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

Gardiner, S.L.

Citation

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

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

Part 4.2

LARGE NORMAL-RANGE *TBP* AND *ATXN7* CAG REPEAT LENGTHS ARE ASSOCIATED WITH INCREASED LIFETIME RISK OF DEPRESSION

Sarah L. Gardiner, MD[1,2]; Martine J. van Belzen, PhD[3];
Merel W. Boogaard[1,3]; Willeke M.C. van Roon-Mom, PhD[2];
Maarten P. Rozing, MD, PhD[5]; Albert M. van Hemert, MD, PhD[4];
Johannes H. Smit, PhD[6]; Aartjan T.F. Beekman, MD, PhD[6];
Gerard van Grootheest, MSc[6]; Robert A. Schoevers, MD, PhD[7];
Richard C. Oude Voshaar, MD, PhD[7]; Hannie C. Comijs, PhD[6];
Brenda W.J.H. Penninx, PhD[6]; Roos C. van der Mast, MD, PhD[4,8];
Raymund A.C. Roos, MD, PhD[1] and N. Ahmad Aziz, MD, PhD[1]

Departments of [1]Neurology, [2]Human Genetics, [3]Clinical Genetics and
[4]Psychiatry, Leiden University Medical Centre, Leiden, the Netherlands
[5]Center for Healthy Aging/ Department of Public Health, Section of
Social Medicine, University of Copenhagen, Copenhagen, Denmark
[6]Department of Psychiatry, and Amsterdam Public Health Research
Institute and Neuroscience Campus Amsterdam, VU University Medical
Centre/GGZ inGeest, Amsterdam, The Netherlands.
[7]University of Groningen, University Medical Centre Groningen,
Department of Psychiatry, Research School Cognitive Behavioural
Neuroscience (BCN), Groningen, The Netherlands
[8]Department of Psychiatry, Collaborative Antwerp Psychiatric Research
Institute (CAPRI), University of Antwerp, Antwerp, Belgium

Translational psychiatry. 2017; 7(6): e1143

ABSTRACT

Depression is one of the most prevalent and debilitating psychiatric disorders worldwide. Recently, we showed that both relatively short and relatively long cytosine-adenine-guanine (CAG) repeats in the huntingtin gene (*HTT*) are associated with an increased risk of lifetime depression. However, to what extent the variations in CAG repeat length in the other eight polyglutamine disease-associated genes (PDAGs) are associated with depression is still unknown. We determined the CAG repeat sizes of *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *ATN1* and *AR* in two well-characterized Dutch cohorts – the Netherlands Study of Depression and Anxiety and the Netherlands Study of Depression in Older Persons – including 2165 depressed and 1058 non-depressed individuals – aged 18-93 years. The association between PDAG CAG repeat size and the risk for depression was assessed via binary logistic regression. We found that the odds ratio (OR) for lifetime depression was significantly higher for individuals with > 10, compared to subjects with ≤ 10, CAG repeats in both *ATXN7* alleles (OR=1.90, CI 1.26-2.85). For *TBP* we found a similar association: A CAG repeat length exceeding the median in both alleles was associated with an increased risk for lifetime depression (OR=1.33, CI 1.00-1.76). In conclusion, we observed that carriers of either *ATXN7* or *TBP* alleles with relatively large CAG repeat sizes in both alleles had a substantially increased risk of lifetime depression. Our findings provide critical evidence for the notion that repeat polymorphisms can act as complex genetic modifiers of depression.

INTRODUCTION

Depression is one of the most common psychiatric disorders with an estimated lifetime prevalence of 14.6% in high-income and 11.1% in low-income countries.[1] Worldwide, depression is responsible for 74.5 million disability adjusted life years (DALYs)[2] and in 2010 the total economic burden of depression in the United States alone was estimated at $210.5 billion,[3] indicating that depression is a major strain on society. In order to create new and more effective therapeutic and preventive strategies, unravelling the pathogenesis of depression is imperative and elucidation of its genetic determinants is a critical step in this process.

The heritability of major depressive disorder (MDD) has been estimated to be between 30 and 50%.[4] Nevertheless, genome wide association studies (GWASs) have shown limited success in identifying its genetic basis. In a mega-analysis with more than 18 thousand subjects from European ancestry, no single-nucleotide polymorphism (SNP) reached genome-wide significance.[4] However recently, a meta-analysis of results from three studies, together containing 180 866 individuals, found two lead SNPs associated with depression.[5] In addition, another study combining three cohorts in a joint analysis with a total of 478 240 subjects found 17 independent SNPs significantly associated with a diagnosis of MDD,[6] suggesting that GWASs can successfully identify genetic associations with highly polygenic phenotypes. However, the identified effect sizes were small. Moreover, to gain sufficient statistical power the sample sizes required were extremely large and as a consequence unavoidably heterogeneous. Therefore, although GWASs have contributed greatly to genetic mapping of complex human traits, their success has been limited by the fact that aside from SNPs, these studies are unsuitable to assess the contribution of other genetic polymorphisms, especially DNA repeat sequences.[7]

Tandem repeats constitute about 3% of the human genome, a higher percentage than the entirety of the protein coding sequences,[8] substantially contributing to genetic variation.[9,10] Many rare hereditary disorders are caused by expansions of simple DNA repeat sequences.[11] However, the association of DNA repeat sequences with more common diseases is largely unknown. Polyglutamine diseases are the most prevalent disorders caused by an expanded DNA repeat sequence.[11,12] These diseases are caused by an expansion of a trinucleotide repeat (cytosine-adenine-guanine (CAG)) in the translated region of various genes. As CAG encodes the amino acid glutamine, the expansion in the trinucleotide repeat sequence results in an elongated polyglutamine domain in the associated proteins.[12] The most common polyglutamine disease is Huntington disease (HD), a severe neurodegenerative disorder characterized by both motor and neuropsychiatric impairment, and is caused by a CAG repeat expansion in exon 1 of the huntingtin (*HTT*) gene.[13,14] Recently, we demonstrated that both relatively short and relatively large CAG repeats in the longer *HTT* allele are associated with lifetime depression, suggesting that repeat polymorphisms could also act as complex genetic modifiers of depression and account for part of its ‘missing heritability’.[15]

Apart from HD, eight other polyglutamine disorders, all of which are neurodegenerative disorders are frequently associated with considerable neuropsychiatric impairment (**Table 1**).[16-35] These rare disorders result from relatively large repeat expansions in polyglutamine disease-associated genes (PDAGs). However, to what extent more common repeat length variations in the normal range of these genes could act as genetic modifiers of depression in the general population is still unknown. Hence, here we aimed to assess the contribution of CAG repeat length variations in these other PDAGs to depression susceptibility using data from two well-defined Dutch cohorts: The Netherlands Study of Depression and Anxiety (NESDA) and the Netherlands Study of Depression in Old Persons (NESDO).

MATERIAL AND METHODS

We genotyped eight PDAGs (*ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATXN7*, *TBP*, *ATN1* and *AR*) in all participants with sufficient amounts of DNA available from blood samples of two well-characterized Dutch cohorts: the NESDA (cohort 1) and the NESDO (cohort 2) cohorts (**Table 1** and **Supplementary Figures 1-17**).

Cohort 1

The NESDA is a cohort study among 2981 participants aged 18-65 years.[36] The participants were recruited from the general population, general practices and mental health care institutes. The sample included 1973 subjects with a lifetime diagnosis of depression and/or dysthymia (including 1925 patients with MDD), 635 subjects with a lifetime anxiety disorder without lifetime depression and 373 healthy controls.[36] Diagnoses were made in accordance with the Diagnostic and Statistical Manuel of Mental Disorders Fourth Edition (DSM-IV) criteria using the WHO Composite International Diagnostic Interview[37].

Cohort 2

The NESDO is a cohort study among 510 participants aged 60-93 years.[38] The participants were recruited from both general practices and mental health care institutes. The sample included 378 depressed individuals (360 with MDD) and 132 healthy controls. The same methods for diagnosing depression were used as in the NESDA.[38]

Genotyping

A polymerase chain reaction (PCR) was performed in a TProfessional thermocycler (Biometra, Westburg) with labelled primers flanking the CAG stretch of the PDAGs (Biolegio) (**Supplementary Table 1**). The PCR was performed using 10 ng of genomic DNA, 1x OneTaq mastermix (New England Biolabs, OneTaq Hot start with GC Buffer

Table 1. Summary genotyped polyglutamine disease associated genes

			CAG repeat ranges								
Gene	Disease	Protein	normal	pathological	Allele	Mean	Median	N	Range	# intermediate¹	# pathological¹
ATXN1	SCA1	Ataxin-1	6-39	41-83	short	29.22	29.00	3228	18-35	0	0
					long	30.76	30.00	3228	26-39	0	0
ATXN2	SCA2	Ataxin-2	14-32	33-500	short	21.91	22.00	3236	14-24	0	0
					long	22.35	22.00	3236	22-36	0	2 (33, 36)
ATXN3	SCA3	Ataxin-3	12-44	52-87	short	18.96	20.00	3235	13-33	0	0
					long	24.27	23.00	3235	13-49	1 (49)	0
CACNA1A	SCA6	CACNA1A	4-18	20-33	short	10.52	11.00	3220	4-14	0	0
					long	12.45	13.00	3220	4-22	0	1 (22)
ATXN7	SCA7	Ataxin-7	3-19	37-460	short	10.04	10.00	2979	7-15	0	0
					long	10.85	10.00	2979	7-19	0	0
TBP	SCA17	TBP	25-43	45-66	short	36.31	37.00	3238	27-39	0	0
					long	37.89	38.00	3238	32-48	1 (44)	1 (48)
ATN1	DRPLA	Atrophin-1	3-38	48-93	short	12.36	14.00	3228	5-22	0	0
					long	15.47	15.00	3228	8-30	0	0
AR	SBMA	Androgen receptor	6-36	38-72	short	20.80	21.00	3228	8-36	0	0
					long	22.90	23.00	3228	8-36	0	0

CACNA1A=calcium channel, voltage-dependent P/Q type, α 1A subunit; TBP=thymine-adenine-thymine-adenine (TATA) box binding protein; SCA=spinocerebellar ataxia; DRPLA=Dentatorubropallidoluysian atrophy; SBMA=spinal bulbar muscular atrophy. ¹) Numbers in the parentheses refer to the actual CAG repeat sizes. Note that all subjects with an intermediate or pathological CAG repeat size were in the group with lifetime depression.

master mix), 1 µl of primer Mix A or B (**Supplementary Table 1**) and Aqua B. Braun water to a final volume of 10 µl. The PCR was run with 27 cycles of 30 seconds, denaturation at 94°C, 1 minute of annealing at 60°C and 2 minutes elongation at 68°C, preceded by 5 minutes of initial denaturation at 94°C. Final elongation was performed at 69°C for 5 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 done with cases and controls randomized on plates and blinding with respect to disease status information.

