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

Gardiner, S.L.

Citation

Gardiner, S. L. (2019, October 10). *Huntington disease and other polyglutamine diseases : using CAG repeat variations to explain missing heritability*. Retrieved from https://hdl.handle.net/1887/79259

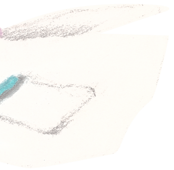
Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/79259

Note: To cite this publication please use the final published version (if applicable).

Cover Page

The handle http://hdl.handle.net/1887/79259 holds various files of this Leiden University dissertation.

Author: Gardiner, S.L.
Title: Huntington disease and other polyglutamine diseases : using CAG repeat variations to explain missing heritability
Issue Date: 2019-10-10



Chapter 3

COGNITION

Part 3.1

REPEAT LENGTH VARIATIONS IN *ATXN1* AND *AR* MODIFY DISEASE EXPRESSION IN ALZHEIMER DISEASE

Sarah L. Gardiner, MD[1,2];
Aster V.E. Harder, MD[1,3]; Yvonne J.M. Campman, MD[1];
Stella Trompet, PhD[4]; Jacobijn Gussekloo, MD, PhD[4,5],
Martine J. van Belzen, PhD[3]; Merel W. Boogaard[1,3];
Raymund A.C. Roos, MD, PhD[1]; Iris E. Jansen, PhD[6];
Yolande A.L. Pijnenburg, MD, PhD[6]; Philip Scheltens, MD, PhD[6];
Wiesje M. Van der Flier, PhD[6]; N. Ahmad Aziz, MD, PhD[1,7]

Departments of [1]Neurology, [2]Human Genetics,[3]Clinical Genetics, [4]Gerontology and Geriatrics, [5]Public Health and Primary Care, Leiden University Medical Centre, Leiden, the Netherlands
[6]Department of Neurology and Alzheimer Centre, VU University Medical Centre, Amsterdam, The Netherlands
[7]German Center for Neurodegenerative Diseases (DZNE), 53127 Bonn, Germany

Neurobiology of Aging. 2019; 73: 230.e9-.e17

ABSTRACT

Genome-wide association studies (GWAS) have contributed greatly to unravelling the genetic basis of Alzheimer disease (AD). However, a large amount of 'missing heritability' remains. In this exploratory study, we investigated the effect of CAG repeats in polyglutamine disease-associated genes (PDAGs) on the risk of AD and its expression. In a cohort of 959 patients diagnosed with AD (Amsterdam Dementia cohort) and 4106 cognitively healthy participants (Leiden 85-plus Study and the Prospective Study of Pravastatin in the Elderly at Risk), we determined the CAG repeat sequences in *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7, TBP*, *HTT*, *ATN1* and *AR*. We did not find a significant association between the risk of AD and variations in CAG repeat numbers of PDAGs. However, we found that differences in CAG repeat numbers in *ATXN1*, *ATXN2* and *AR* were significantly associated with several clinical and imaging features in AD patients. Specifically, the association between memory performance in AD patients and the CAG repeat size in the longer *ATXN1* allele, and the association between atrophy in the medial temporal lobes and the CAG repeat number in the longer *AR* allele remained significant after correction for multiple testing. Our findings suggest that repeat polymorphisms in *ATXN1* and *AR* can act as important genetic modifiers of AD, warranting further scrutiny of their role in its missing heritability and pathogenesis.

Keywords: Alzheimer disease; missing heritability; CAG repeat polymorphisms; polyglutamine diseases; Huntington disease

INTRODUCTION

Dementia currently affects 50 million people worldwide with its total global societal cost estimated at US$ 818 billion in 2015. By 2050 the number of individuals with dementia will have tripled with 152 million being affected. Together these facts indicate that dementia is an increasing burden on society which has prompted the World Health Organization (WHO) to recognize dementia as a public health priority.[1,2] Alzheimer disease (AD) is the most common cause of dementia, accounting for 60-70% of the cases. The overall heritability of AD is estimated to be as high as 60-80% and genome wide association studies (GWAS) have provided crucial information in understanding the genetic architecture of AD.[3,4] However, a relatively large proportion of the genetic determinants of AD remains to be elucidated, the so called 'missing heritability'.[5] A possible cause is that GWAS, aside from single-nucleotide polymorphisms (SNPs), cannot assess the contribution of other important genetic polymorphisms, especially DNA repeat variations. Three percent of the human genome consists of such tandem repeats,[6] a larger proportion than the entire protein coding sequences, which thus substantially contribute to genetic variation.[7-9]

Nine hereditary neurodegenerative diseases, known as polyglutamine diseases, including Huntington disease (HD), are the most prevalent disorders associated with DNA repeat variations.[10,11] 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**).[12-31] Polyglutamine diseases are characterized by progressive motor symptoms, psychiatric disturbances and cognitive deterioration. Increasing evidence implicates the polyglutamine domains of the associated proteins as critical regulators of fundamental homeostatic cellular processes such as transcriptional regulation, mitochondrial energy production and autophagy,[3,9,32] dysregulation of which has been associated with ageing and age-related disorders such as AD.[33-36] However, to what extent more common CAG repeat length variations in PDAGs are associated with the risk of AD and its clinical features is still unknown. As polyglutamine diseases are caused by expanded repeat sequences, we hypothesized that larger repeat numbers in these genes would be associated with a higher AD risk and more severe disease expression. In addition, we also assessed for more complex, non-linear effects of CAG repeat size variations as these have been reported in other neuropsychiatric disorders.[37,38] Therefore, in this exploratory study, we aimed to assess whether CAG repeat length variations in PDAGs 1) were associated with the risk of developing AD, and 2) could act as modifiers of clinical features of AD, including age of onset, cognitive functioning and brain atrophy.

SUBJECTS AND METHODS

Subjects

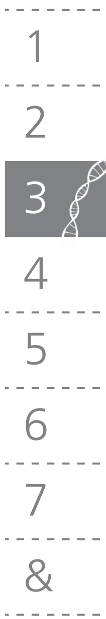
We genotyped the nine known PDAGs (including *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *HTT*, *ATN1* and *AR*) in participants with sufficient amounts of DNA available

Table 1. Number and distribution of genotyped polyglutamine disease-associated genes (PDAGs) per *Cohort*.

