(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
28 March 2013 (28.03.2013)

WIPO | PCT

(10) International Publication Number
WO 2013/041577 A1

(51) International Patent Classification:
C12Q 1/68 (2006.01) *G01N 33/53* (2006.01)
A01K 67/027 (2006.01)

(21) International Application Number:
PCT/EP2012/068456

(22) International Filing Date:
19 September 2012 (19.09.2012)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/626,063 20 September 2011 (20.09.2011) US
61/638,353 25 April 2012 (25.04.2012) US

(71) Applicants: **VIB VZW** [BE/BE]; Rijvisschestraat 120, B-9052 Gent (BE). **UNIVERSITEIT ANTWERPEN** [BE/BE]; Prinsstraat 12, B-2000 Antwerpen (BE).

(72) Inventors: **VAN BROECKHOVEN, Christine**; Koning Albertlei 15, B-2650 Edegem (BE). **CRUTS, Marc**; Belegstraat 73, B-2018 Antwerpen (BE).

(74) Common Representative: **VIB VZW**; Rijvisschestraat 120, B-9052 Gent (BE).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

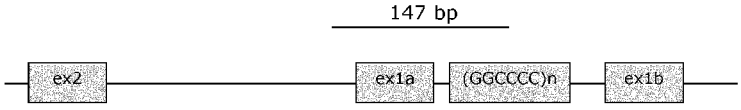
Published:

— *with international search report (Art. 21(3))*

— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: METHODS FOR THE DIAGNOSIS OF AMYOTROPHIC LATERAL SCLEROSIS AND FRONTOTEMPORAL LOBAR DEGENERATION

Figure 1



(57) Abstract: The invention relates to the detection of mutations in the promoter region of the gene C9ORF72, in particular a hexanucleotide expansion, wherein said mutations cause a significant decrease in the expression of gene C9ORF72. The decrease in gene C9ORF72 expression is related to the presence of amyotrophic lateral sclerosis (ALS) or frontotemporal lobar degeneration (FTLD), and the mutations can be used in the diagnosis of ALS and/or FTLD, or in the construction of transgenic animals for studying ALS and/or FTLD.

WO 2013/041577 A1

Methods for the diagnosis of amyotrophic lateral sclerosis and frontotemporal lobar degeneration

Field of the invention

The present invention relates to the field of diagnosis, more particularly to the diagnosis of amyotrophic lateral sclerosis and frontotemporal lobar degeneration. In particular the invention relates to the detection of mutations in the promoter region of the gene C9ORF72, whereby said mutations cause a significant decrease in the expression of gene C9ORF72. The decrease in gene C9ORF72 expression is related to the presence of amyotrophic lateral sclerosis (ALS) or frontotemporal lobar degeneration (FTLD) in a patient. The identified mutations can be used in the diagnosis of ALS and/or FTLD, or in the construction of transgenic animals for studying ALS and/or FTLD.

Introduction to the invention

Amyotrophic lateral sclerosis (ALS) and frontotemporal lobar degeneration (FTLD) are two fatal neurodegenerative diseases for which effective therapies aiming at delaying, halting or preventing the disease are lacking. ALS is the most common motor neuron disorder (Rowland & Shneider, 2001) and FTLD has a prevalence close to that of Alzheimer disease in the population below age 65 years (Rosso et al., 2003). They are considered as both extremes of a spectrum of clinically and pathologically overlapping disorders (Lillo & Hodges, 2009). In addition, there is emerging evidence that FTLD and ALS also share common genetic aetiologies, suggesting that overlapping disease mechanisms are involved in both diseases. Clinically, ALS patients show reduced control of voluntary muscle movement expressed in increased muscle weakness, disturbances of speech, swallowing or breathing, as a result of progressive upper and lower motor neuron degeneration in motor cortex, brainstem and spinal cord, and up to 50% of ALS patients shows mild disturbances in executive functions while a minority also develop overt FTLD (Lomen-Hoerth et al., 2003; Ringholz et al., 2005). FTLD symptoms include behavioural, personality and language disturbances, and also cognitive dysfunctions, due to affected frontal and temporal cortical neurons in the brain. FTLD patients may additionally present with typical clinical signs of ALS in a later stage of the disease (Neary et al., 1998). Pathologically, although in different neuronal cells, TAR DNA-binding protein-43 (TDP-43) is a major constituent of neuronal deposits in both ALS and TDP-43 positive FTLD (FTLD-TDP), the most common pathological FTLD subtype (Arai et al., 2006; Neumann et al., 2006). Five to 10% of ALS patients and up to 50% of FTLD patients has a positive familial history of disease with a Mendelian mode of inheritance indicating a significant contribution of genetic factors in disease aetiology. Although the exact biochemical pathways involved in ALS or FTLD are still unknown, several molecular components were identified in the last twenty years through molecular genetic studies in familial and sporadic patients, which are most likely part of a complex network of cellular mechanisms. Since these genes explain only a minority of patients, further unraveling the genetic heterogeneity is necessary to identify new therapeutic targets. Mutations causing ALS were observed in genes encoding Cu/Zn superoxide dismutase 1 (*SOD1*) (Rosen et al., 1993), TDP-43 (*TARDBP*) (Gitcho et al., 2008; Kabashi et al., 2008; Sreedharan et al., 2008; Van Deerlin et al., 2008; Yokoseki et al., 2008), fused in

sarcoma (*FUS*) (Kwiatkowski, Jr. et al., 2009; Vance et al., 2009) and angiogenin (*ANG*) (Greenway et al., 2006), among other genes, while in familial FTLD patients mutations in the genes encoding granulin (*GRN*) (Baker et al., 2006; Cruts et al., 2006), the microtubule-associated protein tau (*MAPT*) (Hutton et al., 1998), the valosin-containing protein (*VCP*) (Watts et al., 2004) and the charged multivesicular body protein 2B (*CHMP2B*) (Skibinski et al., 2005) were found. Recent family-based linkage and population-based association studies identified genetic factors overlapping between ALS and FTLD. For example, mutations in the ALS genes *TARDBP* and *FUS* are occasionally found in FTLD patients (Kovacs et al., 2009; Van Langenhove et al., 2010) and mutations in the FTLD gene *VCP* were also detected in ALS (Johnson et al., 2010). However, most convincing evidence for the genetic overlap comes from the observation that both ALS and FTLD can occur within the same family or within a single patient of a family. More than 15 autosomal dominant families with ALS and FTLD worldwide are causally linked with a major disease locus at chromosome 9p13-p21 (ALSFTD2 locus) (Boxer et al., 2010; Gijselinck et al., 2010; Le Ber et al., 2009; Luty et al., 2008; Momeni et al., 2006; Morita et al., 2006; Pearson et al., 2011; Valdmanis et al., 2007; Vance et al., 2006). The minimally linked region in all these families is about 3.6 Mb in size containing five known protein-coding genes. Moreover, several recent genome-wide association studies (GWAS) in ALS populations from different European origins showed the presence of a major genetic risk factor for ALS at the same chromosome 9p region (Laaksovirta et al., 2010; Shatunov et al., 2010; van Es et al., 2009). The Finnish study narrowed the associated region to a 232 kb linkage disequilibrium (LD) block containing three known genes (*MOBKL2B*, *IFNK*, *C9orf72*) and suggested the presence of a major risk gene with high penetrance (Laaksovirta et al., 2010). Likewise, a GWAS in FTLD has implicated the same region (Van Deerlin et al., 2010). This finding was further confirmed in other FTLD and ALS-FTLD cohorts (Rollinson et al., 2011). Together, these data demonstrate that ALS and FTLD share a major common genetic factor on chromosome 9p, most likely showing high mutation frequencies. Despite all attempts of several research groups, the genetic defect(s) underlying both genetic linkage and association to this region have not been identified yet.

