CrossMark

ORIGINAL PAPER

Semi-automated quantification of *C9orf72* expansion size reveals inverse correlation between hexanucleotide repeat number and disease duration in frontotemporal degeneration

EunRan Suh[1] · Edward B. Lee[1,2] · Donald Neal[1] · Elisabeth M. Wood[1] ·
Jon B. Toledo[1] · Lior Rennert[3] · David J. Irwin[1,4] · Corey T. McMillan[4] ·
Bryan Krock[5] · Lauren B. Elman[6] · Leo F. McCluskey[6] · Murray Grossman[4] ·
Sharon X. Xie[3] · John Q. Trojanowski[1] · Vivianna M. Van Deerlin[1]

Abstract We investigated whether chromosome 9 open reading frame 72 hexanucleotide repeat expansion (*C9orf72* expansion) size in peripheral DNA was associated with clinical differences in frontotemporal degeneration (FTD) and amyotrophic lateral sclerosis (ALS) linked to *C9orf72* repeat expansion mutations. A novel quantification workflow was developed to measure *C9orf72* expansion size by Southern blot densitometry in a cross-sectional cohort of *C9orf72* expansion carriers with FTD ($n = 39$), ALS ($n = 33$), both ($n = 35$), or who are unaffected ($n = 21$). Multivariate linear regressions were performed to assess whether *C9orf72* expansion size from peripheral DNA was associated with clinical phenotype, age of disease onset, disease duration and age at death. Mode values of *C9orf72* expansion size were significantly shorter in FTD compared to ALS ($p = 0.0001$) but were not associated with

age at onset in either FTD or ALS. A multivariate regression model correcting for patient's age at DNA collection and disease phenotype revealed that *C9orf72* expansion size is significantly associated with shorter disease duration ($p = 0.0107$) for individuals with FTD, but not with ALS. Despite considerable somatic instability of the *C9orf72* expansion, semi-automated expansion size measurements demonstrated an inverse relationship between *C9orf72* expansion size and disease duration in patients with FTD. Our finding suggests that *C9orf72* repeat size may be a molecular disease modifier in FTD linked to hexanucleotide repeat expansion.

Keywords *C9orf72* · FTD · ALS · Southern blot · Expansion

E. R. Suh and E. B. Lee contributed equally to this work.

✉ EunRan Suh
suher@mail.med.upenn.edu

[1] Department of Pathology and Laboratory Medicine, Center for Neurodegenerative Disease Research, Perelman School of Medicine at the University of Pennsylvania, 3600 Spruce Street, Philadelphia, PA 19104-4283, USA

[2] Translational Neuropathology Research Laboratory, Department of Pathology and Laboratory Medicine, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[3] Department of Biostatistics and Epidemiology, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[4] Department of Neurology, Penn Frontotemporal Degeneration Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

[5] Department of Pathology, School of Medicine, University of Utah, Salt Lake City, UT, USA

[6] Department of Neurology, Penn ALS Center, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

Introduction

Frontotemporal degeneration (FTD) is a common cause of dementia in individuals under 65 years of age, and is characterized by a broad range of different clinical phenotypes

including progressive changes in language and/or behavior due to degeneration of the frontal and temporal cortex of the brain [21, 32]. Amyotrophic lateral sclerosis (ALS) is the most common form of motor neuron disease which is characterized by progressive paralysis and respiratory failure leading to death typically within 3 years of onset of disease [1]. Up to 50 and 10 % of individuals with FTD and ALS have a family history of a similar disorder, respectively [7, 8, 41], and both FTD and ALS can co-occur within families and within individuals [5]. The major common genetic cause of FTD and ALS was identified as a repeat expansion in *C9orf72*, located in the noncoding region of the *C9orf72* gene [12, 34]. The pathological mechanism underlying the *C9orf72* expansion remains unclear. The *C9orf72* expansion is also associated with considerable heterogeneity in clinical phenotype including Huntington disease-like disorders [22] and idiopathic Parkinson's disease [11, 23, 24]. Given the wide spectrum of clinical phenotypes associated with *C9orf72* expansions, it is important to identify the molecular modifiers of clinical disease in terms of clinical phenotype, spatial and temporal onset of disease, and pathologic differences in brain and spinal cord. Expansion size has been associated with ages of onset and at sample collection in the frontal cortex of patients with FTD [39], while other studies have observed a correlation between expansion size and ages at onset without observing a relationship between expansion size and clinical phenotype [3]. In the present study, we developed a semi-automated quantification workflow to measure *C9orf72* repeat length based on Southern blot densitometry to minimize variability associated with manual size quantification methods. Using these repeat length measurements, we observed differences in *C9orf72* expansion size in ALS versus FTD in agreement with a recent study [14], and found that hexanucleotide repeat size was associated with shorter disease duration in FTD.

Materials and methods

Study subjects

To identify cases with a *C9orf72* hexanucleotide expansion for further analysis by Southern blot, a total of 851 unrelated subjects meeting selection criteria were recruited. Autopsy cases were selected from the Center for Neurodegenerative Disease Research (CNDR) brain bank at the University of Pennsylvania (Penn) with a neuropathologic diagnosis of frontotemporal lobar degeneration with transactive response DNA-binding protein of 43 kDa (FTLD-TDP) pathology (n = 51) or ALS (n = 124), regardless of the clinical diagnosis or the presence of secondary

neuropathology. Clinical cases evaluated by a board-certified neurologist at the Penn Frontotemporal Degeneration Center or the Penn ALS Center were selected if they had a clinical diagnosis of suspected, possible, probable, or definite ALS using the El Escorial-revised criteria [6] (n = 407), ALS–FTD (n = 31), ALS with mild cognitive impairment (ALS-MCI) (n = 27), or FTD regardless of phenotypic subtype (n = 211). For the cases in which sufficient information was available, the FTD clinical phenotype [behavioral variant FTD (bvFTD), nonfluent–agrammatic primary progressive aphasia (naPPA), semantic variant PPA (svPPA), or logopenic variant PPA (lvPPA)] was determined using established clinical criteria [17, 33] but was not used for case selection. For individuals with both ALS and FTD, the initial presenting symptom was used to categorize patients into ALS versus FTD subgroups in multivariate regression analyses. This categorization was based on the presence of several cases of prolonged FTD where ALS symptoms were a late manifestation of disease, and the previous association with presenting symptoms and clinical disease progression [18]. Information about a family history of FTD, ALS or other neurodegenerative diseases was collected if available [41]. Cases with a known pathogenic mutation in *GRN, MAPT, VCP, SOD1*, or *TARDBP* were excluded. All patients participated in an informed consent procedure that was approved by an Institutional Review Board at the University of Pennsylvania.

