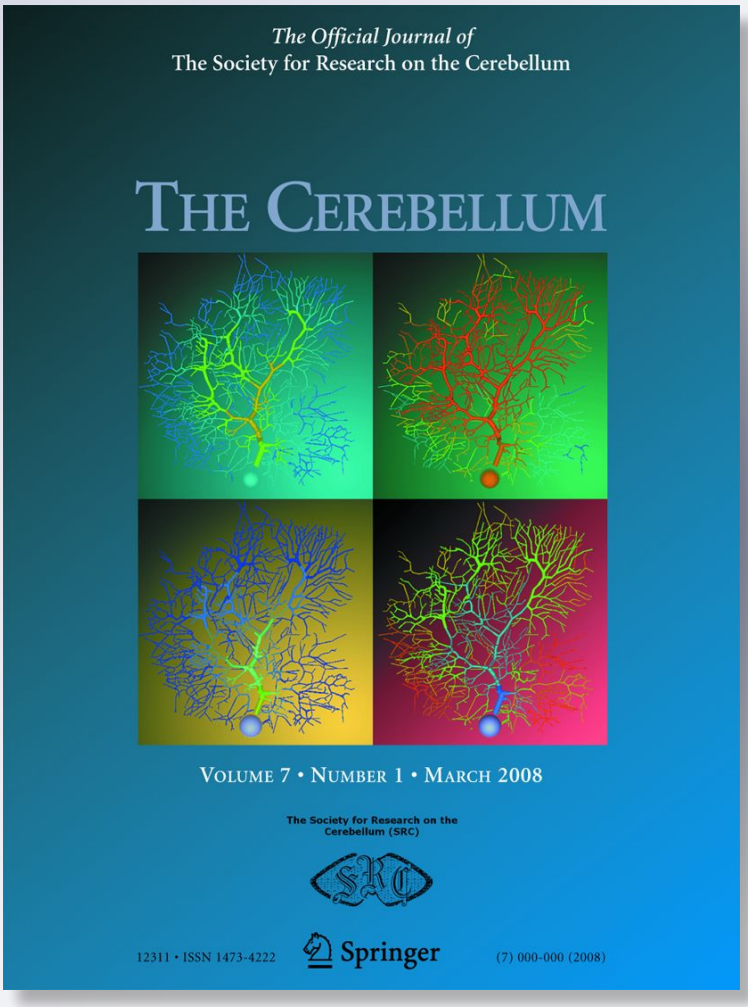
Sequence Analysis of 5# Regulatory Regions of the Machado–Joseph Disease Gene (ATXN3)

Conceição Bettencourt, Mafalda Raposo, Nadiya Kazachkova, Cristina Santos, Teresa Kay, João Vasconcelos, Patrícia Maciel, Karina C. Donis, et al.

The Cerebellum

ISSN 1473-4222

Cerebellum
DOI 10.1007/s12311-012-0373-7



Springer

Cerebellum
DOI 10.1007/s12311-012-0373-7

ORIGINAL PAPER

Sequence Analysis of 5′ Regulatory Regions of the Machado–Joseph Disease Gene (*ATXN3*)

Conceição Bettencourt · Mafalda Raposo · Nadiya Kazachkova · Cristina Santos · Teresa Kay · João Vasconcelos · Patrícia Maciel · Karina C. Donis · Maria Luiza Saraiva-Pereira · Laura B. Jardim · Jorge Sequeiros · Jácome Bruges-Armas · Manuela Lima

Abstract Machado–Joseph disease (MJD) is a late-onset autosomal dominant neurodegenerative disorder, which is caused by a coding $(CAG)_n$ expansion in the *ATXN3* gene (14q32.1). The number of CAG repeats in the expanded alleles accounts only for 50 to 75 % of onset variance, the remaining variation being dependent on other factors. Differential allelic expression of *ATXN3* could contribute to the explanation of different ages at onset in patients displaying

C. Bettencourt · N. Kazachkova · J. Sequeiros · J. Bruges-Armas · M. Lima
Institute for Molecular and Cell Biology (IBMC),
University of Porto,
Porto, Portugal