In the present invention we have identified the causal gene mutation. This mutation can be used for the development of a diagnostic assay and kits for the detection of FTLD and/or ALS.

Figure legends:

Figure 1: Schematic view of the repeat primer-based PCR detection assay on Chr 9

Figure 2: PCR amplification products of healthy person are separated on an Applied Biosystems ABI 3730 and analyzed.

Figure 3: PCR amplification products of ALS patient with repeat expansion on Chr 9 are separated on an Applied Biosystems ABI 3730 and analyzed.

Figure 4: Schematic representation of the chromosome 9p21 ALS-FTLD locus. Upper panel: grey bars indicate the minimal candidate regions in all reported significantly linked ALS-FTLD families, defining a minimal interval of 3.7 Mb between D9S169 and D9S251 containing five protein coding genes, illustrated with grey lines. Lower panel: associated SNPs in ALS and FTLD GWAS are shown in red and LD blocks or finemapped regions of these GWAS are indicated with green lines. Three genes are located in the associated region.

Figure 5: the effect of the repeat expansion on gene expression was evaluated. qPCR analysis of random-primed and poly-A RNA of frontal cortex of two patients carrying the repeat expansion compared to frontal cortex RNA of control individuals demonstrated that in repeat expansion carriers, the C9ORF72 transcript was about 50% reduced after normalization to 4 housekeeping genes.

Figure 6: The DNA-sequence depicts the first 5460 base pairs of SEQ ID NO: 1, in the figure the hexanucleotide repeat which is expanded in patients diagnosed for ALS, FTLD or ALS and FTLD or are susceptible for ALS, FTLD or ALS and FTLD corresponds to nucleotides 5299 to 5321, i.e. 3 $^{5}/_{6}$ repeat units (depicted as bold underlined). In this DNA sequence exon 1A, the first non-coding exon of the longer transcript NM_018325 (GenBank accession number) is located from nucleotides 5362 to 5414 (depicted in UPPER CASE LETTERS).

Figure 7: European Early-Onset Dementia (EOD) Consortium

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. The European EOD consortium currently holds 40 partnering dementia research groups from 16 EU-countries. The European EOD consortium is coordinated by Christine Van Broeckhoven, Neurodegenerative Brain Diseases group, Department of Molecular Genetics, VIB, Antwerp, Belgium. The European EOD consortium member countries are indicated with an asterisk.

Figure 8: Genotyping assays to characterize the *C9orf72* region and G_4C_2 repeat

The *C9orf72* G_4C_2 repeat (dark grey box) is located upstream of the first exon, exon 1a (white box), in the largest transcript 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, we observed in the LCS, is indicated in blue. Below, the primers with their corresponding PCR amplicons are shown for each of the PCR genotyping assays.

Figure 9: Distribution of normal repeat lengths in the Flanders-Belgian patients and control individuals. Histograms of G_4C_2 repeat units sized < 60 repeats in Flanders-Belgian patients, excluding patients with mutations in known causal genes or with a pathological G_4C_2 expansion, compared to control individuals.

Figure 10: Correlation of normal repeat lengths with rs2814707 alleles. Histograms of G_4C_2 repeat units in 610 control individuals homozygous for the rs2814707 C-allele and 53 homozygous for the rs2814707 T-allele.

Detailed description of the invention

In practicing the methods described herein, many conventional techniques in molecular biology, protein biochemistry, cell biology, immunology, microbiology and recombinant DNA are used. These techniques are well-known and are explained in, e.g. Current Protocols in Molecular Biology, Vols. I-III, Ausubel, Ed. (1997); Sambrook et al., Molecular Cloning: A Laboratory Manual, Second Ed. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989); DNA Cloning: A Practical Approach, Vols. I and II, Glover, Ed. (1985);

Oligonucleotide Synthesis, Gait, Ed. (1984); Nucleic Acid Hybridization, Hames & Higgins, Eds. (1985); Transcription and Translation, Hames & Higgins, Eds. (1984); Animal Cell Culture, Freshney, Ed. (1986); Immobilized Cells and Enzymes (IRL Press, 1986); Perbal, A Practical Guide to Molecular Cloning; the series, Meth. Enzymol., (Academic Press, Inc., 1984); Gene Transfer Vectors for Mammalian Cells, Miller & Calos, Eds. (Cold Spring Harbor Laboratory, NY, 1987); and methods in Enzymol., Vols. 154 and 155, Wu & Grossman, and Wu, Eds., respectively.

The definitions of certain terms as used in this specification are provided below. Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs.

As used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. For example, reference to "a nucleic acid" includes a combination of two or more nucleic acids, and the like.

The term "amplification" or "amplify" as used herein means one or more methods known in the art for copying a target nucleic acid, thereby increasing the number of copies of a selected nucleic acid sequence. Amplification may be exponential or linear. A target nucleic acid may be either DNA or RNA. The sequences amplified in this manner form an "amplicon." While the exemplary methods described hereinafter relate to amplification using the polymerase chain reaction ("PCR"), numerous other methods are known in the art for amplification of nucleic acids e.g., isothermal methods, rolling circle methods, etc). The skilled artisan will understand that these other methods may be used either in place of, or together with, PCR methods. See, e.g. Saiki, "Amplification of Genomic DNA" in PCR Protocols, Innis et al, Eds., Academic Press, San Diego, CA 1990, pp. 13-20; Wharam et al, Nucleic Acids Res., 2001, 29(11):E54-E54; Hafner et al, Biotechniques 2001, 30(4):852-6, 858, 860; Zhong et al, Biotechniques, 2001, 30(4):852-6, 858, 860.

The terms "allele" and "allelic variant" are used interchangeably herein. An allele is any one of a number of alternative forms or sequences of the same gene occupying a given locus or position on a chromosome. A single allele for each locus is inherited separately from each parent, resulting in two alleles for each gene. An individual having two copies of the same allele of a particular gene is homozygous at that locus, whereas an individual having two different alleles of a particular gene is heterozygous.

The term "complement" as used herein means the complementary sequence to a nucleic acid according to standard Watson/Crick base pairing rules. A complement sequence can also be a sequence of RNA complementary to the DNA sequence or its complement sequence, and can also be a cDNA. The term "substantially complementary" as used herein means that two sequences hybridize under stringent hybridization conditions. The skilled artisan will understand that substantially complementary sequences need not hybridize along their entire length. In particular, substantially complementary sequences comprise a contiguous sequence of bases that do not hybridize to a target or marker sequence, positioned 3 ' or 5' to a contiguous sequence of bases that hybridize under stringent hybridization conditions to a target or marker sequence.

The term "diagnosing" means determining a disease state or condition in a patient in such a way as to inform a health care provider as to the necessity or suitability of a treatment for the patient.

The term "genomic DNA" refers to some or the entire DNA from the nucleus of a cell. Genomic DNA may be intact or fragmented. In some embodiments, genomic DNA may include sequence from all or a portion of a single gene or from multiple genes, sequence from one or more chromosomes, or sequence from all chromosomes of a cell. In contrast, the term "total genomic nucleic acid" is used herein to refer to the full complement of DNA contained in the genome of a cell. As is well known, genomic nucleic acid includes gene coding regions, introns, 5' and 3' untranslated regions, 5' and 3' flanking DNA and structural segments such as telomeric and centromeric DNA, replication origins, and intergenic DNA. Genomic nucleic acid may be obtained from the nucleus of a cell, or recombinantly produced. Amplification techniques may also be used.

As used herein, "nucleic acid" refers broadly to segments of a chromosome, segments or portions of DNA, cDNA, and/or RNA. Nucleic acid may be derived or obtained from an originally isolated nucleic acid sample from any source e.g. isolated from, purified from, amplified from, cloned from, or reverse transcribed from sample DNA or RNA.