				Cohort 1 (cases)				Cohort 2[a] (controls)				Cohort 3[a] (controls)			
Gene	Disease	Protein	Allele	Mean±SD[b]	Median[b]	N	Range[b]	Mean±SD[b]	Median[b]	N	Range[b]	Mean±SD[b]	Median[b]	N	Range[b]
ATXN1	SCA1	Ataxin-1	short	29.17±1.2	29	952	16-36	29.33±0.8	29	189	27-32	29.22±1.0	29	3875	17-35
			long	30.78±1.7	30	952	26-38	30.74±1.6	30	189	29-37	30.81±1.7	30	3875	26-44
ATXN2	SCA2	Ataxin-2	short	21.95±0.6	22	954	15-27	21.96±0.5	22	190	17-23	21.95±0.6	22	3817	11-27
			long	22.38±1.1	22	954	22-31	22.53±1.5	22	190	22-31	22.43±1.2	22	3817	22-33
ATXN3	SCA3	Ataxin-3	short	19.12±4.5	21	954	12-32	19.07±4.6	21	157	14-28	18.99±4.4	20	3816	14-34
			long	24.58±3.7	23	954	14-40	24.05±3.7	23	157	14-30	24.29±3.7	23	3815	14-46
CACNA1A	SCA6	CACNA1A	short	10.62±2.1	11	954	4-14	10.73±2.1	11	188	4-14	10.64±2.1	11	3876	4-14
			long	12.46±1.1	13	954	7-17	12.47±1.2	13	188	7-14	12.49±1.1	13	3876	7-17
ATXN7	SCA7	Ataxin-7	short	10.01±0.6	10	930	6-13	10.12±0.7	10	150	7-14	10.06±0.5	10	3644	7-14
			long	10.83±1.2	10	930	10-17	10.89±1.3	10	150	10-14	10.77±1.2	10	3644	10-25
TBP	SCA17	TBP	short	36.23±1.8	36	953	27-40	36.27±2.1	37	192	27-38	36.25±1.9	37	3825	27-40
			long	37.80±1.0	38	953	32-42	37.85±0.9	38	192	35-42	37.86±1.0	38	3825	30-47
HTT	HD	Huntingtin	short	16.98±1.9	17	952	9-32	16.79±1.8	17	196	10-24	16.96±2.1	17	3853	9-29
			long	20.04±3.4	19	952	9-36	20.10±3.4	19	196	15-33	20.23±3.5	19	3853	12-38
ATN1	DRPLA	Atrophin-1	short	12.29±3.1	14	954	4-20	12.22±3.2	14	193	8-16	12.31±3.1	14	3874	5-20
			long	15.54±2.0	15	954	8-24	15.32±2.2	15	193	8-23	15.57±2.3	15.5	3874	8-27
AR	SBMA	Androgen receptor	short	21.01±2.8	21	949	9-32	21.03±2.7	21	191	12-31	21.20±2.7	21	3814	7-35
			long	22.77±2.9	23	949	9-33	23.03±2.9	23	191	18-36	22.83±2.8	23	3814	9-39

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. SD=standard deviation. [a]) Participants with an MMSE score > 27 were included (*Cohort* 2: n=205; *Cohort 3*: n=3901). [b]) The numbers reported indicate the number of CAG repeats.

from blood samples of three well-characterized cohorts: The Amsterdam Dementia Cohort (*Cohort 1*), The Leiden 85-plus Study (*Cohort 2*) and the Prospective Study of Pravastatin in the Elderly at Risk (PROSPER, *Cohort 3*).



Cohort 1

The Amsterdam Dementia Cohort is an ongoing memory clinic cohort now including over 5000 patients from VUmc Alzheimer center, the Netherlands.[39] From this cohort, a subset of 959 patients was selected based on a diagnosis of probable AD. Clinical diagnosis of probable AD was established by consensus in a multidisciplinary team, according to the National Institute on Aging and Alzheimer's Association (NIA-AA) criteria.[40] Global Cognition was estimated using the mini-mental state examination (MMSE) and the Cambridge Cognitive Examination (CAMCOG).[41-43] To evaluate memory function, the visual association test (VAT) and the total immediate recall and delayed recall of the Rey auditory verbal learning task (Dutch version) were applied.[44,45] Attention and executive functions were measured through the Trail Making Test (TMT) A and B as well as the Stroop (interference) test.[46,47] MRI images were obtained on 1.0 and 1.5 T MR systems (Siemens MAGNETOM Impact and Sonata, GE Healthcare Signa HDXT) and more recently on a 3T whole body MR systems scanner (MR750, GE Medical Systems, Milwaukee, WI, USA; Ingenuity TF PET/MR, Philips Medical Systems, Best, The Netherlands; Titan, Toshiba Medical Systems Japan). An experienced neuro-radiologist reviewed and scored all scans. Atrophy of the medial temporal lobes was rated on the coronal reconstructions of the T1-weighted MRI scans using a 5-point visual rating scale (0-4).[48] Posterior atrophy was rated on the combination of T1-weighted and FLAIR sequences using a 4-point visual rating scale (0-3), while global cortical atrophy was rated on FLAIR sequences using a 4-point visual rating scale (0-3).[49-51] The Medical Ethical Committee of the VU University Medical Centre approved the protocol and informed consent was obtained from all participants.[39]

Cohort 2 and Cohort 3

The Leiden 85-plus Study was a population-based prospective follow-up study among 599 85-year-old residents of Leiden, The Netherlands, who were recruited between September 1997 and September 1999.[52] The PROSPER originally was a prospective multicentre randomized placebo controlled trial to assess the effect of treatment with pravastatin on the risk of major vascular events among 5786 men and women between 70-83 years old with a pre-existing vascular disease or a raised risk for such a disease. Previous research showed that the treatment with pravastatin in this cohort did not affect cognitive function or brain atrophy. [53,54] PROSPER participants were recruited from three countries with 2517 individuals from Scotland, 2173 individuals from Ireland and 1096 individuals from the Netherlands. In both cohorts, global cognition was assessed using the MMSE.[41,42,55] Further details of both the study protocols have been described before.[52,56,57] We used all participants in *Cohort 2* and *Cohort 3* without a diagnosis of

dementia and with an MMSE > 27 as controls (*Cohort* 2: n=205; *Cohort 3*: n=3901), in this way including only participants with a healthy global cognition.[42] The Medical Ethical Committee of Leiden University Medical Centre approved the Leiden 85-plus study and the institutional ethics review boards of all involved centres approved the PROSPER study. All participants gave informed consent.[52,55]

Genotyping

To determine the CAG repeat length in the nine PDAGs for each included 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 68°C, preceded by five minutes of initial denaturation at 94°C. Final elongation was performed at 68°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

The association of an AD diagnosis and the CAG repeat length variations was assessed via multilevel mixed-effects logistic regression, using the participants in *Cohort 1* as cases and the selected participants in *Cohort 2* and *Cohort 3* as healthy controls. We chose this analysis method in order to correct for population stratification by incorporating country (i.e. The Netherlands, Scotland or Ireland) as a random effect. The association of the clinical and imaging features of AD with CAG repeat variations in PDAGs in *Cohort 1* was assessed using generalized linear regression. All clinical measurements were converted to Z-scores and per cognitive domain (i.e. global cognition, memory, attention and executive function) summary scores were calculated as average Z-scores. The summary scores of these domains and age of onset in AD were subsequently used as dependent variables in the generalized linear regression models. The summary score of global cognition consisted of the average Z-score of MMSE and CAMCOG. Memory was defined as the average Z-score of VAT and the average immediate and delayed recall of the Rey auditory verbal learning task. A summary score of attention included the average Z-scores of TMT A, the colour reading segment of the Stroop test (Stroop

I) and the word reading segment of the Stroop test (Stroop II), and executive function comprised the average Z-score of TMT B and the interference segment of the Stroop test (Stroop III). The data on the TMT and Stroop tests had a skewed distribution and thus were log-transformed and inverted as higher scores indicated a worse performance. Furthermore, atrophy of the medial temporal lobes (MTA) and the posterior lobes (PA) were defined as the average Z-scores of atrophy in the right and left lobe. Global atrophy in itself was also converted to a Z-score.

