

ORIGINAL ARTICLE

The *C9orf72* repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter

I Gijssels^{1,2}, S Van Mossevelde^{1,2}, J van der Zee^{1,2}, A Sieben^{1,2,3}, S Engelborghs^{2,4}, J De Bleeker³, A Ivanou⁵, O Deryck⁶, D Edbauer^{7,8}, M Zhang⁹, B Heeman^{1,2}, V Bäumer^{1,2}, M Van den Broeck^{1,2}, M Mattheijssens^{1,2}, K Peeters^{1,2}, E Rogava^{9,10}, P De Jonghe^{1,2,11}, P Cras^{2,11}, J-J Martin², PP de Deyn^{2,4,12}, M Cruts^{1,2} and C Van Broeckhoven^{1,2} on behalf of the BELNEU CONSORTIUM¹³

Pathological expansion of a G₄C₂ repeat, located in the 5' regulatory region of *C9orf72*, is the most common genetic cause of frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS). *C9orf72* patients have highly variable onset ages suggesting the presence of modifying factors and/or anticipation. We studied 72 Belgian index patients with FTLD, FTLD–ALS or ALS and 61 relatives with a *C9orf72* repeat expansion. We assessed the effect of G₄C₂ expansion size on onset age, the role of anticipation and the effect of repeat size on methylation and *C9orf72* promoter activity. G₄C₂ expansion sizes varied in blood between 45 and over 2100 repeat units with short expansions (45–78 units) present in 5.6% of 72 index patients with an expansion. Short expansions co-segregated with disease in two families. The subject with a short expansion in blood but an indication of mosaicism in brain showed the same pathology as those with a long expansion. Further, we provided evidence for an association of G₄C₂ expansion size with onset age ($P < 0.05$) most likely explained by an association of methylation state of the 5' flanking CpG island and expansion size in blood ($P < 0.0001$) and brain ($P < 0.05$). In several informative *C9orf72* parent–child transmissions, we identified earlier onset ages, increasing expansion sizes and/or increasing methylation states ($P = 0.0034$) of the 5' CpG island, reminiscent of disease anticipation. Also, intermediate repeats (7–24 units) showed a slightly higher methylation degree ($P < 0.0001$) and a decrease of *C9orf72* promoter activity ($P < 0.0001$) compared with normal short repeats (2–6 units). Decrease of transcriptional activity was even more prominent in the presence of small deletions flanking G₄C₂ ($P < 0.0001$). Here we showed that increased methylation of CpGs in the *C9orf72* promoter may explain how an increasing G₄C₂ size lead to loss-of-function without excluding repeat length-dependent toxic gain-of-function. These data provide insights into disease mechanisms and have important implications for diagnostic counseling and potential therapeutic approaches.

Molecular Psychiatry (2016) **21**, 1112–1124; doi:10.1038/mp.2015.159; published online 20 October 2015

INTRODUCTION

A pathological expansion of a hexanucleotide repeat, G₄C₂, located in the 5' regulatory region of *C9orf72* is the most frequent hereditary cause of frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS).^{1–3} Moreover, it is the most frequent causal mutation in patients suffering from both FTLD and ALS and in families segregating both clinical phenotypes (FTLD–ALS). Together with the clinical and pathological overlap between these two disorders, identification of *C9orf72* provided strong evidence for the presence of an FTLD–ALS disease continuum of which the pure clinical forms of FTLD and ALS represent the two extremes.

C9orf72 is transcribed in three major transcripts encoding two protein isoforms (*C9orf72* a and b), which may have a role in autophagy and endosomal trafficking⁴ and might be involved in regulating endoplasmic reticulum stress.⁵ The repeat sequence is part of the functional core promoter of all three *C9orf72* transcripts² (Supplementary Figure S1).

In FTLD and ALS patients carrying a G₄C₂ expansion, allele-specific reduction of *C9orf72* expression in brain tissue^{1,2,6–9} and hypermethylation of the G₄C₂ repeat¹⁰ and the flanking CpG island^{11–15} was observed. Hypermethylation of CpG-rich promoters has been associated with transcriptional silencing in noncoding repeat expansion disorders like, for example, Fragile X syndrome,^{16,17} and CpG methylation state has been directly

¹Neurodegenerative Brain Diseases Group VIB Department of Molecular Genetics, Antwerp, Belgium; ²Institute Born-Bunge, University of Antwerp, Antwerp, Belgium; ³Department of Neurology, University Hospital Ghent and University of Ghent, Ghent, Belgium; ⁴Department of Neurology and Memory Clinic, Hospital Network Antwerp Middelheim and Hoge Beuken, Antwerp, Belgium; ⁵Department of Neurology, Saint-Luc University Hospital and Institute of Neuroscience, Université Catholique de Louvain, Brussels, Belgium; ⁶Department of Neurology, General Hospital Sint-Jan Brugge-Oostende, Bruges, Belgium; ⁷German Center for Neurodegenerative Diseases (DZNE), Munich, Germany; ⁸Adolf Butenandt Institute, Biochemistry and Munich Cluster of Systems Neurology (SyNergy), Ludwig-Maximilians University Munich, Munich, Germany; ⁹Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, Toronto, Ontario, Canada; ¹⁰Division of Neurology, Department of Medicine, University of Toronto, Toronto, Ontario, Canada and ¹¹Department of Neurology, Antwerp University Hospital, Edegem, Belgium. Correspondence: Professor C Van Broeckhoven or Professor M Cruts, Neurodegenerative Brain Diseases Group, VIB Department of Molecular Genetics, University of Antwerp – CDE, Universiteitsplein 1, Antwerp B-2610, Belgium. E-mail: christine.vanbroeckhoven@molgen.vib-ua.be or marc.cruts@molgen.vib-ua.be

¹²Peter P de Deyn is also affiliated to the Department of Neurology and Alzheimer research Center, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands.

¹³Members of the BELNEU consortium are listed at the end of the paper.

Received 12 May 2015; revised 15 July 2015; accepted 5 August 2015; published online 20 October 2015

correlated with repeat expansion size in Friedreich ataxia.¹⁸ Further, both sense and antisense transcription and repeat-associated non-ATG translation of the expanded repeat result in nuclear RNA foci and aggregated dipeptide repeat (DPR) neuropathology.^{1,19–24} However, the exact role of each mechanism in the disease remains unclear. Some observations might represent benign side effects and conversely, the proposed mechanisms might have an effect on specific transcripts leading to degeneration of particular neuronal populations affected in FTLD or ALS.

The pathogenic nature of the repeat depends on its size but the cut-off between normal and pathogenic alleles is not well established. In control populations, the normal repeat sizes ranged between 2 and 24 units with those of 7–24 units potentially acting as risk alleles for disease and hence named intermediate alleles.²⁵ Some studies consider repeats of > 30 units as pathogenic,³ whereas others use a cut-off of 60 units,² depending on the upper limit of the repeat-primed PCR detection method. Exact sizing of the expanded G₄C₂ repeat has been limited because of its 100% GC content, its large size, somatic instability and the repetitive nature of its flanking sequences. Southern blot hybridization studies visualized the expanded alleles and estimated that the size of most repeat expansions ranged between several hundred and several thousand repeat units.^{1,6,26–32} G₄C₂ hexanucleotide repeat sizes of 25–60 repeat units were rarely observed in FTLD, ALS and related disorders;^{6,26,27,30,33–38} however, in most of the studies co-segregation with disease or other arguments for pathogenicity were not observed in families. The apparent size gap between normal and expanded repeat alleles suggests that these repeat sizes might have the propensity to either rapidly expand to pathological sizes or contract to normal sizes. Therefore, the shortest size of a pathological G₄C₂ repeat expansion outside the pathogenic size range remains elusive.

In most repeat expansion disorders, the size of a pathological repeat expansion influences the severity of the clinical symptoms. For example, in myotonic dystrophy type I caused by an expanded (CTG)_n repeat located in the noncoding region of *DMPK*,³⁹ longer expansions correlate with more severe symptoms and an earlier onset age in successive generations because of genetic anticipation. Also, in Friedreich ataxia, the onset age is influenced by the size of a (GAA)_n repeat expansion within the first intron of *FXN*.⁴⁰ In this context, we observed in 26 G₄C₂ expansion carriers a highly variable onset age from 42 to 69 years and a disease duration of 1.5 to 17 years, as well as carriers living up to 76 years without signs of disease, suggesting the presence of disease-modifying factors.^{2,41} A disease liability risk curve calculated that by the age of 70 years, 91% of the carriers will be affected.⁴¹ Further, along with others, we described in FTLD–ALS families segregating a C9orf72 G₄C₂ repeat expansion a decreasing onset age of 7 to 11 years in each younger generation^{2,41–46} suggesting disease anticipation. Whether the clinical variability of FTLD–ALS is due to size variability of the C9orf72 G₄C₂ expansion is still unclear. Studies examining the effect of repeat length on onset age could not establish an inverse correlation.^{6,26,30–32} Studies of genetic anticipation in C9orf72 families are lacking.

In this study, we aimed to address these unresolved concerns by evaluating the correlation of the size of the G₄C₂ repeat and onset age of FTLD–ALS in families and sporadic repeat expansion carriers. In addition, we examined the effect of repeat size variability on CpG methylation of the C9orf72 repeat region and on C9orf72 transcriptional activity, including the effect of the presence of disease-related small insertion/deletion (indels) polymorphisms in the 3' flanking sequences.²⁵

MATERIALS AND METHODS

Participants and study design

The Belgian FTLD and ALS cohorts used in this study consisted of 549 index patients with FTLD of which 35 with concomitant FTLD and ALS (FTLD–ALS), and 210 patients with ALS. Of 72 index patients, blood-derived DNA was used in this study and in addition 61 affected or at-risk relatives with a repeat expansion were collected. Patients were ascertained in Belgium through an ongoing multicenter collaboration of neurology departments and memory clinics partnering in the Belgian Neurology (BELNEU) consortium of neurologists affiliated to specialized memory clinics, neuromuscular reference centers and neurology departments in Belgium.^{2,41} Additional patients were included who had initially been referred to the Diagnostic Service Facility for medical genetic testing. All frontotemporal dementia (FTD) patients were evaluated using a standard diagnostic protocol including detailed recording of clinical and family history, neuropsychological testing, neurological examination, biochemical analyses, and neuroimaging (structural (brain computed tomography or magnetic resonance imaging) and in a selection of patients functional (single photon emission computed tomography or fluorodeoxyglucose-positron emission tomography)) and were followed longitudinally on a regular basis. Clinical diagnoses of behavioral variant FTD, progressive non-fluent aphasia or semantic dementia were based on established clinical criteria for FTD^{47,48} and was made by consensus by at least two neurologists. ALS patients had a clinical diagnosis of definite, probable or possible ALS according to the revisited El Escorial criteria for ALS.^{49,50} The Belgian control cohort consisted of 1044 unrelated individuals free of personal and familial history of neurodegenerative or psychiatric diseases and with a mini-mental state examination score > 24. All participants and/or their legal guardian provided written informed consent for participation in clinical and genetic studies. The clinical study protocol and the informed consent forms used in patient ascertainment were approved by the ethics committee of the respective hospitals. The genetic study protocols and informed consent forms were approved by the ethics committees of the University of Antwerp and the University Hospital Antwerp, Belgium.

