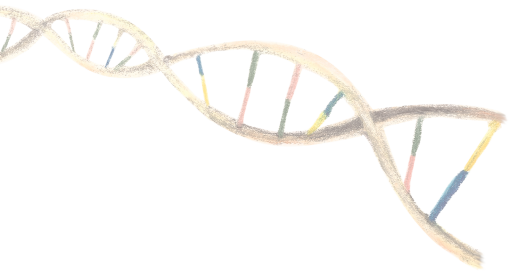
Huntington disease and other polyglutamine diseases : using CAG repeat variations to explain missing heritability

Gardiner, S.L.



Chapter 5

REPEAT LENGTH VARIATIONS IN POLYGLUTAMINE DISEASE-ASSOCIATED GENES AFFECT BODY MASS INDEX

Sarah L. Gardiner,[1,2] Renée de Mutsert,[3] Stella Trompet,[4] Merel W. Boogaard,[1,5] Ko Willems van Dijk,[2,4] P. J. Wouter Jukema,[7] P. Eline Slagboom,[6] Raymund A.C. Roos,[1] Hanno Pijl,[4] Frits R. Rosendaal,[3] N. Ahmad Aziz[1,8]

Departments of [1]Neurology, [2]Human Genetics, [3]Epidemiology ,[4]Internal Medicine, [5]Clinical Genetics, [6]Molecular Epidemiology and [7]Cardiology, Leiden University Medical Centre, Leiden, 2300 RC, The Netherlands [8]Department of Neurodegenerative Disease, UCL Huntington's Disease Centre, University College London Institute of Neurology, London,WC1E 6BT, United Kingdom

This is a post-peer-review, pre-copyedit version of an article published in the International Journal of Obesity. The final authenticated version is available online at: https://doi.org/10.1038/s41366-018-0161-7

International journal of obesity (2005). 2019 Mar;43(3):440-449

ABSTRACT

Background

The worldwide prevalence of obesity, a major risk factor for numerous debilitating chronic disorders, is increasing rapidly. Although a substantial amount of the variation in body mass index (BMI) is estimated to be heritable, the largest meta-analysis of genome-wide association studies (GWAS) to date explained only ~2.7% of the variation. To tackle this 'missing heritability' problem of obesity, here we focused on the contribution of DNA repeat length polymorphisms which are not detectable by GWAS.

Subjects and methods

We determined the cytosine-adenine-guanine (CAG) repeat length in the nine known polyglutamine disease-associated genes (*ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *HTT*, *ATN1* and *AR*) in two large cohorts consisting of 12 457 individuals and analysed their association with BMI and using generalized linear mixed-effect models.

Results

We found a significant association between BMI and the length of CAG repeats in seven polyglutamine disease-associated genes (including *ATXN1, ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP* and *AR*). Importantly, these repeat variations could account for 0.75% of the total BMI variation.

Conclusions

Our findings incriminate repeat polymorphisms as an important novel class of genetic risk factors of obesity and highlight the role of the brain in its pathophysiology.

INTRODUCTION

Obesity is a growing pandemic and acts as a major risk factor for a variety of prevalent chronic disorders, including cardiovascular, metabolic, inflammatory and neoplastic diseases.[1] Several studies have estimated the heritability of body mass index (BMI) at around 40-70%.[2-4] However, the BMI-associated loci identified in the largest meta-analysis of genome-wide association studies (GWAS) to date explained only ~2.7% of the variation,[5] indicating a large degree of 'missing heritability'. The GWAS approach, irrespective of its crucial contribution to the genetic mapping of complex human traits, neglects the effect of dynamic mutations on body composition, in the way trinucleotide expansions for instance associate with neurodegenerative disorders.[6-8] Recent studies have indeed shown that variations in these highly unstable repeat expansions can result in phenotypic consequences for organisms.[9] Nine hereditary neurodegenerative diseases, including Huntington Disease (HD), are caused by protein-coding trinucleotide expansions consisting of cytosine-adenine-guanine (CAG) repeats (**Table 1**).[10,11] Alongside motor impairment and neuropsychiatric disturbances, these disorders are often also accompanied by severe weight loss and metabolic disturbances. Given recent findings that even CAG repeat length variations in the non-mutant range in polyglutamine disease-associated genes (PDAGs) can act as risk factors for neuropsychiatric conditions,[12-14] we hypothesized that these prevalent polymorphisms may also act as genetic risk factors of BMI.

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 two well-characterized cohorts: the Netherlands Epidemiology of Obesity (NEO) study and the Prospective Study of Pravastatin in the Elderly at Risk (PROSPER) study (**Table 1** and **Supplemental Tables 1-3**). The NEO is a cohort study among 6671 men and women aged 45-65 years living in the greater area of Leiden, the Netherlands with an oversampling of overweight or obese individuals. A total of 5217 participants had a BMI of 27 kg/m^2 or higher This study was approved by the medical ethical committee of the Leiden University Centre (LUMC) and written informed consent was obtained from all participants.[15] The PROSPER is a cohort study among 5786 men and women between 70-82 years old with a pre-existing vascular disease or a raised risk for such a disease. Participants were recruited from three countries with 2517 individuals from Scotland, 2173 individuals from Ireland and 1096 individuals from the Netherlands. The study was approved by the institutional ethics review boards of all centres and written informed consent was obtained from all participants.[16] A post-hoc power calculation using the sample sizes of the NEO and PROSPER cohorts combined (n=12 457), showed that, at a significance level of $\alpha = 0.0056$ (0.05/9, because of the 9 tested PDAGs), this

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

			CAG repeat ranges						
Gene	Disease	Protein	normal	pathological	Allele	Mean	Median	N	Range
ATXN1	SCA1	Ataxin-1	6-39	41-83	short	29.21	29.00	12071	17-36
					long	30.79	30.00	12071	22-44
ATXN2	SCA2	Ataxin-2	14-32	33-500	short	21.93	22.00	11937	11-30
					long	22.39	22.00	11937	17-36
ATXN3	SCA3	Ataxin-3	12-44	52-87	short	19.02	20.00	12037	14-36
					long	24.25	23.00	12036	14-62
CACNA1A	SCA6	CACNA1A	4-18	20-33	short	10.58	11.00	12027	4-14
					long	12.47	13.00	12027	4-22
ATXN7	SCA7	Ataxin-7	3-19	37-460	short	10.05	10.00	11641	5-16
					long	10.82	10.00	11641	7-30
TBP	SCA17	TBP	25-43	45-66	short	36.34	37.00	11979	21-40
					long	37.88	38.00	11979	21-47
HTT	HD	Huntingtin	6-26	36-121	short	16.91	17.00	12055	6-31
					long	20.18	19.00	12055	10-40
ATN1	DRPLA	Atrophin-1	3-38	48-93	short	12.36	14.00	12100	3-22
					long	15.53	15.00	12100	8-28
AR	SBMA	Androgen receptor	6-36	38-72	short	21.13	21.00	11849	7-36
					long	22.83	23.00	11849	7-39

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.

sample size enabled detection of a very small effect size equalling to $R^2 = 0.001$ or larger with a statistical power of ≥ 0.78 (calculated using G*Power version 3.1.9.2).[17]

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 4**). 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 4) 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 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

We initially assessed the relation between CAG repeat sizes in the two alleles of each PDAG and BMI for each cohort separately (**Supplemental Table 5** and **6**). Next, to combine the results of both cohorts reliably, we first constructed parsimonious models for each cohort with the CAG repeat lengths of both alleles of each PDAG as independent variables (**Supplemental Table 7** and **8**). Subsequently, we only combined the data for PDAG alleles whose effects on BMI were directionally consistent. We applied a generalized linear mixed-effect model with BMI as the outcome variable and the CAG repeat lengths of both alleles as fixed effects. To assess potential interaction or non-linear effects,[12,18] we also included a product term of both alleles and a quadratic term for each allele. When the effect on BMI of only one allele was consistent between the two cohorts, we only included the quadratic term of that specific allele. Cohort (i.e. NEO or PROSPER) and country (i.e. Scotland, Ireland or the Netherlands) were set as random effects to account for potential population stratification. Non-significant higher order terms were removed from this original model and the analysis was 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.[19,20] The NEO data were weighed to the BMI distribution of the general population (the weight factor given to PROSPER participants was set at 1). To reduce multicollinearity all continuous variables

were centred around their respective means. Furthermore, we calculated the marginal R^2 per PDAG for each model to determine the amount of variance explained by each gene.[21] 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 were excluded. In addition, we excluded related participants and participants with a non-Caucasian ethnicity to increase homogeneity (**Supplemental Table 9-11**). For the results of the combined cohort, we applied a false discovery rate (FDR) correction to account for multiple testing, assuming nine independent tests with q set at .05.[22]

To illustrate the combined effect of the significant CAG repeat size variations in PDAGs on BMI we 1) calculated the residual BMI after regression on age and sex as fixed factors and cohort and country as random factors in a linear mixed-effects model, 2) performed linear regression with CAG repeat sizes in the alleles of the PDAGs significantly associated with BMI (including all interaction and non-linear effects which were identified as significant in the main analyses) as the independent variables and this residual BMI as the outcome, 3) divided the total cohort in four equally sized groups based on quartiles of the predicted values of this regression model, and 4) plotted the average BMI residual for each of these quartiles. 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 11 641 and 12 100 participants of both cohorts for each gene (**Table 1**). The lacking samples were due to too little available DNA and were missing completely at random. Between 6.9 and 7.4% were subsequently excluded due to CAG repeat lengths with a frequency of less than 10, participants being related or of non-Caucasian ethnicity (**Supplemental Table 9-11**), leaving a total of between 10 832 – 11 222 participants per gene for the analyses with 5485-5676 from the NEO cohort and 5276-5615 from the PROSPER cohort.

