Huntington disease and other polyglutamine diseases : using CAG repeat variations to explain missing heritability

Gardiner, S.L.

Part 3.2

REPEAT VARIATIONS IN POLYGLUTAMINE DISEASE-ASSOCIATED GENES AND COGNITIVE FUNCTION IN OLD AGE

Sarah L. Gardiner,[1,2] Stella Trompet,[3] Behnam Sabayan,[4] Merel W. Boogaard,[5] J. Wouter Jukema,[6] P. Eline Slagboom,[7] Raymund A.C. Roos,[1] Jeroen van der Grond,[8] N. Ahmad Aziz[9,10]

Departments of [1]Neurology, [2]Human Genetics, [3]Internal Medicine, section of gerontology and geriatrics, [5]Clinical Genetics,[6]Cardiology, [7]Molecular Epidemiology and [8]Radiology, Leiden University Medical Centre, Leiden, The Netherlands
[4]The Ken and Ruth Davee Department of Neurology, Northwestern University, Chicago, United States
[9]Population Health Sciences, German Centre for Neurodegenerative Diseases (DZNE), Bonn, Germany
[10]Department of Neurology, University of Bonn, Bonn, Germany

Under review at Neurobiology of Aging.

ABSTRACT

Although the heritability of cognitive function in old age is substantial, genome wide-association studies (GWAS) have had limited success in elucidating its genetic basis, leaving a considerable amount of 'missing heritability'. Aside from single nucleotide polymorphisms (SNPs), GWAS are unable to assess other large sources of genetic variation, such as tandem repeat polymorphisms. Therefore, here we studied the association of cytosine-adenine-guanine (CAG) repeat variations in polyglutamine disease-associated genes (PDAGs) with cognitive function in older adults. In a large cohort consisting of 5 786 participants, we found that the CAG repeat number in three PDAGs (*TBP*, *HTT* and *AR*) were significantly associated with cognitive function, which together accounted for 0.49% of the variation. Furthermore, in an MRI sub-study, we found that CAG repeat polymorphisms in four PDAGs (*ATXN2*, *CACNA1A*, *ATXN7* and *AR*) were associated with different imaging characteristics, including brain stem, putamen, thalamus and amygdala volumes. Our findings indicate that tandem repeat polymorphisms are associated with cognitive function in older adults and highlight the importance of PDAGs in elucidating its missing heritability.

Keywords: cognitive function, polyglutamine diseases, tandem repeats, missing heritability, polyglutamine disease-associated genes, Huntington disease

INTRODUCTION

Cognitive function is a key determinant of quality of life and independence in old age. [1-3] The proportion of the world's elderly population is rapidly increasing with the number of people aged 65 years or older estimated to rise from 8.5% of the world's population in 2015 to 16.7% by 2050.[4] Therefore, the necessity to understand and investigate the pathophysiology and risk factors of cognitive decline in the ageing population is urgent. The contribution of genetics to brain function in older adults is substantial.[5-8] Numerous genome-wide association studies (GWAS) have identified several single nucleotide polymorphisms (SNP) associated with age-related cognitive decline.[9] The largest meta-analysis to date, investigating 57 population-based cohorts comprising a total of 300 486 participants aged between 16 and 102 years, reported 148 SNPs associated with general cognitive function at a genome-wide significance level. The maximum proportion of phenotypic variance in general cognition explained in three independent samples by a prediction score derived from these data ranged between 2.63-4.31%,[10] leaving a substantial proportion of unexplained or 'missing heritability'. Despite their essential role in genetic research, GWAS are limited due to the fact that these studies cannot assess polymorphisms in tandem repeat sequences.[11] Nonetheless, 3% of genomic DNA consists of such repetitive DNA stretches and thus have a considerable impact on genetic variation.[12-14]

Nine hereditary neurodegenerative diseases, known as polyglutamine disorders, including Huntington disease (HD), are the most prevalent disorders associated with tandem repeats.[15,16] These diseases are caused by an elongated cytosine-adenine-guanine (CAG) repeat sequence in the protein-coding region of the respective polyglutamine disease-associated gene (PDAG) (**Table 1**).[17-36] The elongated repeat sequence is present from birth. However, symptoms usually do not arise until middle age, making age a prominent risk modifier for disease onset. Furthermore, aside from progressive motor defects and psychiatric disturbances, polyglutamine disorder symptoms often affect cognitive function. Increasing evidence implicates the polyglutamine domains of PDAGs as key regulators of transcriptional regulation, synaptic plasticity, calcium homeostasis and mitochondrial energy production, dysregulation of which have been associated with cognitive ageing.[37-39] However, to what extent more common CAG repeat length variations in PDAGs are associated with cognitive decline in older adults and changes in relevant brain structures, is still unknown. Since polyglutamine diseases are caused by expanded repeat sequences, the most rational hypothesis would be that larger repeat numbers in PDAGs are associated with a worse cognitive function. However, previous studies demonstrated more complex, non-linear associations between CAG repeat size variations, and neurodegenerative and neuropsychiatric disorders.[40-43] Therefore, in this study, we aimed to assess the association of cognitive function in older adults with the CAG repeat number in the nine known PDAGs, investigating different linear and non-linear associations in an unbiased fashion.

Table 1. Summary genotyped polyglutamine disease-associated genes (PDAGs).

			CAG repeat ranges			PROSPER			
PDAG	Disease	Protein	normal	pathological	Allele	Mean±SD	Median	N	Range
ATXN1	SCA1	Ataxin-1	6-38	39-83	short	29.2±1.0	29	5633	17 – 36
					long	30.8±1.7	30	5633	26 – 44
ATXN2	SCA2	Ataxin-2	14-32	33-500	short	21.9±0.6	22	5548	11 – 27
					long	22.4±1.2	22	5548	22 – 27
ATXN3	SCA3	Ataxin-3	12-44	52-87	short	19.0±4.4	20	5544	14 – 34
					long	24.3±3.8	23	5543	14 – 29
CACNA1A	SCA6	CACNA1A	4-18	20-33	short	10.6±2.1	11	5633	4 – 14
					long	12.5±1.1	13	5633	7 – 17
ATXN7	SCA7	Ataxin-7	3-19	37-460	short	10.1±0.5	10	5285	7 – 16
					long	10.8±1.2	10	5285	10 – 25
TBP	SCA17	TBP	25-43	45-66	short	36.3±1.8	37	5559	26 – 40
					long	37.9±1.0	38	5559	30 – 47
HTT	HD	Huntingtin	6-26	36-121	short	16.9±2.1	17	5602	9 – 29
					long	20.2±3.5	19	5602	10 – 38
ATN1	DRPLA	Atrophin-1	3-38	48-93	short	12.3±3.1	14	5633	5 – 20
					long	15.6±2.2	16	5633	8 – 27
AR	SBMA	Androgen receptor	6-34	36-72	short	21.2±2.7	21	5546	7 – 35
					long	22.8±2.9	23	5546	7 - 39

PROSPER= Prospective Study of Pravastatin in the Elderly at Risk. PDAG=polyglutamine disease-associated genes. SD=standard deviation. CACNA1A=calcium channel, voltage-dependent P/Q type, α 1A subunit; TBP=thymine-adenine-thymine-adenine (TATA) box binding protein; SCA=spinocerebellar ataxia; HD=Huntington Disease; DRPLA=Dentatorubropallidoluysian atrophy; SBMA=spinal bulbar muscular atrophy.

SUBJECTS AND METHODS

Subjects

The nine known PDAGs (including *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *HTT*, *ATN1* and *AR*) were genotyped in all participants with sufficient amounts of DNA available from blood samples of the Prospective Study of Pravastatin in the Elderly at Risk (PROSPER) study. This well characterized cohort included 5 786 men and women between 70-83 years old with a pre-existing vascular disease or a raised risk for such a disease. The study originally was a prospective multicentre randomized placebo-controlled trial to assess the effect of treatment with pravastatin on the risk of major vascular events in the elderly. Previous research showed that the treatment with pravastatin in this cohort did not affect cognitive decline or brain volume.[44,45] Participants were recruited from Scotland (n=2517), Ireland (n=2173) and the Netherlands (n=1096). All participants were Caucasian and none were diagnosed with a polyglutamine disease. The study was approved by the institutional ethics review boards of all centres and informed consent was obtained from all participants.[46]

Cognitive function assessment

The Mini-Mental State Examination (MMSE) was used to assess global cognitive function. To measure various cognitive domains, four neuropsychological tests were administered to all subjects. The Stroop-Colour-Word test was applied to assess executive function by evaluating attention span. The outcome parameter for this test was the total number of seconds required to complete the third Stroop card (Stroop III). The Letter-Digit Coding Test (LDT) was applied to assess processing speed with the total number of correct entries completed in 60 sec as outcome. Memory was assessed with the 15-Picture Learning test (PLT) which measured immediate and delayed recall. Immediate recall was defined as the accumulated number of pictures recalled over three learning trials and delayed recall was defined as the number of pictures recalled after 20 min. Cognitive function was measured at baseline, after 9, 18 and 30 months of follow-up, and at the end of the study which varied between participants (36-48 months).[47]

Imaging characteristics

Two successive magnetic resonance imaging (MRI) scans of the brain were obtained from a total of 646 Dutch participants with a follow-up of 30 months. MRI was performed on a system operating at 1.5 T field strength (Philips Medical Systems) and proton desity-T2/dual fast-spin-echo images were collected (TE, 27/120 ms; TR, 3000 ms; echo train length factor, 10; 48 contiguous 3-mm slices; matrix 256x256; FOV, 220).[45] From these images, several MRI parameters were determined by segmenting parenchyma (whole-brain) volume using semi-automated software developed by the Division of Image Processing Department of Radiology of the Leiden University Medical Center. Percentage of atrophy was defined as the difference between intracranial and parenchymal volume divided

by the parenchymal volume. Increase in atrophy percentage per year was determined by dividing the difference in atrophy between the first and the second scan by 2.5 (30 months/ 12 months). Furthermore, baseline values of grey matter volume, white matter volume and total brain volume were reported in mm^3 normalized to Montreal Neurological Institute (MNI) space, while the volume of the brain stem and the subcortical structures (i.e. caudate nucleus, putamen, thalamus, globus pallidus, nucleus accumbens, amygdala and hippocampus) were reported in unnormalized mm^3. The volumes of the right and left subcortical structures were summed to generate one overall volume per subcortical structure.

