



Repeat length variations in polyglutamine disease-associated genes affect body mass index

Sarah L. Gardiner^{1,2} · Renée de Mutsert³ · Stella Trompet⁴ · Merel W. Boogaard^{1,5} · Ko Willems van Dijk^{2,4} · P. J. Wouter Jukema⁶ · P. Eline Slagboom⁷ · Raymund A. C. Roos¹ · Hanno Pijl⁴ · Frits R. Rosendaal³ · N. Ahmad Aziz^{1,8}

Received: 27 November 2017 / Revised: 15 May 2018 / Accepted: 15 June 2018 / Published online: 17 August 2018
© Springer Nature Limited 2018

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 analyzed their association with BMI, 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

Electronic supplementary material The online version of this article (<https://doi.org/10.1038/s41366-018-0161-7>) contains supplementary material, which is available to authorized users.

✉ Sarah L. Gardiner
s.l.gardiner@lumc.nl

¹ Department of Neurology, Leiden University Medical Centre, 2300 RC Leiden, The Netherlands

² Department of Human Genetics, Leiden University Medical Centre, 2300 RC Leiden, The Netherlands

³ Department of Epidemiology, Leiden University Medical Centre, 2300 RC Leiden, The Netherlands

⁴ Department of Internal Medicine, Leiden University Medical

Centre, 2300 RC Leiden, The Netherlands

⁵ Department of Clinical Genetics, Leiden University Medical Centre, 2300 RC Leiden, The Netherlands

⁶ Department of Cardiology, Leiden University Medical Centre, 2300 RC Leiden, The Netherlands

⁷ Department of Molecular Epidemiology, Leiden University Medical Centre, 2300 RC Leiden, The Netherlands

⁸ German Center for Neurodegenerative Diseases (DZNE), Bonn 53127, Germany

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

			CAG repeat ranges							
Gene	Disease	Protein	Normal	Pathological	Allele	Mean	Median	<i>N</i>	Range	
<i>ATXN1</i>	SCA1	Ataxin-1	6–39	41–83	Short	29.21	29.00	12071	17–36	
					Long	30.79	30.00	12071	22–44	
<i>ATXN2</i>	SCA2	Ataxin-2	14–32	33–500	Short	21.93	22.00	11937	11–30	
					Long	22.39	22.00	11937	17–36	
<i>ATXN3</i>	SCA3	Ataxin-3	12–44	52–87	Short	19.02	20.00	12037	14–36	
					Long	24.25	23.00	12036	14–62	
<i>CACNA1A</i>	SCA6	CACNA1A	4–18	20–33	Short	10.58	11.00	12027	4–14	
					Long	12.47	13.00	12027	4–22	
<i>ATXN7</i>	SCA7	Ataxin-7	3–19	37–460	Short	10.05	10.00	11641	5–16	
					Long	10.82	10.00	11641	7–30	
<i>TBP</i>	SCA17	TBP	25–43	45–66	Short	36.34	37.00	11979	21–40	
					Long	37.88	38.00	11979	21–47	
<i>HTT</i>	HD	Huntington	6–26	36–121	Short	16.91	17.00	12055	6–31	
					Long	20.18	19.00	12055	10–40	
<i>ATN1</i>	DRPLA	Atrophin-1	3–38	48–93	Short	12.36	14.00	12100	3–22	
					Long	15.53	15.00	12100	8–28	
<i>AR</i>	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’s disease, *DRPLA* dentatorubropallidolusian atrophy, *SBMA* spinal bulbar muscular atrophy

