# Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3

de Mattos, Eduardo Preusser; Musskopf, Maiara Kolbe; Leotti, Vanessa Bielefeldt; Saraiva-Pereira, Maria Luiza; Jardim, Laura Bannach

*Published in:*
Journal of Neurology, Neurosurgery, and Psychiatry

*DOI:*
10.1136/jnnp-2018-319200



*Document Version*
Publisher's PDF, also known as Version of record

*Publication date:*
2019

RESEARCH PAPER

# Genetic risk factors for modulation of age at onset in Machado-Joseph disease/spinocerebellar ataxia type 3: a systematic review and meta-analysis

Eduardo Preusser de Mattos,[1,2,3] Maiara Kolbe Musskopf,[2] Vanessa Leotti Torman,[4,5] Maria Luiza Saraiva-Pereira,[1,2,6,7] Laura Bannach Jardim[1,2,7,8]

► Additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/jnnp-2018-319200).

For numbered affiliations see end of article.

**Correspondence to**
Dr Laura Bannach Jardim, Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre, Rua Ramiro Barcelos 2350, Brazil; ljardim@hcpa.edu.br

Received 11 July 2018
Revised 24 August 2018
Accepted 11 September 2018



**To cite:** de Mattos EP, Kolbe Musskopf M, Leotti Torman V, *et al*. *J Neurol Neurosurg Psychiatry* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/jnnp-2018-319200

## ABSTRACT

**Objectives** To perform a systematic review and meta-analysis of genetic risk factors for age at onset (AO) in spinocerebellar ataxia type 3/Machado-Joseph disease (SCA3/MJD).

**Methods** Two authors independently reviewed reports on the mathematical relationship between CAG length at the expanded *ATXN3* allele (CAGexp), and other genetic variants if available, and AO. Publications from January 1994 to September 2017 in English, Portuguese or Spanish and indexed in MEDLINE (PubMed), LILACS or EMBASE were considered. Inclusion criteria were reports with >20 SCA3/MJD carriers with molecular diagnosis performed by capillary electrophoresis. Non-overlapping cohorts were determined on contact with corresponding authors. A detailed analysis protocol was registered at the PROSPERO database prior to data extraction (CRD42017073071).

**Results** Eleven studies were eligible for meta-analysis, comprising 10 individual-participant (n=2099 subjects) and two aggregated data cohorts. On average, CAGexp explained 55.2% (95% CI 50.8 to 59.0; p<0.001) of AO variability. Population-specific factors accounted for 8.3% of AO variance. Cohorts clustered into distinct geographic groups, evidencing significantly earlier AO in non-Portuguese Europeans than in Portuguese/South Brazilians with similar CAGexp lengths. Presence of intermediate *ATXN2* alleles (27–33 CAG repeats) significantly correlated with earlier AO. Familial factors accounted for ~10% of AO variability. CAGexp, origin, family effects and CAG length at *ATXN2* together explained 73.5% of AO variance.

**Conclusions** Current evidence supports genetic modulation of AO in SCA3/MJD by CAGexp, *ATXN2* and family-specific and population-specific factors. Future studies should take these into account in the search for new genetic modifiers of AO, which could be of therapeutic relevance.

## INTRODUCTION

Spinocerebellar ataxia type 3/Machado-Joseph disease (SCA3/MJD) is a neurological condition characterised by expansion of a polymorphic trinucleotide CAG tract (CAGexp) at *ATXN3*. SCA3/MJD is the most common dominantly inherited ataxia worldwide,[1 2] and *ATXN3* alleles with ≥45 repeats code for ataxin-3 proteins with abnormally long polyglutamine (polyQ) sequences. PolyQ-expanded ataxin-3 is prone to aggregation into neuronal inclusions and exerts a gain of toxic function, which leads to neuronal toxicity and degeneration,[3] similarly to what happens in Huntington's disease and other SCAs.[4]

The longer the CAGexp at *ATXN3*, the earlier the age at onset (AO) of disease. A large body of evidence has established that AO is not entirely explained by CAGexp, which explains 50% to 60% of the variability in AO,[5–8] and that AO should be modulated by additional genetic and/or environmental factors. Several candidates have been proposed, such as apolipoprotein E genotypic status,[9–11] CAG length at normal *ATXN3*[8 12 13] and *ATXN2*[6 8] alleles, and protein levels of the DNAJB1 chaperone.[14]

Most of the proposed modifiers in SCA3/MJD were not replicated or had small effects that usually improve the explanation of AO variance by not more than 1%. The greater part of the missing variability in AO remains unexplained, suggesting that the main CAGexp-independent modulators of AO are still unknown. Here, we performed a systematic review and meta-analysis of genetic risk factors associated to AO in SCA3/MJD. By analysing both aggregate and individual-participant data of more than 2000 patients from 16 countries across three continents, we were able to detect an important origin-specific effect of CAGexp on AO and to confirm the effect of some putative risk factors published previously.

## METHODS

A detailed methodology protocol for this study was registered at the PROSPERO (International Prospective Register of Systematic Reviews) database prior to data extraction and is available at https://www.crd.york.ac.uk/PROSPERO/ under record CRD42017073071.

### Literature search and data extraction

MEDLINE (PubMed), LILACS and EMBASE were searched from January 1994 to September 2017 for reports on genetic factors related to AO in SCA3/MJD. Search terms employed were 'sca3' OR 'mjd' OR 'spinocerebellar ataxia type 3' OR 'spinocerebellar ataxia type-3' OR 'machado-joseph disease' OR 'machado joseph disease' AND 'age of onset' OR 'age-of-onset' OR 'age at onset' OR 'age-at-onset'. Peer-reviewed articles and meeting abstracts were included, and references were checked to guarantee maximal coverage. Two reviewers

(EPDM and MKM) independently assessed and extracted data into evidence tables. Any disagreement regarding eligibility was discussed with a third reviewer (LBJ).