Procedures

Blood DNA of all 133 G₄C₂ expansion carriers and DNA from frontal cortex, temporal cortex and cerebellum of four patients with a long expansion (> 80 units) and one with a short expansion (< 80 units) was analyzed with a 'short expansion PCR'. We used the KAPA2G Robust HotStart DNA polymerase (Kapa Biosystems, Wilmington, MA, USA) with a primer pair flanking the repeat. The resulting PCR products were size separated on agarose gel and genotyped.

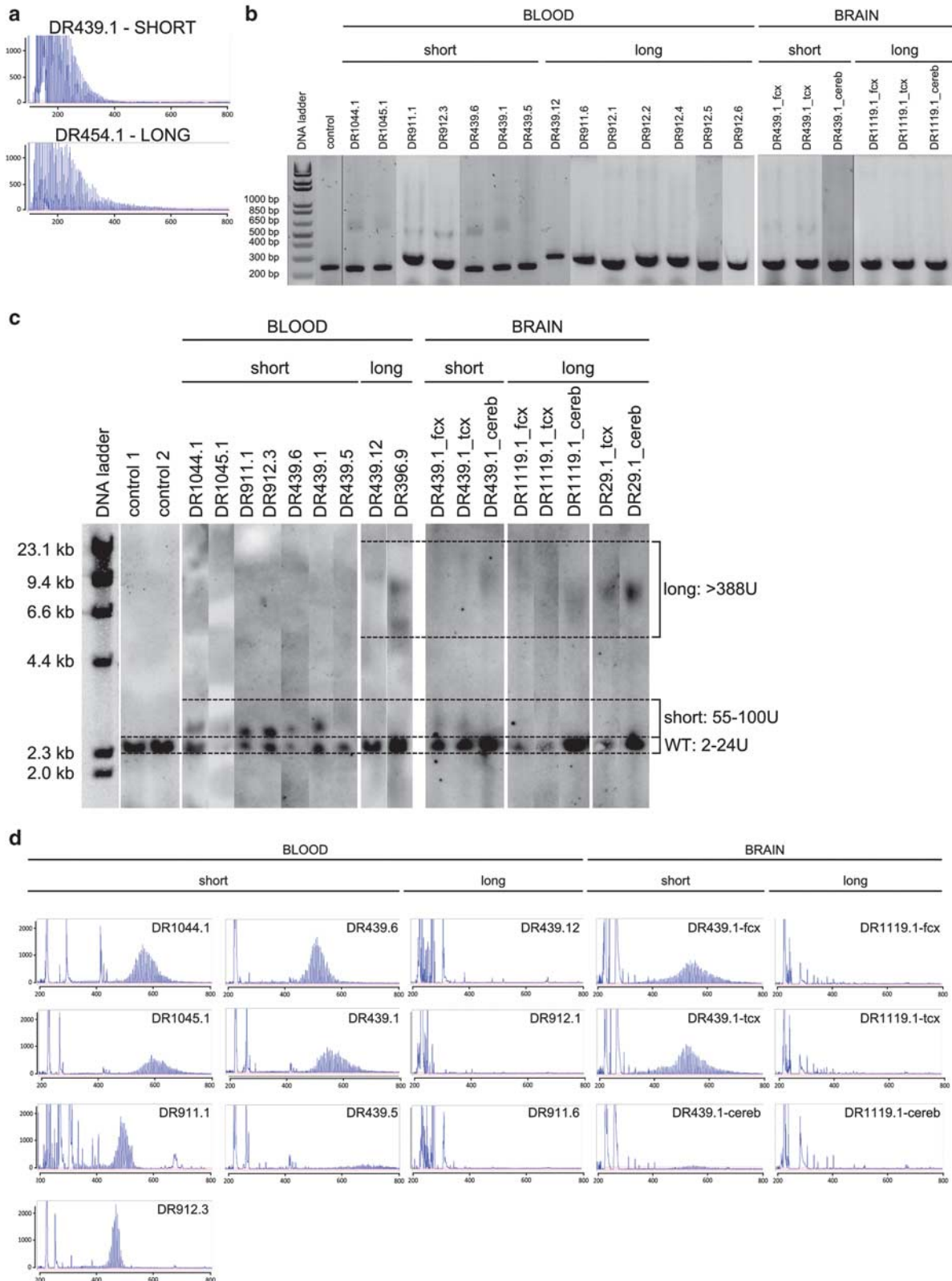
Of 27 samples with sufficient high molecular weight DNA from blood, the expansion size was analyzed by Southern blot hybridization. DNA extracted from frontal cortex, temporal cortex and cerebellum of two long and one short expansion carrier was also analyzed. In all, 20 µg genomic DNA was digested with *Xba*I and separated on a 0.8% agarose gel. Capillary blotting for at least 65 h was followed by ultraviolet crosslinking. Membranes were hybridized with a non-repetitive digoxigenin-labeled PCR probe overnight at 47 °C and subsequently washed in 0.1% sodium dodecyl sulfate buffer containing decreasing concentrations of saline sodium citrate. For probe detection, the procedure as described in the digoxigenin application manual was used using CDPStar (Roche Applied Science, Basel, Switzerland).

Autopsied brains of five patients with a C9orf72 repeat expansion were stained with anti-ubiquitin antibody, AT8 against hyperphosphorylated tau, 4G8 against β-amyloid, anti-FUS antibody and anti-TDP-43 antibody. In addition, poly-(Gly-Ala) DPR proteins and p62 were stained on brain of patient DR439.1 with a short repeat expansion and of two patients with a long repeat expansion.

We developed two methylation sensitive restriction enzyme (MSRE)-based assays followed by quantitative real-time PCR (MSRE-qPCR) to quantify methylation levels of the G₄C₂ repeat and of an upstream flanking CpG island. The G₄C₂ MSRE-qPCR TaqMan assay with primers flanking the G₄C₂ repeat was based on the methylation sensitivity of *Hpa*II evaluating the CpG methylation state of each G₄C₂ unit and two CpGs in the 3' flanking region (Supplementary Figure S2). To analyze the 5' flanking CpG island, a *Hha*I-based MSRE-qPCR assay was developed using primers upstream of the G₄C₂ repeat evaluating 1 of 17 CpGs (Supplementary Figure S2). Digested and non-digested DNA samples were amplified using the TaqMan Universal PCR Master Mix protocol on the ViiA7 Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). In accordance with the OneStep qMethyl kit of Zymo Research (Irvine, CA, USA), we calculated ΔCt

values and the corresponding methylation percentages ($100 \times 2^{-\Delta C_t}$) to estimate the methylation state of the studied amplicons. PCR amplification efficiencies were close to 100%. For the G₄C₂ assay, we selected individuals

homozygous for a normal short (S) (S/S: 2/2 and 5/5 units) or normal intermediate (I) repeat length (I/I: 8/8 and $\geq 8 / > 8$ (up to 21) units). In total, we analyzed the G₄C₂ repeat in blood genomic DNA of 42S/S and 12 I/I



patients, 46S/S and 28 I/I controls (Supplementary Table 1). In addition, only the wild-type allele of expansion carriers was studied because expanded alleles are not amplifiable with repeat flanking primers (42S and 18 I). Further, we included brain frontal cortex DNA of 11S/S patients, 3 I/I patients and 9S/S controls. We analyzed the 5' flanking CpG island in blood DNA of 80 repeat expansion carriers and frontal cortex, temporal cortex and cerebellum of five expansion carriers, of whom one short expansion carrier (DR439.1) and four with an expansion of >80 units (long expansion).

Bisulfite sequencing of the 5' flanking CpG island was performed as previously described¹⁴ on a representative selection of 26 DNA samples from blood with a normal (S/S and I/I) or expanded (short and long) repeat size.

We performed a reporter gene assay using constructs in which reporter gene activity is driven by a C9orf72 promoter fragment. We selected a 2-kb genomic C9orf72 promoter fragment (fragment 1) containing the G₄C₂ repeat and flanking expression regulatory elements and a shorter 1.3-kb genomic fragment (fragment 2) excluding the first noncoding exons of transcript variants 1 and 3 (Supplementary Figure S1). The DNA fragments containing diverse repeat sizes were obtained by PCR amplification of genomic DNA of carriers of differently sized G₄C₂ alleles and were cloned in a *Gaussia* luciferase reporter vector. In addition, we analyzed the effect of two small deletions in the 3' sequence flanking the G₄C₂ repeat: a GTGGT deletion on a G₄C₂ allele with six units and a 5'-CGGGGGCCGCCCCGGGGCGGGCC-3' deletion on a G₄C₂ allele with 12 units.

Reporter gene expression levels were assayed in human embryonic kidney cells (HEK293T) and neuronal SH-SY5Y cells. *Gaussia* luciferase activities were measured relative to *Cypridina* luciferase activities resulting in relative luciferase activities.

Statistical analysis

As a result of small sample sizes and no normal distributions, we performed a two-sided unpaired non-parametric Mann-Whitney *U*-test without assuming equal variances to compare onset age and methylation between short and long expansion carriers and the significance of differences in methylation state between different repeat sizes. Also relative luciferase activities between constructs of different repeat lengths were calculated by a two-sided Mann-Whitney *U*-test. For comparing

methylation state between two generations, we used a two-sided Wilcoxon test.

The webappendix contains further technical details of the procedures used.

RESULTS

Frequent occurrence of short C9orf72 G₄C₂ expansions

Although screening patients for the C9orf72 G₄C₂ repeat expansion using the repeat-primed PCR assay, we observed in some samples an atypical saw-tooth tail not extending beyond 500 bp (60 units) as generally observed in G₄C₂ expansion carriers (Figure 1a), which was suggestive for a shorter expansion. We therefore designed an optimized short expansion PCR assay, and analyzed blood DNA of 133 G₄C₂ expansion carriers. In most expansion carriers, we observed only the wild-type allele, indicating that their repeat expansion size was larger than 80 units, the approximate upper limit of the short expansion PCR assay. However, we detected seven short expansion carriers (7/133, 5.3%) of which six were affected by FTLD or ALS, including four probands (4/72, 5.6%) (Figure 1b; Table 1). Fluorescent fragment analysis showed a series of normally distributed peaks with a 6-bp periodicity (Figure 1d) and a median size ranging between 45 and 78 units (Table 1). We further confirmed the presence of a short expansion in blood DNA of all seven carriers using Southern blot hybridization, and estimated their expansion size between 55 and 100 repeat units (Figure 1c). The size distribution of the expanded allele was within a narrow range in all except one patient (DR439.5), showing additional bands of sizes larger than 440 units (Figure 1c). In total, we assessed blood DNA of 27 expansion carriers by Southern blot hybridization. The size of the G₄C₂ expansion ranged from 55 units to over 2100 units, the latter corresponding to the upper limit of size estimation with acceptable accuracy using standard agarose gel electrophoresis. The signals from blood DNA resulted in a smear or multiple bands

Table 1. Clinical characteristics and size analyses of short G₄C₂ expansion carriers

Individual	Family	Phenotype	AAO/current age (years)	Disease duration (years)	Family history	Median repeat size SB (min-max units)	Median repeat size PCR (min-max units)
DR912.3	DR912	At risk	33	—	F	55 (55-55)	45 (40-49)
DR911.1	DR911	FTLD	72	>2	F	55 (55-55)	47 (44-54)
DR439.6	DR439	FTLD	69	>5	F	55 (55-55)	50 (40-63)
DR439.1	DR439	FTLD	54	15	F	55 (55-55)	56 (47-76)
DR1044.1	DR1044	ALS	63	3	S	55 (55-55)	59 (52-78)
DR1045.1	DR1045	ALS	63	1	S	55 (55-55)	64 (57-82)
DR439.5	DR439	FTLD	52	>11	F	75 (55-100) And >440	78 (65-98)

Abbreviations: AAO, age at onset; ALS, amyotrophic lateral sclerosis; F, familial; FTLD, frontotemporal lobar degeneration; S, sporadic; SB, Southern blot.

