

Supporting Information for this preprint is available from  
the *Human Mutation* editorial office upon request  
([humu@wiley.com](mailto:humu@wiley.com))

**A Pan-European study of the *C9orf72* repeat associated with FTL D: geographic prevalence, genomic instability and intermediate repeats.**

Julie van der Zee<sup>1,2\*</sup>, Ilse Gijselinck<sup>1,2\*</sup>, Lubina Dillen<sup>1,2</sup>, Tim Van Langenhove<sup>1,2</sup>, Jessie Theuns<sup>1,2</sup>, Sebastiaan Engelborghs<sup>2,3</sup>, Stéphanie Philtjens<sup>1,2</sup>, Mathieu Vandenbulcke<sup>4</sup>, Kristel Slegers<sup>1,2</sup>, Anne Sieben<sup>1,2,5</sup>, Veerle Bäumer<sup>1,2</sup>, Githa Maes<sup>1,2</sup>, Ellen Corsmit<sup>1,2</sup>, Barbara Borroni<sup>6</sup>, Alessandro Padovani<sup>6</sup>, Silvana Archetti<sup>6</sup>, Robert Pernecky<sup>7</sup>, Janine Diehl-Schmid<sup>7</sup>, Alexandre de Mendonça<sup>8</sup>, Gabriel Miltenberger-Miltenyi<sup>8</sup>, Sónia Pereira<sup>8</sup>, José Pimentel<sup>9</sup>, Benedetta Nacmias<sup>10</sup>, Silvia Bagnoli<sup>10</sup>, Sandro Sorbi<sup>10</sup>, Caroline Graff<sup>11,12</sup>, Huei-Hsin Chiang<sup>11</sup>, Marie Westerlund<sup>11</sup>, Raquel Sanchez-Valle<sup>12</sup>, Albert Llado<sup>12</sup>, Ellen Gelpi<sup>13</sup>, Isabel Santana<sup>14,15</sup>, Maria Rosário Almeida<sup>16</sup>, Beatriz Santiago<sup>14</sup>, Giovanni Frisoni<sup>17</sup>, Orazio Zanetti<sup>17</sup>, Cristian Bonvicini<sup>17</sup>, Matthis Synofzik<sup>18,19</sup>, Walter Maetzler<sup>18,19</sup>, Jennifer Müller vom Hagen<sup>18,19</sup>, Ludger Schöls<sup>18,19</sup>, Michael T. Heneka<sup>20,21</sup>, Frank Jessen<sup>20,21,22</sup>, Radoslav Matej<sup>23</sup>, Eva Parobkova<sup>23</sup>, Gabor G. Kovacs<sup>24</sup>, Thomas Ströbel<sup>24</sup>, Stayko Sarafov<sup>25</sup>, Ivailo Tournev<sup>25,26</sup>, Albena Jordanova<sup>1,27</sup>, Adrian Danek<sup>28</sup>, Thomas Arzberger<sup>29</sup>, Gian Maria Fabrizi<sup>30</sup>, Silvia Testi<sup>30</sup>, Eric Salmon<sup>31</sup>, Patrick Santens<sup>5</sup>, Jean-Jacques Martin<sup>2</sup>, Patrick Cras<sup>2,32</sup>, Rik Vandenberghe<sup>33</sup>, Peter Paul De Deyn<sup>2,3</sup>, Marc Cruts<sup>1,2+</sup>, Christine Van Broeckhoven<sup>1,2+</sup> on behalf of the European Early-Onset Dementia (EOD) Consortium<sup>†</sup>

© 2012 Wiley Periodicals, Inc.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as [DOI: 10.1002/humu.22244].

\* Joint First authors, † Joint Corresponding authors

<sup>1</sup> Department of Molecular Genetics, VIB, Antwerp, Belgium

<sup>2</sup> Institute Born-Bunge, University of Antwerp, Antwerp, Belgium

<sup>3</sup> Department of Neurology and Memory Clinic, Hospital Network Antwerp, Middelheim and Hoge Beuken, Antwerp, Belgium

<sup>4</sup> Brain and Emotion Laboratory Leuven, Department of Psychiatry, University of Leuven and University Hospitals Leuven Gasthuisberg, Leuven, Belgium

<sup>5</sup> Department of Neurology, University Hospital Ghent and University of Ghent, Ghent, Belgium

<sup>6</sup> Centre for Ageing Brain and Neurodegenerative Disorders, Neurology Unit, University of Brescia, Brescia, Italy

<sup>7</sup> Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany

<sup>8</sup> Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, Lisbon, Portugal

<sup>9</sup> Faculty of Medicine, University of Lisbon, Lisbon, Portugal

<sup>10</sup> Department of Neurological and Psychiatric Sciences, University of Florence, Florence, Italy

<sup>11</sup> Department of Neurobiology, Care Sciences and Society, KI-Alzheimer Disease Research Center, Karolinska Institutet, Stockholm, Sweden

<sup>12</sup> Department of Geriatric Medicine, Genetics unit, Karolinska University Hospital, Stockholm, Sweden

<sup>12</sup> Alzheimer's disease and Other cognitive disorders unit, Department of Neurology, Hospital Clinic, Barcelona, Spain

<sup>13</sup> Biobanc, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), Barcelona, Spain

<sup>14</sup> Neurology Department, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal

<sup>15</sup> Faculty of Medicine, University of Coimbra, Coimbra, Portugal

<sup>16</sup> Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal

- <sup>17</sup> Laboratory of Epidemiology Neuroimaging & Telemedicine & National Centre for Alzheimer's and Mental Diseases, IRCCS Centro San Giovanni di Dio, FBF, Brescia, Italy
- <sup>18</sup> Department of Neurodegeneration, Hertie Institute for Clinical Brain Research and Centre of Neurology, Tuebingen, Germany
- <sup>19</sup> German Research Center for Neurodegenerative Diseases (DZNE), University of Tuebingen, Germany
- <sup>20</sup> Clinical Neuroscience Unit, Department of Neurology, University of Bonn, Germany
- <sup>21</sup> German Center for Neurodegenerative Diseases (DZNE), University of Bonn, Germany
- <sup>22</sup> Department of Psychiatry, University of Bonn, Bonn, Germany
- <sup>23</sup> Department of Pathology and Molecular Medicine, Thomayer Hospital, Prague, Czech Republic
- <sup>24</sup> Institute of Neurology, Medical University Vienna, Vienna, Austria
- <sup>25</sup> Department of Neurology, Medical University Sofia, Sofia, Bulgaria
- <sup>26</sup> Department of Cognitive Science and Psychology, New Bulgarian University, Sofia, Bulgaria.
- <sup>27</sup> Department of Biochemistry, Molecular Medicine Center, Medical University, Sofia, Sofia, Bulgaria
- <sup>28</sup> Neurologische Klinik, Ludwig Maximilians Universität, Munich, Germany
- <sup>29</sup> Zentrum für Neuropathologie und Prionforschung, Ludwig-Maximilians-Universität Munich, Germany
- <sup>30</sup> Section of Neuropathology, Department of Neurological, Neuropsychological, Morphological and Movement Sciences, University of Verona, Italy
- <sup>31</sup> Cyclotron Research Centre and Department of Neurology, University of Liège, Liège, Belgium
- <sup>32</sup> Department of Neurology, Antwerp University Hospital, Edegem, Belgium
- <sup>33</sup> Laboratory for Cognitive Neurology, Department of Neurology, University of Leuven and University Hospitals Leuven Gasthuisberg, Leuven, Belgium

**† Corresponding authors:**

Correspondences, Reprints and Information requests to Prof. dr. Christine Van Broeckhoven, PhD, DSc, Neurodegenerative Brain Diseases Group, VIB Department of Molecular Genetics, University of

Antwerp - CDE, Universiteitsplein 1, B-2610, Antwerp, Belgium; Tel: +3232651102; Fax: +3232651112;

Email: [christine.vanbroeckhoven@molgen.vib-ua.be](mailto:christine.vanbroeckhoven@molgen.vib-ua.be)

Information request to Prof. Dr. Marc Cruts, PhD, Neurodegenerative Brain Diseases Group, VIB

Department of Molecular Genetics, University of Antwerp - CDE, Universiteitsplein 1, B-2610, Antwerp,

Belgium; Tel: +3232651033; Fax: +3232651112; Email: [marc.cruts@molgen.vib-ua.be](mailto:marc.cruts@molgen.vib-ua.be)

‡ EOD Consortium members that contributed to the data in this paper are listed in the appendix

UNCORRECTED ACCEPTED ARTICLE

## ABSTRACT

We assessed the geographical distribution of *C9orf72* G<sub>4</sub>C<sub>2</sub> expansions in a pan-European frontotemporal lobar degeneration (FTLD) cohort (n=1205), ascertained by the European Early-Onset Dementia (EOD) consortium. Next, we performed a meta-analysis of our data and that of other European studies, together 2,668 patients from 15 Western European countries. The frequency of the *C9orf72* expansions in Western Europe was 9.98% in overall FTLD, with 18.52% in familial, and 6.26% in sporadic FTLD patients. Outliers were Finland and Sweden with overall frequencies of respectively 29.33% and 20.73%, but also Spain with 25.49%. In contrast, prevalence in Germany was limited to 4.82%. In addition, we studied the role of intermediate repeats (7 to 24 repeat units), which are strongly correlated with the risk haplotype, on disease and *C9orf72* expression. *In vitro* reporter gene expression studies demonstrated significantly decreased transcriptional activity of *C9orf72* with increasing number of normal repeat units, indicating that intermediate repeats might act as predisposing alleles and in favor of the loss-of-function disease mechanism. Further, we observed a significantly increased frequency of short indels in the GC-rich low complexity sequence (LCS) adjacent to the G<sub>4</sub>C<sub>2</sub> repeat in *C9orf72* expansion carriers (p < 0.001) with the most common indel creating one long contiguous imperfect G<sub>4</sub>C<sub>2</sub> repeat which is likely more prone to replication slippage and pathological expansion.