### Population, exposure, comparators, outcomes, and inclusion and exclusion criteria

SCA3/MJD heterozygotes from diverse geographical origins comprised the population under study. The CAG length at CAGexp was the main exposure considered for meta-analysis; other genetic variants were included in the meta-analysis as risk factors (exposures) if reported at least twice in literature. The outcome was the quantitative variable AO defined as the age at the first symptom, usually gait ataxia. Included studies should report on both (1) molecularly confirmed SCA3/MJD symptomatic and/or asymptomatic heterozygotes and (2) the relationship between *ATXN3* CAGexp (main exposure) and AO (outcome). The term 'carrier' was used here as a synonym for heterozygotes, symptomatic or not, with one *ATXN3* allele with ≥45 CAG repeats. We excluded studies reporting on <20 carriers, in languages other than English, Portuguese or Spanish. If multiple publications reported the same data, the most up-to-date and complete data set was included. Corresponding authors were contacted to check for duplicated data and to grant access to their updated, pseudonymised, individual-participant databases (IPDs), whenever possible. Otherwise, we used summary statistics from aggregated databases (ADs). Besides AO and CAGexp, data on gender, family, length of normal *ATXN3* CAG tracts and at other CAG-containing *loci* and/or additional genetic variants were retrieved, if available.

### Risk of bias assessment and quality control

The outcome AO was poorly defined in some studies. Since most patients with SCA3/MJD develop gait ataxia as the first symptom,[15] we combined in a single model carriers with known AO of gait ataxia (AOga) or of first symptom; when both criteria were available for the same individual, AOga was chosen. Only studies that measured CAG repeats by capillary electrophoresis were considered. Participation in molecular diagnosis quality control programmes was also questioned and informed here.

### Analysis and data synthesis

Boxplots were used to describe the variability on both AO and CAGexp among studies. The meta-analysis was composed of three main models. First, the global relationship between *ATXN3* CAGexp length and AO was investigated using data from both IPDs and ADs,[16] aiming at comparing the degree of explanation of the variability in AO by CAGexp across studies, as reported by the linear $R^2$ measure. A second meta-analysis used IPDs only. Since complex models (quadratic and logarithmic) were only marginally better at explaining the data (see online supplementary file 1), AO was not mathematically transformed, and linear regression was used. A third analysis tested the effects of gender, family and CAG length at the non-expanded *ATXN3* allele and at other CAG-containing *loci*, focusing on the improvement of the $R^2$ measure. Geographical origin and interaction between origin and CAGexp were always included as independent variables. With the exception of *ATXN1*, which was considered a continuous variable, the effect of all CAG-containing *loci* was assessed as both continuous and discrete variables using CAG length groups as published previously[6 8] (online supplementary file 2). The percentage of AO variability explained by belonging to the same family was tested with a fixed-effects model. Analyses were performed using the software R V.3.4.1 with packages lsmeans and lmSupport, and SAS OnDemand for Academics V.3.1 (SAS Institute). Graphs were generated with ggplot2. Results were considered statistically significant when $p<0.05$.

## RESULTS

### Systematic review

The search yielded 641 unique abstracts (online supplementary file 3); 140 studies testing the relationship between AO and CAGexp at *ATXN3* were selected for the systematic review (online supplementary file 4). Thirty-one studies investigated additional modifying effects on AO, including CAG repeat length at the non-expanded *ATXN3* (n=19), *ATXN1* (n=5), *ATXN2* (n=6), *CACNA1A* (n=7), *ATXN7* (n=4), *HTT* (n=4), *TBP* (n=3) and *ATN1* (n=4) alleles. AO differences according to length of GGGGCC repeats at *C9ORF72*, and CAG repeats at *RAI1* and *KCNN3* were each reported once. Another report correlated ataxin-3 and selected chaperones protein levels with AO. Allelic and/or genotypic status of single-nucleotide polymorphisms at 15 genes were also correlated with AO, including variants at *ATXN3* (rs3814834, rs709930 and rs910369; n=1 each), *APOE* (rs429358 and rs7412; n=4) and *ATXN2* (rs7969300), *BDNF* (rs6265), *BECN1* (rs60221525 and rs116943570), *CHIP* (rs6597), *hCAD* (rs12738235), *IL1A* (rs1800587), *IL1B* (rs16944), *IL6* (rs1800795), *MT-ND3* (rs2853826), *OGG1* (rs1052133), *PPARGC1A* (rs7665116), *TNF* (rs1800629) and *UCHL1* (rs5030732; n=1 each). Differences in AO according to the degree of promoter methylation at *ATXN3* were evaluated by two studies, using distinct methodologies. Gender of the affected individual and transmitting parent were correlated with AO in 14 and 6 studies, respectively. Two reports considered the effect of population of origin on AO, and one evaluated the familial dependency of AO. Data extraction, references and detailed information of all AO modifiers reported in the literature, including those not selected for meta-analysis, are described in online supplementary file 5.

After contacting all corresponding authors of studies that met the inclusion and exclusion criteria (n=11), we retrieved updated information on 10 non-overlapping IPDs and 2 ADs of symptomatic individuals only (figure 1). CAGexp and geographical origin were available for all IPDs. Additional data included length of non-expanded CAG tracts at *ATXN3* (n=9 cohorts), and at *ATXN1*, *ATXN2*, *CACNA1A* and *ATXN7* (n=4 cohorts). Information on gender and family effects were available for six and seven cohorts, respectively. Geographical origin, sample sizes and retrieved data for each cohort included in the meta-analysis are summarised in table 1 and detailed in online supplementary file 6.

### Effect of CAGexp and geographical origin

Exposure to diverse CAGexp repeat lengths at *ATXN3* was the most studied risk factor. IPD and AD retrieved from 11 studies comprised four cohorts from Europe,[6 7 17 18] three from Asia,[8 19 20] one from North America,[6] one from Central America[21] and three from Brazil.[5 22 23] Brazilian cohorts comprised the Rio Grande do Sul (Brazil-RS) cohort[23] and cohorts from other Brazilian regions (Brazil-non-RS cohorts): namely, subjects from São Paulo State[22] and those described by Neurogenetics Network, a consortium of Brazilian researchers.[5] Using both IPDs and ADs, the global linear correlation coefficient between CAGexp and $\log_{10}$(AO) was $r=-0.743$ (95% CI -0.768 to -0.713, $p<0.001$), meaning that, on average, the causative mutation determines about 55.2% (50.8%–59.0%) of the AO variability in SCA3/MJD worldwide.