Statistical Analysis

Binary logistic regression was used to assess whether CAG repeat sizes in the two alleles of each PDAG were associated with the risk of lifetime depression. For each PDAG, in an initial model, the presence of lifetime depression (i.e. MDD and/or dysthymia) was set as the dependent variable and the CAG repeat lengths of both alleles were used as the independent variables. To assess interaction effects between the two alleles or non-linear effects, a product term of the two alleles and a quadratic term for each allele were added to the model. In case the product term or the quadratic terms were not significant, these predictors were removed from the model and the analysis was repeated. Subsequently, we adjusted the results for the effects of sex, age and education level (coded as 'basic', 'intermediate' and 'high'[36,38]) in order to assess whether the effect of CAG repeat size variations were independent of these well-established risk factors for depression. The Nagelkerke R^2 was used to assess the proportion of variability explained by the predictors in the model. To account for potential effects of heteroscedasticity and influential points, all statistical significance tests were based on robust estimators of standard errors. Moreover, to visualize our results as well as to assure that the results were not unduly affected by violated model assumptions, we also applied a non-parametric method: we calculated the odds ratios (ORs) for a lifetime diagnosis of depression per CAG repeat group. The CAG repeat groups were defined based on the median CAG repeat size of each allele. A division as such resulted in three adequately sized groups: I= both alleles ≤ their median, II=the relatively shorter allele ≤ the median and the relatively longer allele > the median, III=both alleles > their median). Subsequently, the ORs were compared with the Fisher's exact test. In order to account for multiple testing, we applied a false discovery rate (FDR) correction as described by Benjamini and Hochberg, assuming eight independent tests with a two-sided α of .05.[39] All data are displayed as means and 95% confidence intervals (CIs) unless otherwise specified. All analyses were performed in SPSS version 23.0 (IBM SPSS Statistics for Windows, IBM Corp).

RESULTS

Risk of lifetime depression increases with higher *ATXN7* CAG repeat sizes

In total, we were able to genotype a total of 2979 participants for *ATXN7* (**Table 1**). This included 1998 depressed subjects and 981 non-depressed subjects. For 512 individuals, too little DNA material was available to determine the CAG repeat length in *ATXN7* (**Supplementary Table 2**). These lacking samples were missing completely at random. The CAG repeat lengths ranged from 7-19 repeats (**Table 1**).

We found a significant association between the risk of lifetime depression and the CAG repeat length of *ATXN7* in the shorter allele (β=0.202, p=.006). Adjusting for the effects of gender, age and education, hardly changed these results (β=0.184, p=.013) (**Supplementary Table 3**). After adjusting for multiple testing, this association remained significant. Inclusion of the CAG repeat length of the *ATXN7* allele in the model increased the proportion of explained genetic variation from the baseline model including only gender, age and level of education by 0.004 (i.e. the R^2 increased from 0.024 to 0.028). Considering our control group included subjects with anxiety disorders without co-morbid depression (n=348), we also performed a sensitivity analysis by excluding these individuals from the control group. Neither the parameter estimates nor their significance were materially altered by this procedure (β=0.215, p=.009).

For the non-parametric method, we divided all subjects based on median CAG repeat size for each *ATXN7* allele (median relatively short allele=10, median relatively long allele=10) and compared the ORs of lifetime depression among the groups (**Table 2**). The group containing the largest number of individuals was used as the reference category. Comparing the other groups to the reference category demonstrated the significant effect of the shorter allele on the risk of lifetime depression (**Figure 1**). When the relatively short allele and thus also the relatively long allele both contained a CAG repeat length larger than 10 repeats, the odds for lifetime depression almost doubled compared to the reference category in which both *ATXN7* alleles had a CAG repeat number equal to or smaller than the median.

TBP CAG repeat sizes in both alleles interact to affect the risk of lifetime depression

A total of 3238 individuals were genotyped for *TBP* (**Table 1**). This included 2180 people diagnosed with lifetime depression and 1058 non-depressed people. For 253 individuals, too little DNA material was available to determine the CAG repeat length in *TBP* (**Supplementary Table 2**). The lacking samples were missing completely at random. For *TBP*, we found CAG repeat lengths ranging between 27-48 repeats (**Table 1**).

Table 2. The distribution of participants in three *ATXN7* categories

ATXN7 category	Allele	Mean	Range	Lifetime depression[1]	No lifetime depression[2]	Odds ratio (95% CI)
I[3]	short long	9.94 10.00	7-10 7-10	1220	619	1.00
II	short long	9.97 12.18	7-10 11-19	662	331	1.01 (0.86-1.20)
III	short long	11.90 12.54	11-15 11-18	116	31	1.90 (1.26-2.85)[**]

I=both alleles of *ATXN7* contain a CAG repeat number ≤10; II=the relatively longer *ATXN7* allele contains a CAG repeat number >10 and the relatively shorter allele contains a CAG repeat number ≤10; III=both alleles of *ATXN7* contain a CAG repeat number >10. CI=confidence interval. [1]) Figures represent the number of subjects per *ATXN7* category in the group with lifetime depression. [2]) Figures represent the number of subjects per *ATXN7* category in the group with no lifetime depression. [3]) This category contains the largest number of subjects and was therefore chosen as the reference category for odds ratio calculation. [**]) p<.01 by Fisher's exact test in comparison to the reference category.

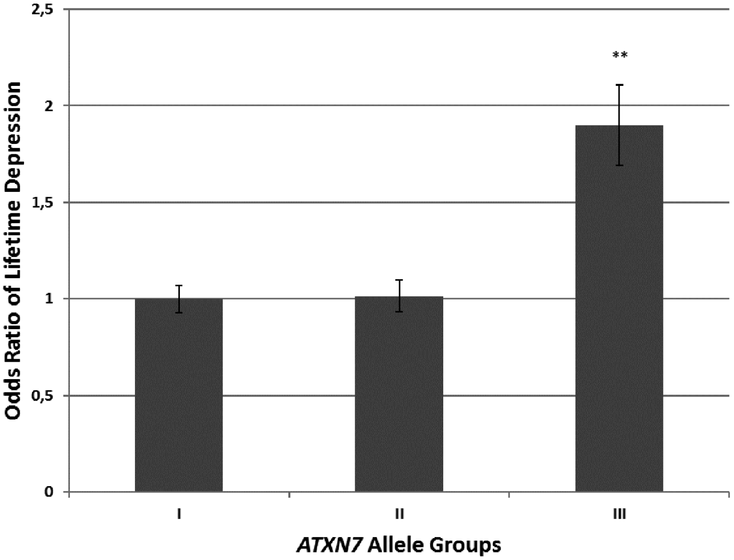


Figure 1. Odds ratio for lifetime depression per *ATXN7* category. The odds ratio for lifetime depression increases significantly when both the relatively long *ATXN7* allele and the relatively short *ATXN7* allele exceed the median CAG repeat number of 10. The odds of having lifetime depression almost doubles. The group having both alleles with a CAG repeat number ≤ 10 was the largest and therefore set as reference category. I=both alleles of *ATXN7* contain a CAG repeat number ≤10; II=the relatively longer *ATXN7* allele contains a CAG repeat number >10 and the relatively shorter allele contains a CAG repeat number ≤10; III=both alleles of *ATXN7* contain a CAG repeat number >10. Error bars indicate ± standard error. ** p<.01 by the Fisher's exact test in comparison to the reference category.

We found the interaction between the number of CAG repeats in the two *TBP* alleles to be significantly associated with the risk of lifetime depression (*TBP* short allele: β=-2.270, p=.004; *TBP* long allele: β=-2.112, p=.006; *TBP* interaction term: β=0.060, p=.005). When adjusting for the effects of gender, age and education the results hardly changed (*TBP* short allele: β=-2.302, p= .004; *TBP* long allele: β-2.150, p=.005; *TBP* interaction term: β=0.061, p=.005) (**Supplementary Table 3**). Furthermore, the results remained significant after applying the FDR correction.[39] The proportion of explained genetic variation increased by 0.005 from the baseline model including only gender, age and level of education (i.e. the R^2 increased from 0.023 to 0.028).

For the non-parametric method, we divided all subjects in three groups based on median CAG repeat size for each *TBP* allele (median relatively short *TBP* allele=37, median relatively long *TBP* allele=38) and compared the ORs of lifetime depression among the groups (**Table 3**). The largest group was set as the reference category. The comparison of the groups demonstrated the significant effect of the interaction between the relatively short *TBP* allele and the relatively long *TBP* allele on the risk of depression. The risk of lifetime depression was significantly higher when both *TBP* alleles had a CAG repeat length exceeding their median compared to the reference category in which both alleles had a CAG repeat number equal to or lower than their median. (**Figure 2**).

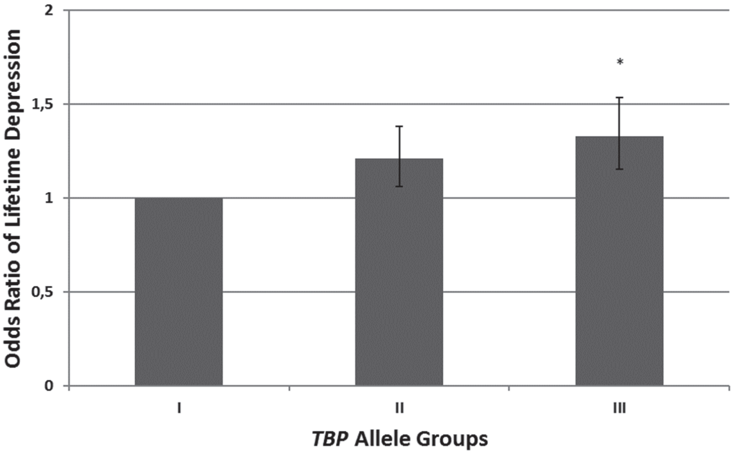


Figure 2. Odds ratio for lifetime depression per *TBP* category. The odds ratio for lifetime depression increases significantly when both the relatively short *TBP* allele and the relatively long *TBP* allele exceed their median CAG repeat number (median short allele=37, median long allele=38). The group with both alleles being equal to or smaller than their medians was the largest and therefore defined as the reference category. I=both *TBP* alleles contain a CAG repeat number ≤ their median; II=the relatively longer *TBP* allele contains a CAG repeat number > the median and the relatively shorter *TBP* allele contains a CAG repeat number ≤ the median; III=both *TBP* alleles contain a CAG repeat number > their median. Error bars indicate ± one standard error. * p<.05 by the Fisher's exact test in comparison to the reference category.