Genotyping

To determine the CAG repeat length in the nine PDAGs for each individual, a polymerase chain reaction (PCR) was performed in a TProfessional thermocycler (Biometra, Westburg) with labelled primers flanking the CAG stretch of the PDAGs (Biolegio) (**Supplemental Table 1**). The PCR was performed using 10 ng of genomic DNA, 1x OneTaq mastermix (New England Biolabs, OneTaq Hot start with GC Buffer master mix), 1 µl of primer Mix A or B (**Supplemental Table 1**) and Aqua B. Braun water to a final volume of 10 µl. The PCR was run with 27 cycles of 30 seconds, denaturation at 94°C, one minute of annealing at 60°C and two minutes elongation at 69°C, preceded by five minutes of initial denaturation at 94°C. Final elongation was performed at 69°C for five minutes. Every PCR included a negative control without genomic DNA and a reference sample of CEPH 1347-02 genomic DNA. The PCR products were run on an ABI 3730 automatic DNA sequencer (Applied Biosystems) and analysed using the GeneMarker software version 2.4.0. For every analysis, we included three controls with known CAG repeat lengths for each PDAG to assure every run was performed reliably. All assessments were performed by randomizing study participants across batches while researchers were blinded with respect to the clinical information.

Statistical Analysis

To analyse the association between the CAG repeat number in PDAGs and cognitive function in older adults, we initially performed a principal component analysis to aggregate the different cognitive test scores into one outcome. However, the first principal component of 'cognition at baseline' and 'cognition over time' both did not explain more than 52.7 – 55.8% of the variance. Therefore, we did not find these components to summarize the actual outcomes reliably. Instead, we calculated Z-scores for all cognitive assessment test outcomes, including MMSE score, total number of seconds to complete Stroop III, total number of correct entries in the LDT and total number of pictures recalled in the immediate and delayed recall of the PLT. The data of Stroop III had a skewed distribution and thus was log-transformed, and inverted as higher scores indicated a worse performance. We summed up the Z-scores of the data

derived at baseline to create a baseline summary score of cognitive function. A higher score indicated a better cognitive function. Generalized linear mixed-effect models with this summary score as dependent variable were applied to assess the association between the CAG repeat numbers in the PDAGs and cognitive function at baseline. Country (i.e. Scotland, Ireland or the Netherlands) was set as random effect to account for potential population stratification and the CAG repeat number of both alleles were set as fixed effects. Each allele was defined as either 'short' or 'long' relative to the CAG repeat number in the other allele. The alleles were defined as such, because previous research has shown the 'shorter' and the 'longer' allele to have different associations with various outcomes, such as age of onset in polyglutamine diseases, body mass index and depression.[43,48-51] The CAG repeat size in the X-linked *AR* gene was examined by 1) analysing men and women separately and 2) investigating either the shorter or the longer *AR* allele separately in men and women combined. To assess potential interaction or non-linear effects,[51,52] we also included a product term of both alleles and a quadratic term for each allele as fixed effects.

To assess the association between the CAG repeat number in the PDAGs and the decline in cognitive function over time, we created a summary score of the cognitive function over time by adding the Z-scores of the data derived per visit (the Z-scores were created with respect to the baseline distribution). Consequently, each individual had one cognitive summary score per visit. We then applied similar generalized linear mixed-effects models with this summary score as dependent variable. The time of assessment, defined as the number of months of follow-up (i.e. 0, 9, 18, 30 or 36-48), and product terms of this follow-up time and the CAG repeat number in both alleles were added as fixed effects, in addition to the variables in the previous model. To model the clustering by country as well as the longitudinal repeated measurements within each individual, we applied a random intercept-and-slope model: Apart from a random intercept for country, a random slope for the time of assessment was added to account for the longitudinal inter-individual variation in the rate of cognitive decline (i.e. the random slope for the time effect) and to allow treatment of missing values at each time point under the missing-at-random assumption.

MRIs were only performed in Dutch participants and MRI parameters were not assessed at more than one point in time. Therefore, linear regression instead of generalized linear mixed-effect modelling, was applied to analyse the association between the different imaging parameters and the CAG repeat number in the PDAGs. The MRI variables were set as dependent variables and the CAG repeat numbers of both alleles, including the product term and the quadratic terms were included as independent variables.

In order to assess whether CAG repeat variation in the PDAGs could explain a significant additional amount of variation in the respective dependent variable, we first performed

an omnibus test (i.e., a restricted F-test) per PDAG by including as independent variables all terms associated with the respective locus (i.e., CAG repeat size of both alleles, their interaction, a quadratic term for each allele and the product of follow-up time and the CAG repeat number in both alleles if applicable). We applied a false discovery rate (FDR) correction to account for multiple testing, assuming nine independent tests with the FDR q set at .05, to determine whether an omnibus test result was significant.[53] Only in cases where an omnibus test was statistically significant after multiple testing correction, we performed post hoc tests by assessing the effect of the individual PDAG alleles and their associated higher-order terms in order to gain more insight into the nature of the association. Non-significant higher order terms were removed and the analyses were repeated to arrive at a final model. All final models were corrected for age, sex and population structure using principal components generated from genome wide genotyping data.[54] Although having a pre-existing vascular disease or a raised risk for such a disease probably has an effect on cognition and cognitive decline, this variable is unlikely to be associated with CAG repeat polymorphisms in PDAGs and can thus not be considered a confounder. Therefore, we did not correct for this covariate. To reduce multicollinearity, the CAG repeat numbers were centred around their respective means. To account for potential effects of heteroscedasticity and influential points, all statistical significance tests were based on robust estimators of standard errors and all CAG repeat lengths with a frequency of less than ten for the analyses of cognitive function or less than five for the analyses of imaging characteristics, were excluded (**Supplemental Table 2 and 3**). We calculated the R^2 per PDAG for each final model to determine the amount of variance explained by each gene.[55]

We determined the amount of variance in cognitive function explained by the significant repeat polymorphisms in total by incorporating the fixed effects of the final significant models with cognitive function over time as outcome into one model and calculating its marginal R^2. To illustrate this combined effect we 1) calculated the residual cognitive function after regression on age, sex, follow-up time and principal components generated from genome wide genotyping data as fixed factors, and country and the follow-up time as random factors in a generalized mixed-effects model, 2) performed linear regression with the CAG repeat sizes of the significant PDAGs, including all interaction and non-linear effects which were identified as significant in the main analyses as independent variables and the residual cognitive function as the outcome, 3) divided the total cohort in two equally sized groups based on the predicted values of this regression model, 4) plotted the average residual cognitive function of these two groups. All data are displayed as means and 95% confidence intervals (CIs) unless otherwise specified. All analyses were performed in STATA/SE version 14.2 (StataCorp LLC).

RESULTS

We were able to determine the CAG repeat length of between 5285-5633 individuals for each gene. The lacking samples were due to too little available DNA and were missing completely at random (**Table 1**). In total, we found 79 participants with CAG repeat numbers within the pathological range of at least one PDAG, including *ATXN1* (n=4), *ATXN2* (n=2), *TBP* (n=66) and *HTT* (n=7). For a more extensive report on these findings, please see our previous publication.[56] Between 0.1 to 2.9% of the cases per gene were excluded from the analyses of cognitive function because of CAG repeat numbers with a frequency of less than ten and between 0.8 to 7.2% of the cases were excluded from the analyses of the imaging parameters due to CAG repeat numbers with a frequency of less than five (**Supplemental Table 2 and 3**). The cohort characteristics are summarized in **Table 2**. The summary scores of baseline cognitive function and cognitive function over time ranged from -12.55 to 9.92 and from -16.30 to 10.15, respectively.

Baseline cognitive function was associated with CAG repeat variations in *AR* in women

Initial omnibus tests with cognitive function at baseline as the outcome were significant for *ATXN2* and *TBP* ($p \leq .008$). In addition, the omnibus tests for the shorter *AR* allele assessed in both sexes and the longer *AR* allele assessed in women only, were statistically significant ($p \leq .004$). The post hoc tests subsequently indicated a significant non-linear association between cognitive function at baseline and the CAG repeat number in the longer *AR* allele in women (**Table 3**). Both relatively small and relatively large CAG repeat numbers in the longer *AR* allele in women were associated with a higher cognitive function at baseline (**Figure 1**). A CAG repeat number of about 22-23 in the longer *AR* allele was associated with the lowest cognitive function at baseline. These results did not change materially after correcting for age, sex and population structure (**Table 3**). Post hoc tests for *ATXN2*, *TBP* and the shorter *AR* allele assessed in both sexes did not indicate a significant association with cognitive function at baseline.

The CAG repeat variations in *TBP, HTT* and *AR* were associated with cognitive function over time

The omnibus tests with cognitive function over time as outcome were significant for *ATXN2*, *TBP* and *HTT* ($p < .005$). Furthermore, the omnibus tests were statistically significant, when the longer and the shorter *AR* allele were assessed separately in both sexes, and when both *AR* alleles were assessed in women only ($p < .009$). Subsequent post hoc tests indicated a significant association between cognitive function over time and CAG repeat variations in the longer *TBP* allele, the longer *HTT* allele, the longer *AR* allele assessed in only women and the shorter *AR* allele assessed in both sexes (**Table 3**). Larger CAG repeat numbers in the longer *TBP* allele were associated with a lower summary score of cognitive function at each point in time that cognitive function was

Table 2. Summary PROSPER characteristics.