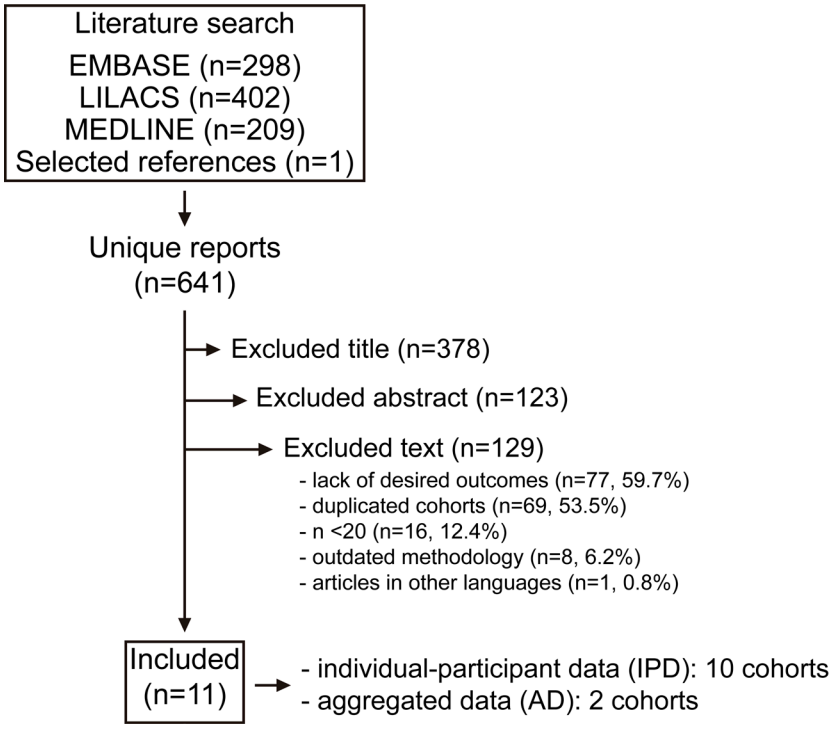


**Figure 1** Workflow of the studies selected for the present meta-analysis.

Subsequent analyses used IPD cohorts only, totaling 2099 patients. Variability of CAGexp length among the 10 cohorts (figure 2A) was wider than variability in AO (figure 2B). Inclusion of geographical origin increased in 8.34% the explanation of AO variability (adjusted $R^2$=0.556; $F_{10,2091}$=263.8, p<0.001; online supplementary file 1). CAGexp significantly interacted with origin, which improved the model by an additional 1.02% (adjusted $R^2$=0.564; $F_{19,2082}$=144.1, p<0.001). The differential effect of CAGexp on AO among the 10 cohorts was evidenced by differences in slope and position of regression lines (figure 3A and B). Pairwise analysis of cohorts with similar slopes and/or intercepts allowed for data aggregation into three main

**Table 1** Studies selected for meta-analysis of the modulation of age at onset in SCA3/MJD by CAG repeat length at *ATXN3* and additional factors

| Cohort | N, published* | N, available† | Data type | QC programme‡ | Available data (risk factors) | Reference |
|---|---|---|---|---|---|---|
| Rio Grande do Sul, Brazil | 463 | 507 | IPD | Yes | Family, gender, origin, *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A*, *ATXN7* | Souza *et al*[23] 2016 |
| EUROSCA, Europe | 403 | 403 | IPD | Yes | Family, gender, origin, *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A*, *ATXN7* | Tezenas du Montcel *et al*[6] 2014 |
| Taiwan | 48 | 347 | IPD | NS | Gender, origin, *ATXN3* | Wang *et al*[19] 2012 |
| Portugal (Mainland) | 48 | 226 | IPD | Yes | Origin, *ATXN3* | Silveira *et al*[17] 1998 |
| China | 141 | 141 | IPD | NS | Family, origin, *ATXN3* (expanded allele only) | Wang *et al*[20] 2017 |
| Netherlands | 342§ | 133 | IPD | Yes | Family, origin, *ATXN3* | van de Warrenburg *et al*[18] 2005 |
| São Paulo, Brazil | 34 | 110 | IPD | No | Origin, *ATXN3* (expanded allele only) | França Jr *et al*[22] 2009 |
| Portugal (Azorean Islands) | 93 | 106 | IPD | Yes | Family, gender, origin, *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A*, *ATXN7* | Raposo *et al*[7] 2015 |
| Neurogenetics network, Brazil | 481¶ | 104 | IPD | Yes | Family, gender, origin, *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A*, *ATXN7* | de Castilhos *et al*[5] 2014 |
| Cuba | 22 | 22 | IPD | Yes | Origin, *ATXN3* | González-Zaldívar *et al*[21] 2015 |
| China | 802 | | AD | NS | Origin, *ATXN3* (expanded allele only) | Chen *et al*[8] 2016 |
| USA | 110 | | AD | NS | Origin, *ATXN3* (expanded allele only) | Tezenas du Montcel *et al*[6] 2014 |

*Number of patients reported on the original publication.
†Number of patients whose data were obtained after contacting the corresponding author of the selected publication.
‡Participation in molecular diagnosis quality control programme.
§This study reports on Dutch and French patients, but only the Dutch cohort was available for analysis.
¶This study includes patients from the Rio Grande do Sul cohort, but only patients from other Brazilian regions were retrieved for analysis.
AD, aggregated database; EUROSCA, European Consortium on Spinocerebellar Ataxias; IPD, individual-participant database; NS, not specified; SCA3/MJD, spinocerebellar ataxia type 3/Machado-Joseph disease.