**Key words:** FTLD, *C9orf72*, Repeat expansion, Intermediate alleles, European Early-Onset Dementia consortium

## INTRODUCTION

A pathological expansion of a hexanucleotide G<sub>4</sub>C<sub>2</sub> repeat in the promoter region of the gene *C9orf72* (MIM# 614260) was recently identified as the long sought-after underlying gene defect [DeJesus-Hernandez et al., 2011; Renton et al., 2011; Gijssels et al., 2012] of linkage [Vance et al., 2006; Morita et al., 2006; Valdmanis et al., 2008; Luty et al., 2008; Le Ber et al., 2009; Gijssels et al., 2010; Pearson et al., 2011; Boxer et al., 2011] and association [van Es et al., 2009; Laaksovirta et al., 2010; Van Deerlin et al., 2010; Shatunov et al., 2010] of frontotemporal lobar degeneration (FTLD) and amyotrophic lateral sclerosis (ALS) to the chromosome 9p21 region [Van Langenhove et al., 2012a]. In the Flanders-Belgian population we calculated that the pathological G<sub>4</sub>C<sub>2</sub> expansion is the second most common genetic cause of FTLD [Gijssels et al., 2012; Van Langenhove et al., 2012b]. Particularly in the subgroup of familial patients presenting with ALS also (FTLD-ALS), *C9orf72* was the first causal gene explaining up to 85.71% of patients [Gijssels et al., 2012; Van Langenhove et al., 2012b]. Since then, several patient cohorts of different geographical regions have been screened for this repeat expansion mutation establishing *C9orf72* as a major gene for FTLD with frequencies of 7-11% in total FTLD and 12-25% in familial FTLD patients [Boeve et al., 2012; DeJesus-Hernandez et al., 2011; Gijssels et al., 2012; Hsiung et al., 2012; Mahoney et al., 2012; Renton et al., 2011; Simon-Sanchez et al., 2012; Snowden et al., 2012; Majounie 2012; Ferrari 2012].

Different possible disease mechanisms have been proposed including haploinsufficiency and RNA toxicity [DeJesus-Hernandez et al., 2011; Gijssels et al., 2012] and extensive genotype-phenotype correlation studies are being reported [DeJesus-Hernandez et al., 2011; Renton et al., 2011; Murray et al., 2011; Al-Sarraj et al., 2011; Troakes et al., 2011; Chio et al., 2012; Majounie et al., 2012; Ferrari et al., 2012; Arighi et al., 2012; Bigio, 2012; Simon-Sanchez et al., 2012; Mahoney et al., 2012; Boeve et al., 2012; Whitwell et al., 2012; Snowden et al., 2012; Hsiung et al., 2012; Gijssels et al., 2012]. However, very little or nothing is known about the mutation spectrum, the genomic mechanism by which the G<sub>4</sub>C<sub>2</sub> repeat is expanding and the impact of repeat length on disease susceptibility and gene expression.

In the present study, we aimed at expanding our *C9orf72* observations in the Flanders-Belgian FTLN and FTLN-ALS cohort (n= 360) with a larger European cohort of 845 FTLN and FTLN-ALS patients, in which we determined the geographical distribution and prevalence of the pathological G<sub>4</sub>C<sub>2</sub> expansion. Furthermore, we provide the first evidence for a role of G<sub>4</sub>C<sub>2</sub> intermediate repeat length on *C9orf72* expression and hypothesize on the genomic mechanisms favoring pathological expansion of the G<sub>4</sub>C<sub>2</sub> repeat.

## MATERIALS AND METHODS

### Study populations

The European FTLN cohort was collected through the European Early-Onset Dementia (EOD) consortium (Supp. Table S1). The European EOD consortium was launched in August 2011 to centralize and harmonize epidemiological, clinical, and biological data together with biomaterial of early-onset dementia patients throughout Europe to stimulate high-profile translational dementia research. Supp. Table S1 describes the number of patients per country and per clinical subgroup contributed by the European EOD consortium members. We received DNA and clinical and demographic information on 917 unrelated FTLN and FTLN-ALS patients as well as histopathology data of 46 patients obtained at autopsy. The 917 patients also included 10 patients from Wallonia, the French speaking part of Belgium, and 6 more patients from Italy, Spain, and Sweden, that were referred for clinical genetic testing to the Diagnostic Service Facility in our Department of Molecular Genetics (DMG DSF). Patients had been diagnosed according to established clinical diagnostic Neary criteria [Neary et al., 1998] and to the Mackenzie consensus criteria for neuropathology diagnosis [Mackenzie et al., 2010].

The Flanders-Belgian cohort consisted of 337 unrelated patients with FTLN and 23 with FTLN-ALS. These patients were recruited through the Belgian Neurology (BELNEU) consortium, a collaboration with neurologists affiliated to 9 different specialized memory clinics and neurology departments in Belgium [Gijssels et al., 2012; Van Langenhove et al., 2012b] (Supp. Table S2). In addition to the

patient cohort a Flanders-Belgian control cohort was assembled (n=1083). For more detailed description see Supp. Materials and Methods.

### **Histopathology of *C9orf72* expansion carriers**

From 11 *C9orf72* G<sub>4</sub>C<sub>2</sub> expansion carriers formalin-fixed brain was available for immunohistochemistry. Five µm slices were obtained from frontal cortex, temporal neocortex, hippocampus, area striata, neostriatum, mesencephalon, pons and cerebellum. Of seven cases, additional samples were provided from thalamus and spinal cord. Slides were stained against Ubiquitin, p62, hyperphosphorylated tau, β-amyloid, TDP43, and FUS. For technical details see Supp. Materials and Methods.

### ***C9orf72* G<sub>4</sub>C<sub>2</sub> genetic PCR assays for genotyping**

We developed an alternative repeat-primed PCR assay (reverse RP-PCR; Figure 1) and a STR fragment length assay with flanking primers optimized for alleles with high GC content (STR-PCR, Figure 1) allowing reliable identification of G<sub>4</sub>C<sub>2</sub> expansion carriers and exact sizing of normal lengths. These assays were performed in both cohorts and in relatives of the younger generation of index patients carrying an intermediate repeat allele or a variation in the flanking LCS without expansion. For technical details on primers and amplification protocols see Supp. Materials and Methods.

### **Sequencing of the *C9orf72* GC-rich low complexity sequence**

We used the product of an alternative forward repeat-primed PCR (RP-PCR for sequencing; Figure 1) and sequenced the LCS with the locus-specific reverse primer.

The Flanders-Belgian patient (N=317) and control (N=752) cohorts and 57 expansion carriers and 114 non-expansion carriers of the European cohort were successfully screened. Co-segregation of LCS variations with the presence of a G<sub>4</sub>C<sub>2</sub> expansion was analyzed in two available families. For technical details see Supp. Materials and Methods.



### ***C9orf72* exon sequencing and dosage analysis**

Both cohorts were screened for coding and splice-site mutations in *C9orf72*. The frequency of rare mutations was determined in 400 controls. Further, we screened the Flanders-Belgian cohort for exonic deletions or duplications using the MAQ technique [Kumps et al., 2010]. For technical details see Supp. Materials and Methods.

### **Genetic association studies**

We calculated association with disease of *C9orf72* intermediate G<sub>4</sub>C<sub>2</sub> alleles and of SNP rs2814707 tagging the chromosome 9p21 risk haplotype stratifying for the presence or absence of *C9orf72* intermediate repeats. Odds ratios (OR) with 95% confidence interval (CI) were calculated in a logistic regression model, adjusted for age and gender. In addition, we studied the correlation between the minor risk T-allele and intermediate repeat length. Further, in patients of both cohorts without a G<sub>4</sub>C<sub>2</sub> expansion, we calculated correlation between age at onset and normal repeat length using a Kruskal-Wallis test. Further, we compared age at onset between short (2 - 6) and intermediate (7 - 24) repeat length in a Kaplan-Meier survival analysis. All analyses were done in IBM SPSS Statistics 20. For details see the Supp. Materials and Methods.

### **Luciferase reporter assays**

We selected a 2kb *C9orf72* promoter fragment (chr9:27,572,414-27,574,451; NCBI Build37 – hg19) containing the G<sub>4</sub>C<sub>2</sub> repeat and enriched for histone marks, DNaseI hypersensitivity clusters and transcription factor binding sites based on ENCODE transcription data [Gijssels et al., 2012]. The fragment was obtained by PCR of individuals carrying different numbers of normal repeat units (2, 8, 17 and 24 units) using primers with flanked attB-sites. PCR products were cloned into the pDONR221 vector (Invitrogen) by a BP recombination reaction (Invitrogen) and the integrity of all inserts was confirmed by sequence analysis. Correct entry clones were selected and cloned into an in-house developed promoterless destination vector containing the *Gaussia* luciferase reporter gene downstream of

a Gateway cassette, by use of a LR recombination reaction (Invitrogen). Human HEK293T cells were propagated and seeded for transient transfection in 24-well tissue-culture dishes, at  $2 \cdot 10^5$  cells per well, and were allowed to recover for 24 h. Cells were co-transfected with 40 ng of pSV40-CLuc plasmid that encodes the *Cypridina* luciferase gene with a SV40 promoter (New England Biolabs) and 1000 ng of three independent *C9orf72* promoter constructs per unit, with use of 2.4  $\mu$ l Lipofectamine 2000 (Invitrogen), in duplo. After 24h, Gaussia luciferase activities ( $LA_G$ ) and *Cypridina* luciferase activities ( $LA_C$ ) were measured in duplo in the growth medium using a BioLux *Gaussia* and *Cypridina* Luciferase Assay Kit (New England Biolabs) and a Veritas Microplate Luminometer with Dual Reagent Injectors Luminometer (Promega). To correct for transfection efficiency and DNA uptake, the relative luciferase activity (RLA) was calculated as  $RLA = LA_G / LA_C$ . This experiment was repeated 3 times resulting in 36 measurements for each construct. RLAs between different repeat lengths were calculated by a Mann-Whitney U test. For details see Supp. Materials and Methods.

## RESULTS

### The European EOD Consortium

The European cohort included 917 unrelated patients, of which 845 had a possible or probable diagnosis of FTLN (n=781) or FTLN-ALS (n=64) (Table 1). In an additional 72 other patients, clinical presentation showed indications of FTLN together with symptomatology of other neurodegenerative brain diseases such as Alzheimer or Parkinson disease. A pathological diagnosis on autopsied brain was obtained for 45 patients, comprising FTLN-TDP (n=28), FTLN-MND-TDP (n=15), FTLN-UPS (n=1), and FTLN-TAU (n=1) diagnoses. Information on family history of disease was available for 609 (72.07%) of the 845 patients, of which 251 had a positive family history of disease and 358 were considered sporadic patients (Table 2). Average onset age and range were comparable between FTLN ( $62.7 \pm 9.0$ , range 28-88 years) and FTLN-ALS patient groups ( $60.9 \pm 9.8$ , range 31-83 years).

### *C9orf72* pathological $G_4C_2$ expansions

To determine the impact and distribution of the G<sub>4</sub>C<sub>2</sub> expansion across Europe, G<sub>4</sub>C<sub>2</sub> repeat lengths were determined by forward repeat-primed PCR (Figure 1). A G<sub>4</sub>C<sub>2</sub> expansion was observed in 8.64% (73/845) of the European cohort (Table 2). Per clinical subgroup, 6.40% of FTLN and 35.94% of FTLN-ALS patients carried a pathological G<sub>4</sub>C<sub>2</sub> expansion. In familial patients, the overall frequency increased to 11.68%, with frequencies of 8.37% (21/251) in FTLN, and 47.83% (11/23) in FTLN-ALS patients. In sporadic patients, the overall frequency decreased to 4.68%, with frequencies of 3.63% (13/358) in FTLN, and 18.52% (5/27) in FTLN-ALS.

