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. Worldwide, depression is responsible for 74.5 million disability-adjusted life years and in 2010 the total economic burden of depression in the United States alone was estimated at $210.5 billion, 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%. Nevertheless, genome-wide association studies (GWASs) have shown limited success in identifying its genetic basis. In a meta-analysis with more than 18 000 subjects from European ancestry, no single-nucleotide polymorphism (SNP) reached genome-wide significance. However, recently, a meta-analysis of results from three studies, together containing 180 866 individuals, found two lead SNPs associated with depression. 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, 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.

Tandem repeats constitute ~3% of the human genome, a higher percentage than the entirety of the protein-coding sequences, substantially contributing to genetic variation. Many rare hereditary disorders are caused by expansions of simple DNA repeat sequences. 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. 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 polyglutamine domain in the associated proteins is characterized by both motor and neurodegenerative impairment. Huntington disease is caused by a CAG...
repeat expansion in exon 1 of the huntingtin (HTT) gene. Ap1-14, Recent studies have 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’. Apart from Huntington disease, 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 these other PDAGs to depression remains 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).

MATERIALS 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 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, Leusden, the Netherlands) with labelled primers flanking the CAG stretch of the PDAGs (Bioregio, Nijmegen, the Netherlands; Supplementary Table 1). The PCR was performed using 10 ng of genomic DNA, 1 x OneTaq mastermix (New England Biolabs, Ipswich, MA, USA, OneTaq Hot start with GC Buffer mastermix), 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 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, Foster City, CA, USA) and analysed using the GeneMarker software version 2.4.0. (SoftGenetics, State College, PA, USA). For every analysis, we included three controls with known CAG repeat lengths for each PDAG to ensure that every run was performed reliably. All assessments were made 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 (that is, 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 nonlinear 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’).52-54 In order to assess whether the effect of CAG repeat size variations was independent of these well-established risk factors for depression, the Nagelkerke R2 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 s.e.’s. Moreover, to visualise our results as well as to assure

<table>
<thead>
<tr>
<th>Gene</th>
<th>Disease</th>
<th>Protein</th>
<th>CAG repeat ranges</th>
<th>Allele</th>
<th>Mean</th>
<th>Median</th>
<th>N</th>
<th>Range</th>
<th>No. of intermediate*</th>
<th>No. of pathological*</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATXN1</td>
<td>SCA1</td>
<td>Ataxin-1</td>
<td>6–39</td>
<td>Short</td>
<td>29.22</td>
<td>29.00</td>
<td>3228</td>
<td>18–35</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ATXN2</td>
<td>SCA2</td>
<td>Ataxin-2</td>
<td>14–32</td>
<td>Short</td>
<td>21.91</td>
<td>22.00</td>
<td>3236</td>
<td>14–24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ATXN3</td>
<td>SCA3</td>
<td>Ataxin-3</td>
<td>12–44</td>
<td>Short</td>
<td>18.96</td>
<td>20.00</td>
<td>3235</td>
<td>13–33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CACNA1A</td>
<td>SCA6</td>
<td>CACNA1A</td>
<td>4–18</td>
<td>Short</td>
<td>10.52</td>
<td>11.00</td>
<td>3220</td>
<td>2–14</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ATXN7</td>
<td>SCA7</td>
<td>Ataxin-7</td>
<td>3–19</td>
<td>Short</td>
<td>10.04</td>
<td>10.00</td>
<td>2979</td>
<td>7–15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TBP</td>
<td>SCA17</td>
<td>TBP</td>
<td>25–43</td>
<td>Short</td>
<td>36.31</td>
<td>37.00</td>
<td>3238</td>
<td>27–39</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ATN1</td>
<td>DRPLA</td>
<td>Atrophin-1</td>
<td>3–38</td>
<td>Short</td>
<td>12.36</td>
<td>14.00</td>
<td>3228</td>
<td>5–22</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AR</td>
<td>SMBA</td>
<td>Androgen receptor</td>
<td>6–36</td>
<td>Short</td>
<td>20.80</td>
<td>21.00</td>
<td>3228</td>
<td>8–36</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Abbreviations: CACNA1A, calcium channel, voltage-dependent P/Q type, α 1A subunit; CAG, cytosine–adenine–guanine; DRPLA, dentatorubropallidolysian atrophy; SBMA, spinal bulbar muscular atrophy; SCA, spinocerebellar ataxia; TBP, thymine-adenine-thymine-adenine (TATA) box-binding protein.
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 correction as described by Benjamini and Hochberg, assuming eight independent tests with a two-sided \( \alpha \) of 0.05.39 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, Armonk, NY, USA).