As used herein, the term "oligonucleotide" refers to a short polymer composed of deoxyribonucleotides, ribonucleotides or any combination thereof. Oligonucleotides are generally between about 10 and about 100 nucleotides in length. Oligonucleotides are preferably 15 to 70 nucleotides long, with 20 to 26 nucleotides being the most common. The single letter code for nucleotides is as described in the U.S. Patent Office Manual of Patent Examining Procedure, section 2422, table 1. An oligonucleotide may be used as a primer or as a probe.

An oligonucleotide is "specific" for a nucleic acid if the oligonucleotide has at least 50% sequence identity with a portion of the nucleic acid when the oligonucleotide and the nucleic acid are aligned. An oligonucleotide that is specific for a nucleic acid is one that, under the appropriate hybridization or washing conditions, is capable of hybridizing to the target of interest and not substantially hybridizing to nucleic acids which are not of interest. Higher levels of sequence identity are preferred and include at least 75%, at least 80%, at least 85%, at least 90%, at least 95% and more preferably at least 98% sequence identity.

As used herein, a "primer" for amplification is an oligonucleotide that specifically anneals to a target or marker nucleotide sequence. The 3' nucleotide of the primer should be identical to the target or marker sequence at a corresponding nucleotide position for optimal primer extension by a polymerase. As used herein, a "forward primer" is a primer that anneals to the anti-sense strand of dsDNA. A "reverse primer" anneals to the sense-strand of dsDNA.

As used herein, the term "promoter" refers to a nucleic acid sequence sufficient to direct transcription of a gene. Also included in the invention are those promoter elements which are sufficient to render promoter dependent gene expression controllable for cell type specific, tissue specific or inducible by external signals or agents. The term "neuron-specific promoter" refers to a promoter that results in a higher level of transcription of a gene in cells of neuronal lineage compared to the transcription level observed in cells of a non-neuronal lineage.

Diagnostic Methods of the present invention

The present inventors identified that the presence of a hexanucleotide expansion (5'-GGCCCC-3') in SEQ ID NO: 1, particularly in the promoter region of SEQ ID NO: 1, more particularly in exon 1A, shows a strong association with ALS and/or FTLD disease risk. In accordance with the present invention, there are provided methods of detecting a particular nucleic acid segment of interest in a sample of nucleic acids. In particular embodiments, the nucleic acid segment of interest includes a hexanucleotide expansion (5'-GGCCCC-3') in the gene SEQ ID NO: 1. SEQ ID NO: 1 is the genomic sequence of ORF (open reading frame 72) present on human chromosome 9 (in short C9ORF72). The complete nucleotide sequence of SEQ ID NO: 1 is depicted below. This information may be used to determine if an individual is suffering from or is susceptible to ALS and/or FTLD. The wording 'ALS and/or FTLD' means that in one embodiment a patient having a hexanucleotide expansion in SEQ ID NO: 1 has ALS or is susceptible to ALS, in another embodiment a patient having said hexanucleotide expansion in SEQ ID NO: 1 has FTLD or is susceptible to FTLD. In yet another embodiment a patient having said hexanucleotide expansion in SEQ ID NO: 1 has combined disease hallmarks of FTLD and ALS. In particular embodiments the information may also be used in genetic counseling to determine if the individual could have offspring with an increased risk of ALS and/or FTLD.

SEQ ID NO: 1 depicts the genomic sequence of C9ORF72, preceded by 5000 bp upstream and followed by 2000 bp downstream regulatory sequence. In other words SEQ ID NO: 1 codes for the gene C9ORF72. C9ORF72 is a gene with hitherto unknown function. The sequence SEQ ID NO: 1 corresponds to the reverse complement of the region spanning nucleotides 27532544 to 27568842 in the human reference genome sequence version GRCh37; in GenBank the sequence is referred to as Accession NT_008413. Figure 6 depicts the first 5460 base pairs of SEQ ID NO: 1, in Figure 6 the hexanucleotide repeat which is expanded in patients corresponds to nucleotides 5299 to 5321, i.e. 3 $^{5}/_{6}$ repeat units (depicted as bold underlined). In this DNA sequence exon 1A, the first non-coding exon of the longer transcript NM_018325 (GenBank accession number) is located from nucleotides 5362 to 5414 (depicted in UPPER CASE LETTERS).

The hexanucleotide repeat in SEQ ID NO: 1 is a repeat of the sequence 5'-GGGGCC-3'. Because SEQ ID NO: 1 is depicted in the complement with respect to the orientation of human chromosome 9 the diagnostic method can be described as a detection of a 5'-GGCCCC-3' repeat expansion wherein said repeat expansion is measured with respect to the orientation of the chromosome (human Chromosome 9). In the alternative the diagnostic method can be described as a detection of a 5'-GGGGCC-3' repeat expansion wherein said repeat expansion is measured with respect to the orientation of the C9ORF72 gene on Chromosome 9 (in particular since the orientation of the C9ORF72 gene on Chromosome 9 is reversed). In the scientific literature the detection of a sequence (here the repeat expansion) is described (as a scientific consensus) with respect to the gene. The claims of the application are therefore citing the detection of a 5'-GGGGCC-3' repeat expansion in the promoter region of the C9ORF72 gene or in the alternative the detection of a 5'-GGGGCC-3' repeat expansion in SEQ ID NO: 1.

In a specific embodiment the invention provides a method for diagnosing ALS or FTLD or the combined presence of ALS and FTLD or a method for diagnosing the susceptibility to ALS or FTLD or the combined presence of ALS and FTLD comprising detecting the presence of a

hexanucleotide repeat expansion of 5'-GGGGCC-3' in SEQ ID NO: 1 in a sample comprising nucleic acids from a subject, and diagnosing the subject as having or being susceptible to ALS or FTLD or the combined presence of ALS and FTLD when said hexanucleotide repeat expansion is present in SEQ ID NO: 1.

In another embodiment the invention provides a method for diagnosing ALS or FTLD or the combined presence of ALS and FTLD or a method for diagnosing the susceptibility to ALS or FTLD or the combined presence of ALS and FTLD comprising detecting the presence of a hexanucleotide repeat expansion of 5'-GGGGCC-3' of more than 60 repeats in SEQ ID NO: 1 in a sample comprising nucleic acids from a subject, and diagnosing the subject as having or being susceptible to ALS or FTLD or the combined presence of ALS and FTLD when said hexanucleotide repeat expansion is present in SEQ ID NO: 1.

In another embodiment the invention provides a method for diagnosing ALS or FTLD or the combined presence of ALS and FTLD or a method for diagnosing the susceptibility to ALS or FTLD or the combined presence of ALS and FTLD comprising detecting the presence of a hexanucleotide repeat expansion of 5'-GGGGCC-3' of more than 80 repeats in SEQ ID NO: 1 in a sample comprising nucleic acids from a subject, and diagnosing the subject as having or being susceptible to ALS or FTLD or the combined presence of ALS and FTLD when said hexanucleotide repeat expansion is present in SEQ ID NO: 1.

In a particular embodiment the diagnostic method is based on a PCR method.

In yet another particular embodiment the diagnostic method is based on a repeat primer-based PCR detection assay.