expansions can result in phenotypic consequences for organisms [9]. Nine hereditary neurodegenerative diseases, including Huntington’s 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 Supplementary 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/m² 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 nine tested PDAGs), this 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 (Bioglio) (Supplementary Table 4). The PCR was performed using 10 ng of genomic DNA, 1× OneTaq mastermix (New England Biolabs, OneTaq Hot start with GC Buffer master mix), 1 µl of primer Mix A or B (Supplementary Table 4) and Aqua B. Braun water to a final volume of 10 µl. The PCR was run for 27 cycles of 30 s, denaturation at 94 °C, 1 min of annealing at 60 °C and 2 min elongation at 68 °C, preceded by 5 min of initial denaturation at 94 °C. Final elongation was performed at 69 °C for 5 min. 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 analyzed 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 (Supplementary Tables 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 (Supplementary Tables 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 principle 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 (Supplementary Tables 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 0.05 [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 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 ten, participants being related or of non-Caucasian ethnicity (Supplementary Tables 9–11), leaving a total of 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*) (Supplementary Table 5). Seven PDAGs in the PROSPER cohort were significantly associated with BMI (including *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP* and *HTT*) (Supplementary Table 6). Between the two cohorts, the effect on BMI of at

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	<i>t</i>	<i>p</i> -Value	95% CI		<i>R</i> ²
<i>ATXN1</i>	<i>ATXN1_s</i>	-0.058	0.047	-1.24	0.214	-0.150	0.034	1.84×10 ⁻³
	<i>ATXN1_l</i>	0.078	0.029	2.74	0.006	0.022	0.134	
	<i>ATXN1_l2</i>	-0.036	0.011	-3.26	0.001	-0.058	-0.015	
<i>ATXN2</i>	<i>ATXN2_s</i>	-0.068	0.063	-1.08	0.282	-0.191	0.056	0.55×10 ⁻³
	<i>ATXN2_l</i>	0.081	0.024	3.35	0.001	0.034	0.129	
<i>ATXN3</i>	<i>ATXN3_s</i>	-0.039	0.011	-3.58	<0.001	-0.061	-0.018	2.56×10 ⁻³
	<i>ATXN3_l</i>	0.048	0.018	2.73	0.006	0.014	0.082	
	<i>ATXN3_l2</i>	0.003	0.001	4.14	<0.001	0.002	0.004	
	<i>ATXN3_sl</i>	-0.020	0.000	-76.09	<0.001	-0.021	-0.020	
<i>CACNA1A</i>	<i>CACNA1A_s</i>	-0.007	0.013	-0.53	0.599	-0.033	0.019	0.31×10 ⁻³
	<i>CACNA1A_l</i>	-0.038	0.006	-5.99	<0.001	-0.050	-0.025	
	<i>CACNA1A_l2</i>	0.010	0.005	2.24	0.025	0.001	0.019	
<i>ATXN7</i>	<i>ATXN7_s</i>	0.122	0.013	9.14	<0.001	0.095	0.148	0.56×10 ⁻³
	<i>ATXN7_l</i>	-0.019	0.026	-0.73	0.466	-0.071	0.033	
	<i>ATXN7_s2</i>	0.057	0.008	7.14	<0.001	0.041	0.073	
<i>TBP</i>	<i>TBP_s</i>	0.011	0.001	10.87	<0.001	0.009	0.014	1.24×10 ⁻³
	<i>TBP_l</i>	-0.126	0.087	-1.46	0.145	-0.296	0.044	
	<i>TBP_s2</i>	-0.012	0.004	-3.17	0.002	-0.019	-0.005	
<i>ATN1</i>	<i>ATN1_s</i>	-0.008	0.011	-0.70	0.486	-0.030	0.014	
	<i>ATN1_l</i>	0.032	0.021	1.51	0.130	-0.009	0.073	
<i>AR</i> ♂	<i>AR</i>	-0.018	0.011	-1.68	0.093	-0.039	0.003	0.23×10 ⁻³
	<i>AR_2</i>	-0.003	0.001	-5.00	<0.001	-0.003	-0.002	
<i>AR</i> ♀	<i>AR_s</i>	-0.054	0.028	-1.95	0.052	-0.109	0.000	1.41×10 ⁻³
	<i>AR_l</i>	0.025	0.010	2.45	0.014	0.005	0.045	
	<i>AR_l2</i>	-0.012	0.006	-2.11	0.035	-0.024	-0.001	
<i>AR</i> (long)	<i>AR_l</i>	-0.011	0.003	-3.41	0.001	-0.017	-0.005	0.05×10 ⁻³