Familial cases referred to the CNDR for genetic analysis only (i.e., not clinically assessed at FTD center/ALS center) and found to have *C9orf72* hexanucleotide expansions (n = 6 FTD, n = 2 ALS) were also included in the Southern blot analysis. In addition, family members of patients identified to have *C9orf72* expansions, whether affected (4 FTD and 1 ALS) or asymptomatic, for which DNA was available (n = 21) were included in the Southern blot analysis.

DNA for PCR and Southern blot analysis was extracted from blood, saliva, or brain tissue using commercial DNA extraction Kits (Autogen, Qiagen, or Oragene). Only samples from expansion cases with sufficient DNA concentration (>50 ng/μl) were used for Southern blot analysis (n = 128).

Genotyping of *C9orf72* hexanucleotide repeats and repeat-primed PCR analysis

A two-primer PCR assay was performed to determine the number of repeats in normal and small expansion alleles according to Dejesus-Hernandez et al. [12] with minor protocol modifications to improve the sensitivity for detection of alleles of 30–46 repeats. In brief, PCR was performed using 50 ng of DNA in a reaction containing 1X Amplitaq Gold buffer, 5 % DMSO, 1 M betaine, dNTP mixture with

7-deazaGTP instead of dGTP (0.25 mM each), 0.9 mM MgCl$_2$, FAM-labeled forward (5′-FAM-CAAGGAGGG AAACAACCGCAGCC) and reverse primers (5′-GCA GGCACCGCAACCGCAG) (1 μM each), and Amplitaq Gold polymerase 0.5 U/reaction (Life Technologies). Molecular analysis to detect the *C9orf72* expansion was performed as described by Renton et al. [35] with modifications. In brief, repeat-primed PCR was performed using 500–1000 ng of DNA in a reaction containing 200 μM each dNTP, 7.1 % DMSO (Sigma-Aldrich), 0.93 M betaine (Sigma-Aldrich), 0.18 mM deazaGTP (Roche), 0.9 mM MgCl$_2$ (Roche), 1.4 μM FAM-labeled forward primer (5′-FAM-AGTCGCTAGAGGCGAAAGC), 0.175 μM reverse repeat primer (5′-TACGCATCCCAGTTTGAGAC GGGGGCCGGGGCCGGGGCCGGGG), and 1.4 μM anchor tail reverse primer (5′-TACGCATCCCAGTTTG AGACG), and 1.125 units FastStart polymerase (Roche). Both the two-primer PCR and repeat-primed PCR were performed with touchdown PCR cycling conditions consisting of 4 min at 95 °C followed by cycles of 95 °C for 30 s, annealing starting at 70 °C for 1 min, and extension at 72 °C for 3 min, ending with a final extension step of 10 min at 72 °C. The annealing temperature was decreased in 2 °C steps as follows: 70 °C for two cycles, 68 °C for three cycles, 66 °C for four cycles, 64 °C for five cycles, 62 °C for six cycles, 60 °C for seven cycles, 58 °C for eight cycles, and 56 °C for five cycles. The ramp rate was set to 18 %. 2 μl diluted (1:10) two-primer or undiluted repeat-primed PCR products, respectively, were separated by capillary electrophoresis with 23 s injection time using POP-7 polymer and 36-cm 16-capillary array on an ABI PRISM 3130xl Genetic Analyzer (Applied Biosystems). PCR fragment size was determined by size standard Genescan ROX-500 with GeneMapper software. The two-primer method has been confirmed to detect the size expansions as large as 46 repeats; whereas the repeat-primed method can detect but not determine the size of large expansions.

Penetrance of *C9orf72*

The penetrance of *C9orf72* was estimated using the etm package [2] by a cumulative incidence function in 111 affected and 17 prodromal mutation carriers (total $n = 128$). It was calculated based on the age at onset of the affected cases and the age at the last follow-up of the subjects at risk.

C9orf72 southern blotting

Southern blotting was performed on subjects with sufficient DNA available with concentration higher than 50 ng/μl and a *C9orf72* expansion detected by repeat-primed PCR ($n = 128$) as previously described [3] with modifications.

In brief, 3 μg genomic DNA was digested at 37 °C for 16 h with AluI and DdeI (NEB), denatured at 95 °C for 5 min, and run on a 0.8 % agarose gel at 100 V for 4 h. 4 ng DIG-labeled DNA Molecular Marker II (Roche Life Science) was loaded on the first and last lanes on each gel as a size standard for the repeat size determination. DNA was transferred to a positively charged nylon membrane (GE Life Sciences) by capillary blotting and cross-linked to the membrane via UV irradiation. Filter hybridization was undertaken as recommended in the DIG Application Manual (Roche Life Science), with the supplementation of DIG Easy Hyb buffer with 100 μg/ml of denatured fragmented salmon sperm DNA. After prehybridization at 48 °C, the blot was hybridized overnight with a total of 15 ng/μl of DIG-labeled $(G_4C_2)_5$ oligonucleotide probe (Integrated DNA Technologies) per milliliter of hybridization solution. Membranes were washed in 2X standard sodium citrate (SSC) and 0.1 % sodium dodecyl sulfate (SDS) at room temperature while the shaking incubator ramped from 48 to 65 °C followed by three additional washes at 65 °C for 15 min each with fresh 2X SSC/0.1 % SDS, 0.5X SSC/0.1 % SDS and 0.2X SSC/0.1 % SDS. Anti-digoxigenin antibody (1:10,000, Roche Life Science) was used to detect the probe, which was visualized with CSPD (Roche Life Science) as a chemiluminescent substrate. Blots were exposed in a LAS-3000 Luminescent Image Analyzer (Fujifilm) for times ranging between 1 and 16 h.