Variable	Observations (n)	Percentage		
Sex				
men	2798	48.4		
women	2988	51.6		
Country				
the Netherlands	1096	18.9		
Scotland	2517	43.5		
Ireland	2173	37.6		
Education				
elementary School	131	2.3		
> elementary School	5655	97.7		
Variable	**Observations (n)**	**Mean**	**SD**	**Range**
Age at baseline	5786	75.33	3.35	69.37 – 83.39
Cognitive function				
Baseline measurements				
MMSE (score)	5718	28.03	1.54	22 – 30
Stroop III (sec.)	5370	66.59	26.91	20.09 – 346.30
LDT (correct #)	5407	23.07	7.83	5 – 54
PLTi (correct #)	5444	9.32	1.95	4 – 14.33
PLTd (correct #)	5444	10.14	2.59	3 - 15
Measurements from all time points				
MMSE (score)	25401	28.10	1.81	8 – 30
Stroop III (sec.)	23803	65.64	27.17	10.09 – 346.30
LDT (correct #)	24340	22.90	7.81	0 – 55
PLTi (correct #)	24670	9.38	2.00	1 – 15
PLTd (correct #)	24670	10.08	2.80	0 - 15
Summary Scores[a]				
Baseline cognitive function	5003	0.18	3.57	-12.55 – 9.92
Decline in cognitive function	22399	0.21	3.63	-16.30 – 10.15
MRI variables (*The Netherlands*)				
Baseline atrophy (%)	536	26.21	3.10	12.74 – 37.04
Atrophy after 30 months (%)	542	27.23	3.43	7.61 – 38.41
Increase in atrophy per year (%)	527	-0.42	0.84	-5.34 – 4.81
Grey matter volume (mm^3)[b]	494	$5.91*10^5$	$4.45*10^4$	$4.24*10^5$ – $7.16*10^5$
White matter volume (mm^3)[b]	494	$7.68*10^5$	$3.85*10^4$	$6.46*10^5$ – $8.98*10^5$

Table 2. (continued)

Variable	Observations (n)	Mean	SD	Range
Total brain volume (mm^3)[b]	494	$1.36*10^6$	$6.57*10^4$	$1.14*10^6$ – $1.56*10^6$
Nucleus caudatus (mm^3)	423	$7.52*10^3$	$9.89*10^2$	$5.17*10^3$ – $1.11*10^4$
Putamen (mm^3)	409	$1.05*10^4$	$1.10*10^3$	$7.68*10^3$ – $1.39*10^4$
Thalamus (mm^3)	428	$1.62*10^4$	$1.29*10^3$	$1.28*10^4$ – $1.92*10^4$
Globus pallidus (mm^3)	387	$3.87*10^3$	$7.23*10^2$	$2.21*10^3$ – $6.26*10^3$
Nucleus accumbens (mm^3)	209	$1.13*10^3$	$2.53*10^2$	$5.84*10^2$ – $1.98*10^3$
Amygdala (mm^3)	423	$4.00*10^3$	$5.80*10^2$	$2.36*10^3$ – $5.38*10^3$
Hippocampus (mm^3)	413	$9.28*10^3$	$1.06*10^3$	$6.71*10^3$ – $12.2*10^4$
Brain stem (mm^3)	429	$2.34*10^4$	$2.60*10^3$	$1.72*10^4$ – $30.0*10^4$

MMSE=mini mental state examination. Stroop III=interference segment of the Stroop test. LDT=letter digit test. PLTi=average total word learning test immediate recall. PLTd=word learning test delayed recall. SD=standard deviation. [a]) Calculated by adding the Z scores derived from the results of the cognitive tests. [b]) Normalized to Montreal Neurological Institute (MNI) space.

assessed. This association is visualized by observing that as individuals age the summary score of cognitive function declines, but the negative association with the CAG repeat number in the longer *TBP* allele remains apparent in each age group (**Figure 2A**). Similarly, in each age group the CAG repeat number in the longer *HTT* allele and the longer *AR* allele in women had a non-linear associations with cognitive function over time. Both relatively small and relatively large CAG repeat numbers in the longer *HTT* allele were associated with a lower cognitive function (**Figure 2B**). Conversely and similar to its association with cognitive function at baseline, smaller and larger CAG repeat numbers in the longer *AR* allele assessed in women were associated with a higher cognitive function (**Figure 2C**). The CAG repeat number in the shorter *AR* allele in both sexes was associated with the rate of decline in cognitive function, based on the significant association between the interaction term of the CAG repeat number in the shorter *AR* allele and follow-up time (**Table 3**). Larger CAG repeat numbers in the shorter *AR* allele assessed in both sexes was associated with a faster decline in cognitive function over time (**Figure 2D**). These associations remained significant and largely unaltered after correction for age, sex and population structure. Post hoc tests for *ATXN2* and the longer *AR* allele assessed in both sexes did not indicate a significant association with cognitive function over time.

In combination, CAG repeat polymorphisms in *TBP*, *HTT* and the shorter *AR* allele assessed in both sexes explained 0.49% of the total variance in cognitive function (**Figure 3**). In this calculation, we did not incorporate the effect of the CAG repeat polymorphisms in the longer *AR* allele, because this effect was only significant when assessed in women.

Table 3. The association between summary scores of cognitive function and the CAG repeat number in polyglutamine disease-associated genes.

Outcome Variable[a]	β-coefficient[d]	SE	t	p-value	95% CI		R^2
Baseline cognitive function							
AR_s ♀	$7.78*10^{-2}$	$4.51*10^{-2}$	1.72	.085	$-1.07*10^{-2}$	$1.66*10^{-1}$	$8.95*10^{-3}$
AR_l ♀	$1.52*10^{-2}$	$5.50*10^{-2}$	0.28	.782	$-9.27*10^{-2}$	$1.23*10^{-1}$	
AR_l2 ♀	**$1.70*10^{-2}$**	**$4.14*10^{-3}$**	**4.12**	**<.001***	**$8.93*10^{-3}$**	**$2.52*10^{-2}$**	
AR_s ♀[b]	$7.51*10^{-2}$	$3.15*10^{-2}$	2.38	.017	$1.33*10^{-2}$	$1.37*10^{-1}$	
AR_l ♀[b]	$-8.20*10^{-4}$	$5.45*10^{-2}$	-0.02	.988	$-1.08*10^{-1}$	$1.06*10^{-1}$	
AR_l2 ♀[b]	**$2.01*10^{-2}$**	**$4.18*10^{-3}$**	**4.82**	**<.001***	**$1.19*10^{-2}$**	**$2.83*10^{-2}$**	
Cognitive function over time							
TBP_s	$9.67*10^{-3}$	$5.03*10^{-3}$	1.93	.054	$-1.76*10^{-4}$	$1.95*10^{-2}$	$4.78*10^{-3}$
TBP_l	**$-4.79*10^{-2}$**	**$2.29*10^{-2}$**	**-2.10**	**.036***	**$-9.27*10^{-2}$**	**$-3.10*10^{-3}$**	
TBP_s[c]	$2.01*10^{-2}$	$9.13*10^{-3}$	2.20	.028	$2.22*10^{-3}$	$3.80*10^{-2}$	
TBP_l[c]	**$-6.08*10^{-2}$**	**$2.29*10^{-2}$**	**-2.65**	**.008***	**$-1.06*10^{-1}$**	**$1.59*10^{-2}$**	
HTT_s	$-1.90*10^{-2}$	$3.32*10^{-2}$	-0.57	.566	$-8.40*10^{-2}$	$4.59*10^{-2}$	$3.06*10^{-3}$
HTT_l	$9.40*10^{-3}$	$1.23*10^{-2}$	0.77	.443	$-1.46*10^{-2}$	$3.34*10^{-2}$	
HTT_l2	**$-4.74*10^{-3}$**	**$1.71*10^{-3}$**	**-2/78**	**.005***	**$8.08*10^{-3}$**	**$1.40*10^{-3}$**	
HTT_s[c]	$-8.76*10^{-3}$	$3.04*10^{-2}$	-0.29	.773	$-6.84*10^{-2}$	$5.08*10^{-2}$	
HTT_l[c]	$1.48*10^{-2}$	$1.46*10^{-2}$	1.02	.310	$-1.38*10^{-2}$	$4.34*10^{-2}$	
HTT_l2[c]	**$-5.20*10^{-3}$**	**$1.83*10^{-3}$**	**-2.84**	**.004***	**$-8.78*10^{-3}$**	**$-1.62*10^{-3}$**	
AR_s ♀	$5.87*10^{-2}$	$2.24*10^{-2}$	2.63	.009	$1.49*10^{-2}$	$1.03*10^{-1}$	$2.37*10^{-2}$
AR_l ♀	$1.24*10^{-2}$	$1.94*10^{-2}$	0.64	.523	$-2.57*10^{-2}$	$5.05*10^{-2}$	
AR_l2 ♀	**$1.66*10^{-2}$**	**$7.51*10^{-3}$**	**2.20**	**.028***	**$1.83*10^{-3}$**	**$3.13*10^{-2}$**	
AR_s ♀[b]	$5.72*10^{-2}$	$1.84*10^{-2}$	3.12	.002	$2.12*10^{-2}$	$9.32*10^{-2}$	
AR_l ♀[b]	$-2.47*10^{-3}$	$1.58*10^{-2}$	-0.16	.876	$-3.35*10^{-2}$	$2.85*10^{-2}$	
AR_l2 ♀[b]	**$1.91*10^{-2}$**	**$1.77*10^{-3}$**	**10.83**	**<.001***	**$1.57*10^{-2}$**	**$2.26*10^{-2}$**	
AR_s	$2.94*10^{-3}$	$1.54*10^{-2}$	-0.19	.851	$-3.37*10^{-2}$	$2.78*10^{-2}$	$7.86*10^{-4}$
AR_s_m	**$-5.48*10^{-4}$**	**$-5.48*10^{-4}$**	**-4.02**	**<.001***	**$-8.14*10^{-4}$**	**$-2.81*10^{-4}$**	
AR_s[c]	$-3.24*10^{-3}$	$1.63*10^{-2}$	-0.20	.842	$-3.52*10^{-2}$	$2.87*10^{-2}$	
AR_s_m[c]	**$-4.00*10^{-4}$**	**$1.36*10^{-4}$**	**-2.94**	**.003***	**$-6.66*10^{-4}$**	**$-1.33*10^{-4}$**	

s=relatively shorter allele. l=relatively longer allele. s2=quadratic term relatively shorter allele. l2=quadratic term relatively longer allele. _m= interaction with month. PDAGs= polyglutamine disease-associated genes, SE=standard error. CI=confidence interval. R^2 = estimated variance explained by the significant association. **Bold** indicates the variables that drive the significant association. [a]) Only the variables of PDAGs are reported in the cases where the omnibus test was statistically significant after multiple testing correction. [b]) Corrected for age and population structure using principal components. [c]) Corrected for age, sex and population structure using principal components. [d]) This column indicates the change in summary score of cognitive function per unit CAG repeat size increase. *) p-value statistically significant.

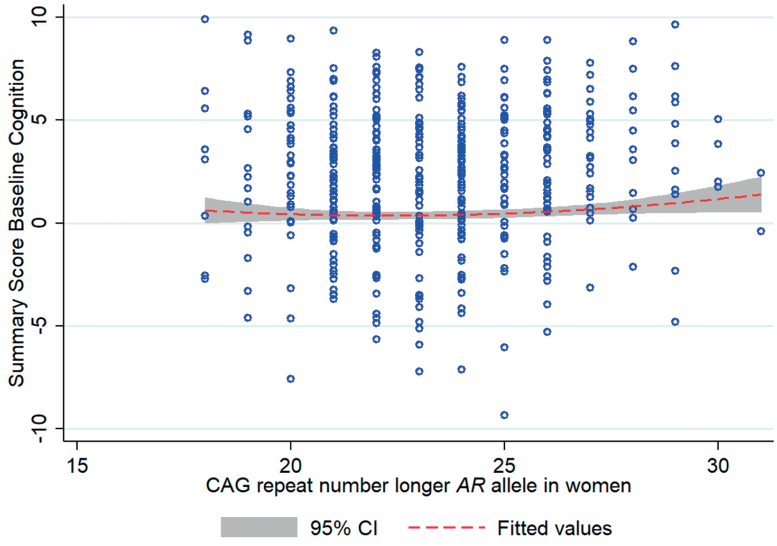


Figure 1. The association between cognitive function at baseline and the CAG repeat number in the longer *AR* allele in women. The cognitive function assessed at baseline had a non-linear association with the CAG repeat number in the longer *AR* allele in women. In women, both relatively small en relatively large CAG repeat numbers were associated with a higher cognitive function at baseline.