Figure 1. Sizing of C9orf72 repeat expansions. (a) Repeat-primed PCR results of one short repeat expansion carriers (DR439.1) and one patient with a long expansion (DR454.1) showing the typical saw-tooth pattern. Short repeat expansions were recognized by a prematurely ending tail, whereas the peaks of the long expansion extended far beyond 60 units. (b) Agarose gel showing short expansion PCR results of short repeat expansion carriers and their relatives with a long repeat expansion (pedigrees in Figure 2), in blood DNA (left). PCR results on brain DNA (right) are shown for three brain regions of one short and one long expansion carrier (fcx, frontal cortex; tx, temporal cortex; cereb, cerebellum). Strong lower bands represent the wild-type allele and weak upper bands the short expansion alleles. (c) Southern blot hybridization results of blood DNA from seven short expansion carriers, two long expansion carriers and from two controls, and of brain DNA from one short expansion carrier and two long expansion carriers. The wild-type allele is indicated with black dotted lines and the expansion alleles with white dotted lines. Depending on the number of units on the wild-type allele, the detected fragment length of the non-expanded allele varied between 2370 and 2466 bp. The specificity of the hybridization probe is demonstrated by the presence of two wild-type alleles only in the controls. (d) Selected chromatograms resulting from short expansion PCR as shown in Figure 1b. The sample is indicated for each chromatogram. The wild-type allele showed a discrete peak between 224 and 272 bp, whereas the expanded allele is represented by a range of subsequent peaks with a Gaussian distribution between 446 and 794 bp.

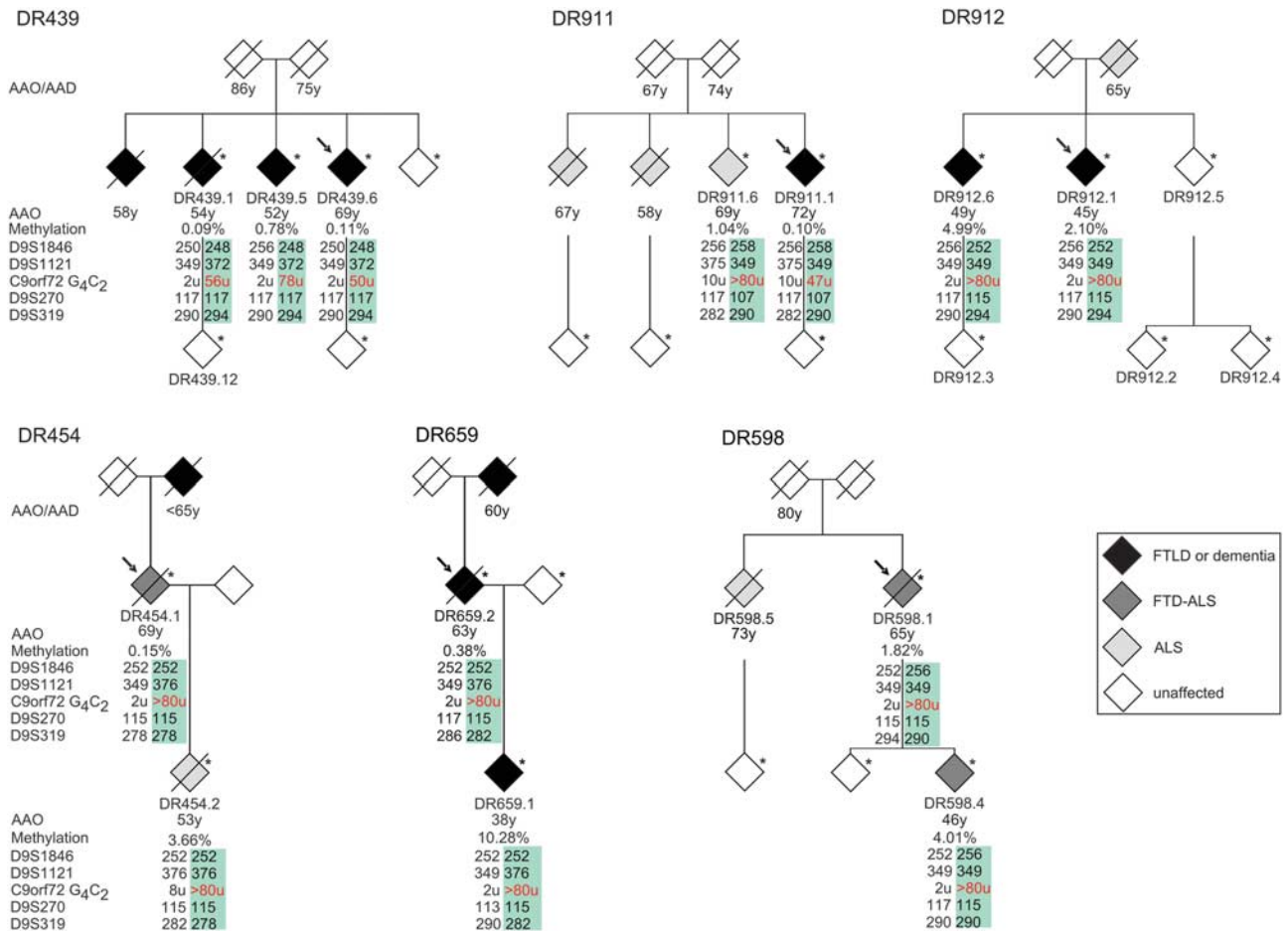


Figure 2. Segregation of short repeat expansions and possible anticipation in G₄C₂ expansion families. Three families are shown in which a short repeat expansion was detected (DR439, DR911 and DR912). Filled symbols are indicating patients with below their respective age at onset (AAO) in years (y). Age at death (AAD) is shown for individuals who died at old age without symptoms. Individuals with DNA are depicted with an asterisk right of the symbol. The C9orf72 repeat length is included in the haplotype indicated in number of units (u). The numbers in diamonds indicate the number of unaffected at-risk individuals. The disease haplotype is shown with light green bars. In families DR439 and DR911, a short repeat expansion segregates with disease on a specific disease haplotype. Clinical characteristics of presented individuals can be found in Table 1 and Supplementary Table 2. Four parent–offspring pairs (DR439, DR454, DR659 and DR598) presented with evidence for anticipation, including methylation (delta Ct) increase (DR439, DR659 and DR598), age at onset decrease (DR454, DR659 and DR598) and/or repeat size increase (DR439) across two generations.

(Figure 1c), making exact repeat expansion sizing unreliable. Based on the 27 samples analyzed, a gap between short expansions (up to 80 repeat units) and expansions of > 388 repeat units (long expansions) was apparent. Short expansion PCR easily discriminated these two groups of short and long expansion carriers.

Segregation analysis of short G₄C₂ expansions and association with DPR pathology

Segregation analysis in the families of the four index patients carrying a short repeat expansion in blood demonstrated co-segregation with disease in family DR439 and DR911 (Figure 2). In family DR439, blood DNA of three sibs with FTLD carried a short expansion of 50, 56 and 78 repeat units, respectively (Figure 2, Table 1). As it was indicated above, one of them (DR439.5) also had additional bands on Southern blot. DR439.1 had a pre-symptomatic child DR439.12 aged 34 years with a long expansion of > 1100 repeat units (Figure 1c). Using short tandem repeat markers flanking C9orf72, we demonstrated that the repeat expansion alleles segregated on the same disease haplotype (Figure 2). Immunohistochemical analysis of autopsied

brain of patient DR439.1 carrying a short expansion of 56 repeat units in blood, showed TDP-43-positive neuronal cytoplasmic inclusions in the hippocampus and frontal cortex compatible with TDP-43 type B proteinopathy.⁵¹ Dot-like neuronal cytoplasmic inclusion in the frontal lobe (premotor cortex), hippocampal dentate gyrus and the granular cell layer of the cerebellum were observed after p62 and poly-GA DPR immunostaining (Figure 3 and Supplementary Figure S3). Other aggregating DPR proteins (poly-GP, poly-GR, poly-AP and poly-PR) had been observed in the frontal cortex of this patient.^{22,24} Short expansion PCR of DNA extracted from the frontal and temporal cortex of patient DR439.1 were consistent with the short expansion size determined in blood DNA, however, the weak signal on Southern blot of the short expansion allele and additional signals corresponding to larger expansion sizes indicated mosaicism in these brain areas (Figure 1). The signal of the short expansion was weakest or even absent in cerebellum in favor of a relatively strong signal corresponding to an expansion of > 1100 repeat units on Southern blot hybridization (Figure 1). Two other patients with a long expansion in blood DNA did not show a short expanded allele in brain (Figure 1). We compared the DPR/p62 and TDP-43 pathology of DR439.1 with the pathology of two patients with a