Semi-automated determination of expanded *C9orf72* hexanucleotide repeat length

A novel semi-automated quantification workflow to measure *C9orf72* expansion size based on Southern blot was developed to minimize variability associated with manual size quantification methods (Supplementary Fig. 1). Hexanucleotide repeat number was estimated by densitometric analysis of the blot using ImageJ (http://imagej.nih. gov/ij/). To generate the best-fit standard curve per lane, lines representing size maker bands were drawn between the DIG molecular markers on both sides of the blot. The profile of the samples on each lane was plotted with the relationship between the distance of the migration from the well on the *X*-axis and the amplitudes on the *Y*-axis. A best-fit formula per lane was generated using the Eureqa Pro software (Nutonian Inc.) that related the distance of each DIG-labeled DNA molecular weight marker bands per lane to its corresponding distance in base pairs. Using this formula, the distance in base pairs for the measured mode, maximum, and minimum of each sample was calculated. The accuracy of sizing fragments that migrated slower than the 23 kb fragment of the molecular weight marker was limited by a high percent coefficient of variation. Therefore, for fragments greater than 23 kb, we

used 3800 hexanucleotide repeat units (23 kb) for the statistical analysis. The mode was defined as the point(s) on the Southern blot with the highest intensity or maximum amplitude, the maximum and minimum were defined as the points on the Southern blot where the intensity of the signal dropped to 10 % of the maximum signal amplitude on the high and low ends, respectively.

SNP genotyping

The *C9orf72* SNP rs3849942 was genotyped on cases with a *C9orf72* expansion using a custom multiplex genotyping panel, PANDoRA (Pan Neurodegenerative Disease-oriented Risk Allele) panel using a MassArray iPLEX system (Agena Bioscience) as previously described [29]. Data were analyzed with the MassArray Typer 4.0.5 by post-processing cluster analysis to refine real-time genotype calls.

Statistical analysis

We sought to determine whether *C9orf72* repeat length from peripheral blood was associated with various clinical parameters including clinical phenotype, age of disease onset, disease duration and age at death. Using either age at onset, disease duration or age at death as the outcome and both *C9orf72* repeat length and clinical phenotype (diagnosis) as predictors, we performed multiple linear regression models. Analyses were adjusted for age at collection to account for age-dependent changes in *C9orf72* repeat length in peripheral blood. The regression analyses require normality assumptions for age at onset, disease duration and age at death. These normality assumptions were checked and are appropriate. However, since the *C9orf72* repeat length violated normality assumptions, Mann–Whitney U test and Spearman correlations were performed to examine the relationships between *C9orf72* repeat length, tissue type, and age at collection. Since our study was exploratory rather than confirmatory, multiple testing adjustment was not performed based on the recommendation by Bender et al. [4]. All analyses were conducted using the statistical software package SAS version 9.3 (SAS Institute Inc., Cary, North Carolina) or the R statistical package, and results were visualized using GraphPad Prism Software (GraphPad, San Diego, CA). All statistical tests were two-sided. Statistical significance was set at the $p < 0.05$ level unless otherwise noted.

Results

Haplotype and penetrance of *C9orf72* expansion

The *C9orf72* expansion is known to be associated with a risk haplotype covering a 110-kb region between *MOB3B*

and *C9orf72* on chromosome 9p21.2 [37], which is partially tagged by the rs3849942 minor allele. The rs3849942 SNP was genotyped in our cohort of *C9orf72* expansion carriers which confirmed that the rs3849942 minor allele A was observed in all *C9orf72* mutation cases. The non-expanded allele size was also determined by PCR where we observed a trimodal distribution of the non-expanded hexanucleotide repeat size with lengths of 2, 5 and 8 corresponding to 58.2, 12.3 and 10.7 % of the cohort, respectively (Fig. 1a). Moreover, non-expanded *C9orf72* repeat lengths were significantly higher in individuals homozygous for the rs3849942 risk allele (Fig. 1a, $p < 0.0001$, Mann–Whitney test). The non-expanded allele repeat length size was not different between individuals presenting with ALS vs. FTD ($p = 0.81$, Mann–Whitney test), and did not correlate with age at onset ($r = -0.10$, $p = 0.32$, Spearman correlation), disease duration ($r = -0.06$, $p = 0.63$, Spearman correlation), or age at death ($r = -0.15$, $p = 0.20$, Spearman correlation). There was no correlation between the unexpanded versus expanded repeat length size ($r = 0.03$, $p = 0.76$, Spearman correlation, see Supplementary Fig. 2).

A total of 128 affected and prodromal unaffected cases with a *C9orf72* expansion were used to determine the penetrance of the mutation. The penetrance of the *C9orf72* expansion, regardless of clinical phenotype of family members, is 19.4, 50.6 and 96.1 % at 50, 57, and 72 years of age at onset, respectively (Fig. 1b) indicative of a high but incomplete disease penetrance.

A novel quantification workflow reduced the variation in determination of *C9orf72* expansion size