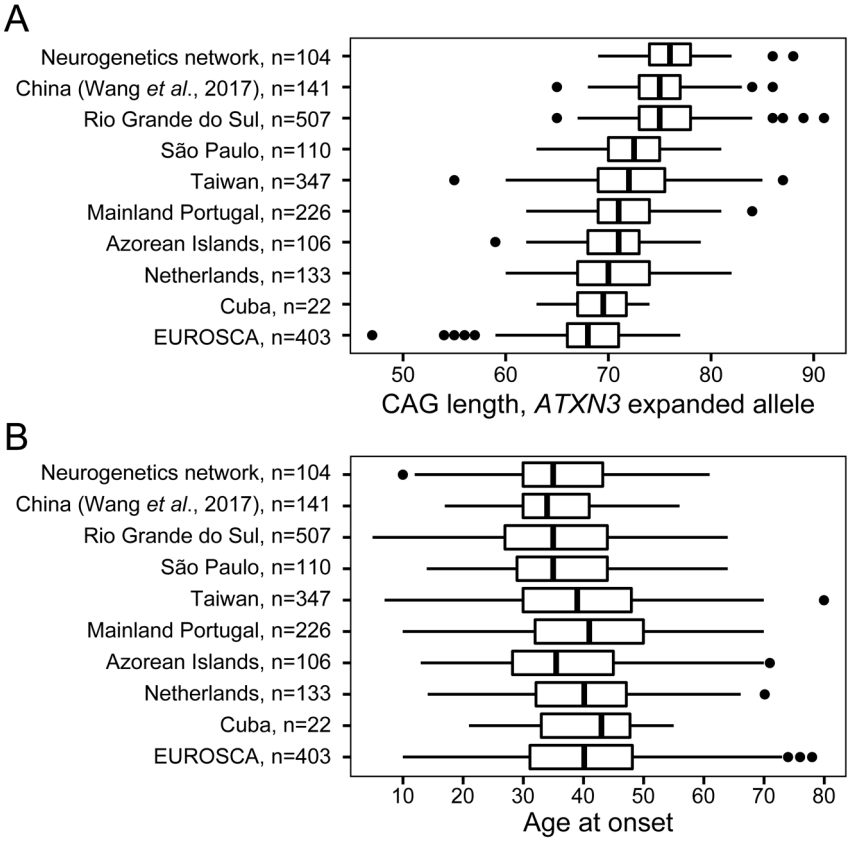


**Figure 2** Variability of CAG length at the expanded ATXN3 allele and age at onset in SCA3/MJD. Data on CAG length at the expanded *ATXN3* allele and age at onset are shown for the 10 original, individual-participant data cohorts selected for meta-analysis. EUROSCA, European Consortium on Spinocerebellar Ataxias; SCA3/MJD, spinocerebellar ataxia type 3/Machado-Joseph disease.

geographical/ethnic groups with differential CAGexp modulation of AO: an average group with heterogeneous origins (China, Cuba, Brazil-non-RS and Taiwan cohorts), the group of non-Portuguese Europe (EUROSCA and Netherlands cohorts) and the group with clear Portuguese origin (Azorean Islands, mainland Portugal and Brazil-RS cohorts) (figure 3C and D; online supplementary file 7). Table 2 presents mean AO predictions for each geographically distinct group as a function of CAGexp tracts of three different lengths.

## Effect of non-expanded CAG repeats at *ATXN3* and other CAG-containing *loci*

Data on CAG length of non-expanded *ATXN3*, *ATXN1*, *ATXN2*, *CACNA1A* and *ATXN7* alleles were available from 944 patients from four cohorts (EUROSCA, Azorean Islands, Brazil-RS and Brazil-non-RS) for meta-analysis. Inclusion of CAG length at the non-expanded *ATXN3* allele did not significantly improve the correlation between CAGexp and AO (p=0.327, continuous variable; p=0.388, discrete variable; online supplementary file 1). From the remaining candidate *loci*, only *ATXN2* significantly improved the explanation of AO variability, with longer *ATXN2* CAG tracts correlating with earlier AO (adjusted $R^2$=0.630; $F_{10,933}$=161.6, p<0.001; online supplementary file 1). There was a significant interaction between length of the longest CAG tract at *ATXN2* and CAGexp, which contributed an additional 0.39% to the explanation of AO variability (p=0.020; online supplementary file 1). Presence of at least one intermediate *ATXN2* allele (27–33 CAGs; 5% of alleles) significantly correlated with earlier AO (adjusted $R^2$=0.632; $F_{13,930}$=125.7, p<0.001; figure 4A and online supplementary file 1) in individuals with CAGexp tracts of up to 73 repeats (table 3).

## Family and gender effects

Information on family effects was available for 1368 patients from 565 families (online supplementary file 8). Among these, CAGexp and origin alone explained ~60% of AO variance (adjusted $R^2$=0.599; $F_{11,1356}$=186.8; p<0.001). Inclusion of family data in a fixed-effects model increased the explanation by ~10% (adjusted $R^2$=0.702; $F_{888,479}$=4.6; p<0.001; online supplementary file 1). Data on gender were available for 1468 patients, and its inclusion contributed an additional 0.3% increase in the explanation of variability in AO (adjusted $R^2$=0.590; $F_{10,1457}$=211.9; p<0.001; online supplementary file 1). On average, male patients had younger ages at onset, especially among individuals with longer CAGexp tracts (figure 4B and online supplementary file 9). However, when considered together with CAG repeat length at *ATXN2* (n=942 individuals), the effect of gender was not significant (p=0.08; online supplementary file 1).

## Combined effects

The final and best regression model considered CAGexp, origin, family effects and *ATXN2* genotypes, and explained 73.5% (95% CI 68.2 to 77.6) of the AO variance (adjusted $R^2$=0.735; $F_{682,245}$=4.8; p<0.001; online supplementary file 1).