Table 3. The distribution of participants in three *TBP* categories

TBP category	Allele	Mean	Range	Lifetime depression[1]	No lifetime depression[2]	Odds ratio (95% CI)
I[3]	short long	35.78 37.47	27-37 32-38	1462	716	1.00
II	short long	36.04 39.30	30-37 39-48	230	93	1.21 (0.94-1.57)
III	short long	38.09 39.28	38-39 39-44	201	74	1.33 (1.00-1.76)*

I=both *TBP* alleles contain a CAG repeat number ≤ their median; II=the relatively longer *TBP* allele contains a CAG repeat number > the median and the relatively shorter *TBP* allele contains a CAG repeat number ≤ the median; III=both *TBP* alleles contain a CAG repeat number > their median (median short allele=37, median long allele=38). [1]) Figures represent the number of subjects per *TBP* category in the group with lifetime depression. [2]) Figures represent the number of subjects per *TBP* category in the group with no lifetime depression. [3]) This category contains the largest number of subjects and was therefore chosen as the reference category for odds ratio calculation. [*]) p<.05 by Fisher's exact test in comparison to the reference category.

CAG repeat sizes in other PDAGs were not associated with risk of lifetime depression

The number of CAG repeats in *ATXN1*, *ATXN2*, *ATXN3*, *CACNA1A*, *ATN1* and *AR* were not associated with the risk of lifetime depression. Neither the main effect of the alleles, nor the interactions between the two alleles, or the quadratic terms were significantly associated with the presence of lifetime depression in the combined cohort. Since *AR* is located on the X-chromosome, we performed one analysis stratifying the data by gender and another analysis using only either the relatively short or the relatively long allele in the model. None of the two approaches demonstrated a significant association between CAG repeat number in *AR* and depression (**Supplementary Table 3**).

The effects of *HTT, ATXN7* and *TBP* CAG repeat sizes are independent

In previous research, we demonstrated that both the main term of the longer *HTT* allele and the quadratic term of the longer *HTT* allele are also significantly associated with a diagnosis of lifetime depression.[15] To assess the contribution of the CAG repeat polymorphisms in all PDAGs to depression risk, we examined the degree to which *HTT*, *ATXN7* and *TBP* CAG repeat sizes explain depression heritability. To this end, we applied a multivariate model with the CAG repeat sizes in these three PDAGs as predictors. In addition, we adjusted for the effects of gender, age and education. We found that the parameter estimates as well as their associated statistical significances hardly changed, indicating that the effects of *HTT, ATXN7* and *TBP* CAG repeat sizes on depression susceptibility are mutually independent (*HTT* long allele: β=-0.286, p=.018; *HTT* long allele quadratic term β=0.006, p=.017; *ATXN7* short allele: β=0.210, p=.005;

TBP short allele: β=-2.728, p=.001; *TBP* long allele, β=-2.566; p=.001; *TBP* interaction term: β=0.072, p=.001). Inclusion of the CAG repeat sizes in *HTT, ATXN7 and TBP* in the model increased R^2 by 0.014 from the baseline model including only gender, age and level of education (from 0.024 to 0.038), indicating that the CAG repeat lengths in these PDAGs can account for an additional 1.4% of the genetic variation on the observed probability scale. We also derived the R^2 on the liability scale as described previously.[40] Assuming that depression has a lifetime prevalence of about 15% in the Netherlands[41] and adjusting for the oversampling of patients with depression in our cohort, the R^2 on the liability scale was 0.0191, indicating that CAG repeat size polymorphisms in *HTT, ATXN7* and *TBP* together can account for about 1.9% of depression heritability.

The prevalence of intermediate and pathological PGDA alleles

In total, four genotyped subjects had a CAG repeat number in the pathological range of a PDAG. One individual had a CAG repeat number of 22 in *CACNA1A*, one subject had a CAG repeat number of 48 in *TBP* and two people had a CAG repeat number of 33 and 36 in *ATXN2* (**Table 1**). All four were depressed, but at the ages of 53 years for CACNA1A, 59 years for *TBP*, and 26 and 31 years for *ATXN2*, none had been diagnosed with the respective diseases (i.e. SCA6, SCA17 and SCA2).

One person had a CAG repeat number in the longer *TBP* allele of 44, belonging to the intermediate range. This individual was also diagnosed with lifetime depression. Furthermore, one depressed individual had an intermediate CAG repeat number in the longer allele of *ATXN3* (i.e. 49 repeats). For *ATXN1, ATXN7, ATN1* and *AR* all subjects had a CAG repeat number within the normal range (**Table 1**).

When we included the two depressed individuals with a CAG repeat length in the reduced penetrance range of *HTT* found previously[15], we observed a trend for a higher proportion of subjects with a CAG repeat in the intermediate or pathological range of a PDAG in the depressed group (p=.059).

DISCUSSION

To our knowledge this is the first study assessing the influence of CAG repeat size variations in the normal range of PDAGs – other than *HTT* – on the risk of lifetime depression. Interestingly, we found an association between the presence of lifetime depression and the CAG repeat length in two PDAGs, i.e. *ATXN7* and *TBP*. The main relationship we found was that when the CAG repeat number of both alleles in either *ATXN7* or *TBP* was relatively large, the odds for a diagnosis of lifetime depression markedly increased. Moreover, six genotyped individuals had a CAG repeat number in at least one of their PDAGs that extended into the intermediate or pathological range (2 intermediate, 4 pathological). All of these subjects were diagnosed with lifetime depression. These findings are in support of the hypothesis that repeat polymorphisms may act as complex genetic modifiers of depression and thus could account for part of its 'missing heritability'.

In our study, we found that the odds of lifetime depression almost doubled in individuals with a CAG repeat length of >10 in both *ATXN7* alleles. Similarly, we found that the risk of lifetime depression markedly increased (OR = 1.33) in individuals that had a CAG repeat length exceeding the median in both *TBP* alleles. These effect sizes are considerable, especially when compared to the effect sizes of the two most significant SNPs found in a recent GWAS: OR=0.955 (95% CI 0.943-0.968) and OR=1.051 (95% CI 1.036-1.067).[6] Although we cannot fully exclude potential modifying effects of SNPs in linkage disequilibrium (LD) with *ATXN7* and *TBP*, the fact that such SNPs in these PDAGs have not been detected before in GWASs suggests that the influence of other genetic variants in LD with the CAG repeat size in these genes is likely to be minimal.[42] Furthermore, we found that CAG repeat size polymorphisms in *HTT, ATXN7* and *TBP* together can account for about 1.9% of depression heritability. Using the same population prevalence of 15% for depression, the meta-analysis in the recent GWAS calculated a heritability of 5.9%.[6] The amount of variance explained by these SNPs is larger than the amount explained by the CAG repeat polymorphisms we found. However, the meta-analysis in this GWAS examined SNPs in the entire genome of 326 113 individuals[6] whereas our estimation was based on the effects of CAG repeat variations in only nine PDAGs. This fact suggests that investigating additional repeat polymorphisms in the human genome could lead to the identification of many more novel genetic determinants of depression. The importance of investigating repeat polymorphisms in association with health and disease has been described in previous literature and begins to gain more recognition within the field of genetic research.[9,10,43-46]

CAG repeat numbers exceeding 36 in *ATXN7* are responsible for the severe neurodegenerative disorder spinocerebellar ataxia type 7 (SCA7).[34] SCA7 is a progressive autosomal dominant neurodegenerative disorder primarily characterized by cerebellar ataxia and macular degeneration.[47] Although depressive symptoms are a frequent finding in many neurodegenerative disorders, the prevalence of depression has hardly been assessed in SCA7.[48-51] Our findings suggest that depression might be an underappreciated feature of SCA7 which needs further characterization. A CAG repeat number in *TBP* larger than 48, causes the severe progressive neurodegenerative disorder SCA17[34]. Aside from cerebellar ataxia, dementia and pyramidal symptoms, in 67% of the cases SCA17 is also accompanied by psychiatric signs and symptoms, including depression, behavioural changes as well as psychosis.[52] Initial symptoms of SCA17 have been described to include depression.[53]

The gene *ATXN7* encodes the protein ataxin-7. Apart from several important cellular functions[51], ataxin-7 is an integral part of the TATA-binding protein-free TAF-containing

(TFTC) complex.[54] TFTC allows for the initiation of transcription via RNA polymerase II in the absence of the RNA polymerase II transcription factor D (TFIID)[55]. Interestingly, the TATA-box binding protein (TBP) that is encoded by *TBP,* is the DNA-binding subunit of TFIID. TBP anchors TFIID to the TATA-box upstream of the first codon, allowing for the initiation of transcription via RNA polymerase II.[56] Thus, TBP as well as ataxin-7 seem to play a part in the initiation of transcription via RNA polymerase II. RNA polymerase II catalyses the transcription of DNA to synthesize precursors of mRNA, snRNA and microRNA.[57,58] Numerous studies indicate an altered mRNA expression of proteins such as the serotonin 5-HT$_{1A}$ receptor,[59,60] brain –derived neurotropic factor (BDNF)[61-64] and corticotrophin-releasing factor (CRF)[65,66] in subjects suffering from MDD compared to controls. Furthermore, the evidence supporting a role for the dysfunction of microRNA mediated regulated gene expression in MDD is increasing.[67-71] Therefore, CAG repeat variations in either *ATXN7* or *TBP* could modulate the function of RNA polymerase II and thereby lead to changes in mRNA and microRNA expression that have previously been associated with MDD.[59-71] Hence, we intend to investigate the effect of CAG repeat length variations within the normal range in *ATXN7* and *TBP* on the expression and function of proteins thought to be associated with depression in a cellular model.