C. Bettencourt (✉) · M. Raposo · N. Kazachkova · M. Lima
Center of Research in Natural Resources (CIRN) and Department of Biology, University of the Azores,
9501-801 Ponta Delgada, Azores, Portugal
e-mail: mcbettencourt@uac.pt

C. Bettencourt
Laboratorio de Biología Molecular, Instituto de Enfermedades Neurológicas, Fundación Socio-Sanitaria de Castilla-La Mancha,
Guadalajara, Spain

C. Santos
Unitat Antropologia Biològica, Dep. Biologia Animal, Biologia Vegetal i Ecologia, Universitat Autònoma de Barcelona,
Bellaterra,
Barcelona, Spain

T. Kay
Department of Clinical Genetics, Hospital of D. Estefania,
Lisbon, Portugal

J. Vasconcelos
Department of Neurology, Hospital of Divino Espirito Santo,
Ponta Delgada, Portugal

P. Maciel
Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho,
Braga, Portugal

P. Maciel
ICVS/3B's—PT Government Associate Laboratory,
Guimarães, Braga, Portugal

K. C. Donis · M. L. Saraiva-Pereira · L. B. Jardim
Medical Genetics Service, Hospital de Clínicas de Porto Alegre,
Porto Alegre, Brazil

M. L. Saraiva-Pereira
Department of Biochemistry,
Universidade Federal do Rio Grande do Sul,
Porto Alegre, Brazil

L. B. Jardim
Department of Internal Medicine,
Universidade Federal do Rio Grande do Sul,
Porto Alegre, Brazil

J. Sequeiros
ICBAS, University of Porto,
Porto, Portugal

J. Bruges-Armas
Specialized Service of Epidemiology and Molecular Biology (SEEBMO), Hospital of Santo Espirito,
Angra do Heroismo, Portugal

Published online: 16 March 2012

similar CAG repeat sizes. Variation in 5′ regulatory regions of the *ATXN3* gene may have the potential to influence expression levels and, ultimately, modulate the MJD phenotype. The main goal of this work was to analyze the extent of sequence variation upstream of the *ATXN3* start codon. A fragment containing the core promoter and the 5′ untranslated region (UTR) was sequenced and analyzed in 186 patients and 59 controls (490 chromosomes). In the core promoter, no polymorphisms were observed. In the 5′ UTR, only one SNP (rs3814834) was found, but no improvements on the explanation of onset variance were observed, when adding its allelic state in a linear model. Accordingly, in silico analysis predicted that this SNP lays in a nonconserved position for CMYB binding. Therefore, no functional effect could be predicted for this variant.

Keywords Ataxin-3 · 5′ regulatory regions · 5′ UTR · MJD · Promoter · SCA3

Introduction

Machado–Joseph disease (MJD) or spinocerebellar ataxia type 3 (OMIM #109150) is an autosomal dominant neurodegenerative disorder. Most patients show the first symptoms in adulthood (mean ~40 years), but onset extremes of 4 and 70 years have been described [1, 2]. The MJD phenotype is complex and pleomorphic, involving the cerebellar, ocular motor, pyramidal, extrapyramidal, and peripheral motor systems, at variable degrees [2].

MJD's causative mutation corresponds to an expansion of a $(CAG)_n$ motif in exon 10 of the *ATXN3* gene (14q32.1) [3, 4]. Normal alleles range from 12 to 44 CAG repeats, while expanded alleles present more than 52 repeats [5, 6]. The size of the expanded tract is incompletely correlated with the age at onset, explaining from 50 to 75 % of its variance [7, 8]. Other factors (genetic, environmental, or both) should contribute to the clinical variability observed; to date, only the *APOE* gene has been established as a modifier of the age at onset in MJD [9], while *GRIK2*, *IL1B*, and *NEDD8* genes have been suggested as possible modifiers [10].

