Genetic variation in *ATXN3* (ataxin-3) 3'UTR: insights into the downstream regulatory elements of the causative gene of Machado-Joseph disease/spinocerebellar ataxia type 3

* Ana Rosa Vieira Melo[1,2]; Mafalda Raposo[1,2]; Marta Ventura[3]; Sandra Martins[4,5]; Sara Pavão[1]; Isabel Alonso[2]; Conceição Bettencourt[6]; Manuela Lima[1,2]

[1]Faculdade de Ciências e Tecnologia, Universidade dos Açores, Ponta Delgada, Portugal; [2]UniGENe, Instituto de Biologia Molecular e Celular (IBMC), Instituto de Investigação e Inovação em Saúde (I3S), Universidade do Porto, Portugal; [3]Food and Nutrition Department, National Institute of Health Doutor Ricardo Jorge (INSA, IP), Lisbon, Portugal ;[4] IPATIMUP - Institute of Molecular Pathology and Immunology of the University of Porto, Porto, Portugal; [5]i3S - Instituto de Investigação e Inovação em Saúde, Universidade do Porto, Porto, Portugal; [6]Department of Neurodegenerative Disease and Queen Square Brain Bank for Neurological Disorders, UCL Queen Square Institute of Neurology, London

*Corresponding author:

Ana Rosa Vieira Melo

Faculty of Sciences and Technology

University of the Azores

Rua da Mãe de Deus, s/n

9500-321 Ponta Delgada

Tel: +351 296 650 170

Email: anarosavieiramelo@gmail.com

CONFLICT OF INTEREST (COI) STATEMENT

All the authors have no conflict of interest to declare.

ABSTRACT

Untranslated regions are involved in the regulation of transcriptional and post-transcriptional processes. Characterization of these regions remains poorly explored for *ATXN3,* the causative gene of Machado-Joseph disease (MJD). Although a few genetic modifiers have been identified for MJD age at onset (AO), they only explain a small fraction of the AO variance. Our aim was to analyse variation at the 3'UTR of *ATXN3* in MJD patients, analyse its impact on AO and attempt to build haplotypes that might discriminate between normal and expanded alleles.

After assessing *ATXN3* 3'UTR variants in molecularly confirmed MJD patients, *in silico* analysis were conducted to predict their functional impact (e.g. their effect on miRNAs binding sites). Alleles in *cis* with the expanded $(CAG)_n$ were inferred from family data and haplotypes were built. The effect of the alternative alleles on the AO, and on SARA and NESSCA ataxia scales was tested.

Nine variants, all previously described, were found. For eight variants, in silico analyses predicted: a) deleterious effects (rs10151135; rs55966267); b) changes on miRNAs binding sites (rs11628764; rs55966267; rs709930) and c) alterations of RNA binding proteins (RBPs) binding sites (rs1055996; rs910369; rs709930; rs10151135; rs3092822; rs7158733). Patients harbouring the alternative allele at rs10151135 had significantly higher SARA Axial subscores (p = 0.023), comparatively with those homozygous for the reference allele. Ten different haplotypes were obtained, one of which was exclusively found in *cis* with the expanded and four with the normal allele. These findings, which are relevant for the design of allele-specific therapies, warrant further investigation in independent MJD cohorts.

KEYWORDS: Genetic modifiers, untranslated regions, transcription regulation, polyglutamine disease, miRNA, gametic phase, MJD, SCA3

INTRODUCTION

Upstream and downstream untranslated regions of genes (5'UTRs and 3'UTRs, respectively), along with *cis*-acting elements (including gene promoters), are crucial in the regulation of gene expression [1-2]. 5'UTRs contain numerous binding sites for transcription factors responsible for regulating transcription initiation and stability [2-3]. 3'UTRs contain (1) targets for microRNAs (miRNAs), which act by physically blocking translation or by inducing transcripts degradation; (2) recognition motifs for RNA binding proteins (RBP), involved in the regulation of protein expression levels; and (3) alternative polyadenylation (APA) signals, which can alter the length of 3'UTRs and affect the binding of regulatory elements [4]. Genetic variation in 5' and 3'UTRs of disease-causing genes can have a significant impact on the regulation of their expression in an allele-specific manner, potentially modulating the disease phenotype. Furthermore, because therapeutic modulation of mRNA expression holds great promise for the treatment of many hereditary diseases [5], a comprehensive understanding of genetic variation in these regions is crucial.

The *ATXN3* gene (14q32.12), encoding for ataxin-3, is the causative gene of Machado-Joseph disease (MJD)/spinocerebellar ataxia type 3 (SCA3). MJD is a currently untreatable late onset disease; MJD belong to the group of autosomal dominant polyglutamine (polyQ) neurodegenerative disorders. The *ATXN3* gene contains 13 exons [6-7], spanning a genomic region of around 48 kilobases (kb) and presenting a 3'UTR of variable length, reflecting the alternative use of exons 10 and 11. Exon 10 contains the (CAG)n repeat tract, which is abnormally expanded in MJD mutation carriers/patients. Normal *ATXN3* alleles contain 12 to 44 CAG repeats, whereas expanded alleles display between 52 and 86 CAGs [8]. The $(CAG)_n$ number in the expanded allele explains only 45-70% of the variance of age of the first symptoms (age at onset – AO) [9]; the remaining 30-55% should be explained by other factors, including genetic modifiers [10-15]. Although few studies have investigated genetic variants at

ATXN3 3'UTR and its potential impact on AO [16-17], none of these studies has comprehensively covered the whole extension of the *ATXN3* 3'UTR. Previous reports on the interactions of endogenous miRNAs (namely miR-25) with *ATXN3* 3'UTR, using cell and animal models, suggested that the overexpression of specific miRNAs contributed to the reduction of mutant ATXN3 protein levels, and to a decrease in protein aggregates [18-19]. Following up this evidence, Long and collaborators [17] studied two substitution variants (rs910369 and rs709930) located upstream and downstream a miR-25 binding site in a Chinese cohort of MJD patients. They reported an association between rs709930 and the disease AO. Such findings, however, have not been replicated in independent cohorts.