Southern blot analysis was performed to determine the correlation of repeat size with phenotype measures. However, due to considerable somatic instability of the *C9orf72* expansion, as evidenced by a smear pattern observed in Southern blots for the hexanucleotide repeats, accurate measurement of the *C9orf72* expansion size was difficult. Comparison of various *C9orf72* Southern blotting protocols revealed that the AluI–DdeI double digest protocol described by Beck et al., by virtue of yielding a smaller-sized repeat-containing DNA fragment, yielded somewhat better size resolution compared to other protocols (data not shown). Using this Southern blot protocol, we first tested routine densitometric analysis using DNA from the ND11836 lymphoblastoid cell line (Coriell Cell Repositories) as a *C9orf72* expansion control. ND11836 DNA reproducibly exhibits six discrete expansion fragments on Southern blot (Fig. 2a), consistent with the oligoclonal nature of lymphoblastoid cell lines. Initially, we determined the expansion size by measuring the distance of migration of the peak for each lane of the blot using MultiGauge software (Fujifilm) and expressed relative to the migration of
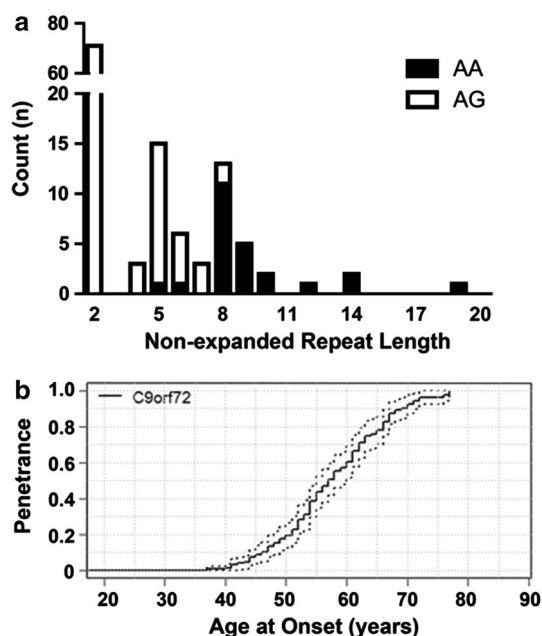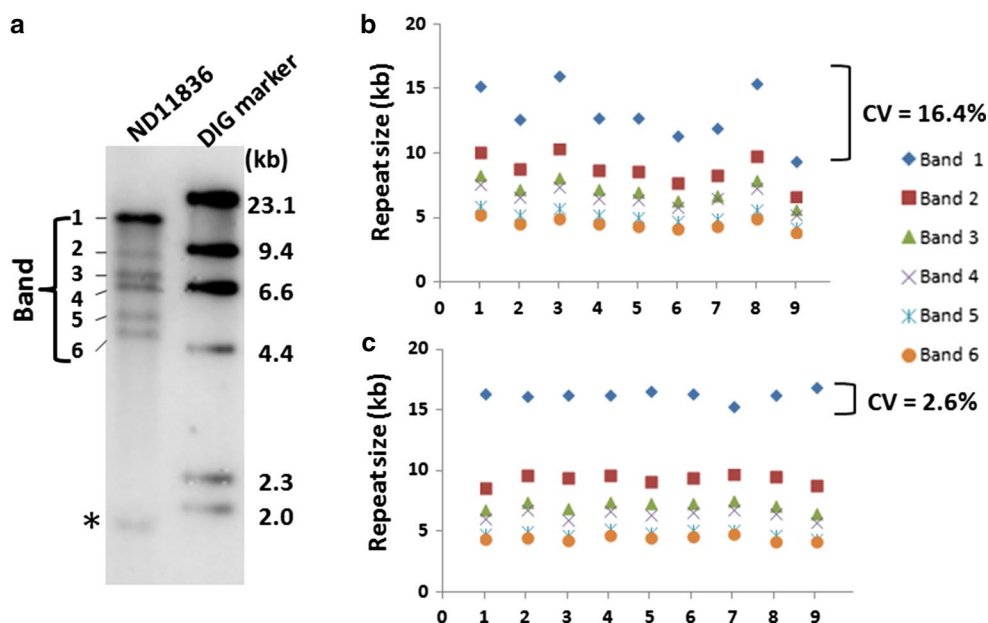
Fig. 1 **a** Frequency of repeat lengths of the non-expanded allele, broken down into individuals heterozygous (*open bars*) or homozygous (*black bars*) for the rs3849942 risk allele (*n* = 122). **b** The penetrance of *C9orf72* hexanucleotide repeat expansions in our cohort of 128 expansion carriers is shown (*solid line*) with 95 % confidence intervals (*dotted lines*)

the bands of the molecular marker on a logarithmic scale. However, this standard densitometry led to a high percent coefficient of variation (CV) in determination of the repeat size of these six distinct bands. For example, the highest band (band 1) was measured over nine independent

Southern blots yielding a CV of 16.4 % (Fig. 2b). Therefore, in an attempt to improve the accuracy of repeat size measurements, we developed a workflow that applied a best-fit formula for each lane using a combination of Image J and Eureqa Pro software. This new semi-automated quantification reduced variability; the CV associated with measuring the size of the largest band (Band 1) was reduced to 2.6 % using this method (Fig. 2c; Supplementary Table 1).

Peripheral versus central nervous system *C9orf72* repeat size

The semi-automated quantification method was used to measure *C9orf72* hexanucleotide repeat size of the expansion carriers (*n* = 128). To facilitate statistical analysis, cases were classified as ALS vs. FTD based on the individual's presenting clinical symptoms as we have previously described [18]. This cohort included 61 patients with ALS, 46 patients with FTD and 21 prodromal cases (i.e., unaffected by neurologic disease, Table 1). The *C9orf72* expansion has been reported to be shorter in cerebellum possibly due to differences in repeat length instability across cell types [39]. We confirmed that *C9orf72* expansion size in peripheral DNA (blood or saliva) was significantly higher compared to that in brain DNA (cerebellum) (Fig. 3a, b; *p* < 0.0001, Mann–Whitney test). Paired DNA samples were available from 17 cases which revealed a moderate correlation between *C9orf72* expansion size in blood vs. cerebellum (Fig. 3e, *r* = 0.60, *p* = 0.011, Spearman correlation). However, in these paired samples, *C9orf72* expansion size in the cerebellum was significantly lower than that in blood (Fig. 3c, d; *p* = 0.0003; Paired *t* test). Indeed, we

Fig. 2 *C9orf72* repeat length measurements on Southern blots as a function of replicate testing. *C9orf72* expansion band pattern of lymphoblast cell line ND11836 is shown in **a**. Six distinct *C9orf72* expansion bands on 9 independent Southern blots were quantified by **b** standard/manual and **c** customized/semi-automated best-fit formula-based densitometric methods. The *asterisk* marks a non-specific band that is seen in both non-expanded and expanded cases. Comparison of coefficient of variation (CV) between standard and semi-automated quantifications is shown in Supplementary Table 1

Table 1 Clinical characteristics of *C9orf72* expansion cohort (*n* = 128)

	ALS (*n* = 61)	FTD (*n* = 46)	Prodromal (*n* = 21)
Deaths (*n*)	48	29	0
Gender, number male (%)	38 (61.3 %)	27 (56.3 %)	7 (33.3 %)
Age at onset, mean (IQR)	56.1 (51–63)	57.5 (50–63)	–
Age at death, mean (IQR)	59.8 (55–67)	64.9 (56–75)	–
Disease duration, mean (IQR)	2.2 (1.3–3)	6.7 (3.6–9.7)	–
Peripheral DNA only, *n*	39	29	21
Cerebellum DNA only, *n*	10	12	0
Paired peripheral and cerebellum DNA, *n*	12	5	0
Peripheral *C9orf72* modal repeat number, mean (IQR)	2840 (2107–31799)	2537 (1499–3577)	2474 (1922–3504)
Cerebellum *C9orf72* modal repeat number, mean (IQR)	1645 (1246–1748)	1474 (1255–1705)	–