To evaluate the distribution of the G<sub>4</sub>C<sub>2</sub> expansion we calculated, overall and per clinical phenotype, mutation frequencies per country (Supp. Table S2, Table 3). In Belgium, Portugal, and Italy, the pathological G<sub>4</sub>C<sub>2</sub> expansion mutation showed a comparable overall frequency ranging between 6.09 and 7.86%, and close to the average overall European frequency of 8.38%. However, a marked enrichment was observed in the Spanish (25.49%) and Swedish (21.33%) patient cohorts. In contrast, in the German patients only 3.52% were carriers.

### ***C9orf72*-associated clinical and pathological phenotype**

Of the 73 G<sub>4</sub>C<sub>2</sub> expansion carriers in the European cohort, 50 received a clinical diagnosis of FTLN and 23 of FTLN-ALS. The average onset age in all carriers was  $58.0 \pm 7.5$  years with an onset age range of 40 to 75 years and was comparable in the FTLN and FTLN-ALS subgroups (Supp. Table S3). The average disease duration in 31 deceased carriers was  $5.3 \pm 3.5$  years (range 1 to 14 years) but differed in the clinical subgroups. Survival was on average 1.4 years shorter for the FTLN-ALS carriers with  $4.7 \pm 3.5$  years (n=17, range 1 to 14 years), compared to the FTLN patients with  $6.1 \pm 3.6$  years (n=14, range 1 to 14 years). Of 23 G<sub>4</sub>C<sub>2</sub> expansion carriers, more extensive clinical information was available to allow sub-classification to the different FTLN phenotypes. In 22, clinical presentation was conform bvFTD (95.65%) and one presented with PNFA. Autopsied brain was available for 11 European G<sub>4</sub>C<sub>2</sub> expansion carriers (2 Portuguese, 5 Spanish, 1 Austrian, 1 Czech, and 2 Swedish patients). In all 11 cases, we found TDP-43 pathology in the frontal and temporal neocortex, in the hippocampus and neostriatum, which was

compatible with type B TDP proteinopathy [Mackenzie et al., 2010]. Accordingly, TDP-43 immunoreactive neuronal cytoplasmic inclusions (NCI) and dystrophic neurites (DN) were widespread over the entire cortical thickness, but NCI were more pronounced in the pyramidal cells of layer 2 compared to deeper cortical layers. In addition to the TDP-43 positive pathology, p62 immunoreactive NCI were present in the granular layer of the dentate gyrus of the hippocampus, and in the granular layer of the cerebellar cortex. Further, p62 positive irregular granular NCI were observed in pyramidal neurons of the CA4 and CA3 region of the hippocampus.

### **Genomic complexity in the *C9orf72* region**

The *C9orf72* G<sub>4</sub>C<sub>2</sub> repeat is contiguous with a GC-rich, low complexity sequence (LCS), comprising exon 1 of the *C9orf72* transcript NM\_018325.3 (Figure 1). We sequenced the GC-rich LCS in 317 unrelated patients of the Flanders-Belgian cohort and in 752 controls (Figure 1). We observed heterozygous deletions of 5 to 23 base pairs (bp) in a total of 19 individuals of which 10 patients carrying a G<sub>4</sub>C<sub>2</sub> expansion (10/27 = 37.04%), 5 non-carrier patients (5/290 = 1.72%) and 4 control persons (4/752 = 0.53%) (Table 4). These variable deletions were significantly more frequently observed in carriers of a G<sub>4</sub>C<sub>2</sub> expansion compared to the group of non-carrier patients (OR = 33.53; 95% CI 10.31 - 109.09; p<0.001) and controls (OR = 93.50; 95% CI 28.53 - 306.39; p<0.001). Remarkably, 9 of 10 (90.00%) expansion carriers presented with the same heterozygous 10-bp GTGGTCGGGG deletion (g.26747\_26756 delGTGGTCGGGG) (Supp. Figure S1) which was not observed in patient non-carriers and controls (Table 4). This 10-bp deletion is contiguous with the G<sub>4</sub>C<sub>2</sub> repeat and joins two 100% GC sequences thereby extending the GC-rich motif of the G<sub>4</sub>C<sub>2</sub> repeat with imperfect repeats (Figure 1). In this context, it is striking that deletion of the GTGGT motif (Figure 1) was seen in all 10 deletion carriers of the 27 unrelated patients carrying an expanded G<sub>4</sub>C<sub>2</sub> repeat (37.04%), only once in the non-carrier patients (1/290 = 0.34%) and once in control individuals (1/752=0.13%) (Table 4). To replicate these findings, we successfully sequenced the LCS in 57 unrelated patient carriers and 114 patient non-carriers from the European cohort. We observed a comparable high frequency of deletions and insertions (indels)

in the patient carriers (14/57 = 24.56%) and no indels in the non-carriers (0/114, <0.88%) (OR = 36.79; 95% CI 4.69 - 288.35; p = 0.001) (Table 4). Seven of 14 (50.00%) patient carriers presented the same 10-bp deletion and in 10 of 14 (71.43%) patient carriers GTGGT was deleted (Table 4). Of note, children of the FTL D patient with the g.26747\_26751delGTGGT deletion but without expansion did not show *de novo* expansions. This deletion is located on an allele of 5 repeat units.

The LCS is comprised in the PCR fragments produced by the forward RP-PCR assay to identify G<sub>4</sub>C<sub>2</sub> expansion carriers (Figure 1) [Gijssels et al., 2012]. To eliminate influences of LCS variability in G<sub>4</sub>C<sub>2</sub> expansion detection and for sizing normal repeat alleles, we developed a reverse RP-PCR assay on the sense strand (Figure 1). This assay confirmed the presence of the G<sub>4</sub>C<sub>2</sub> expansion in patients of both cohorts.

#### **Sizing of normal *C9orf72* G<sub>4</sub>C<sub>2</sub> repeat lengths**

We used a STR genotyping assay (STR-PCR, Figure 1), to size both normal alleles in non-carriers and the wild-type allele in G<sub>4</sub>C<sub>2</sub> expansion carriers. In addition, we used the reverse RP-PCR assay (Figure 1) to validate the size of the longest allele in non-expansion carriers. When we compared the allele lengths between the two assays, we obtained 99% concordance. Discordant allele scoring could be explained by the presence of a LCS deletion. This implies that the STR-PCR can be used for correct sizing of normal repeat alleles. The observed lengths of the G<sub>4</sub>C<sub>2</sub> repeat ranged from 2 to 24 units in non-expansion carriers of the Flanders-Belgian cohort (g.26724GGGGCC(2\_24); relative to reverse complement of AL451123.12) (Figure 2A) and between 2 and 21 units in the European cohort. The shortest allele contained one unit less than the reference genome (NCBIbuild37 - hg19) (Figure 1). Frequency distribution of normal repeat alleles showed a trimodal allele distribution (Figure 2A). Using an estimation-maximization (EM) algorithm implemented in the R Package MCLUST, the three groups could be defined by a cut-off at 4 and 7 units (2 to 3 units, 4 to 6 units and 7 to 24 units), with a probability of good classification of 0.9944.

### Intermediate repeat length and the chromosome 9 risk haplotype

In the Flanders-Belgian cohort, we previously had calculated a significant association of disease with the risk T-allele of rs2814707 ( $p=0.008$ ) [Gijssels et al., 2012] tagging the chromosome 9p risk haplotype. Repeat expansion carriers contributed the largest fraction of the attributed risk but residual association signal remained in the non-expansion carriers homozygous for the T-allele (OR = 1.75; 95% CI 1.02 - 3.01;  $p = 0.042$ ) (Supp. Table S4) [Gijssels et al., 2012]. We used three different PCR genotyping assays to test for pathological  $G_4C_2$  expansions, making it unlikely that the residual association could be explained by missed mutation carriers. Also, we analyzed *C9orf72* for other mutations by sequencing all exons in 493 patients and by exon-based analysis of putative deletions/duplications in 413 patients. Except for one patient-specific missense mutations (c.196A>T in NM\_018325.3 (p.Thr66Ser)) without a clear *in silico* deleterious effect, no other pathogenic mutations were identified. When we compared the distribution of normal repeat lengths between the rare T-allele and the common C-allele, we observed that alleles of at least 7 repeat units were strongly correlated with the T-allele ( $p<0.001$ ) (Figure 2B), corresponding with allele group 3 of the trimodal frequency distribution. We defined this group (7-24 units) as intermediate repeat alleles. When we recalculated genetic association of rs2814707 with disease after excluding individuals homozygous for intermediate alleles, the residual association disappeared ( $p=0.121$ ) (Supp. Table S4). However, we could not demonstrate direct association of these intermediate alleles with disease (Supp. Table S5). Though, a slight increase in frequency of intermediate alleles (25.1%) was observed compared to control individuals (22.6%) (Supp. Table S5). Also homozygous intermediate repeat carriers were slightly more frequent in the patient group (6.1% versus 4.6%). We could not observe evidence in favor of the hypothesis that intermediate repeats are unstable and might trigger pathological  $G_4C_2$  expansion, as the younger generation in 20 families of index patients with an intermediate repeat allele up to 22 units, did not show longer intermediate repeat alleles or pathological expansions.

## Reporter gene analysis

To evaluate the effect of repeat length on *C9orf72* promoter activity, we measured the relative luciferase activity (RLA) of constructs containing a *C9orf72* promoter fragment with 2, 9, 17 and 24 units. We demonstrated a highly significant decrease of the RLA between fragments with 9, 17 and 24 units compared to the wild type allele comprising 2 units ( $p < 0.001$ ) with a maximum decrease of 52% in the 24 units containing promoter (Figure 3). These data show that intermediate repeat alleles result in a significantly reduced *C9orf72* promoter activity compared to wild type alleles.

## DISCUSSION

In the present study, we have assessed the geographical distribution of the pathological G<sub>4</sub>C<sub>2</sub> expansions in an extended pan-European patient cohort of FTLD patients, originating from Italy, Germany, Portugal, Sweden, Spain, Czech Republic, Bulgaria, Austria, and Belgium, ascertained within a newly formed, European Early-Onset Dementia (EOD) consortium. The prevalence of pathological G<sub>4</sub>C<sub>2</sub> expansions in the total European cohort (73/845, 8.64%) as well as in the FTLD-only (50/781, 6.40%) and FTLD-ALS (23/64, 35.94%) subgroups, was comparable to frequencies others and we published in the three original gene identification reports on *C9orf72* [Dejesus-Hernandez et al., 2011; Renton et al., 2011; Gijssels et al., 2012].