^aThis column indicates the amount of BMI change in kg/m² per unit CAG repeat size increase

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

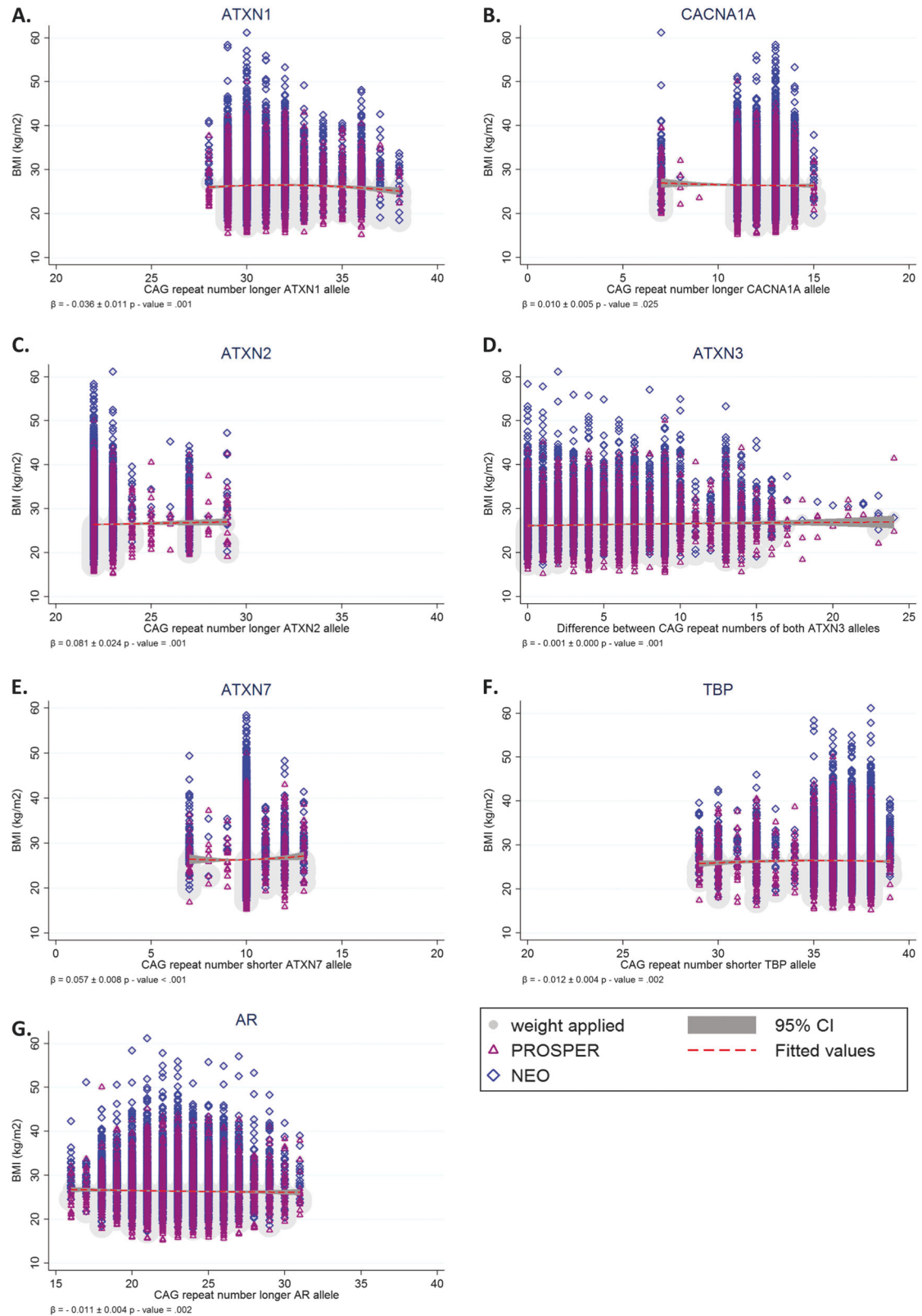
least one allele was in the same direction for eight PDAGs (Supplementary Table 7 and 8). The data of only these directionally consistent alleles were combined. The effects of both *HTT* alleles were not consistent and therefore not combined (Table 2). After combining the data of the directionally consistent alleles, we found a total of seven PDAGs (including *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP* and *AR*) to be significantly associated with BMI (Table 2). For 5744 participants in the NEO and 5244 participants in the PROSPER cohorts, we obtained principle components generated from genome-wide genotyping data as described before [19, 20]. We corrected for age, sex and population structure using these principle components. This correction did not substantially alter our results (Supplementary 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 (Fig. 1a, b). The longer allele of *ATXN2* was associated with BMI in a linear fashion. Higher numbers of CAG repeats were associated with a higher BMI (Fig. 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 (Fig. 1d). Furthermore, the shorter alleles of both *ATXN7* and *TBP* had a quadratic association with BMI (Fig. 1e, f). Lastly, we examined the effect on BMI of the CAG repeat size in the X-linked *AR* gene, for which we (1) analyzed 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