Two cases (one each of ALS and FTD) with *C9orf72* repeat number <40 were excluded in the multiple regression analysis
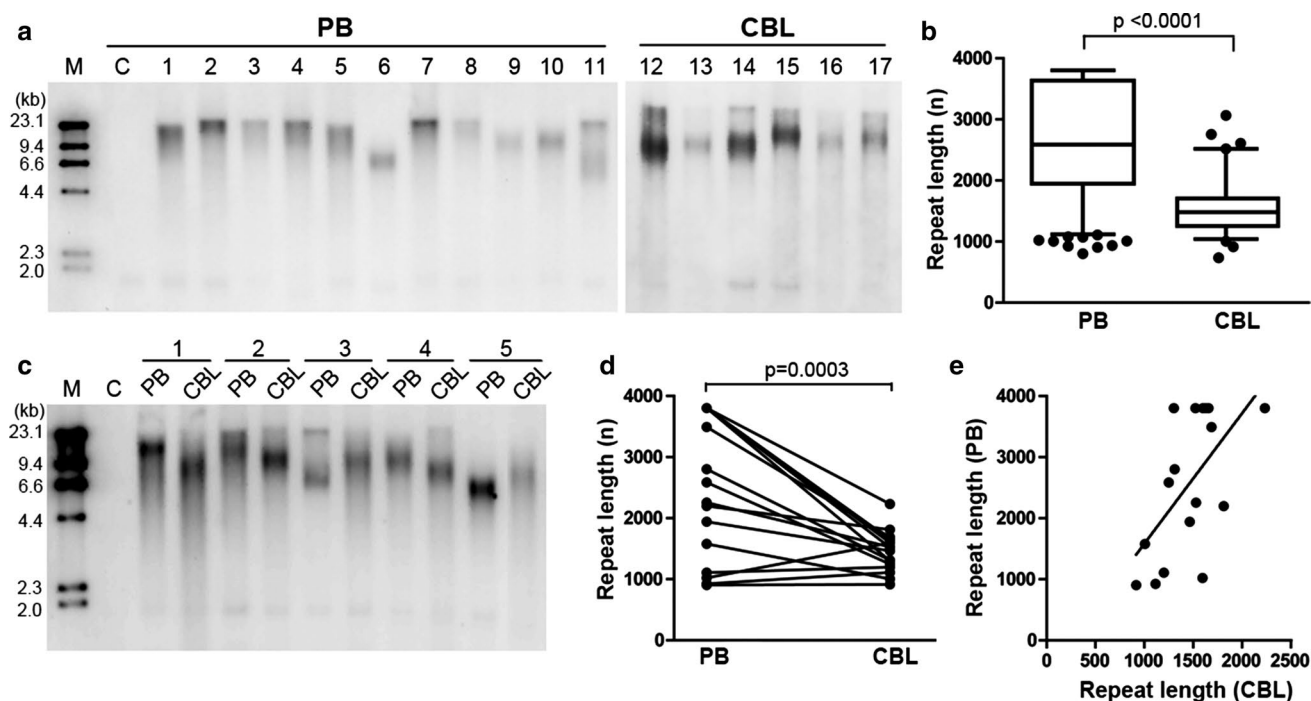

Fig. 3 *C9orf72* repeat length in peripheral blood and cerebellum. **a**, **c** Representative Southern blot images show *C9orf72* expansion size and patterns in peripheral blood and cerebellum DNA. **b** *Box-and-whisker plot* showing comparison of *C9orf72* repeat numbers between peripheral blood (*n* = 83) and cerebellum (*n* = 39) samples using Mann–Whitney test (*p* < 0.0001). *Dots* represent the data points above or below the 90th or 10th percentile, respectively. **c** Southern blot showing representative paired peripheral blood and cerebellum samples (*n* = 17). *C9orf72* repeat numbers of paired samples were compared by **d** paired *t* test with *p* = 0.0003 and by **e** Spearman correlation with correlation coefficient *r* = 0.60, *p* = 0.011. *PB* peripheral blood, *CBL* cerebellum, *M* DIG molecular marker II, *C* DNA sample from an unaffected, healthy subject homozygous for a 2-repeat *C9orf72* allele as a negative control

observed a significant correlation between *C9orf72* expansion size and age at the time of peripheral DNA collection from patients with ALS/FTD or prodromal cases (Fig. 4a, *r* = 0.32, *p* = 0.0007, Spearman correlation). The correlation between age of peripheral DNA collection and repeat length was not dependent on clinical phenotype as a similar correlation was observed in the subset of subjects with either ALS or FTD (ALS *r* = 0.4371, *p* = 0.0012; FTD *r* = 0.4338, *p* = 0.0092, Spearman correlations). Notably, there were three prodromal individuals in the cohort with relatively low *C9orf72* expansion size at age <30 (Fig. 4a). However, the correlation between age of peripheral DNA collection and *C9orf72* expansion size was similar after excluding these three individuals (*r* = 0.30, *p* = 0.0017, Spearman correlation). In contrast with peripheral DNA, there was no significant correlation between cerebellar

Fig. 4 *C9orf72* repeat length as a function of age. **a** Correlation of *C9orf72* repeat length in peripheral blood with age at collection (*n* = 50 ALS, 34 FTD, and 21 prodromal cases). **b** Correlation of *C9orf72* repeat length in cerebellum with age at autopsy (*n* = 22 ALS, 17 FTD). Correlation coefficient was calculated by Spearman analysis
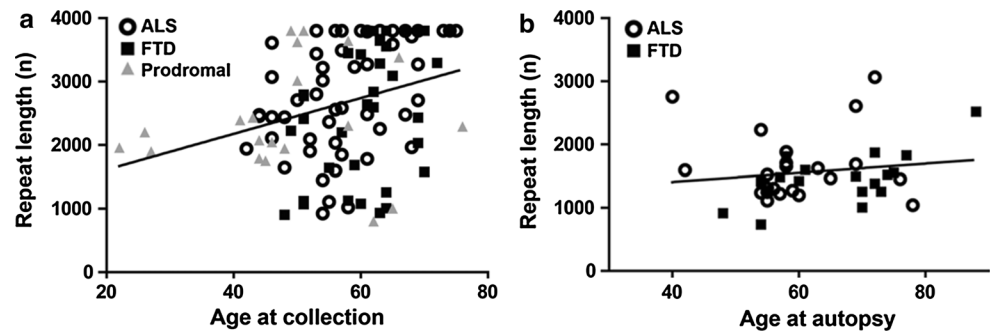

Table 2 Association of *C9orf72* expansion repeat length and age at onset (*n* = 50 ALS, 32 FTD) from multiple regression model

Variables	Regression coefficient	Standard error	95 % confidence limits		Chi-square (DF)	*p* value
Intercept	−4.17	1.93	−7.96	−0.3	4.66 (1)	0.031
Age at collection	1.0	0.034	0.93	1.07	841.46 (1)	<0.0001
Diagnosis (ALS)	2.90	0.51	1.90	3.90	32.08 (1)	<0.0001
Repeat length	−0.0002	0.0003	−0.0007	0.0004	0.38 (1)	0.54

DF degrees of freedom

C9orf72 expansion size and age at autopsy (Fig. 4b, $r = 0.19$, $p = 0.24$, Spearman correlation). These results suggest that the lower expansion size in cerebellum relative to peripheral DNA is in part due to reduced hexanucleotide repeat instability in cerebellum; that is, the expansion size appears to be more stable in cerebellum.