The presence of one subject with a CAG repeat size in the pathological range of *CACNA1A*, one subject with a CAG repeat size in the pathological range of *TBP* and two subjects with a pathological CAG repeat size in *ATXN2* within our two cohorts is remarkable. For these individuals, depression is apparently the first symptom with which SCA6, SCA17 or SCA2 manifests itself. The prevalence of an autosomal dominant cerebral ataxia among Europeans is estimated to be between 1-3 per 100 000 of which 2% are diagnosed with SCA6, < 1% with SCA17 and 10% with SCA2[72], resulting in a prevalence of 0.02-0.06 per 100 000, < 0.01-0.03 per 100 000 and 0.10-0.30 per 100 000 for SCA6, SCA17 and SCA2, respectively. These numbers are in stark contrast to the substantially higher prevalence estimates in patients with depression in our study, 90 per 100 000 for SCA6, 50 per 100 000 for SCA17 and 50 per 100 000 for SCA2. Previously we also found two individuals with incompletely penetrant *HTT* alleles (i.e. those containing 36-39 CAG repeats) in this same population, both of whom suffered from depression.[15] The difference in proportion of individuals with intermediate or pathological CAG repeat lengths in PDAGs indeed tended to be higher in depressed compared to control subjects, although likely due to the relative rarity of these expanded alleles the results did not reach statistical significance. Our findings are in line with another study which estimated the prevalence of incompletely penetrant *HTT* alleles in MDD patients to be about 3 in 1000, whereas such alleles were absent in the control group.[73] Together these findings suggest that depression could be the first manifestation of polyglutamine diseases and that, conversely, polyglutamine diseases might be underdiagnosed in patients with depression.

In this study, we analysed a homogenous population by using samples from two Dutch cohorts for our analysis, thereby minimizing the impact of population stratification.

However, the use of a uniform group of individuals could also be seen as a limitation of our study, as it might have consequences for the generalizability of our findings. However, our results appear robust and consistent, although confirmation of our findings in other populations, as well as the elucidation of their pathophysiological basis is warranted.

In conclusion, we observed a significant association between the risk of lifetime depression and CAG repeat size in *ATXN7* and *TBP*. A relatively large CAG repeat number in both alleles of either *ATXN7* or *TBP* substantially increased depression risk. Our findings add more critical evidence to the notion that repeat polymorphisms could act as complex genetic modifiers of depression and, therefore, could partially account for its 'missing heritability'. In addition, our findings indicate that the role of DNA repeat polymorphisms as potential genetic modifiers of other psychiatric disorders also needs further scrutiny.

ACKNOWLEDGEMENTS

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 infrastructure for the NESDA study (www.nesda.nl) has been funded through the Geestkracht program of the Netherlands Organisation for Health Research and Development (Zon-Mw, grant number 10-000-1002) and participating universities (VU University Medical Center, Leiden University Medical Centre, University Medical Centre Groningen). The infrastructure for the NESDO study (http://nesdo.amstad.nl) is funded through the Fonds NutsOhra (project 0701-065), Stichting tot Steun VCVGZ, NARSAD The Brain and Behaviour Research Fund (grant ID 41080), and the participating universities and mental health care organizations (VU University Medical Centre, Leiden University Medical Center, University Medical Centre Groningen, UMC St Radboud, and GGZ inGeest, GG Net, GGZ Nijmegen, and Parnassia).

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Bromet E, Andrade LH, Hwang I, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC medicine* 2011; **9**: 90.
2. Whiteford HA, Degenhardt L, Rehm J, et al. Global burden of disease attributable to mental and substance use disorders: findings from the Global Burden of Disease Study 2010. *The Lancet* 2013; **382**(9904): 1575-86.
3. Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *The Journal of clinical psychiatry* 2015; **76**(2): 155-62.
4. Major Depressive Disorder Working Group of the Psychiatric GC, Ripke S, Wray NR, et al. A mega-analysis of genome-wide association studies for major depressive disorder. *Molecular psychiatry* 2013; **18**(4): 497-511.
5. Okbay A, Baselmans BM, De Neve JE, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature genetics* 2016; **48**(6): 624-33.
6. Hyde CL, Nagle MW, Tian C, et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nature genetics* 2016.
7. Manolio TA. Genomewide association studies and assessment of the risk of disease. *The New England journal of medicine* 2010; **363**(2): 166-76.
8. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature* 2001; **409**(6822): 860-921.
9. Hannan AJ. Tandem repeat polymorphisms: modulators of disease susceptibility and candidates for ‘missing heritability’. *Trends in genetics : TIG* 2010; **26**(2): 59-65.
10. 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.
11. Mirkin SM. Expandable DNA repeats and human disease. *Nature* 2007; **447**(7147): 932-40.
12. Shao J, Diamond MI. Polyglutamine diseases: emerging concepts in pathogenesis and therapy. *Human molecular genetics* 2007; **16 Spec No. 2**: R115-23.
13. Bates G, Harper P, Jones L. Huntington’s Disease. Third edition ed. New York: Oxford University Press; 2002.
14. THsDCR G. A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington’s disease chromosomes. The Huntington’s Disease Collaborative Research Group. *Cell* 1993; **72**(6): 971-83.
15. Gardiner SL, van Belzen MJ, Boogaard MW, et al. Huntingtin gene repeat size variations affect risk of lifetime depression *(submitted for publication)*.
16. Orr HT, Chung MY, Banfi S, et al. Expansion of an unstable trinucleotide CAG repeat in spinocerebellar ataxia type 1. *Nature genetics* 1993; **4**(3): 221-6.
17. Chung MY, Ranum LP, Duvick LA, Servadio A, Zoghbi HY, Orr HT. Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. *Nature genetics* 1993; **5**(3): 254-8.
18. Matilla T, Volpini V, Genis D, et al. Presymptomatic analysis of spinocerebellar ataxia type 1 (SCA1) via the expansion of the SCA1 CAG-repeat in a large pedigree displaying anticipation and parental male bias. *Human molecular genetics* 1993; **2**(12): 2123-8.
19. Pulst SM, Nechiporuk A, Nechiporuk T, et al. Moderate expansion of a normally biallelic trinucleotide repeat in spinocerebellar ataxia type 2. *Nature genetics* 1996; **14**(3): 269-76.
20. Sanpei K, Takano H, Igarashi S, et al. Identification of the spinocerebellar ataxia type 2 gene using a direct identification of

repeat expansion and cloning technique, DIRECT. *Nature genetics* 1996; **14**(3): 277-84.

21. Giunti P, Sweeney MG, Harding AE. Detection of the Machado-Joseph disease/spinocerebellar ataxia three trinucleotide repeat expansion in families with autosomal dominant motor disorders, including the Drew family of Walworth. *Brain : a journal of neurology* 1995; **118 (Pt 5)**: 1077-85.
22. Maciel P, Gaspar C, DeStefano AL, et al. Correlation between CAG repeat length and clinical features in Machado-Joseph disease. *American journal of human genetics* 1995; **57**(1): 54-61.
23. Maruyama H, Nakamura S, Matsuyama Z, et al. Molecular features of the CAG repeats and clinical manifestation of Machado-Joseph disease. *Human molecular genetics* 1995; **4**(5): 807-12.
24. Zhuchenko O, Bailey J, Bonnen P, et al. Autosomal dominant cerebellar ataxia (SCA6) associated with small polyglutamine expansions in the alpha 1A-voltage-dependent calcium channel. *Nature genetics* 1997; **15**(1): 62-9.
25. Matsuyama Z, Kawakami H, Maruyama H, et al. Molecular features of the CAG repeats of spinocerebellar ataxia 6 (SCA6). *Human molecular genetics* 1997; **6**(8): 1283-7.
26. David G, Abbas N, Stevanin G, et al. Cloning of the SCA7 gene reveals a highly unstable CAG repeat expansion. *Nature genetics* 1997; **17**(1): 65-70.
27. Benton CS, de Silva R, Rutledge SL, Bohlega S, Ashizawa T, Zoghbi HY. Molecular and clinical studies in SCA-7 define a broad clinical spectrum and the infantile phenotype. *Neurology* 1998; **51**(4): 1081-6.
28. Del-Favero J, Krols L, Michalik A, et al. Molecular genetic analysis of autosomal dominant cerebellar ataxia with retinal degeneration (ADCA type II) caused by CAG triplet repeat expansion. *Human molecular genetics* 1998; **7**(2): 177-86.
29. Johansson J, Forsgren L, Sandgren O, Brice A, Holmgren G, Holmberg M. Expanded CAG repeats in Swedish spinocerebellar ataxia type 7 (SCA7) patients: effect of CAG repeat length on the clinical manifestation. *Human molecular genetics* 1998; **7**(2): 171-6.
30. Nakamura K, Jeong SY, Uchihara T, et al. SCA17, a novel autosomal dominant cerebellar ataxia caused by an expanded polyglutamine in TATA-binding protein. *Human molecular genetics* 2001; **10**(14): 1441-8.
31. Koide R, Kobayashi S, Shimohata T, et al. A neurological disease caused by an expanded CAG trinucleotide repeat in the TATA-binding protein gene: a new polyglutamine disease? *Human molecular genetics* 1999; **8**(11): 2047-53.
32. Nagafuchi S, Yanagisawa H, Sato K, et al. Dentatorubral and pallidoluysian atrophy expansion of an unstable CAG trinucleotide on chromosome 12p. *Nature genetics* 1994; **6**(1): 14-8.
33. La Spada AR, Wilson EM, Lubahn DB, Harding AE, Fischbeck KH. Androgen receptor gene mutations in X-linked spinal and bulbar muscular atrophy. *Nature* 1991; **352**(6330): 77-9.
34. 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.
35. Fan HC, Ho LI, Chi CS, et al. Polyglutamine (PolyQ) diseases: genetics to treatments. *Cell transplantation* 2014; **23**(4-5): 441-58.
36. Penninx BW, Beekman AT, Smit JH, et al. The Netherlands Study of Depression and Anxiety (NESDA): rationale, objectives and methods. *International journal of methods in psychiatric research* 2008; **17**(3): 121-40.
37. Wittchen HU, Robins LN, Cottler LB, Sartorius N, Burke JD, Regier D. Cross-cultural feasibility, reliability and sources of variance of the Composite International Diagnostic Interview (CIDI). The Multicentre WHO/ADAMHA Field Trials. *The British*

journal of psychiatry : the journal of mental science 1991; **159**: 645-53, 58.