CAG repeat variations in *ATXN2, CACNA1A, ATXN7* and *AR* were significantly associated with brain stem and subcortical structure volumes

The MRI variables of which omnibus tests and subsequent post hoc tests indicated significant associations with CAG repeat polymorphisms in the PDAGs are presented in **Table 4**. In women, the CAG repeat number in the longer *AR* allele had a non-linear association with the volume of the amygdala. Both smaller and larger CAG repeat numbers were associated with a higher volume of the amygdala (**Supplemental Figure 1A**). The longer *AR* allele in both sexes had a similar non-linear association with the volume of the amygdala, the putamen and the thalamus (**Supplemental Figures 1B, C and D**). In contrast, larger CAG repeat numbers in the shorter *AR* allele assessed in both sexes were associated with a smaller volume of the amygdala (**Supplemental Figure 2**).

When the shorter *CACNA1A* allele had less than 11 repeats, larger CAG repeat numbers in the longer *CACNA1A* allele were associated with an increased volume the amygdala (**Supplemental Figure 3A**). CAG repeat variations in *CACNA1A* were also associated with the volume of the thalamus. Larger CAG repeat numbers in the shorter *CACNA1A* allele were associated with a larger thalamic volume (**Supplemental Figure 3B**). The CAG repeat number in the longer *ATXN2* allele and the shorter *ATXN7* allele had inverse associations with the volume of the brain stem (**Supplemental Figure 4A and**

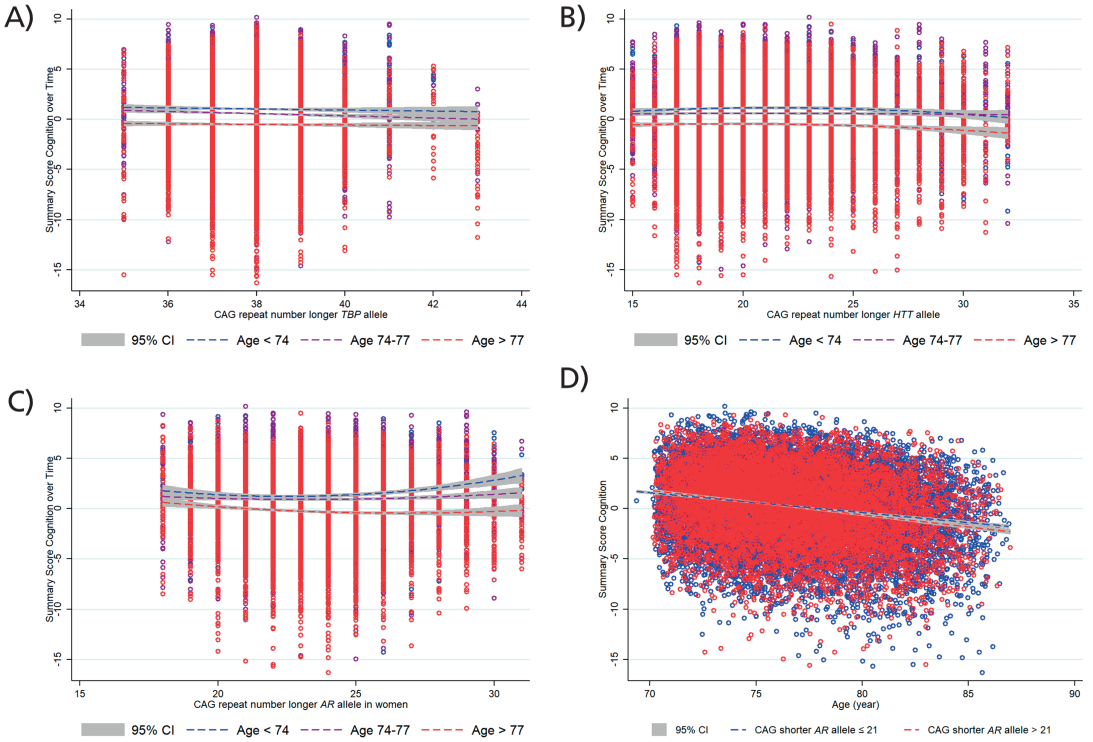


Figure 2. The association between cognitive function over time and the CAG repeat number in *TBP, HTT* and *AR*. A. Larger CAG repeat numbers in the longer *TBP* allele were associated with a lower cognitive function. This association remained significant throughout the entire assessment period and is thus visible in all age groups. **B.** Both relatively small and relatively large CAG repeat numbers in the longer *HTT* allele were associated with a decreased cognitive function. This association was also present at all assessment time-points and can be observed over the three different age groups **C.** Relatively small and relatively large CAG repeat numbers in the longer *AR* allele assessed in women were associated with a higher cognitive function. This association was again consistent over the entire assessment period and all age groups. **D.** CAG repeat numbers > 21 CAG repeats were associated with a faster decline in cognitive function over time compared to CAG repeat numbers ≤ 21.

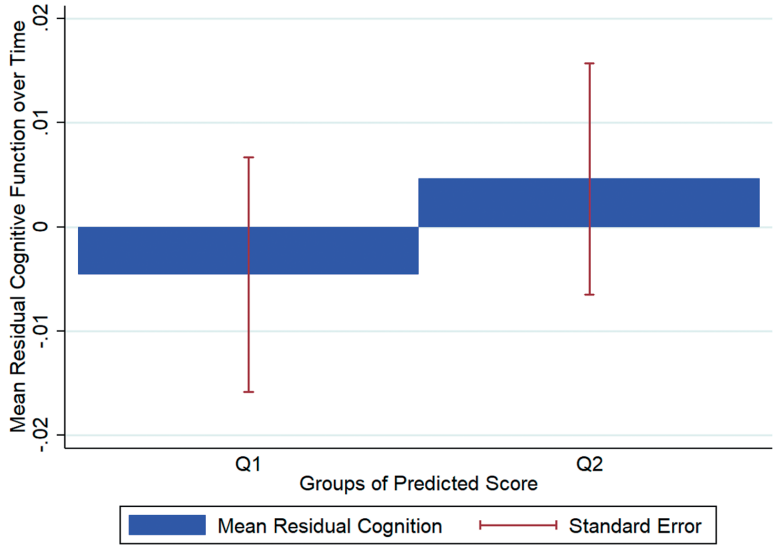


Figure 3. The effect of CAG repeat size variations in polyglutamine disease-associated genes (PDAGs) on the summary score of cognitive function over time. This plot illustrates that the combination of CAG repeat size variations in the three PDAGs can account for a variation of up to 0.02 summary score points.

B). The CAG repeat number in the shorter *ATXN2* allele had non-linear associations with the volume of the putamen and the thalamus (**Supplemental Figure 4C and D**). Both smaller and larger CAG repeat numbers in the shorter *ATXN2* allele were associated with smaller volumes of both subcortical structures. The association between the CAG repeat number in *ATXN2* and the increase in percentage of atrophy per year, the volume of the amygdala and the volume of the globus pallidus, and the association between the interaction of both *ATXN7* alleles and the volume of the nucleus accumbens were mainly driven by a few influential points and, therefore, unlikely to be robust (**Supplemental Figures 5A, B, C and 6**). CAG repeat variations in the PDAGs were not significantly associated with atrophy at baseline, grey matter volume, white matter volume, total brain volume, the volume of the hippocampus or the volume of the nucleus caudatus. All results remained materially unaltered after correcting for age, sex and population structure (**Table 4**).

DISCUSSION

To our knowledge, this study is the first to investigate the association between CAG repeat variations within the 'normal' range of PDAGs and cognitive function in older adults. We found that CAG repeat number variations in *TBP*, *HTT* and *AR* were significantly associated with cognitive function independent of age. In addition, we

Table 4. The association between MRI variables and the CAG repeat number in polyglutamine disease-associated genes.

Outcome	Variable[a]	β-coefficient[d]	SE	t	p-value	95% CI		R^2
Increase in Atrophy								
	ATXN2_s	$1.04*10^{-1}$	5.82*10-2	1.79	.075	$-1.04*10^{-2}$	$2.18*10^{-1}$	$5.30*10^{-3}$
	ATXN2_l	$-5.83*10^{-2}$	$5.48*10^{-2}$	-1.06	.288	$-1.66*10^{-1}$	$4.94*10^{-2}$	
	ATXN2_sl	**$2.24*10^{-1}$**	**$1.06*10^{-1}$**	**2.12**	**.035***	**$1.62*10^{-2}$**	**$1.62*10^{-2}$**	
	ATXN2_s[b]	$8.22*10^{-2}$	$4.21*10^{-2}$	1.95	.052	$-5.66*10^{-4}$	$1.65*10^{-1}$	
	ATXN2_l[b]	$-5.14*10^{-2}$	$5.61*10^{-2}$	-0.92	.360	$-1.62*10^{-1}$	5.89*10-2	
	ATXN2_sl[b]	**$1.44*10^{-1}$**	**$5.88*10^{-2}$**	**2.44**	**.015***	**$2.81*10^{-2}$**	**$2.59*10^{-1}$**	
Subcortical structures								
Thalamus								
	ATXN2_s	-1040.46	299.23	-3.48	.001	-1628.72	-452.21	$1.92*10^{-2}$
	ATXN2_l	-25.54	61.43	-0.42	.678	-146.31	95.23	
	ATXN2_s2	**-245.40**	**67.15**	**-3.65**	**<.001***	**-377.10**	**-113.39**	
	ATXN2_s[b]	-1074.56	363.73	-2.95	.003	-1789.95	-359.18	
	ATXN2_l[b]	-10.65	67.34	-0.16	.874	-143.11	121.80	
	ATXN2_s2[b]	**-246.22**	**81.53**	**-3.02**	**.003***	**-406.58**	**-85.86**	
	AR_l	-38.11	26.23	-1.45	.147	-89.68	13.45	$1.77*10^{-2}$
	AR_l2	**20.34**	**7.64**	**2.66**	**.008***	**5.31**	**35.37**	
	AR_l[b]	-66.02	26.36	-2.50	.013	-117.87	-14.16	
	AR_l2[b]	**21.20**	**7.66**	**2.77**	**.006***	**6.13**	**36.26**	
Nucleus Accumbens								
	ATXN7_s	-1513.85	88.52	-17.10	<.001	-1688.48	-1339.21	$4.03*10^{-2}$
	ATXN7_l	21.31	17.29	1.23	.219	-12.80	55.41	