C9orf72 expansion size in FTD versus ALS

We sought to determine whether *C9orf72* expansion size from peripheral blood was associated with various clinical phenotypes. Given the effect of age of peripheral DNA collection on *C9orf72* expansion size, age of DNA collection was included as a covariate in all subsequent statistical analyses to account for peripheral DNA repeat instability. The patients with FTD in our cohort presented with a later age at onset, had a later age at death, and exhibited longer disease duration compared with individuals with ALS (Table 1). Multivariate linear regression analysis showed that after correcting for age of collection, *C9orf72* expansion size was significantly lower in FTD compared to ALS (diagnosis of FTD $\beta = -409.79$, $p = 0.037$; age at collection $\beta = 51.38$, $p < 0.0001$).

Age of disease onset data were available from 82 individuals with peripheral blood expansion size measurements. A multivariate regression model correcting for age of DNA collection and disease phenotype (ALS vs. FTD) showed no association between *C9orf72* expansion size and age at onset ($p = 0.54$, Table 2). However, we note that age of DNA collection was highly co-linear with age of disease onset due to the fact that DNA collection typically

occurs around the time of initial clinical presentation (Supplementary Fig. 3). This co-linearity makes it statistically difficult to determine whether there is truly a lack of association between *C9orf72* expansion size and age at onset.

C9orf72 expansion size, age at death, and disease duration

We then determined the association between *C9orf72* expansion size in peripheral blood and age of disease onset, disease duration, and age at death (*n* = 38 ALS, 17 FTD). Individuals who presented with ALS died significantly earlier than patients with FTD ($p = 0.0001$). However, a multivariate regression model correcting for age of DNA collection and disease phenotype revealed that *C9orf72* expansion size is not significantly associated with age at death in individuals who presented with FTD or ALS ($p = 0.14$, Table 3).

We then generated a multivariate regression model to determine whether *C9orf72* expansion size was associated with disease duration from symptom onset to death in the same cohort of 55 individuals. We described above that age of DNA collection was highly co-linear with age of disease onset in our cohort. A similar relationship was not seen in terms of disease duration. Rather, it appeared that disease duration was longer as a function of age at onset (and also age of DNA collection) in individuals with FTD. Given this observation, an interaction between clinical diagnosis (ALS vs. FTD) and age at collection was included in our multivariate regression model with the goal of determining whether *C9orf72* expansion size was associated with

Table 3 Association of *C9orf72* expansion repeat length and age at death (n = 38 ALS, 17 FTD) from multiple regression model

Variables	Regression coefficient	Standard error	95 % confidence limits		Chi-square (DF)	p value
Intercept	0.91	1.67	−2.41	4.24	0.29 (1)	0.59
Age at collection	1.03	0.03	0.97	1.08	1370.60 (1)	<0.0001
Diagnosis (ALS)	−2.83	1.12	−5.02	−0.65	6.44 (1)	0.011
Repeat length	−0.0002	0.0003	−0.0009	−0.0004	0.56 (1)	0.45
Repeat length × diagnosis ALS	0.0006	0.0004	−0.0002	0.0014	2.17 (1)	0.14

DF degrees of freedom

Table 4 Association of *C9orf72* expansion repeat length and disease duration (n = 38 ALS, 17 FTD) from multiple regression model

Variables	Regression coefficient	Standard error	95 % confidence limits		Chi-square (DF)	p value
Intercept	−12.31	3.81	−19.79	−4.84	10.44 (1)	0.0012
Age at collection	0.35	0.068	0.22	0.49	26.44 (1)	<0.0001
Diagnosis (ALS)	13.999	4.42	5.34	22.65	10.05 (1)	0.0015
Repeat length	−0.0012	0.0005	−0.002	−0.0003	7.43 (1)	0.0064
Age at collection × diagnosis ALS	−0.35	0.0806	−0.51	−0.19	19.03 (1)	<0.0001
Repeat length × diagnosis ALS	0.0015	0.0006	0.0003	0.0027	6.51 (1)	0.0107

DF degrees of freedom

disease duration. Indeed, including this additional interaction term was justified as the model revealed that for individuals who present with FTD, a later age of collection (which corresponds to a later age at onset) is associated with significantly longer disease duration ($p < 0.0001$, Table 4).

With regard to the association between expansion size and disease duration, this multivariate regression model also demonstrated a significant interaction such that longer repeat expansion size is associated with shorter disease duration in individuals who presented with FTD ($p = 0.0107$, Table 4), while *C9orf72* expansion size had little effect on disease duration in individuals who presented with ALS. These analyses suggest that *C9orf72* expansion size is a disease modifier in FTD in that longer expansion length is associated with shorter disease duration.

Discussion

Since the *C9orf72* expansion was discovered as the most common genetic cause of FTD and ALS in 2011, the question as to whether *C9orf72* expansion size is associated with different clinical phenotypes and/or progression of disease remains uncertain. Previous studies have observed that *C9orf72* expansion size correlates with age at onset in the frontal cortex, with shorter disease duration in the cerebellum in FTD [39], with age at onset in ALS [19], and with age at onset regardless of clinical phenotype [3]. Given these conflicting results from the analysis of the *C9orf72* mutation, one issue may be related to technical differences

in ascertainment of expansion size. In the present study, we provide evidence that *C9orf72* expansion size in peripheral DNA of patients with FTD correlates with shorter disease duration. Differences in terms of choice of restriction enzyme digest for Southern blotting, size estimation procedures, and statistical modeling may account for some of the variability among the various studies. For example, the high degree of somatic instability in peripheral DNA requires that age at peripheral DNA collection be included as a covariate in regression models. With these technical and statistical challenges in mind, we were able to find a significant relationship between increased *C9orf72* expansion size and shorter disease duration in FTD, which one may expect if the repeat expansion results in a toxic gain of function. The sample size of our FTD cohort was relatively small ($n = 17$) because we only included *C9orf72* expansion cases in the regression model if information was available for age at onset, age at peripheral blood collection, age at death, and repeat length in DNA obtained from peripheral blood. Despite the small sample size, the confidence interval for the regression coefficient ($\beta = -0.0012$) was narrow (95 % CI = −0.02 and −0.0003). The reasons why the association between repeat length and disease duration was not seen in ALS remain enigmatic but could reflect the overall shorter duration of ALS or ALS-specific genetic modifiers of the effects of *C9orf72* expansion on the progression of ALS.