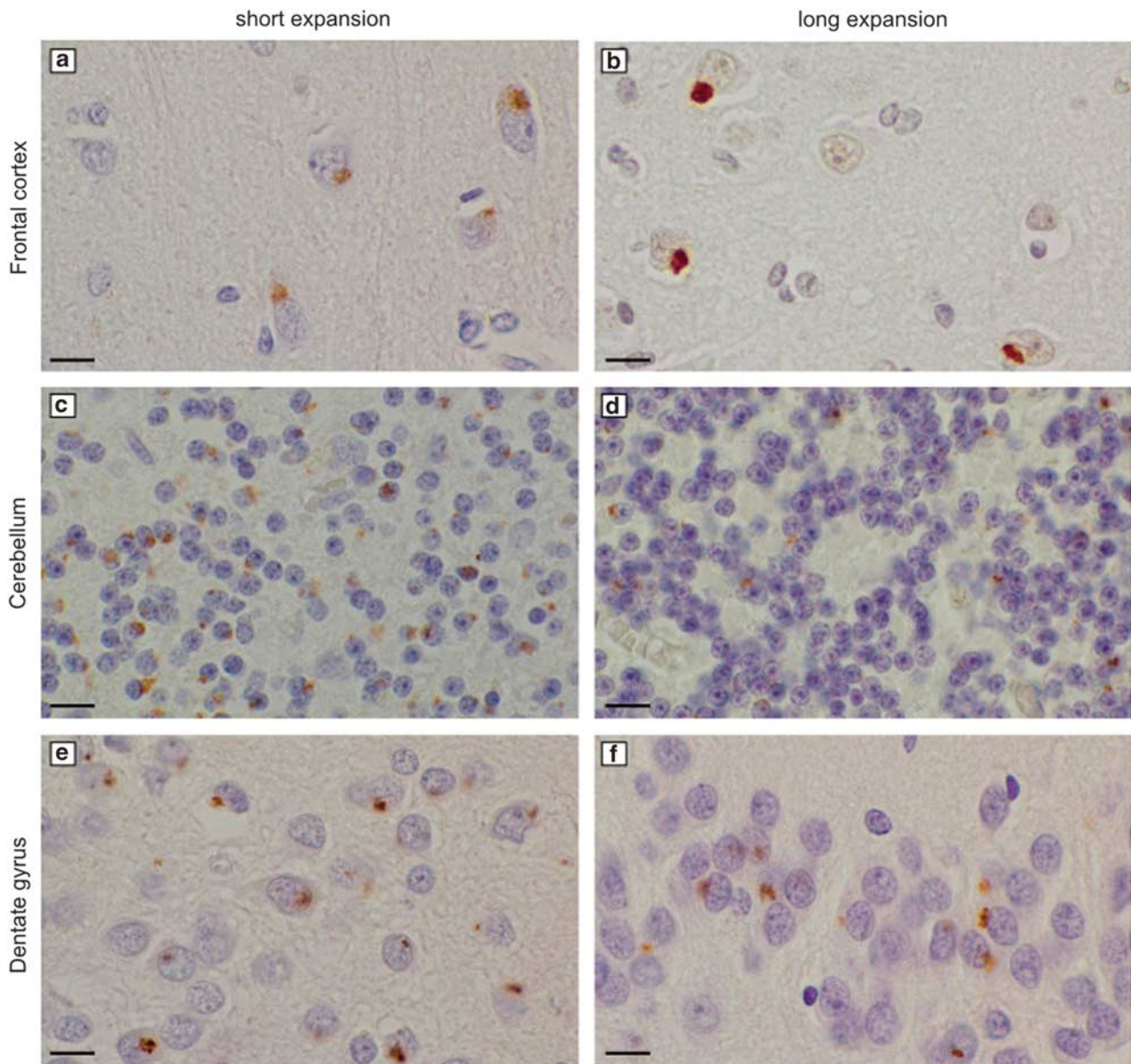


Figure 3. DPR pathology of a patient with a short expansion compared with a patient with a long expansion. Immunohistochemistry with poly-GA-specific antibodies detects DPR-positive dot-like neuronal cytoplasmic inclusions in Brodmann area 6 of the frontal cortex (**a-b**), the granular cell layer of the cerebellum (**c-d**) and the dentate gyrus of the hippocampus (**e-f**) of a patient with a short repeat expansion (56 units) (**a-c-e**), comparable with the inclusions in a patient with a long repeat expansion (> 80 units) (**b-d-f**). Scale bars denote 10 μ m.

long expansion.² Visual inspection and semiquantitative analysis of DPR/p62 and TDP-43 inclusion load in frontal cortex, cerebellum and hippocampus did not show striking differences between the carrier of a short expansion in blood with some indication of mosaicism in brain and long expansion carriers (Figure 3 and Supplementary Figure S3).

In family DR911, the index patient (DR911.1) presented with late-onset FTLN (72 years), while the three affected sibs were diagnosed with ALS and onset ages 69, 67 and 58 years (Figure 2, Table 1). The parents died at age 67 and 74 years without documented dementia or ALS symptoms. We obtained DNA of the index patient and of one affected sib (DR911.6) and showed the presence of a short expansion of 47 units and a long expansion of > 80 units, respectively (Figure 2). One of the asymptomatic at-risk individuals in the youngest generation carried a repeat expansion of > 80 units. Repeat expansions in the other unaffected children

were excluded. Haplotype analysis confirmed that the expansions segregated on the same disease haplotype (Figure 2).

In family DR912, there were two affected sibs with early-onset FTLN (DR912.1, 45 years and DR912.6, 49 years). One of the parents died from ALS at age 66 years after 1 year of disease (Figure 2). We obtained DNA of the two affected sibs (DR912.1 and DR912.6) and four at-risk relatives, all carrying a repeat expansion. Only the asymptomatic carrier DR912.3 showed a short expansion of 45 units, and the short and the long expansions were located on the same haplotype (Figure 2).

Onset age and DNA methylation are associated with G₄C₂ expansion size

When we compared onset age differences between 6 patients with a short expansion and 51 patients with a long expansion of

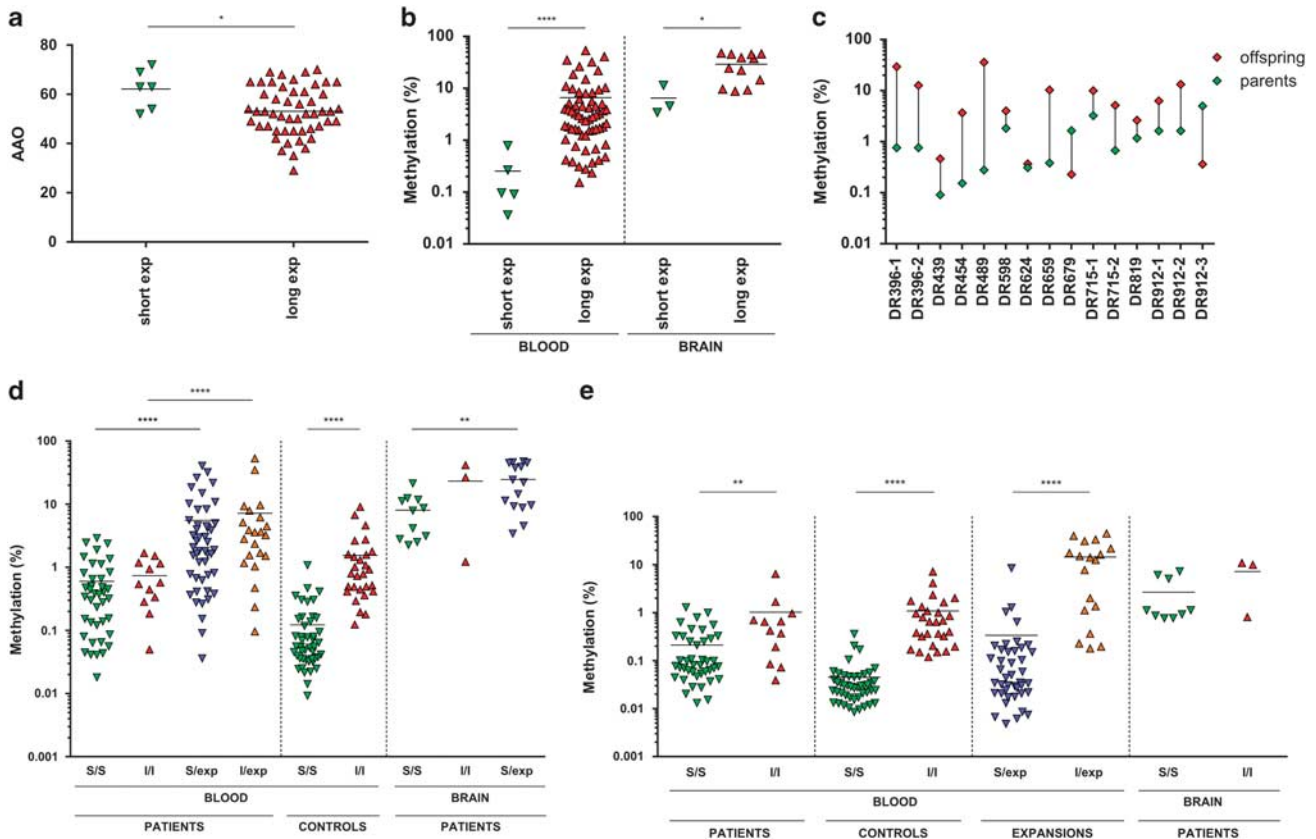


Figure 4. Association of G_4C_2 size with disease onset age and DNA methylation of the G_4C_2 repeat and the 5' flanking CpG island. (a) Comparison of age at onset (AAO) between patients with a short expansion (< 80 units) and patients with a long expansion (> 80 units). (b) Comparison of the methylation level of the 5' flanking CpG island between patients with a short expansion (< 80 units) and patients with a long expansion (> 80 units) in DNA from blood and brain. Brain regions include frontal cortex, temporal cortex and cerebellum. (c) Methylation differences of the 5' flanking CpG island in 15 parent-offspring pairs of 11 families. Methylation levels are shown in red for the parents and in green for the offspring. Clinical characteristics of presented individuals can be found in Supplementary Table 2. (d) HhaI MSRE-qPCR results of the 5' flanking CpG island are presented for expansion carriers (S/exp and I/exp) versus patients without expansion with short/short (S/S) and intermediate/intermediate (I/I) genotype stratified for normal short (S) or intermediate (I) repeat length of the normal alleles (a), and for controls with S/S and I/I genotype (a). (e) HpaII MSRE-qPCR results of the G_4C_2 repeat are shown for patients and controls with S/S and I/I genotype and for expansion carriers with a normal short or intermediate wild-type allele. Disease status and genotype are indicated on the X axis, whereas the Y axis shows the methylation ratio (in %) for each sample. The mean is represented by black horizontal bars for each subcategory of samples. The significance of differences in methylation was calculated using the Mann-Whitney U-test. P-values are presented above the bars (**** $P < 0.0001$; ** $P < 0.01$; * $P < 0.05$).

whom the age at onset was recorded (mean 53.1 years, range 29-70 years), we calculated a significantly later onset age in the short expansion carriers (mean 62.2 years, range 52-72 years; $P = 0.037$; Figure 4a). Notably, the patient DR439.5 carrying a pool of short and long expansion sizes had the earliest onset age (52 years).

As suggested in previous DNA methylation studies of G_4C_2 expansion carriers,¹¹⁻¹⁵ the qPCR assay in the 5' flanking CpG island revealed a significant increase of methylation in blood DNA ($P < 0.0001$) and brain DNA ($P = 0.0044$) of patients carrying an expansion compared with patients without expansion, normalized for the repeat size on their normal allele(s) (Figure 4d).

Further, we examined whether the observed differences in onset age between short and long expansion carriers could be explained by altered DNA methylation of the 5' flanking CpG island. Indeed, the methylation levels in the 5' flanking CpG island (Supplementary Figure S2) were significantly lower in short expansion carriers in blood DNA ($P < 0.0001$) and brain DNA ($P = 0.031$; Figure 4b). Patient DR439.5, carrying a pool of short and long expansion sizes in blood, showed the highest methylation level in blood DNA. Three different brain regions of patient

DR439.1 showing a pool of short and long expansion sizes were examined (frontal cortex, temporal cortex and cerebellum), of which the cerebellar region showed the highest methylation consistent with untraceable quantity of short expansion sizes in this brain region (Figure 1c).

Parent-offspring pairs show disease anticipation

As exact expansion sizes were lacking in most of the parent-offspring pairs and because the young age of most of the asymptomatic expansion carriers in the youngest generation, no robust conclusions regarding repeat amplification and disease anticipation could be drawn based on onset ages or expansion sizes only. In one parent-offspring pair (DR439.1 and DR439.12), an increase in expansion size with about 1000 units was observed, although the offspring, aged 34 years, was not yet affected (Figures 1 and 2; Table 1; Supplementary Table 2). Three other affected parent-offspring pairs in families DR454, DR659 and DR598 with DNA available (Figure 2), showed a 16, 25 and 19 years earlier onset age, respectively, in the affected child compared with the parent (Figure 2, Supplementary Table 2). In each of

these pairs, short expansion PCR excluded a short expansion (>80 units).