In yet another embodiment the present invention also provides variant nucleic acids derived from C9ORF72 associated with ALS and/or FTLD. In one embodiment, the detection of a hexanucleotide expansion (5'-GGGGCC-3') present in SEQ ID NO: 1 has a strong association with ALS and/or FTLD risk. In a specific embodiment said hexanucleotide expansion is between 25 and 200 repeats. In yet another embodiment said hexanucleotide expansion is between 25 and 50 repeats. In yet another embodiment said hexanucleotide expansion is between 30 and 200 repeats. In yet another embodiment said hexanucleotide expansion is between 30 and 50 repeats. In yet another embodiment said hexanucleotide expansion is between 40 and 200 repeats. In yet another embodiment said hexanucleotide expansion is between 50 and 200 repeats. In yet another embodiment said hexanucleotide expansion is between 60 and 200 repeats. In yet another embodiment said hexanucleotide expansion is between 70 and 200 repeats. In yet another embodiment said hexanucleotide expansion is between 80 and 200 repeats. In yet another embodiment said hexanucleotide expansion is between 90 and 200 repeats. In yet another embodiment said hexanucleotide expansion is between 100 and 200 repeats. In yet another embodiment said hexanucleotide expansion is higher than 25 repeats. In yet another embodiment said hexanucleotide expansion is higher than 30 repeats. In yet another embodiment said hexanucleotide expansion is higher than 35 repeats. In yet another embodiment said hexanucleotide expansion is higher than 40 repeats. In yet another embodiment said hexanucleotide expansion is between 80 and 140 repeats. In yet another embodiment said hexanucleotide expansion is between 30 and 80 repeats. In yet another embodiment said hexanucleotide expansion in SEQ ID NO: 1 is longer than 1000 base pairs. In yet another embodiment said hexanucleotide expansion is longer than 5000 base pairs. In yet another embodiment said hexanucleotide expansion is longer than 10.000 base

pairs. In yet another embodiment said hexanucleotide expansion is between 5000 and 10.000 base pairs. In yet another embodiment the number of repeats reflects the onset of the disease of ALS and/or FTLD, *id est* the higher the number of hexanucleotide repeats are identified in SEQ ID NO: 1 the earlier the onset of ALS and/or FTLD can occur in a patient (i.e. the presence of a higher expansion of the hexanucleotide sequence results in a younger disease onset of ALS and/or FTLD).

Those skilled in the art will readily recognize that nucleic acid molecules may be double-stranded molecules and that reference to a particular site on one strand refers, as well, to the corresponding site on a complementary strand. In defining a variant position, allele, or nucleotide sequence, reference to an adenine, a thymine (uridine), a cytosine, or a guanine at a particular site on one strand of a nucleic acid molecule also defines the thymine (uridine), adenine, guanine, or cytosine (respectively) at the corresponding site on a complementary strand of the nucleic acid molecule. Thus, reference may be made to either strand in order to refer to a particular variant position, allele, or nucleotide sequence. Probes and primers may be designed to hybridize to either strand and variant genotyping methods disclosed herein may generally target either strand. Throughout the specification, in identifying a variants position, reference is generally made to the protein-encoding strand, only for the purpose of convenience.

References to variant peptides, polypeptides, or proteins of the present invention include peptides, polypeptides, proteins, or fragments thereof, that contain at least one nucleotide or amino acid residue that differs from the corresponding nucleotide or amino acid sequence of the art-known nucleotide/protein (the art-known protein may be interchangeably referred to as the "wild-type", "reference", or "normal" protein). Such variant proteins can result from a codon change caused by a non-synonymous nucleotide substitution at a protein-coding SNP position (i.e., a missense mutation) disclosed by the present invention. Variant proteins can also result from a nonsense mutation, i.e., a SNP that creates a premature stop codon, a SNP that generates a read-through mutation by abolishing a stop codon, or a SNP that otherwise alters the structure, function/activity, or expression of a protein, such as a SNP in a regulatory region (e.g., a promoter or enhancer) or a SNP that leads to alternative or defective splicing, such as a SNP in an intron or a SNP at an exon/intron boundary.

The methods and compositions described herein may be used to detect mutations in the gene depicted in SEQ ID NO: 1, more particularly in Exon 1A present in SEQ ID NO: 1, more particularly in the promoter region of C9ORF72 using a biological sample obtained from an individual. Nucleic acid (DNA or RNA) may be isolated from the sample according to any methods well known to those of skill in the art. Examples include tissue samples or any cell-containing bodily fluid. Biological samples may be obtained by standard procedures and may be used immediately or stored, under conditions appropriate for the type of biological sample, for later use. Methods of obtaining test samples are well known to those of skill in the art and include, but are not limited to, aspirations, tissue sections, drawing of blood or other fluids, surgical or needle biopsies, and the like. The test sample may be obtained from a patient. The test sample may contain cells, tissues or fluid obtained from a patient suspected being afflicted with or a carrier for an allele associated with increased susceptibility to ALS and/or FTLD. The test sample may be a cell-containing liquid or a tissue. Samples may include, but are not limited to, amniotic fluid, cerebrospinal fluid, biopsies, blood, blood cells, bone marrow, fine needle biopsy samples, peritoneal fluid, amniotic fluid, plasma, pleural fluid, saliva, semen,

serum, tissue or tissue homogenates, frozen or paraffin sections of tissue. Samples may also be processed, such as sectioning of tissues, fractionation, purification, or cellular organelle separation.

The methods can also be used to perform prenatal diagnosis using any type of embryonic or fetal cell or nucleic acid containing body fluid. Fetal cells can be obtained through the pregnant female, or from a sample of an embryo. Fetal cells are present in amniotic fluid obtained by amniocentesis, chorionic villi aspirated by syringe, percutaneous umbilical blood, a fetal skin biopsy, a blastomere from a four-cell to eight-cell stage embryo (pre- implantation), or a trophectoderm sample from a blastocyst (pre -implantation or by uterine lavage). In yet another particular embodiment the mutation detection can be carried out on the basis of free-floating DNA which is genomic DNA derived from the fetus and which is present in the blood of a pregnant female.

If necessary, the sample may be collected or concentrated by centrifugation and the like. The cells of the sample may be subjected to lysis, such as by treatments with enzymes, heat, surfactants, ultrasonication, or a combination thereof. The lysis treatment is performed in order to obtain a sufficient amount of nucleic acid derived from the individual's cells to detect using polymerase chain reaction. Alternatively, variants in the C9ORF72 gene may be detected using an acellular bodily fluid.

Methods of plasma and serum preparation are well known in the art. Either "fresh" blood plasma or serum, or frozen (stored) and subsequently thawed plasma or serum may be used. Frozen (stored) plasma or serum should optimally be maintained at storage conditions of -20 to -70°C until thawed and used. "Fresh" plasma or serum should be refrigerated or maintained on ice until used, with nucleic acid (e.g., RNA, DNA or total nucleic acid) extraction being performed as soon as possible. Exemplary methods are described below.

Blood can be drawn by standard methods into a collection tube, typically siliconized glass, either without anticoagulant for preparation of serum, or with EDTA, sodium citrate, heparin, or similar anticoagulants for preparation of plasma. When preparing plasma or serum for storage, although not an absolute requirement, is that plasma or serum is first fractionated from whole blood prior to being frozen. This reduces the burden of extraneous intracellular RNA released from lysis of frozen and thawed cells which might reduce the sensitivity of the amplification assay or interfere with the amplification assay through release of inhibitors to PCR such as porphyrins and hematin. "Fresh" plasma or serum may be fractionated from whole blood by centrifugation, using gentle centrifugation at 300-800 times gravity for five to ten minutes, or fractionated by other standard methods. High centrifugation rates capable of fractionating out apoptotic bodies should be avoided. Since heparin may interfere with RT- PCR, use of heparinized blood may require pretreatment with heparanase, followed by removal of calcium prior to reverse transcription. Imai, H., et al., J. Virol. Methods 36:181-184, (1992). Thus, EDTA is a suitable anticoagulant for blood specimens in which PCR amplification is planned.