38. Comijs HC, van Marwijk HW, van der Mast RC, et al. The Netherlands study of depression in older persons (NESDO); a prospective cohort study. *BMC research notes* 2011; **4**: 524.

39. 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.

40. Lee SH, Goddard ME, Wray NR, Visscher PM. A better coefficient of determination for genetic profile analysis. *Genetic epidemiology* 2012; **36**(3): 214-24.

41. Andrade L, Caraveo-Anduaga JJ, Berglund P, et al. The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *International journal of methods in psychiatric research* 2003; **12**(1): 3-21.

42. Okbay A, Baselmans BM, De Neve JE, Turley P, Nivard MG. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. 2016; **48**(6): 624-33.

43. Kashi Y, King D, Soller M. Simple sequence repeats as a source of quantitative genetic variation. *Trends in genetics : TIG* 1997; **13**(2): 74-8.

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

45. Sawaya S, Bagshaw A, Buschiazzo E, et al. Microsatellite tandem repeats are abundant in human promoters and are associated with regulatory elements. *PloS one* 2013; **8**(2): e54710.

46. Sawaya SM, Lennon D, Buschiazzo E, Gemmell N, Minin VN. Measuring microsatellite conservation in mammalian evolution with a phylogenetic birth-death model. *Genome biology and evolution* 2012; **4**(6): 636-47.

47. Lebre AS, Brice A. Spinocerebellar ataxia 7 (SCA7). *Cytogenet Genome Res* 2003; **100**(1-4): 154-63.

48. Pla P, Orvoen S, Saudou F, David DJ, Humbert S. Mood disorders in Huntington's disease: from behavior to cellular and molecular mechanisms. *Frontiers in behavioral neuroscience* 2014; **8**: 135.

49. Chaudhuri KR, Schapira AH. Non-motor symptoms of Parkinson's disease: dopaminergic pathophysiology and treatment. *The Lancet Neurology* 2009; **8**(5): 464-74.

50. Schmitz-Hubsch T, Coudert M, Tezenas du Montcel S, et al. Depression comorbidity in spinocerebellar ataxia. *Movement disorders : official journal of the Movement Disorder Society* 2011; **26**(5): 870-6.

51. Rub U, Schols L, Paulson H, et al. Clinical features, neurogenetics and neuropathology of the polyglutamine spinocerebellar ataxias type 1, 2, 3, 6 and 7. *Progress in neurobiology* 2013; **104**: 38-66.

52. Stevanin G, Brice A. Spinocerebellar ataxia 17 (SCA17) and Huntington's disease-like 4 (HDL4). *Cerebellum (London, England)* 2008; **7**(2): 170-8.

53. Bauer P, Laccone F, Rolfs A, et al. Trinucleotide repeat expansion in SCA17/TBP in white patients with Huntington's disease-like phenotype. *Journal of medical genetics* 2004; **41**(3): 230-2.

54. Helmlinger D, Hardy S, Sasorith S, et al. Ataxin-7 is a subunit of GCN5 histone acetyltransferase-containing complexes. *Human molecular genetics* 2004; **13**(12): 1257-65.

55. Wieczorek E, Brand M, Jacq X, Tora L. Function of TAF(II)-containing complex without TBP in transcription by RNA polymerase II. *Nature* 1998; **393**(6681): 187-91.

56. Lescure A, Lutz Y, Eberhard D, et al. The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA

polymerase II and III promoters. *The EMBO journal* 1994; **13**(5): 1166-75.

57. Kornberg RD. Eukaryotic transcriptional control. *Trends in cell biology* 1999; **9**(12): M46-9.

58. Sims RJ, 3rd, Mandal SS, Reinberg D. Recent highlights of RNA-polymerase-II-mediated transcription. *Current opinion in cell biology* 2004; **16**(3): 263-71.

59. Lopez-Figueroa AL, Norton CS, Lopez-Figueroa MO, et al. Serotonin 5-HT1A, 5-HT1B, and 5-HT2A receptor mRNA expression in subjects with major depression, bipolar disorder, and schizophrenia. *Biol Psychiatry* 2004; **55**(3): 225-33.

60. Lopez JF, Chalmers DT, Little KY, Watson SJ. A.E. Bennett Research Award. Regulation of serotonin1A, glucocorticoid, and mineralocorticoid receptor in rat and human hippocampus: implications for the neurobiology of depression. *Biol Psychiatry* 1998; **43**(8): 547-73.

61. Cattaneo A, Gennarelli M, Uher R, et al. Candidate genes expression profile associated with antidepressants response in the GENDEP study: differentiating between baseline 'predictors' and longitudinal 'targets'. *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 2013; **38**(3): 377-85.

62. D'Addario C, Dell'Osso B, Galimberti D, et al. Epigenetic modulation of BDNF gene in patients with major depressive disorder. *Biol Psychiatry* 2013; **73**(2): e6-7.

63. Pandey GN, Dwivedi Y, Rizavi HS, Ren X, Zhang H, Pavuluri MN. Brain-derived neurotrophic factor gene and protein expression in pediatric and adult depressed subjects. *Progress in neuro-psychopharmacology & biological psychiatry* 2010; **34**(4): 645-51.

64. Zhou L, Xiong J, Lim Y, et al. Upregulation of blood proBDNF and its receptors in major depression. *J Affect Disord* 2013; **150**(3): 776-84.

65. Austin MC, Janosky JE, Murphy HA. Increased corticotropin-releasing hormone immunoreactivity in monoamine-containing pontine nuclei of depressed suicide men. *Mol Psychiatry* 2003; **8**(3): 324-32.

66. Raadsheer FC, van Heerikhuize JJ, Lucassen PJ, Hoogendijk WJ, Tilders FJ, Swaab DF. Corticotropin-releasing hormone mRNA levels in the paraventricular nucleus of patients with Alzheimer's disease and depression. *The American journal of psychiatry* 1995; **152**(9): 1372-6.

67. Geaghan M, Cairns MJ. MicroRNA and Posttranscriptional Dysregulation in Psychiatry. *Biol Psychiatry* 2015; **78**(4): 231-9.

68. Hommers LG, Domschke K, Deckert J. Heterogeneity and individuality: microRNAs in mental disorders. *Journal of neural transmission (Vienna, Austria : 1996)* 2015; **122**(1): 79-97.

69. Hunsberger JG, Austin DR, Chen G, Manji HK. MicroRNAs in mental health: from biological underpinnings to potential therapies. *Neuromolecular medicine* 2009; **11**(3): 173-82.

70. Issler O, Chen A. Determining the role of microRNAs in psychiatric disorders. *Nature reviews Neuroscience* 2015; **16**(4): 201-12.

71. O'Connor RM, Grenham S, Dinan TG, Cryan JF. microRNAs as novel antidepressant targets: converging effects of ketamine and electroconvulsive shock therapy in the rat hippocampus. *The international journal of neuropsychopharmacology/officialscientific journal of the Collegium Internationale Neuropsychopharmacologicum (CINP)* 2013; **16**(8): 1885-92.

72. Durr A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *The Lancet Neurology* 2010; **9**(9): 885-94.

73. Perlis RH, Smoller JW, Mysore J, et al. Prevalence of incompletely penetrant Huntington's disease alleles among individuals with major depressive disorder. *The American journal of psychiatry* 2010; **167**(5): 574-9.

SUPPLEMENTARY MATERIAL

Supplementary Table 1. Polyglutamine disease associated genes and primers

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

Supplementary Table 2. Genotyped subjects per gene

Gene	Lifetime depression[1]	No lifetime depression[2]	Total
ATXN1	2170	1058	3228
ATXN2	2178	1058	3236
ATXN3	2180	1057	3237
CACNA1A	2165	1055	3220
ATXN7	1998	981	2979
TBP	2180	1058	3238
ATN1	2169	1059	3228
AR	2174	1054	3228

[1]) Figures represent the number of subjects per gene that were genotyped in the group with lifetime depression; [2]) Figures represent the number of subjects per gene that were genotyped in the group with no lifetime depression.

Supplementary Table 3. Analysis of genes via binary logistic regression

Gene	Step[1]	Variable	β-coefficient	SE	95% CI		p-value
ATXN1	1	ATXN1_s	-0.054	0.622	-1.273	1.165	.931
		ATXN1_l	0.122	0.635	-1.123	1.367	.848
		ATXN1_s2	0.001	0.010	-0.020	0.021	.939
		ATXN1_l2	-0.002	0.010	-0.022	0.018	.849
		ATXN1_sl	-0.000	0.021	-0.042	0.042	.996
	2	ATXN1_s	-0.011	0.037	-0.084	0.063	.291
		ATXN1_l	-0.004	0.023	-0.049	0.042	.777
ATXN2	1	ATXN2_s	0.394	2.600	-4.702	5.489	.880
		ATXN2_l	-2.224	2.607	-7.333	2.885	.393
		ATXN2_s2	-0.058	0.031	-0.120	0.003	.064
		ATXN2_l2	0.009	0.012	-0.014	0.032	.435
		ATXN2_sl	0.082	0.120	-0.153	0.317	.492
	2	ATXN2_s	-0.038	0.060	-0.156	0.079	.524
		ATXN2_l	0.038	0.034	-0.029	0.105	.270
ATXN3	1	ATXN3_s	-0.142	0.103	-0.344	0.059	.166
		ATXN3_l	-0.056	0.059	-0.172	0.059	.339
		ATXN3_s2	0.003	0.003	-0.003	0.009	.318
		ATXN3_l2	0.001	0.002	-0.002	0.004	.686
		ATXN3_sl	0.002	0.004	-0.006	0.009	.681
	2	ATXN3_s	0.009	0.009	-0.009	0.028	.320
		ATXN3_l	-0.001	0.011	-0.022	0.020	.952
CACNA1A	1	CACNA1A_s	-0.128	0.245	-0.608	0.352	.601
		CACNA1A_l	0.353	0.335	-0.304	1.010	.293
		CACNA1A_s2	-0.002	0.009	-0.019	0.015	.814