As a result of the correlation between expansion size, onset age and methylation, we analyzed methylation differences across generations in 15 parent–offspring pairs of 11 families (Supplementary Table 2). A methylation increase up to 35.6% was apparent in 13 pairs including DR439, DR454, DR659 and DR598 ($P=0.0034$) (Figure 4c; Supplementary Table 2), supporting an intergenerational repeat amplification as shown in DR439 and possibly resulting in an earlier onset age as shown in DR454, DR659 and DR598.

Epigenetic analysis of the C9orf72 promoter in intermediate repeat carriers

We previously suggested that intermediate repeats (7–24 units) might act as risk alleles predisposing the carriers to G₄C₂ expansions.²⁵ In a follow-up study, we observed in a group of ALS and FTL–ALS patients ($N=135$) that homozygous carriers of intermediate repeat alleles showed a significantly increased risk ($P=0.038$; 95% confidence interval = 2.075 (1.041–4.137)). Therefore, we compared the 5' CpG island methylation of homozygous normal intermediate repeat carriers (I/I) (7–24 units) with homozygous normal short repeat carriers (S/S) (2–6 units) and demonstrated a slight but significantly higher methylation in I/I carriers in controls ($P < 0.0001$) but not in patients ($P=0.1471$; Figure 4d).

In addition, we compared methylation states of normal short and intermediate G₄C₂ repeats itself using a G₄C₂ MSRE-qPCR assay. We showed a slight but significant increase in methylation level of the G₄C₂ repeat in blood DNA from I/I carriers compared with S/S carriers, in the patient ($P=0.005$) and control groups ($P < 0.0001$; Figure 4e). Also intermediate wild-type alleles of expansion carriers are significantly more methylated than normal short wild-type alleles ($P < 0.0001$; Figure 4e).

Notably, the degree of methylation is very low (<5%) in the groups of individuals without long expansions versus ~10% in expansion carriers. Sequencing of bisulfite-treated DNA samples detected 5' CpG methylation in long expansion carriers with at least 5% methylation and no 5' CpG methylation in short expansion carriers or normal repeat carriers (Supplementary Figure S4).

Increased DNA methylation in patients with a normal repeat compared with control persons with a normal repeat was apparent in both MSRE-qPCR assays in blood ($P=0.0133$ for the 5' CpG island and $P=0.0004$ for the G₄C₂ repeat). Also, 5' CpG and G₄C₂ methylation was significantly higher in brain than in blood of expansion carriers (Figure 4b) and patients with a normal repeat ($P < 0.0001$; Figures 4d and 4e).

Variability in repeat sizes and flanking sequences is associated with decreased C9orf72 transcriptional activity

We further investigated whether the G₄C₂ repeat size has a direct effect on the transcriptional activity of the C9orf72 promoter. In addition, as we observed a significantly higher amount of indels in the 3' flanking sequence in patients without C9orf72 expansion (7/379 = 1.85%) than in controls (4/752 = 0.53%; $P=0.033$; $N=379$), we also evaluated the effect of these indels on transcriptional activity.

In HEK293T cells, we observed a highly significant decrease of transcriptional activity of promoter fragment 1, containing G₄C₂ repeat alleles with intermediate length within the normal range (9, 14, 19 and 24 units) or within the size range of unclear pathogenicity (31 and 38 units), compared with the normal reference allele with two repeat units ($P < 0.0001$; Figure 5a). The promoter activity of the 24 units fragment gradually dropped to 57%, which is significantly lower than the 19 units fragment ($P=0.010$). However, the promoter activity of the fragments with

31 and 38 repeat units was not significantly changing (Figure 5a), possibly due to repeat instability (Supplementary Figure S5). For fragment 2, the same trend was observed, although the most significant drop in transcriptional activity was observed for the fragment with 14 repeat units compared with the fragments with 2 and 9 units ($P < 0.0001$; Figure 5c). For both fragments 1 and 2, the presence of the 5-bp deletion (6U-del) or the 23-bp deletion (12U-del), resulted in a drastic decrease in transcriptional activity compared with the reference fragment with 6 units ($P=0.0002$ for fragment 1; $P=0.0034$ for fragment 2) or 12 units ($P < 0.0001$ for fragment 1; $P=0.0004$ for fragment 2) (Figures 5b and d).

Similarly, in SH-SY5Y cells the transcriptional activity of the promoter fragment 1 was significantly decreased for all repeat sizes compared with the normal reference repeat of two units ($P < 0.0001$) with a maximal drop to 53% for the 38 units fragment (Figure 5e). Again, the 6U-del and the 12U-del in the 3' flanking sequence resulted in a significant transcriptional decrease compared with the respective reference fragment with 6 repeat units ($P=0.0002$) or 12 repeat units ($P < 0.0001$; Figure 5f).

DISCUSSION

The basis for clinical heterogeneity and suggestive anticipation among C9orf72 patients is not well understood. The effect of the repeat size on the clinical characteristics of repeat expansion disorders is well known. In our study, G₄C₂ expansion sizes in blood varied considerably between and within families (45 to >2100 units), which is largely comparable with previous reports^{1,6,26–32} and indicate a high degree of instability of the expanded repeat. We identified 5.3% carriers of a short expansion with a repeat size between 45 and 78 units in blood. We were able to amplify the expanded allele, and the short expansions could also be visualized by Southern blot hybridization without other larger bands in six of seven patients (Figure 1c). In two families, DR439 and DR911, we provided evidence that G₄C₂ expansion carriers with a repeat as short as 47 repeat units in blood can present with FTL or ALS (Table 1; Figure 2). For instance, patient DR439.1 showed numerous p62/DPR-positive neuronal cytoplasmic inclusion in different affected brain regions, a characteristic pathological hallmark of C9orf72 patients^{22–24} without remarkable differences compared with long expansion carriers. However, the short expansion in this subject was only partially apparent in the frontal and temporal cortex and absent in cerebellum. Several studies reported instability of the repeat from 16 units on due to somatic mosaicism across different tissues,^{26,30,38} which complicates determination of the repeat pathogenicity based on repeat sizing in blood-derived DNA. However, intra-individual variation of repeat number between tissues was higher than the variation within each tissue group.³⁸ Although other studies reported short expansions of <100 units,^{6,27,38,52} this is the first report demonstrating co-segregation of FTL and ALS with an expansion outside the pathogenic size range from 400 to 4400 repeat units.^{26,30} A short C9orf72 repeat expansion of 66 units causes TDP-43 pathology, neuronal loss and behavioral deficits in a recent mouse model, corroborating our findings.⁵³ Although expansions of any size can be recognized by the repeat-primed PCR assay, the detection of expansions as short as 47 units underscores the importance of the short expansion PCR assay as a fast and reliable screening method for identifying expansions up to about 80 repeat units.²⁶

Expansion carriers without a detectable expanded allele on short expansion PCR were classified as 'long expansions' (>80 units), as exact sizing of expansions in the higher size range was inappropriate because of the smeared signals because of mosaicism and the limited resolution of an agarose gel above 15 kb. In contrast to other studies,^{1,6,15,26–30,32,38} we were able to provide a significant later age of disease onset in patients with a short expansion than those with a long expansion (Figure 4a). To

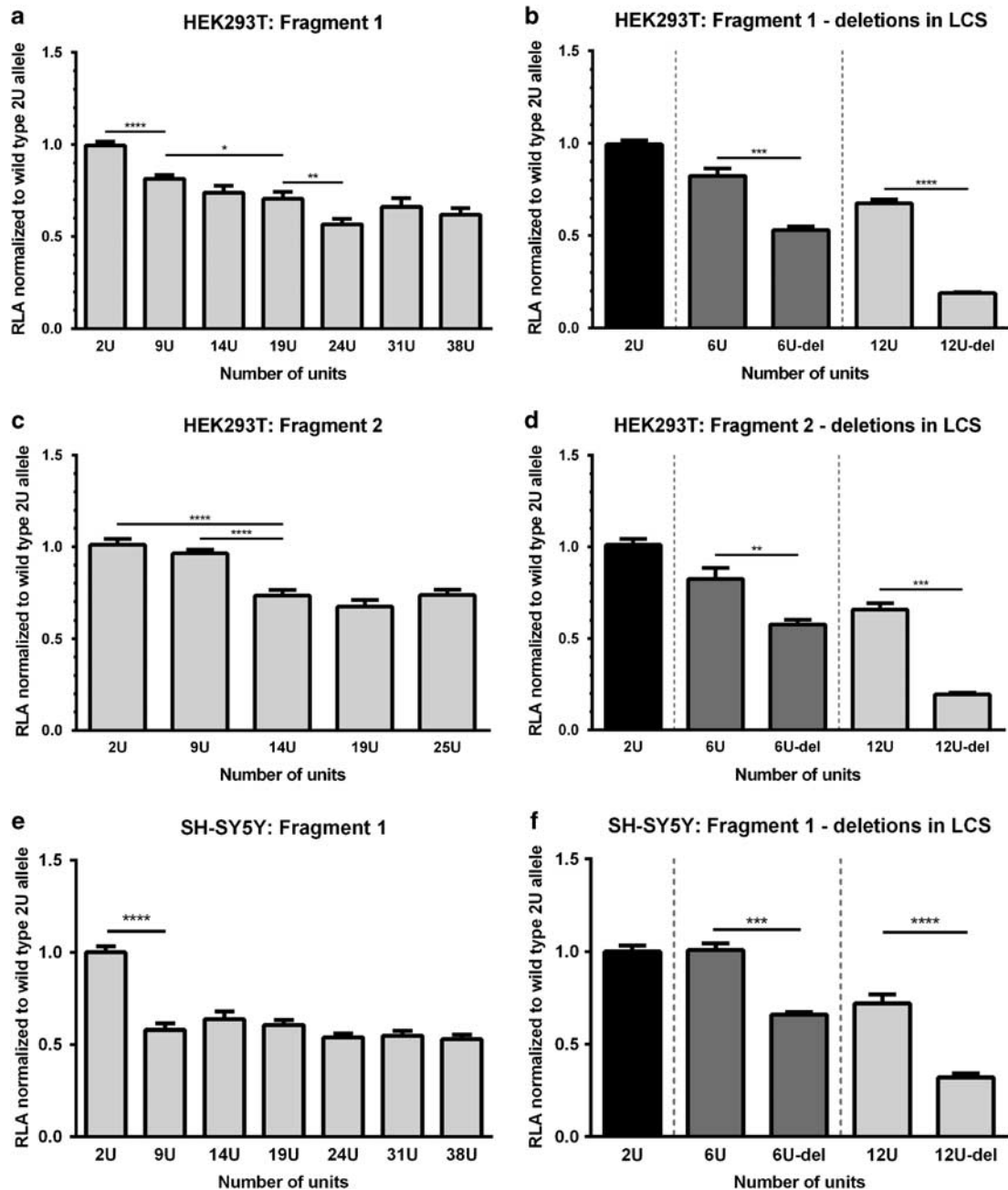


Figure 5. C9orf72 reporter gene analyses in HEK293T and SH-SY5Y cells. Bars represent relative Gaussia/Cypridina luciferase activities (RLA) for each of the C9orf72 constructs containing different G₄C₂ repeat lengths in fragment 1 (a-e) or fragment 2 (c) or different deletions in the 3' flanking sequence in fragment 1 (b-f) and fragment 2 (d), as indicated in the X axis. Values and error bars represent the mean ± S.E.M. relative to the wild-type reference allele of two units (Y axis). The significance of differences in expression was calculated using the Mann-Whitney U-test. P-values are presented above the bars (****P < 0.0001; ***P < 0.001; **P < 0.01; *P < 0.05). Significant P-values are only indicated with the first subsequent allele that is significantly different.

reach this result, we used only blood and brain-derived DNA and no DNA of cell lines. In addition, a standardized clinical diagnostic protocol ensured a consistent onset age estimate in all patients. Further, we showed that this difference in age at onset might be explained by lower methylation of the CpG island upstream of G₄C₂ in patients with short expansion sizes in blood, as well as in brain (Figure 4b). Others also found hypermethylation of the CpG island in expansion carriers¹¹⁻¹⁵ but a positive association with expansion size was never reported. Only one study found a

marginally significant association of hypermethylation with shorter repeat size in blood only.¹⁵ Further investigation and replication studies of age at onset and methylation correlations with repeat size in large cohorts are needed.

