Age of onset in Huntington disease is influenced by CAG repeat variations in other polyglutamine disease-associated genes

Authors & affiliations:

Geerte Stuitje, MSc[1,2]; Martine J. van Belzen, PhD[2]; Sarah L. Gardiner, MD[1,3]; Willeke M.C. van Roon-Mom, PhD[3]; Merel W. Boogaard[2]; REGISTRY Investigators of the European Huntington Disease Network[4]; Sarah J. Tabrizi, MD, PhD[5]; Raymund A.C. Roos, MD, PhD[1]; N. Ahmad Aziz, MD, PhD[1,5]

Departments of [1]Neurology, [2]Clinical Genetics and [3]Human Genetics, Leiden University Medical Centre, Leiden, the Netherlands

[4]Membership of the REGISTRY Investigators of the European Huntington's Disease Network is provided in the Acknowledgments

[5]Department of Neurodegenerative Disease, UCL Huntington's Disease Centre, University College London Institute of Neurology, London, United Kingdom.

Corresponding author:

N.A. Aziz, MD, PhD

Department of Neurology, Leiden University Medical Centre, Leiden, the Netherlands & Department of Neurodegenerative Disease, UCL Huntington's Disease Centre, University College London Institute of Neurology, London, United Kingdom

Tel.: (+31) (0) 71 526 2197

Fax: (+31) (0) 71 524 8253

E-mail: N.A.Aziz@lumc.nl

Sir,

We read with great interest the recent article by Tezenas du Montcel et al. who showed that the age of onset in several spinocerebellar ataxias (SCAs) is modulated by CAG repeat sizes in the normal range in other polyglutamine disease-associated genes (Tezenas du *et al.*, 2014). Interestingly, the age of onset in patients with SCA3 was also influenced by the CAG repeat size in the *HTT* gene: Long normal *HTT* CAG repeat size was associated with a delayed age of onset in SCA3 patients (Tezenas du *et al.*, 2014). Similarly, in a subsequent study in patients with SCA3 from mainland China, it was shown that the difference in CAG repeat size between the two *HTT* alleles interacted with the *ATXN3* expansion and affected age of onset in these patients (Chen *et al.*, 2016). A CAG repeat expansion in the *HTT* gene is the cause of Huntington disease (HD), the most common polyglutamine disease worldwide. Like other polyglutamine diseases, the age of onset in HD is inversely associated with the CAG repeat expansion size in the mutant allele which accounts for between 47 to 72% of the variance in age of onset in different HD populations (Cazeneuve and Durr, 2014). However, there is a wide distribution of age of onset in individuals carrying a mutation with an identical number of CAG repeats, suggesting the existence of other important (epi)genetic and/or environmental factors (Arning and Epplen, 2012; GeM-HD consortium, 2015). Given that the age of onset in SCA3 patients was recently found to be influenced by the *HTT* CAG repeat size, we wondered whether the age of onset in HD patients could also be influenced by the CAG repeat size variations in other polyglutamine disease-associated genes (PDAGs), particularly *ATXN3*. Therefore, we assessed the association between the number of CAG repeats in all known PDAGs and age of onset in a large cohort of HD patients.

We obtained clinical data and DNA samples from a subset (n=1000) of manifest HD patients participating in the European Huntington Disease Network REGISTRY study (http://www.euro-hd.net/html/registry). All these participants had entered the study before July 21, 2015. All participants in whom data on age of onset were available and in whom CAG repeat numbers in both alleles of each PDAG could be determined were included in the analyses. Using 10 ng of genomic DNA, two multiplex PCRs were performed in a TProfessional thermocycler (Biometra, Westburg) with labelled primers flanking the CAG stretch for *ATN1*, *ATXN1*, *ATXN7*, *CACNA1A* and *HTT* in one mix and *AR*, *ATXN2*, *ATXN3*, and *TBP* in a second mix (Biolegio) (primers and PCR conditions are available upon request). Every PCR included a negative control without genomic DNA, a reference sample of CEPH 1347-02 genomic DNA and two positive control samples with predetermined 40 and 47 *HTT* CAG repeats (Applied Biosystems). Repeat size determination was performed by running the PCR products on an ABI 3730/3130 automatic DNA sequencer (Applied Biosystems) and analyzing the results with GeneMarker software (version 2.4.0).