Table 4. (continued)

Outcome	Variable[a]	β-coefficient[d]	SE	t	p-value	95% CI		R^2
	ATXN7_s2	599.53	21.66	27.68	<.001	556.80	642.26	
	ATXN7_sl	**320.05**	**35.37**	**9.05**	**<.001***	**250.27**	**389.83**	
	ATXN7_s[b]	-1564.11	96.44	-16.22	<.001	-1754.60	-1373.62	
	ATXN7_l[b]	14.86	17.47	0.85	.396	-19.64	49.37	
	ATXN7_s2[b]	617.60	38.68	15.97	<.001	541.20	694.01	
	ATXN7_sl[b]	**326.38**	**43.21**	**7.55**	**<.001***	**241.02**	**411.73**	
Amygdala								
	ATXN2_s	-10.94	29.72	-0.37	.713	-69.36	47.48	$2.16*10^{-2}$
	ATXN2_l	-214.06	78.92	-2.71	.007	-369.22	-58.90	
	ATXN2_l2	45.54	21.39	2.13	.034	3.49	87.60	
	ATXN2_sl	**34.65**	**10.18**	**3.40**	**.001***	**14.63**	**54.67**	
	ATXN2_s[a]	-33.95	27.27	-1.24	.214	-87.58	19.69	
	ATXN2_l[b]	195.89	89.94	-2.18	.030	-372.80	-18.99	
	ATXN2_l2[b]	39.42	23.51	1.68	.094	-6.81	85.66	
	ATXN2_sl[b]	**46.42**	**11.08**	**4.19**	**<.001***	**24.62**	**68.22**	
	CACNA1A_s	76.64	23.93	3.20	.001*	29.59	123.69	$5.37*10^{-2}$
	CACNA1A_l	-43.38	37.83	-1.15	.252	-117.75	31.00	
	CACNA1A_s2	16.06	6.60	2.43	.015	3.09	29.04	
	CACNA1A_sl	**-33.74**	**9.94**	**-3.39**	**.001***	**-53.28**	**-14.20**	
	CACNA1A_s[b]	75.32	25.78	2.92	.004*	24.63	126.02	
	CACNA1A_l[b]	-36.80	40.46	-0.91	.346	-116.38	42.78	
	CACNA1A_s2[b]	17.04	6.92	2.46	.014	3.43	30.66	
	CACNA1A_sl[b]	**-37.61**	**10.23**	**-3.68**	**<.001***	**-57.74**	**-17.48**	

Table 4. (continued)

Outcome	Variable[a]	β-coefficient[d]	SE	t	p-value	95% CI		R^2
	AR_s ♀	-32.12	21.60	-1.49	.139	-74.76	10.52	$4.59*10^{-2}$
	AR_l ♀	12.00	19.57	0.61	.541	-26.63	50.63	
	AR_l2 ♀	**17.01**	**7.03**	**2.42**	**.017***	**3.13**	**30.88**	
	AR_s ♀[c]	-41.08	24.45	-1.68	.095	-89.39	7.23	
	AR_l ♀[c]	16.92	21.56	0.78	.434	-25.69	59.52	
	AR_l2 ♀[c]	**22.37**	**7.05**	**3.17**	**.002***	**8.44**	**36.30**	
	AR_s	**-31.62**	**11.06**	**-2.86**	**.004***	**-53.37**	**-9.88**	$2.00*10^{-2}$
	AR_s[b]	**-33.25**	**12.21**	**-2.72**	**.007***	**-57.27**	**-9.24**	
	AR_l	-26.32	10.95	-2.40	.017	-47.85	-4.79	$2.07*10^{-2}$
	AR_l2	**6.15**	**3.01**	**2.05**	**.042***	**$2.38*10^{-1}$**	**12.06**	
	AR_l[b]	-29.67	12.16	-2.44	.015	-53.60	-5.75	
	AR_l2[b]	**7.38**	**3.24**	**2.28**	**.023***	**1.00**	**13.75**	
Brain Stem								
	ATXN2_s	21.61	114.39	0.19	.850	-203.26	246.49	$1.06*10^{-2}$
	ATXN2_l	**-283.18**	**106.49**	**-2.66**	**.008***	**-492.52**	**-73.83**	
	ATXN2_s[b]	67.65	122.75	0.55	.582	-173.76	309.07	
	ATXN2_l[b]	**-277.09**	**123.48**	**-2.24**	**.025***	**-519.95**	**-34.23**	
	ATXN7_s	**-748.72**	**361.24**	**-2.07**	**.039***	**-1458.91**	**-38.52**	**$9.20*10^{-3}$**
	ATXN7_l	8.23	126.19	0.07	.948	-239.86	256.33	
	ATXN7_s[b]	**-1078.27**	**283.32**	**-3.81**	**<.001***	**-1635.55**	**-512.00**	
	ATXN7_l[b]	97.66	119.33	0.82	.414	-137.05	332.38	

s=relatively shorter allele. l=relatively longer allele. s2=quadratic term relatively shorter allele. l2=quadratic term relatively longer allele. sl=interaction term relatively shorter and longer allele. PDAGs= polyglutamine disease-associated genes, SE=standard error. CI=confidence interval. R^2 = estimated variance explained by the significant association. **Bold** indicates the variables that drive the significant association. [a]) Only the variables of PDAGs are reported in the cases where the omnibus test was statistically significant after multiple testing correction. [b]) Corrected for age, sex and population structure using principal components. [c]) Corrected for age and population structure using principal components. [d]) This column indicates the change in volume of the MRI variable per unit CAG repeat size increase. *) p-value statistically significant.

found that the CAG repeat variations in the shorter *AR* allele were associated with the rate of cognitive decline. Together the CAG repeat polymorphisms in these genes explained about 0.49% of the variance in cognitive function. Furthermore, we found that CAG repeat polymorphisms in *ATXN2*, *CACNA1A, ATXN7* and *AR* were significantly associated with the volume of the brain stem and several subcortical structures, including the amygdala, the thalamus and the putamen. Remarkably, the associations of *AR* repeat polymorphisms with cognitive function closely matched the associations between the *AR* repeat polymorphisms and MRI variables.

The maximum proportion of phenotypic variance in general cognition explained in three independent samples by a prediction score derived from data of 57 population-based cohorts evaluated in a meta-analysis, ranged between 2.63-4.31%.[10] In this study, we found that CAG repeat polymorphisms within the 'normal' range in the nine known PDAGs explained almost 0.49% of the variance in cognitive function. Although external validation of our results is warranted, this remarkable finding indicates the enormous potential of tandem repeat polymorphisms in the genetic research of cognitive function.

The polyglutamine stretch in *HTT* has been implicated previously as relevant for cognitive function. *Lee et al.* reported that in children aged 6-18 years, larger CAG repeat numbers within the 'normal' range in the longer *HTT* allele were associated with a higher general intelligence (IQ).[57] In addition, larger CAG repeat numbers in *HTT* were associated with increased grey matter volume in the globus pallidus.[58] Furthermore, organisms with a more complex central nervous system possess a higher number of repeats in *HTT*.[59,60] However, CAG repeat numbers in *HTT* beyond 35 repeats cause HD, which is characterized among other things by cognitive deterioration. These facts imply a curvilinear association between cognitive function and the number of CAG repeats in *HTT* with the suggestion of an optimal number of CAG repeats. Indeed, here we found that CAG repeat numbers in the longer *HTT* allele had a non-linear association with cognitive function in older adults with relatively small and large CAG repeat numbers being associated with a lower cognitive function, and the optimal CAG repeat number being around 22 repeats. This non-linear association has also been suggested by other groups.[61]

Comparable to *HTT*, the number of CAG repeats in the X-linked *AR* gene increases as organisms gain a more complex central nervous system throughout evolution.[62-64] In addition, very short CAG repeat sequences in *AR* were associated with very severe mental retardation.[65] However, in older men, larger CAG repeat numbers in *AR* were found to accelerate cognitive decline.[40,66] Thus, these findings suggest a complex association between cognitive function and CAG repeat variations in *AR*, which is reflected in our findings. CAG repeat variations in the longer *AR* allele had a non-linear association with cognitive function and the volumes of the amygdala, the putamen and the thalamus. Both relatively small and relatively large CAG repeat numbers were associated with

a higher cognitive function and larger volumes of the respective subcortical structures. The first part of this non-linear association is reflected in the association between the CAG repeat number in the shorter *AR* allele, and the rate of cognitive decline and the volume of the amygdala. Relatively short CAG repeat numbers in the shorter *AR* allele were associated with a slower decline in cognitive function and a larger volume of the amygdala. As the CAG repeat number in the shorter *AR* allele increased the rate of cognitive decline increased and the volume of the amygdala decreased, similar to the first part of the non-linear association between the longer *AR* allele and cognitive function. This complex association between cognitive function and CAG repeat variation over the entire 'normal' range in *AR* is intriguing and warrants more detailed investigation in future experiments.

CAG repeat numbers in *TBP* exceeding 44 repeats are associated with the development of spinocerebellar ataxia type 17 (SCA17), an autosomal dominant inherited neurodegenerative disease characterized by cerebellar ataxia, involuntary movements, psychiatric symptoms and cognitive decline eventually resulting in dementia.[67] Previously, we demonstrated that larger CAG repeat numbers in *TBP* within the 'normal' range were associated with a higher risk of lifetime depression [41]. This finding supports the notion that the cut-off of disease causing CAG repeats in SCA17 might not be as rigorous as previously suggested and that depressed individuals with a CAG repeat number within this large 'normal' CAG repeat range could have a less severe form of SCA17 characterized only by psychiatric symptoms. Here we found that larger CAG repeat numbers in the longer *TBP* allele were associated with a decreased cognitive function. Similarly, this finding suggests that the pathology caused by larger CAG repeat numbers in *TBP* operate in a more continuous and gradual fashion rather than on a dichotomous scale.