The aims of this study were to describe genetic variation at the *ATXN3* 3'UTR, to predict its functional effects, to investigate the impact of 3'UTR variants on the MJD AO, and to build (CAG)n allele-specific haplotypes with the variants identified.

PATIENTS AND METHODS

Patients

One hundred DNA samples (50 males: 50 females; total cohort), extracted from peripheral blood of molecularly confirmed Azorean MJD patients, were used in this study. Data from an extended whole-exome sequencing was available for a sub-group of nineteen molecularly confirmed Azorean MJD patients (discovery sample). Sizing of the CAG tract at *ATXN3* was performed according to Bettencourt and collaborators [20]. AO was defined as the self-reported age of the first symptoms of the disease. In the total cohort, the average AO was 37.37 ± 11.24 (mean ± standard deviation) and the average size of the CAG tract was 21.42 ± 4.59 for the normal allele, and 70.56 ± 3.63 for the expanded allele. For a subset of patients, from the total cohort, scores for Scale for the Assessment and Rating of Ataxia (SARA) (N=34), and Neurological Examination Score for Spinocerebellar Ataxia (NESSCA) (N=65) were available. For those patients the average SARA score was 14.56 ± 8.00 and the average NESSCA score was 12.25 ± 5.11. For total cohort the mean disease duration was 21.99 ± 9.76 years.

This study was approved by the Ethics Committee of the University of the Azores (Parecer 2/2016 and Parecer 5/2018); written informed consent was obtained from all participants.

Genetic variation at ATXN3 3'UTR

An extended whole-exome sequencing (WES) covering the UTRs was performed for nineteen MJD patients (discovery sample), using the *Agilent SureSelect XT V5 + UTR* 75Mb capture kit and run in the Illumina *HiSeq4000* platform with a 100 bp paired end reads. Sequence alignment and variant calling were performed against the Reference Consortium Human Build 37 (GRCh37) using the Burrows-Wheeler Aligner (bwa-mem 0.7.17) [21], and the Genome Analysis Toolkit (GATK 3.7) [22]. A mean target coverage of 54.40X was obtained for *ATXN3*

3'UTR and all substitution variants (from now on designated as variants) were covered at least 10X.

Variants at the *ATXN3* 3'UTR identified from WES data were validated by Sanger sequencing, and genotyped, after the alignment of the sequences with the sequence NM 004993.4, in 81 additional samples (total of 100 samples – total cohort) also by Sanger sequencing. Further details, including primers sequences and conditions, are available upon request.

Allelic and genotypic frequencies, Hardy-Weinberg equilibrium tests with Bonferroni correction, and linkage disequilibrium test between the identified variants were determined using Arlequin software v.3.5 [23]. The polymorphism information content (PIC), a quantitative measure of the degree of polymorphism [24], was calculated for all variants using a codominant model in Gene-Calc tool (https://gene-calc.pl/pic). PIC values for co-dominant systems range from 0 (monomorphic) to 1 (very highly informative, with several alleles of equal frequency).

Population differentiation tests using genotypic frequencies were performed to compare the total cohort of Azorean patients with healthy populations from Europe (HapMap-CEU and AFD_EUR_PANEL; frequencies assessed on November 17th, 2020 in dbSNP - https://www.ncbi.nlm.nih.gov/snp/) – European controls. Comparisons were only performed using populations with N>20. For rs910369 and rs709930, Azorean genotypic frequencies were also compared with those previously reported by Long *et al*. for Chinese MJD patients and controls [17].

The allelic phase between each of the nine analysed variants and the normal/expanded *ATXN3* allele was inferred, whenever possible, by segregation analysis in the MJD families. In patients homozygous for all variants, haplotypes in *cis* with the expanded CAG repeat were directly assessed.

In silico predictions for the functional impact of variants in ATXN3 3'UTR

Functional effects of variants found in the *ATXN3* 3'UTR were predicted using Functional Analysis Through Hidden Markov Models (FATHMM) (http://fathmm.biocompute.org.uk/) [25], Genome Wide Annotation of Variants (GWAVA) (https://www.sanger.ac.uk/sanger/StatGen_Gwava) [26], and Variant Effect Predictor – VEP (http://www.ensembl.org/Tools/VEP). For FATHMM and GWAVA, a cut-off of 0.5 was considered, with variants showing scores higher than 0.5 being predicted to have a deleterious effect. Predicted *ATXN3* miRNA binding sites were identified using *miRmap* [27] and *Target Scan v.7.1* tools [28]. Variants upstream and downstream (~200 bp) of miRNA binding sites, which could have a potential impact on miRNA binding, were searched using *MirSNP* (http://bioinfo.bjmu.edu.cn/mirsnp/search/) [29]. For this analysis only predicted miRNAs binding sites with a score > 150 and an energy score < -7 were considered [26]. To explore the results obtained by these tools further, the miRNAs previously identified were compared to those found, using the software Ingenuity Pathways Analysis (QIAGEN, IPA), to bind or were predicted bind to *ATXN3* gene. RBPs motifs were explored using the tool *RBP-Var2* (http://www.rbp-var.biols.ac.cn/) [30] with default settings. The search for alternative polyadenylation (polyA) motifs generated or abolished by the *ATXN3* 3' UTR variants was performed by using PolyASite 2.0 [31], with default settings.