Nucleic Acid Extraction and Amplification: the nucleic acid to be amplified may be from a biological sample such as a patient, cell culture, tissue sample, and the like. The biological sample can be from a subject which includes any animal, preferably a mammal. A preferred subject is a human, which may be a patient presenting to a medical provider for diagnosis or treatment of a disease. The biological sample may be obtained from a stage of life such as a

fetus, young adult, adult, and the like. Subjects may be humans being tested for the existence of a hexanucleotide expansion in SEQ ID NO: 1. The volume of plasma or serum used in the extraction may be varied dependent upon clinical intent, but volumes of 100 µl to one milliliter of plasma or serum are usually sufficient.

Various methods of extraction are suitable for isolating the DNA or RNA. Suitable methods include phenol and chloroform extraction. See Maniatis et al, Molecular Cloning, A Laboratory Manual, 2d, Cold Spring Harbor Laboratory Press, page 16.54 (1989). Numerous commercial kits also yield suitable DNA and RNA including, but not limited to, QIAamp™ mini blood kit, Agencourt Genfind™, Roche Cobas® Roche MagNA Pure® or phenol: chloroform extraction using Eppendorf Phase Lock Gels®, and the NucliSens extraction kit (Biomérieux, Marcy l'Etoile, France). In other methods, mRNA may be extracted from patient blood/bone marrow samples using MagNA Pure LC mRNA HS kit and Mag NA Pure LC Instrument (Roche Diagnostics Corporation, Roche Applied Science, Indianapolis, IN). Nucleic acid extracted from tissues, cells, plasma or serum can be amplified using nucleic acid amplification techniques well known in the art. Many of these amplification methods can also be used to detect the presence of the hexanucleotide expansion (id est the mutations in C9ORF72) simply by designing oligonucleotide primers or probes to interact with or hybridize to a particular target sequence in a specific manner. By way of example, but not by way of limitation, these techniques can include the polymerase chain reaction (PCR), reverse transcriptase polymerase chain reaction (RT-PCR), nested PCR, ligase chain reaction. See Abravaya, K., et al, Nucleic Acids Research, 23:675-682, (1995), branched DNA signal amplification, Urdea, M. S., et al, AIDS, 7 (suppl 2):S11-S 14, (1993), amplifiable RNA reporters, Q-beta replication, transcription-based amplification, boomerang DNA amplification, strand displacement activation, cycling probe technology, isothermal nucleic acid sequence based amplification (NASBA). See Kievits, T. et al, J Virological Methods, 35:273-286, (1991), Invader Technology, or other sequence replication assays or signal amplification assays. These methods of amplification are well-known in the art. In particular embodiments long template PCR can be used for the detection of the hexanucleotide repeat of the invention. In yet another embodiment sequencing (e.g. Sanger sequencing or other) can be used for the detection of the hexanucleotide repeat of the invention. In yet another embodiment classical sequencing or bisulfite sequencing, in which the template DNA is first chemically modified using bisulfite converting unmethylated C nucleotides into T, can be used to detect the hexanucleotide repeat of the invention. Without limiting the diagnostic methods to a particular mechanism it is thought that the latter methods are more suitable for the detection of huge (id est larger than 1000 base pairs) repeats of the hexanucleotide of the invention, in one mechanism because the bisulfite reaction prevents the formation of excessive secondary structures in the template DNA comprising C9ORF72, in particular the hexanucleotide repeat region. In yet another embodiment Southern blot restriction fragment length analysis can be used for the detection of the hexanucleotide repeat of the invention.

Thus in suitable embodiments PCR is used to amplify a target or marker sequence of interest. The skilled artisan is capable of designing and preparing primers that are appropriate for amplifying a target or marker sequence. The length of the amplification primers depends on several factors including the nucleotide sequence identity and the temperature at which these nucleic acids are hybridized or used during in vitro nucleic acid amplification. The considerations necessary to determine a preferred length for an amplification primer of a particular sequence identity are well-known to a person of ordinary skill. For example, the

length of a short nucleic acid or oligonucleotide can relate to its hybridization specificity or selectivity.

For analyzing variant nucleic acids, it may be appropriate to use oligonucleotides specific for alternative alleles. Such oligonucleotides which detect single nucleotide variations in target sequences may be referred to by such terms as "allele-specific probes", or "allele-specific primers". The design and use of allele-specific probes for analyzing polymorphisms is described in, e.g., Mutation Detection A Practical Approach, ed. Cotton et al. Oxford University Press, 1998; Saiki et al, Nature, 324: 163-166 (1986);

In one embodiment, a probe or primer may be designed to hybridize to a segment of target DNA such that the herein identified hexanucleotide repeat sequence aligns with either the 5' most end or the 3' most end of the probe or primer.

In some embodiments, the amplification may include a labeled primer, thereby allowing detection of the amplification product of that primer. In particular embodiments, the amplification may include a multiplicity of labeled primers; typically, such primers are distinguishably labeled, allowing the simultaneous detection of multiple amplification products.

In a specific embodiment, the primer or probe is labeled with a fluorogenic reporter dye that emits a detectable signal. While a suitable reporter dye is a fluorescent dye, any reporter dye that can be attached to a detection reagent such as an oligonucleotide probe or primer is suitable for use in the invention. Such dyes include, but are not limited to, Acridine, AMCA, BODIPY, Cascade Blue, Cy2, Cy3, Cy5, Cy7, Dabcyl, Edans, Eosin, Erythrosin, Fluorescein, 6-Fam, Tet, Joe, Hex, Oregon Green, Rhodamine, Rhodol Green, Tamra, Rox, and Texas Red.

In yet another embodiment, the detection reagent may be further labeled with a quencher dye such as Tamra, especially when the reagent is used as a self-quenching probe such as a TaqMan® (U.S. Pat. Nos. 5,210,015 and 5,538,848) or Molecular Beacon probe (U.S. Pat. Nos. 5,118,801 and 5,312,728), or other stemless or linear beacon probe (Livak et al, 1995, PCR Method Appl, 4:357-362; Tyagi et al, 1996, Nature Biotechnology, 14:303- 308; Nazarenko et al, 1997, Nucl. Acids Res., 25:2516-2521.

In some embodiments, the target sequence comprising the hexanucleotide repeat of the invention is amplified and the resulting amplicon is detected by electrophoresis. In some embodiments, the specific mutation or variant is detected by sequencing the amplified nucleic acid. In some embodiments, the target sequence is amplified using a labeled primer such that the resulting amplicon is detectably labeled. In some embodiments, the primer is fluorescently labeled.

In one embodiment, detection of a variant nucleic acid, such as the hexanucleotide repeat of the invention, is performed using the TaqMan® assay, which is also known as the 5' nuclease assay (see U.S. Pat. Nos. 5,210,015 and 5,538,848). The TaqMan® assay detects the accumulation of a specific amplified product during PCR. The TaqMan® assay utilizes an oligonucleotide probe labeled with a fluorescent reporter dye and a quencher dye. The reporter dye is excited by irradiation at an appropriate wavelength, it transfers energy to the quencher dye in the same probe via a process called fluorescence resonance energy transfer (FRET).

When attached to the probe, the excited reporter dye does not emit a signal. The proximity of the quencher dye to the reporter dye in the intact probe maintains a reduced fluorescence for the reporter. The reporter dye and quencher dye may be at the 5' most and the 3' most ends, respectively or vice versa. Alternatively, the reporter dye may be at the 5 ' or 3' most end while the quencher dye is attached to an internal nucleotide, or vice versa. In yet another embodiment, both the reporter and the quencher may be attached to internal nucleotides at a distance from each other such that fluorescence of the reporter is reduced. During PCR, the 5' nuclease activity of DNA polymerase cleaves the probe, thereby separating the reporter dye and the quencher dye and resulting in increased fluorescence of the reporter. Accumulation of PCR product is detected directly by monitoring the increase in fluorescence of the reporter dye. The DNA polymerase cleaves the probe between the reporter dye and the quencher dye only if the probe hybridizes to the target hexanucleotide repeat expansion containing template which is amplified during PCR, and the probe is designed to hybridize to the target hexanucleotide repeat expansion site only if the particular hexanucleotide repeat expansion allele is present. TaqMan® primer and probe sequences can readily be determined using the variant and associated nucleic acid sequence information provided herein. A number of computer programs, such as Primer Express (Applied Biosystems, Foster City, Calif), can be used to rapidly obtain optimal primer/probe sets. It will be apparent to one of skill in the art that such primers and probes for detecting the variants of the present invention are useful in diagnostic assays for ALS and/or FTLD can be readily incorporated into a kit format. The present invention also includes modifications of the TaqMan® assay well known in the art such as the use of Molecular Beacon probes (U.S. Pat. Nos. 5,118,801 and 5,312,728) and other variant formats (U.S. Pat. Nos. 5,866,336 and 6,117,635).