analyzing the *AR* gene in men and women combined, a longer *AR* CAG repeat size in the longer allele was also associated with lower BMI (Fig. 1g). To estimate the total



◀ **Fig. 1** Scatterplots of the association between body mass index (BMI) 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. Beta-coefficient \pm SE. CI confidence interval

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/m² (corresponding to 1.29 kg for an individual 1.75 m in height) (Fig. 2).

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

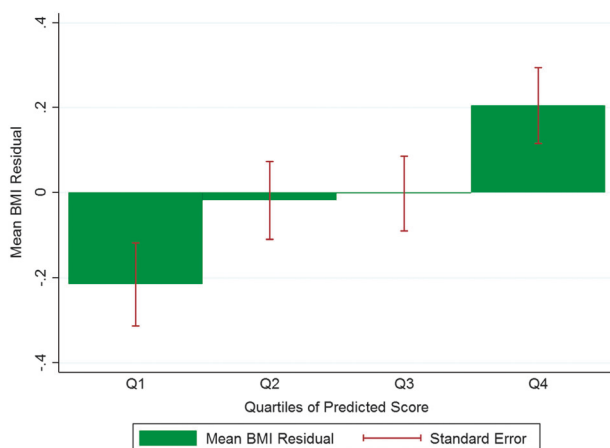


Fig. 2 The effect of CAG repeat size variations in polyglutamine disease-associated genes (PDAGs) on body mass index (BMI). This plot illustrates that in combination, CAG repeat size variations in only seven PDAGs can account for a variation of up to ~ 0.42 kg/m² in BMI. Please refer to the Methods section for details on how the ‘Predicted Score’ was constructed