To assess whether CAG repeat variations in the PDAGs could explain a significant additional amount of variation in the respective dependent variable, we first performed an omnibus test per dependent variable (i.e. a likelihood ratio test for the risk of AD and a restricted F-test for age of onset, global cognition, memory, attention, executive function, global atrophy, MTA and PA) with the CAG repeat lengths in both alleles of all PDAGs as independent variables, including interaction and quadratic terms for both alleles at each locus to assess potential non-linear effects [58,59]. All continuous variables were centred around their respective means to reduce multicollinearity. Subsequently, only in cases where an omnibus test for the respective outcome was statistically significant (degrees of freedom = 45, false discovery rate (FDR) q-value < 0.05), [60] we performed a second omnibus test per PDAG, where we included as independent variables only the five terms associated with each locus (i.e. CAG repeat sizes in both alleles, their interaction and a quadratic term for each allele). Since *AR* is linked to the X-chromosome, we analysed the shorter and longer allele of *AR* in separate models. All models, except the models with age of onset as the outcome, included age, sex and education level as covariates. The model with age of onset as the outcome did not include age at assessment because age at assessment by definition occurs after the age of onset and, therefore, cannot be expected to have influenced age of onset in any biologically plausible way. The models with age of onset, atrophy or cognitive summary scores as outcome were additionally corrected for APOE genotype. Furthermore, the models with atrophy as outcome were also adjusted for duration of symptoms, while the models with cognitive summary scores were adjusted for both duration of symptoms and brain atrophy, defined as the average Z-score of all atrophy measurements. Only when the second omnibus test was nominally significant (degrees of freedom = 5, $p < 0.05$), we performed post-hoc tests by assessing the effect of the individual PDAG alleles and their associated higher-order terms. For these post-hoc tests we applied an FDR with a $q < 0.05$ considered to be statistically significant. Non-significant higher order terms were removed from this original model to arrive at a final model. For all final models including a significant PDAG effect, the R^2 was calculated to determine the amount of variance explained by each gene. 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 five were excluded (**Supplemental Table 2** and **3**). Due to

the exploratory nature of this study, we also report results that did not remain significant after correction for multiple testing as assessing these associations further could be of interest in future research. 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).

A post-hoc sensitivity analysis showed that, at a significance level of $\alpha = 0.0056$ (0.05/9), for each PDAG our sample size (4106 controls and 959 AD patients) would be sufficient to detect an odds ratio of ≥ 1.05 for the association between CAG repeat number and the risk of AD with a statistical power of 0.80. Similarly, under the same assumptions, our AD sample size would allow for the detection of an effect size (defined as Cohen's f) of at least 1.37×10^{-2} for the association between CAG repeat number in each PDAG and clinical/ imaging measures of AD (n=959) (calculated using G*Power version 3.1.9.2).[61]

RESULTS

AD risk was not associated with CAG repeat size

From the 959 cases in *Cohort 1*, 205 controls in *Cohort 2* and 3901 controls in *Cohort 3* we were able to determine the CAG repeat length of between 5021-4724 individuals for each gene. The lacking samples were due to too little available DNA material and were missing completely at random (112-409, **Table 1**). Between 0.0 to 0.4% of the cases per gene were excluded from the analyses because of CAG repeat numbers with a frequency lower than five (**Supplemental Table 2**). In *Cohort 1*, the patients were younger and had a lower average MMSE score compared to the participants in *Cohort 2* and *Cohort 3* (controls) (**Table 2**) In *Cohort 2*, more females participated and the level of education was lower compared to the other cohorts. The initial omnibus test assessing the association between the risk of AD and the CAG repeat variations in the PDAGs was not significant, indicating that the CAG repeat variations in the PDAGs did not explain any additional variation in the risk of AD ($p = .635$).

CAG repeat size variations and disease severity in AD

In AD patients, we assessed associations between CAG repeat sizes and both clinical and imaging features. The clinical and imaging scores significantly associated with the CAG repeat sizes in PDAGs after correction for potential confounders, including age, sex and level of education are displayed in **Table 3**. Only the association between memory and the CAG repeat number in the longer *ATXN1* allele, and the association between MTA and the CAG repeat number in the longer *AR* allele remained significant after correction for multiple testing and robust after visual inspection.

Table 2. Summary characteristics *Cohort 1* (cases), *Cohort 2* (controls) and *Cohort 3* (controls).

	Cohort1 (cases, n = 959)			Cohort 2 (controls, n = 205)[a]			Cohort 3 (controls, n = 802)[a]			
Variable	**mean±SD**	**median**	**range**	**mean±SD**	**median**	**range**	**mean±SD**	**median**	**range**	**p-value[b]**
Age	66.0±7.9	66.1	35.4-87.3	85	85	-	75.09±3.3	74.6	70.15-83.30	<.001*
MMSE score	20.2±5.1	21	2-30	28.8±0.8	29	28-30	28.90±0.8	29	28-30	<.001*
	number		**percent**	**number**		**percent**	**number**		**percent**	
Sex										.008*
male	443		46.2	75		36.6	1 939		48.9	
female	516		53.8	130		63.4	2 030		51.2	
Education										<.001*
low (ISCED 0-1)	9		0.9	92		44.9	80		2.0	
high (ISCED 2-6)	950		99.1	113		55.1	3889		98.0	

SD = standard deviation. MMSE = mini-mental state examination. ISCED = International Standard Classification of Education. [a]) Participants with an MMSE score > 27 were included (n=205). [b]) Tested with Kruskal-Wallis H Test. *) p-value statistically significant.

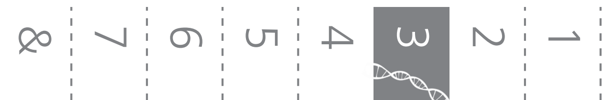


Table 3. Final corrected models significantly associated with clinical and imaging features of Alzheimer disease in *Cohort 1.*

Outcome	Variable	β-coefficient[a]	SE	t	p-value	95% CI		R^2
Clinical features								
Attention	ATXN1_s[b]	0.057	0.039	1.46	.145	-0.020	0.134	12.3*10^{-3}
	ATXN1_l[b]	-0.069	0.026	-2.68	.008	-0.119	-0.018	
	ATXN1_sl[b]	0.037	0.014	2.53	.012	0.008	0.065	
Memory	ATXN1_s[b]	0.041	0.034	1.19	.234	-0.026	0.107	9.3*10^{-3}
	ATXN1_l[b]	-0.064	0.021	-3.06	.002*	-0.104	-0.023	
Executive function	ATXN2_s[b]	0.612	0.112	5.48	<.001*	0.392	0.831	7.9*10^{-3}
	ATXN2_l[b]	0.077	0.052	1.47	.142	-0.026	0.179	
	ATXN2_s2[b]	0.161	0.028	5.70	<.001*	0.106	0.217	
	ATXN2_sl[b]	-0.270	0.067	-4.04	<.001*	-0.401	-0.138	
Imaging features								
MTA	AR_s[c]	0.030	0.013	2.42	.016	0.006	0.055	6.0*10^{-3}
	AR_l[c]	0.031	0.011	2.66	.008*	0.008	0.053	7.8*10^{-3}