The FTLD syndrome in the European G<sub>4</sub>C<sub>2</sub> expansion carriers was most often characterized by behavioral disturbances (95.7%), which was in line with our observations from an in-depth genotype-phenotype correlation study in the Belgian *C9orf72* carriers [Van Langenhove et al., 2012b]. Thirty-two percent of expansion carriers also developed concomitant ALS symptomatology. Further, we observed a wide range in onset age and disease duration, suggestive of genetic and/or environmental (epigenetic) disease modifiers [Van Langenhove et al., 2012b]. Of 13 G<sub>4</sub>C<sub>2</sub> expansion carriers, post-mortem neuropathological diagnosis confirmed brain deposition of TDP-43 inclusions, explaining 30.23% (13/43) of TDP-43 positive patients. Because up to 70% of the TDP pathology patients remained unresolved after *C9orf72* analysis, we performed mutation profiling of the other TDP-43 associated genes *GRN*, *VCP*, and

*TARDBP*. This revealed one additional *TARDBP* mutation, p.Ile383Val, in a patient from the Barcelona Brain Bank. Interestingly, this mutation located in the C-terminal of the protein was reported once before in one US ALS family [Rutherford et al., 2008], yet in this Spanish patient – diagnosis SD, age at onset 60 years, age at death 75 years, father with dementia - no clinical nor pathological signs of MND were recorded. Taken together, of the 43 TDP-43 pathology patients 67.44% (29/43) remained unresolved by the known genes. In the 17 familial TDP patients, this represented 58.82% (10/17) of patients without a known mutation, pointing to at least one, but likely more, other TDP-43-associated genes yet to be discovered. Eleven *C9orf72* positive TDP-43 pathology carriers were submitted to further in-depth histopathological evaluation. Although the neuropathological signature of *C9orf72* has proven to be more diverse than initially anticipated (reports ranged from strong TDP proteinopathy reminiscent of FTLD-TDP type A [Murray et al., 2011; Snowden et al., 2012; Stewart et al., 2012], to complete absence of TDP-43 pathology [Murray et al., 2011; Gijssels et al., 2012] (reviewed in Cruts et al., Trends Neurosci, under review), patients in our study all fitted the criteria of TDP proteinopathy type B [Mackenzie et al., 2010], albeit with relatively low lesion load. Eight of the 11 patients displayed concomitant AD-pathology, yet, AD stages were relatively mild (Braak stage I to III for neurofibrillary tangle (NFT) pathology and Stage B for  $\beta$ -amyloid pathology) and probably this finding was age related more than related to the *C9orf72* expansion. In the five cases stained with p62 antibody, 100% showed immunoreactive NCI in the dentate gyrus of the hippocampus, in the CA4 and CA23 pyramidal neurons of the hippocampus, and the granular cortical layer of the cerebellum. As this lesion load is far more pronounced than the TDP-43 lesions, it can be assumed that the presence of p62 pathology is a highly distinctive feature of *C9orf72*-associated FTLD [Al-Sarraj et al., 2011]. In another recent large-scale screening of the *C9orf72* repeat expansion by Majounie and colleagues [Majounie et al., 2012], frequencies were reported for a total of 4448 ALS patients and 1425 FTLD patients from European and US Caucasian populations, and compared to smaller sets of patients from other ethnic backgrounds. Of interest with respect to the present study, the Majounie study included FTLD patients from the UK, the Netherlands, France, Finland, Germany, Sardinia, and Sweden. Except for a small number of German



(n=29) and Swedish (n=7) patients, these countries were not represented in our study and vice versa. Further, a study conducted in Denmark identified 10 *C9orf72* expansion carriers in a cohort of 82 Danish FTLN patients [Lindquist et al., 2012]. To get an even more representative picture of the prevalence and distribution of the *C9orf72* repeat expansion in Western Europe we performed a meta-analysis of the FTLN cohorts from our European EOD consortium study (n=845), our previously reported Flanders-Belgian cohort (n=360) [Gijssels et al., 2012], the Majounie study (n=1381) [Majounie et al., 2012], and the Lindquist study (n=82) [Lindquist et al., 2012]. This resulted in prevalence numbers on 2668 FTLN patients from 15 European countries (Table 3, countries with patients groups smaller than 20 were not included in the meta-analysis). The overall frequency of the *C9orf72* pathological G<sub>4</sub>C<sub>2</sub> expansion in Western Europe was 9.98% in total FTLN, 18.52% in familial FTLN, and 6.26% in sporadic FTLN. Looking at the per country distribution, Belgium, Italy, The Netherlands, Portugal, and the UK showed an average total prevalence ranging from 6.09% to 10.29%. Denmark showed a slightly increased prevalence of 12.20%, and in France the prevalence was about double the average prevalence with 18.00%. In line with the hypothesis of a Scandinavian founder effect for the chromosome 9 – *C9orf72* haplotype [Mok et al., 2011], frequencies were highest in Finland (29.33%) and high in Sweden (20.73%). Yet, in contrast with this hypothesis and the expected North-South axis for the founder haplotype prevalence, frequencies for *C9orf72* reached 25.49% in Spain. To the other extreme, prevalence in Germany was just 4.82%. Of note, Ireland, Luxembourg, Switzerland, Poland, , Norway, and Greece are not yet represented in this meta-analysis of Western Europe.

We exactly sized the normal range of repeat alleles in both the Flanders-Belgian and European cohorts using two different assays and observed a range between 2 and 24 units. The allele frequencies showed a trimodal distribution with a long tail containing the rare longer alleles. Alleles in group 3 of this trimodal distribution (intermediate alleles; 7 to 24 units) are almost solely present on the chromosome 9 risk haplotype tagged by the rs2814707 T-allele. We asked the question whether these alleles might act as a risk factor on FTLN disease susceptibility or could have a mild effect on gene expression. Although intermediate repeat alleles were not significantly associated with disease in the Flanders-Belgian cohort,

the slight increase in frequency in patients compared to control individuals explained the residual association we had observed for homozygous carriers of the rs2814707 T-allele (OR = 1.75; 95% CI 1.02 - 3.01; p = 0.042). A group of 16 FTLN-ALS patients without pathological expansion were included in this cohort. Although this is a small group, it is remarkable that 7 of 32 alleles (21.9%) have at least 10 units, which is significantly higher than 7.2% in control individuals (OR = 3.685; 95% CI 1.560 - 8.706; p = 0.003). This suggests that intermediate alleles might rather have a risk effect in diseases involving ALS. Since the European cohort was collected for a frequency study of the repeat expansion, we did not have ethnicity matched control individuals for an association study.

Flanders-Belgian and European patient carriers of intermediate repeats (7 to 24) did not show an earlier onset age compared to carriers of shorter repeats (2 to 6) (data not shown). Also, the variable onset age in patients with a pathological G<sub>4</sub>C<sub>2</sub> expansion [Gijssels et al., 2012], could not be explained by a modifying effect of number of repeat units in the wild-type allele (data not shown), in contrast to what has been shown for patients with Huntington's disease [Aziz et al., 2009].

The mechanism by which the G<sub>4</sub>C<sub>2</sub> intermediate repeats might contribute to disease pathogenesis remained elusive. In this study, we performed reporter gene expression studies to evaluate the effect of intermediate repeats according to their size on *C9orf72* promoter activity. We provided compelling evidence that transcriptional activity of the *C9orf72* promoter significantly decreases with an increasing amount of repeat units with a maximal reduction of 52% of a 24 units containing promoter compared to 2 units on the wild type allele (Figure 3). To better discriminate the cut-off from which decreased transcriptional activity is apparent, more different length alleles should be investigated, although from these data it is clear that intermediate alleles affect the normal transcriptional activity of the *C9orf72* promoter. With the cloned intermediate alleles we did not reach an expression level of virtually 0 as expected in pathogenic expansion alleles. Therefore, individuals with the number of units higher than 24 are needed to better interpret the grey zone between normal and expanded alleles. These *in vitro* data confirms that the G<sub>4</sub>C<sub>2</sub> repeat is located in the promoter region of *C9orf72* as we previously suggested [Gijssels et al., 2012]. It also favors the loss-of-function hypothesis that was based on 50% reduction of

*C9orf72* levels in brain of pathological expansion carriers preventing expression of the mutant allele [Gijssels et al., 2012]. However, in this study, we did not identify other simple sequence mutations or deletions/duplications pointing to a loss-of-function mechanism, except for one patient-specific missense mutation (p.Thr66Ser) without clear pathogenic nature. Therefore, other molecular mechanisms might also contribute to disease including sequestration of RNA-binding proteins and RNA toxicity [DeJesus-Hernandez et al., 2011], which can also lead to a decrease of the mRNA level. These two potential disease mechanisms are not mutually exclusive, and might also lead together to degeneration of neuronal populations in either the frontal cortex (FTLD) or the spinal cord (ALS).

By what mechanisms the G<sub>4</sub>C<sub>2</sub> repeat expands to a pathological size range remains to be discovered. In our first *C9orf72* study in the Flanders-Belgian cohort [Gijssels et al., 2012], we showed that the majority of the pathological G<sub>4</sub>C<sub>2</sub> expansions were located on the same risk haplotype tagged by the rare T-allele of SNP rs2814707, and strongly associated with FTLD and ALS. These observations were confirmed in the European study with 100% of European carriers of a pathological G<sub>4</sub>C<sub>2</sub> expansion carrying at least one T-allele. It has been proposed that this tight genetic association might be explained by a single founder mutation on this risk haplotype [Mok et al., 2011; Majounie et al., 2012]. An alternative hypothesis is that a specific genomic context on this risk haplotype is rendering the G<sub>4</sub>C<sub>2</sub> repeat less stable and making it more prone to expansion into the pathological size range. In this context, we and others [DeJesus-Hernandez et al., 2011] showed that the G<sub>4</sub>C<sub>2</sub> intermediate repeats are also highly significantly overrepresented on the same risk haplotype suggesting that these intermediate repeats are more prone to replication slippage and unstable inheritance and thus triggering pathological expansions. The observation that intermediate alleles have the same linkage disequilibrium with a specific SNP allele as the expanded alleles, is also previously made in other repeat expansion disorders, e.g. Fragile X Syndrome [Gunter et al., 1998]. The appearance of pathological G<sub>4</sub>C<sub>2</sub> expansions in apparently sporadic patients in this and previous studies [DeJesus-Hernandez et al., 2011; Renton et al., 2011; Gijssels et al., 2012], seems to support this hypothesis. However, we could not observe increasing G<sub>4</sub>C<sub>2</sub> repeat length or *de novo* expansions in the younger generations of families of index patients with intermediate repeat

length up to 22 units. Therefore, intermediate alleles might rather be considered as predisposing alleles for further stepwise expansion over probably many generations instead of pre-mutations. A study using sporadic ALS trios did also not show evidence of repeat instability between two generations [Pamphlett et al., 2012].