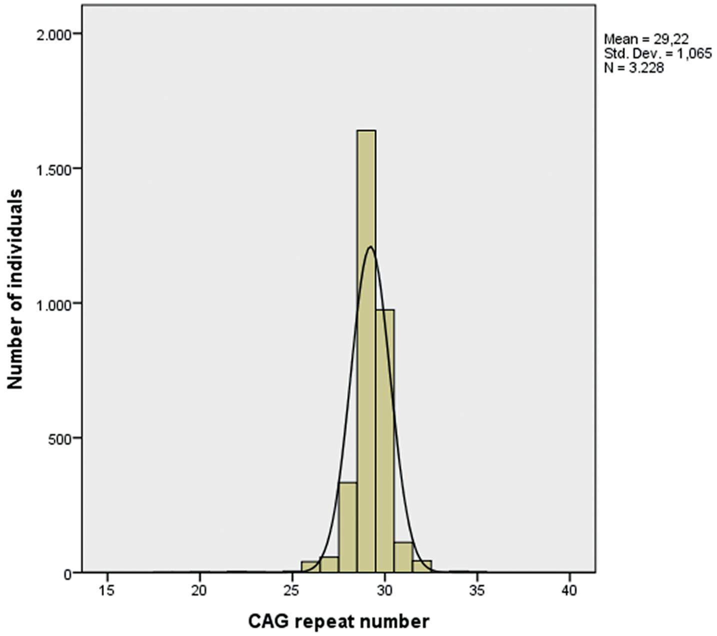
Supplementary Table 3. (continued)

Gene	Step[1]	Variable	β-coefficient	SE	95% CI		p-value
		CACNA1A_l2	-0.021	0.020	-0.060	0.018	.296
		CACNA1A_sl	0.014	0.022	-0.028	0.056	.520
	2	CACNA1A_s	0.008	0.019	-0.029	0.046	.660
		CACNA1A_l	-0.007	0.036	-0.079	0.064	.843
ATXN7	1	ATXN7_s	0.485	0.673	-0.833	1.804	.470
		ATXN7_l	1.209	0.941	-0.636	3.053	.199
		ATXN7_s2	0.089	0.046	-0.002	0.180	.054
		ATXN7_l2	0.024	0.024	-0.023	0.071	.311
		ATXN7_sl	-0.176	0.088	-0.349	-0.003	.046*
	2	ATXN7_s	0.669	0.588	-0.484	1.822	.256
		ATXN7_l	0.418	0.510	-0.582	1.418	.413
		ATXN7_sl	-0.040	0.051	-0.139	0.059	.426
	3	ATXN7_s	0.202	0.073	0.059	0.345	.006**
		ATXN7_l	0.012	0.034	-0.054	0.078	.713
	4[2]	ATXN7_s	0.184	0.074	0.039	0.328	.013*
		ATXN7_l	0.018	0.034	-0.049	0.085	.604
TBP	1	TBP_s	-1.873	0.832	-3.504	-0.242	.024*
		TBP_l	-2.878	1.293	-5.412	-0.344	.026*
		TBP_s2	-0.011	0.008	-0.027	0.006	.203
		TBP_l2	0.006	0.017	-0.027	0.038	.727
		TBP_sl	0.069	0.025	0.021	0.118	.005**
	2	TBP_s	-2.270	0.796	-3.831	-0.709	.004**
		TBP_l	-2.112	0.761	-3.604	-0.621	.006**
		TBP_sl	0.060	0.021	0.019	0.101	.005**
	3[2]	TBP_s	-2.302	0.796	-3.861	-0.743	.004**
		TBP_l	-2.150	0.761	-3.641	-0.658	.004**
		TBP_sl	0.061	0.021	0.019	0.102	.005**
ATN1	1	ATN1_s	-0.036	0.169	-0.368	0.296	.833
		ATN1_l	0.138	0.093	-0.045	0.321	.139
		ATN1_s2	0.006	0.007	-0.008	0.020	.411
		ATN1_l2	-0.002	0.004	-0.010	0.006	.610
		ATN1_sl	-0.007	0.008	-0.023	0.009	.402
	2	ATN1_s	-0.003	0.283	-0.029	0.023	.820
		ATN1_l	-0.003	0.013	-0.041	0.035	.866
AR ♂	1	AR_s	0.047	0.188	-0.322	0.416	.804
		AR_s2	-0.001	0.004	-0.009	0.007	.796
	2	AR_s	-0.001	0.022	-0.044	0.041	.945
AR ♀	1	AR_s	0.179	0.214	-0.241	0.599	.404
		AR_l	-0.143	0.256	-0.644	0.358	.576

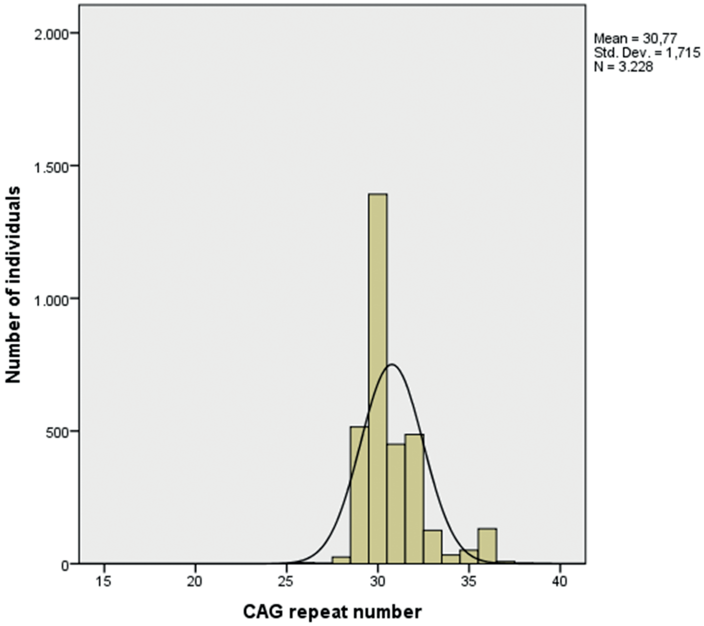
Supplementary Table 3. (continued)

Gene	Step[1]	Variable	β-coefficient	SE	95% CI		p-value
		AR_s2	-0.003	0.006	-0.014	0.007	.545
		AR_l2	0.004	0.006	-0.008	0.015	.539
		AR_sl	-0.001	0.010	-0.020	0.018	.908
	2	AR_s	0.017	0.023	-0.028	0.062	.460
		AR_l	0.007	0.020	-0.032	0.046	.735
AR_s	1	AR_s	0.142	0.119	-0.090	0.374	.231
		AR_s2	-0.003	0.003	-0.009	0.002	.222
	2	AR_s	-0.002	0.014	-0.030	0.026	.889
AR_l	1	AR_l	0.030	0.139	-0.242	0.302	.827
		AR_l2	0.000	0.003	-0.006	0.006	.917
	2	AR_l	0.016	0.013	-0.010	0.042	.234

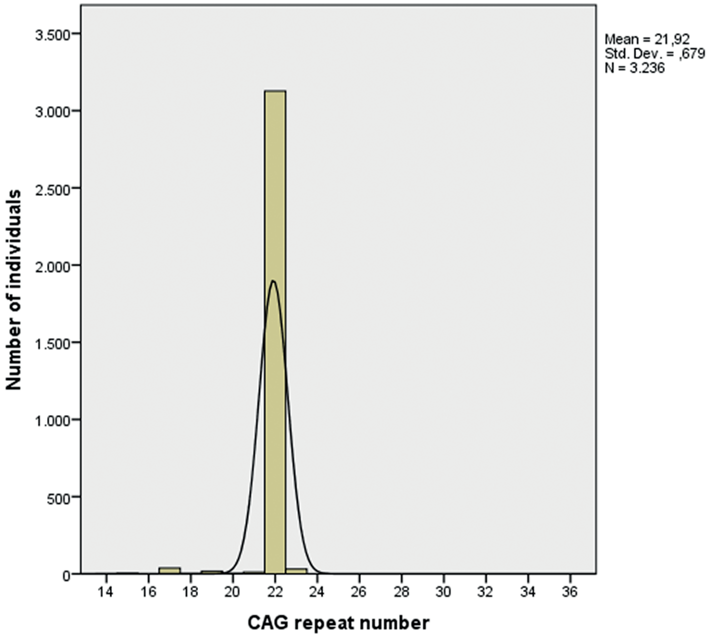
s=relatively shorter allele; l=relatively longer allele; s2=quadratic term relatively shorter allele; l2=quadratic term relatively longer allele; sl=interaction term relatively shorter and longer allele. SE=standard error. CI=confidence interval. [1]) Numbers represent the sequential steps in the binary logistic analysis in which non-significant quadratic and interaction terms were sequentially removed to arrive at the final model. [2]) The step in which the results were adjusted for age, gender and level of education.



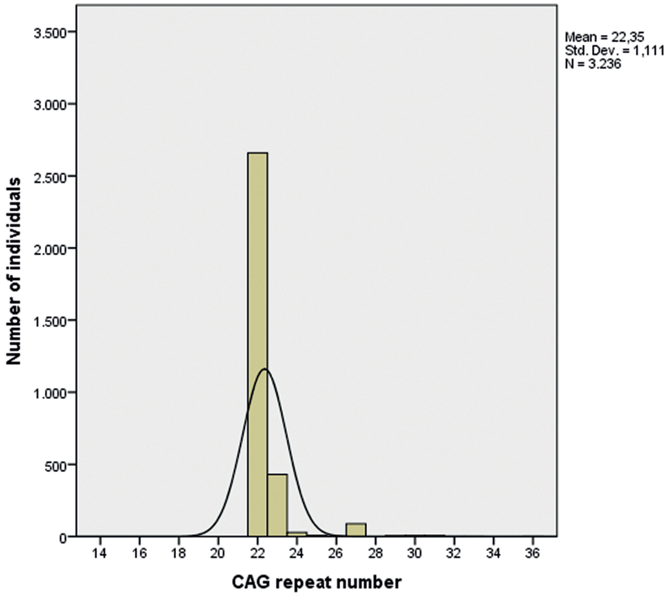
Supplementary Figure 1. The distribution of the CAG repeat number in the shorter *ATXN1* allele. The CAG repeat number in the shorter allele of *ATXN1* shows a normal distribution among individuals.