The *ATXN3* gene encodes for ataxin-3, which displays distinct isoforms [11], and presents, in its mutated form, an abnormal elongation of the polyglutamine tract near its C-terminus [3]. This gene is ubiquitously expressed, in neuronal and non-neuronal tissues, thus the presence of the mutated gene cannot explain by itself the selective neuronal degeneration observed in MJD (reviewed in [6]). The idea that the levels of expression of the mutant allele may contribute to neurodegeneration has been supported by observations from model-based studies [12, 13], as well as by the earlier onset and faster progression of the disease seen in homozygous patients [1, 14, 15]. Therefore, we can hypothesize that higher expression levels of mutant ataxin-3 in heterozygous patients may also be associated with a more severe phenotype, namely with a decrease in age at onset.

Variation in noncoding regulatory regions, i.e., those at 5′, may affect gene expression levels in an allele-specific manner [16]. Besides the promoter, gene expression levels may be further influenced by *cis*-regulatory elements located in the 5′ untranslated region (UTR) [17]. The *ATXN3* promoter has been described as being directly upstream of the gene. Its core region of transcriptional activation, containing consensus sequences for a CCAAT box and SP1 binding sites, appears to be located between −291 and the transcription start site [18]. Although the genomic structure of the *ATXN3* 5′ flanking region has already been described, the extent of variation in this region, which could be involved in transcriptional and translational regulation, remains unexplored.

The main goal of this work was to describe the variation upstream of the *ATXN3* gene start codon, in its core promoter and 5′ UTR, in MJD patients and in controls. Furthermore, this work aimed to investigate potential modulating effects of such putative variation on the age at onset of MJD.

Patients and Methods

Subjects

After informed consent, blood samples were collected from 186 MJD patients (59 from the Azores, 67 from mainland Portugal, and 60 from Brazil) and 59 apparently healthy Caucasian controls (molecularly excluded for the CAG expansion). DNA was extracted from whole blood using standard procedures. The age at onset was available for the series of MJD patients in the study, and it was defined as the age at which the first complaints of gait instability or diplopia were noticed (more details in [9]). The $(CAG)_n$ tract size in *ATXN3* gene was also available, and had been determined according to Bettencourt et al. [19].

Amplification and Sequencing

A fragment of 816 bp, containing the core promoter and the 5′ UTR (Fig. 1), as well as downstream sequences of the *ATXN3* gene, was amplified using the following set of primers: P-1aF (5′-CAAGTGCTGGGTTTTGGGAG-3′) and P-1R (5′-AAAGCGATGGAAAGTGACGG-3′). An initial denaturation step (95 °C, 5 min) was followed by 37 cycles of denaturation (94 °C, 30 s), annealing (56 °C, 30 s), and extension (72 °C, 45 s), and a last step of extension (72 °C, 7 min). PCR products were purified using a Spin Kit (Genomed). All samples were sequenced in both forward and reverse directions, using primers P-1aF and P-1aR (5′-ACGCGGACACTCACTTTCTC-3′), and whenever necessary using P-1R. Sequence reactions were carried out using the

BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and run in an ABI 310 sequencer. All sequence alignments, namely of the obtained sequences with the *Homo sapiens* genomic DNA sequence (GenBank AB038653.1), were performed with BioEdit v7.0.9.0 [20].

In Silico Sequence Analysis

Two predicting tools, TFBIND (http://tfbind.ims.u-tokyo.ac.jp) [21], and MATCH™ (http://www.gene-regulation.com/pub/programs.html#match) [22], were used to find differences in putative transcription factor (TF) binding sites due to rs3814834.

Statistical Analysis

Genotypic and allelic frequencies of rs3814834 were determined. Conformity with the Hardy–Weinberg equilibrium expectations was tested. An exact test of differentiation evaluated differences in genotypic frequencies for this SNP between patients and controls. Analyses were performed using the Arlequin software v.3.1 [23], with a level of significance of 0.05. OpenEpi v.2 proportion calculator (www.openepi.com) was used to determine the 95 % confidence intervals for the allele frequencies present in the Reference SNP Cluster Report of rs3814834 (available at http://www.ncbi.nlm.nih.gov), for the European population (EGP_CEPH-PANEL).

Linear regression analyses were used to test the effect of several variables on age at onset of MJD patients. The multivariate linear model included the following covariates: CAG number in expanded and in normal alleles, genotypes for rs3814834, geographical origin, and gender of the patient. Age at onset, stratified by rs3814834 genotypes, was adjusted for the mean number of CAGs in the expanded *ATXN3* allele, after fitting a linear regression model. Regression analyses were performed using SPSS v.15.0 [24].