In our cohort, the *C9orf72* expansions were generally between approximately 300 and 3800 units (Supplementary Fig. 4). The diffuse smear seen in Southern blots is consistent with a high degree of somatic repeat instability.

Somatic instability is a feature of several microsatellite expansion loci including myotonic dystrophy type 1 (DM1) locus, Huntingtin (HTT), and spinocerebellar ataxia 10 (SCA10) [31]. Repeat instability is a dynamic process within tissues and over generations resulting in somatic and intergenerational heterogeneity. The molecular mechanisms that underlie instability of *C9orf72* expansion are not clear. However, it may be assumed that dysfunction of DNA metabolism plays a role in promoting repeat expansion based on molecular studies on other repeat expansion neurological disorders [30]. Potential mechanisms include errors in DNA replication [10, 26, 43], mismatch repair pathways, or excision repair pathways [13, 20, 27]. We confirm previous reports which found that cerebellar DNA exhibits shorter repeat size relative to peripheral DNA and to other brain regions (Supplementary Fig. 5). A similar pattern has been observed for trinucleotide repeat expansion diseases where cerebellar CAG tracts are shorter than other brain and peripheral tissues in transgenic models for HD [28] and human tissues with neurological disorder including HD and SCA1 [9, 38]. The low degree of repeat instability in cerebellum may be related to the post-mitotic nature of neurons which is particularly relevant for the cerebellum which exhibits the highest neuronal density of any tissue or brain region. Notably, CAG–CTG trinucleotide repeats have been demonstrated to expand in non-dividing mouse cells, suggesting that cell division-independent expansion can also play a role in repeat instability [16].

C9orf72 expansions are associated with a wide spectrum of neurological diseases, which implies that there are disease modifiers which affect clinical phenotypes and disease course. *TMEM106B* is a genetic modifier for *C9orf72* expansion carriers with FTD and *TMEM106B* rs1990622 major allele T is associated with later age at death and age at onset [15, 40]. While additional genetic modifiers await identification, a recent study on monozygotic twins discordant for ALS strongly suggests that genetic modifiers may not be the only factors that cause clinical heterogeneity among *C9orf72* expansion carriers [42]. For example, our group demonstrated that *C9orf72* promoter hypermethylation is associated with longer disease duration in patients with FTD and that epigenetic repression of the *C9orf72* expansion may be protective [25, 36]. We also previously observed that *C9orf72* promoter hypermethylation is associated with smaller repeat lengths, which is consistent with our current results. We speculate that there may be an interaction such that *C9orf72* repeat length and *C9orf72* promoter methylation levels may both affect disease progression, although this, as well as replication of our finding, will require further analyses of larger cohorts.

In conclusion, we demonstrate that longer *C9orf72* expansion length in peripheral DNA is correlated with shorter disease duration in patients with FTD but not ALS.

Our finding suggests that *C9orf72* expansion size may be a disease modifier in FTD.

References

1. Al-Chalabi A, Hardiman O (2013) The epidemiology of ALS: a conspiracy of genes, environment and time. Nat Rev Neurol 9:617–628. doi:10.1038/nrneurol.2013.203
2. Allignol A, Schumacher M, Beyersmann J (2011) Empirical transition matrix of multi-state models: the etm package. J Stat Softw 38:1–15
3. Beck J, Poulter M, Hensman D et al (2013) 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 92:345–353. doi:10.1016/j.ajhg.2013.01.011
4. Bender R, Lange S (2001) Adjusting for multiple testing—when and how? J Clin Epidemiol 54:343–349
5. Brettschneider J, Van Deerlin VM, Robinson JL et al (2012) Pattern of ubiquilin pathology in ALS and FTLD indicates presence of *C9ORF72* hexanucleotide expansion. Acta Neuropathol 123:825–839. doi:10.1007/s00401-012-0970-z
6. Brooks BR (1994) El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "clinical limits of amyotrophic lateral sclerosis" workshop contributors. J Neurol Sci 124(Suppl):96–107
7. Byrne S, Elamin M, Bede P et al (2012) Cognitive and clinical characteristics of patients with amyotrophic lateral sclerosis carrying a *C9orf72* repeat expansion: a population-based cohort study. Lancet Neurol 11:232–240. doi:10.1016/S1474-4422(12)70014-5
8. Chio A, Calvo A, Mazzini L et al (2012) Extensive genetics of ALS: a population-based study in Italy. Neurology 79:1983–1989. doi:10.1212/WNL.0b013e3182735d36
9. Chong SS, McCall AE, Cota J et al (1995) Gametic and somatic tissue-specific heterogeneity of the expanded SCA1 CAG repeat in spinocerebellar ataxia type 1. Nat Genet 10:344–350. doi:10.1038/ng0795-344
10. Cleary JD, Pearson CE (2005) Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. Trends Genet 21:272–280. doi:10.1016/j.tig.2005.03.008
11. Cooper-Knock J, Shaw PJ, Kirby J (2014) The widening spectrum of *C9ORF72*-related disease; genotype/phenotype correlations and potential modifiers of clinical phenotype. Acta Neuropathol 127:333–345. doi:10.1007/s00401-014-1251-9
12. DeJesus-Hernandez M, Mackenzie IR, Boeve BF et al (2011) Expanded GGGGCC hexanucleotide repeat in noncoding region of *C9ORF72* causes chromosome 9p-linked FTD and ALS. Neuron 72:245–256. doi:10.1016/j.neuron.2011.09.011
13. Dion V, Wilson JH (2009) Instability and chromatin structure of expanded trinucleotide repeats. Trends Genet 25:288–297. doi:10.1016/j.tig.2009.04.007

14. Dols-Icardo O, Garcia-Redondo A, Rojas-Garcia R et al (2014) Characterization of the repeat expansion size in *C9orf72* in amyotrophic lateral sclerosis and frontotemporal dementia. Hum Mol Genet 23:749–754. doi:10.1093/hmg/ddt460

15. Gallagher MD, Suh E, Grossman M et al (2014) TMEM106B is a genetic modifier of frontotemporal lobar degeneration with *C9orf72* hexanucleotide repeat expansions. Acta Neuropathol 127:407–418. doi:10.1007/s00401-013-1239-x