Impact of ATXN3 3'UTR variants in phenotype of MJD

An analysis of covariance (ANCOVA), using the number of CAGs in the expanded *ATXN3* allele as covariate, was conducted to evaluate the effect of presence/absence of the identified 3'UTR variants on MJD AO. For all cases, patients carrying the alternative allele were compared to the patients homozygous for the reference allele. Using the number of CAGs in the expanded *ATXN3* allele as covariate, an ANCOVA was also performed in the subset of samples with

known allelic phase, to evaluate the effect of the presence of the alternative allele in *cis* with the expanded CAG repeat. Using disease duration as covariate, an ANCOVA was performed in the subset of samples with NESSCA and SARA scores and subscores, to further evaluate the effect of the presence of the alternative allele; NESSCA subscores included the following items: 1) gait ataxia; 2) limb ataxia; 3) dysphagia; 4) dysarthria; 5) fasciculation and 6) distal amyotrophies. For SARA, two different subscores were considered: a) axial score, constituted by the items sitting, stance and gait; and b) the appendicular score, constituted by the items finger chase, nose finger and fast alternative movement. Statistical analysis was performed using *IBM SPSS Statistics v.25.0*; *p* values lower than 0.05 were considered statistically significant

RESULTS

Genetic variation at the ATXN3 3'UTR and their predicted functional impact

In the discovery sample (N=19), analysed by extended whole exome sequencing, a total of nine variants - rs1055996, rs11628764, rs55966267, rs910369, rs709930, rs10151135, rs7158238, rs3092822 and rs7158733 (Figure 1) were found in the *ATXN3* 3'UTR. All variants were previously described in dbSNP and were subsequently validated in a cohort of 100 MJD patients. Allelic and genotypic frequencies for the 3'UTR variants along with the functional predictions from the *in-silico* analyses are shown in Table 1. The frequency of alternative alleles was >5% for all variants, with rs55966267 showing the lowest minor allele frequency (MAF = 6%). PIC values ranged from 0.13 to 0.46 (Table 1), with only rs11628764, rs55966267 and rs10151135 showing PIC values lower than 0.40, suggesting these three variants would be less informative when trying to discriminate normal and expanded alleles in MJD patients. Genotypic frequencies for variants rs1055996, rs11628764, rs910369, rs709930, rs10151135, rs7158238, rs3092822 and rs7158733 were in conformity with Hardy-Weinberg equilibrium expectations.

The genotypic frequencies for rs910369 and rs10151135 differed significantly between MJD patients (total cohort) and Europeans controls (Supplementary table 1). These two variants, which are not in LD with each other (D'=0.18, r^2=0, *p*=0.80; Supplementary table 2), present higher heterozygosity in patients than in European controls and consequently higher frequencies of their corresponding minor alleles.

For variants rs910369 and rs709930, previously studied by Long *et al.* [17], no differences in genotypic frequencies between Azorean and Chinese MJD patients were observed. However, there were significant differences in the genotypic frequencies between the Chinese MJD cohort and both Chinese and European controls (Supplementary table 1). Similarly, and as

above mentioned, the rs910369 also differed between the Azorean MJD cohort and control groups.

For a subset of samples, it was possible to construct haplotypes containing all nine variants analysed (Supplementary table 3). The 10 different haplotypes obtained and their absolute frequencies when in *cis* with the expanded or the normal allele are shown in Figure 2 a) and 2 b), respectively. Six different haplotypes were identified as being in *cis* with the expanded *ATXN3* allele; noteworthy, haplotype 10 (GTTATGTGT) was only observed in *cis* with the expanded allele. Moreover, the frequency of haplotype 2 (ATTCCGCTG) in the expanded allele is 2-fold higher than in the normal allele. Nine haplotypes were associated with the normal allele; four of which were only present in the normal alleles: haplotype 4 (GCTCCTCTG), haplotype 6 (ACTCCTCGG), haplotype 8 (GTTATTTGT), and haplotype 9 (GTTCCGCTG). Overall, these results indicate that haplotype 10 might be valuable in distinguishing the expanded *ATXN3* allele and haplotypes 2, 4, 6, 8 and 9 might be valuable in distinguishing the normal *ATXN3* allele.

Regarding *in silico* predictions (Table 1), Variant Effect Prediction tool confirmed that variant rs7158733 leads to the formation of a premature stop codon as previously reported [32], whereas a deleterious effect was predicted for rs10151135. None of the variants identified in the 3'UTR region of *ATXN3* is predicted to alter polyadenylation signals. A total of 1137 miRNA binding sites were identified in the *ATXN3* 3'UTR. mirSNP predicts that the presence of three variants (rs11628764, rs55966267 and rs709930) can modify the affinity of miRNAs to a binding site (Table 1). In the case of rs11628764 and rs709930, one new miRNAs binding sites per each variant, would be created. Except for hsa-miR-586 we were able to confirm the predictions obtained for the miRNAs binding sites using the software Ingenuity Pathways Analysis (QIAGEN, IPA).

Furthermore, six out of the nine variants were predicted to affect the secondary structure of mRNA, the binding of RBPs and the expression of the RNA (score = 1e, Table 1). Six variants map to target sites of RBPs, which are expected to alter the RNA structure. These RBPs include: WDR33 (rs1055996 and rs910369); RBM47, FUS, ZC3H7B, ELAVL1, AGO2 and PTBP1 (rs910369); PTBP1 (rs3092822 and rs7158733); eIF4AIII (rs709930 and rs10151135); and AGO2 and ELAVL1 (rs709930).