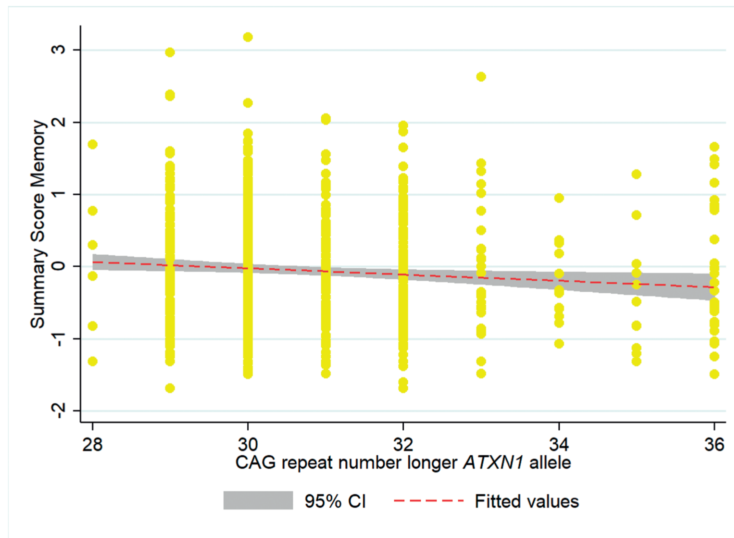
PDAG = polyglutamine disease-associated gene. s = relatively shorter allele. l = relatively longer allele. s2 = quadratic term relatively shorter allele. sl = interaction term relatively shorter and longer allele. SE = standard error. CI = confidence interval. R^2 = estimated variation explained by the significant association. *AR* (short) = the relatively shorter X-linked *AR* allele assessed in both men and women. *AR* (long) = the relatively longer X-linked *AR* allele assessed in both men and women. [a]) Increase in summary score per CAG repeat number increase. [b]) Estimates corrected for age, sex, level of education, duration of AD symptoms, atrophy summary score and APOE genotype. [c]) Estimates corrected for age, sex, level of education, duration of AD symptoms and APOE genotype.[*]) p-value statistically significant after correction for multiple comparisons.

Age of onset

Although the omnibus tests indicated that age of onset in AD was significantly related to *ATXN2* and *TBP* CAG repeat sizes, the individual coefficients of neither gene were statistically significant (**Supplemental Table 4**), potentially indicating an effect size too small to be detected with our sample size.

Memory, attention, executive and global cognitive function

We found that larger CAG repeat numbers in the longer *ATXN1* allele were significantly associated with worse memory (**Figure 1**). This association remained significant and robust after correction for multiple testing, explaining almost 1% of the variation in memory among AD patients. Although attention also appeared to be significantly associated with the interaction between the shorter and longer *ATXN1* alleles CAG repeat size (**Supplemental Figure 1**), this association did not remain significant after correction for multiple testing. For *ATXN2*, the models indicated a significant effect on executive



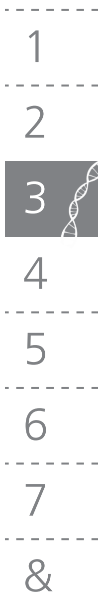


Figure 1. The association between the summary score of memory and the CAG repeat number in the longer *ATXN1* allele. The summary score of memory was significantly associated with the CAG repeat number in the longer *ATXN1* allele. Larger CAG repeat numbers in the longer *ATXN1* allele are associated with a decreased summary score of memory. This association remained significant after correction for age, sex, level of education, duration of AD symptoms and summary score atrophy (β = -0.063 ± 0.021, 95% CI -0.104 to -0.022, p-value = .002). This association remained significant after correction for multiple testing.

function, however, visual inspection of this association showed that this relation was mainly driven by a few influential points and, thus, unlikely to be robust (**Supplemental Figure 2**). Lastly, the initial omnibus test with global cognitive function as dependent variable was not significant (p = .076).

Brain atrophy

Global brain atrophy was not significantly associated with CAG repeat size in any of the PDAGs (all p > .066). In contrast, larger CAG repeat sizes in either of the *AR* alleles were significantly associated with more severe MTA (**Figure 2A** and **B**). The association between MTA and the CAG repeat number in the longer *AR* allele remained significant after correction for multiple testing and explained almost 0.8% of the variation in MTA among AD patients. The initial omnibus test with PA as dependent variable was not significant (p = .103).

DISCUSSION

Our study is the first to investigate the effect of CAG repeat size variations in PDAGs on AD risk and its clinical and imaging features. We did not find a significant association between the risk of AD and the CAG repeat number in any of the PDAGs. However, CAG repeat numbers in *ATXN1* and *AR* had robust associations with different disease features

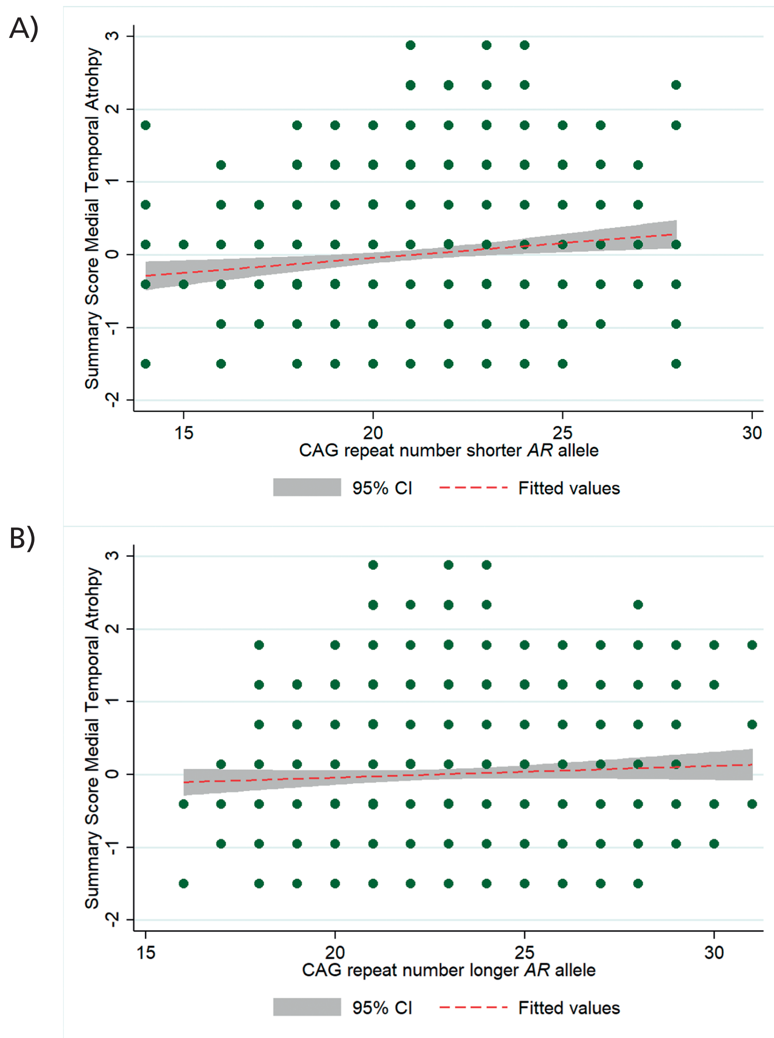


Figure 2. The association between the summary score of atrophy in the medial temporal lobes and the CAG repeat number in the shorter and longer *AR* alleles. Both larger CAG repeat numbers in the shorter *AR* allele (**A**) and the longer *AR* allele (**B**) were associated with higher summary scores of medial temporal atrophy. These associations remained significant after correction for age, sex, level of education and duration of AD symptoms; short: $\beta = 0.029 \pm 0.013$, 95% CI 0.003 to 0.056, p-value = .016, long: $\beta = 0.031 \pm 0.011$, 95% CI 0.008 to 099.053, p-value = .008). After correction for multiple testing, the association between medial temporal atrophy and the CAG repeat number in the longer *AR* allele remained significant.

in AD, including memory, attention and MTA. Although only the association between memory and the CAG repeat number in the longer *ATXN1* allele, and the association between MTA and the CAG repeat number in the longer *AR* allele remained significant after correction for multiple testing, our findings suggest that the role of these repeat polymorphisms as complex genetic modifiers of AD merits further scrutiny.