The ageing brain is characterized by several cellular and molecular changes, many of which overlap with pathways affected in polyglutamine disorders.[38,39] For instance, processes dysfunctional in HD, such as transcriptional abnormalities, dysregulation of the chaperone network, alterations of cellular protein degradation systems, mitochondrial deficits, unbalanced redox-homeostasis, and changes in axonal transport and synaptic function have all been linked to ageing.[38]. Aside from the polyglutamine disease spinocerebellar ataxia type 2, polymorphisms in *ATXN2* have also been associated with other neurodegenerative diseases, such as amyotrophic lateral sclerosis and progressive supranuclear palsy.[20,68,69] In addition, in a large GWAS, the *ATXN2* locus was associated with longevity.[70] Therefore, we can hypothesize that polymorphisms in *ATXN2* increase the susceptibility for certain neurodegenerative diseases and neurodegeneration in general through similar cellular pathways and perhaps also resulting in alterations in different cerebral structures, such as the volume of the brain stem, the putamen and the thalamus. Yeast strains lacking Sgf73 are exceptionally long-lived. Sgf73 is a yeast orthologue of ataxin-7 (ATXN7), the protein encoded by *ATXN7*.[71,72] ATXN7 is a member

of the TBP (TATA-binding protein)-free TAF (TBP-associated factor) complex (TFTC) and the SPT3/TAF9/GCN5 acetyltransferase (STAGA) complex, co-activators required for the transcription of RNA polymerase II-dependent genes.[72] We found that larger CAG repeat sequences in *ATXN7* were associated with a smaller volume of the brain stem. This finding suggests that polyglutamine elongations might very well affect ATXN7 function within TFTC and STAGA and consequently affect the volume of cerebral structures and longevity. Several studies have also shown that aged neurons suffer from an increased Ca^{2+} conduction.[73,74] *CACNA1A* encodes the α1A subunit of P/Q-type voltage-dependent calcium channel.[25] This subunit contains the pore forming structure of the calcium channel and is responsible for channel gating, permeability, and voltage dependent activation and inactivation.[75] Although the role of the polyglutamine domain in the function of the channel continues to be debated, the associations we found between the CAG repeat number in *CACNA1A* and the volume of the brain stem in the older adults, indicate a potential modifying role. Collectively, these findings illustrate the close similarities between characteristics of the ageing brain and the pathophysiology of polyglutamine diseases. Combined with the associations we found between the CAG repeat number in *ATXN2*, *CACNA1A*, *ATXN7*, *TBP*, *HTT* and *AR*, and cognitive function, the volume of the brain stem as well as several other subcortical structures in older adults, these findings implicate an important role for PDAGs in the normal ageing process of the human brain.

In total, we found 79 individuals with a CAG repeat number in the pathological range of at least one PDAG. None of the included participants had been diagnosed with a polyglutamine disease. Unfortunately, however, follow-up data on these participants was not available. Therefore, we were neither able to determine whether these participants would have developed a polyglutamine disease at a later time, nor distinguish between the cognitive decline as an early sign of a late-onset polyglutamine disease or part of 'normal' aging. However, the fact that we found the association between cognitive function in old age and CAG repeat polymorphisms in different PDAGs to extend well into the normal range, suggests that independent of polyglutamine disease development CAG repeat polymorphisms are associated with cognitive function in old age.

A limitation of our work is that we conducted our study in a large cohort containing participants from three European countries, a relatively heterogeneous population. To adjust for this, we corrected for potential population stratification. Results from heterogeneous cohorts can be generalized more easily. Nonetheless, to increase the robustness of our findings, additional research should be conducted in other study populations. Another limitation involves the fact that MRI data were only available for a subset of our cohort and the available data consisted solely of structural measurements. Unfortunately, we did not have diffusion tensor imaging data or functional measurements at our disposal, which could potentially have led to more insight into the biological foundation of the significant associations we found between CAG repeat polymorphisms and cognitive function.

In conclusion, we found that CAG repeat number variations in *TBP*, *HTT* and *AR* were significantly associated with cognitive function in older adults, jointly explaining nearly 0.49% of the variance. Furthermore, CAG repeat variations in *ATXN2*, *CACNA1A, ATXN7* and *AR* were associated with the volume of the brain stem and several subcortical structures. Our results demonstrate the importance of tandem repeat polymorphisms as novel, but hitherto underappreciated, modifiers of cognitive ageing and emphasize the role of PDAGs in healthy brain function.

ACKNOWLEDGEMENTS

We would like to express our gratitude to all individuals who participated in the PROSPER study as well as all research staff for collecting the data.

FUNDING

This study was supported by a VENI-grant (#91615080) from the Netherlands Organization of Scientific Research, a Marie Sklodowska-Curie Individual Fellowship grant from the European Union (Horizon 2020, #701130; NAA) and an investigator-initiated grant obtained from Bristol-Myers Squibb.

CONFLICT OF INTEREST

Dr. R.A.C. Roos reported being an advisor for UniQure. No other disclosures were reported.

REFERENCES

1. Cigolle CT, Langa KM, Kabeto MU, Tian Z, Blaum CS. Geriatric conditions and disability: the Health and Retirement Study. *Ann Intern Med* 2007; **147**(3): 156-64.
2. Blazer DG, Yaffe K, Karlawish J. Cognitive aging: a report from the Institute of Medicine. *Jama* 2015; **313**(21): 2121-2.
3. Pusswald G, Tropper E, Kryspin-Exner I, et al. Health-Related Quality of Life in Patients with Subjective Cognitive Decline and Mild Cognitive Impairment and its Relation to Activities of Daily Living. *J Alzheimers Dis* 2015; **47**(2): 479-86.
4. He W, Goodkind D, Kowal P. An Aging World: 2015. Washington, DC: U.S. Government Publishing Office; 2016.
5. Singer JJ, MacGregor AJ, Cherkas LF, Spector TD. Genetic influences on cognitive function using The Cambridge Neuropsychological Test Automated Battery. *Intelligence* 2006; (34): 421-8.
6. Deary IJ, Johnson W, Houlihan LM. Genetic foundations of human intelligence. *Hum Genet* 2009; **126**(1): 215-32.
7. Plomin R, DeFries JC, McClearn GE, McGuffin P. Behavioral Genetics 5th ed. New York: Worth Publishers; 2008.
8. Lee T, Henry JD, Trollor JN, Sachdev PS. Genetic influences on cognitive functions in the elderly: a selective review of twin studies. *Brain Res Rev* 2010; **64**(1): 1-13.
9. Lin CH, Lin E, Lane HY. Genetic Biomarkers on Age-Related Cognitive Decline. *Front Psychiatry* 2017; **8**: 247.
10. Davies G, Lam M, Harris SE, et al. Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nat Commun* 2018; **9**(1): 2098.
11. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001; **409**(6822): 860-921.
12. Hannan AJ. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. *Trends Genet* 2010; **26**(2): 59-65.
13. Hannan AJ. TRPing up the genome: Tandem repeat polymorphisms as dynamic sources of genetic variability in health and disease. *Discovery medicine* 2010; **10**(53): 314-21.
14. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. *Nat Rev Genet* 2018.
15. Mirkin SM. Expandable DNA repeats and human disease. *Nature* 2007; **447**(7147): 932-40.
16. Shao J, Diamond MI. Polyglutamine diseases: emerging concepts in pathogenesis and therapy. *Hum Mol Genet* 2007; **16 Spec No. 2**: R115-23.
17. Orr HT, Chung MY, Banfi S, et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nat Genet* 1993; **4**(3): 221-6.
18. Chung MY, Ranum LP, Duvick LA, Servadio A, Zoghbi HY, Orr HT. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nat Genet* 1993; **5**(3): 254-8.
19. Matilla T, Volpini V, Genis D, et al. Presymptomatic analysis of spinocerebellar ataxia type 1 (SCA1) via the expansion of the SCA1 CAG-repeat in a large pedigree displaying anticipation and parental male bias. *Hum Mol Genet* 1993; **2**(12): 2123-8.
20. Pulst SM, Nechiporuk A, Nechiporuk T, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nat Genet* 1996; **14**(3): 269-76.
21. Sanpei K, Takano H, Igarashi S, et al. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of repeat expansion and cloning technique, DIRECT. *Nat Genet* 1996; **14**(3): 277-84.

22. Giunti P, Sweeney MG, Harding AE. Detection of the Machado-Joseph disease/ spinocerebellar ataxia three trinucleotide repeat expansion in families with autosomal dominant motor disorders, including the Drew family of Walworth. *Brain* 1995; **118 (Pt 5)**: 1077-85.

23. Maciel P, Gaspar C, DeStefano AL, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *Am J Hum Genet* 1995; **57**(1): 54-61.

24. Maruyama H, Nakamura S, Matsuyama Z, et al. Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Hum Mol Genet* 1995; **4**(5): 807-12.

25. Zhuchenko O, Bailey J, Bonnen P, et al. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nat Genet* 1997; **15**(1): 62-9.

26. Matsuyama Z, Kawakami H, Maruyama H, et al. Molecular features of the CAG repeats of spinocerebellar ataxia 6 (SCA6). *Hum Mol Genet* 1997; **6**(8): 1283-7.

27. David G, Abbas N, Stevanin G, et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nat Genet* 1997; **17**(1): 65-70.

28. Benton CS, de Silva R, Rutledge SL, Bohlega S, Ashizawa T, Zoghbi HY. Molecular and clinical studies in SCA-7 define a broad clinical spectrum and the infantile phenotype. *Neurology* 1998; **51**(4): 1081-6.

29. Del-Favero J, Krols L, Michalik A, et al. Molecular genetic analysis of autosomal dominant cerebellar ataxia with retinal degeneration (ADCA type II) caused by CAG triplet repeat expansion. *Hum Mol Genet* 1998; **7**(2): 177-86.

30. Johansson J, Forsgren L, Sandgren O, Brice A, Holmgren G, Holmberg M. Expanded CAG repeats in Swedish spinocerebellar ataxia type 7 (SCA7) patients: effect of CAG repeat length on the clinical manifestation. *Hum Mol Genet* 1998; **7**(2): 171-6.

31. Nakamura K, Jeong SY, Uchihara T, et al. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Hum Mol Genet* 2001; **10**(14): 1441-8.

32. Koide R, Kobayashi S, Shimohata T, et al. A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Hum Mol Genet* 1999; **8**(11): 2047-53.

33. Nagafuchi S, Yanagisawa H, Sato K, et al. Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nat Genet* 1994; **6**(1): 14-8.

34. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 1991; **352**(6330): 77-9.

35. Bettencourt C, Hensman-Moss D, Flower M, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann Neurol* 2016; **79**(6): 983-90.

36. Fan HC, Ho LI, Chi CS, et al. Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell Transplant* 2014; **23**(4-5): 441-58.