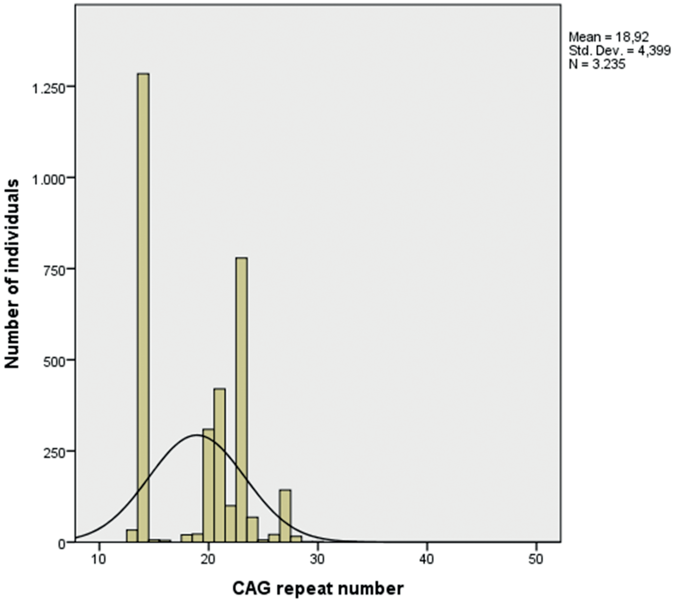
Supplementary Figure 2. The distribution of the CAG repeat number in the longer *ATXN1* allele. The CAG repeat number in the longer allele of *ATXN1* shows a normal distribution among individuals.



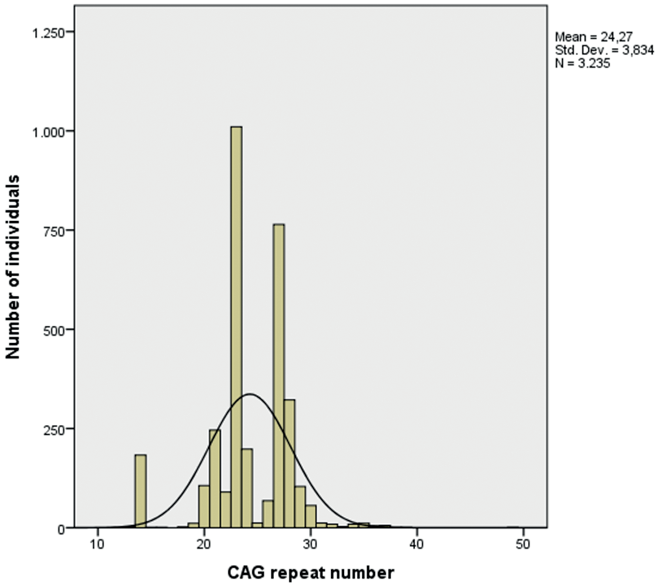
Supplementary Figure 3. The distribution of the CAG repeat number in the shorter *ATXN2* allele. There is little variation in the distribution of the CAG repeat number in the shorter allele of *ATXN2* among individuals. Nonetheless, the distribution appears to be relatively normal.



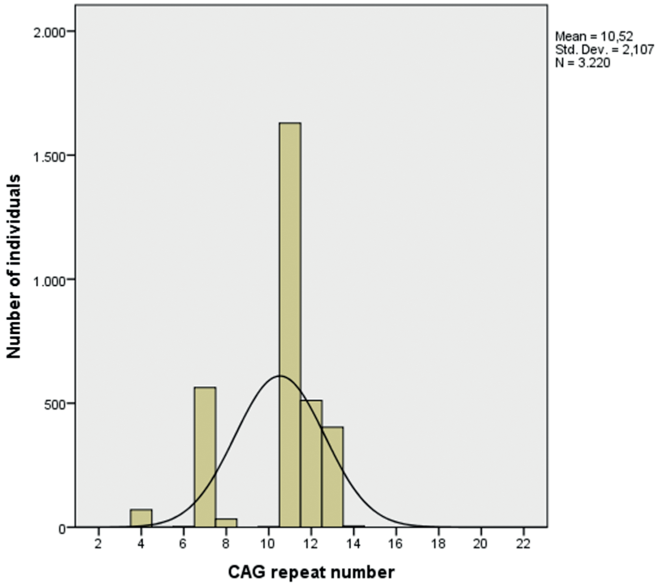
Supplementary Figure 4. The distribution of the CAG repeat number in the longer *ATXN2* allele. There is little variation in the distribution of the CAG repeat number in the longer allele of *ATXN2* among individuals. Nonetheless, the distribution appears to be relatively normal.



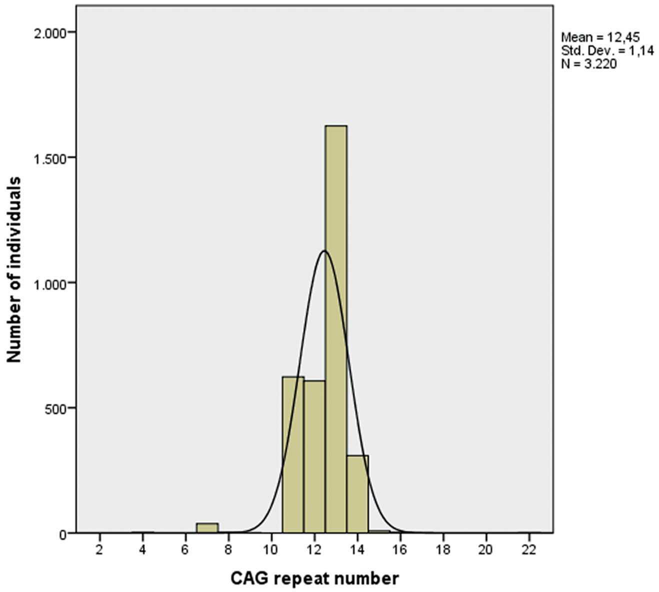
Supplementary Figure 5. The distribution of the CAG repeat number in the shorter *ATXN3* allele. In the distribution of CAG repeat number in the shorter *ATXN3* allele there appears to be a spike in the frequency of alleles with 14 CAG repeats. This spike disrupts the normal distribution of the CAG repeat number.



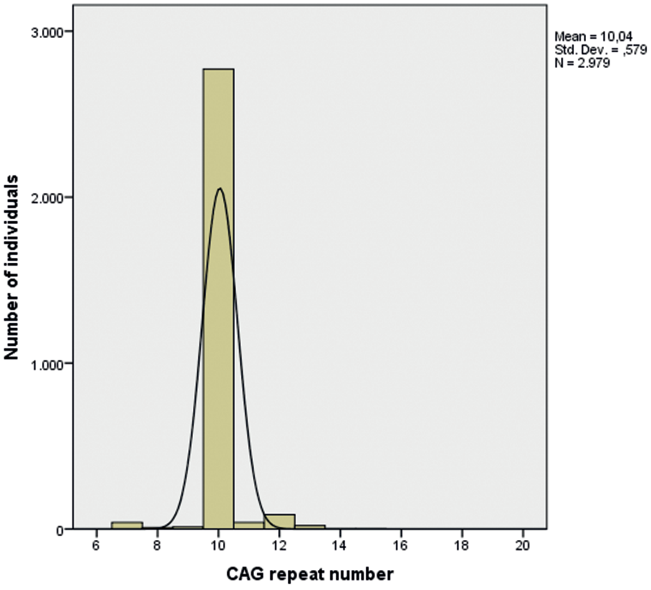
Supplementary Figure 6. The distribution of the CAG repeat number in the longer *ATXN3* allele. In the distribution of the CAG repeat number in the longer *ATXN3* allele there appears to be a spike in the frequency of alleles with 14 CAG repeats. This spike disrupts the normal distribution of the CAG repeat number. Nonetheless, the distribution remains relatively normal.



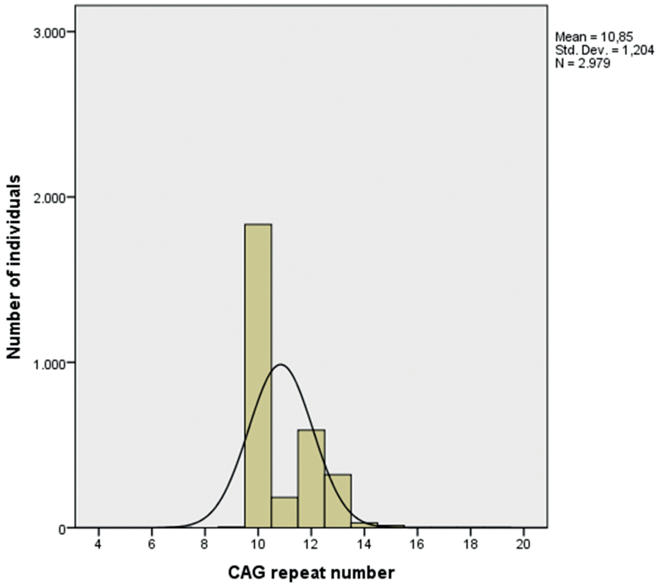
Supplementary Figure 7. The distribution of the CAG repeat number in the shorter *CACNA1A* allele. In the distribution of CAG repeat number in the shorter *CACNA1A* allele the otherwise normal distribution appears to be disrupted by missing frequencies around a CAG repeat number of 8-10.



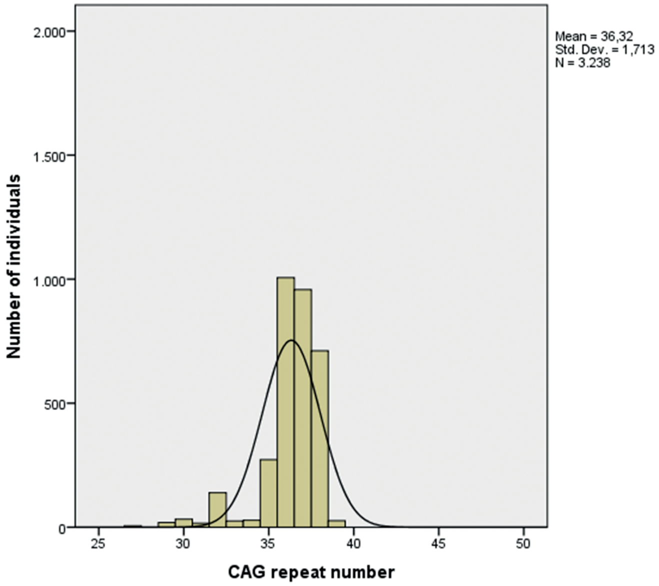
Supplementary Figure 8. The distribution of the CAG repeat number in the longer *CACNA1A* allele. In the distribution of CAG repeat number in the longer *CACNA1A* allele appears to be normal among individuals.