### RESULTS

Risk of lifetime depression increases with higher \( \text{ATXN7} \) CAG repeat sizes

In total, we were able to genotype a total of 2979 participants for \( \text{ATXN7} \) (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 \( \text{ATXN7} \) (Supplementary Table 2). These lacking samples were missing completely at random. The CAG repeat lengths ranged from 7 to 19 repeats (Table 1).

We found a significant association between the risk of lifetime depression and the CAG repeat length of \( \text{ATXN7} \) in the shorter allele (\( \beta = 0.202, P = 0.006 \)). Adjusting for the effects of gender, age and education hardly changed these results (\( \beta = 0.184, P = 0.013 \); Supplementary Table 3). After adjusting for multiple testing, this association remained significant. Inclusion of the CAG repeat length of the \( \text{ATXN7} \) 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 (that is, the \( R^2 \) increased from 0.024 to 0.028). Considering our control group included subjects with anxiety disorders without comorbid 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 (\( \beta = 0.215, P = 0.009 \)).

For the non-parametric method, we divided all subjects based on the median CAG repeat size for each \( \text{ATXN7} \) 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 with the reference category in which both \( \text{ATXN7} \) alleles had a CAG repeat number equal to or smaller than the median.

### \( \text{TBP} \) CAG repeat sizes in both alleles interact to affect the risk of lifetime depression

A total of 3238 individuals were genotyped for \( \text{TBP} \) (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 \( \text{TBP} \) (Supplementary Table 2). The lacking samples were missing completely at random. For \( \text{TBP} \), we found CAG repeat lengths ranging between 27 and 48 repeats (Table 1).

We found the interaction between the number of CAG repeats in the two \( \text{TBP} \) alleles to be significantly associated with the risk of lifetime depression (\( \text{TBP} \) short allele: \( \beta = -2.270, P = 0.004 \); \( \text{TBP} \) long allele: \( \beta = -2.112, P = 0.006 \); \( \text{TBP} \) interaction term: \( \beta = 0.060, P = 0.005 \)). When adjusting for the effects of gender, age and
protein. Bars indicate ± both shorter and longer alleles contain a CAG repeat number equal to or smaller than their median; II, the relatively longer TBP allele contains a CAG repeat number equal to or larger than their median; and III, both TBP alleles contain a CAG repeat number equal to or larger than their median. Error bars indicate ± one s.e. *P < 0.05 by Fisher’s exact test in comparison with the reference category.

Table 3. The distribution of participants in three TBP categories

<table>
<thead>
<tr>
<th>TBP category</th>
<th>Allele</th>
<th>Mean</th>
<th>Range</th>
<th>Lifetime depression</th>
<th>No lifetime depression</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Short</td>
<td>35.78</td>
<td>27–37</td>
<td>1462</td>
<td>716</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>37.47</td>
<td>32–38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>Short</td>
<td>36.04</td>
<td>30–37</td>
<td>230</td>
<td>93</td>
<td>1.21 (0.94–1.57)</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>39.30</td>
<td>39–48</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Short</td>
<td>38.09</td>
<td>38–39</td>
<td>201</td>
<td>74</td>
<td>1.33 (1.00–1.76)</td>
</tr>
<tr>
<td></td>
<td>Long</td>
<td>39.28</td>
<td>39–44</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

I, both TBP alleles contain a CAG repeat number ≤ their median; II, the relatively longer TBP allele contains a CAG repeat number > their median and the relatively shorter TBP allele contains a CAG repeat number ≤ their median; III, both TBP alleles contain a CAG repeat number > their median (median short allele = 37, median long allele = 38); CAG, cytosine–adenine–guanine; CI, confidence interval. *Figures represent the number of subjects per TBP category in the group with lifetime depression. This category contains the largest number of subjects and was therefore chosen as the reference category for odds ratio calculation.