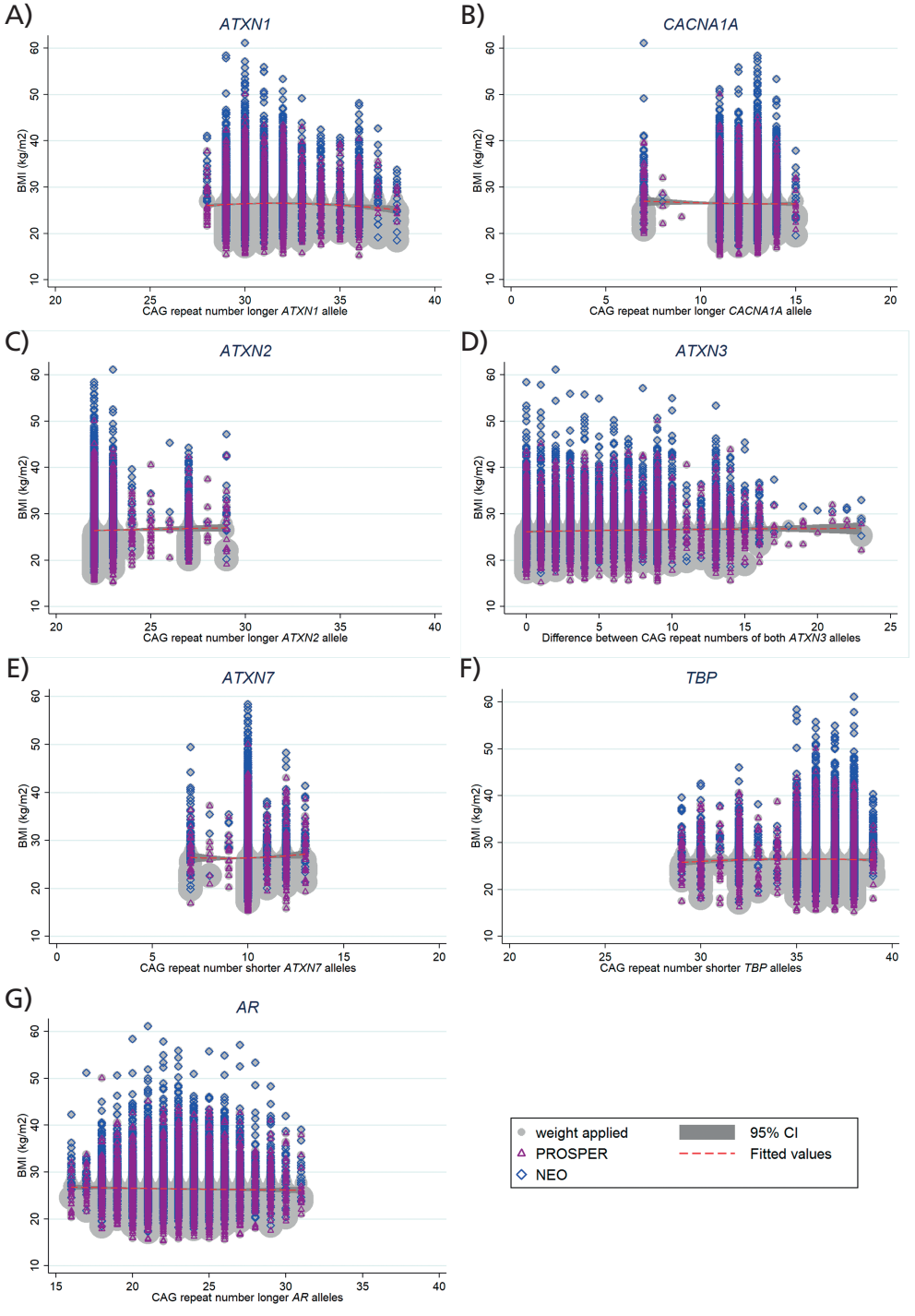
In the NEO cohort, we found four PDAGs that were significantly associated with BMI (including *ATXN1*, *ATXN2*, *ATXN3* and *TBP*) (**Supplemental Table 5**). Seven PDAGs in the PROSPER cohort were significantly associated with BMI (including *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A, ATXN7*, *TBP* and *HTT*) (**Supplemental Table 6**). Between the two cohorts, the effect on BMI of at least one allele was in the same direction for eight PDAGs (**Supplemental Table 7** and **8**). The data of only these directionally consistent alleles were combined. The effects of both *HTT* alleles were not consistent and therefore not combined (**Table 2**). After combining the data of the directionally consistent alleles, we found a total of seven PDAGs (including *ATXN1, ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*,

Table 2. The association between polyglutamine disease-associated genes (PDAGs) and body mass index (BMI) in the combined cohort

Gene	Variable	β-coefficient$_a$	SE	t	p-value	95% CI		R^2
ATXN1	ATXN1_s	-0.058	0.047	-1.24	.214	-0.150	0.034	$1.84*10^{-3}$
	ATXN1_l	0.078	0.029	2.74	.006	0.022	0.134	
	ATXN1_l2	-0.036	0.011	-3.26	.001	-0.058	-0.015	
ATXN2	ATXN2_s	-0.068	0.063	-1.08	.282	-0.191	0.056	$0.55*10^{-3}$
	ATXN2_l	0.081	0.024	3.35	.001	0.034	0.129	
ATXN3	ATXN3_s	-0.039	0.011	-3.58	<.001	-0.061	-0.018	$2.56*10^{-3}$
	ATXN3_l	0.048	0.018	2.73	.006	0.014	0.082	
	ATXN3_l2	0.003	0.001	4.14	<.001	0.002	0.004	
	ATXN3_sl	-0.020	0.000	-76.09	<.001	-0.021	-0.020	
CACNA1A	CACNA1A_s	-0.007	0.013	-0.53	.599	-0.033	0.019	$0.31*10^{-3}$
	CACNA1A_l	-0.038	0.006	-5.99	<.001	-0.050	-0.025	
	CACNA1A1_l2	0.010	0.005	2.24	.025	0.001	0.019	
ATXN7	ATXN7_s	0.122	0.013	9.14	<.001	0.095	0.148	$0.56*10^{-3}$
	ATXN7_l	-0.019	0.026	-0.73	.466	-0.071	0.033	
	ATXN7_s2	0.057	0.008	7.14	<.001	0.041	0.073	
TBP	TBP_s	0.011	0.001	10.87	<.001	0.009	0.014	$1.24*10^{-3}$
	TBP_l	-0.126	0.087	-1.46	.145	-0.296	0.044	
	TBP_s2	-0.012	0.004	-3.17	.002	-0.019	-0.005	
ATN1	ATN1_s	-0.008	0.011	-0.70	.486	-0.030	0.014	
	ATN1_l	0.032	0.021	1.51	.130	-0.009	0.073	
AR ♂	AR	-0.018	0.011	-1.68	.093	-0.039	0.003	$0.23*10^{-3}$
	AR_2	-0.003	0.001	-5.00	<.001	-0.003	-0.002	
AR ♀	AR_s	-0.054	0.028	-1.95	.052	-0.109	0.000	$1.41*10^{-3}$
	AR_l	0.025	0.010	2.45	.014	0.005	0.045	
	AR_l2	-0.012	0.006	-2.11	.035	-0.024	-0.001	
AR (long)	AR_l	-0.011	0.003	-3.41	.001	-0.017	-0.005	$0.05*10^{-3}$

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. BMI = body mass index, PDAGs= polyglutamine disease-associated genes, SE=standard error. CI=confidence interval. *AR* ♂=*AR* assessed in males. *AR* ♀=*AR* assessed in females. *AR* (long)= the longer *AR* allele assessed in both males and females. $_a$) This column indicates the amount of BMI change in kg/m2 per unit CAG repeat size increase.

TBP and *AR*) to be significantly associated with BMI (**Table 2**). For 5744 participants in the NEO and 5244 participants in the PROSPER we obtained principal components generated from genome wide genotyping data as described before .[19,20] We corrected for


A)
ATXN1
BMI (kg/m2)
CAG repeat number longer ATXN1 allele
B)
CACNA1A
BMI (kg/m2)
CAG repeat number longer CACNA1A allele
C)
ATXN2
BMI (kg/m2)
CAG repeat number longer ATXN2 allele
D)
ATXN3
BMI (kg/m2)
Difference between CAG repeat numbers of both ATXN3 alleles
E)
ATXN7
BMI (kg/m2)
CAG repeat number shorter ATXN7 alleles
F)
TBP
BMI (kg/m2)
CAG repeat number shorter TBP alleles
G)
AR
BMI (kg/m2)
CAG repeat number longer AR alleles
weight applied
PROSPER
NEO
95% CI
Fitted values

age, sex and population structure using these principal components. This correction did not substantially alter our results (**Supplemental Table 12**).