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 single nucleotide polymorphism (SNP) located in the *A2BPI* 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 single nucleotide polymorphism 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 promoters, 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/m² and explained a variance in BMI of $\sim 1\%$, 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/m² per allele and explained $\sim 0.14\%$ 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 linkage disequilibrium (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 linkage disequilibrium 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 linkage disequilibrium 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 to 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 Skłodowska-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'.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. WHO. 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. *Curr Opin Lipidol.* 1997;8:205–11.
3. Chung WK. An overview of monogenic and syndromic obesities in humans. *Pediatr Blood Cancer.* 2012;58:122–8.
4. Goran MI. Genetic influences on human energy expenditure and substrate utilization. *Behav Genet.* 1997;27:389–99.
5. Locke AE, Kahali B, Berndt SI, Justice AE, Pers TH, Day FR, et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature.* 2015;518:197–206.
6. Hannan AJ. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for 'missing heritability'. *Trends Genet.* 2010;26:59–65.
7. Hannan AJ. TRPing up the genome: tandem repeat polymorphisms as dynamic sources of genetic variability in health and disease. *Discov Med.* 2010;10:314–21.
8. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med.* 2010;363:166–76.
9. Duitama J, Zablotskaya A, Gemayel R, Jansen A, Belet S, Vermeesch JR, et al. Large-scale analysis of tandem repeat variability in the human genome. *Nucleic Acids Res.* 2014;42:5728–41.
10. Bettencourt C, Hensman-Moss D, Flower M, Wiethoff S, Brice A, Goizet C, et al. DNA repair pathways underlie a common genetic mechanism modulating onset in polyglutamine diseases. *Ann Neurol.* 2016;79:983–90.
11. Fan HC, Ho LI, Chi CS, Chen SJ, Peng GS, Chan TM, et al. Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell Transplant.* 2014;23:441–58.
12. Gardiner SL, van Belzen MJ, Boogaard MW, van Roon-Mom WMC, Rozing MP, van Hemert AM, et al. Large normal-range TBP and ATXN7 CAG repeat lengths are associated with increased lifetime risk of depression. *Transl Psychiatry.* 2017;7:e1143.
13. Gardiner SL, van Belzen MJ, Boogaard MW, van Roon-Mom WMC, Rozing MP, van Hemert AM, et al. Huntington gene repeat size variations affect risk of lifetime depression. *Transl Psychiatry.* 2017;7:1277–84.
14. Killoran A, Biglan KM, Jankovic J, Eberly S, Kayson E, Oakes D, et al. Characterization of the Huntington intermediate CAG repeat expansion phenotype in PHAROS. *Neurology.* 2013;80:2022–7.
15. de Mutsert R, den Heijer M, Rabelink TJ, Smit JW, Romijn JA, Jukema JW, et al. The Netherlands Epidemiology of Obesity (NEO) study: study design and data collection. *Eur J Epidemiol.* 2013;28:513–23.
16. Shepherd J, Blauw GJ, Murphy MB, Bollen EL, Buckley BM, Cobbe SM, et al. Pravastatin in elderly individuals at risk of vascular disease (PROSPER): a randomised controlled trial. *Lancet.* 2002;360:1623–30.
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. *Behav Res Methods.* 2007;39:175–91.
18. Tezenas du MS, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain.* 2014;137:2444–55.
19. Shiffman D, Trompet S, Louie JZ, Rowland CM, Catanese JJ, Iakoubova OA, et al. Genome-wide study of gene variants associated with differential cardiovascular event reduction by pravastatin therapy. *PLoS ONE.* 2012;7:e38240.
20. Blauw LL, Noordam R, Trompet S, Berbee JFP, Rosendaal FR, van Heemst D, et al. Genetic variation in the obesity gene FTO is not associated with decreased fat oxidation: the NEO study. *Int J Obes (Lond).* 2017;41:1594–600.
21. Nakagawa S, Schielzeth H. A general and simple method for obtaining R² from generalized linear mixed-effects models. *Methods Ecol Evol.* 2013;4:9.
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B (Methodol).* 1995;57:289–300.
23. Aziz NA, van der Marck MA, Pijl H, Olde Rikkert MG, Bloem BR, Roos RA. Weight loss in neurodegenerative disorders. *J Neurol.* 2008;255:1872–80.
24. Saute JA, da Silva AC, Muller AP, Hansel G, de Mello AS, Maeda F, et al. Serum insulin-like system alterations in patients with spinocerebellar ataxia type 3. *Mov Disord.* 2011;26:731–5.
25. Saute JA, Silva AC, Souza GN, Russo AD, Donis KC, Vedolin L, et al. Body mass index is inversely correlated with the expanded CAG repeat length in SCA3/MJD patients. *Cerebellum.* 2012;11:771–4.
26. Schmitz-Hubsch T, Coudert M, Bauer P, Giunti P, Globas C, Baliko L, et al. Spinocerebellar ataxia types 1, 2, 3, and 6: disease severity and nonataxia symptoms. *Neurology.* 2008;71:982–9.
27. Mahler A, Steiniger J, Endres M, Paul F, Boschmann M, Doss S. Increased catabolic state in spinocerebellar ataxia type 1 patients. *Cerebellum.* 2014;13:440–6.
28. Ashkenazi A, Bento CF, Ricketts T, Vicinanza M, Siddiqi F, Pavel M, et al. Polyglutamine tracts regulate beclin 1-dependent autophagy. *Nature.* 2017;545:108–11.
29. Gustafson DR, Wen MJ, Koppanati BM. Androgen receptor gene repeats and indices of obesity in older adults. *Int J Obes Relat Metab Disord.* 2003;27:75–81.