Further, we observed in patients of both the Flanders-Belgian and European cohorts with a pathological  $G_4C_2$  expansion, a significantly higher frequency of short indels in the GC-rich LCS adjacent to the  $G_4C_2$  repeat (24/84, 28.57%) compared to patient non-carriers (5/404, 1.24%;  $p < 0.001$ ) and control persons (4/752, 0.53%;  $p < 0.001$ ). Remarkably, in 23.81% (20/84) of repeat expansion carriers a deletion of the GTGGT motif contiguous with the  $G_4C_2$  repeat was comprised within the indel (Figure 1), compared to 0.25% in the non-carrier patients and 0.13% in control individuals (Table 1). This GTGGT deletion joins the two GC-rich sequences, increasing the overall GC-content of the LCS. This has already been reported for CAG expansion diseases, where the expandability of the repeat increases with a higher GC-content of the surrounding DNA sequence [Brock et al., 1999; Nestor and Monckton, 2011]. Also, the GTGGT deletion creates one long contiguous imperfect  $G_4C_2$  repeat which is likely more prone to replication slippage. Therefore, the GTGGT sequence is potentially an important stabilizer of the  $G_4C_2$  repeat. This resembles the loss of AGG interspersions in unstable Fragile X alleles [Kunst and Warren, 1994; Larsen et al., 2000]. Moreover, all carriers of a GTGGT deletion in the LCS, had at least one rs2814707 T-allele and the 10-bp deletion was co-transmitted with the  $G_4C_2$  expansion in relatives of two expansion carriers, which indicates that it is likely located on the risk haplotype and might trigger the expansion.

Nevertheless, it should be noted that not all carriers of a pathological  $G_4C_2$  expansion also have an LCS indel and thus we cannot exclude that the indels are rather a consequence of the close neighborhood of the expansion destabilizing the genomic region. Of note, the indels found in non-expansion carriers are mostly not affecting the GTGGT sequence but rather stabilizing the repeat by deleting 100% GC-rich sequence. The GTGGT deletion in one FTLN patient without repeat expansion is located on a short allele with 5 units, which is probably not unstable enough to expand. A GTGGT deletion on an intermediate allele will most likely be not stable and hence not observed.

A better understanding of the underlying mechanism causing the G<sub>4</sub>C<sub>2</sub> instability will be essential to assess disease risk and improve clinical benefits. Further, elucidating how G<sub>4</sub>C<sub>2</sub> expansions lead to disease will be crucial to unveil biological pathways and key molecules in the disease process as targets for future therapies.

UNCORRECTED ACCEPTED ARTICLE

## ACKNOWLEDGMENTS

The authors are grateful to the participants contributing to this study from within the European EOD consortium and the BELNEU consortium. Also, we thank the personnel of the Genetic Service Facility of VIB (<http://www.vibgeneticservicefacility.be>) for their support of the genetic analyses. The data generation for this paper was in part funded by the Interuniversity Attraction Poles (IAP) program P6/43 of the Belgian Science Policy Office; the Europe Initiative on Centers of Excellence in Neurodegeneration (CoEN) and the Methusalem program supported by the Flemish Government; the Foundation for Alzheimer Research (SAO/FRMA); the Medical Foundation Queen Elisabeth; the Research Foundation Flanders (FWO); the Agency for Innovation by Science and Technology Flanders (IWT) and the Special Research Fund of the University of Antwerp, Belgium. *Czech sample collection*: the IGA NT 12094-5/2011 grant from the Czech Ministry of Health; the Cassa di Risparmio PT e Pescaia (n. 2011.0264). *Italy-Verona sample collection*: S.T. is a recipient of a post-doctoral research fellowship funded by Fondazione CariVerona (grant No. 2009.1026 'Cognitive and behavioral disability in Dementias and Psychosis'). *Swedish sample collections*: The contributors of the Swedish cohort acknowledges the financial support by Swedish Brain Power, Swedish Research Council, the King Gustaf V and Queen Victoria's Foundation of Freemasons and the foundations of Marianne and Marcus Wallenberg, Knut and Alice Wallenberg, Gun and Bertil Stohne, Gamla tjänarinnor, and Swedish Alzheimer Foundation. The IWT provided a PhD fellowship to T.V.L. and the FWO to S.P., postdoctoral fellowships to I.G., and J.v.d.Z. were granted by the FWO.

### Conflict of interest

The authors report no conflict of interest with regard to the reported findings.

## APPENDIX

European EOD Consortium members that contributed to the data in this paper:

*Department of Molecular Genetics, VIB, Antwerp, Belgium, Neurodegenerative Brain Diseases Group:*

Julie van der Zee, Ilse Gijselinck, Lubina Dillen, Tim Van Langenhove, Jessie Theuns, Stéphanie Philtjens, Kristel Slegers, Veerle Bäumer, Githa Maes, Ellen Corsmit, Marc Cruets, Christine Van Broeckhoven; *Neurogenomics Group:* Albena Jordanova

*Institute Born-Bunge, University of Antwerp, Antwerp, Belgium, Laboratory of Neurogenetics:* Julie van der Zee, Ilse Gijselinck, Lubina Dillen, Tim Van Langenhove, Stéphanie Philtjens, Jessie Theuns, Kristel Slegers, Veerle Bäumer, Githa Maes, Marc Cruets, Christine Van Broeckhoven; *Laboratory of Neurochemistry and Behavior:* Sebastiaan Engelborghs, Peter P. De Deyn; *Laboratory of Neurobiology:* Patrick Cras

*Department of Neurology and Memory Clinic, Hospital Network Antwerp, Middelheim and Hoge Beuken, Antwerp, Belgium:* Sebastiaan Engelborghs, Peter P. De Deyn

*University of Leuven, Leuven, Belgium, Brain and Emotion Laboratory:* Mathieu Vandenbulcke; *Laboratory for Cognitive Neurology:* Rik Vandenberghe

*University Hospitals Leuven Gasthuisberg, Leuven, Belgium, Department of Psychiatry:* Mathieu Vandenbulcke; *Department of Neurology:* Rik Vandenberghe

*Centre for Ageing Brain and Neurodegenerative Disorders, Neurology Unit, University of Brescia, Brescia, Italy:* Barbara Borroni, Alessandro Padovani, Silvana Archetti

*Department of Psychiatry and Psychotherapy, Technische Universität München, Munich, Germany:*

Robert Perneczky, Janine Diehl-Schmid

*Department of Neurodegeneration, Hertie Institute for Clinical Brain Research and Centre of Neurology,*

*Tuebingen, Germany:* Matthis Synofzik, Walter Maetzler, Jennifer Müller vom Hagen, Ludger Schöls

*German Research Center for Neurodegenerative Diseases (DZNE), University of Tuebingen, Germany:*

Matthis Synofzik, Walter Maetzler, Jennifer Müller vom Hagen, Ludger Schöls; *University of Bonn,*

*Germany:* Michael T. Heneka, Frank Jessen, Alfredo Ramirez, Delja Kurzwelly, Carmen Sachtleben,

Wolfgang Mairer

*Institute of Molecular Medicine, Faculty of Medicine, University of Lisbon, Lisbon, Portugal:* Alexandre

de Mendonça, Gabriel Miltenberger-Miltenyi, Sónia Pereira, Clara Firmo

*Faculty of Medicine, University of Lisbon, Lisbon, Portugal:* José Pimentel

*Alzheimer's disease and Other cognitive disorders unit, Department of Neurology, Hospital Clinic,*

*Barcelona, Spain:* Raquel Sanchez-Valle, Albert Llado, Anna Antonell, Jose Molinuevo

*Biobanc, Hospital Clinic, Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS),*

*Barcelona, Spain:* Ellen Gelpi

*KI-Alzheimer Disease Research Center, Karolinska Institutet, Stockholm, Sweden, Department of*

*Neurobiology, Care Sciences and Society:* Caroline Graff, Huei-Hsin Chiang, Marie Westerlund;

*Department of NVS:* Charlotte Forsell, Lena Lillius



*Karolinska University Hospital, Stockholm, Sweden, Department of Geriatric Medicine, Genetics unit:*  
Caroline Graff, Anne Kinhult Ståhlbom, Håkan Thonberg; *Department clinical cytology and pathology:*  
Inger Nennesmo

*Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska  
Academy, University of Gothenburg:* Anne Börjesson-Hanson

*Department of Neurological and Psychiatric Sciences, University of Florence, Florence, Italy:* Benedetta  
Nacmias, Silvia Bagnoli, Sandro Sorbi, Valentina Bessi, Irene Piaceri

*Neurology Department, Centro Hospitalar Universitário de Coimbra, Coimbra, Portugal:* Isabel Santana,  
Beatriz Santiago

*University of Coimbra, Coimbra, Portugal, Faculty of Medicine:* Isabel Santana, Maria Helena Ribeiro;  
*Center for Neuroscience and Cell Biology:* Maria Rosário Almeida, Catarina Oliveira

*Department of Neurology Centro Hospitalar de São João, and Department of Clinical Neuroscience and  
Mental Health, Faculty of Medicine University of Porto, Porto, Portugal:* João Massano, Carolina Garret

*Hospital de Angra do Heroísmo, Azores, Portugal:* Paula Pires

*Laboratory of Epidemiology Neuroimaging & Telemedicine & National Centre for Alzheimer's and  
Mental Diseases, IRCCS Centro San Giovanni di Dio, FBF, Brescia, Italy:* Giovanni Frisoni, Orazio  
Zanetti, Cristian Bonvicini

*Medical University Sofia, Sofia, Bulgaria, Department of Neurology: Stayko Sarafov, Ivailo Tournev;*

*Molecular Medicine Center, Department of Biochemistry: Albena Jordanova*

*Department of Cognitive Science and Psychology, New Bulgarian University, Sofia, Bulgaria: Ivailo*

*Tournev*

*Institute of Neurology, Medical University Vienna, Vienna, Austria: Gabor G. Kovacs, Thomas Ströbel*

*University of Bonn, Germany, Department of Neurology, Clinical Neuroscience Unit: Michael T. Heneka,*

*Frank Jessen, Alfredo Ramirez, Delia Kurzwelly, Carmen Sachtleben, Wolfgang Mairer; Department of  
Psychiatry: Frank Jessen*

*Department of Pathology and Molecular Medicine, Thomayer Hospital, Prague, Czech Republic:*

*Radoslav Matej, Eva Parobkova*

*Ludwig-Maximilians-Universität Munich, Germany, Neurologische Klinik: Adrian Danel; Zentrum für*

*Neuropathologie und Prionforschung: Thomas Arzberger*

*Section of Neuropathology, Department of Neurological, Neuropsychological, Morphological and*

*Movement Sciences, University of Verona, Italy: Gian Maria Fabrizi, Silvia Testi*

*Unità Operativa Complessa di Neuropatologia d'U, Azienda Ospedaliera Universitaria Integrata di*

*Verona: Sergio Ferrari, Tiziana Cavallaro*

*Cyclotron Research Centre and Department of Neurology, University of Liège, Liège, Belgium: Eric*

*Salmon*

*Department of Neurology, University Hospital Ghent and University of Ghent, Ghent, Belgium: Patrick Santens*

*Department of Neurology, Antwerp University Hospital, Edegem, Belgium: Patrick Cras*