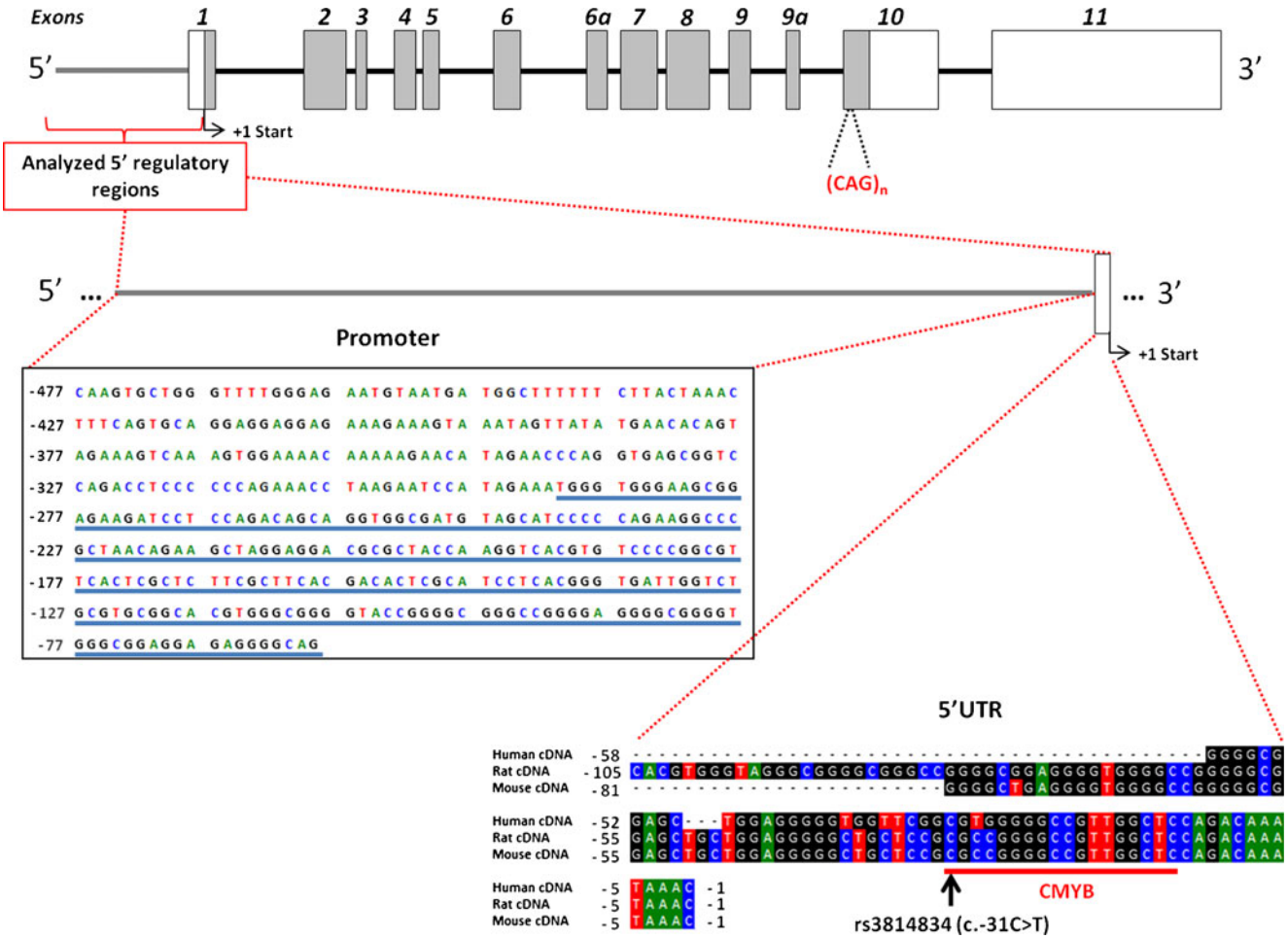


Fig. 1 Diagram representing the genomic region containing the *ATXN3* gene (*grey boxes* representing the coding exons, and *white boxes* the UTRs), with illustration of the analyzed 5′ regulatory regions, including a portion of the promoter (with the core promoter *underlined in blue*, from −59 to −291) and the 5′ UTR (with the rs3814834 at −31, and the binding site for CMYB *underlined in red*). Indications of the start codon (+1 start), and the $(CAG)_n$ tract location are also given