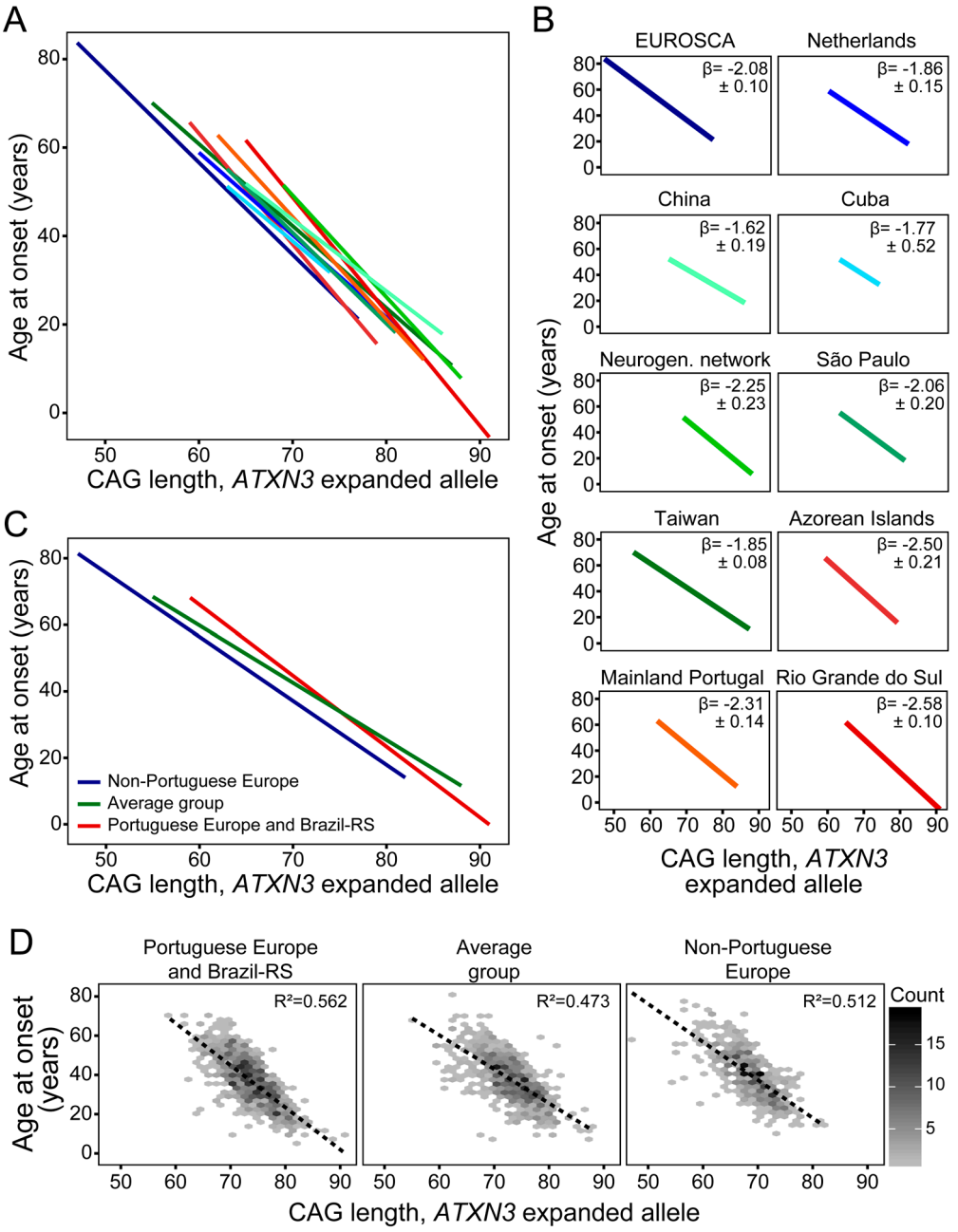


**Figure 3** Population-specific modulation of the age at onset (AO) in spinocerebellar ataxia type 3/Machado-Joseph disease. (A) Linear regression of CAGexp and AO in the 10 distinct cohorts included in the present meta-analysis (cohorts are colour-coded as in panel B). (B) Same as in (A), but with individual plots for each cohort. The β coefficient of the linear regression, expressed as mean±SE, is also shown for each cohort. (C) Same as in (A), but with patients grouped in three main geographical origins, as a result of pairwise comparisons of intercepts and slopes of lines presented in (A). (D) Same as in (C), but with geographical origins depicted separately. Hexagons depict groups of patients, according to the colour scheme, and dashed lines represent linear regressions. For each plot, the $R^2$ correlation coefficient is also shown. EUROSCA, European Consortium on Spinocerebellar Ataxias.

## DISCUSSION

This worldwide systematic investigation of risk factors for AO in SCA3/MJD detected that the CAGexp determines, on average, 55.2% of the phenotypic variability in AO. Additional modulation of AO by family factors, gender and CAG length at *ATXN2* were confirmed. Interestingly, currently unknown effectors related to geographical origin were also shown to modify AO. Although several more candidates have already been proposed, data were not robust enough to be meta-analysed, and further replication studies are necessary to assess their validity as phenotypic modulators in SCA3/MJD.

### Effect of CAGexp and geographical/ethnic and family background

Clear geographical/ethnic differences on the effect of CAGexp on AO tell us that a universal correlation might not applicable to all carrier populations. Choice of statistical modelling might further evidence how populational differences in CAGexp can impact AO determination. Significant increase in explanation of AO variability was detected in Han Chinese,[8] European carriers from non-Portuguese populations and Americans[6] using quadratic models. Here, the quadratic modelling of CAGexp

**Table 2** Effect of population of origin on the age at onset (AO) in SCA3/MJD

| Parameter* | Portuguese Europe and Brazil-RS | Average group | Non-Portuguese Europe |
|---|---|---|---|
| Intercept | 193.95 (184.36–203.55)[a] | 163.12 (153.71–172.54)[b] | 171.83 (160.78–182.88)[b] |
| Slope | −2.13 (−2.26 to −2.00)[a] | −1.72 (−1.85 to −1.59)[b,c] | −1.92 (−2.08 to −1.76)[a,c] |
| Predicted age at onset (years) | | | |
| For CAGexp=65 | 55.36 (54.10–56.62)[a] | 51.20 (50.00–52.39)[b] | 46.75 (45.84–47.66)[c] |
| For CAGexp=75 | 34.04 (33.45–34.62)[a] | 33.98 (33.32–34.63)[a] | 27.51 (26.27–28.75)[b] |
| For CAGexp=85 | 12.71 (11.14–14.29)[a] | 16.76 (15.10–18.41)[b] | 8.27 (5.55–10.99)[c] |

Patients with SCA3/MJD from 10 cohorts with distinct geographical origins were grouped according to similarity of linear regression parameters. Mean AOs are presented for expanded CAG tracts at *ATXN3* (CAGexp) of different lengths.

*Data are presented as mean (95% CI). For each parameter, means sharing the same letter are not statistically different (Tukey-adjusted comparisons).

†

CAGexp, CAG length at the expanded ATXN3 allele; SCA3/MJD, spinocerebellar ataxia type 3/Machado-Joseph disease.

from IPDs yielded only a marginal improvement when compared with a simpler, linear regression modelling of AO variance. This is likely attributed to presence of individuals with larger CAGexp tracts (figure 2A), which correlate more strongly with AO,[7] compared with previous publications.[6 8]

Variation of CAGexp distribution was markedly larger than variation of AO among populations (figure 2). SCA3/MJD populations with larger CAGexp belonged to Brazilian and Asian cohorts. Inversely, subjects from Austria, Belgium, France, Germany, Hungary, Italy, Netherlands, Poland, Spain and UK (EUROSCA cohort)[6] had the shortest mean CAGexp. Reasons for such differences are still unknown. Although ascertainment bias usually operates in favour of recruiting more severe cases (ie, longer CAGexp tracts), this bias was unlikely in at least one cohort (Brazil-RS) with large CAGexp tracts since coverage in this population was shown to be very high.[23] Substantial differences in CAGexp determination are also unlikely since most included studies were performed in laboratories engaged in molecular diagnosis quality control programmes. Therefore, distinct CAGexp patterns likely represent true differences related to population of origin.