Individuals with CAG repeat sequences of 39 or more in *ATXN1* suffer from the hereditary neurodegenerative disorder spinocerebellar ataxia type 1 (SCA1). SCA1 is characterized by cerebellar ataxia, dysarthria, bulbar degeneration and in advanced stages also cognitive impairment.[62] Reported impaired cognitive domains include executive function, visuospatial thinking, attention and memory.[63-67] Our results parallel these cognitive symptoms in SCA1 patients as larger CAG repeat numbers in the longer *ATXN1* allele were associated with worse memory and attention in AD patients as well. Furthermore, a GWAS study identified a single-nucleotide polymorphism in an intron of *ATXN1* as a susceptibility marker for AD. Subsequent functional analysis studies demonstrated that *ATXN1* loss of function potentiates β-amyloid precursor protein (APP) processing and subsequently increases levels of both Aβ40 and Aβ42,[68] that can lead to learning and neurobehavioral abnormalities.[69,70] Although the contribution of the *ATXN1* polyglutamine domain to this process is unknown, we hypothesize that elongation of the polyglutamine tract could contribute to a loss-of-function effect of the ataxin-1 protein (ATXN1) (**Figure 3A**). Supporting this notion is the fact that an elongated *ATXN1* leads to loss of the ATXN1-CIC complex, which in turn is associated with learning and memory deficits in

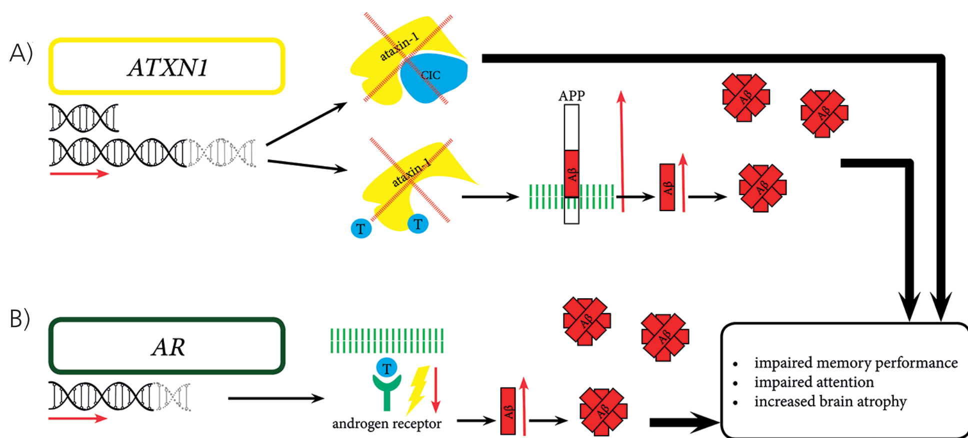


Figure 3. A model of how increased numbers of CAG repeats in *ATXN1* and *AR* could potentially affect the pathology of Alzheimer disease. A. An increased number of CAG repeats in the longer allele of *ATXN1* is known to lead to a loss of the ATXN1-CIC complex, which in turn was associated with memory deficits in mice and humans.[71,72] In addition, larger CAG repeat numbers in *ATXN1* impair the function of the ataxin-1 protein, which has been associated with potentiation of the β-amyloid precursor protein (APP) and the subsequent increase in levels of β-amyloid (Aβ) that can lead to the formation of Aβ plaques, and learning and neurobehavioral abnormalities.[68,70,80] **B.** Larger CAG repeat numbers in *AR* result in a loss of androgen receptor sensitivity and decreased transcriptional activity of androgen responsive genes. Since decreased levels of androgen were associated with increases in Aβ accumulation, we hypothesize that this decrease in androgen receptor sensitivity could also lead to an increase in Aβ, the subsequent formation of plaques and an increase in imaging features of Alzheimer disease.[76]

mice and humans.[71,72] Collectively, these results suggest that larger CAG repeat sequences within the normal range of *ATXN1* could affect and worsen clinical features of AD.

The progressive neuromuscular disorder spinal and bulbar muscular atrophy (SBMA) or Kennedy's disease is caused by a CAG repeat sequence longer than 35 repeats in the X-linked *AR* gene. *AR* encodes the androgen receptor (AR). Longer CAG repeat sequences in *AR* are associated with an earlier onset of motor-symptoms and differences in disease manifestations.[73-75] In addition, CAG repeat polymorphisms in exon 1 of the *AR* gene modulate androgen sensitivity and elongated repeat sequences result in decreased androgen receptor action and decreased transcriptional activity of androgen responsive genes. Androgens have been shown to influence brain β-amyloid levels and deposition. Decreased androgen levels were associated with region-specific increases in β-amyloid accumulation in mice and decreased testosterone levels were found in male AD brains.[76-78] In accordance with these findings, we found that a higher CAG repeat number in *AR* was associated with more MTA in patients with AD, indicating that larger CAG repeat numbers in *AR* and the resulting decline in sensitivity of the androgen receptor for androgens might lead to an increased deposition of β-amyloid and thus more neuronal cell death and brain atrophy (**Figure 3B**).

As our results indicate, the associations between characteristics of AD and the CAG repeat polymorphisms in the different PDAGs were complex. Larger CAG repeat numbers in *ATXN1* led to a worse memory, while the interaction between both *ATXN1* alleles affected attention. Furthermore, CAG repeat numbers in both *AR* alleles had an inverse association with MTA. However, no associations between the clinical characteristics of AD and the CAG repeat polymorphisms in *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *HTT* and *ATN1* were found. The variation in types of association or the absence of an association could be due to the subtle and complex effects that the polyglutamine stretch has on the respective genes and proteins. Variations in CAG repeat number can change associated protein properties, such as flexibility and binding affinity. In addition, these stretches can alter the local DNA structure and transcription activity. These effects are very likely to be both gene and protein context specific. [79] However, much is still unknown about the specific influence of repeat polymorphisms on normal protein transcription and function. Therefore, studies investigating the precise molecular effects of CAG repeat polymorphisms within the normal range are needed to pinpoint the exact pathophysiological mechanisms involved.

The major limitations of our study were the relatively small number of subjects included in *Cohort 1* (n=959). Therefore, our study likely lacked sufficient statistical power to detect associations between the clinical and imaging features of AD and the different CAG repeat polymorphisms that would remain significant after correction for multiple testing.

In order to obtain more robust results, future studies in larger cohorts of patients with AD are warranted.

In conclusion, we were unable to find a significant association between the risk of AD and CAG repeat numbers in the PDAGs. However, we found tentative evidence that CAG repeat polymorphism in *ATXN1* and *AR* can act as complex genetic modifiers of AD phenotype. Specifically, larger CAG repeat numbers in the longer *ATXN1* allele were associated with a worse memory and larger CAG repeat numbers in the longer *AR* allele were associated with more severe MTA, both findings remaining significant after correction for multiple testing and accounting for almost 1% and 0.8% of the variation in memory and atrophy among AD patients, respectively. Therefore, further research in larger cohorts is warranted to delineate the role of these tandem repeat polymorphisms in the pathogenesis of AD and their contribution to its missing heritability.