Figure 2. Odd's ratio for lifetime depression per TBP allele. The odd’s ratio for lifetime depression increases significantly when both the relatively shorter TBP allele and the relatively longer 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, is 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 s.e. *P < 0.05 by Fisher’s exact test in comparison with the reference category. CAG, cytosine–adenine–guanine; TBP, thymine–adenine–thymine–adenine (TATA) box-binding protein.

education, the results hardly changed (TBP short allele: $\beta = -2.302, P = 0.004$; TBP long allele: $\beta = -2.150, P = 0.005$; TBP interaction term: $\beta = 0.061, P = 0.005$; Supplementary Table 3). Furthermore, the results remained significant after performing the sensitivity analysis and applying the false discovery rate correction.$^{39}$ The proportion of explained genetic variation increased by 0.005 from the baseline model including only gender, age and level of education (that is, the $R^2$ increased from 0.023 to 0.028).

For the non-parametric method, we divided all subjects in three groups based on the 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 with the reference category, in which both alleles had a CAG repeat number equal to or lower than their median (Figure 2).

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. As 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: $\beta = -0.286, P = 0.018$; HTT long allele quadratic term $\beta = -0.006, P = 0.017$; ATXN7 short allele: $\beta = 0.210, P = 0.005$; TBP short allele: $\beta = -2.728, P = 0.001$; TBP long allele: $\beta = -2.566, P = 0.001$; TBP interaction term: $\beta = 0.072, P = 0.001$). 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 ~ 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 ~ 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 (that is, 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 (that is, 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 = 0.059).

**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, that is, 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 (two intermediate, four 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 who had a CAG repeat length exceeding the median in both TBP alleles. These effect sizes are considerable, especially when compared with 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). 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 ~ 1.9% of depression heritability. Using the same population prevalence of 1.5% for depression, the meta-analysis in this GWAS calculated a heritability of 5.9%.5

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).33 SCA7 is a progressive autosomal-dominant neurodegenerative disorder primarily characterized by cerebellar ataxia and macular degeneration.52 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 characterisation. 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 complex.68 TATA-binding protein-free TAF-containing complex allows for the initiation of transcription via RNA polymerase II in the absence of the RNA polymerase II transcription factor D.55 Interestingly, the TATA-box-binding protein (TBP) that is encoded by TBP, is the DNA-binding subunit of RNA polymerase II transcription factor D. TBP anchors RNA polymerase II transcription factor D 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 have a part in the initiation of transcription via RNA polymerase II. RNA polymerase II catalyses the transcription of DNA to synthesise precursors of mRNA, snRNA and microRNA.57,58 Numerous studies indicate an altered mRNA expression of proteins such as the serotonin 5-HT1A receptor,59–61 brain-derived neurotrophic factor,62–64 and corticotrophin-releasing factor in subjects suffering from MDD compared with 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.55–57 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 and 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 10,000 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 (that is, 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 with 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 that estimated the prevalence of incompletely penetrant HTT alleles in MDD patients to be ~3 in 1000, whereas such alleles were absent in the control group. 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 minimising 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 generalisability 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.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

ACKNOWLEDGMENTS
This study was supported by a VENI-grant (#916.15.080) from the Netherlands Organization of Scientific Research and a Marie Sklodowska-Curie Individual Fellowship grant from the European Union (Horizon 2020, 7401130; NAA). The infrastructure for the NESDA study (www.nesda.nl) has been funded through the Geestkracht programme 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 organisations (VU University Medical Centre, Leiden University Medical Centre, University Medical Centre Groningen, UMC St Radboud, and GGZ inGeest, GG Net, GGZ Nijmegen and Parnassia).

REFERENCES
24 Zhuchenko O, Bailey J, Bonnen P, Ashizawa T, Zoghbi HY. Molecular and clinical studies in SCA-7 de
termination (ADCA type II) caused by CAG triplet repeat expansion in families with autosomal dominant motor disorders, including the Drew family of Walworth.
33 Nagafuchi S, Yanagisawa H, Sato K, Shirayama T, Ohsaki E, Bundo M et al. Den

Translational Psychiatry (2017), 1 – 7
Supplementary Information accompanies the paper on the Translational Psychiatry website (http://www.nature.com/tp).