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;46:31–9.
31. Kiehl TR, Nechiporuk A, Figueroa KP, Keating MT, Huynh DP, Pulst SM. Generation and characterization of Sca2 (ataxin-2) knockout mice. *Biochem Biophys Res Commun*. 2006;339:17–24.
32. Lastres-Becker I, Brodessa S, Lutjohann D, Azizov M, Buchmann J, Hintermann E, et al. Insulin receptor and lipid metabolism pathology in ataxin-2 knock-out mice. *Hum Mol Genet*. 2008;17:1465–81.
33. Shibata H, Huynh DP, Pulst SM. A novel protein with RNA-binding motifs interacts with ataxin-2. *Hum Mol Genet*. 2000;9:1303–13.
34. Carmo-Silva S, Nobrega C, Pereira de Almeida L, Cavadas C. Unraveling the role of ataxin-2 in metabolism. *Trends Endocrinol Metab*. 2017;28:309–18.
35. Arkova OV, Ponomarenko MP, Rasskazov DA, Drachkova IA, Arshinova TV, Ponomarenko PM, 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:S5. Suppl 13
36. Luppino FS, de Wit LM, Bouvy PF, Stijnen T, Cuijpers P, Penninx BW, et al. Overweight, obesity, and depression: a systematic review and meta-analysis of longitudinal studies. *Arch Gen Psychiatry*. 2010;67:220–9.
37. Helmlinger D, Hardy S, Eberlin A, Devys D, Tora L. Both normal and polyglutamine-expanded ataxin-7 are components of TFIIIC-type GCN5 histone acetyltransferase-containing complexes. *Biochem Soc Symp*. 2006;73:155–63.
38. Emery CF, Fondow MD, Schneider CM, Christofi FL, Hunt C, Busby AK, et al. Gastric bypass surgery is associated with reduced inflammation and less depression: a preliminary investigation. *Obes Surg*. 2007;17:759–63.
39. Shoelson SE, Herrero L, Naaz A. Obesity, inflammation, and insulin resistance. *Gastroenterology*. 2007;132:2169–80.
40. Vaccarino V, Johnson BD, Sheps DS, Reis SE, Kelsey SF, Bittner V, et al. Depression, inflammation, and incident cardiovascular disease in women with suspected coronary ischemia: the National Heart, Lung, and Blood Institute-sponsored WISE study. *J Am Coll Cardiol*. 2007;50:2044–50.
41. Bremner MA, Beekman AT, Deeg DJ, Penninx BW, Dik MG, Hack CE, et al. Inflammatory markers in late-life depression: results from a population-based study. *J Affect Disord*. 2008;106:249–55.
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;65:973–8.
43. Pasquali R, Vicennati V. Activity of the hypothalamic-pituitary-adrenal axis in different obesity phenotypes. *Int J Obes Relat Metab Disord*. 2000;24 Suppl 2:S47–9.
44. Walker BR. Activation of the hypothalamic-pituitary-adrenal axis in obesity: cause or consequence?. *Growth Horm IGF Res*. 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. *Am J Psychiatry*. 2001;158:1612–6.
46. Holsboer F. The corticosteroid receptor hypothesis of depression. *Neuropsychopharmacology*. 2000;23:477–501.
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. *Obes Surg*. 2008;18:1119–25.
48. Huber JD. Diabetes, cognitive function, and the blood-brain barrier. *Curr Pharm Des*. 2008;14:1594–600.
49. Ajilore O, Haroon E, Kumaran S, Darwin C, Binesh N, Mintz J, et al. Measurement of brain metabolites in patients with type 2 diabetes and major depression using proton magnetic resonance spectroscopy. *Neuropsychopharmacology*. 2007;32:1224–31.