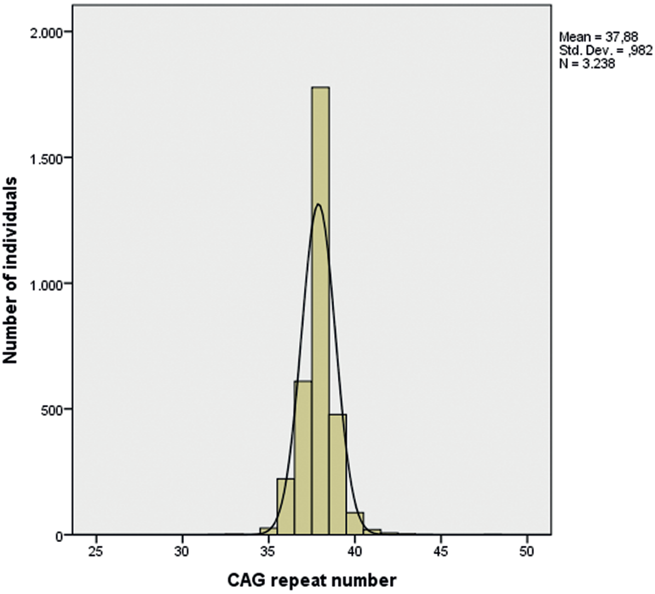
Supplementary Figure 9. The distribution of the CAG repeat number in the shorter *ATXN7* allele. There appears to be little variation in the distribution of CAG repeat number in the shorter *ATXN7* allele. Nonetheless, the distribution seems to be relatively normal.



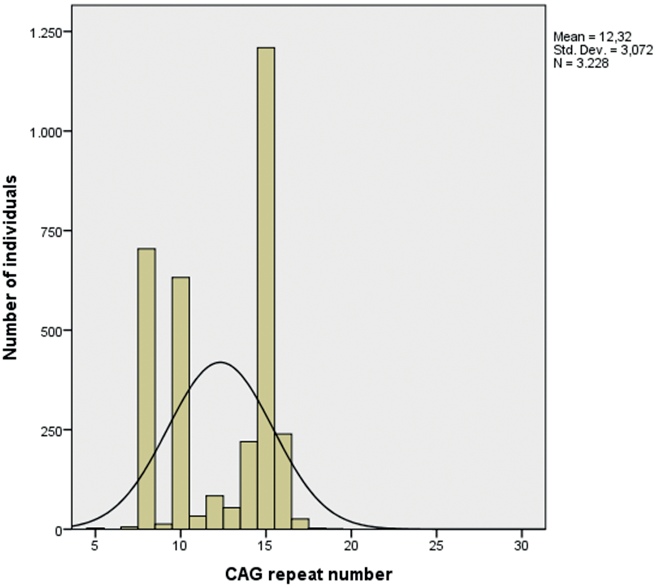
Supplementary Figure 10. The distribution of the CAG repeat number in the longer *ATXN7* allele. The distribution of CAG repeat number in the longer *ATXN7* allele seems to be skewed to the left.



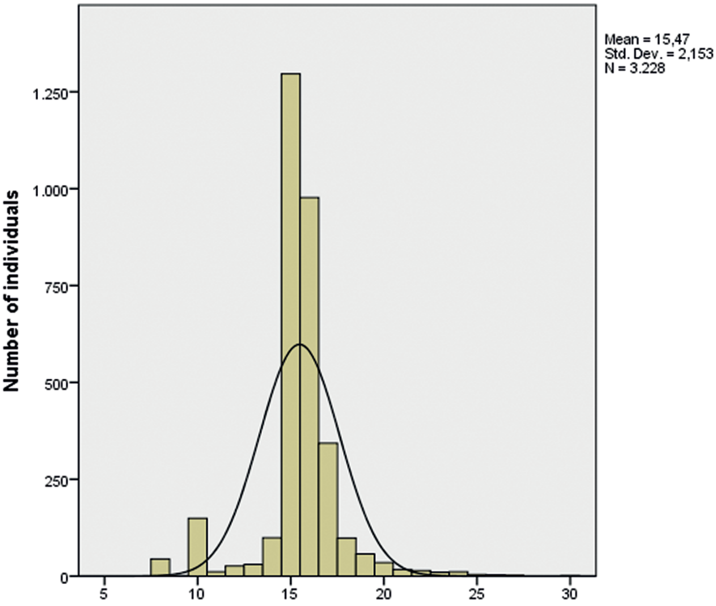
Supplementary Figure 11. The distribution of the CAG repeat number in the shorter *TBP* allele. The distribution of CAG repeat number in the shorter *TBP* allele seems to be skewed a bit to the right.



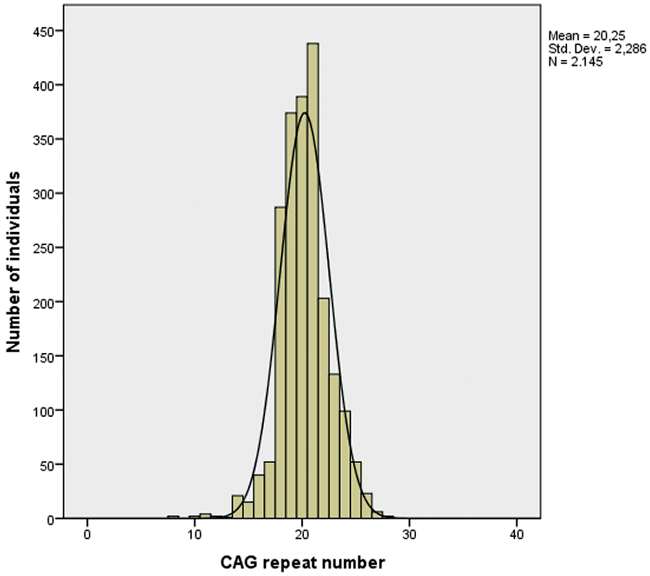
Supplementary Figure 12. The distribution of the CAG repeat number in the longer *TBP* allele. The distribution of CAG repeat number in the longer *TBP* allele seems to normal among individuals.



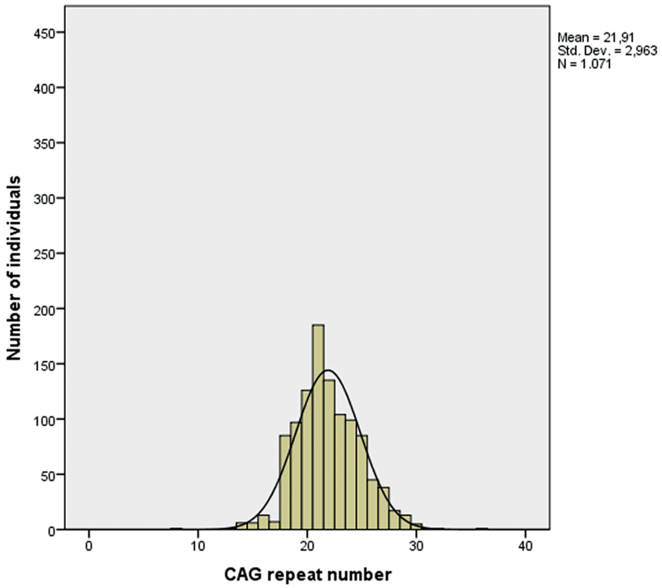
Supplementary Figure 13. The distribution of the CAG repeat number in the shorter *ATN1* allele. The distribution of CAG repeat number in the shorter *ATN1* allele seems to have two spikes in the frequencies around the CAG repeat numbers of 8 and 10.



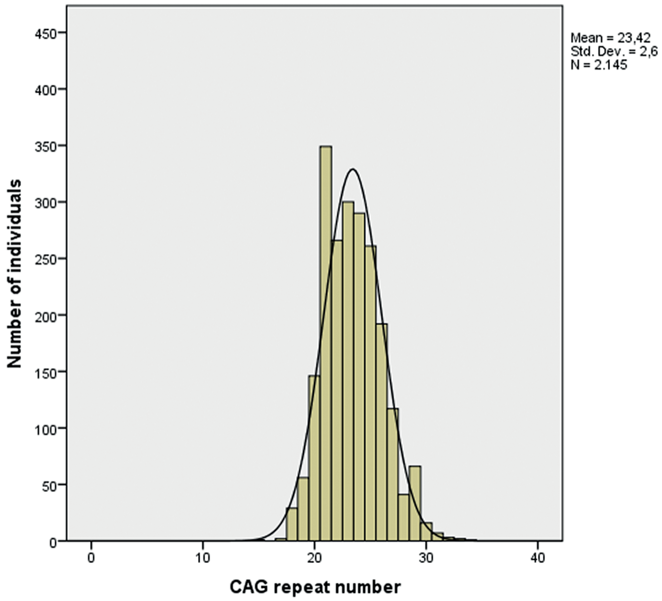
Supplementary Figure 14. The distribution of the CAG repeat number in the longer *ATN1* allele. The distribution of CAG repeat number in the longer *ATN1* allele seems to have a normal distribution among individuals.



Supplementary Figure 15. The distribution of the CAG repeat number in the shorter *AR* allele among women. The distribution of CAG repeat number in the shorter *AR* allele among women seems to have a normal distribution.



Supplementary Figure 16. The distribution of the CAG repeat number in the longer *AR* allele among women. The distribution of CAG repeat number in the longer *AR* allele among women seems to have a normal distribution.



Supplementary Figure 17. The distribution of the CAG repeat number in the *AR* allele among men. The distribution of CAG repeat number in the *AR* allele among men seems to have a normal distribution.