ACKNOWLEDGMENTS

The authors are grateful to all individuals who participated in the Amsterdam Dementia Cohort, the Leiden 85-plus Study and the PROSPER study as well as all research staff for collecting the data. Research of the VUmc Alzheimer center is part of the neurodegeneration research program of the Amsterdam Neuroscience. The VUmc Alzheimer center is supported by Alzheimer Nederland and Stichting VUmc fonds. The clinical database structure was developed with funding from Stichting Dioraphte. NAA was supported by a VENI-grant (#91615080) from the Netherlands Organization of Scientific Research and a Marie Sklodowska-Curie Individual Fellowship grant from the European Union (Horizon 2020, #701130; NAA). The Leiden 85-plus Study was funded in part by an unrestricted grant from the Dutch Ministry of Health, Welfare and Sports.

CONFLICT OF INTEREST

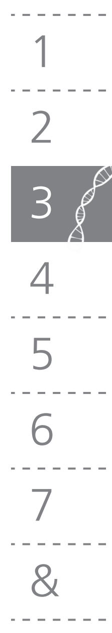
The authors declare no conflict of interest.

REFERENCES

1. Dementia Fact sheet. September 2017. http://www.who.int/mediacentre/factsheets/fs362/en/.
2. Prince M, Wimo A, Guerchet M, et al. World Alzheimer report 2015: the global impact of dementia, an analysis of prevalence, incidence, costs and trends, 2015.
3. Lee SH, Harold D, Nyholt DR, et al. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Human molecular genetics* 2013; **22**(4): 832-41.
4. Ridge PG, Mukherjee S, Crane PK, Kauwe JS. Alzheimer's disease: analyzing the missing heritability. *PloS one* 2013; **8**(11): e79771.
5. Gatz M, Reynolds CA, Fratiglioni L, et al. Role of genes and environments for explaining Alzheimer disease. *Archives of general psychiatry* 2006; **63**(2): 168-74.
6. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001; **409**(6822): 860-921.
7. Hannan AJ. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. *Trends in genetics : TIG* 2010; **26**(2): 59-65.
8. 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.
9. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. *Nature reviews Genetics* 2018.
10. Mirkin SM. Expandable DNA repeats and human disease. *Nature* 2007; **447**(7147): 932-40.
11. Shao J, Diamond MI. Polyglutamine diseases: emerging concepts in pathogenesis and therapy. *Human molecular genetics* 2007; **16 Spec No. 2**: R115-23.
12. Orr HT, Chung MY, Banfi S, et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature genetics* 1993; **4**(3): 221-6.
13. 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. *Nature genetics* 1993; **5**(3): 254-8.
14. 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. *Human molecular genetics* 1993; **2**(12): 2123-8.
15. Pulst SM, Nechiporuk A, Nechiporuk T, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nature genetics* 1996; **14**(3): 269-76.
16. 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. *Nature genetics* 1996; **14**(3): 277-84.
17. 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 : a journal of neurology* 1995; **118 (Pt 5)**: 1077-85.
18. Maciel P, Gaspar C, DeStefano AL, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *American journal of human genetics* 1995; **57**(1): 54-61.
19. Maruyama H, Nakamura S, Matsuyama Z, et al. Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Human molecular genetics* 1995; **4**(5): 807-12.
20. 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. *Nature genetics* 1997; **15**(1): 62-9.

21. Matsuyama Z, Kawakami H, Maruyama H, et al. Molecular features of the CAG repeats of spinocerebellar ataxia 6 (SCA6). *Human molecular genetics* 1997; **6**(8): 1283-7.
22. David G, Abbas N, Stevanin G, et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nature genetics* 1997; **17**(1): 65-70.
23. 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.
24. 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. *Human molecular genetics* 1998; **7**(2): 177-86.
25. 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. *Human molecular genetics* 1998; **7**(2): 171-6.
26. Nakamura K, Jeong SY, Uchihara T, et al. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Human molecular genetics* 2001; **10**(14): 1441-8.
27. 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? *Human molecular genetics* 1999; **8**(11): 2047-53.
28. Nagafuchi S, Yanagisawa H, Sato K, et al. Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nature genetics* 1994; **6**(1): 14-8.
29. 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.
30. Bettencourt C, Hensman-Moss D, Flower M, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Annals of neurology* 2016; **79**(6): 983-90.
31. Fan HC, Ho LI, Chi CS, et al. Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell transplantation* 2014; **23**(4-5): 441-58.
32. Ashkenazi A, Bento CF, Ricketts T, et al. Polyglutamine tracts regulate beclin 1-dependent autophagy. *Nature* 2017; **545**(7652): 108-11.
33. Gardner A, Boles RG. Beyond the serotonin hypothesis: mitochondria, inflammation and neurodegeneration in major depression and affective spectrum disorders. *Progress in neuro-psychopharmacology & biological psychiatry* 2011; **35**(3): 730-43.
34. Klempan TA, Sequeira A, Canetti L, et al. Altered expression of genes involved in ATP biosynthesis and GABAergic neurotransmission in the ventral prefrontal cortex of suicides with and without major depression. *Molecular psychiatry* 2009; **14**(2): 175-89.
35. Caldeira GL, Ferreira IL, Rego AC. Impaired transcription in Alzheimer's disease: key role in mitochondrial dysfunction and oxidative stress. *Journal of Alzheimer's disease : JAD* 2013; **34**(1): 115-31.
36. Spiegelman BM. Transcriptional control of mitochondrial energy metabolism through the PGC1 coactivators. *Novartis Foundation symposium* 2007; **287**: 60-3; discussion 3-9.
37. 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. *Translational psychiatry* 2017; **7**(6): e1143.
38. 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.



39. van der Flier WM, Pijnenburg YA, Prins N, et al. Optimizing patient care and research: the Amsterdam Dementia Cohort. *Journal of Alzheimer's disease : JAD* 2014; **41**(1): 313-27.

40. Arber C, Precious SV, Cambray S, et al. Activin A directs striatal projection neuron differentiation of human pluripotent stem cells. *Development* 2015; **142**(7): 1375-86.

41. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research* 1975; **12**(3): 189-98.

42. Tombaugh TN, McIntyre NJ. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society* 1992; **40**(9): 922-35.

43. Derix MM, Hofstede AB, Teunisse S, et al. [CAMDEX-N: the Dutch version of the Cambridge Examination for Mental Disorders of the Elderly with automatic data processing]. *Tijdschrift voor gerontologie en geriatrie* 1991; **22**(4): 143-50.

44. Rey A. L'Examen Clinique En Psychologie; 1958.

45. Lindeboom J, Schmand B, Tulner L, Walstra G, Jonker C. Visual association test to detect early dementia of the Alzheimer type. *Journal of neurology, neurosurgery, and psychiatry* 2002; **73**(2): 126-33.

46. Reitman RM. Validity of the Trail Making Test as an indicator of organic brain damage. *Percept Mot Skills* 1958; **8**: 271-6.

47. Ridley Stroop J. Studies of Interference in Serial Verbal Reactions; 1992.

48. Scheltens P, Leys D, Barkhof F, et al. Atrophy of medial temporal lobes on MRI in "probable" Alzheimer's disease and normal ageing: diagnostic value and neuropsychological correlates. *Journal of neurology, neurosurgery, and psychiatry* 1992; **55**(10): 967-72.