Pairwise comparisons allowed us to categorise carriers into three main geographical/ethnic groups, reflecting distinct relationships between CAGexp and AO (figure 3 and table 2), and suggesting that CAGexp does not have the same effect on AO of all SCA3/MJD carriers worldwide. Assuming that the 'average' group (figure 3C,D) represents the worldwide average relationship between CAGexp and AO in SCA3/MJD, our analysis suggests the existence of AO modifiers with opposing effects on non-Portuguese European carriers versus subjects with Portuguese ancestry (mainland Portuguese, Azorean and South Brazilians). There seems to be factors among non-Portuguese Europeans and carriers of Portuguese ancestry that effectively predispose to earlier and later AO, respectively, given a CAGexp of same length. It is also possible that the geographical/ethnic effect uncovered here reflects, at least partially, distinct *ATXN3* haplotypes and mutational origins, as different SCA3/MJD populations show distinct haplotypic frequencies.[24] Further research will be necessary to establish a causal link, if any, between CAGexp haplotypes and AO.

Familial effects might also be due to genetic AO modifiers, although the effect of shared environmental exposures within a family cannot be excluded. A significant decrease in residual AO variance within families, compared with that between families, was observed previously.[18 25] The ~10% improvement in $R^2$ observed here was smaller than the 25% observed in a French and Dutch cohort[18]; whether this was due to presence of several small families with one or two individuals in the meta-analysis remains to be established (online supplementary file 8).

## Effect of the non-expanded *ATXN3* allele and of other non-expanded CAG-containing *loci*

In agreement with most original studies, there was no association between length of the non-expanded *ATXN3* allele and AO (online supplementary files 1 and 5) and that was also the case

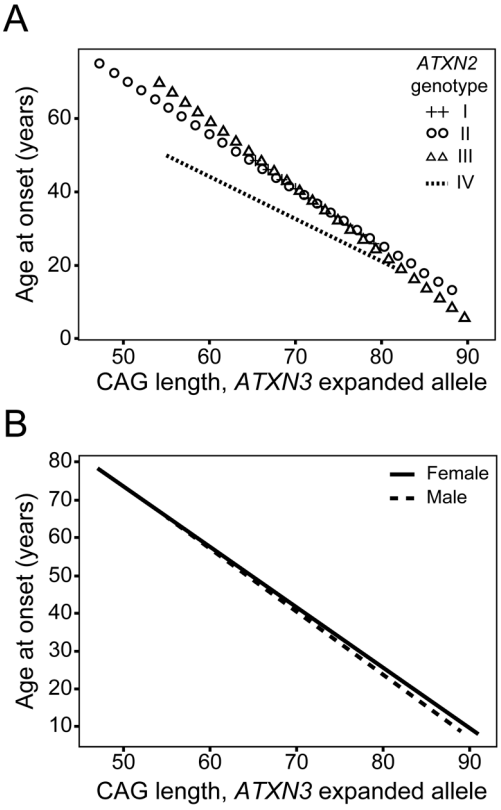


**Figure 4** Effect of co-modifiers of the age at onset (AO) in SCA3/MJD. (A) Linear regression of CAG repeat length at the expanded *ATXN3* allele (CAGexp) and AO in patients with spinocerebellar ataxia type 3/Machado-Joseph disease, divided into four categories according to length of CAG tracts at both *ATXN2* alleles. I, at least one short (<22) allele; II, homozygous medium (22) alleles; III, at least one short intermediate allele (23–26) with or without a medium (=22) allele; IV, at least one intermediate (27–33) allele. (B) Same as in (A) but with individuals divided according to gender.

**Table 3** Effect of *ATXN2* genotypes on the age at onset (AO) in SCA3/MJD

| | *ATXN2* genotype* | | | |
|---|---|---|---|---|
| **Parameter** | **I** | **II** | **III** | **IV** |
| Individuals, n (%) | 32 (3.4) | 634 (67.2) | 47 (5.0) | 231 (24.4) |
| Intercept† | 228.03 (179.66–276.40)[a,b] | 213.00 (199.41–226.60)[a] | 213.85 (199.02–228.68)[a] | 169.38 (142.50–196.27)[b] |
| Slope† | −2.64 (−3.31 to −1.97)[a,b] | −2.42 (−2.60 to −2.23)[a] | −2.43 (−2.64 to 2.23)[a] | −1.90 (−2.28 to −1.52)[b] |
| Predicted age at onset† | | | | |
| For CAGexp=65 | 56.54 (51.21–61.87)[a] | 55.69 (53.94–57.43)[a] | 55.61 (53.80–57.43)[a] | 45.64 (42.62–48.65)[b] |
| For CAGexp=70 | 43.35 (40.50–46.20)[a] | 43.59 (42.54–44.64)[a] | 43.44 (42.30–44.58)[a] | 36.12 (33.90–38.33)[b] |
| For CAGexp=73 | 35.43 (32.83–38.03)[a] | 36.33 (35.45–37.20)[a] | 36.14 (35.08–37.20)[a] | 30.41 (27.98–32.83)[b] |
| For CAGexp=74 | 32.80 (29.95–35.64)[a,b] | 33.90 (33.02–34.79)[a] | 33.70 (32.59–34.81)[a] | 28.50 (25.90–31.11)[b] |
| For CAGexp=75 | 30.16 (26.94–33.37)[a,b] | 31.48 (30.55–32.42)[a] | 31.27 (30.08–32.46)[a] | 26.60 (23.78–29.42)[b] |
| For CAGexp=85 | 3.77 (−5.30 to 12.84)[a] | 7.28 (4.92–9.64)[a] | 6.92 (4.11–9.74)[a] | 7.56 (1.58–13.55)[a] |