Results

Sequence analysis of 420 bp of the *ATXN3* gene promoter failed to reveal any polymorphism in this region (N=490 chromosomes). At the 5′ UTR, a previously described polymorphism, c.−31C>T (rs3814834), was observed (Fig. 1), in both patients and controls. Genotypic frequencies (Table 1) were in conformity with Hardy–Weinberg expectations. In control samples, all three possible genotypes (C/C, C/T, and T/T) were observed (Table 1). However, in MJD samples, the T/T genotype was absent (Table 1), resulting in statistically significant differences in genotypic frequencies between patients and controls (p=0.00355). Nevertheless, allelic frequencies of 97.7 % ($CI_{95\ \%}$=92.7–99.6) for the C variant, and of 2.3 % ($CI_{95\ \%}$=0.4–7.3) for the T variant, are reported for the European population (EGP_CEPH-PANEL), being similar to those obtained for the MJD patients.

For the studied series of patients, a summary of descriptive statistics is shown in Table 2. Although patients with the C/T genotype (rs3814834) presented an earlier mean age at onset than those with the C/C genotype, such difference disappeared after adjusting the age at onset for the $(CAG)_n$ size (Table 2). The expected negative correlation between the size of the $(CAG)_n$ tract in the expanded alleles and the age at onset was confirmed in this MJD series (r=−0.773, p< 0.001), explaining by itself 59.5 % of onset variance. In relation to the linear model accounting only for the CAG repeat size in expanded alleles, the multivariate model improved the predictability of the age at onset in about 5.4 % (accounting now for 64.9 % of onset variance). This model counted with significant contributions from the geographical origin (p<0.001) and gender of the patients (p=0.005), with male patients presenting earlier ages at onset. Both the $(CAG)_n$ tract size in the normal alleles as well as the allelic state of rs3814834 showed no significant effects.

In silico analysis showed that the rs3814834 falls within the recognition site for the binding of CMYB. However, it lies in a nonconserved position for CMYB binding, not altering the binding of this TF.

Table 1 Genotype and allele frequencies for the c.−31C>T polymorphism (rs3814834) in MJD patients and controls

Frequency (%)	MJD patients (N=186)	Controls (N=59)
Genotypes		
C/C	91.9	78.0
C/T	8.1	18.6
T/T	0.0	3.4
Alleles		
C	96.0	87.3
T	4.0	12.7

Table 2 Descriptive statistics for the MJD patients studied

Characteristics	rs3814834		Total
	C/C	C/T	
Number of patients (%)	171 (91.9)	15 (8.1)	186
Population			
Portugal			
Azores	54 (91.5)	5 (8.5)	59
Mainland	62 (92.5)	5 (7.5)	67
Brazil	55 (91.7)	5 (8.3)	60
Gender			
Male	84 (90.3)	9 (9.7)	93
Female	87 (93.5)	6 (6.5)	93
Age at onset			
Mean±SE (years)	36.97±0.94	34.07±2.94	38.58±0.90
Adjusted onset[a]±SE (years)	38.66±0.61	38.54±1.56	
Range (years)	12–70	12–57	12–70
CAG repeat length			
Normal			
Mean±SE	21.71±0.38	25.07±0.74	21.98±0.36
Range	14–37	19–28	14–37
Expanded			
Mean±SE	72.73±0.29	74.27±0.83	72.85±0.27
Range	63–82	68–81	63–82

[a] Adjusted for the mean size of expanded CAG repeats in the patient sample

Discussion

The absence of polymorphisms in the 490 analyzed chromosomes supports the idea of a conserved *ATXN3* promoter. This is in agreement with what was previously obtained by Costa et al. [25], who analyzed the *atxn3* promoter in mice and found it to be highly conserved even between species. In the 5′ UTR, the variation found does not seem to alter the putative binding of CMYB, thus no functional effect can be predicted for this variant. Results from in silico analysis are in agreement with the absence of improvement of the linear model when adding the allelic state for the detected SNP (rs3814834). The use of age at onset as the only available measure of phenotype severity constitutes a limitation of our study, given its approximate nature, and especially because variants in regulatory regions may be associated with the disease course (in terms of progression rate and severity, for example), but not with the disease onset.