Impact of the ATXN3 3'UTR variants in phenotype of MJD

For all nine variants, after adjusting for the size of expanded (CAG)n allele, no significant differences were observed on AO between patients carrying the alternative allele (heterozygous or homozygous) and reference allele (Supplementary table 3). Taking into consideration that in heterozygous patients it is difficult to dissect allele-specific effects without knowing the allelic phase relative to the expanded (CAG)n allele, we restricted this comparison to patients with known allelic phase which was possible to determine individually for six variants. In this smaller set of patients (N=47 to N=88, depending on which variant), no significant effects could be observed. These results could be a consequence of the limitation of the smaller sample size of this subset of samples.

Regarding rs709930, a variant previously described by Long and collaborators [17] as being associated with AO in Chinese MJD patients, we were not able to replicate the observation of an earlier AO in patients that carried the alternative allele, as previously described [17].

The analysis exploring the association between the variants identified and NESSCA and SARA scores and subscores revealed that for rs10151135 patients carrying the alternative allele presented a higher SARA Axial subscore than patients homozygous for the reference allele (p = 0.023).

DISCUSSION

Our work reports nine variants present in the 3' UTR of MJD causative gene - *ATXN3*, which may be involved in the regulation of the expression of this gene and may help to discriminate between *ATXN3* normal and expanded alleles.

In silico predictions revealed several functional impacts for the studied variants: a) two new miRNA binding sites, which can have implications in modulation of gene expression; b) changes in the secondary structure of mRNA, the binding of RBPs and expression of the RNA; and c) the creation of a premature stop codon that may lead to a truncated form of the ATXN3 protein. These predictions confirm the complexity and potential importance of the 3'UTR regulation of *ATXN3* expression. As previously referred, the presence of a variant could affect the binding of a miRNA leading to a change in the normal regulation of protein expression [4]. Prior studies in MJD revealed that the overexpression of several miRNAs (miR-9, miR-25 miR-181a and miR-494) contributed to the reduction of the ATXN3 protein levels and to a decrease in this protein aggregates [18-19]. In another study, miR-25 as well as miR-29a, miR-34b, miR-125b and miR-7014 were shown to have a potential as biomarkers of disease severity of importance in future studies about the pathogenesis of the disease as well as in drug development efforts [32-33]. We were able to confirm the results obtained for the miRNAs binding sites using the software Ingenuity Pathways Analysis (QIAGEN, IPA), and only one of the miRNAs was not found: hsa-miR-586. As this miRNA binding site is predicted to be created by the tool that was previously used, two reasons could be pointed out: a) it could be a false result or b) the database of the software used did not have the information for this miRNA binding site. None of the miRNAs predicted to target binding sites altered by the presence of the variants identified in the present study were reported in these previous studies [18-19; 32-33]; although *in silico* results are consistent across prediction tools, functional studies concerning the miRNAs aspects should be performed.

A significant modulatory effect of the alternative allele at rs10151135 on SARA Axial subscore, was identified. Noteworthy, this variant is predicted to have a deleterious effect according to several prediction tools and therefore its functional effect deserves further investigation.

Despite possible functional changes/deleterious effects due to variants in 3'UTR, none of the identified variants had significant modulatory effects on the disease age at onset. This can be due to sample size limitations, the low frequency of alleles with predicted deleterious effects (e.g. rs55966267 MAF = 0.06), and/or the small magnitude of effects for these variants. Also, the limited size of the discovery sample could influence the results, as the remaining 81 different samples could have additional variants that could affect the age at onset. For rs709930, previously associated to the MJD AO [17], we were not able to obtain the result of an earlier onset in patients whose alternative allele was in *cis* with expanded CAG. Noteworthy, in the study by Long and collaborators [17], 67.6% of the patients carried the alternative allele for rs709930 compared to 29.8% in our study. Such population-specific differences in allelic frequencies may have facilitated the detection of an effect in the Chinese MJD cohort as we previously reported for other modifier genes on MJD age at onset [11].

Costa *et al*. [34] highlight the importance, in the context of the development of allele-specific therapies, of describing disease-associated haplotypes that could discriminate between normal and expanded alleles. The design of therapies targeting allele-specific inhibition (affecting the expanded allele only) is based on the assumption that expanded alleles can be discriminated from their wild-type counterparts, offering advantages in relation to the non-allele-specific silencing strategy previously applied in a rat model of MJD [35]; an example could be the development of small interfering RNAs, which based on the presence of a specific variant, or haplotype, could discriminate between normal allele and expanded allele and silence/modulate the expression of mutant *ATXN3* [36]. In this work we describe haplotypes exclusively found in *cis* with the expanded (N=1) or with the normal allele (N=4), and that should be of discriminative utility; we further identified a haplotype whose frequency is

markedly higher in *cis* with the expanded allele. These findings, which are highly relevant for the design of allele-specific therapies, warrant further investigation in independent cohorts of MJD patients.

CONCLUSION

In summary, some the observed variants are predicted to have functional impact on the expression of *ATXN3*. Several factors could justify our inability to find significant associations between the presence of these variants and age at onset for the disease, namely limitations on sample size; access to additional cohorts of MJD patients should allow us to confirm this hypothesis. We were further able to identify disease-associated haplotypes with discriminative ability for the expanded allele, which could be of relevance for future design of allele-specific silencing strategies.