*ATXN2* genotypes (*ATXN2*gen) were included in a linear regression model of AO as a function of CAG length at the expanded *ATXN3* allele (CAGexp) and geographical origin as follows: AO ~CAGexp+Origin+CAGexp*Origin+*ATXN2* gen+*ATXN2*gen*CAGexp. Mean AOs are presented for expanded CAG tracts at *ATXN3* of different lengths.
CAGexp, CAG length at the expanded *ATXN3* allele; SCA3/MJD, spinocerebellar ataxia type 3/Machado-Joseph disease.
**ATXN2* genotypes were defined as follows: (I) at least one short (<22) allele, (II) homozygous medium (=22) alleles, (III) at least one short intermediate (23–26) allele with or without a medium allele, and (IV) at least one intermediate (27–33) allele. †Data are presented as mean (95% CI). For each parameter, means sharing the same letter are not statistically different (Tukey-adjusted comparisons).
[†]Data are presented as mean (95% CI). For each parameter, means sharing the same letter are not statistically different (Tukey-adjusted comparisons).

for *ATXN1*, *ATXN7* and *CACNA1A*. However, since we did not have access to IPD from the large Chinese cohort that reported the associations between AO and *CACNA1A* and *ATXN7*,[8] population-specific differences in the range of CAG tracts at these loci—and in power to detect their potential effects—should not be overruled.

In contrast, we confirmed the association between non-expanded *ATXN2* alleles of intermediate CAG length (27–33 repeats) and earlier AO in SCA3/MJD (figure 4A, table 3 and online supplementary file 1), as reported previously.[6 8] Lack of confirmation in other cohorts is most likely attributed to small sample sizes[6 7] or inclusion of *ATXN2* in regression analysis as a continuous variable.[5] Whether the modulatory effect of *ATXN2* would be due to the CAG tract directly, or another genetically linked variant, is still unknown. Several observations support a biologically significant role for *ATXN2* and the normal ataxin-2 protein in neurodegenerative diseases. For instance, longer non-expanded *ATXN2* alleles have been related to increased risk of developing amyotrophic lateral sclerosis,[26] progressive supranuclear palsy,[27] frontotemporal dementia[28] and multiple systems atrophy.[29] Outside the CAG tract, a correlation between a missense polymorphism at *ATXN2* and earlier AO in Chinese patients with SCA3/MJD was recently shown.[30] Moreover, lower ataxin-2 levels were detected in brains of patients with SCA3/MJD and transgenic mice compared with healthy controls.[31] Importantly, restoration of ataxin-2 levels in affected mice led to significant morphological and behavioural improvements.[31] Therefore, the current evidence suggests that ataxin-2 is a strong candidate modifier of AO (and, maybe, disease progression) in SCA3/MJD.

## Study limitations

Although great care was taken to control for potential biases and confounding factors, the present study is not without methodological limitations. Importantly, due to its retrospective assessment, it is possible that AO was not precisely defined for some of the individuals included. However, recalling biases were likely present in all patient cohorts, thus arguing in favour of true differences in AO among carriers from distinct populations/ethnicities. Moreover, different studies selected for meta-analysis had distinct definitions of AO, namely AO of the first symptom or AO of gait ataxia. Even though gait ataxia is usually the first symptomatic manifestation of SCA3/MJD, other symptoms might present before gait abnormalities.[15] While some of the largest patient cohorts included in this study had gait ataxia clearly stated as the parameter of choice for AO, which might have contributed to reduce AO heterogeneity, it is possible that distinct AO parameters are differentially modulated by CAGexp and/or other genetic factors.

## Concluding remarks

The present analysis estimated that CAGexp is globally responsible for 55.2% of AO variance, on average. Gender and shared familial characteristics (most likely genetic) were confirmed as factors that influence AO in SCA3/MJD. Among candidate genes, CAG length at *ATXN2* was the only variant confirmed by the meta-analysis; future studies on the ataxin-3/ataxin-2 interactions might disclose promising discoveries.

Moreover, the IPD meta-analysis suggested protective factors in SCA3/MJD geographical groups of Portuguese origin, and probably a lack of some protective factors in non-Portuguese Europeans. Studying selected SCA3/MJD carrier groups, such as cohorts from specific geographical origins, or families with disease onset markedly different from the expected AO for their location, could significantly boost the search for genetic AO modulators.

The best model to assess the effect of the confirmed independent variables on AO determination in SCA3/MJD included CAGexp, geographical origin, family and CAG length at *ATXN2*: this model explained 73.5% of AO variability. That does not mean that factors responsible for the remaining variance should not have a genetic nature as well. In fact, several studies reviewed here assessed the effect of other genetic variants on AO, and some had promising results. Unfortunately, most were unique studies that were not qualified for meta-analysis. However, one of the advantages of meta-analyses is that updates can be performed in the future. Hopefully, further evidences on modifiers could increase the explanation of AO variability in SCA3/MJD, disclosing factors with potential therapeutic roles.

**Author affiliations**
[1]Programa de Pós-Graduação em Genética e Biologia Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[2]Laboratório de Identificação Genética, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
[3]Department of Biomedical Sciences of Cells & Systems, Section of Molecular Cell Biology, University Medical Center Groningen/Groningen University, Groningen, The Netherlands
[4]Departamento de Estatística, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[5]Programa de Pós-Graduação em Epidemiologia, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[6]Departamento de Bioquímica, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil
[7]Serviço de Genética Médica, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
[8]Departamento de Medicina Interna, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

**Acknowledgements** This work would not be possible without the collaborative efforts of many people. We would like to thank all patients and corresponding authors of selected studies for sharing their data with us. This work was supported by the following Brazilian agencies: CAPES—Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Project 99999.01528/2015-2017; CNPq—Conselho Nacional de Desenvolvimento Científico e Tecnológico, Project 402968/2012-3; FIPE-HCPA—Fundo de Incentivo à Pesquisa do Hospital de Clínicasde Porto Alegre, Projects GPPG HCPA 13-0303 and 14-0204. EPDM, MLS-P and LBJ were supported by CNPq.

**Contributors** EPDM, MLS-P and LBJ designed the study. EPDM, MKM and LBJ performed the systematic review. EPDM and VBL performed the statistical analysis. EPDM, MKM, VBL, MLS-P and LBJ wrote the manuscript.

**Funding** This study was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (grant no. 402968/2012-3), Fundo de Incentivo à Pesquisa do Hospital de Clínicas de Porto Alegre (grant no. GPPG HCPA 13-0303GPPG HCPA 14-0204) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (grant no. 99999.01528/2015-2017).