In the combined cohort, the longer alleles of *ATXN1*, *ATXN2* and *CACNA1A* were significantly associated with BMI. The association between BMI and the longer alleles of *ATXN1* and *CACNA1A* was quadratic, implying that both shorter and longer CAG repeat lengths were associated with a lower or higher BMI, respectively (**Figure 1A** and **1B**). The longer allele of *ATXN2* was associated with BMI in a linear fashion. Higher numbers of CAG repeats were associated with a higher BMI (**Figure 1C**). For *ATXN3*, the interaction between the two alleles affected BMI (**Table 2**). Given that the effect of CAG repeat size in the shorter and longer *ATXN3* allele on BMI was in opposite direction, we calculated the difference in CAG repeat size between the longer and shorter *ATXN3* alleles and found this difference to have a quadratic association with BMI (**Figure 1D**). Furthermore, the shorter alleles of both *ATXN7* and *TBP* had a quadratic association with BMI (**Figure 1E** and **1F**). Lastly, we examined the effect on BMI of the CAG repeat size in the X-linked *AR* gene, for which we 1) analysed men and women separately, and 2) investigated either the shorter or the longer *AR* allele in men and women combined. In men, long CAG repeat lengths resulted in an exponential decrease of BMI, whereas in women, the longer *AR* allele had a quadratic association with BMI (**Table 2**). When analysing the *AR* gene in men and women combined, a longer *AR* CAG repeat size in the longer allele was also associated with lower BMI (**Figure 1G**). To estimate the total percentage of variation in BMI explained by these seven PDAGs, we calculated the marginal R^2 for the final model including all the alleles which were significantly associated with BMI in the per gene analysis (**Table 2**). For *AR*, we included only the longer allele. The seven PDAGs that were significantly associated with BMI accounted for 0.75% of its variation in the combined cohort. Additional analysis of the combined effect showed that the difference in BMI between the lowest and highest quartile of the prediction score calculated based on the CAG repeat sizes in these seven PDAGs was about 0.42 kg/m^2 (corresponding to 1.29 kg for an individual 1.75 m in height) (**Figure 2**).

Figure 1. Scatterplots of the association between body mass index (BM)I and polyglutamine disease-associated genes (PDAGs). Shorter and longer CAG repeats lengths in the longer alleles of *ATXN1* (**A**) and *CACNA1A* (**B**) were associated with a lower and higher BMI, respectively. **C.** Larger CAG repeat numbers in longer allele of *ATXN2* were associated with a higher BMI. **D.** The difference in CAG repeat number between the shorter and longer *ATXN3* alleles had a quadratic association with BMI. Larger and smaller differences between these alleles were associated with a lower BMI. **E.** Shorter and longer CAG repeats in the shorter *ATXN7* allele (**E**) and the shorter TBP allele (**F**) were associated with a higher and lower BMI, respectively. **G.** The longer allele of *AR* had a quadratic association with BMI. Shorter and longer CAG repeats were associated with a higher BMI. the bèta-coefficient ± SE. CI=Confidence Interval.

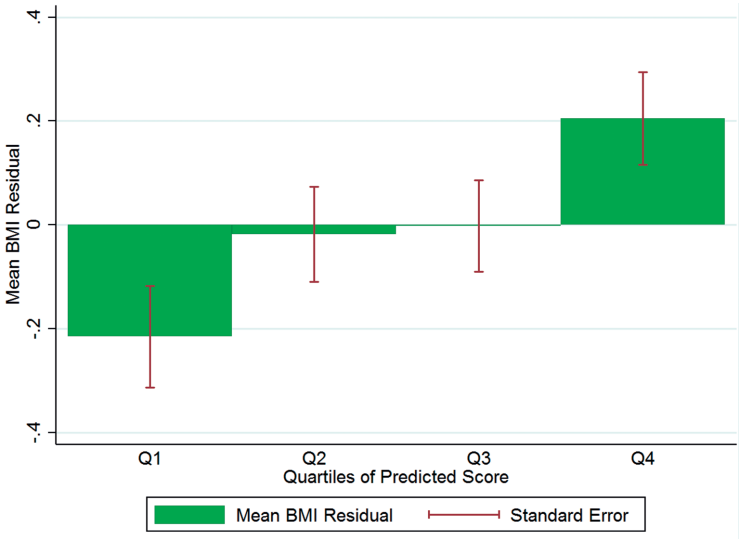


Figure 2. The effect of CAG repeat size variations in polyglutamine disease-associated genes (PDAGs) on body mass index (BMI). This plot illustrates that in combination CAG repeat size variations in only seven PDAGs can account for a variation of up to ~0.42 kg/m^2 in BMI. Please refer to the methods section for details on how the 'Predicted Score' was constructed.

DISCUSSION

Metabolic disturbances occur in many neurodegenerative diseases, including polyglutamine disorders.[23] Spinocerebellar ataxia type 3 (SCA3), one of the most prevalent polyglutamine diseases worldwide, is frequently complicated by unintended weight loss. The number of CAG repeats in the longer *ATXN3* allele was shown to have an inverse association with BMI in SCA3 patients.[24,25] We found that a larger difference between both *ATXN3* alleles was associated with a lower BMI. These results are consistent with the decreased BMI in SCA3 patients as the longer *ATXN3* allele needs to have a relatively large number of CAG repeats in order for the difference with the shorter allele to be large. Furthermore, amyotrophy has been reported in SCA1 and SCA6 patients with SCA1 patients displaying a higher resting state energy expenditure and fat oxidation compared to age, sex and body composition matched controls.[26,27] Consistent with these characteristics, the curvilinear association between BMI and the CAG repeat number in the longer *ATXN1* allele indicated that larger CAG repeat numbers also led to a lower BMI. The association between BMI and the CAG repeat length in *CACNA1A* was not consistent with the reported SCA6 characteristics, suggesting that the relationship between *CACNA1A* and BMI is different for the 'healthy' range compared to the diseased range. Including the diseased range in future research could provide additional insights into the overall effect of *CACNA1A* on BMI. Together, these results indicate that the effects of PDAGs on metabolism are not confined to the pathological range and may represent a homeostatic property of the polyglutamine domains of the encoded proteins in systemic energy regulation.[28]

The other PDAGs have also been suggested to be involved in the regulation of BMI and metabolism. For instance, normal ranged *AR* CAG repeat sizes, which determine androgen receptor sensitivity to testosterone, have been associated with body fat mass and blood lipid levels before.[29,30] Recent research also implicates *ATXN2* in metabolic processes. *ATXN2* knockout or transgenic mice display changes in body weight, insulin sensitivity and fertility.[31,32] Furthermore, a SNP located in the *A2BP1* gene which encodes the ataxin-2 binding protein 1 (also known as FOX-1) was associated with percentage of total body fat in Pima Indians,[33] while a SNP in *ATXN2L* encoding ataxin-2 like protein which interacts with ataxin-2, has been related to BMI.[5,34] Other obesity-related SNPs change the affinity of the thymine-adenine-thymine-adenine (TATA) box binding protein (TBP) encoded by *TBP* for human gene promotors, suggesting a possible pathophysiological mechanism for obesity involving *TBP*.[35]

Cognitive and behavioural changes are key characteristics of polyglutamine disorders. However, little is known about the extent to which repeat variations within the 'healthy' range result in similar deficits and whether these could cause changes in BMI. In previous research, we found a significant association between the risk of lifetime depression and the CAG repeat numbers in *ATXN7* and *TBP*.[12] The association between depression and obesity has been well established and a meta-analysis of longitudinal studies showed that obese individuals had a 55% increased risk of depression and depressed individuals had a 58% increased risk of becoming obese.[36] Interestingly, the association of the CAG repeat number in the shorter *ATXN7* allele with BMI and depression was consistent with larger CAG repeat numbers leading to both a higher risk of lifetime depression and a higher BMI.[12] *ATXN7* encodes ataxin-7 (ATXN7), a member of the TATA-binding protein-free TAF complex (TFTC) and the SPT3/TAF9/GCN5 acetyltransferase (STAGA) complex. These complexes are coactivators involved in the initiation of gene transcription via RNA polymerase II.[37] Through modification of the transcription of RNA polymerase II dependent genes, *ATXN7* repeat variations could cause obesity resulting in depression via metabolic pathways, such as inflammatory responses, dysregulation of the hypothalamic-pituitary-adrenal axis (HPA axis) and alterations in the brain due to diabetes mellitus and insulin resistance.[38-49] In addition, increased psychological stress, body dissatisfaction, physical pain and a decreased self-esteem due to obesity could also cause depression.[50-52] Conversely, repeat polymorphisms in *ATXN7* could cause depression leading to obesity through the adoption of an unhealthy lifestyle, including insufficient physical exercise and unhealthy dietary preferences.[53] *AR* CAG repeat variations have also been previously associated with depression in men. Larger CAG repeat numbers in *AR* lead to lower transcriptional effects of testosterone and were associated with depressive symptoms.[54-56] Furthermore, larger CAG repeat numbers in *AR* were associated with lower test scores on three cognitive tests in older white men and decreased effects of testosterone have been associated with cognitive problems in rodents, such as decreased performances in spatial learning, memory and inhibitory avoidance tasks. Different studies have shown

that working memory and episodic memory are core cognitive processes critical for food-related decision-making, and that disruption of these processes contributes to problems with appetite control and weight gain.[57] Therefore, high CAG repeat numbers in *AR* and the resulting decreased transcriptional effects of testosterone, might lead to cognitive deficits that in turn could result in changes in appetite control and BMI.