Supporting our rationale about the effect of variants in 5′ regulatory regions, several studies have reported that variants in these regions of certain genes can be associated with the onset of neurodegenerative diseases, as is the case of Alzheimer's disease (e.g., [26]) or Parkinson's disease (e.g.,

[27]). Similarly to our results, polymorphisms in the 5′ regulatory regions of the *HTT* gene have not been significantly associated with the age at onset of Huntington's disease (HD) [28]. Nevertheless, contributions of other variants, further upstream of the analyzed regions or even in other regulatory regions (e.g., 3′ UTR), to differential allelic expression of the *ATXN3* gene, have not been a matter of study.

Although it was out of the scope of this paper, results from the regression analysis suggested that geographical origin (reflecting environmental and/or other population specific factors) as well as gender of the patients (suggesting the existence of sex-linked factors) may be modulating the toxic effect of the polyglutamine expansion in this series of patients. Several examples of environmental and/or sex-linked factors have been reported for other polyglutamine diseases, such as HD (e.g., [29–31]). Specific analyses concerning the effect of such factors in MJD warrant further investigation in future studies.

Summary

In this first report on the extent of genetic variation upstream of the *ATXN3* start codon in MJD patients and controls, the core promoter region was shown to be highly conserved. Although the present series of patients previously enabled the detection of a modifier effect of *APOE* ε2, it was not possible to demonstrate an impact on disease onset related to the SNP (rs3814834) found in the 5′ UTR.

Acknowledgments This work was supported by the grants "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), and "High Prevalence Diseases in the Azores Islands" (M2.1.2/I/026/2008, funded by Institute of Biotechnology and Biomedicine of the Azores). C.B. [SFRH/BPD/63121/2009] is a postdoctoral fellow from FCT. N.K. [M3.1.3/F/004/2009] is a postdoctoral fellow from "Secretaria Regional da Ciência, Tecnologia e Equipamentos". M.L.S.-P. and L.B.J. are supported by CNPq, Brazil. The authors are grateful to Vanessa Emmel and health professionals for their help in the sample and data collection.

Conflict of interest The authors declare that they have no conflict of interests.