To assess whether CAG repeat lengths in PDAGs were associated with age of onset in HD we applied multiple linear regression. Given the known exponential association between age of onset and mutant *HTT* CAG repeat size, the natural logarithmic transformation of age of onset was used as the dependent variable (Langbehn et al., 2004). We modelled the effect of each PDAG on age of onset separately by including its two alleles (with both linear and quadratic terms in order to account for potential non-linear effects) as well as their interaction as predictor variables while also adjusting for the effects of sex and CAG repeat sizes in both *HTT* alleles and their interaction (Aziz *et al.*, 2009). For the *AR* gene only

CAG repeat size in the longer allele was used since men carry only one allele of this X-linked gene. Next, to assess whether the effect of mutant *HTT* CAG repeat size on age of onset was modified by CAG repeat lengths in other PDAG, the interaction between CAG repeat size in the mutant *HTT* allele and CAG repeat size in each of the two alleles of the other PDAG was additionally included. To reduce multicollinearity, particularly with respect to the interaction terms, all continuous predictors were centred around their respective means. To account for the effects of heteroscedasticity and influential points all statistical significance tests were based on robust estimators of standard errors. Moreover, to assure that the results were not unduly affected in case of deviations from model assumptions we also applied a non-parametric method by dividing the group based on median values of each PDAG and comparing differences in age of onset by the non-parametric Mann-Whitney *U*-test. Given the exploratory nature of this study no specific correction for multiple comparisons was applied. All tests were two-sided and significance level was set at $p < 0.05$. All analyses were performed in SPSS version 23.0 (IBM SPSS Statistics for Windows, IBM Corp).

The mean age of onset was 48.8 with a standard deviation of 12.2 years. The number of assessed samples per gene are summarized in Supplementary Table 1. The distribution of CAG repeat lengths followed a unique pattern for each gene and in some cases had a strong preference for a particular range of repeat lengths (Supplementary Table 1 and Supplementary Fig. 1). As expected age of onset was inversely associated with CAG repeat length in the expanded *HTT* allele (β = -0.060, p < 0.001) which accounted for 66.1% of the variation in age of onset in this cohort. Longer CAG repeat size in the larger *ATXN3* allele was associated with a later age of onset in HD patients (β=0.003, p=0.048). Non-parametric comparison of age of onset between participants with CAG repeat sizes below the median versus those with CAG repeat sizes above the median in the larger *ATXN3* allele confirmed this association (median age of onset: 47.6 *vs.* 50.0 years, p=0.025; Fig. 1). There was not a significant interaction between either of the *ATXN3* alleles and the expanded *HTT* allele ($p \geq 0.20$). However, there was a significant interaction between the CAG repeat size in the expanded *HTT* allele and the larger *CACNA1A* allele (β=-3.87 × 10^{-3} and p=0.011 for the interaction effect). Further scrutiny of this interaction revealed that for patients with a below median number of CAG repeats in the expanded *HTT* allele more repeats in the longer *CACNA1A* allele resulted in a later age of onset (median age of onset: 56.1 *vs.* 61.1 years, p=0.003), while for patients with an above median expansion the *CACNA1A* CAG repeat had little influence on the age-of-onset (Fig. 2). There was also a significant interaction between the CAG repeat size in the expanded *HTT* allele and the larger *AR* allele, with a model including a quadratic term for the *AR* CAG repeat size providing the best fit (β= -2.54 × 10^{-4} and p=0.035 for the interaction effect). Comparison of the medians in the total group showed that for patients with a below median number of CAG repeats in the expanded *HTT* allele more repeats on the longer *AR* allele tended to delay age of onset, while for patients with an above median expansion the longer *AR* CAG repeats tended to advance age of onset (Supplementary Fig.2). However, given that *AR* encodes for the androgen receptor, we also performed an additional analyses stratified by sex which demonstrated that the actual effect differed between males and females: In males a longer *AR* allele tended to delay age of onset in subjects with a relatively low expanded *HTT* CAG repeat size (median age of onset: 58.5 *vs.* 55.3 years, p=0.004), while in females a longer *AR* allele resulted in an earlier age of onset in subjects with a relatively larger expanded *HTT* CAG repeat size (median age of onset: 39.2 *vs.* 42.1 years, p=0.009) (Supplementary Fig. 3). Although

regression analysis suggested an effect of CAG repeat size in the smaller alleles of *ATXN7* and *TBP* and a non-linear effect of the longer *ATXN1* CAG repeat size, these effects were statistically non-significant when tested non-parametrically (data not shown).