We gathered evidence that anticipation might have a role in C9orf72 families segregating a G₄C₂ expansion by comparing onset ages, expansion sizes and methylation states across generations of parent-offspring pairs. First, a parent-offspring pair of family DR439 showed an increment of about 1000 repeat

units and a methylation increase from the affected parent to the asymptomatic child (Figure 2, Supplementary Table 2). Remarkably is that the parents of the four affected sibs died at 75 and 86 years without documented dementia or ALS symptoms, potentially suggesting that they did not carry a repeat expansion. The data obtained in family DR439 suggest genetic anticipation over three generations. Other modifying factors may also influence the clinical phenotype in this family since the index patient had a later onset age (69 years) than his two affected sibs (52 and 54 years). This is in line with previously reported data of an identical twin discordant for ALS but showing the same repeat expansion size and methylation level.⁵⁴ Second, three affected parent-offspring pairs in three families (DR454, DR659 and DR598) showed a 16, 25 and 19 years earlier onset age, respectively and a higher methylation state of the flanking CpG island in the child compared with the affected parent (Figure 2). In 9 of 11 other parent-offspring pairs without repeat sizing or onset age data, a methylation increase was apparent from parent to offspring (Supplementary Table 2).

Several reports described an earlier onset age in each younger generation suggesting disease anticipation, however, not all C9orf72 families showed a decrease in onset age across generations.^{2,41–46} Also, the anticipation data obtained in the families was largely not validated by attempts to size the repeat expansion. In our study, we could also not correlate an earlier onset age with an increase in repeat size in each parent-child pair, for example, in families DR454 and DR659, because of the technical limitations of the current methods used for sizing the C9orf72 repeat expansions. Nonetheless, our data are highly supportive of genetic anticipation in C9orf72 families for several reasons. We analyzed onset age data in individual affected parent-child pairs instead of between onset age data that was pooled per generation. Also, we made optimal use of the detection of short C9orf72 repeat expansions that were segregating in families, to directly observe in individual generations the amplification of the C9orf72 repeat. Further, we showed association between expansion size, onset age and methylation state (see above). The observation of the repeat size contraction as observed in one parent-child pair of family DR912 (DR912-3; Figures 2 and 4c), and possibly other modifying factors, may explain the occurrences of onset age decrease and methylation increase (Figure 4c; Russ *et al.*¹⁵) in some families.

Genetic anticipation because of repeat instability, together with the incomplete penetrance and wide range of onset age of disease in C9orf72 carriers,⁴¹ has important consequences for genetic counseling of relatives of a C9orf72 carrier. Likely one of the two unaffected parents carries a repeat expansion, potentially a shorter one, in the absence or the presence of undetected mild symptoms hampering reasonable predictions of risk of relatives as well as of onset of the disease symptoms or clinical phenotype in carriers.^{55,56}

The emergence of two clinically and pathologically distinct diseases,⁵⁷ FTD and ALS, in the C9orf72 carriers is remarkable and might potentially be explained by modifying factors. It is possible that differences in repeat expansion sizes or in the presence of interruptions in the C9orf72 repeat sequence might influence the clinical manifestation. To date, this hypothesis cannot be adequately tested as there is no method available that allows to determine the precise size or repeat content of the expansion. The current methods, repeat-based PCR amplification or Southern blotting, provide at the best an approximate size measure of the expansion without information on the purity of the repeat sequence. Another possibility is that differences in the degree or pattern of methylation of the C9orf72 promoter region can explain the clinical divergence. In our study, we did not observe differences between FTD and ALS expansion carriers in blood (Supplementary Figure S6). However, methylation differences in other CpG islands, including the G₄C₂ repeat itself, or in other

tissues than blood cannot be excluded. Further, it is possible that genetic variation in other genes may drive the disease pathology toward a specific side of the clinical FTD/ALS continuum. Genetic modifiers for clinical expression of FTD have already been reported, including *TMEM106B*, which exerted a protective effect for FTD in C9orf72 carriers⁵⁸ and *ATXN2*, which increased risk of C9orf72 carriers for ALS.^{59,60} Also, the co-occurrence of mutations in multiple FTD and ALS genes might modify the disease phenotype (reviewed in Lattante *et al.*⁶¹). A hypothesis-free approach to find genetic modifiers of the FTD-ALS continuum could be the use of single-nucleotide polymorphism data obtained in genome-wide association studies comprising FTD and ALS patients,⁶² or the exploration of genome or exome data generated in large patient cohorts.⁶³

Further, we observed slight but significant G₄C₂ repeat methylation differences in intermediate (I) G₄C₂ risk allele carriers versus normal short (S) repeat carriers in blood (Figure 4). Although the methylation percentages of normal repeat carriers were in a very low range and not detected in previous reports using qualitative methods,^{13,14} significant differences were calculated in both independent quantitative assays measuring methylation of the 5' CpG island and of the G₄C₂ repeat itself, supporting the correctness of the data. Notably, the higher methylation levels in I/I versus S/S carriers were mostly derived from the homozygous intermediate carriers with at least one allele of >8 units. As these longer intermediate repeat alleles are relatively under-represented in the small I/I patient group (Supplementary Table 1), the methylation of the flanking CpG island was not significantly different between S/S and I/I patient groups. Methylation of pathological G₄C₂ expansions could not be studied by our method as PCR amplification is technically difficult. Nevertheless, the correlation between the methylation state of the G₄C₂ repeat and that of the 5' flanking region in normal repeat carriers suggested that hypermethylation of the G₄C₂ repeat could also have a role in carriers of G₄C₂ expansions. This assumption was recently shown by a qualitative assay.¹⁰ Although a later age at inclusion was previously associated with longer repeat sizes^{15,30} and DNA methylation levels might change with aging,⁶⁴ in this study, differences in age at inclusion could not explain the observed correlations in S/S, I/I and expansion carriers (Supplementary Figure S6). Of interest, DNA methylation was higher in patients than in controls in blood and brain suggesting a risk-increasing effect of methylation caused by other factors than repeat length. Also of note, C9orf72 methylation was remarkably higher in brain than in blood. Despite these tissue-specific methylation levels, the repeat size effect on methylation was also observed in brain, as methylation in blood and brain are positively correlated, as previously described.¹⁵

To evaluate the effect of the hypermethylated intermediate repeat sizes on C9orf72 promoter activity, we developed an *in vitro* reporter gene assay. As our preliminary results already suggested,²⁵ here we established a highly significant gradual decrease of C9orf72 promoter activity with an increasing number of G₄C₂ repeat units in human kidney and neuroblastoma cell lines (Figure 5), the latter most closely related to the affected neurons in patients. Our promoter study showed that increased methylation of CpG sequences in larger repeats and hence transcriptional silencing of the promoter is most likely one of the mechanisms of how an increasing repeat length can lead to decreased promoter activity, but other possibilities might also have a role. First, an excess of transcription factor binding sites in the lengthened G₄C₂ motif and the tight complex formation of G-quadruplexes⁶⁵ might hamper proper transcription. Second, the DNA loop formed between a potential distant promoter element and the C9orf72 promoter complex might become too large to correctly connect both components.

Further, the transcriptional decreasing effect of intermediate repeat lengths and small deletions in the 3' flanking sequence

could probably explain the observed nominally significant association of intermediate repeat alleles and indels in the 3' flanking sequence with increased disease risk in the Belgian ALS and FTLD-ALS population, possibly through subtle changes in C9orf72 expression. The 23-bp deletion in the 3' flanking sequence has a more drastic effect on decreased promoter activity than the GTGGT deletion. The GTGGT deletion joins the G₄C₂ repeat with the 3' flanking sequence thereby creating an imperfect G₄C₂ repeat with 6 units more, mimicking the effect of an intermediate 12 units repeat. Alternatively, both deletions could result in deficiency of essential core promoter elements.

Altogether, our data are in favor of a loss-of-function hypothesis. Indeed, epigenetic changes including the observed DNA hypermethylation by this and other studies^{11–15} and histone trimethylation⁹ could explain the decreased C9orf72 brain expression in repeat expansion carriers^{1,2,6–9} through transcriptional silencing, as seen in other repeat expansion disorders, for example, Fragile X syndrome.^{16,17,66} An association between hypermethylation of a CpG island flanking an expanded repeat and repeat expansion size has also been found in the first intron of the Friedreich ataxia syndrome gene *FXN*, which could possibly be correlated with reduced mRNA levels.^{18,67} Accordingly, expression of C9orf72 transcripts 2 and 3 was lower in frontal and temporal cortex of two long expansion carriers, than the C9orf72 expression of the short expansion carrier DR439.1, the latter located within the large expression range of control individuals without expansion (Supplementary Figure S7). In favor of this loss-of-function hypothesis are two C9orf72 models: a zebrafish knock-down model showing axonal degeneration of motor neurons⁷ and a *C. elegans* knock-out model displaying age-dependent paralysis and neurodegeneration of GABAergic motor neurons.⁶⁸ Alternatively, the later onset age in short repeat expansion carriers might also be in favor of an RNA toxicity or DPR protein toxicity mechanism as unstably growing repeat lengths will result in a gradually more harmful effect, as seen in other repeat expansion diseases, for example, myotonic dystrophy type 1.³⁹ This hypothesis is supported by studies identifying sense and antisense RNA foci and DPR protein aggregates formed by the expanded C9orf72 G₄C₂ repeat in human neurons of different tissues involved in FTLD and ALS *in vivo*^{1,20–24} and RNA foci in induced pluripotent stem cell-derived human neurons,^{69,70} and by the sequestration of RNA-binding proteins in the pathological deposits of repeat expansion carriers.^{19,21,71} The disease causing role of DPR protein deposits was shown in a *Drosophila* model⁷² and DPR neurotoxicity was observed in primary neurons.⁷³ Moreover, RNA foci burden in the frontal cortex showed a significant inverse correlation with onset age⁷⁴ and repeat length⁷² corroborating our findings. Some observations might have benign side effects but, more likely, different mechanisms together are involved in the disease process. Alternatively, hypermethylation might be a rescue mechanism to prevent the formation of RNA foci¹² and might therefore be neuroprotective.^{15,75}