Oligonucleotide probes can be designed which are between about 10 and about 100 nucleotides in length and hybridize to the amplified region. Oligonucleotides probes are preferably 12 to 70 nucleotides; more preferably 15-60 nucleotides in length; and most preferably 15-25 nucleotides in length. The probe may be labeled. Amplified fragments may be detected using standard gel electrophoresis methods. For example, in preferred embodiments, amplified fractions are separated on an agarose gel and stained with ethidium bromide by methods known in the art to detect amplified fragments.

Assay controls may be used in the assay for detecting carriers and individuals afflicted with or susceptible to a ALS, FTLD or to ALS and FTLD. Positive controls for normal or wild type C9ORF72 gene may be used.

Kits

In a further aspect, the invention disclosure provides kits for diagnosing ALS, FTLD, or ALS and FTLD or prognosing patients for having an increased risk of ALS, FTLD, or ALS and FTLD, the kit comprising a set of reagents for determining the presence or absence, or differential presence, of the hexanucleotide repeat expansion.

Protein detection

Gene C9ORF72 or SEQ ID NO: 1 codes for at least two different splice forms, a shorter and a longer isoform wherein the shorter isoform is essentially a fragment of the longer isoforms. These two splice forms are depicted in SEQ ID NO: 2 and SEQ ID No: 3.

The gene C9ORF72 has 2 major, validated transcripts, but more transcripts may exist. The validated transcripts have GenBank Accession Numbers NM_018325 (the longer transcript, 3233 nucleotides) and NM_145005 (the shorter transcript, 1879 nucleotides). The encoded proteins have GenPept Accession Numbers NP_060795 (the longer protein isoform, 481 amino acids, depicted in SEQ ID NO: 2) and NP_659442 (the shorter protein isoform, 222 amino acids, depicted in SEQ ID NO: 3).

For convenience the term further used as “protein C9ORF72” refers to SEQ ID NO: 2 and SEQ ID NO: 3.

A level of protein C9ORF72 expression is reduced due to the presence of the identified mutation in the promoter region of gene C9ORF72 that results in little or no expression of C9ORF72 RNA and little or no presence of the C9ORF72 protein. The presence of the identified mutations in only one C9ORF72 allele can result in a level of wild-type C9ORF72 polypeptide that is intermediate between the level of wild-type C9ORF72 polypeptide typically observed when both C9ORF72 alleles are wild-type and the level typically observed when both alleles contain the mutation.

The term "reduced level" as used herein with respect to a level of C9ORF72 expression is any level of C9ORF72 expression that is less than a median level of wild-type C9ORF72 polypeptide or C9ORF72 RNA expression in a random population of humans (e.g. a random population of 10, 20, 30, 40, 50, 100, 500, 1000 or more humans) having homozygous wild-type C9ORF72 alleles. In some cases, a "reduced level" of C9ORF72 expression can be any level of wild-type C9ORF72 polypeptide or C9ORF72 RNA expression that is less than a median level of wild-type C9ORF72 polypeptide or RNA expression, respectively, in a random population of humans (e.g. a random population of 10, 20, 30, 40, 50, 100, 500, 1000 or more humans) not having been diagnosed with ALS and/or FTLD. In some cases, a reduced level of C9ORF72 expression can be a level of wild-type C9ORF72 expression that is at least one (e.g. at least 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, or 2.2) standard deviation less than a mean level of wild-type C9ORF72 expression in a random population of humans (e.g. having homozygous wild-type C9ORF72 alleles and/or not having been diagnosed with ALS and/or FTLD). In some cases, a reduced level of C9ORF72 expression can be a level of wild-type C9ORF72 expression that is less than a median level of wild-type C9ORF72 expression in a random population of humans (e.g., a random population of 10, 20, 30, 40, 50, 100, 500, 1000 or more humans) that are age-matched and/or who are race-matched. Humans that are age-matched can be the same age or can be in the same age range (e.g., 15 to 35 years of age, 35 to 75 years of age, 75 to 100 years of age, 35 to 45 years of age, 60 to 80 years of age, 20 to 35 years of age, or 40 to 50 years of age). In some case, a reduced level of C9ORF72 expression can be a level of wild-type C9ORF72 expression that is less than a median level of wild-type C9ORF72 expression in a random population of humans (e.g. a random population of 10, 20, 30, 40, 50, 100, 500, 1000 or more humans) having homozygous wild-type C9ORF72 alleles that are age-matched and/or who are race-matched. In some cases, a reduced level of C9ORF72 expression can be little or no detectable wild-type C9ORF72 expression.

It will be appreciated that C9ORF72 expression levels from comparable samples (e.g. blood samples) are used when determining whether or not a particular C9ORF72 expression level is a reduced level. For example, a mRNA level of C9ORF72 expression in a skin biopsy from a human is compared to the median mRNA level of C9ORF72 expression in skin biopsies from a