In conclusion, we found that age of onset in patients with HD is modulated by CAG repeat sizes in the normal range of *ATXN3*, *CACNA1A* and *AR*. Our findings extend those of recent reports in SCAs (Tezenas du *et al.*, 2014; Raposo *et al.*, 2015; Chen *et al.*, 2016), and provide further support for the notion that there may be a biological interaction between different PDAGs (Tezenas du *et al.*, 2014; Tezenas du Montcel, 2015). However, given the exploratory nature of this study, larger studies are needed to confirm these preliminary findings in other cohorts of HD patients.

Supplementary material
- Supplementary table 1
- Supplementary figures 1, 2 and 3
- Acknowledgement list REGISTRY Investigators of the European Huntington Disease Network investigators

Conflict of interest
The authors declare no conflicts of interest.

References

Arning L, Epplen JT. Genetic modifiers of Huntington's disease: beyond CAG. Future Neurology 2012; 7(1): 93-109.

Aziz NA, Jurgens CK, Landwehrmeyer GB, van Roon-Mom WM, van Ommen GJ, Stijnen T, *et al.* Normal and mutant HTT interact to affect clinical severity and progression in Huntington disease. Neurology 2009; 73(16): 1280-5.

Cazeneuve C, Durr A. Genetic and molecular studies. In: Bates GP, Tabrizi SJ, Jones L, editors. Huntington's disease. 4 ed. New York: Oxford University Press; 2014. p. 109-30.

Chen Z, Zheng C, Long Z, Cao L, Li X, Shang H, *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.

GeM-HD consortium. Identification of Genetic Factors that Modify Clinical Onset of Huntington's Disease. Cell 2015; 162(3): 516-26.

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.

Tezenas du Montcel S. Reply: Replicating studies of genetic modifiers in spinocerebellar ataxia type 3: can homogeneous cohorts aid? Brain 2015; 138(Pt 12): e399.

Tezenas du MS, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, *et al.* Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. Brain 2014; 137(Pt 9): 2444-55.

Figure 1: Association between *ATXN3* CAG repeat size and age of onset. Boxplots comparing the age-of-onset between participants with a below or above median number of CAG repeats in their larger *ATXN3* allele (Mann-Whiney *U*-test p=0.025). Black horizontal lines represent medians, boxes display interquartile ranges and whiskers are 1.5 × interquartile range. Circles represent individual patient data with horizontally added jitter.
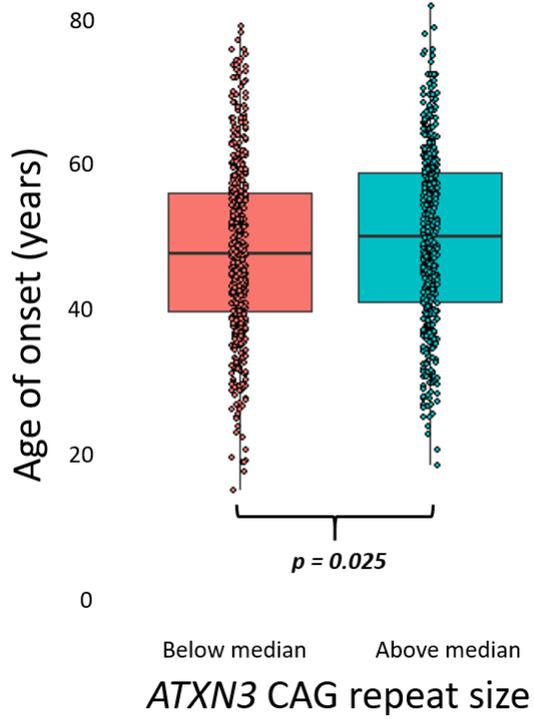
Figure 2: Interaction between the larger *CACNA1A* allele and mutant *HTT* CAG repeat size. Only in patients with mutant *HTT* CAG repeat size below median higher *CACNA1A* CAG repeat size was associated with a higher age-of-onset (Mann-Whiney *U*-test p=0.003). Black horizontal lines represent medians, boxes display interquartile ranges and whiskers are 1.5 × interquartile range. Circles represent individual patient data with horizontally added jitter.