This study is meaningful in assessing disease risk and severity and to provide better diagnostic guidelines for genetic testing and counseling. Our data indicate that methylation might serve as a potential biomarker. Aberrant DNA methylation is becoming a promising therapeutic target in FTLD and ALS because abnormal DNA methylation might be involved in FTLD⁷⁶ and ALS.^{77,78} If RNA toxicity also has a role in the disease process, the use of antisense oligonucleotides targeting and decreasing RNA foci will be promising as a potential therapeutic approach.^{70,79,80}

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

The research was in part funded by the Belgian Science Policy Office Interuniversity program; the Flemish government supported European Initiative on Centers of Excellence in Neurodegeneration (CoEN); the Flemish government initiated Methusalem excellence program; the Flemish government initiated Flanders Impulse Program on Networks for Dementia Research (VIND); the Alzheimer Research Foundation (SAO-FRA); the Medical Foundation Queen Elisabeth (QEMF); the Research Foundation Flanders (FWO); the Agency for Innovation by Science and Technology Flanders (IWT) and the University of Antwerp Research Fund; Belgium; and the MetLife Foundation Award for Medical Research, USA. We thank the personnel of the Genetic Service Facility of the VIB Department of Molecular Genetics (<http://www.vibgeneticservicefacility.be>) and the Antwerp Biobank of the Institute Born-Bunge for their expert support. The FWO provided a postdoctoral fellowship to IG. DE received funding from Helmholtz Young Investigator program (HZ-NG-607) and the European Community's Health Seventh Framework Programme FP7/2013–2018 under grant agreement no. 617198 (DPR-MODELS). This work was in part supported by the W Garfield Weston Foundation (ER).

ROLE OF THE FUNDING SOURCE

The sponsors of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

BELNEU CONSORTIUM

The BELNEU consortium coordinated by Christine Van Broeckhoven: Dirk Nuytten (Hospital Network Antwerp, Antwerp, Belgium); Tim Van Langenhove, Katrien Smets, Jonathan Baets (Antwerp University Hospital, Edegem, Belgium); Rik Vandenberghe, Mathieu Vandenbulcke, Wim Robberecht, Philip Van Damme (University Hospitals Leuven Gasthuisberg, Leuven, Belgium); Patrick Santens, Bart Dermout (University Hospital Ghent, Ghent, Belgium); Olivier Deryck, Bruno Bergmans (AZ Sint-Jan Brugge, Brugge, Belgium); Alex Michotte, Jan Versijpt (University Hospital Brussels, Brussels, Belgium); Christiana Willems (Jessa Hospital, Hasselt, Belgium); Eric Salmon (University of Liège, Liège, Belgium).

REFERENCES

- 1 DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ *et al*. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* 2011; **72**: 245–256.
- 2 Gijssels I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G *et al*. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol* 2012; **11**: 54–65.
- 3 Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR *et al*. A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* 2011; **72**: 257–268.
- 4 Farg MA, Sundaramoorthy V, Sultana JM, Yang S, Atkinson RA, Levina V *et al*. C9ORF72, implicated in amyotrophic lateral sclerosis and frontotemporal dementia, regulates endosomal trafficking. *Hum Mol Genet* 2014; **23**: 3579–3595.
- 5 Zhang YJ, Jansen-West K, Xu YF, Gendron TF, Bieniek KF, Lin WL *et al*. Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol* 2014; **128**: 505–524.
- 6 Waite AJ, Baumer D, East S, Neal J, Morris HR, Ansorge O *et al*. Reduced C9orf72 protein levels in frontal cortex of amyotrophic lateral sclerosis and frontotemporal degeneration brain with the C9ORF72 hexanucleotide repeat expansion. *Neurobiol Aging* 2014; **35**: 1779.
- 7 Ciura S, Lattante S, Le BI, Latouche M, Tostivint H, Brice A *et al*. Loss of function of C9orf72 causes motor deficits in a zebrafish model of Amyotrophic Lateral Sclerosis. *Ann Neurol* 2013; **74**: 180–187.
- 8 Fratta P, Poulter M, Lashley T, Rohrer JD, Polke JM, Beck J *et al*. Homozygosity for the C9orf72 GGGGCC repeat expansion in frontotemporal dementia. *Acta Neuropathol* 2013; **126**: 401–409.
- 9 Belzil VV, Bauer PO, Prudencio M, Gendron TF, Stetler CT, Yan IK *et al*. Reduced C9orf72 gene expression in c9FTD/ALS is caused by histone trimethylation, an epigenetic event detectable in blood. *Acta Neuropathol* 2013; **126**: 895–905.

- 10 Xi Z, Zhang M, Bruni AC, Maletta RG, Colao R, Fratta P *et al*. The C9orf72 repeat expansion itself is methylated in ALS and FTLD patients. *Acta Neuropathol* 2015; **129**: 715–727.
- 11 Belzil VV, Bauer PO, Gendron TF, Murray ME, Dickson D, Petrucelli L. Characterization of DNA hypermethylation in the cerebellum of c9FTD/ALS patients. *Brain Res* 2014; **1584**: 15–21.
- 12 Liu EY, Russ J, Wu K, Neal D, Suh E, McNally AG *et al*. C9orf72 hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. *Acta Neuropathol* 2014; **128**: 525–541.
- 13 Xi Z, Rainero I, Rubino E, Pinessi L, Bruni AC, Maletta RG *et al*. Hypermethylation of the CpG island near the C9orf72 G4C2-repeat expansion in FTLD patients. *Hum Mol Genet* 2014; **23**: 5630–5637.
- 14 Xi Z, Zinman L, Moreno D, Schymick J, Liang Y, Sato C *et al*. Hypermethylation of the CpG island near the G4C2 repeat in ALS with a C9orf72 expansion. *Am J Hum Genet* 2013; **92**: 981–989.
- 15 Russ J, Liu EY, Wu K, Neal D, Suh E, Irwin DJ *et al*. Hypermethylation of repeat expanded C9orf72 is a clinical and molecular disease modifier. *Acta Neuropathol* 2014; **129**: 39–52.
- 16 Oberle I, Rousseau F, Heitz D, Kretz C, Devys D, Hanauer A *et al*. Instability of a 550-base pair DNA segment and abnormal methylation in fragile X syndrome. *Science* 1991; **252**: 1097–1102.
- 17 Knight SJ, Flannery AV, Hirst MC, Campbell L, Christodoulou Z, Phelps SR *et al*. Trinucleotide repeat amplification and hypermethylation of a CpG island in FRAXE mental retardation. *Cell* 1993; **74**: 127–134.
- 18 Castaldo I, Pinelli M, Monticelli A, Acquaviva F, Giacchetti M, Filla A *et al*. DNA methylation in intron 1 of the frataxin gene is related to GAA repeat length and age of onset in Friedreich ataxia patients. *J Med Genet* 2008; **45**: 808–812.
- 19 Xu Z, Poidevin M, Li X, Li Y, Shu L, Nelson DL *et al*. Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc Natl Acad Sci USA* 2013; **110**: 7778–7783.
- 20 Gendron TF, Bieniek KF, Zhang YJ, Jansen-West K, Ash PE, Caulfield T *et al*. Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol* 2013; **126**: 829–844.
- 21 Lee YB, Chen HJ, Peres JN, Gomez-Deza J, Attig J, Stalekar M *et al*. Hexanucleotide repeats in ALS/FTD form length-dependent RNA foci, sequester RNA binding proteins, and are neurotoxic. *Cell Rep* 2013; **5**: 1178–1186.
- 22 Mori K, Weng SM, Arzberger T, May S, Rentzsch K, Kremmer E *et al*. The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS. *Science* 2013; **339**: 1335–1338.
- 23 Ash PE, Bieniek KF, Gendron TF, Caulfield T, Lin WL, DeJesus-Hernandez M *et al*. Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* 2013; **77**: 639–646.
- 24 Mori K, Arzberger T, Grasser FA, Gijssels I, May S, Rentzsch K *et al*. Bidirectional transcripts of the expanded C9orf72 hexanucleotide repeat are translated into aggregating dipeptide repeat proteins. *Acta Neuropathol* 2013; **126**: 881–893.
- 25 van der Zee J, Gijssels I, Dillen L, Van Langenhove T, Theuns J, Engelborghs S *et al*. A pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability, and intermediate repeats. *Hum Mutat* 2013; **34**: 363–373.
- 26 Beck J, Poulter M, Hensman D, Rohrer JD, Mahoney CJ, Adamson G *et al*. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am J Hum Genet* 2013; **92**: 345–353.
- 27 Dobson-Stone C, Hallupp M, Loy CT, Thompson EM, Haan E, Sue CM *et al*. C9ORF72 repeat expansion in Australian and Spanish frontotemporal dementia patients. *PLoS One* 2013; **8**: e56899.
- 28 Ishiura H, Takahashi Y, Mitsui J, Yoshida S, Kihira T, Kokubo Y *et al*. C9ORF72 repeat expansion in amyotrophic lateral sclerosis in the Kii peninsula of Japan. *Arch Neurol* 2012; **69**: 1154–1158.
- 29 Buchman VL, Cooper-Knock J, Connor-Robson N, Higginbottom A, Kirby J, Razinskaya OD *et al*. Simultaneous and independent detection of C9ORF72 alleles with low and high number of GGGGCC repeats using an optimised protocol of Southern blot hybridisation. *Mol Neurodegener* 2013; **8**: 12.
- 30 van Blitterswijk M, DeJesus-Hernandez M, Niemantsverdriet E, Murray ME, Heckman MG, Diehl NN *et al*. Association between repeat sizes and clinical and pathological characteristics in carriers of C9ORF72 repeat expansions (Xpansize-72): a cross-sectional cohort study. *Lancet Neurol* 2013; **12**: 978–988.
- 31 Dols-Icardo O, Garcia-Redondo A, Rojas-Garcia R, Sanchez-Valle R, Noguera A, Gomez-Tortosa E *et al*. Characterization of the repeat expansion size in C9orf72 in amyotrophic lateral sclerosis and frontotemporal dementia. *Hum Mol Genet* 2013; **23**: 749–754.
- 32 Hubers A, Marroquin N, Schmoll B, Vielhaber S, Just M, Mayer B *et al*. Polymerase chain reaction and Southern blot-based analysis of the C9orf72 hexanucleotide repeat in different motor neuron diseases. *Neurobiol Aging* 2014; **35**: 1214–1216.
- 33 Dobson-Stone C, Hallupp M, Bartley L, Shepherd CE, Halliday GM, Schofield PR *et al*. C9ORF72 repeat expansion in clinical and neuropathologic frontotemporal dementia cohorts. *Neurology* 2012; **79**: 995–1001.
- 34 Garcia-Redondo A, Dols-Icardo O, Rojas-Garcia R, Esteban-Perez J, Cordero-Vazquez P, Munoz-Blanco JL *et al*. Analysis of the C9orf72 gene in patients with amyotrophic lateral sclerosis in Spain and different populations worldwide. *Hum Mutat* 2013; **34**: 79–82.
- 35 Ratti A, Corrado L, Castellotti B, Del BR, Fogh I, Cereda C *et al*. C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect. *Neurobiol Aging* 2012; **33**: 2528–14.
- 36 Simon-Sanchez J, Doppler EG, Cohn-Hokke PE, Hukema RK, Nicolaou N, Seelaar H *et al*. The clinical and pathological phenotype of C9orf72 hexanucleotide repeat expansions. *Brain* 2012; **135**: 723–735.
- 37 Xi Z, Zinman L, Grinberg Y, Moreno D, Sato C, Bilbao JM *et al*. Investigation of c9orf72 in 4 neurodegenerative disorders. *Arch Neurol* 2012; **69**: 1583–1590.
- 38 Nordin A, Akimoto C, Wuolikainen A, Alstermark H, Jonsson P, Birve A *et al*. Extensive size variability of the GGGGCC expansion in C9orf72 in both neuronal and non-neuronal tissues in 18 patients with ALS or FTD. *Hum Mol Genet* 2015; **24**: 3133–3142.
- 39 Harley HG, Rundle SA, MacMillan JC, Myring J, Brook JD, Crow S *et al*. Size of the unstable CTG repeat sequence in relation to phenotype and parental transmission in myotonic dystrophy. *Am J Hum Genet* 1993; **52**: 1164–1174.
- 40 Filla A, De MG, Cavalcanti F, Pianese L, Monticelli A, Campanella G *et al*. The relationship between trinucleotide (GAA) repeat length and clinical features in Friedreich ataxia. *Am J Hum Genet* 1996; **59**: 554–560.
- 41 Van Langenhove T, van der Zee J, Gijssels I, Engelborghs S, Vandenberghe R, Vandenberghe M *et al*. Distinct clinical characteristics of C9orf72 expansion carriers compared with GRN, MAPT, and nonmutation carriers in a Flanders-Belgian FTLD cohort. *JAMA Neurol* 2013; **70**: 365–373.
- 42 Chio A, Borghero G, Restagno G, Mora G, Drepper C, Traynor BJ *et al*. Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72. *Brain* 2012; **135**: 784–793.
- 43 Hsiung GY, DeJesus-Hernandez M, Feldman HH, Sengdy P, Bouchard-Kerr P, Dwosh E *et al*. Clinical and pathological features of familial frontotemporal dementia caused by C9ORF72 mutation on chromosome 9p. *Brain* 2012; **135**: 709–722.
- 44 Stewart H, Rutherford NJ, Briemberg H, Krieger C, Cashman N, Fabros M *et al*. Clinical and pathological features of amyotrophic lateral sclerosis caused by mutation in the C9ORF72 gene on chromosome 9p. *Acta Neuropathol* 2012; **123**: 409–417.
- 45 Boeve BF, Boylan KB, Graff-Radford NR, DeJesus-Hernandez M, Knopman DS, Pedraza O *et al*. Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72. *Brain* 2012; **135**: 765–783.
- 46 Arighi A, Fumagalli GG, Jacini F, Fenoglio C, Ghezzi L, Pietroboni AM *et al*. Early onset behavioral variant frontotemporal dementia due to the C9ORF72 hexanucleotide repeat expansion: psychiatric clinical presentations. *J Alzheimers Dis* 2012; **31**: 447–452.
- 47 Gorno-Tempini ML, Hillis AE, Weintraub S, Kertesz A, Mendez M, Cappa SF *et al*. Classification of primary progressive aphasia and its variants. *Neurology* 2011; **76**: 1006–1014.
- 48 Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J *et al*. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain* 2011; **134**: 2456–2477.
- 49 Brooks BR, Miller RG, Swash M, Munsat TL. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. *Amyotroph Lateral Scler Other Motor Neuron Disord* 2000; **1**: 293–299.
- 50 deCarvalho M, Dengler R, Eisen A, England JD, Kaji R, Kimura J *et al*. Electrodiagnostic criteria for diagnosis of ALS. *Clin Neurophysiol* 2008; **119**: 497–503.
- 51 Mackenzie IR, Neumann M, Baborie A, Sampathu DM, Du PD, Jaros E *et al*. A harmonized classification system for FTLT-TDP pathology. *Acta Neuropathol* 2011; **122**: 111–113.
- 52 Fratta P, Polke JM, Newcombe J, Mizielinska S, Lashley T, Poulter M *et al*. Screening a UK amyotrophic lateral sclerosis cohort provides evidence of multiple origins of the C9orf72 expansion. *Neurobiol Aging* 2015; **36**: 546–547.
- 53 Chew J, Gendron TF, Prudencio M, Sasaguri H, Zhang YJ, Castanedes-Casey M *et al*. Neurodegeneration. C9ORF72 repeat expansions in mice cause TDP-43 pathology, neuronal loss, and behavioral deficits. *Science* 2015; **348**: 1151–1154.
- 54 Xi Z, Yunusova Y, van BM, Dib S, Ghani M, Moreno D *et al*. Identical twins with the C9orf72 repeat expansion are discordant for ALS. *Neurology* 2014; **83**: 1476–1478.
- 55 Cruts M, Engelborghs S, van der Zee J, Van Broeckhoven C. C9orf72-related amyotrophic lateral sclerosis and frontotemporal dementia. *GeneReviews* 2015. Online bookchapter <http://www.ncbi.nlm.nih.gov/books/NBK268647/>.