random population of humans (e.g. having homozygous wild-type C9ORF72 alleles and/or not having been diagnosed with ALS and/or FTLD). In addition, C9ORF72 expression levels are compared to a median C9ORF72 expression level measured using the same or a comparable method. Any appropriate method can be used to determine a C9ORF72 expression level. For example, Northern blotting, RT-PCR, or quantitative PCR can be used to determine a level of RNA molecules encoding a wild-type C9ORF72 polypeptide. In some cases, mass spectrometry can be used to determine a level of a wild-type C9ORF72 polypeptide. In some cases, a level of C9ORF72 polypeptide can be detected using a method that relies on an anti-C9ORF72 polypeptide antibody. It is envisaged here that an antibody against SEQ ID NO: 2 will also have a specificity for SEQ ID NO: 3 because SEQ ID NO: 3 can be considered as a fragment of SEQ ID NO: 2. Such methods include, without limitation, FACS, Western blotting, ELISA, immunohistochemistry, and immunoprecipitation. Antibody based assays (e.g. sandwich enzyme-linked immunosorbent assays) can include using combinations of antibodies that bind to one or more sites of the amino-terminal, central, and carboxy-terminal portions of a C9ORF72 polypeptide or a fragment thereof. An anti-C9ORF72 polypeptide antibody can be labeled for detection. For example, an anti-C9ORF72 polypeptide antibody can be labeled with a radioactive molecule, a fluorescent molecule, or a bioluminescent molecule. C9ORF72 polypeptides can also be detected indirectly using a labeled antibody that binds to an anti-C9ORF72 polypeptide antibody that binds to a C9ORF72 polypeptide. An antibody can be, without limitation, a polyclonal, monoclonal, human, humanized, chimeric, or single-chain antibody, or an antibody fragment having binding activity, such as a Fab fragment, F(ab') fragment, Fd fragment, fragment produced by a Fab expression library, fragment comprising a VL or VH domain, or epitope binding fragment of any of the above. An antibody can be of any type (e.g., IgG, IgM, IgD, IgA or IgY), class (e.g., IgGl, IgG4, or IgA2), or subclass. In a particular embodiment an antibody is derived from camels, in another particular embodiment an antibody derived from camels is a variable heavy chain also designated as a nanobody®. In addition, an antibody can be from any animal including birds and mammals. For example, an antibody can be a human, rabbit, sheep, camelid or goat antibody. An antibody can be naturally occurring, recombinant, or synthetic. Antibodies can be generated and purified using any suitable methods known in the art. For example, monoclonal antibodies can be prepared using hybridoma, recombinant, or phage display technology, or a combination of such techniques. In some cases, antibody fragments can be produced synthetically or recombinantly from a gene encoding the partial antibody sequence. An anti-C9ORF72 polypeptide antibody can bind to a C9ORF72 polypeptide at an affinity of at least 10^4 mol^{-1} (e.g., at least 10^5, 10^6, 10^7, 10^8, 10^9, 10^{10}, 10^{11}, or 10^{12} mol^{-1}). An anti-C9ORF72 polypeptide antibody provided herein can be prepared using any appropriate method. For example, any substantially pure C9ORF72 polypeptide, or fragment thereof can be used as an immunogen to elicit an immune response in an animal such that specific antibodies are produced. Thus, a human C9ORF72 polypeptide or a fragment thereof can be used as an immunizing antigen. In addition, the immunogen used to immunize an animal can be chemically synthesized or derived from translated cDNA. Further, the immunogen can be conjugated to a carrier polypeptide, if desired. Commonly used carriers that are chemically coupled to an immunizing polypeptide include, without limitation, keyhole limpet hemocyanin (KLH), thyroglobulin, bovine serum albumin (BSA), and tetanus toxoid. The preparation of polyclonal antibodies is well-known to those skilled in the art. See, e.g., Green et al, Production of Polyclonal Antisera, in IMMUNOCHEMICAL PROTOCOLS (Manson, ed.), pages 1-5 (Humana Press 1992) and Coligan et al, Production of Polyclonal Antisera in Rabbits, Rats, Mice and Hamsters, in CURRENT PROTOCOLS IN IMMUNOLOGY, section 2.4.1 (1992). In addition, those of skill in

the art will know of various techniques common in the immunology arts for purification and concentration of polyclonal antibodies, as well as monoclonal antibodies (Coligan, et al., Unit 9, CURRENT PROTOCOLS IN IMMUNOLOGY, Wiley Interscience, 1994). The preparation of monoclonal antibodies also is well-known to those skilled in the art. See, e.g., Kohler & Milstein, Nature 256:495 (1975); Coligan et al, sections 2.5.1 2.6.7; and Harlow et al, ANTIBODIES: A LABORATORY MANUAL, page 726 (Cold Spring Harbor Pub. 1988). Briefly, monoclonal antibodies can be obtained by injecting mice with a composition comprising an antigen, verifying the presence of antibody production by analyzing a serum sample, removing the spleen to obtain B lymphocytes, fusing the B lymphocytes with myeloma cells to produce hybridomas, cloning the hybridomas, selecting positive clones that produce antibodies to the antigen, and isolating the antibodies from the hybridoma cultures. Monoclonal antibodies can be isolated and purified from hybridoma cultures by a variety of well-established techniques. Such isolation techniques include affinity chromatography with Protein A Sepharose, size exclusion chromatography, and ion exchange chromatography. See, e.g., Coligan et al , sections 2.7.1 2.7.12 and sections 2.9.1 2.9.3; Barnes et al, Purification of Immunoglobulin G (IgG) , in METHODS IN MOLECULAR BIOLOGY, Vol. 10, pages 79-104 (Humana Press 1992).

Once hybridoma clones that produce antibodies to an antigen of interest have been selected, further selection can be performed for clones that produce antibodies having a particular specificity.

In particular embodiments the methods for diagnosis herein described before can also be used for the detection of somatic mutations in patients which are suspected to suffer from ALS and/or FTLD. Such mutations (which are generally known as non-familial or non-hereditary or *de novo* mutations) can occur because of the expansion of the 5'-GGCCCC-3' repeat present in SEQ ID NO: 1 in several cells including neuronal cells. In a particular embodiment cerebrospinal fluid (CSF) is used to identify the hexanucleotide expansion repeat of the invention. Identification can be based on DNA based methods, protein based detection or mRNA based detection of the lower expression of C9ORF72 or the presence of the hexanucleotide repeat expansion in C9ORF72.

The invention also provides kits that can be used to perform a method provided herein (e.g. to determine whether or not the promoter of the C9ORF72 gene comprises the mutation of the invention). Such kits can include nucleic acid molecules (e.g., primer pairs or probes), antibodies (e.g., anti-C9ORF72 polypeptide antibodies), secondary antibodies, control nucleic acid molecules (e.g. C9ORF72 nucleic acids that do or do not contain a mutation), control polypeptides (e.g. wild type or mutant C9ORF72 polypeptides), DNA aptamers, microarrays, ELISA plates, or data analysis software optionally together with any other appropriate reagent, tool, or instruction for performing the methods described herein. Appropriate informational material can be descriptive, instructional, marketing or other material that relates to the methods described herein and/or the use of the reagents for the methods described herein. For example, the informational material can relate to performing a genetic analysis on a human and subsequently diagnosing the human as being at risk (or not) for ALS, FTLD or ALS and FTLD, and/or delivering a prognosis of the human relating to survival time, likelihood of responding to therapy, or quality of life. In addition, or in an alternative, the informational material of a kit can be contact information, e.g., a physical address, email address, website, or telephone number, where a user of the kit can obtain substantive information about

performing a genetic analysis and interpreting the results, particularly as they apply to a human's likelihood of developing dementia and a subsequent prognosis. The informational material of the kits can be in any form. In many cases, the informational material, e.g. instructions, can be provided in printed matter, e.g. a printed text, drawing, and/or photograph, e.g., a label or printed sheet. Informational material can be provided in other formats, such as Braille, computer readable material, video recording, or audio recording. Informational material can also be provided in any combination of formats. The kit can include one or more containers for the reagents for performing a genetic analysis, such as reagents for performing PCR, FISH, CGH, or any other method described herein. The kit can contain separate containers, dividers, or compartments for the reagents and informational material. A container can be labeled for use for the diagnosis and/or prognosis of a human relating to the development and treatment of dementia.

This document also provides methods and materials to assist medical or research professionals in determining whether or not a mammal has a hexanucleotide repeat expansion in the C9ORF72 nucleic acid. Medical professionals can be, for example, doctors, nurses, medical laboratory technologists, and pharmacists. Research professionals can be, for example, principle investigators, research technicians, postdoctoral trainees, and graduate students. A professional can be assisted by (1) determining the presence or absence of a mutation (*id est* the hexanucleotide repeat expansion of the invention) in a C9ORF72 nucleic acid in a sample, and (2) communicating information about the presence or absence of that mutation to that professional. Any appropriate method can be used to communicate information to another person (e.g., a professional). For example, information can be given directly or indirectly to a professional. In addition, any type of communication can be used to communicate the information. For example, mail, e-mail, telephone, and face-to-face interactions can be used. The information also can be communicated to a professional by making that information electronically available to the professional. For example, the information can be communicated to a professional by placing the information on a computer database such that the professional can access the information. In addition, the information can be communicated to a hospital, clinic, or research facility serving as an agent for the professional.

Still another aspect of the invention is the use of a mutation according to the invention in the construction of a non-human transgenic animal. Indeed, introduction of a mutation according to the invention in a model organism such as a mouse would lead to an reduction of SEQ ID NO: 2 and SEQ ID NO: 3 production and development of an ALS disease like phenotype or in the development of an FTLD disease like phenotype or the development of a combined ALS disease like and FTLD disease like phenotype. Methods to make mutant SEQ ID NO: 1 based transgenic animals have been described, amongst others, in GB2380196 and in WO9640895. Such model organism could be used for screening compounds and testing medication useful for treatment of ALS and/or FTLD. In yet another particular embodiment a knock-out transgenic animal can be constructed having a loss of C9ORF72. In a particular embodiment a knock-out animal is made being hemizygous for a loss of function of C9ORF72.