We recognize that our cohort size was relatively small compared to the sample sizes usually included in GWAS. However, the fact that we were able to find many tandem repeat polymorphisms in the PDAGs significantly associated with BMI implies that our study was sufficiently powered to detect these effects. In addition, our sample size allowed us to find relatively small effects similar to for instance the effect of the type 2 diabetes-associated A allele at rs9939609 linked to the *FTO* gene that was associated with a median per-allele change of ~0.36 kg/m^2 and explained a variance in BMI of ~1%, or the effect of the C allele at rs17782313 linked to the *MC4R* gene that was associated with a difference in BMI of ~0.22 kg/m^2 per allele and explained ~0.14% of the variance in BMI.[58,59] Although increasing the sample size might have resulted in the detection of even more, and even smaller effects, we must affirm that determining the repeat numbers in these genes was not a straightforward process, could not be automated and was extremely laborious. This fact also compelled us to focus on a set of predefined and promising genes with repeat variations which are known to be 1) related to changes in protein function, and 2) causative of (brain) disorders which are accompanied by profound metabolic disturbances. Nonetheless, many more interesting tandem repeat polymorphisms exist in the human genome and future research is warranted to delineate the effects of these other repeat polymorphisms on BMI.[60] Recently, a method was described that could allow genome-wide imputation of short tandem repeats (STRs) from SNP data using a phased SNP/STR haplotype panel generated from available whole genome sequencing datasets.[61] However, these SNP/STR haplotypes have not been published yet, but once these data become publicly available, this panel could be used to test the association between many STR variations and BMI within the myriad of existing data.

To our knowledge, the association between normal ranged CAG repeat polymorphisms in the nine PDAGs and BMI was not assessed before and the SNPs previously found associated with BMI were not located in or near the investigated PDAGs.[5] Through LD analysis, several studies found haplotypes associated with expanded or large 'healthy' ranged CAG repeat numbers in *ATXN1*, *CACNA1A*, *ATXN7* and *AR*.[62-67] However, these associated haplotypes differed substantially per investigated population. In addition, the CAG repeat sequence in PDAGs are directly translated in the respective proteins and have important functional consequences.[68] Therefore, the CAG repeat sequence itself is likely to lead to the variation in BMI. Although we cannot fully exclude potential modifying effects of other genetic loci in LD with PDAGs, the fact that tagging SNPs in or

around PDAGs have not been related to BMI before suggests that the influence of other genetic variants in LD with these triplet repeats is likely to be minimal.[5]

In summary, we found the CAG repeat size in seven PDAGs to be significantly associated with BMI in two large study populations accounting for 0.75% of the total variation. As PDAGs are known be critically implicated in processes which recently were identified through pathway analysis to be involved in obesity susceptibility, including synaptic function and glutamate signalling, and can be specifically targeted by promising therapeutics currently in development for polyglutamine disorders, including gene suppression strategies,[69] our results open a novel therapeutic avenue for obesity treatment. In conclusion, we demonstrate the relevance of trinucleotide repeats as a new class of genetic risk factors of obesity and provide further evidence for the fundamental link between the brain and metabolism.

ACKNOWLEDGEMENTS

We would like to express our gratitude to all individuals who participated in the NEO and PROSPER study as well as all research staff for collecting the data and I. de Jonge for data management. Dr. Anton de Craen's guidance was crucial for the design of this study for which we posthumously wish to express our utmost appreciation. This study 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 PROSPER study was supported by an investigator-initiated grant obtained from Bristol-Myers Squibb. The NEO study is supported by the participating Departments, the Division and the Board of Directors of the Leiden University Medical Centre, and by the Leiden University, Research Profile Area 'Vascular and Regenerative Medicine'.

CONFLICT OF INTEREST

There are no financial conflicts of interest to be disclosed.

REFERENCES

1. Obesity: preventing and managing the global epidemic. Report of a WHO consultation. *World Health Organization technical report series* 2000; **894**: i-xii, 1-253.
2. Chagnon YC, Perusse L, Bouchard C. Familial aggregation of obesity, candidate genes and quantitative trait loci. *Current opinion in lipidology* 1997; **8**(4): 205-11.
3. Chung WK. An overview of mongenic and syndromic obesities in humans. *Pediatric blood & cancer* 2012; **58**(1): 122-8.
4. Goran MI. Genetic influences on human energy expenditure and substrate utilization. *Behavior genetics* 1997; **27**(4): 389-99.
5. Locke AE, Kahali B, Berndt SI, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* 2015; **518**(7538): 197-206.
6. Hannan AJ. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. *Trends in genetics : TIG* 2010; **26**(2): 59-65.
7. 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.
8. Manolio TA. Genomewide association studies and assessment of the risk of disease. *The New England journal of medicine* 2010; **363**(2): 166-76.
9. Duitama J, Zablotskaya A, Gemayel R, et al. Large-scale analysis of tandem repeat variability in the human genome. *Nucleic acids research* 2014; **42**(9): 5728-41.
10. 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.
11. Fan HC, Ho LI, Chi CS, et al. Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell transplantation* 2014; **23**(4-5): 441-58.
12. 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.
13. Gardiner SL, van Belzen MJ, Boogaard MW, et al. Huntingtin gene repeat size variations affect risk of lifetime depression *Translational psychiatry* 2017.
14. Killoran A, Biglan KM, Jankovic J, et al. Characterization of the Huntington intermediate CAG repeat expansion phenotype in PHAROS. *Neurology* 2013; **80**(22): 2022-7.
15. de Mutsert R, den Heijer M, Rabelink TJ, et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. *European journal of epidemiology* 2013; **28**(6): 513-23.
16. 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.
17. 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.
18. 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.
19. 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.
20. Blauw LL, Noordam R, Trompet S, et al. Genetic variation in the obesity gene FTO is not associated with decreased fat oxidation: the NEO study. *International journal of obesity (2005)* 2017; **41**(10): 1594-600.
21. 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.

22. 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.
23. Aziz NA, van der Marck MA, Pijl H, Olde Rikkert MG, Bloem BR, Roos RA. Weight loss in neurodegenerative disorders. *Journal of neurology* 2008; **255**(12): 1872-80.
24. Saute JA, da Silva AC, Muller AP, et al. Serum insulin-like system alterations in patients with spinocerebellar ataxia type 3. *Movement disorders : official journal of the Movement Disorder Society* 2011; **26**(4): 731-5.
25. Saute JA, Silva AC, Souza GN, et al. Body mass index is inversely correlated with the expanded CAG repeat length in SCA3/MJD patients. *Cerebellum (London, England)* 2012; **11**(3): 771-4.
26. Schmitz-Hubsch T, Coudert M, Bauer P, et al. Spinocerebellar ataxia types 1, 2, 3, and 6: disease severity and nonataxia symptoms. *Neurology* 2008; **71**(13): 982-9.
27. Mahler A, Steiniger J, Endres M, Paul F, Boschmann M, Doss S. Increased catabolic state in spinocerebellar ataxia type 1 patients. *Cerebellum (London, England)* 2014; **13**(4): 440-6.
28. Ashkenazi A, Bento CF, Ricketts T, et al. Polyglutamine tracts regulate beclin 1-dependent autophagy. *Nature* 2017; **545**(7652): 108-11.
29. Gustafson DR, Wen MJ, Koppanati BM. Androgen receptor gene repeats and indices of obesity in older adults. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* 2003; **27**(1): 75-81.
30. Zitzmann M, Gromoll J, von Eckardstein A, Nieschlag E. The CAG repeat polymorphism in the androgen receptor gene modulates body fat mass and serum concentrations of leptin and insulin in men. *Diabetologia* 2003; **46**(1): 31-9.
31. Kiehl TR, Nechiporuk A, Figueroa KP, Keating MT, Huynh DP, Pulst SM. Generation and characterization of Sca2 (ataxin-2) knockout mice. *Biochemical and biophysical research communications* 2006; **339**(1): 17-24.
32. Lastres-Becker I, Brodesser S, Lutjohann D, et al. Insulin receptor and lipid metabolism pathology in ataxin-2 knock-out mice. *Human molecular genetics* 2008; **17**(10): 1465-81.
33. Shibata H, Huynh DP, Pulst SM. A novel protein with RNA-binding motifs interacts with ataxin-2. *Human molecular genetics* 2000; **9**(9): 1303-13.
34. Carmo-Silva S, Nobrega C, Pereira de Almeida L, Cavadas C. Unraveling the Role of Ataxin-2 in Metabolism. *Trends in endocrinology and metabolism: TEM* 2017; **28**(4): 309-18.
35. Arkova OV, Ponomarenko MP, Rasskazov DA, et al. Obesity-related known and candidate SNP markers can significantly change affinity of TATA-binding protein for human gene promoters. *BMC genomics* 2015; **16 Suppl 13**: S5.
36. Luppino FS, de Wit LM, Bouvy PF, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Archives of general psychiatry* 2010; **67**(3): 220-9.
37. 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. *Biochemical Society symposium* 2006; (73): 155-63.
38. Emery CF, Fondow MD, Schneider CM, et al. Gastric bypass surgery is associated with reduced inflammation and less depression: a preliminary investigation. *Obesity surgery* 2007; **17**(6): 759-63.
39. Shoelson SE, Herrero L, Naaz A. Obesity, inflammation, and insulin resistance. *Gastroenterology* 2007; **132**(6): 2169-80.