**Disclaimer** Funding organisations had no role in study design; collection, analysis and interpretation of data; writing of the report; or decision to submit the article for publication.

**Competing interests** None declared.

**Patient consent** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

## REFERENCES

1 Schöls L, Bauer P, Schmidt T, *et al*. Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis. *Lancet Neurol* 2004;3:291–304.
2 Sequeiros J, Martins S, Silveira I. Epidemiology and population genetics of degenerative ataxias. *Handb Clin Neurol* 2012;103:227–51.
3 Matilla-Dueñas A, Ashizawa T, Brice A, *et al*. Consensus paper: pathological mechanisms underlying neurodegeneration in spinocerebellar ataxias. *Cerebellum* 2014;13:269–302.
4 Kuiper EFE, de Mattos EP, Jardim LB, *et al*. Chaperones in polyglutamine aggregation: beyond the q-stretch. *Front Neurosci* 2017;11:1–11.
5 de Castilhos RM, Furtado GV, Gheno TC, *et al*. Spinocerebellar ataxias in Brazil—frequencies and modulating effects of related genes. *Cerebellum* 2014;13:17–28.
6 Tezenas du Montcel S, Durr A, Bauer P, *et al*. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain* 2014;137(Pt 9):2444–55.
7 Raposo M, Ramos A, Bettencourt C, *et al*. Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid? *Brain* 2015;138(Pt 12):e398.
8 Chen Z, Zheng C, Long Z, *et al*. (CAG)n loci as genetic modifiers of age-at-onset in patients with Machado-Joseph disease from mainland China. *Brain* 2016;139(Pt 8):e41.
9 Bettencourt C, Raposo M, Kazachkova N, *et al*. The APOE ε2 allele increases the risk of earlier age at onset in Machado-Joseph disease. *Arch Neurol* 2011;68:1580–3.
10 Peng H, Wang C, Chen Z, *et al*. APOE ε2 allele may decrease the age at onset in patients with spinocerebellar ataxia type 3 or Machado-Joseph disease from the Chinese Han population. *Neurobiol Aging* 2014;35:2179.e15–2179.e18.
11 Zhou Q, Ni W, Dong Y, *et al*. The role of apolipoprotein E as a risk factor for an earlier age at onset for Machado-Joseph disease is doubtful. *PLoS One* 2014;9:e111356.
12 Dürr A, Stevanin G, Cancel G, *et al*. Spinocerebellar ataxia 3 and Machado-Joseph disease: clinical, molecular, and neuropathological features. *Ann Neurol* 1996;39:490–9.
13 França MC, Emmel VE, D'Abreu A, *et al*. Normal ATXN3 allele but not CHIP polymorphisms modulates age at onset in Machado-Joseph disease. *Front Neurol* 2012;3:164.
14 Zijlstra MP, Rujano MA, Van Waarde MA, *et al*. Levels of DNAJB family members (HSP40) correlate with disease onset in patients with spinocerebellar ataxia type 3. *Eur J Neurosci* 2010;32:760–70.
15 Globas C, du Montcel ST, Baliko L, *et al*. Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. *Mov Disord* 2008;23:2232–8.
16 Pigott TD. *Advances in meta-analysis*. Boston, MA: Springer US, 2012.
17 Silveira I, Coutinho P, Maciel P, *et al*. Analysis of SCA1, DRPLA, MJD, SCA2, and SCA6 CAG repeats in 48 Portuguese ataxia families. *Am J Med Genet* 1998;81:134–8.
18 van de Warrenburg BP, Hendriks H, Dürr A, *et al*. Age at onset variance analysis in spinocerebellar ataxias: a study in a Dutch-French cohort. *Ann Neurol* 2005;57:505–12.
19 Wang PS, Chen HC, Wu HM, *et al*. Association between proton magnetic resonance spectroscopy measurements and CAG repeat number in patients with spinocerebellar ataxias 2, 3, or 6. *PLoS One* 2012;7:e47479.
20 Wang C, Peng H, Li J, *et al*. Alteration of methylation status in the ATXN3 gene promoter region is linked to the SCA3/MJD. *Neurobiol Aging* 2017;53:192.e5–192.e10.
21 González-Zaldívar Y, Vázquez-Mojena Y, Laffita-Mesa JM, *et al*. Epidemiological, clinical, and molecular characterization of Cuban families with spinocerebellar ataxia type 3/Machado-Joseph disease. *Cerebellum Ataxias* 2015;2:1.
22 França MC, D'Abreu A, Nucci A, *et al*. Progression of ataxia in patients with Machado-Joseph disease. *Mov Disord* 2009;24:1387–90.
23 Souza GN, Kersting N, Krum-Santos AC, *et al*. Spinocerebellar ataxia type 3/Machado-Joseph disease: segregation patterns and factors influencing instability of expanded CAG transmissions. *Clin Genet* 2016;90:134–40.
24 Martins S, Calafell F, Gaspar C, *et al*. Asian origin for the worldwide-spread mutational event in Machado-Joseph disease. *Arch Neurol* 2007;64:1502–8.
25 DeStefano AL, Cupples LA, Maciel P, *et al*. A familial factor independent of CAG repeat length influences age at onset of Machado-Joseph disease. *Am J Hum Genet* 1996;59:119–27.
26 Elden AC, Kim HJ, Hart MP, *et al*. Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature* 2010;466:1069–75.
27 Ross OA, Rutherford NJ, Baker M, *et al*. Ataxin-2 repeat-length variation and neurodegeneration. *Hum Mol Genet* 2011;20:3207–12.
28 Lattante S, Millecamps S, Stevanin G, *et al*. Contribution of ATXN2 intermediary polyQ expansions in a spectrum of neurodegenerative disorders. *Neurology* 2014;83:990–5.
29 Zhou X, Wang C, Ding D, *et al*. Analysis of $(CAG)_n$ expansion in ATXN1, ATXN2 and ATXN3 in Chinese patients with multiple system atrophy. *Sci Rep* 2018;8:3889.
30 Ding D, Li K, Wang C, *et al*. ATXN2 polymorphism modulates age at onset in Machado-Joseph disease. *Brain* 2016;139:aww176.
31 Nóbrega C, Carmo-Silva S, Albuquerque D, *et al*. Re-establishing ataxin-2 downregulates translation of mutant ataxin-3 and alleviates Machado-Joseph disease. *Brain* 2015;138(Pt 12):3537–54.