ACKNOWLEDGEMENTS

The clinical evaluation of patients included in this study was performed by João Vasconcelos (MD). This work was financed by FEDER - Fundo Europeu de Desenvolvimento Regional funds through the COMPETE 2020 - Operacional Programme for Competitiveness and Internationalisation (POCI), Portugal 2020, and by Portuguese funds through FCT - Fundação para a Ciência e Tecnologia in the framework of the EXOS3 project: PTDC/DTP-PIC/2638/2017 (POCI-01-0145-FEDER-016592). ARVM is supported by a PhD grant (SFRH/BD/129547/2017) financed by FCT – Fundação para a Ciência e Tecnologia. MR and SM are supported by Fundação para a Ciência e Tecnologia (CEECIND/03018/2018 and CEECIND/00684/2017, respectively). CB is supported by the Multiple System Atrophy Trust, and Alzheimer's Research UK.

REFERENCES

1. Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. Genome Biol. 2002;3(3): REVIEWS0004. doi: 10.1186/gb-2002-3-3-reviews0004.
2. Barrett LW, Fletcher S & Wilton SD (2013). Untranslated Gene Regions and Other Non-coding Elements, *Springer Briefs in Biochemistry and Molecular Biology.*
3. Liu H, Yin J, Xiao M, Gao C, Mason AS, Zhao Z, Liu Y, Li J, Fu D. Characterization and evolution of 5' and 3' untranslated regions in eukaryotes. Gene. 2012; 507(2):106-11. doi: 10.1016/j.gene.2012.07.034.
4. Wanke KA, Devanna P, Vernes SC. Understanding Neurodevelopmental Disorders: The Promise of Regulatory Variation in the 3'UTRome. Biol Psychiatry. 2018; 83(7):548-557. doi: 10.1016/j.biopsych.2017.11.006.
5. Asrani KH, Farelli JD, Stahley MR, Miller RL, Cheng CJ, Subramanian RR, Brown JM. Optimization of mRNA untranslated regions for improved expression of therapeutic mRNA. RNA Biol. 2018;15(6):756-762. doi: 10.1080/15476286.2018.1450054.
6. Ichikawa Y, Goto J, Hattori M, Toyoda A, Ishii K, Jeong SY, Hashida H, Masuda N, Ogata K, Kasai F, Hirai M, Maciel P, Rouleau GA, Sakaki Y, Kanazawa I. The genomic structure and expression of MJD, the Machado-Joseph disease gene. J Hum Genet. 2001;46(7):413-22. doi: 10.1007/s100380170060.
7. Bettencourt C, Santos C, Montiel R, Costa Mdo C, Cruz-Morales P, Santos LR, Simões N, Kay T, Vasconcelos J, Maciel P, Lima M. Increased transcript diversity: novel splicing variants of Machado-Joseph disease gene (*ATXN3*). Neurogenetics. 2010;11(2):193-202. doi: 10.1007/s10048-009-0216-y.
8. Bettencourt C, Lima M. Machado-Joseph Disease: from first descriptions to new perspectives. Orphanet J Rare Dis. 2011; 6:35. doi: 10.1186/1750-1172-6-35.
9. de Mattos EP, Kolbe Musskopf M, Bielefeldt Leotti V, Saraiva-Pereira ML, Jardim LB. Genetic risk factors for modulation of age at onset in Machado-Joseph

disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. 2019; 90(2):203-210. doi: 10.1136/jnnp-2018-319200.

10. Bettencourt C, Raposo M, Kazachkova N, Cymbron T, Santos C, Kay T, Vasconcelos J, Maciel P, Donis KC, Saraiva-Pereira ML, Jardim LB, Sequeiros J, Lima M. The APOE ε2 allele increases the risk of earlier age at onset in Machado-Joseph disease. Arch Neurol. 2011; 68(12):1580-3. doi: 10.1001/archneurol.2011.636.
11. Raposo M, Ramos A, Bettencourt C, Lima M. Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid? Brain. 2015; 138(Pt 12):e398. doi: 10.1093/brain/awv206.
12. Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, Stevanin G, Koutsis G, Karadima G, Panas M, Yescas-Gómez P, García-Velázquez LE, Alonso-Vilatela ME, Lima M, Raposo M, Traynor B, Sweeney M, Wood N, Giunti P; SPATAX Network, Durr A, Holmans P, Houlden H, Tabrizi SJ, Jones L. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. Ann Neurol. 2016;79(6):983-90. doi: 10.1002/ana.24656
13. de Mattos EP, Kolbe Musskopf M, Bielefeldt Leotti V, Saraiva-Pereira ML, Jardim LB. Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis. J Neurol Neurosurg Psychiatry. 2019;90(2):203-210. doi: 10.1136/jnnp-2018-319200.
14. Mergener R, Furtado GV, de Mattos EP, Leotti VB, Jardim LB, Saraiva-Pereira ML. Variation in DNA Repair System Gene as an Additional Modifier of Age at Onset in Spinocerebellar Ataxia Type 3/Machado-Joseph Disease. Neuromolecular Med. 2020; 22(1):133-138. doi: 10.1007/s12017-019-08572-4.
15. Akçimen F, Martins S, Liao C, Bourassa CV, Catoire H, Nicholson GA, Riess O, Raposo M, França MC, Vasconcelos J, Lima M, Lopes-Cendes I, Saraiva-Pereira ML, Jardim LB, Sequeiros J, Dion PA, Rouleau GA. Genome-wide association study identifies genetic

factors that modify age at onset in Machado-Joseph disease. Aging (Albany NY). 2020; 12(6):4742-4756. doi: 10.18632/aging.102825.