40. Vaccarino V, Johnson BD, Sheps DS, et al. Depression, inflammation, and incident cardiovascular disease in women with suspected coronary ischemia: the National Heart, Lung, and Blood Institute-sponsored WISE study. *Journal of the American College of Cardiology* 2007; **50**(21): 2044-50.

41. Bremmer MA, Beekman AT, Deeg DJ, et al. Inflammatory markers in late-life depression: results from a population-based study. *Journal of affective disorders* 2008; **106**(3): 249-55.

42. Milaneschi Y, Corsi AM, Penninx BW, Bandinelli S, Guralnik JM, Ferrucci L. Interleukin-1 receptor antagonist and incident depressive symptoms over 6 years in older persons: the InCHIANTI study. *Biol Psychiatry* 2009; **65**(11): 973-8.

43. Pasquali R, Vicennati V. Activity of the hypothalamic-pituitary-adrenal axis in different obesity phenotypes. *International journal of obesity and related metabolic disorders : journal of the International Association for the Study of Obesity* 2000; **24 Suppl 2**: S47-9.

44. Walker BR. Activation of the hypothalamic-pituitary-adrenal axis in obesity: cause or consequence? *Growth hormone & IGF research : official journal of the Growth Hormone Research Society and the International IGF Research Society* 2001; **11 Suppl A**: S91-5.

45. Belanoff JK, Kalehzan M, Sund B, Fleming Ficek SK, Schatzberg AF. Cortisol activity and cognitive changes in psychotic major depression. *The American journal of psychiatry* 2001; **158**(10): 1612-6.

46. Holsboer F. The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 2000; **23**(5): 477-501.

47. Lee WJ, Lee YC, Ser KH, Chen JC, Chen SC. Improvement of insulin resistance after obesity surgery: a comparison of gastric banding and bypass procedures. *Obesity surgery* 2008; **18**(9): 1119-25.

48. Huber JD. Diabetes, cognitive function, and the blood-brain barrier. *Current pharmaceutical design* 2008; **14**(16): 1594-600.

49. Ajilore O, Haroon E, Kumaran S, et al. Measurement of brain metabolites in patients with type 2 diabetes and major depression using proton magnetic resonance spectroscopy. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 2007; **32**(6): 1224-31.

50. Atlantis E, Ball K. Association between weight perception and psychological distress. *International journal of obesity (2005)* 2008; **32**(4): 715-21.

51. Derenne J, Beresin E. Body Image, Media, and Eating Disorders-a 10-Year Update. *Academic psychiatry : the journal of the American Association of Directors of Psychiatric Residency Training and the Association for Academic Psychiatry* 2018; **42**(1): 129-34.

52. Beesdo K, Jacobi F, Hoyer J, Low NC, Hofler M, Wittchen HU. Pain associated with specific anxiety and depressive disorders in a nationally representative population sample. *Social psychiatry and psychiatric epidemiology* 2010; **45**(1): 89-104.

53. Paans NPG, Bot M, van Strien T, Brouwer IA, Visser M, Penninx B. Eating styles in major depressive disorder: Results from a large-scale study. *Journal of psychiatric research* 2018; **97**: 38-46.

54. Zitzmann M. Mechanisms of disease: pharmacogenetics of testosterone therapy in hypogonadal men. *Nature clinical practice Urology* 2007; **4**(3): 161-6.

55. Colangelo LA, Sharp L, Kopp P, et al. Total testosterone, androgen receptor polymorphism, and depressive symptoms in young black and white men: the CARDIA Male Hormone Study. *Psychoneuroendocrinology* 2007; **32**(8-10): 951-8.

56. Vermeersch H, T'Sjoen G, Kaufman JM, Vincke J, Van Houtte M. Testosterone,

androgen receptor gene CAG repeat length, mood and behaviour in adolescent males. *European journal of endocrinology* 2010; **163**(2): 319-28.

57. Higgs S, Spetter MS. Cognitive Control of Eating: the Role of Memory in Appetite and Weight Gain. *Current obesity reports* 2018; **7**(1): 50-9.

58. Frayling TM, Timpson NJ, Weedon MN, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science (New York, NY)* 2007; **316**(5826): 889-94.

59. Loos RJ, Lindgren CM, Li S, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nature genetics* 2008; **40**(6): 768-75.

60. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. *Nature reviews Genetics* 2018.

61. Saini S, Mitra I, Gymrek M. A reference haplotype panel for genome-wide imputation of short tandem repeats. *bioRxiv* 2018.

62. Mittal U, Sharma S, Chopra R, et al. Insights into the mutational history and prevalence of SCA1 in the Indian population through anchored polymorphisms. *Human genetics* 2005; **118**(1): 107-14.

63. Krysa W, Sulek A, Rakowicz M, Szirkowiec W, Zaremba J. High relative frequency of SCA1 in Poland reflecting a potential founder effect. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology* 2016; **37**(8): 1319-25.

64. Craig K, Takiyama Y, Soong BW, et al. Pathogenic expansions of the SCA6 locus are associated with a common CACNA1A haplotype across the globe: founder effect or predisposing chromosome? *European journal of human genetics : EJHG* 2008; **16**(7): 841-7.

65. Faruq M, Magana JJ, Suroliya V, et al. A Complete Association of an intronic SNP rs6798742 with Origin of Spinocerebellar Ataxia Type 7-CAG Expansion Loci in the Indian and Mexican Population. *Annals of human genetics* 2017; **81**(5): 197-204.

66. Magana JJ, Gomez R, Maldonado-Rodriguez M, et al. Origin of the spinocerebellar ataxia type 7 gene mutation in Mexican population. *Cerebellum (London, England)* 2013; **12**(6): 902-5.

67. Terry KL, De Vivo I, Titus-Ernstoff L, Shih MC, Cramer DW. Androgen receptor cytosine, adenine, guanine repeats, and haplotypes in relation to ovarian cancer risk. *Cancer research* 2005; **65**(13): 5974-81.

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

69. Keiser MS, Kordasiewicz HB, McBride JL. Gene suppression strategies for dominantly inherited neurodegenerative diseases: lessons from Huntington's disease and spinocerebellar ataxia. *Human molecular genetics* 2016; **25**(R1): R53-64.

SUPPLEMENTARY MATERIAL

Supplemental Table 1. Baseline characteristics NEO and PROSPER cohort.

Variable		NEO (n=6671)	PROSPER (n=5786)	Total (n=12 457)
Age (mean ± SD)		55.80 ± 6.0	75.33 ± 3.4	64.87 ± 10.9
Sex n (%)	male	3156 (47.3)	2798 (48.4)	5954 (47.8)
	female	3515 (52.7)	2988 (51.6)	6503 (52.2)
Country n (%)	The Netherlands	6671 (100)	1096 (18.9)	7767 (62.4)
	Scotland	-	2517 (43.5)	2517 (20.2)
	Ireland	-	2173 (37.6)	2173 (17.4)
BMI (mean ± SD)		30.08 ± 4.9	26.82 ± 4.2	28.57 ± 4.8

n=number of individuals; SD=standard deviation.

Supplemental Table 2. Summary genotyped polyglutamine disease-associated genes (PDAGs) in the NEO cohort.

Gene	Allele	N	Mean ± SD	Median	Mode	Range
ATXN1	short	6438	29.21 ± 1.1	29	29	17-36
	long	6438	30.77 ± 1.8	30	30	22-40
ATXN2	short	6389	21.92 ± 0.7	22	22	13-30
	long	6389	22.36 ± 1.2	22	22	17-36
ATXN3	short	6494	19.03 ± 4.3	20	14	14-35
	long	6494	24.22 ± 3.7	23	23	14-62
CACNA1A	short	6394	10.52 ± 2.1	11	11	4-14
	long	6394	12.46 ± 1.2	13	13	4-22
ATXN7	short	6356	10.04 ± 0.5	10	10	5-13
	long	6356	10.85 ± 1.3	10	10	7-30
TBP	short	6418	36.38 ± 1.7	37	37	23-40
	long	6418	37.88 ± 1.0	38	38	30-45
HTT	short	6453	16.88 ± 1.9	17	17	6-31
	long	6453	20.13 ± 3.5	19	17	11-40
ATN1	short	6467	12.42 ± 3.0	14	15	3-22
	long	6467	15.49 ± 2.1	15	15	8-28
AR	short	6369	21.09 ± 2.7	21	21	8-36
	long	6368	22.83 ± 2.9	23	21	11-38

n=number of genotyped participants; SD=standard deviation.