UNCORRECTED ACCEPTED ARTICLE

## REFERENCES

- Al-Sarraj S, King A, Troakes C, Smith B, Maekawa S, Bodi I, Rogelj B, Al-Chalabi A, Hortobagyi T, Shaw CE. 2011. p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLN and MND/ALS. *Acta Neuropathol* 122:691-702.
- Arighi A, Fumagalli GG, Jacini F, Fenoglio C, Ghezzi L, Pietroboni AM, De RM, Serpente M, Ridolfi E, Bonsi R, Bresolin N, Scarpini E et al. 2012. Early Onset Behavioral Variant Frontotemporal Dementia due to the C9ORF72 Hexanucleotide Repeat Expansion: Psychiatric Clinical Presentations. *J Alzheimers Dis*.
- Aziz NA, Jurgens CK, Landwehrmeyer GB, van Roon-Mom WM, van Ommen GJ, Stijnen T, Roos RA. 2009. Normal and mutant HTT interact to affect clinical severity and progression in Huntington disease. *Neurology* 73:1280-1285.
- Bigio EH. 2012. Motor neuron disease: the C9orf72 hexanucleotide repeat expansion in FTD and ALS. *Nat Rev Neurol* 8:249-250.
- Boeve BF, Boylan KB, Graff-Radford NR, DeJesus-Hernandez M, Knopman DS, Pedraza O, Vemuri P, Jones D, Lowe V, Murray ME, Dickson DW, Josephs KA et al. 2012. Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72. *Brain* 135:765-783.
- Boxer AL, Mackenzie IR, Boeve BF, Baker M, Seeley WW, Crook R, Feldman H, Hsiung GY, Rutherford N, Laluz V, Whitwell J, Foti D et al. 2011. Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. *J Neurol Neurosurg Psychiatry* 82:196-203.

- Brock GJ, Anderson NH, Monckton DG. 1999. Cis-acting modifiers of expanded CAG/CTG triplet repeat expandability: associations with flanking GC content and proximity to CpG islands. *Hum Mol Genet* 8:1061-1067.
- Chio A, Borghero G, Restagno G, Mora G, Drepper C, Traynor BJ, Sendtner M, Brunetti M, Ossola I, Calvo A, Pugliatti M, Sotgiu MA et al. 2012. Clinical characteristics of patients with familial amyotrophic lateral sclerosis carrying the pathogenic GGGGCC hexanucleotide repeat expansion of C9ORF72. *Brain* 135:784-793.
- DeJesus-Hernandez M, Mackenzie IR, Boeve C, Boxer AL, Baker M, Rutherford NJ, Nicholson AM, Finch NA, Flynn H, Adamson J, Kouri N, Wojtas A et al. 2011. Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* doi:10.1016/j.neuron.2011.09.011.
- Ferrari R, Mok K, Moreno JH, Cosentino S, Goldman J, Pietrini P, Mayeux R, Tierney MC, Kapogiannis D, Jicha GA, Murrell JR, Ghetti B et al. 2012. Screening for C9ORF72 repeat expansion in FTL. *Neurobiol Aging* 33:1850.
- Gijssels I, Engelborghs S, Maes G, Cuijt I, Peeters K, Mattheijssens M, Joris G, Cras P, Martin JJ, De Deyn PP, Kumar-Singh S, Van Broeckhoven C et al. 2010. Identification of 2 Loci at chromosomes 9 and 14 in a multiplex family with frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Arch Neurol* 67:606-616.
- Gijssels I, Van Langenhove T, van der Zee J, Slegers K, Philtjens S, Kleinberger G, Janssens J, Bettens K, Van Cauwenberghe C, Pereson S, Engelborghs S, Sieben A et al. 2012. 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* 11:54-65.

- Gunter C, Paradee W, Crawford DC, Meadows KA, Newman J, Kunst CB, Nelson DL, Schwartz C, Murray A, Macpherson JN, Sherman SL, Warren ST. 1998. Re-examination of factors associated with expansion of CGG repeats using a single nucleotide polymorphism in FMR1. *Hum Mol Genet* 7:1935-1946.
- Hsiung GY, Dejesus-Hernandez M, Feldman HH, Sengdy P, Bouchard-Kerr P, Dwosh E, Butler R, Leung B, Fok A, Rutherford NJ, Baker M, Rademakers R et al. 2012. Clinical and pathological features of familial frontotemporal dementia caused by C9ORF72 mutation on chromosome 9p. *Brain* 135:709-722.
- Kumps C, Van RN, Heyrman L, Goossens D, Del-Favero J, Noguera R, Vandesompele J, Speleman F, De PK. 2010. Multiplex Amplicon Quantification (MAQ), a fast and efficient method for the simultaneous detection of copy number alterations in neuroblastoma. *BMC Genomics* 11:298.
- Kunst CB, Warren ST. 1994. Cryptic and polar variation of the fragile X repeat could result in predisposing normal alleles. *Cell* 77:853-861.
- Laaksovirta H, Peuralinna T, Schymick JC, Scholz SW, Lai SL, Myllykangas L, Sulkava R, Jansson L, Hernandez DG, Gibbs JR, Nalls MA, Heckerman D et al. 2010. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol* 9:978-985.
- Larsen LA, Armstrong JS, Gronskov K, Hjalgrim H, Macpherson JN, Brondum-Nielsen K, Hasholt L, Norgaard-Pedersen B, Vuust J. 2000. Haplotype and AGG-interspersion analysis of FMR1 (CGG)(n) alleles in the Danish population: implications for multiple mutational pathways towards fragile X alleles. *Am J Med Genet* 93:99-106.

- Le Ber I, Camuzat A, Berger E, Hannequin D, Laquerriere A, Golfier V, Seilhean D, Viennet G, Couratier P, Verpillat P, Heath S, Camu W et al. 2009. Chromosome 9p-linked families with frontotemporal dementia associated with motor neuron disease. *Neurology* 72:1669-1676.
- Lindquist S, Duno M, Batbayli M, Puschmann A, Braendgaard H, Mardosiene S, Svenstrup K, Pinborg L, Vestergaard K, Hjermand L, Stokholm J, Andersen B et al. 2012. Corticobasal and ataxia syndromes widen the spectrum of C9ORF72 hexanucleotide expansion disease. *Clin Genet* 9999.
- Luty AA, Kwok JB, Thompson EM, Blumbergs P, Brooks WS, Loy CT, Dobson-Stone C, Panegyres PK, Hecker J, Nicholson GA, Halliday GM, Schofield PR. 2008. Pedigree with frontotemporal lobar degeneration--motor neuron disease and Tar DNA binding protein-43 positive neuropathology: genetic linkage to chromosome 9. *BMC Neurol* 8:32.
- Mackenzie IR, Neumann M, Bigio EH, Cairns NJ, Alafuzoff I, Kril J, Kovacs GG, Ghetti B, Halliday G, Holm IE, Ince PG, Kamphorst W et al. 2010. Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update. *Acta Neuropathol* 119:1-4.
- Mahoney CJ, Beck J, Rohrer JD, Lashley T, Mok K, Shakespeare T, Yeatman T, Warrington EK, Schott JM, Fox NC, Rossor MN, Hardy J et al. 2012. Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features. *Brain* 135:736-750.
- Majounie E, Renton AE, Mok K, Dopper EG, Waite A, Rollinson S, Chio A, Restagno G, Nicolaou N, Simon-Sanchez J, Van Swieten JC, Abramzon Y et al. 2012. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol* 11:323-330.

- Mok K, Traynor BJ, Schymick J, Tienari PJ, Laaksovirta H, Peuralinna T, Myllykangas L, Chio A, Shatunov A, Boeve BF, Boxer AL, Dejesus-Hernandez M et al. 2011. The chromosome 9 ALS and FTD locus is probably derived from a single founder. *Neurobiol Aging*.
- Morita M, Al Chalabi A, Andersen PM, Hosler B, Sapp P, Englund E, Mitchell JE, Habgood JJ, de Belleruche J, Xi J, Brown RH. 2006. A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. *Neurology* 66:839-844.
- Murray ME, Dejesus-Hernandez M, Rutherford NJ, Baker M, Duara R, Graff-Radford NR, Wszolek ZK, Ferman TJ, Josephs KA, Boylan KB, Rademakers R, Dickson DW. 2011. Clinical and neuropathologic heterogeneity of c9FTD/ALS associated with hexanucleotide repeat expansion in C9ORF72. *Acta Neuropathol* 122:673-690.
- Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, Freedman M, Kertesz A, Robert PH, Albert M, Boone K, Miller BL et al. 1998. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology* 51:1546-1554.
- Nestor CE, Monckton DG. 2011. Correlation of inter-locus polyglutamine toxicity with CAG\*CTG triplet repeat expandability and flanking genomic DNA GC content. *PLoS One* 6:e28260.
- Pamphlett R, Cheong PL, Trent RJ, Yu B. 2012. Transmission of C9orf72 hexanucleotide repeat expansions in sporadic amyotrophic lateral sclerosis: an Australian trio study. *Neuroreport* 23:556-559.
- Pearson JP, Williams NM, Majounie E, Waite A, Stott J, Newsway V, Murray A, Hernandez D, Guerreiro R, Singleton AB, Neal J, Morris HR. 2011. Familial frontotemporal dementia with amyotrophic lateral sclerosis and a shared haplotype on chromosome 9p. *J Neurol* 258:647-655.



- Renton AE, Majounie E, Waite A, Simon-Sanchez J, Rollinson S, Gibbs JR, Schymick JC, Laaksovirta H, van Swieten JC, Myllykangas L, Kalimo H, Paetau A et al. 2011. A hexanucleotide repeat expansion in C9ORF72 Is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* doi:10.1016/j.neuron.2011.09.010.
- Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365-386.
- Rutherford NJ, Zhang YJ, Baker M, Gass JM, Finch NA, Xu YF, Stewart H, Kelley BJ, Kuntz K, Crook RJ, Sreedharan J, Vance C et al. 2008. Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis. *PLoS Genet* 4:e1000193.
- Shatunov A, Mok K, Newhouse S, Weale ME, Smith B, Vance C, Johnson L, Veldink JH, van Es MA, van den Berg LH, Robberecht W, Van DP et al. 2010. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol* 9:986-994.
- Simon-Sanchez J, Dopper EG, Cohn-Hokke PE, Hukema RK, Nicolaou N, Seelaar H, de Graaf JR, de K, I, van Schoor NM, Deeg DJ, Smits M, Raaphorst J et al. 2012. The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions. *Brain* 135:723-735.
- Snowden JS, Rollinson S, Thompson JC, Harris JM, Stopford CL, Richardson AM, Jones M, Gerhard A, Davidson YS, Robinson A, Gibbons L, Hu Q et al. 2012. Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations. *Brain* 135:693-708.
- Stewart H, Rutherford NJ, Briemberg H, Krieger C, Cashman N, Fabros M, Baker M, Fok A, Dejesus-Hernandez M, Eisen A, Rademakers R, Mackenzie IR. 2012. Clinical and pathological features of

amyotrophic lateral sclerosis caused by mutation in the *C9ORF72* gene on chromosome 9p. *Acta Neuropathol.*

Troakes C, Maekawa S, Wijesekera L, Rogelj B, Siklos L, Bell C, Smith B, Newhouse S, Vance C, Johnson L, Hortobagyi T, Shatunov A et al. 2011. An MND/ALS phenotype associated with *C9orf72* repeat expansion: Abundant p62-positive, TDP-43-negative inclusions in cerebral cortex, hippocampus and cerebellum but without associated cognitive decline. *Neuropathology.*

Valdmanis PN, Kabashi E, Dion PA, Rouleau GA. 2008. ALS predisposition modifiers: knock NOX, who's there? *SOD1* mice still are. *Eur J Hum Genet* 16:140-142.