50. Atlantis E, Ball K. Association between weight perception and psychological distress. *Int J Obes (Lond)*. 2008;32:715–21.
51. Derenne J, Beresin E. Body image, media, and eating disorders—a 10-year update. *Acad Psychiatry*. 2018;42:129–34.
52. Beesdo K, Jacobi F, Hoyer J, Low NC, Hoffer M, Wittchen HU. Pain associated with specific anxiety and depressive disorders in a nationally representative population sample. *Soc Psychiatry Psychiatr Epidemiol*. 2010;45:89–104.
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. *J Psychiatr Res*. 2018;97:38–46.
54. Zitzmann M. Mechanisms of disease: pharmacogenetics of testosterone therapy in hypogonadal men. *Nat Clin Pract Urol*. 2007;4:161–6.
55. Colangelo LA, Sharp L, Kopp P, Scholtens D, Chiu BC, Liu K, et al. Total testosterone, androgen receptor polymorphism, and depressive symptoms in young black and white men: the CARDIA Male Hormone Study. *Psychoneuroendocrinology*. 2007;32:951–8.
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. *Eur J Endocrinol*. 2010;163:319–28.
57. Higgs S, Spetter MS. Cognitive control of eating: the role of memory in appetite and weight gain. *Curr Obes Rep*. 2018;7:50–9.
58. Frayling TM, Timpson NJ, Weedon MN, Zeggini E, Freathy RM, Lindgren CM, et al. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science*. 2007;316:889–94.
59. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet*. 2008;40:768–75.
60. Hannan AJ. Tandem repeats mediating genetic plasticity in health and disease. *Nat Rev Genet*. 2018;19:286–98.
61. Saini S, Mitra I, Gymrek M. A reference haplotype panel for genome-wide imputation of short tandem repeats. *bioRxiv*. 2018 1–23.
62. Mittal U, Sharma S, Chopra R, Dheeraj K, Pal PK, Srivastava AK, et al. Insights into the mutational history and prevalence of SCA1 in the Indian population through anchored polymorphisms. *Hum Genet*. 2005;118:107–14.
63. Krysa W, Sulek A, Rakowicz M, Szirkowicz W, Zaremba J. High relative frequency of SCA1 in Poland reflecting a potential founder effect. *Neurol Sci*. 2016;37:1319–25.
64. Craig K, Takiyama Y, Soong BW, Jardim LB, Saraiva-Pereira ML, Lythgow K, et al. Pathogenic expansions of the SCA6 locus are associated with a common CACNA1A haplotype across the globe: founder effect or predisposing chromosome? *Eur J Human Genet*. 2008;16:841–7.
65. Faruq M, Magana JJ, Suroliya V, Narang A, Murillo-Melo NM, Hernandez-Hernandez O, 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. *Ann Hum Genet*. 2017;81:197–204.
66. Magana JJ, Gomez R, Maldonado-Rodriguez M, Velazquez-Perez L, Tapia-Guerrero YS, Cortes H, et al. Origin of the spinocerebellar ataxia type 7 gene mutation in Mexican population. *Cerebellum*. 2013;12:902–5.
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 Res.* 2005;65:5974–81.
68. Kashi Y, King DG. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 2006;22:253–9.
69. Keiser MS, Kordasiewicz HB, McBride JL. Gene suppression strategies for dominantly inherited neurodegenerative diseases: lessons from Huntington's disease and spinocerebellar ataxia. *Hum Mol Genet.* 2016;25:R53–64.