Supplemental Table 3. Summary genotyped polyglutamine disease-associated genes (PDAGs) in the PROSPER cohort.

Gene	Allele	N	Mean ± SD	Median	Mode	Range
ATXN1	short	5633	29.21 ± 1.0	29	29	17-36
	long	5633	30.81 ± 1.7	30	30	26-44
ATXN2	short	5548	21.94 ± 0.6	22	22	11-27
	long	5548	22.43 ± 1.2	22	22	22-33
ATXN3	short	5543	18.99 ± 4.4	20	14	14-36
	long	5543	24.27 ± 3.8	23	23	14-49
CACNA1A	short	5633	10.64 ± 2.1	11	11	4-14
	long	5633	12.48 ± 1.1	13	13	7-17
ATXN7	short	5285	10.07 ± 0.5	10	10	7-16
	long	5285	10.78 ± 1.2	10	10	10-25
TBP	short	5561	36.30 ± 1.9	37	37	21-40
	long	5561	37.87 ± 1.0	38	38	21-47
HTT	short	5602	16.95 ± 2.1	17	17	9-29
	long	5602	20.22 ± 3.5	19	17	10-38
ATN1	short	5633	12.29 ± 3.1	14	15	8-20
	long	5633	15.57 ± 2.2	16	15	8-27
AR	short	5546	21.18 ± 2.7	21	21	7-35
	long	5546	22.84 ± 2.9	23	21	7-39

n=number of genotyped participants; SD=standard deviation.

Supplemental Table 4. 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 5. The association between polyglutamine disease-associated genes (PDAGs) and body mass index (BMI) in the NEO cohort.

Gene	Variable	β-coefficient[a]	SE	t	p-value	95% CI		R^2
ATXN1	ATXN1_s	-0.090	0.096	-0.94	.349	-0.279	0.099	$2.3*10^{-3}$
	ATXN1_l	0.099	0.069	1.44	.149	-0.036	0.235	
	ATXN1_l2	-0.044	0.016	-2.72	.007	-0.075	-0.012	
	ATXN1_s[b]	-0.040	0.091	-0.44	.658	-0.217	0.137	
	ATXN1_l[b]	0.108	0.065	1.66	.098	-0.020	0.235	
	ATXN1_l2[b]	-0.044	0.015	-2.88	.004	-0.073	-0.014	
ATXN2	ATXN2_s	-0.750	0.286	-2.62	.009	-1.310	-0.189	$9.3*10^{-3}$
	ATXN2_l	0.093	0.169	0.55	.584	-0.239	0.425	
	ATXN2_sl	-0.284	0.101	-2.82	.005	-0.481	-0.086	
	ATXN2_s[b]	-0.776	0.234	-3.31	.001	-1.237	-0.316	
	ATXN2_l[b]	0.109	0.161	0.68	.495	-0.206	0.425	
	ATXN2_sl[b]	-0.255	0.106	-2.42	.016	-0.463	-0.048	
ATXN3	ATXN3_s	-0.049	0.019	-2.61	.009	-0.086	-0.012	$2.3*10^{-3}$
	ATXN3_l	0.053	0.022	2.38	.017	0.009	0.096	
	ATXN3_s[b]	-0.049	0.018	-2.72	.006	-0.084	-0.014	
	ATXN3_l[b]	0.055	0.021	2.59	.010	0.013	0.097	
CACNA1A	CACNA1A_s	-0.016	0.037	-0.44	.659	-0.088	0.056	
	CACNA1A_l	-0.056	0.074	-0.75	.451	-0.200	0.089	
	CACNA1A_s[b]	-0.004	0.035	-0.12	.904	-0.073	0.064	
	CACNA1A_l[b]	-0.055	0.071	-0.77	.439	-0.195	0.085	
ATXN7	ATXN7_s	0.114	0.126	0.90	.367	-0.133	0.360	
	ATXN7_l	0.010	0.062	0.17	.869	-0.111	0.132	
	ATXN7_s[b]	0.082	0.120	0.68	.498	-0.154	0.317	
	ATXN7_l[b]	0.001	0.059	0.01	.993	-0.114	0.116	
TBP	TBP_s	0.061	0.047	1.31	.189	-0.030	0.152	$1.7*10^{-3}$
	TBP_l	-0.205	0.085	-2.40	.016	-0.372	-0.038	
	TBP_s[b]	0.034	0.045	0.74	.459	-0.055	0.122	
	TBP_l[b]	-0.199	0.084	-2.38	.017	-0.362	-0.035	
HTT	HTT_s	0.060	0.040	1.48	.139	-0.019	0.138	
	HTT_l	-0.041	0.023	-1.79	.074	-0.086	0.004	
	HTT_s[b]	0.047	0.038	1.23	.219	-0.028	0.121	
	HTT_l[b]	-0.039	0.022	-1.74	.082	-0.083	0.005	
ATN1	ATN1_s	-0.000	0.026	-0.01	.995	-0.050	0.050	
	ATN1_l	0.046	0.037	1.25	.211	-0.026	0.118	
	ATN1_s[b]	-0.003	0.025	-0.14	.891	-0.052	0.045	
	ATN1_l[b]	0.041	0.036	1.12	.262	-0.030	0.112	

Supplemental Table 5. (continued)

Gene	Variable	β-coefficient[a]	SE	t	p-value	95% CI		R^2
AR ♂	AR_s	-0.027	0.035	-0.75	.452	-0.096	0.043	
	AR_s[c]	-0.030	0.034	-0.88	.379	-0.097	0.037	
AR ♀	AR_s	-0.061	0.054	-1.12	.261	-0.166	0.045	
	AR_l	0.017	0.046	0.36	.718	-0.074	0.107	
	AR_s[c]	-0.078	0.053	-1.46	.144	-0.182	0.027	
	AR_l[c]	0.032	0.046	0.69	.488	-0.058	0.121	
AR_s	AR_s	-0.035	0.029	-1.23	.220	-0.092	0.021	
	AR_s[b]	-0.034	0.028	-1.19	.235	-0.089	0.022	
AR_l	AR_l	-0.012	0.027	-0.44	.658	-0.066	0.042	
	AR_l[b]	-0.013	0.026	-0.48	.632	-0.064	0.039	

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. SE=standard error. CI=confidence interval. [a]) This column indicates the amount of BMI change in kg/m2 per unit CAG repeat size increase. [b]) Results were adjusted for age, sex and population structure using principal components. [c]) Results were adjusted for age only and population structure using principal components.

Statistical analysis

We applied a linear regression model with the CAG repeat lengths of the shorter and longer alleles as independent variables and BMI as dependent variable. To assess potential interaction or non-linear effects,[12,18] we included a product term of both alleles and a quadratic term for each allele as independent variables. If the product term or the quadratic terms were not significantly associated with BMI, these variables were removed from the original model and the analysis was repeated until a final model was established. The final model was corrected for age and sex. Due to oversampling of individuals with overweight or obesity (≥ 27 kg/m^2), the data were weighed to the BMI distribution of the general population. Multicollinearity was reduced by centring all continuous variables around their respective means. Furthermore, we calculated the R^2 per PDAG per final model 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 <10 were excluded. In addition, we excluded related participants and participants with a non-Caucasian ethnicity to increase homogeneity (**Supplemental Table 9**).

Supplemental Table 6. The association between polyglutamine disease-associated genes (PDAGs) and body mass index (BMI) in the PROSPER cohort.