Van Deerlin VM, Sleiman PM, Martinez-Lage M, Chen-Plotkin A, Wang LS, Graff-Radford NR, Dickson DW, Rademakers R, Boeve BF, Grossman M, Arnold SE, Mann DM et al. 2010. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet* 42:234-239.

van Es MA, Veldink JH, Saris CG, Blauw HM, van Vught PW, Birve A, Lemmens R, Schelhaas HJ, Groen EJ, Huisman MH, Van Der Kooi AJ, De VM et al. 2009. Genome-wide association study identifies 19p13.3 (*UNC13A*) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet* 41:1083-1087.

Van Langenhove T, Van Broeckhoven C, van der Zee J. 2012a. The molecular basis of the frontotemporal lobar degeneration - amyotrophic lateral sclerosis spectrum. *Ann Med.*

Van Langenhove T, van der Zee J, Gijssels I, Engelborghs S, Vandenberghe R, Vandenbulcke M, De Bleeckere J, Sieben A, Versijpt J, Ivanoiu A, Deryck O, Willems C et al. 2012b. Distinct clinical characteristics of *C9orf72* expansion carriers compared to *GRN*, *MAPT* and non-mutation carriers in a Flanders-Belgian FTL cohort. *Arch Neurol.*

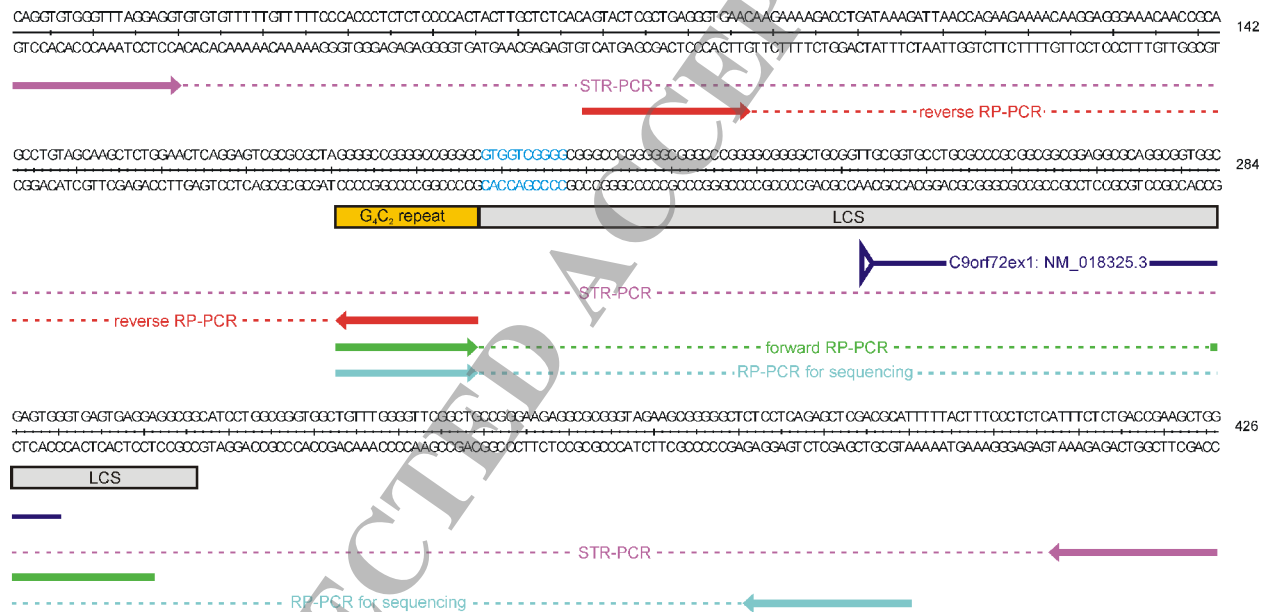
- Vance C, Al Chalabi A, Ruddy D, Smith BN, Hu X, Sreedharan J, Siddique T, Schelhaas HJ, Kusters B, Troost D, Baas F, de Jong V et al. 2006. Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. *Brain* 129:868-876.
- Weckx S, Del Favero J, Rademakers R, Claes L, Cruts M, De Jonghe P, Van Broeckhoven C, De Rijk P. 2005. novoSNP, a novel computational tool for sequence variation discovery. *Genome Res* 15:436-442.
- Whitwell JL, Weigand SD, Boeve BF, Senjem ML, Gunter JL, DeJesus-Hernandez M, Rutherford NJ, Baker M, Knopman DS, Wszolek ZK, Parisi JE, Dickson DW et al. 2012. Neuroimaging signatures of frontotemporal dementia genetics: C9ORF72, tau, progranulin and sporadics. *Brain* 135:794-806.

UNCORRECTED ACCEPTED ARTICLE

## FIGURE LEGENDS

### Figure 1. Genotyping assays to characterize the *C9orf72* region and G<sub>4</sub>C<sub>2</sub> repeat

The *C9orf72* G<sub>4</sub>C<sub>2</sub> repeat (yellow box) is located upstream of the first exon of isoform NM\_018325.3 (dark blue arrow) and adjacent to a GC-rich low-complexity sequence (LCS; light grey box) with their nucleotide sequences shown above. The sequence of the recurrent 10-bp deletion g.26747\_26756delGTGGTCTGGGG (Table 4, Supp. Figure S1), we observed in the LCS, is indicated in blue. Below the sequence, the primers with their corresponding PCR amplicons are shown for each of the PCR genotyping assays: STR-PCR in pink, forward RP-PCR in green, reverse RP-PCR in red and RP-PCR for sequencing in blue.

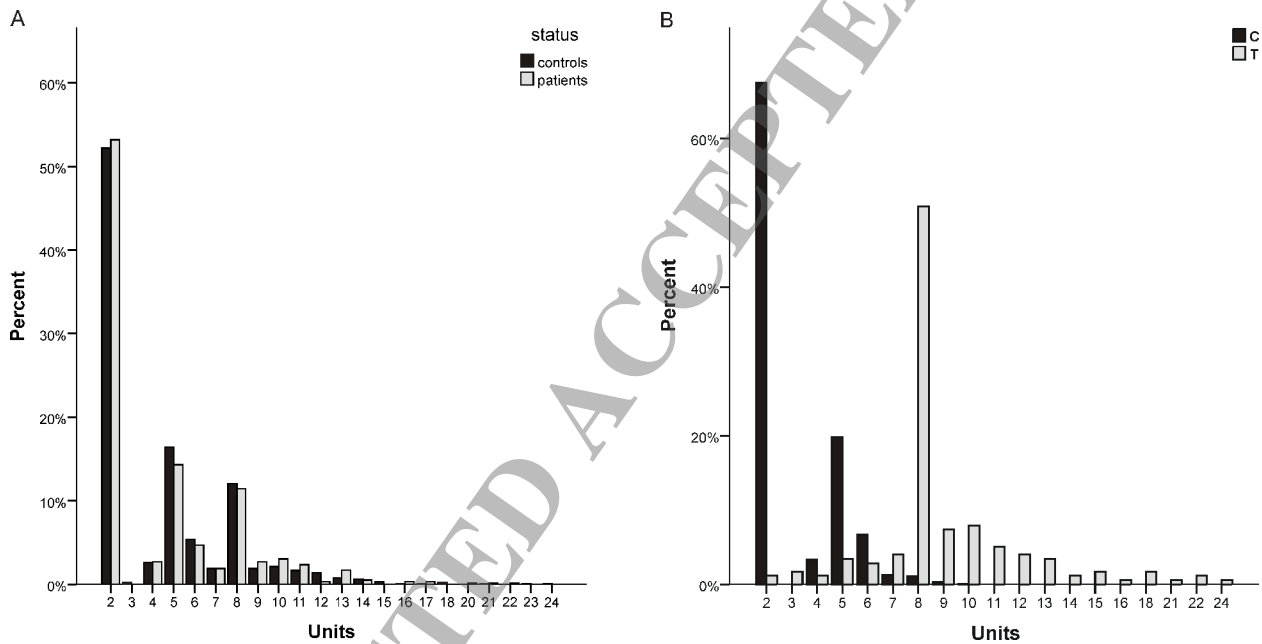


**Figure 2A. Distribution of normal repeat lengths in the Flanders-Belgian patients and control individuals.**

Histograms of G<sub>4</sub>C<sub>2</sub> repeat units sized < 60 repeats in Flanders-Belgian patients, excluding patients with mutations in known causal genes or with a pathological G<sub>4</sub>C<sub>2</sub> expansion, compared to control individuals.