16. Gomes-Pereira M, Hilley JD, Morales F, Adam B, James HE, Monckton DG (2014) Disease-associated CAG.CTG triplet repeats expand rapidly in non-dividing mouse cells, but cell cycle arrest is insufficient to drive expansion. Nucleic Acids Res 42:7047–7056. doi:10.1093/nar/gku285

17. Gorno-Tempini ML, Hillis AE, Weintraub S et al (2011) Classification of primary progressive aphasia and its variants. Neurology 76:1006–1014. doi:10.1212/WNL.0b013e31821103e6

18. Hu WT, Seelaar H, Josephs KA et al (2009) Survival profiles of patients with frontotemporal dementia and motor neuron disease. Arch Neurol 66:1359–1364. doi:10.1001/archneurol.2009.253

19. Hubers A, Marroquin N, Schmoll B et al (2014) Polymerase chain reaction and southern blot-based analysis of the *C9orf72* hexanucleotide repeat in different motor neuron diseases. Neurobiol Aging 35(5):1214 e1–6. doi:10.1016/j.neurobiolaging.2013.11.034

20. Hubert L Jr, Lin Y, Dion V, Wilson JH (2011) Topoisomerase 1 and single-strand break repair modulate transcription-induced CAG repeat contraction in human cells. Mol Cell Biol 31:3105–3112. doi:10.1128/MCB.05158-11

21. Irwin DJ, Cairns NJ, Grossman M et al (2015) Frontotemporal lobar degeneration: defining phenotypic diversity through personalized medicine. Acta Neuropathol 129:469–491. doi:10.1007/s00401-014-1380-1

22. Kostic VS, Dobricic V, Stankovic I, Ralic V, Stefanova E (2014) *C9orf72* expansion as a possible genetic cause of Huntington disease phenocopy syndrome. J Neurol 261:1917–1921. doi:10.1007/s00415-014-7430-8

23. Lesage S, Le Ber I, Condroyer C et al (2013) *C9orf72* repeat expansions are a rare genetic cause of parkinsonism. Brain 136:385–391. doi:10.1093/brain/aws357

24. Lindquist SG, Duno M, Batbayli M et al (2013) Corticobasal and ataxia syndromes widen the spectrum of *C9ORF72* hexanucleotide expansion disease. Clin Genet 83:279–283. doi:10.1111/j.1399-0004.2012.01903.x

25. Liu EY, Russ J, Wu K et al (2014) *C9orf72* hypermethylation protects against repeat expansion-associated pathology in ALS/FTD. Acta Neuropathol 128:525–541. doi:10.1007/s00401-014-1286-y

26. Liu G, Chen X, Bissler JJ, Sinden RR, Leffak M (2010) Replication-dependent instability at (CTG) × (CAG) repeat hairpins in human cells. Nat Chem Biol 6:652–659. doi:10.1038/nchembio.416

27. Lopez Castel A, Cleary JD, Pearson CE (2010) Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol 11:165–170. doi:10.1038/nrm2854

28. Mangiarini L, Sathasivam K, Mahal A, Mott R, Seller M, Bates GP (1997) Instability of highly expanded CAG repeats in mice transgenic for the Huntington's disease mutation. Nat Genet 15:197–200. doi:10.1038/ng0297-197

29. McMillan CT, Toledo JB, Avants BB et al (2014) Genetic and neuroanatomic associations in sporadic frontotemporal lobar

30. McMurray CT (2010) Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet 11:786–799. doi:10.1038/nrg2828

31. Pearson CE, Nichol Edamura K, Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 6:729–742. doi:10.1038/nrg1689

32. Rascovsky K, Grossman M (2013) Clinical diagnostic criteria and classification controversies in frontotemporal lobar degeneration. Int Rev Psychiatry 25:145–158. doi:10.3109/09540261.2013.763341

33. Rascovsky K, Hodges JR, Knopman D et al (2011) Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. Brain 134:2456–2477. doi:10.1093/brain/awr179

34. Renton AE, Chio A, Traynor BJ (2014) State of play in amyotrophic lateral sclerosis genetics. Nat Neurosci 17:17–23. doi:10.1038/nn.3584

35. Renton AE, Majounie E, Waite A et al (2011) A hexanucleotide repeat expansion in *C9ORF72* is the cause of chromosome 9p21-linked ALS–FTD. Neuron 72:257–268. doi:10.1016/j.neuron.2011.09.010

36. Russ J, Liu EY, Wu K et al (2015) Hypermethylation of repeat expanded *C9orf72* is a clinical and molecular disease modifier. Acta Neuropathol 129:39–52. doi:10.1007/s00401-014-1365-0

37. Smith BN, Newhouse S, Shatunov A et al (2013) The *C9ORF72* expansion mutation is a common cause of ALS ± FTD in Europe and has a single founder. Eur J Hum Genet 21:102–108. doi:10.1038/ejhg.2012.98

38. Tanaka F, Sobue G, Doyu M et al (1996) Differential pattern in tissue-specific somatic mosaicism of expanded CAG trinucleotide repeats in dentatorubral–pallidoluysian atrophy, Machado–Joseph disease, and X-linked recessive spinal and bulbar muscular atrophy. J Neurol Sci 135:43–50

39. van Blitterswijk M, DeJesus-Hernandez M, Niemantsverdriet E et al (2013) Association between repeat sizes and clinical and pathological characteristics in carriers of *C9ORF72* repeat expansions (Xpansize-72): a cross-sectional cohort study. Lancet Neurol 12:978–988. doi:10.1016/S1474-4422(13)70210-2

40. van Blitterswijk M, Mullen B, Nicholson AM et al (2014) *TMEM106B* protects *C9ORF72* expansion carriers against frontotemporal dementia. Acta Neuropathol 127:397–406. doi:10.1007/s00401-013-1240-4

41. Wood EM, Falcone D, Suh E et al (2013) Development and validation of pedigree classification criteria for frontotemporal lobar degeneration. JAMA neurology 70:1411–1417. doi:10.1001/jamaneurol.2013.3956

42. Xi Z, Yunusova Y, van Blitterswijk M et al (2014) Identical twins with the *C9orf72* repeat expansion are discordant for ALS. Neurology 83:1476–1478. doi:10.1212/WNL.0000000000000886

43. Yang Z, Lau R, Marcadier JL, Chitayat D, Pearson CE (2003) Replication inhibitors modulate instability of an expanded trinucleotide repeat at the myotonic dystrophy type 1 disease locus in human cells. Am J Hum Genet 73:1092–1105. doi:10.1086/379523