49. Lehmann M, Koedam EL, Barnes J, et al. Posterior cerebral atrophy in the absence of medial temporal lobe atrophy in pathologically-confirmed Alzheimer's disease. *Neurobiology of aging* 2012; **33**(3): 627.e1-.e12.

50. Koedam EL, Lehmann M, van der Flier WM, et al. Visual assessment of posterior atrophy development of a MRI rating scale. *European radiology* 2011; **21**(12): 2618-25.

51. Pasquier F, Leys D, Weerts JG, Mounier-Vehier F, Barkhof F, Scheltens P. Inter- and intraobserver reproducibility of cerebral atrophy assessment on MRI scans with hemispheric infarcts. *European neurology* 1996; **36**(5): 268-72.

52. der Wiel AB, van Exel E, de Craen AJ, et al. A high response is not essential to prevent selection bias: results from the Leiden 85-plus study. *Journal of clinical epidemiology* 2002; **55**(11): 1119-25.

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

54. 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.

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

56. den Elzen WP, Willems JM, Westendorp RG, et al. Effect of erythropoietin levels on mortality in old age: the Leiden 85-plus Study. *CMAJ* 2010; **182**(18): 1953-8.

57. 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. *The American journal of cardiology* 1999; **84**(10): 1192-7.

58. 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. *Translational psychiatry* 2017; **7**: e1143.

59. 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.

60. 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.

61. Faul F, Erdfelder E, Lang AG, Buchner A. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 2007; **39**(2): 175-91.

62. Opal P, Ashizawa T. Spinocerebellar Ataxia Type 1. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. GeneReviews((R)). Seattle (WA): University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.

63. Klinke I, Minnerop M, Schmitz-Hubsch T, et al. Neuropsychological features of patients with spinocerebellar ataxia (SCA) types 1, 2, 3, and 6. *Cerebellum (London, England)* 2010; **9**(3): 433-42.

64. Fancellu R, Paridi D, Tomasello C, et al. Longitudinal study of cognitive and psychiatric functions in spinocerebellar ataxia types 1 and 2. *Journal of neurology* 2013; **260**(12): 3134-43.

65. Ma J, Wu C, Lei J, Zhang X. Cognitive impairments in patients with spinocerebellar ataxia types 1, 2 and 3 are positively correlated to the clinical severity of ataxia symptoms. *International journal of clinical and experimental medicine* 2014; **7**(12): 5765-71.

66. Moriarty A, Cook A, Hunt H, Adams ME, Cipolotti L, Giunti P. A longitudinal investigation into cognition and disease progression in spinocerebellar ataxia types 1, 2, 3, 6, and 7. *Orphanet journal of rare diseases* 2016; **11**(1): 82.

67. Burk K. Cognition in hereditary ataxia. *Cerebellum (London, England)* 2007; **6**(3): 280-6.

68. Zhang C, Browne A, Child D, Divito JR, Stevenson JA, Tanzi RE. Loss of function of ATXN1 increases amyloid beta-protein levels by potentiating beta-secretase processing of beta-amyloid precursor protein. *The Journal of biological chemistry* 2010; **285**(12): 8515-26.

69. Bertram L, Tanzi RE. Thirty years of Alzheimer's disease genetics: the implications of systematic meta-analyses. *Nature reviews Neuroscience* 2008; **9**(10): 768-78.

70. Lesne S, Koh MT, Kotilinek L, et al. A specific amyloid-beta protein assembly in the brain impairs memory. *Nature* 2006; **440**(7082): 352-7.

71. Lu HC, Tan Q, Rousseaux MW, et al. Disruption of the ATXN1-CIC complex causes a spectrum of neurobehavioral phenotypes in mice and humans. *Nature genetics* 2017; **49**(4): 527-36.

72. Lim J, Crespo-Barreto J, Jafar-Nejad P, et al. Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. *Nature* 2008; **452**(7188): 713-8.

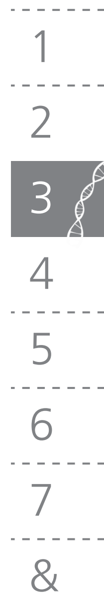
73. Igarashi S, Tanno Y, Onodera O, et al. Strong correlation between the number of CAG repeats in androgen receptor genes and the clinical onset of features of spinal and bulbar muscular atrophy. *Neurology* 1992; **42**(12): 2300-2.

74. Nishiyama A, Sugeno N, Tateyama M, Nishiyama S, Kato M, Aoki M. Postural leg tremor in X-linked spinal and bulbar muscular atrophy. *Journal of clinical neuroscience : official journal of the Neurosurgical Society of Australasia* 2014; **21**(5): 799-802.

75. Finsterer J, Soraru G. Onset Manifestations of Spinal and Bulbar Muscular Atrophy (Kennedy's Disease). *Journal of molecular neuroscience : MN* 2016; **58**(3): 321-9.

76. Carroll JC, Rosario ER, Kreimer S, et al. Sex differences in beta-amyloid accumulation in 3xTg-AD mice: role of neonatal sex steroid hormone exposure. *Brain research* 2010; **1366**: 233-45.

77. Rosario ER, Chang L, Head EH, Stanczyk FZ, Pike CJ. Brain levels of sex steroid hormones in men and women during normal aging



and in Alzheimer's disease. *Neurobiology of aging* 2011; **32**(4): 604-13.

78. Rosario ER, Chang L, Stanczyk FZ, Pike CJ. Age-related testosterone depletion and the development of Alzheimer disease. *Jama* 2004; **292**(12): 1431-2.

79. Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. *Trends in genetics : TIG* 2006; **22**(5): 253-9.

80. Bertram L, Lange C, Mullin K, et al. Genome-wide association analysis reveals putative Alzheimer's disease susceptibility loci in addition to APOE. *American journal of human genetics* 2008; **83**(5): 623-32.

Supplemental Table 1. Polyglutamine disease associated genes and primers.

Gene	Primer 1 (forward)	Primer 2 (reverse)	Mix
ATXN1	CCCCAACCGCCAACCCC	GTGGGATCATCGTCTGGTGGG	A
ATXN2	CGTTCCGGCGTCTCCTTGG	ACCGAGGAGGGAGCCGT	B
ATXN3	GTAATCTGTATCAGACTAACTGCTCTTG	GGGAATGAAGAATAATGTAAAGCAAAAATCAC	B
CACNA1A	CGTGTCCTATTCCCCTGTGATCC	CCTGGGTACCTCCGAGG	A
ATXN7	GAACTCCCTGGCGCCTCC	GATTCCACGACTGTCCCAGCAT	A
TBP	CCACAGCCTATTCAGAACACC	TGGGACGTTGACTGCTGAAC	B
HTT	ATGAAGGCCTTCGAGTCCCTCAAGTCCTTC	GGCGGTGGCGGCTGTTGCTGCTGC	A
ATN1	CCACCCACCAGTCTCAACACATC	CCAGTGGGTGGGGAAATGCTC	A
AR	ACCGAGGAGCTTTCCAGAAT	CTCATCCAGGACCAGGTAGC	B

Supplemental Table 2. Genotyped subjects per gene and number of excluded cases for *Cohort 1, Cohort 2* and *Cohort 3* combined

Gene	Total # of Subjects[a]	Included range shorter allele[b]	Included range longer allele[b]	# Excluded cases (%)[c]
ATXN1	5016	25-33	28-38	20 (0.4)
ATXN2	4961	17-23	22-31	16 (0.3)
ATXN3	4927	14-28	14-38	15 (0.3)
CACNA1A	5018	4-14	7-17	0 (0.0)
ATXN7	4724	7-13	10-15	7 (0.1)
TBP	4970	27-40	30-42	12 (0.2)
HTT	5001	9-26	15-35	15 (0.3)
ATN1	5021	5-18	8-24	21 (0.4)
AR shorter allele	4932	13-31	-	18 (0.4)
AR longer allele	4932	-	14-32	16 (0.3)

[a]) Due to insufficient amounts of DNA, we were unable to genotype all subjects for each gene. The number of lacking samples per gene ranged between 112-409 and were missing completely at random. [b]) Range of CAG repeats with frequencies of 5 or more. [c]) Number of cases excluded due to CAG repeat lengths not being within the range of CAG repeats with frequencies of 5.