Gene	Variable	β-coefficient[a]	SE	t	p-value	95% CI		R^2
ATXN1	ATXN1_s	0.060	0.009	6.36	<.001	0.042	0.079	$0.82*10^{-3}$
	ATXN1_l	-0.002	0.020	-0.08	.937	-0.040	0.037	
	ATXN1_s2	0.033	0.013	2.50	.012	0.007	0.059	
	ATXN1_sl	-0.056	0.016	-3.53	<.001	-0.087	-0.025	
	ATXN1_s[b]	0.094	0.018	5.20	<.001	0.058	0.129	
	ATXN1_l[b]	0.009	0.027	0.33	.745	-0.044	0.062	
	ATXN1_s2[b]	0.023	0.021	1.08	.282	0.019	0.065	
	ATXN1_sl[b]	-0.065	0.017	-3.84	<.001	-0.098	-0.032	
ATXN2	ATXN2_s	0.671	0.136	4.92	<.001	0.403	0.938	$4.79*10^{-3}$
	ATXN2_l	0.106	0.035	3.00	.003	0.037	0.175	
	ATXN2_s2	0.147	0.042	3.49	<.001	0.065	0.230	
	ATXN2_s[b]	0.696	0.213	3.28	.001	0.280	1.113	
	ATXN2_l[b]	0.097	0.039	2.49	.013	0.021	0.172	
	ATXN2_s2[b]	0.150	0.062	2.40	.016	0.028	0.272	
ATXN3	ATXN3_s	-0.018	0.003	-6.61	<.001	-0.024	-0.013	$1.32*10^{-3}$
	ATXN3_l	0.015	0.011	1.37	.171	-0.006	0.036	
	ATXN3_l2	0.004	0.001	5.44	<.001	0.003	0.005	
	ATXN3_sl	-0.020	0.007	-2.99	.003	-0.034	-0.007	
	ATXN3_s[b]	-0.018	0.005	-3.68	<.001	-0.028	-0.009	
	ATXN3_l[b]	0.016	0.005	3.27	.001	0.007	0.026	
	ATXN3_l2[b]	0.005	0.001	6.93	<.001	0.004	0.007	
	ATXN3_sl[a]	-0.025	0.014	-1.87	.061	-0.052	0.001	
CACNA1A	CACNA1A_s	0.018	0.003	6.91	<.001	0.013	0.023	$0.18*10^{-3}$
	CACNA1A_l	-0.054	0.008	-6.39	<.001	-0.070	-0.037	
	CACNA1A_s[b]	0.015	0.008	1.91	.057	-0.000	0.030	
	CACNA1A_l[b]	-0.032	0.015	-2.14	.032	-0.061	-0.003	
ATXN7	ATXN7_s	0.130	0.014	9.03	<.001	0.102	0.158	$1.22*10^{-3}$
	ATXN7_l	-0.115	0.014	-8.51	<.001	-0.142	-0.089	
	ATXN7_l2	0.043	0.018	2.36	.018	0.007	0.078	
	ATXN7_s[b]	0.139	0.047	2.94	.003	0.046	0.231	
	ATXN7_l[b]	-0.108	0.016	-6.73	<.001	-0.139	-0.077	
	ATXN7_l2[b]	0.054	0.014	3.94	<.001	0.027	0.081	
TBP	TBP_s	0.031	0.015	2.03	.042	0.001	0.061	$0.67*10^{-3}$
	TBP_l	0.051	0.031	1.64	.101	-0.010	0.112	
	TBP_sl	0.022	0.006	3.52	<.001	0.010	0.034	
	TBP_s[b]	0.043	0.013	3.37	.001	0.018	0.067	
	TBP_l[b]	0.012	0.042	0.30	.764	-0.069	0.094	
	TBP_sl[b]	0.018	0.011	1.56	.120	-0.005	0.040	

Supplemental Table 6. (continued)

Gene	Variable	β-coefficient[a]	SE	t	p-value	95% CI		R^2
HTT	HTT_s	-0.056	0.012	-4.73	<.001	-0.079	-0.033	$2.15*10^{-3}$
	HTT_l	0.044	0.010	4.54	<.001	0.025	0.064	
	HTT_s2	-0.015	0.001	-10.86	<.001	-0.018	-0.012	
	HTT_sl	0.021	0.005	4.45	<.001	0.012	0.030	
	HTT_s[b]	-0.065	0.016	-4.04	<.001	-0.096	-0.033	
	HTT_l[b]	0.042	0.012	3.34	.001	0.017	0.066	
	HTT_s2[b]	-0.016	0.006	-2.59	.009	-0.029	-0.004	
	HTT_sl[b]	0.026	0.003	7.72	<.001	0.019	0.032	
ATN1	ATN1_s	-0.029	0.020	-1.44	.149	-0.069	0.010	
	ATN1_l	-0.005	0.054	-0.10	.924	-0.111	0.101	
	ATN1_s[b]	-0.041	0.027	-1.51	.131	-0.093	0.012	
	ATN1_l[b]	0.004	0.057	0.07	.942	-0.108	0.116	
AR ♂	AR_s	-0.001	0.011	-0.13	.899	-0.022	0.019	
	AR_s[c]	-0.003	0.006	-0.45	.652	-0.014	0.009	
AR ♀	AR_s	0.004	0.061	0.06	.952	-0.116	0.124	
	AR_l	0.002	0.065	0.03	.974	-0.125	0.129	
	AR_s[c]	0.001	0.057	0.01	.991	-0.112	0.113	
	AR_l[c]	-0.007	0.057	-0.12	.904	-0.118	0.104	
AR_s	AR_s	0.006	0.019	0.30	.763	-0.031	0.042	
	AR_s[b]	-0.000	0.020	-0.00	.999	-0.039	0.039	
AR_l	AR_l	-0.004	0.028	-0.14	.886	-0.059	0.051	
	AR_l[b]	-0.007	0.030	-0.23	.816	-0.066	0.052	

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. SE=standard error. CI=confidence interval. [a]) This column indicates the amount of BMI change in kg/m2 per unit CAG repeat size increase. [b]) Results were adjusted for age, sex and population structure using principal components. [c]) Results were adjusted for age only and population structure using principal components.

Statistical analysis

We applied a generalized mixed effect model with the CAG repeat length of the shorter and longer alleles as fixed effects and BMI as the outcome variable. Country (i.e. Scotland, Ireland or the Netherlands) was set as random effect to account for potential population stratification. To assess potential interaction or non-linear effects,[12,18] we included a product term of both alleles and a quadratic term for each allele as fixed effects. If the product term or the quadratic terms were not significantly associated with BMI, these variables were removed from the original model and the analysis was repeated until a final model was established. The final model was corrected for age and sex. Multicollinearity was reduced by centring all continuous variables around their respective means. Furthermore, we calculated the marginal R^2 per PDAG per final model

as described previously to determine the amount of variance explained by that gene.[21] 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 <10 were excluded. In addition, we excluded related participants and participants with a non-Caucasian ethnicity to increase homogeneity (**Supplemental Table 10**).

Supplemental Table 7. The association between the shorter and longer alleles of polyglutamine disease-associated genes (PDAGs) and body mass index (BMI) in the NEO cohort.

Gene	Variable	β-coefficient$_a$	SE	t	p-value	95% CI	
ATXN1	ATXN1_s	-0.033	0.096	-0.34	.731	-0.220	0.154
	ATXN1_l$_b$	-0.050	0.045	-1.11	.266	-0.137	0.038
ATXN2	ATXN2_s	-0.759	0.334	-2.27	.023	-1.415	-0.103
	ATXN2_l$_b$	0.086	0.174	0.50	.619	-0.254	0.427
ATXN3	ATXN3_s$_b$	-0.049	0.019	-2.61	.009	-0.086	-0.012
	ATXN3_l$_b$	0.053	0.022	2.38	.017	0.009	0.096
CACNA1A	CACNA1A_s	-0.016	0.037	-0.44	.659	-0.088	0.056
	CACNA1A_l$_b$	-0.056	0.074	-0.75	.451	-0.200	0.089
ATXN7	ATXN7_s$_b$	0.114	0.126	0.90	.367	-0.133	0.360
	ATXN7_l	0.010	0.062	0.17	.869	-0.111	0.132
TBP	TBP_s$_b$	0.061	0.047	1.31	.189	-0.030	0.152
	TBP_l	-0.205	0.085	-2.40	.016	-0.372	-0.038
HTT	HTT_s	0.060	0.040	1.48	.139	-0.019	0.138
	HTT_l	-0.041	0.023	-1.79	.074	-0.086	0.004
ATN1	ATN1_s$_b$	-0.000	0.026	-0.01	.995	-0.050	0.050
	ATN1_l	0.046	0.037	1.25	.211	-0.026	0.118
AR ♂	AR$_b$	-0.027	0.035	-0.75	.452	-0.096	0.043
AR ♀	AR_s	-0.061	0.054	-1.12	.261	-0.166	0.045
	AR_l$_b$	0.017	0.046	0.36	.718	-0.074	0.107
AR (short)	AR_s	-0.035	0.029	-1.23	.220	-0.092	0.021
AR (long)	AR_l$_b$	-0.012	0.027	-0.44	.658	-0.066	0.042

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. SE=standard error. CI=confidence interval. *AR* ♂=*AR* assessed in males. *AR* ♀=*AR* assessed in females. *AR* (short)= the shorter *AR* allele assessed in both males and females. *AR* (long)= the longer *AR* allele assessed in both males and females. $_a$) This column indicates the amount of BMI change in kg/m2 per unit CAG repeat size increase. $_b$) The allele that was directionally consistent to allele in the PROSPER cohort.

Statistical analysis

We applied a linear regression model with the CAG repeat lengths of the shorter and longer alleles as independent variables and BMI as dependent variable. Due to oversampling of individuals with overweight or obesity (≥ 27 kg/m^2), the data were weighed to the BMI distribution of the general population. Multicollinearity was reduced by centring all continuous variables 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 <10 were excluded. In addition, we excluded related participants and participants with a non-Caucasian ethnicity to increase homogeneity (**Supplemental Table 9**).

Supplemental Table 8. The association between the shorter and longer alleles of polyglutamine disease-associated genes (PDAGs) and body mass index (BMI) in the PROSPER cohort.