16. Ventura M (2011). Doença de Machado-Joseph: análise da região 3'UTR do gene *ATXN3* (Master thesis, University of the Azores, Ponta Delgada, S.Miguel, Azores, Portugal). Retrieved from http://hdl.handle.net/10400.3/1183.
17. Long Z, Chen Z, Wang C, Huang F, Peng H, Hou X, Ding D, Ye W, Wang J, Pan Q, Li J, Xia K, Tang B, Ashizawa T, Jiang H. Two novel SNPs in ATXN3 3' UTR may decrease age at onset of SCA3/MJD in Chinese patients. PLoS One. 2015; 10(2): e0117488. doi: 10.1371/journal.pone.0117488.
18. Huang F, Zhang L, Long Z, Chen Z, Hou X, Wang C, Peng H, Wang J, Li J, Duan R, Xia K, Chuang DM, Tang B, Jiang H. miR-25 alleviates polyQ-mediated cytotoxicity by silencing ATXN3. FEBS Lett. 2014; 588(24):4791-8. doi: 10.1016/j.febslet.2014.11.013.
19. Carmona V, Cunha-Santos J, Onofre I, Simões AT, Vijayakumar U, Davidson BL, Pereira de Almeida L. Unravelling Endogenous MicroRNA System Dysfunction as a New Pathophysiological Mechanism in Machado-Joseph Disease. Mol Ther. 2017;25(4):1038-1055. doi: 10.1016/j.ymthe.2017.01.021.
20. Bettencourt C, Fialho RN, Santos C, Montiel R, Bruges-Armas J, Maciel P, Lima M. 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):333-339. doi: 10.1007/s10038-008-0261-7.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009; 25(14):1754-60. doi: 10.1093/bioinformatics/btp324.
22. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20(9):1297-303. doi: 10.1101/gr.107524.110.

23. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010; 10(3):564-7. doi: 10.1111/j.1755-0998.2010.02847.x.

24. Serrote C, Reiniger L, Silva K, Rabaiolli S, Stefanel C. Determining the Polymorphism Information Content of a molecular marker. *Gene*. 2020; 726:144175. doi: 10.1016/j.gene.2019.144175.

25. Shihab HA, Rogers MF, Gough J, Mort M, Cooper DN, Day IN, Gaunt TR, Campbell C. An integrative approach to predicting the functional effects of non-coding and coding sequence variation. Bioinformatics. 2015; 31(10):1536-43. doi: 10.1093/bioinformatics/btv009.

26. Ritchie GR, Dunham I, Zeggini E, Flicek P. Functional annotation of noncoding sequence variants. Nat Methods. 2014; 11(3):294-6. doi: 10.1038/nmeth.2832.

27. Vejnar CE, Zdobnov EM. MiRmap: comprehensive prediction of microRNA target repression strength. Nucleic Acids Res. 2012; 40(22):11673-83. doi: 10.1093/nar/gks901.

28. Agarwal V, Bell GW, Nam JW, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. Elife. 2015; 12(4): e05005. doi: 10.7554/eLife.05005.

29. Liu C, Zhang F, Li T, Lu M, Wang L, Yue W, Zhang D. MirSNP, a database of polymorphisms altering miRNA target sites, identifies miRNA-related Variants in GWAS Variants and eQTLs. BMC Genomics. 2012; 13:661. doi: 10.1186/1471-2164-13-661.

30. Mao F, Xiao L, Li X, Liang J, Teng H, Cai W, Sun ZS. RBP-Var: a database of functional variants involved in regulation mediated by RNA-binding proteins. Nucleic Acids Res. 2016; 44(D1): D154-63. doi: 10.1093/nar/gkv1308.

31. Herrmann CJ, Schmidt R, Kanitz A, Artimo P, Gruber AJ, Zavolan M. PolyASite 2.0: a consolidated atlas of polyadenylation sites from 3' end sequencing. Nucleic Acids Res. 2020; 48(D1): D174-D179. doi: 10.1093/nar/gkz918.

32. Shi Y, Huang F, Tang B, Li J, Wang J, Shen L, Xia K, Jiang H. MicroRNA profiling in the serums of SCA3/MJD patients. Int J Neurosci. 2014;124(2):97-101. doi: 10.3109/00207454.2013.827679.

33. Hou X, Gong X, Zhang L, Li T, Yuan H, Xie Y, Peng Y, Qiu R, Xia K, Tang B, Jiang H. Identification of a potential exosomal biomarker in spinocerebellar ataxia Type 3/Machado-Joseph disease. Epigenomics. 2019 ;11(9):1037-1056. doi: 10.2217/epi-2019-0081.

34. Costa IPD, Almeida BC, Sequeiros J, Amorim A, Martins S. A Pipeline to Assess Disease-Associated Haplotypes in Repeat Expansion Disorders: The Example of MJD/SCA3 *Locus*. Front Genet. 2019; 10: 38. doi: 10.3389/fgene.2019.00038.

35. Alves S, Nascimento-Ferreira I, Auregan G, Hassig R, Dufour N, Brouillet E, Pedroso de Lima MC, Hantraye P, Pereira de Almeida L, Déglon N. Allele-specific RNA silencing of mutant ataxin-3 mediates neuroprotection in a rat model of Machado-Joseph disease. PLoS One. 2008 8;3(10): e3341. doi: 10.1371/journal.pone.0003341.