- 56 Fong JC, Karydas AM, Goldman JS. Genetic counseling for FTD/ALS caused by the C9ORF72 hexanucleotide expansion. *Alzheimers Res Ther* 2012; **4**: 27.
- 57 Mackenzie IR, Arzberger T, Kremmer E, Troost D, Lorenzl S, Mori K et al. Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations. *Acta Neuropathol* 2013; **126**: 859–879.
- 58 van Blitterswijk M, Mullen B, Nicholson AM, Bieniek KF, Heckman MG, Baker MC et al. TMEM106B protects C9ORF72 expansion carriers against frontotemporal dementia. *Acta Neuropathol* 2014; **127**: 397–406.
- 59 van Blitterswijk M, Mullen B, Heckman MG, Baker MC, DeJesus-Hernandez M, Brown PH et al. Ataxin-2 as potential disease modifier in C9ORF72 expansion carriers. *Neurobiol Aging* 2014; **35**: 2421–2427.
- 60 Lattante S, Millicamps S, Stevanin G, Rivaud-Pechoux S, Moigneu C, Camuzat A et al. Contribution of ATXN2 intermediary polyQ expansions in a spectrum of neurodegenerative disorders. *Neurology* 2014; **83**: 990–995.
- 61 Lattante S, Ciura S, Rouleau GA, Kabashi E. Defining the genetic connection linking amyotrophic lateral sclerosis (ALS) with frontotemporal dementia (FTD). *Trends Genet* 2015; **31**: 263–273.
- 62 Ferrari R, Hernandez DG, Nalls MA, Rohrer JD, Ramasamy A, Kwok JB et al. Frontotemporal dementia and its subtypes: a genome-wide association study. *Lancet Neurol* 2014; **13**: 686–699.
- 63 Cirulli ET, Lasseigne BN, Petrovski S, Sapp PC, Dion PA, Leblond CS et al. Exome sequencing in amyotrophic lateral sclerosis identifies risk genes and pathways. *Science* 2015; **347**: 1436–1441.
- 64 Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S et al. Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum Mol Genet* 2011; **20**: 1164–1172.
- 65 Fratta P, Mizielinska S, Nicoll AJ, Zloh M, Fisher EM, Parkinson G et al. C9orf72 hexanucleotide repeat associated with amyotrophic lateral sclerosis and frontotemporal dementia forms RNA G-quadruplexes. *Sci Rep* 2012; **2**: 1016.
- 66 Sutcliffe JS, Nelson DL, Zhang F, Pieretti M, Caskey CT, Saxe D et al. DNA methylation represses FMR-1 transcription in fragile X syndrome. *Hum Mol Genet* 1992; **1**: 397–400.
- 67 Evans-Galea MV, Carroddus N, Rowley SM, Corben LA, Tai G, Saffery R et al. FXN methylation predicts expression and clinical outcome in Friedreich ataxia. *Ann Neurol* 2012; **71**: 487–497.
- 68 Therrien M, Rouleau GA, Dion PA, Parker JA. Deletion of C9ORF72 results in motor neuron degeneration and stress sensitivity in *C. elegans*. *PLoS One* 2013; **8**: e83450.
- 69 Almeida S, Gascon E, Tran H, Chou HJ, Gendron TF, Degroot S et al. Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol* 2013; **126**: 385–399.
- 70 Sareen D, O'Rourke JG, Meera P, Muhammad AK, Grant S, Simpkinson M et al. Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci Transl Med* 2013; **5**: 208ra149.
- 71 Mori K, Lammich S, Mackenzie IR, Forne I, Zilow S, Kretschmar H et al. hnRNP A3 binds to GGGGCC repeats and is a constituent of p62-positive/TDP43-negative inclusions in the hippocampus of patients with C9orf72 mutations. *Acta Neuropathol* 2013; **125**: 413–423.
- 72 Mizielinska S, Gronke S, Niccoli T, Ridler CE, Clayton EL, Devoy A et al. C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* 2014; **345**: 1192–1194.
- 73 May S, Hornburg D, Schludi MH, Arzberger T, Rentzsch K, Schwenk BM et al. C9orf72 FTL/ALS-associated Gly-Ala dipeptide repeat proteins cause neuronal toxicity and Unc119 sequestration. *Acta Neuropathol* 2014; **128**: 485–503.
- 74 Mizielinska S, Lashley T, Norona FE, Clayton EL, Ridler CE, Fratta P et al. C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci. *Acta Neuropathol* 2013; **126**: 845–857.
- 75 McMillan CT, Russ J, Wood EM, Irwin DJ, Grossman M, McCluskey L et al. C9orf72 promoter hypermethylation is neuroprotective: neuroimaging and neuropathologic evidence. *Neurology* 2015; **84**: 1622–1630.
- 76 Banzhaf-Strathmann J, Claus R, Mücke O, Rentzsch K, van der Zee J, Engelborghs S et al. Promoter DNA methylation regulates progranulin expression and is altered in FTL. *Acta Neuropathol* 2013; **1**: 16.
- 77 Chestnut BA, Chang Q, Price A, Lesuisse C, Wong M, Martin LJ. Epigenetic regulation of motor neuron cell death through DNA methylation. *J Neurosci* 2011; **31**: 16619–16636.
- 78 Morahan JM, Yu B, Trent RJ, Pamphlett R. A genome-wide analysis of brain DNA methylation identifies new candidate genes for sporadic amyotrophic lateral sclerosis. *Amyotroph Lateral Scler* 2009; **10**: 418–429.
- 79 Donnelly CJ, Zhang PW, Pham JT, Heusler AR, Mistry NA, Vidensky S et al. RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention. *Neuron* 2013; **80**: 415–428.
- 80 Lagier-Tourenne C, Baughn M, Rigo F, Sun S, Liu P, Li HR et al. Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc Natl Acad Sci USA* 2013; **110**: E4530–E4539.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)