Gene	Variable	β-coefficient[a]	SE	t	p-value	95% CI	
ATXN1	ATXN1_s	0.040	0.012	3.40	.001	0.017	0.063
	ATXN1_l[b]	-0.006	0.022	-0.27	.791	-0.049	0.037
ATXN2	ATXN2_s	0.072	0.071	1.02	.309	-0.066	0.210
	ATXN2_l[b]	0.126	0.033	3.79	<.001	0.061	0.191
ATXN3	ATXN3_s[b]	-0.020	0.003	-7.69	<.001	-0.026	-0.015
	ATXN3_l[b]	0.003	0.008	0.39	.699	-0.013	0.019
CACNA1A	CACNA1A_s	0.018	0.003	6.91	<.001	0.013	0.023
	CACNA1A_l[b]	-0.054	0.008	-6.39	<.001	-0.070	-0.037
ATXN7	ATXN7_s[b]	0.130	0.014	9.05	<.001	0.102	0.158
	ATXN7_l	-0.063	0.035	-1.81	.070	-0.131	0.005
TBP	TBP_s[b]	0.027	0.014	1.90	.057	-0.001	0.055
	TBP_l	0.044	0.030	1.45	.147	-0.015	0.103
HTT	HTT_s	-0.014	0.005	-2.73	.006	-0.025	-0.004
	HTT_l	0.038	0.011	3.36	.001	0.016	0.061
ATN1	ATN1_s[b]	-0.029	0.020	-1.44	.149	-0.069	0.010
	ATN1_l	-0.005	0.054	-0.10	.924	-0.111	0.101
AR ♂	AR[b]	-0.001	0.011	-0.13	.899	-0.022	0.019
AR ♀	AR_s	0.004	0.061	0.06	.952	-0.116	0.124
	AR_l[b]	0.002	0.065	0.03	.974	-0.125	0.129
AR (short)	AR_s	0.006	0.019	0.30	.763	-0.031	0.042
AR (long)	AR_l[b]	-0.004	0.028	-0.14	.886	-0.059	0.051

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. SE=standard error. CI=confidence interval. *AR* ♂=*AR* assessed in males. *AR* ♀=*AR* assessed in females. *AR* (short)= the shorter *AR* allele assessed in both males and females. *AR* (long)= the longer *AR* allele assessed in both males and females. [a]) This column indicates the amount of BMI change in kg/m2 per unit CAG repeat size increase. [b]) The allele that was directionally consistent to allele in the NEO cohort.

Statistical analysis

We applied a generalized mixed effect model with the CAG repeat length of the shorter and longer alleles as fixed effects and BMI as the outcome variable. Country (i.e. Scotland, Ireland or the Netherlands) was set as random effect to account for potential population stratification. Multicollinearity was reduced by centring all continuous variables 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 <10 were excluded. In addition, we excluded related participants and participants with a non-Caucasian ethnicity to increase homogeneity (**Supplemental Table 10**).

Supplemental Table 9. Genotyped subjects per gene and number of excluded cases for the NEO cohort.

Gene	Total # of Subjects[a]	Included range shorter allele[b]	Included range longer allele[b]	# Excluded cases (%)[c]
ATXN1	6438	26-32	28-38	839 (13.0)
ATXN2	6389	17-23	22-29	831 (13.0)
ATXN3	6494	14-28	14-37	818 (12.6)
CACNA1A	6394	4-14	7-15	807 (12.6)
ATXN7	6356	7-13	10-15	800 (14.4)
TBP	6418	29-39	35-42	827 (12.9)
HTT	6453	9-24	15-32	857 (13.3)
ATN1	6467	8-17	8-24	844 (13.1)
AR (short)	6340	15-29	-	855 (13.5)
AR (long)	6340	-	16-31	826 (13.0)
AR ♂	3010	16-29	-	390 (13.0)
AR ♀	3330	14-26	18-31	486 (14.6)

AR (short)= the shorter *AR* allele assessed in both males and females. *AR* (long)= the longer *AR* allele assessed in both males and females. *AR* ♂=*AR* assessed in males. *AR* ♀=*AR* assessed in females. [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 357-609 and were missing completely at random. [b]) Range of CAG repeats with frequencies of 10 or more. [c]) Number of cases excluded due to CAG repeat lengths not being within the range of CAG repeats with frequencies of 10 or more, participants being related (n=419) or of non-Caucasian ethnicity (n=417).

Supplemental Table 10. Genotyped subjects per gene and number of excluded cases for the PROSPER cohort.

Gene	Total # of Subjects[a]	Included range shorter allele[b]	Included range longer allele[b]	# Excluded cases (%)[c]
ATXN1	5633	26-32	28-38	54 (1.0)
ATXN2	5548	17-23	22-29	43 (0.8)
ATXN3	5543	14-28	14-37	31 (0.6)
CACNA1A	5633	4-14	7-15	18 (0.3)
ATXN7	5285	7-13	10-15	9 (0.2)
TBP	5561	29-39	35-42	46 (0.8)
HTT	5602	9-24	15-32	59 (1.1)
ATN1	5633	8-17	8-24	34 (0.6)
AR (short)	5509	15-29	-	68 (1.2)
AR (long)	5509	-	16-31	42 (0.8)
AR ♂	2662	16-29	-	41 (1.5)
AR ♀	2847	14-26	18-31	41 (1.4)

AR (short)= the shorter *AR* allele assessed in both males and females. *AR* (long)= the longer *AR* allele assessed in both males and females. *AR* ♂=*AR* assessed in males. *AR* ♀=*AR* assessed in females. [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 357-609 and were missing completely at random. [b]) Range of CAG repeats with frequencies of 10 or more. [c]) Number of cases excluded due to CAG repeat lengths not being within the range of CAG repeats with frequencies of 10 or more (no participants were related and all participants were of Caucasian ethnicity).

Supplemental Table 11. Genotyped subjects per gene and number of excluded cases for the combined cohort.

Gene	Total # of Subjects[a]	Included range shorter allele[b]	Included range longer allele[b]	# Excluded cases (%)[c]
ATXN1	12 071	26-32	28-38	893 (7.4)
ATXN2	11 937	17-23	22-29	874 (7.3)
ATXN3	12 036	14-28	14-37	848 (7.0)
CACNA1A	12 027	4-14	7-15	825 (6.9)
ATXN7	11 641	7-13	10-15	809 (6.9)
TBP	11 979	29-39	35-42	873 (7.3)
HTT	12 055	9-24	15-32	916 (7.6)
ATN1	12 100	8-17	8-24	878 (7.3)
AR (short)	11 849	15-29	-	923 (7.8)
AR (long)	11 849	-	16-31	868 (7.3)
AR ♂	5 672	16-29	-	431 (7.6)
AR ♀	6 177	14-26	18-31	527 (8.5)

AR (long)= the longer *AR* allele assessed in both males and females. *AR* ♂=*AR* assessed in males. *AR* ♀=*AR* assessed in females. [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 357-609 and were missing completely at random. [b]) Range of CAG repeats with frequencies of 10 or more. [c]) Number of cases excluded due to CAG repeat lengths not being within the range of CAG repeats with frequencies of 10 or more, participants being related (n=419) or of non-Caucasian ethnicity (n=417).

Supplemental Table 12. The association between polyglutamine disease-associated genes (PDAGs) and body mass index (BMI) corrected for age, sex and population structure in the combined cohort.

Gene	Variable	β-coefficient[a]	SE	t	p-value	95% CI	
ATXN1	ATXN1_s	-0.031	0.023	-1.33	.182	-0.076	0.014
	ATXN1_l	0.087	0.028	3.12	.002	0.032	0.142
	ATXN1_l2	-0.036	0.011	-3.24	.001	-0.058	-0.014
ATXN2	ATXN2_s	-0.062	0.067	-0.92	.358	-0.193	0.070
	ATXN2_l	0.111	0.004	29.80	<.001	0.104	0.119
ATXN3	ATXN3_s	-0.040	0.010	-4.10	<.001	-0.059	-0.021
	ATXN3_l	0.050	0.017	2.94	.003	0.017	0.083
	ATXN3_l2	0.003	0.001	2.78	.005	0.001	0.005
	ATXN3_sl	-0.021	0.002	-13.81	<.001	-0.024	-0.018
CACNA1A	CACNA1A_s	-0.002	0.004	-0.53	.595	-0.011	0.006
	CACNA1A_l	-0.047	0.014	-3.42	.001	-0.074	0.020
	CACNA1A_l2	0.001	0.002	0.39	.696	-0.003	0.005
ATXN7	ATXN7_s	0.110	0.012	9.44	<.001	0.087	0.132
	ATXN7_l	-0.024	0.015	-1.59	.112	-0.054	0.006
	ATXN7_s2	0.068	0.001	63.61	<.001	0.066	0.070
TBP	TBP_s	-0.010	0.020	-0.51	.607	-0.048	0.028
	TBP_l	-0.132	0.066	-1.99	.046	-0.261	-0.002
	TBP_s2	-0.014	0.004	-3.36	.001	-0.023	-0.006
ATN1	ATN1_s	-0.014	0.015	-0.93	.352	-0.043	0.015
	ATN1_l	0.032	0.014	2.35	.019	0.005	0.059
AR ♂	AR	-0.022	0.011	-1.90	.057	-0.044	0.001
	AR_2	-0.001	0.000	-3.93	<.001	-0.001	-0.000
AR ♀	AR_s	-0.067	0.036	-1.85	.064	-0.138	0.004
	AR_l	0.036	0.019	1.95	.051	-0.000	0.073
	AR_l2	-0.013	0.007	-1.90	.057	-0.014	0.000
AR (long)	AR_l	-0.010	0.003	-3.12	.002	-0.017	-0.004

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. BMI = body mass index, PDAGs= polyglutamine disease-associated genes, SE=standard error. CI=confidence interval. [a]) This column indicates the amount of BMI change in kg/m2 per unit CAG repeat size increase.