 15 Sinden RR, Potaman VN, Oussatcheva EA, Pearson CE, Lyubchenko YL, Shlyakhtenko LS: Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. *J Biosci* 2002; **27** (Suppl 1): 53–65.
- 16 Ito Y, Tanaka F, Yamamoto M *et al*: Somatic mosaicism of the expanded CAG trinucleotide repeat in mRNAs for the responsible gene of Machado-Joseph disease (MJD), dentatorubral-pallidoluysian atrophy (DRPLA), and spinal and bulbar muscular atrophy (SBMA). *Neurochem Res* 1998; **23**: 25–32.