36. Prudencio M, Garcia-Moreno H, Jansen-West KR, Al-Shaikh RH, Gendron TF, Heckman MG, Spiegel MR, Carlomagno Y, Daughrity LM, Song Y, Dunmore JA, Byron N, Oskarsson B, Nicholson KA, Staff NP, Gorcenco S, Puschmann A, Lemos J, Januário C, LeDoux MS, Friedman JH, Polke J, Labrum R, Shakkottai V, McLoughlin HS, Paulson HL, Konno T, Onodera O, Ikeuchi T, Tada M, Kakita A, Fryer JD, Karremo C, Gomes I, Caviness JN, Pittelkow MR, Aasly J, Pfeiffer RF, Veerappan V, Eggenberger ER, Freeman WD, Huang JF, Uitti RJ, Wierenga KJ, Marin Collazo IV, Tipton PW, van Gerpen JA, van Blitterswijk M, Bu G, Wszolek ZK, Giunti P, Petrucelli L. Toward allele-specific targeting therapy and pharmacodynamic marker for spinocerebellar ataxia type 3. Sci Transl Med. 2020; 21;12(566): eabb7086. doi: 10.1126/scitranslmed.abb7086.

TABLES

Table 1. Characterization of the *ATXN3* 3'UTR variants identified in the present study.

Variant	Allelic Frequency	Genotypic Frequency (PIC)	Functional effect FATHMM/GWAVA (Scores)*	MirSNP - miRNA binding sites (BS)**	RBP binding - RBP-Var2 scores***
rs1055996 (N=88) (NC_000014.8: g.92527399A>G)	A: 0.71 G: 0.29	AA: 0.56 AG: 0.31 GG: 0.13 (PIC = 0.41)	Benign or neutral (0.17/0.33)	No predicted modifications	1e
rs11628764 (N=90) (NC_000014.8: g.92527977T>C)	T: 0.87 C: 0.13	TT :0.74 TC :0.26 CC: 0.00 (PIC = 0.23)	Benign or neutral (0.17/0.43)	hsa-miR-374c-5p (New BS) hsa-miR-3688-3p (- binding) hsa-miR-4694-3p (- binding)	NA
rs55966267 (N=79) (NC_000014.8: g.92529558T>C)	T: 0.94 C: 0.06	TT: 0.91 TC: 0.05 CC: 0.04 (PIC = 0.13)	Benign or neutral/Deleterious (0.18/**0.59**)	hsa-miR-335-3p (+ binding)	6
rs910369 (N=100) (NC_000014.8: g.92530282C>A)	C: 0.66 A: 0.34	CC: 0.42 CA: 0.48 AA: 0.10 (PIC = 0.45)	Benign or neutral (0.09/0.29)	NA	1e
rs709930 (N=79) (NC_000014.8: g.92530513C>T)	C: 0.65 T: 0.35	CC: 0.42 CT: 0.47 TT: 0.11 (PIC = 0.45)	Benign or neutral (0.15/0.49)	hsa-miR-586 (New BS)	1e
rs10151135 (N=78) (NC_000014.8: g.92530551G>T)	G: 0.83 T: 0.17	GG: 0.67 GT: 0.30 TT: 0.03 (PIC = 0.28)	Deleterious (**0.80/0.51**)	No predicted modifications	1e
rs7158238 (N=100) (NC_000014.8: g.92537105C>T)	C: 0.64 T: 0.36	CC: 0.43 CT: 0.41 TT: 0.16 (PIC = 0.46)	Benign or neutral 0.07/0.15	No predicted modifications	4
rs3092822 (N=100) (NC_000014.8: g.92537163T>G)	T: 0.67 G: 0.33	TT: 0.43 TG: 0.45 GG: 0.12 (PIC = 0.45)	Benign or neutral (0.15/0.16)	NA	1e
rs7158733 (N=100) (NC_000014.8: g.92537223G>T)	G: 0.66 T: 0.34	GG: 0.46 GT: 0.40 TT: 0.14 (PIC = 0.45)	Benign or neutral (0.07/0.28) ◊	NA	1e

For all substitution variants, alleles are provided on the forward (or plus) genomic strand.

*Scores in bold highlight predicted deleterious effects.

**NA – no information available; + binding: enhanced miRNA affinity for the binding site, - binding: decreased miRNA affinity for the binding site.

***Scores vary from 1 to 6: a score of 1 (a,b,c,d and e) is indicative of a higher impact in the RNA structure and in the modification of the binding of RBPs; score 6 is the one with a lower impact in the modification of the binding site of RBPs.

◊ According to Variant Effect Predictor (VEP) this variant was predicted to originate a stop codon.

Table 2. Age at onset adjusted for the size of expanded $(CAG)_n$ alleles observed in MJD patients with known allelic phase.

Variants ID	Patients carrying the reference allele in *cis* with the expanded $(CAG)_n$ allele		Patients carrying the alternative allele in *cis* with the expanded $(CAG)_n$ allele		*p* value
	n	Adj. mean AO	n	Adj. mean AO	
rs1055996	49	36.68 ± 0.83	39	36.25 ± 1.04	*0.76*
rs910369	42	35.15 ± 1.05	18	35.31 ± 1.61	*0.94*
rs709930	33	35.38 ± 1.21	14	34.76 ± 1.86	*0.78*
rs7158238	42	37.26 ± 1.09	19	36.64 ± 1.64	*0.76*
rs3092822	45	36.71 ± 1.09	17	37.14 ± 1.79	*0.84*
rs7158733	46	35.94 ± 1.09	16	37.85 ± 1.88	*0.61*