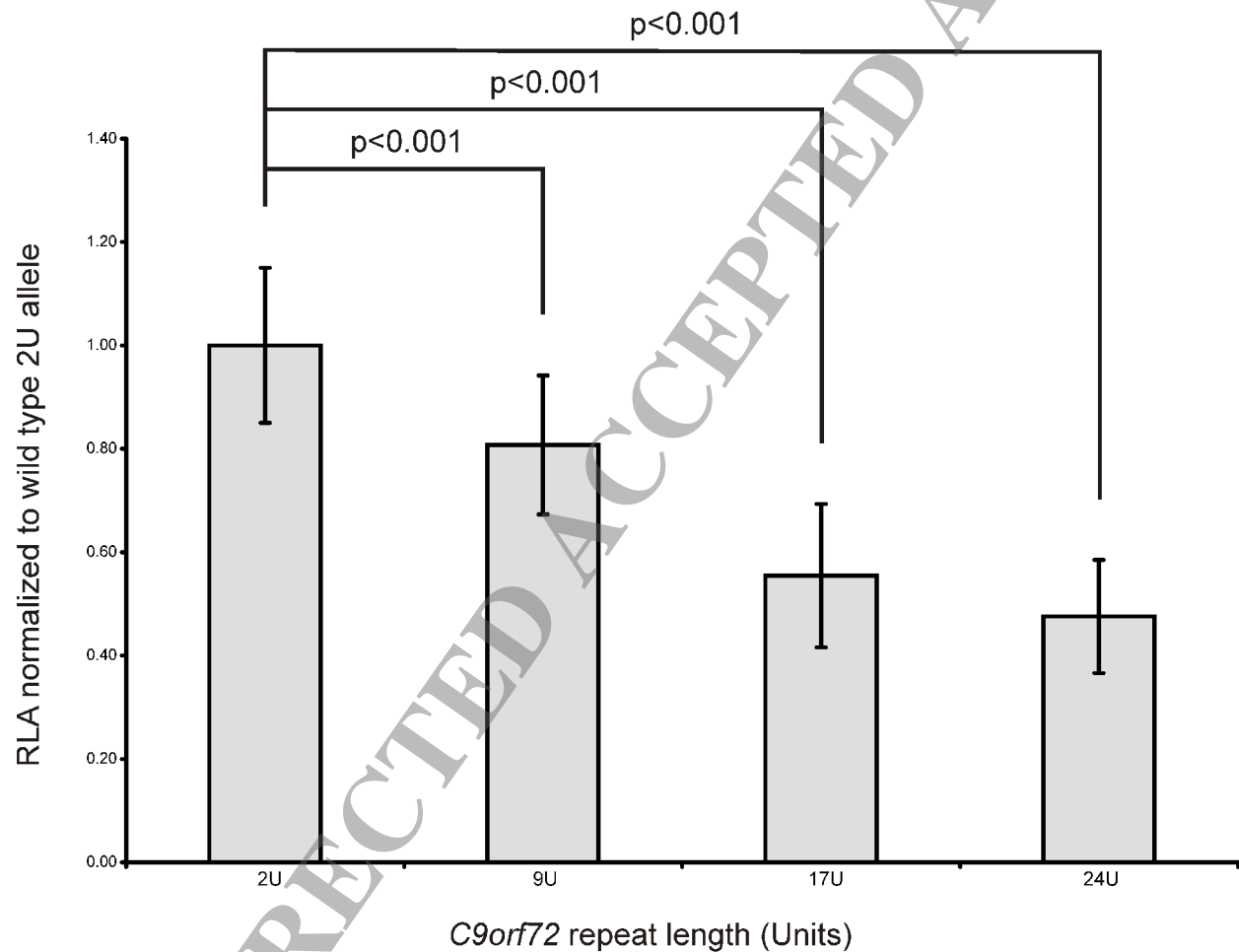
**Figure 2B. Correlation of normal repeat lengths with rs2814707 alleles**

Histograms of G<sub>4</sub>C<sub>2</sub> repeat units in 610 control individuals homozygous for the rs2814707 C-allele and 53 homozygous for the rs2814707 T-allele.



**Figure 3. Transcriptional activity of *C9orf72* promoter with alleles of different repeat length**

Bars represent relative *Gaussia/Cypridina* luciferase activities (RLA) for the different *C9orf72* constructs compared to the wild type allele of 2 units, for an increasing amount of repeat units. Values represent the mean ( $\pm$ SDEV) of 36 independent measurements relative to the 2 units wild type allele. The significance of differences in expression was calculated using the Mann-Whitney U test. P-values are presented above the bars.



**Table 1. Descriptive characteristics of the European and Flanders-Belgian cohorts**

<b>Clinical diagnosis cohorts</b>	<b>Total n</b>	<b>Familial n (%)</b>	<b>Disease onset ± SD (range)</b>	<b>Pathology n (%)</b>
<b>European Cohort</b>				
<b>Total</b>	845	274 (32.43)	62.5 ± 9.0 (28-88)	45 (5.33)
<b>FTLD</b>	781	251 (32.14)	62.7 ± 9.0 (28-88)	28 (3.59)
<b>FTLD-ALS</b>	64	23 (35.94)	60.9 ± 9.8 (31-83)	17 (26.56)
<b>Other <sup>(1)</sup></b>	72	40 (55.56)	60.0 ± 12.1 (35-95)	N.A.
<b>Flanders-Belgian Cohort</b>				
<b>Total</b>	360	108 (30.00)	62.9 ± 9.7 (29-85)	24 (6.67)
<b>FTLD</b>	337	101 (29.97)	63.0 ± 9.7 (29-85)	21 (6.23)
<b>FTLD-ALS</b>	23	7 (30.43)	62.7 ± 10.0 (39-75)	3 (13.04)
<b>Combined European and Flanders-Belgian cohorts</b>				
<b>Total</b>	1205	382 (31.70)	62.7 ± 9.2 (28-88)	70 (5.81)
<b>FTLD</b>	1118	352 (31.48)	62.8 ± 9.2 (28-88)	49 (4.38)
<b>FTLD-ALS</b>	87	30 (34.48)	61.4 ± 9.8 (31-83)	21 (24.14)

Note: <sup>(1)</sup>This group of other 72 patients was not included in the total since they did not fulfill the criteria for possible or probable diagnosis.

**Table 2. Frequencies of the *C9orf72* pathological G<sub>4</sub>C<sub>2</sub> expansion in the European and Flanders-Belgian cohorts**

	FTLD		FTLD-ALS		Total	
<b>European Cohort</b>						
<b>Total</b>	50/781	6.40%	23/64	35.94%	73/845	8.64%
<b>Familial</b>	21/251	8.37%	11/23	47.83%	32/274	11.68%
<b>Sporadic</b>	13/358	3.63%	5/27	18.52%	18/385	4.68%
<b>Other <sup>(1)</sup></b>	16/172	9.30%	7/14	50.00%	23/186	12.37%
<b>Flanders-Belgian Cohort</b>						
<b>Total</b>	21/337	6.23%	7/23	30.43%	28/360	7.78%
<b>Familial</b>	12/101	11.88%	6/7	85.71%	18/108	16.67%
<b>Sporadic</b>	9/236	3.81%	1/16	6.25%	10/252	3.97%
<b>Combined European and Flanders-Belgian cohorts</b>						
<b>Total</b>	71/1118	6.35%	30/87	34.48%	101/1205	8.38%
<b>Familial</b>	33/352	9.38%	17/30	56.67%	50/382	13.09%
<b>Sporadic</b>	22/594	3.70%	6/43	13.95%	28/637	4.40%
<b>Other <sup>(1)</sup></b>	16/172	9.30%	7/14	50.00%	23/186	12.37%

*Note: <sup>(1)</sup>Other includes patients of which there was no family history data available in the European cohort.*



**Table 3. Meta-analysis of *C9orf72* prevalence studies in Western Europe**

Country	Total			Familial			Sporadic			Studies represented in the meta-analysis
	total n	carriers n	carriers %	total n	carriers n	carriers %	total n	carriers n	carriers %	
<b>Total<sup>(1)</sup></b>	<b>2636</b>	<b>263</b>	<b>9.98</b>	<b>756</b>	<b>140</b>	<b>18.52</b>	<b>1804</b>	<b>113</b>	<b>6.26</b>	
<b>Belgium</b>	369	29	7.86	115	19	16.52	254	10	3.94	<i>Gijselinck et al., 2012</i> <i>European EOD consortium</i>
<b>Denmark</b>	82	10	12.20	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	<i>Lindquist et al., 2012</i>
<b>Finland</b>	75	22	29.33	27	13	48.15	48	9	18.75	<i>Majounie et al., 2012</i>
<b>France</b>	200	36	18.00	50	22.00	44	150	14	9.33	<i>Majounie et al., 2012</i>
<b>Germany</b>	228	11	4.82	66	5	7.58	162	6	3.7	<i>European EOD consortium</i> <i>Majounie et al., 2012</i>
<b>Italy</b>	345	21	6.09	86	5	5.81	259	16	6.18	<i>European EOD consortium</i>
<b>The Netherlands</b>	340	35	10.29	116	30	25.86	224	5	2.23	<i>Majounie et al., 2012</i>
<b>Portugal</b>	151	10	6.62	92	4	4.35	59	6	10.17	<i>European EOD consortium</i>
<b>Spain</b>	51	13	25.49	20	6	30	31	7	22.58	<i>European EOD consortium</i>
<b>Sweden</b>	82	17	20.73	14	8	57.14	74	9	12.16	<i>European EOD consortium</i> <i>Majounie et al., 2012</i>
<b>UK</b>	713	59	8.27	170	28	16.47	543	31	5.71	<i>Majounie et al., 2012</i>

Note: Meta-analysis combined prevalences from the European EOD consortium study, the Gijselinck et al., LancetNeurol 2012 study [Gijselinck et al., 2012], the Majounie et al., Lancet Neurol 2012 study [Majounie et al., 2012], and the Lindquist et al., Clin Neurol study [Lindquist et al., 2012]. <sup>(1)</sup>Patient samples from the Czech Republic, Bulgaria, and Austria from the European EOD consortium study, and from Sardinia from the Majounie study, contained less than 20 patients and were therefor excluded from the meta-analysis. N.A.: information not available.

**Table 4. Low Complexity Sequence (LCS) indels in carriers of a pathological G<sub>4</sub>C<sub>2</sub> expansion and in non-carriers of Flanders-Belgian and European cohorts**

Diagnosis	Number of individuals	Indel Genomic mutation name <sup>(1)(2)</sup>
<b>Flanders-Belgian cohort</b>		
<b>Patients expansion carriers (10/27)</b>		
FTLD	7	g.26747_26756del <b>GTGGT</b> CGGGG
FTLD-ALS	2	g.26747_26756del <b>GTGGT</b> CGGGG
FTLD-ALS	1	g.26742_26764delGGGG <b>GTGGT</b> CGGGGCGGGCCCG
<b>Patients without expansion (5/290)</b>		
FTLD	1	g.26747_26751del <b>GTGGT</b>
FTLD	3	g.26752_26762delCGGGGCGGGCC
FTLD-ALS	1	g.26752_26774delCGGGGCGGGCCCGGGGCGGGCC
<b>Controls (4/752)</b>		
control	3	g.26752_26762delCGGGGCGGGCC
control	1	g.26746_26761del <b>CGTGGT</b> CGGGGCGGGC
<b>European cohort</b>		
<b>Patients expansion carriers (14/57)</b>		
FTLD	6	g.26747_26756del <b>GTGGT</b> CGGGG
FTLD	1	g.26747_26751del <b>GTGGT</b>
FTLD	1	g.26747_26768del <b>GTGGT</b> CGGGGCGGGCCCGGGG
FTLD	1	g.26753_26764delGGGGCGGGCCCG
FTLD	1	g.26775_26776insG
FTLD	1	g.26775_26776insGGGGCGGGCCCG
FTLD-ALS	1	g.26747_26756del <b>GTGGT</b> CGGGG
FTLD-ALS	1	g.26746_26773del <b>CGTGGT</b> CGGGGCGGGCCCGGGGCGGGC
FTLD-ALS	1	g.26752_26762delCGGGGCGGGCC

Note: <sup>(1)</sup>The nucleotide sequence GTGGT is most frequently deleted in the LCS adjacent to the G<sub>4</sub>C<sub>2</sub> repeat in C9orf72 (indicated in bold); <sup>(2)</sup>gDNA numbering relative to reverse complement of contig AL451123.12 and starting at nucleotide 1.