37. Paulson HL, Shakkottai VG, Clark HB, Orr HT. Polyglutamine spinocerebellar ataxias - from genes to potential treatments. *Nat Rev Neurosci* 2017; **18**(10): 613-26.

38. Hands S, Sinadinos C, Wyttenbach A. Polyglutamine gene function and dysfunction in the ageing brain. *Biochim Biophys Acta* 2008; **1779**(8): 507-21.

39. Burke SN, Barnes CA. Neural plasticity in the ageing brain. *Nat Rev Neurosci* 2006; **7**(1): 30-40.

40. Gardiner SL, Harder AVE, Campman YJM, et al. Repeat length variations in ATXN1 and AR modify disease expression in Alzheimer's disease. *Neurobiol Aging* 2019; **73**: 230.e9-.e17.

41. Gardiner SL, van Belzen MJ, Boogaard MW, et al. Large normal-range TBP and ATXN7 CAG repeat lengths are associated with increased lifetime risk of depression. *Transl Psychiatry* 2017; **7**(6): e1143.

42. Gardiner SL, van Belzen MJ, Boogaard MW, et al. Huntingtin gene repeat size variations affect risk of lifetime depression. *Transl Psychiatry* 2017; **7**(12): 1277.

43. Stuitje G, van Belzen MJ, Gardiner SL, et al. Age of onset in Huntington's disease is influenced by CAG repeat variations in other polyglutamine disease-associated genes. *Brain* 2017; **140**(7): e42.

44. Trompet S, van Vliet P, de Craen AJ, et al. Pravastatin and cognitive function in the elderly. Results of the PROSPER study. *J Neurol* 2010; **257**(1): 85-90.

45. ten Dam VH, Box FM, de Craen AJ, et al. Lack of effect of pravastatin on cerebral blood flow or parenchymal volume loss in elderly at risk for vascular disease. *Stroke* 2005; **36**(8): 1633-6.

46. Shepherd J, Blauw GJ, Murphy MB, et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. *Lancet* 2002; **360**(9346): 1623-30.

47. Shepherd J, Blauw GJ, Murphy MB, et al. The design of a prospective study of Pravastatin in the Elderly at Risk (PROSPER). PROSPER Study Group. PROspective Study of Pravastatin in the Elderly at Risk. *Am J Cardiol* 1999; **84**(10): 1192-7.

48. Aziz NA, Jurgens CK, Landwehrmeyer GB, et al. Normal and mutant HTT interact to affect clinical severity and progression in Huntington disease. *Neurology* 2009; **73**(16): 1280-5.

49. Gardiner SL, de Mutsert R, Trompet S, et al. Repeat length variations in polyglutamine disease-associated genes affect body mass index. *Int J Obes (Lond)* 2018.

50. Gardiner SL, van Belzen MJ, Boogaard MW, et al. Huntingtin gene repeat size variations affect risk of lifetime depression. *Translational Psychiatry* 2017; **7**(12): 1277.

51. Gardiner SL, van Belzen MJ, Boogaard MW, et al. Large normal-range TBP and ATXN7 CAG repeat lengths are associated with increased lifetime risk of depression. *Transl Psychiatry* 2017; **7**: e1143.

52. Tezenas du MS, 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.

53. Benjamini Y, Hochberg Y. Controling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* 1995; **57**(1): 289-300.

54. Shiffman D, Trompet S, Louie JZ, et al. Genome-wide study of gene variants associated with differential cardiovascular event reduction by pravastatin therapy. *PLoS One* 2012; **7**(5): e38240.

55. Nakagawa S, Schielzeth H. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* 2013; **4**: 9.

56. Gardiner SL, Boogaard MW, Trompet S, et al. Prevalence of Carriers of Intermediate and Pathological Polyglutamine Disease-Associated Alleles Among Large Population-Based Cohorts. *JAMA neurology* 2019.

57. Lee JK, Ding Y, Conrad AL, et al. Sex-specific effects of the Huntington gene on normal neurodevelopment. *J Neurosci Res* 2017; **95**(1-2): 398-408.

58. Muhlau M, Winkelmann J, Rujescu D, et al. Variation within the Huntington's disease gene influences normal brain structure. *PLoS One* 2012; **7**(1): e29809.

59. Cattaneo E, Zuccato C, Tartari M. Normal huntingtin function: an alternative approach to Huntington's disease. *Nat Rev Neurosci* 2005; **6**(12): 919-30.

60. Tartari M, Gissi C, Lo Sardo V, et al. Phylogenetic comparison of huntingtin homologues reveals the appearance of a primitive polyQ in sea urchin. *Mol Biol Evol* 2008; **25**(2): 330-8.

61. Lee JK, Conrad A, Epping E, et al. Effect of Trinucleotide Repeats in the Huntington's Gene on Intelligence. *EBioMedicine* 2018.

62. Djian P, Hancock JM, Chana HS. Codon repeats in genes associated with human diseases: fewer repeats in the genes of nonhuman primates and nucleotide substitutions concentrated at the sites of reiteration. *Proc Natl Acad Sci U S A* 1996; **93**(1): 417-21.

63. Choong CS, Kemppainen JA, Wilson EM. Evolution of the primate androgen receptor: a structural basis for disease. *J Mol Evol* 1998; **47**(3): 334-42.

64. Hong KW, Hibino E, Takenaka O, et al. Comparison of androgen receptor CAG and GGN repeat length polymorphism in humans and apes. *Primates* 2006; **47**(3): 248-54.

65. Kooy RF, Reyniers E, Storm K, et al. CAG repeat contraction in the androgen receptor gene in three brothers with mental retardation. *Am J Med Genet* 1999; **85**(3): 209-13.

66. Yaffe K, Edwards ER, Lui LY, Zmuda JM, Ferrell RE, Cauley JA. Androgen receptor CAG repeat polymorphism is associated with cognitive function in older men. *Biol Psychiatry* 2003; **54**(9): 943-6.

67. Zuhlke C, Burk K. Spinocerebellar ataxia type 17 is caused by mutations in the TATA-box binding protein. *Cerebellum (London, England)* 2007; **6**(4): 300-7.

68. Sproviero W, Shatunov A, Stahl D, et al. ATXN2 trinucleotide repeat length correlates with risk of ALS. *Neurobiol Aging* 2017; **51**: 178.e1-.e9.

69. Ross OA, Rutherford NJ, Baker M, et al. Ataxin-2 repeat-length variation and neurodegeneration. *Hum Mol Genet* 2011; **20**(16): 3207-12.

70. Fortney K, Dobriban E, Garagnani P, et al. Genome-Wide Scan Informed by Age-Related Disease Identifies Loci for Exceptional Human Longevity. *PLoS genetics* 2015; **11**(12): e1005728.

71. McCormick MA, Mason AG, Guyenet SJ, et al. The SAGA histone deubiquitinase module controls yeast replicative lifespan via Sir2 interaction. *Cell Rep* 2014; **8**(2): 477-86.

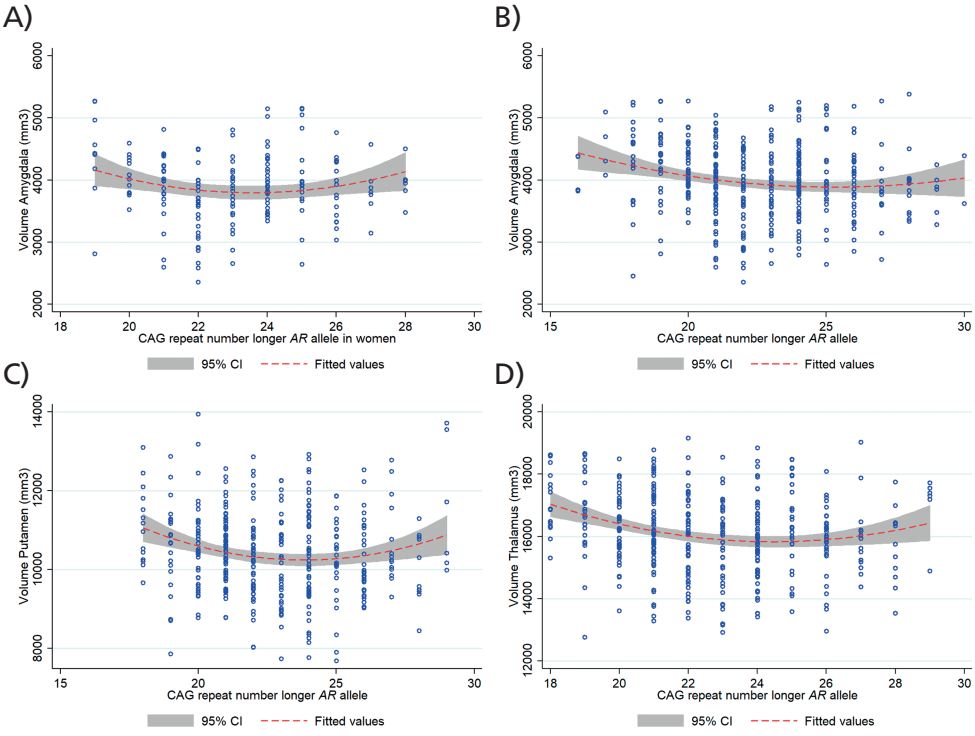
72. Helmlinger D, Hardy S, Eberlin A, Devys D, Tora L. Both normal and polyglutamine-expanded ataxin-7 are components of TFTC-type GCN5 histone acetyltransferase-containing complexes. *Biochem Soc Symp* 2006; (73): 155-63.

73. Toescu EC, Verkhratsky A, Landfield PW. Ca2+ regulation and gene expression in normal brain aging. *Trends Neurosci* 2004; **27**(10): 614-20.

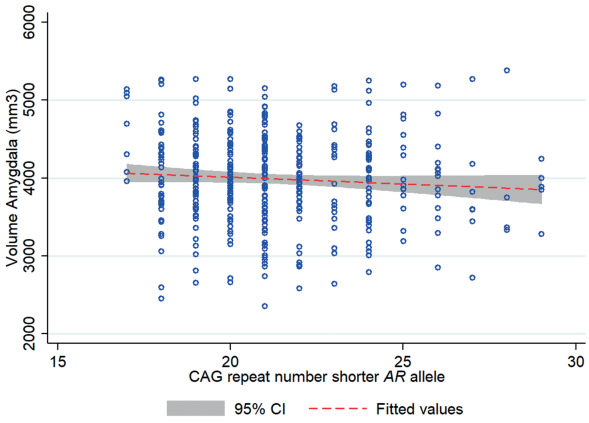
74. Foster TC, Norris CM. Age-associated changes in Ca(2+)-dependent processes: relation to hippocampal synaptic plasticity. *Hippocampus* 1997; **7**(6): 602-12.