References

1. Carvalho DR, La Rocque-Ferreira A, Rizzo IM, Imamura EU, Speck-Martins CE. Homozygosity enhances severity in spinocerebellar ataxia type 3. Pediatr Neurol. 2008;38(4):296–9.
2. Coutinho P. Doença de Machado-Joseph: Tentativa de definição: PhD Dissertation, Instituto de Ciências Biomédicas Abel Salazar, Porto; 1992.
3. 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(3):221–8.
4. 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(7):413–22.
5. Maciel P, Costa MC, Ferro A, et al. Improvement in the molecular diagnosis of Machado–Joseph disease. Arch Neurol. 2001;58(11):1821–7.
6. Bettencourt C, Lima M. Machado–Joseph disease: from first descriptions to new perspectives. Orphanet J Rare Dis. 2011;6:35.
7. 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(1):54–61.
8. Bettencourt C, Santos C, Montiel R, et al. The $(CAG)_n$ tract of Machado–Joseph Disease gene (*ATXN3*): a comparison between DNA and mRNA in patients and controls. Eur J Hum Genet. 2010;18(5):621–3.
9. Bettencourt C, Raposo M, Kazachkova N, et al. The ε2 allele of APOE increases the risk of earlier age-at-onset in Machado–Joseph Disease (MJD/SCA3). Arch Neurol. 2011;68(12):1580–3.
10. Emmel VE, Donis KC, Gheno TC, et al. GRIK2, IL1B, NEDD8 and NEDD9 genes may act as modifiers of the Machado Joseph Disease/SCA 3 phenotype [Abs]. In: 12th ICHG. Montreal; 2011. p 498.
11. Bettencourt C, Santos C, Montiel R, et al. Increased transcript diversity: novel splicing variants of Machado–Joseph Disease gene (*ATXN3*). Neurogenetics. 2010;11(2):193–202.
12. Bilen J, Bonini NM. *Drosophila* as a model for human neurodegenerative disease. Annu Rev Genet. 2005;39:153–71.
13. Goti D, Katzen SM, Mez J, et al. A mutant ataxin-3 putative-cleavage fragment in brains of Machado–Joseph disease patients and transgenic mice is cytotoxic above a critical concentration. J Neurosci. 2004;24(45):10266–79.
14. Lerer I, Merims D, Abeliovich D, Zlotogora J, Gadoth N. Machado–Joseph disease: correlation between the clinical features, the CAG repeat length and homozygosity for the mutation. Eur J Hum Genet. 1996;4(1):3–7.
15. Sobue G, Doyu M, Nakao N, et al. Homozygosity for Machado–Joseph disease gene enhances phenotypic severity. J Neurol Neurosurg Psychiatry. 1996;60(3):354–6.
16. Wang X, Tomso DJ, Liu X, Bell DA. Single nucleotide polymorphism in transcriptional regulatory regions and expression of environmentally responsive genes. Toxicol Appl Pharmacol. 2005;207(2 Suppl):84–90.
17. Chatterjee S, Pal JK. Role of 5′- and 3′-untranslated regions of mRNAs in human diseases. Biol Cell. 2009;101(5):251–62.
18. Schmitt I, Evert BO, Khazneh H, Klockgether T, Wuellner U. The human *MJD* gene: genomic structure and functional characterization of the promoter region. Gene. 2003;314:81–8.
19. 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(4):333–9.
20. Hall TA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser. 1999;41:95–8.
21. Tsunoda T, Takagi T. Estimating transcription factor bindability on DNA. Bioinformatics. 1999;15(7–8):622–30.
22. Kel AE, Gossling E, Reuter I, Cheremushkin E, Kel-Margoulis OV, Wingender E. MATCH: a tool for searching transcription factor binding sites in DNA sequences. Nucleic Acids Res. 2003;31(13):3576–9.
23. Schneider S, Roessli D, Excoffier L. Arlequin: a software for population genetics data analysis, version 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland 2000.
24. Inc. S. SPSS for Windows: Release 15.0. In. Chicago: SPSS Inc.; 2006.

25. do Carmo Costa M, Gomes-da-Silva J, Miranda CJ, Sequeiros J, Santos MM, Maciel P. Genomic structure, promoter activity, and developmental expression of the mouse homologue of the Machado–Joseph disease (MJD) gene. Genomics. 2004;84(2):361–73.
26. Theuns J, Del-Favero J, Dermaut B, et al. Genetic variability in the regulatory region of presenilin 1 associated with risk for Alzheimer's disease and variable expression. Hum Mol Genet. 2000;9(3):325–31.
27. Sutherland G, Mellick G, Sue C, et al. A functional polymorphism in the parkin gene promoter affects the age of onset of Parkinson's disease. Neurosci Lett. 2007;414(2):170–3.
28. Coles R, Leggo J, Rubinsztein DC. Analysis of the 5′ upstream sequence of the Huntington's disease (HD) gene shows six new rare alleles which are unrelated to the age at onset of HD. J Med Genet. 1997;34(5):371–4.
29. Byars JA, Beglinger LJ, Moser DJ, Gonzalez-Alegre P, Nopoulos P. Substance abuse may be a risk factor for earlier onset of Huntington disease. J Neurol 2012 (in press).
30. Friedman JH, Trieschmann ME, Myers RH, Fernandez HH. Monozygotic twins discordant for Huntington disease after 7 years. Arch Neurol. 2005;62(6):995–7.
31. Wexler NS, Lorimer J, Porter J, et al. Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. Proc Natl Acad Sci U S A. 2004;101(10):3498–503.