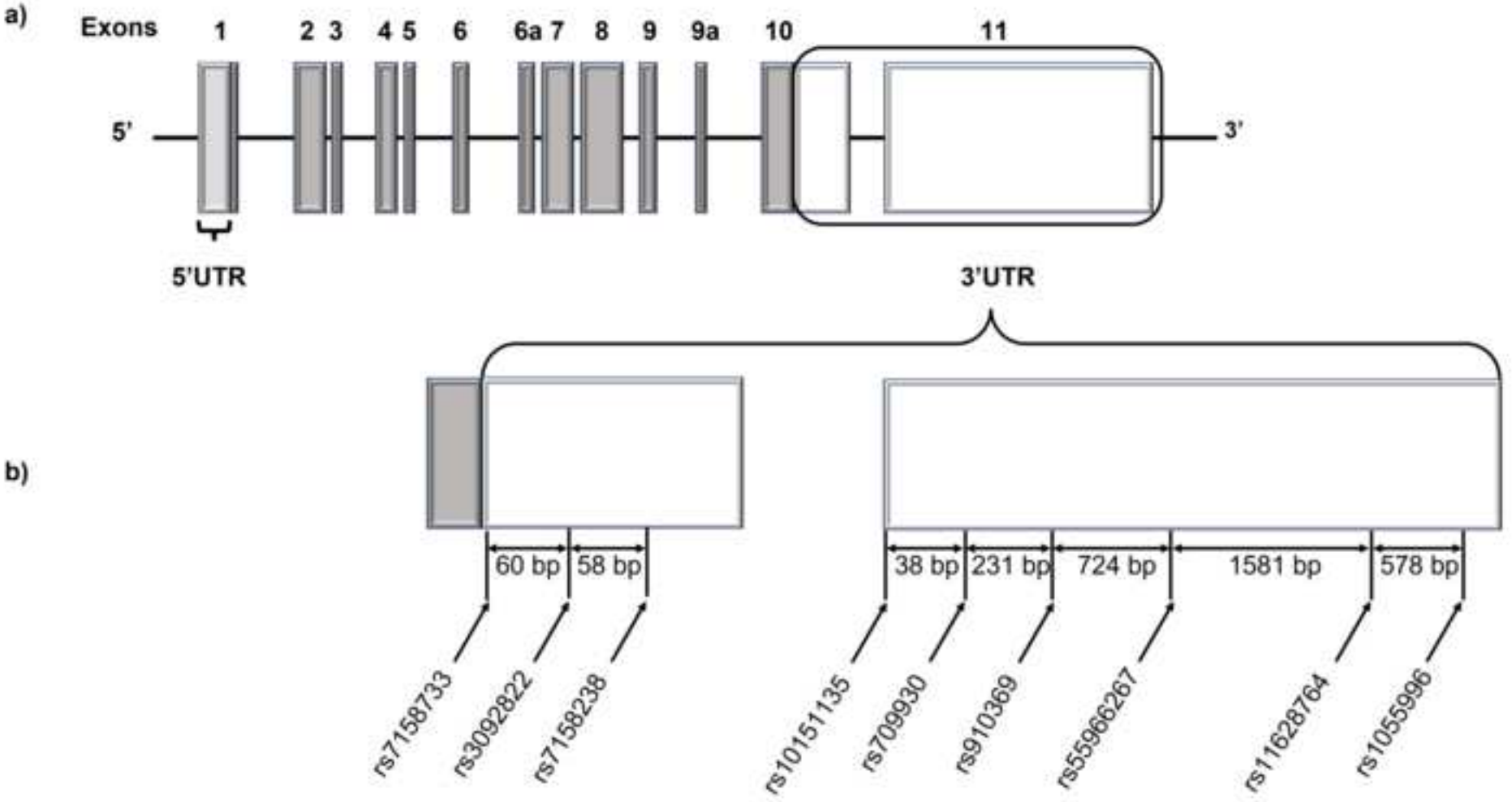
Adj. mean AO is presented as mean ± standard error. Three variants (rs10151135, rs11628764 and rs55966267) were excluded from the analysis due to the low frequency of the minor allele.

FIGURES CAPTIONS

Figure 1. Representation of the *ATXN3* gene showing the relative position (distance between the identified variants in base pairs - bp) of the nine variants identified in the 3'UTR of MJD patients. a) Exons (in medium gray) of *ATXN3* are numbered from 1 to 11. Five prime (5') and three prime (3') untranslated regions (UTRs) are in light gray and in white pattern, respectively; b) Three of the nine variants (rs7158733, rs3092822 and rs7158238) are in the initial portion of 3'UTR, in exon 10; the remaining six (rs10151135, rs709930, rs910369, rs55966267, rs11628764 and rs1055996) are in exon 11.

Figure 2. Representation of the comparison of the obtained haplotypes for both alleles: a) Description of the haplotypes obtained for the subset of samples which the haplotype for normal and expanded allele was complete; b) Frequency (absolute number) of haplotypes by normal allele (white columns) and expanded allele (grey columns) (N=36 chromosomes).

Figure



Figure

a)

Haplotype number	rs1055996	rs11628764	rs55966267	rs910369	rs709930	rs10151135	rs7158238	rs3092822	rs7158733
1	A	T	T	C	T	G	C	T	G
2	A	T	T	C	C	G	C	T	G
3	A	T	T	C	C	T	C	T	G
4	G	C	T	C	C	T	C	T	G
5	A	T	C	C	C	G	C	T	G
6	A	C	T	C	C	T	C	G	G
7	A	T	T	A	T	G	T	G	T
8	G	T	T	A	T	T	T	G	T
9	G	T	T	C	C	G	C	T	G
10	G	T	T	A	T	G	T	G	T

b)