75. Catterall WA. Structure and regulation of voltage-gated Ca2+ channels. *Annu Rev Cell Dev Biol* 2000; **16**: 521-55.

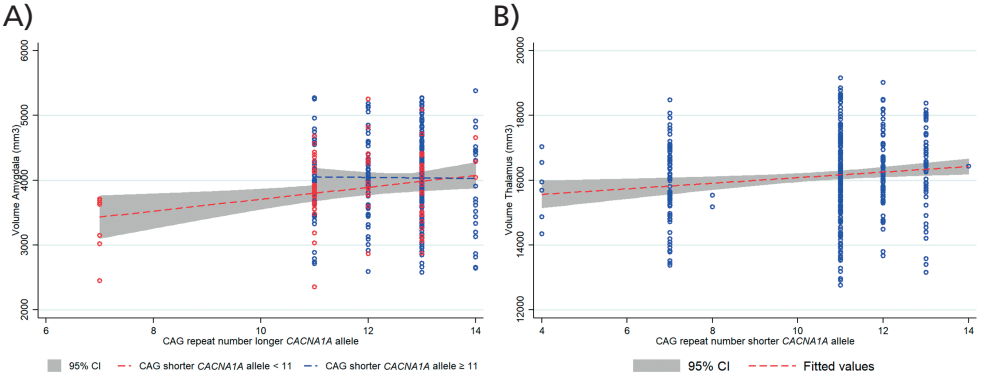
SUPPLEMENTARY MATERIAL



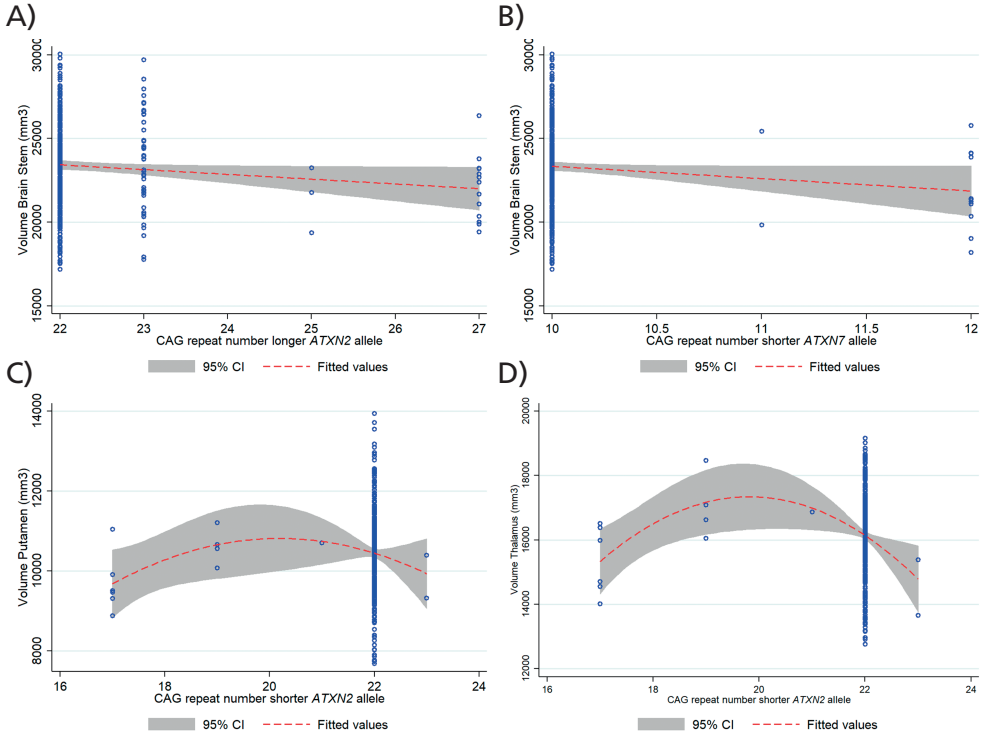
Supplemental Figure 1. The association between the volumes of the amygdala, the putamen and the thalamus, and the CAG repeat number in the longer *AR* allele. A. The volume of the amygdala had a non-linear association with the CAG repeat number in the longer *AR* allele in women. Both relatively small and relatively large CAG repeat numbers were associated with a larger volume. The volumes of the amygdala (**B**) the putamen (**C**) and the thalamus (**D**) all had a non-linear association with the longer *AR* allele in both sexes. Again, relatively small and relatively large CAG repeat numbers were associated with a larger volume.



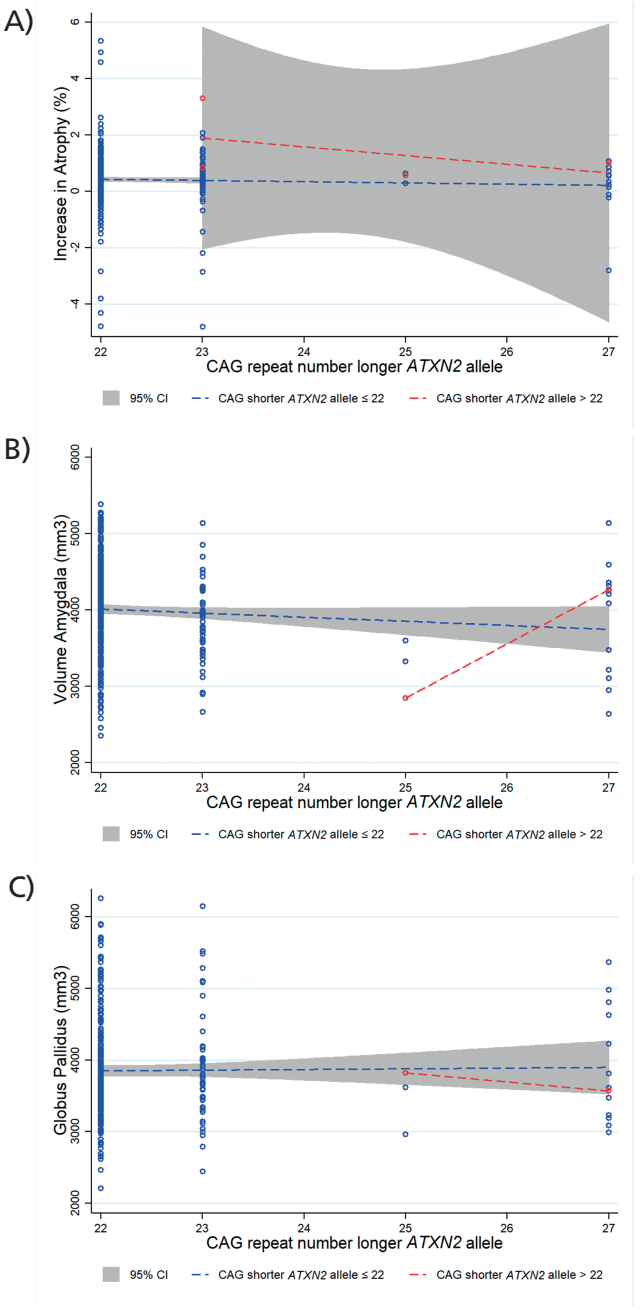
Supplemental Figure 2. Associations between the volume of the amygdala and the CAG repeat number in the shorter *AR* allele. The volume of the amygdala had an inverse association with the CAG repeat number in the shorter *AR* allele assessed in both sexes. Larger CAG repeat numbers were associated with a decreased volume.



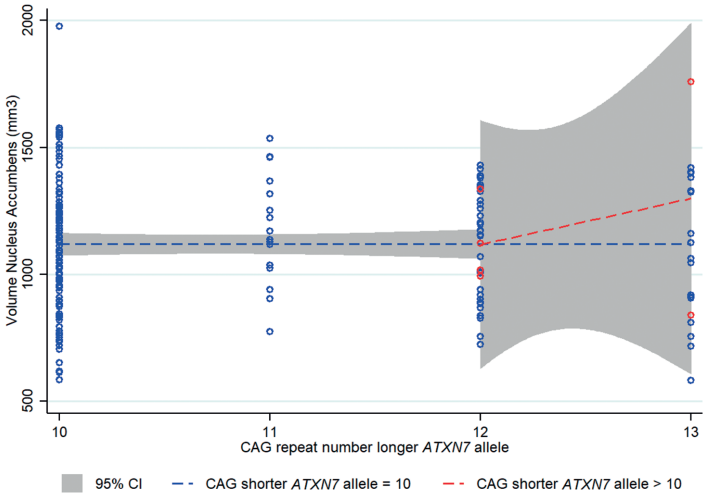
Supplemental Figure 3. The associations between the volume of the amygdala and the thalamus, and the CAG repeat number in *CACNA1A*. A. When the CAG repeat number in the shorter *CACNA1A* allele was smaller than 11 repeats, larger CAG repeat numbers in the longer *CACNA1A* allele were associated with an increased volume of the amygdala. This positive associations was no longer present, when the CAG repeat number in the shorter *CACNA1A* allele was equal to or larger than 11 repeats. **B.** Larger CAG repeat numbers in the shorter *CACNA1A* allele were associated with an increased thalamic volume.



Supplemental Figure 4. The associations between the volume of the brainstem, the pumtamen and the thalamus, and the CAG repeat number in *ATXN2* allele and the shorter *ATXN7* allele. **A.** Larger CAG repeat numbers in the longer *ATXN2* allele were associated with a smaller volume of the brain stem. **B.** Larger CAG repeat numbers in the shorter *ATXN7* allele were associated with a smaller volume of the brain stem. The volumes of the putamen (**C**) and the thalamus (**D**) both had a non-linear association with the CAG repeat number in the shorter *ATXN2* allele. Smaller and larger CAG repeat numbers were associated with smaller volumes of the respective subcortical structure.



Supplemental Figure 5. The association between the increase in atrophy per year, the volume of the amygdala and the volume of the globus pallidus, and the CAG repeat number in *ATXN2*. The interaction between both *ATXN2* alleles was associated with the increase in atrophy per year (**A**), the volume of the amygdala (**B**) and the volume of the globus pallidus (**C**). However, these associations seemed to be based on a few influential points and therefore, cannot be considered robust.



Supplemental Figure 6. The association between the volume of the nucleus accumbens and the CAG repeat number in *ATXN7*. The volume of the nucleus accumbens was significantly associated with the interaction between both *ATXN7* alleles. However, this association was driven by a few influential points and therefore, cannot be considered robust.