Supplementary Table 3. Genotyped subjects per gene and number of excluded cases for *Cohort 1*

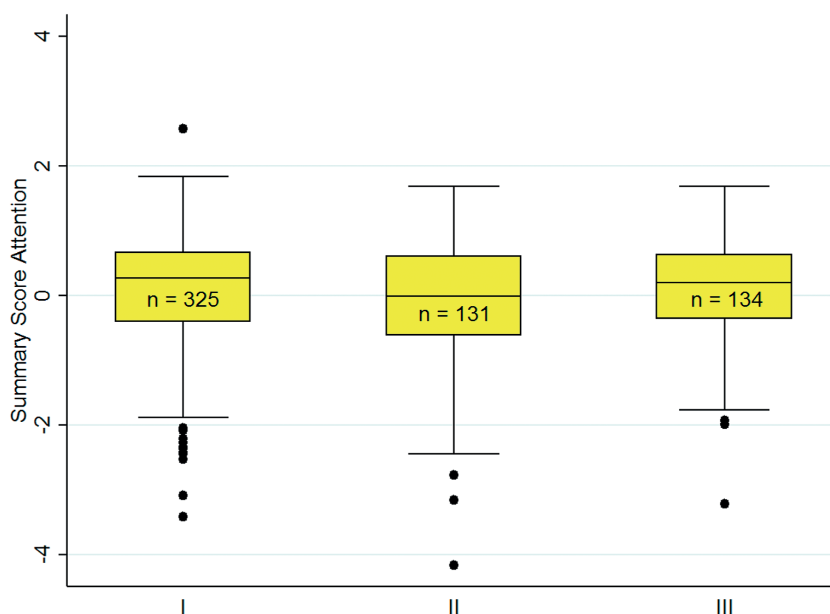
Gene	Total # of Subjects[a]	Included range shorter allele[b]	Included range longer allele[b]	# Excluded cases (%)[c]
ATXN1	952	26-32	28-36	19 (2.0)
ATXN2	954	17-26	22-28	7 (0.7)
ATXN3	954	13-31	14-39	3 (0.3)
CACNA1A	954	4-13	7-14	8 (0.8)
ATXN7	930	7-13	10-16	2 (0.2)
TBP	953	29-39	35-40	13 (1.4)
HTT	952	9-24	15-30	15 (1.6)
ATN1	954	8-18	8-23	8 (0.8)
AR shorter allele	952	14-28	-	15 (1.6)
AR longer allele	952	-	16-31	14 (1.5)

[a]) Due to insufficient amounts of DNA, we were unable to genotype all subjects for each gene. The number of lacking samples per gene ranged between 5-29 and were missing completely at random. [b]) Range of CAG repeats with frequencies of 5 or more. [c]) Number of cases excluded due to CAG repeat lengths not being within the range of CAG repeats with frequencies of 5.

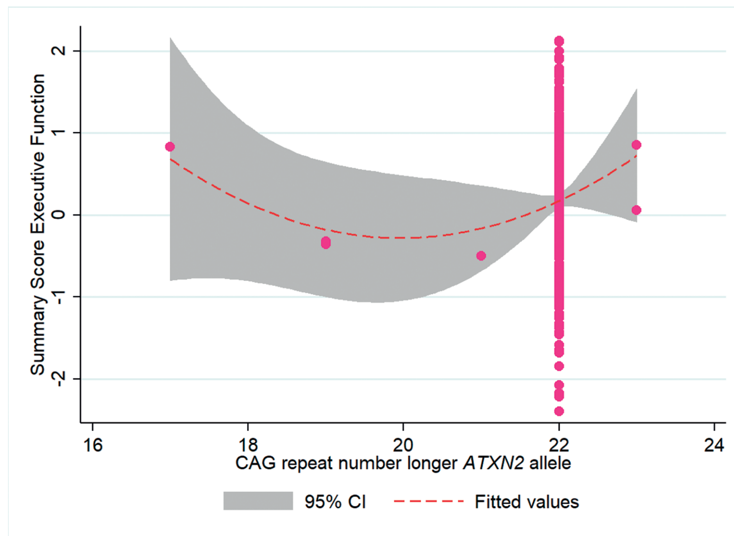
Supplemental Table 4. The association between age of onset in Alzheimer disease and the CAG repeat number in polyglutamine disease-associated genes of which both omnibus tests were significant.

Gene	Variable	β-coefficient[a]	SE	t	p-value	95% CI		R^2
ATXN2	ATXN2_s	0.617	0.554	1.11	.266	-0.471	1.704	
	ATXN2_l	-0.296	0.251	-1.18	.239	-0.789	0.197	
TBP	TBP_s	0.320	0.156	2.05	.041[b]	0.014	0.627	
	TBP_l	0.053	0.329	0.16	.872	-0.592	0.698	

PDAG = polyglutamine disease-associated gene. s = relatively shorter allele. l = relatively longer allele. sl = sl = interaction term relatively shorter and longer allele. SE = standard error. CI = confidence interval. *AR* (short) = the relatively shorter X-linked *AR* allele assessed in both men and women. *AR* (long) = the relatively longer X-linked *AR* allele assessed in both men and women. [a]) Increase in cognitive summary score per CAG repeat number increase corrected for age, sex, level of education, duration of AD symptoms, atrophy summary score and APOE genotype. [b]) Not significant before correction for age, sex, level of education, duration of AD symptoms, atrophy summary score and APOE genotype.



Supplemental Figure 1. The association between the summary score of attention and the interaction between the CAG repeat numbers in both *ATXN1* alleles. When both the relatively short and the relatively long *ATXN1* alleles were smaller than or equal to their medians, the summary score of attention was higher. The summary score of attention decreased when the relatively longer *ATXN1* allele became larger than its median, but when both alleles were larger than their medians, the attention summary score increased again. I = both *ATXN1* alleles ≤ median. II = the relatively shorter *ATXN1* allele ≤ median and the relatively longer *ATXN1* allele > median. III = both alleles > median. median shorter *ATXN1* allele = 29 CAG repeats. median longer *ATXN1* allele = 30 CAG repeats. After correction for multiple testing, this association did not remain significant.



Supplemental Figure 2. The association between the summary score of executive function and the CAG repeat number in the shorter *ATXN2* allele. The summary score of executive function had a non-linear association with the CAG repeat number in the shorter *ATXN2* allele. Both smaller and larger CAG repeat numbers in the shorter *ATXN2* allele were associated with a higher summary score of executive function. The result remained significant after correction for age, sex, level of education, duration of AD symptoms and summary score of atrophy (β = 0.151 ± 0.026, 95% CI 0.099; 0.202, p-value < .001). However, this finding was dependent on several influential points, casting doubt on the validity of this association.