Examples

1. Family-based linkage to ALSFTD2 locus on chromosome 9p

Since the original reports of a Dutch and a Scandinavian ALS-FTLD family linked with chromosome 9p21 (Morita et al., 2006; Vance et al., 2006), a growing number of families with inherited ALS and FTLD are reported with significant linkage to the ALSFTD2 locus on chromosome 9p21 (Boxer et al., 2010; Gijselinck et al., 2010; Le Ber et al., 2009; Luty et al., 2008; Valdmanis et al., 2007) (table 1). In all these families patients show similar clinical and pathological characteristics. Clinically, individuals may present with symptoms of both ALS and FTLD, or with ALS or FTLD alone. Pathologically, autopsied patients have TDP-43 positive type 2 (Sampathu et al., 2006) brain inclusions (Boxer et al., 2010; Gijselinck et al., 2010; Le Ber et al., 2009; Luty et al., 2008; Morita et al., 2006; Vance et al., 2006) - (see Table 1).

Family	Origin	Max LOD score at 9p21	Mean onset age in years (range)	Mean disease duration in years (range)	TDP-43+	# ALS	# ALS + FTLD	# FTLD	References
Luty	Australian	3.41	53 (43-68)	9 (1-16)	+	2	2	7	(Luty et al., 2008)
DR14	Belgian	3.38	58.1 (51-65)	6.4 (1-17)	+	1	0	10	(Gijselinck et al., 2010)
F2	Dutch	3.02	60.3 (39-72)	3.0 (1-8)	ND	7	3	2	(Vance et al., 2006)
Que23	Canadian	3.01	55.8 (46-58)	2.4 (1.5-3)	ND	5	0	3	(Valdmanis et al., 2007)
VSM20	Irish	3.01	45.7 (35-57)	5.4 (3-10)	+	2	3	5	(Boxer et al., 2010)
F438	Scandinavian	3.00	55.3 (45-64)	4.3 (1-9)	ND	5	0	9	(Morita et al., 2006)
6 families	French	8.0[1]	57.9 (40-84)	3.6 (1-8)	+	9	12	10	(Le Ber et al., 2009)
Que1	French-Canadian	2.51	54.3 (45-63)	4.8 (2-9)	ND	5	3	0	(Valdmanis et al., 2007)
Fr104	Spanish	1.55	ND	ND	ND	4	1	0	(Valdmanis et al., 2007)
F2	North-American	1.5	ND	ND	ND	0	7	0	(Momeni et al., 2006)
Gwent	Brittish	ND	42.2 (31-52)	3.6 (1-13)	+	3	6	0	(Pearson et al., 2011)
F476	North-American	ND	ND	ND	ND	2	3	0	(Momeni et al., 2006)
ALS_A	American	ND	? (35-73)	? (0.5-5)	ND	6	0	0	(Krueger et al., 2009)

Table 1: Genetic, clinical and pathological characteristics of ALS-FTLD families linked or associated with chromosome 9p21 (ND: not determined; [1]summed LODscore in 6 small families, not linked separately)

The minimal candidate region was previously defined by D9S169 (Luty et al., 2008) and D9S1805 (Valdmanis et al., 2007) spanning 7 Mb and was recently reduced to 3.6 Mb between D9S169 (Luty et al., 2008) and D9S251 by Boxer and colleagues (2010) (figure 4).

2. Identification of a hexanucleotide repeat expansion in C9ORF72

In addition to conclusive genetic linkage of ALS/FTD to chromosome 9p in family DR14, we found highly significant genetic association with single nucleotide polymorphisms (SNPs) in cohorts of patients with ALS and ALS/FTD recruited from the Flanders region in Belgium. Haplotype association identified a 130 kb region that substantially reduced the published Finnish 42-SNP risk haplotype (Laaksovirta, H. PubMed: 20801718) leaving only three known positional candidate genes: MOBKL2B, IFNK and C9ORF72. Detailed association mapping using 58 SNPs tagging over 400 SNPs in this region, including sliding window association analysis indicated highly significant association with the 5' region of C9ORF72. Sanger sequencing of the coding regions of the three candidate genes and all evolutionary conserved elements in the candidate region (summed size > 60 kb) in patients of conclusively linked ALS/FTD family DR14 identified no mutations that segregated with disease while being absent in neurologically healthy control individuals. Also, second-generation sequencing using the cPAL technology of the complete region could not identify putative mutations. Sequence analyses were successful for the complete region except of C9ORF72 exon1A defined as the first, noncoding exon of the longer gene transcript (chr9:27573429-27573481 in the GRCh37/hg19 reference sequence of the human genome). Analysis of patient genomic DNA consequently obtained sequence of only 1 chromosomal copy due to failed PCR or absence of sequence reads. Also quantitative studies of the C9ORF72 exon1A genomic region in patients resulted in ambiguous data. Together, data obtained in genetic association studies, sequence analysis and quantitative genomic analyses suggested that a structural variation in or near exon1A was the most likely cause of ALS/FTLD. After exclusion the most obvious possibility of a genomic deletion, the possibility of an insertion was explored. Sequence inspection of the C9ORF72 exon1A genomic region indicated a low-complexity GC-rich region, partly overlapping the exon sequence (chr9:27573413-27573544 in the GRCh37/hg19 reference sequence of the human genome) and in the promoter region of the gene. The most 5' part of the low-complexity region was composed of 3 5/6 units of hexanucleotide repeat sequence GGGGCC. Therefore, the possibility of a hexanucleotide repeat expansion as the cause of chromosome 9p-linked ALS/FTD was explored using a repeat primer-based PCR detection assay. In patients of family DR14, the repeat was expanded from less than 30 repeat units in more than 800 control individuals of the Flanders region in Belgium to more than 80 repeat units, being the upper detection limit of the assay. The size range of the repeat did not show overlap between patients and controls. Analysis of the repeat size in 101 ALS patients, 23 patients with ALS and FTLD and 267 FTLD patients identified 33 additional patients with an expanded hexanucleotide repeat sized > 80 hexanucleotide units. Segregation analysis in 4 additional ALS/FTLD families demonstrated complete segregation of the expanded repeat with disease, suggesting an age-dependent, but complete penetrance of the mutation. The frequency of the repeat expansion in familial patients was 44% in ALS, 71% in ALS/FTLD and 10% in FTLD patients (see table 2). Genetic association of the repeat expansion with disease was more significant than the association with the risk haplotype and could fully explain the latter association, leaving no residual association evidence due to another risk variant in the same genomic region. Based on data on regulatory elements obtained in the ENCODE project (Encyclopedia of DNA Elements), the expanded repeat was located inside the promoter driving expression of C9ORF72. Therefore, the hypothesis of an effect of the repeat expansion on gene expression was evaluated (see Figure 5). qPCR analysis of random-primed and poly-A RNA of frontal cortex of two patients carrying the repeat expansion compared to frontal cortex RNA of control individuals demonstrated that in repeat expansion carriers, the C9ORF72

transcript was about 50% reduced after normalization to 4 housekeeping genes. Moreover, analysis of allele-specific expression of C9ORF72 in frontal cortex of repeat expansion carriers who are heterozygous for coding SNP rs10122902 demonstrated substantially reduced expression of the gene copy having the expanded repeat in its promoter. This demonstrates a direct, cis-acting effect of the repeat expansion on expression of C9ORF72.

Together these data suggest that haploinsufficiency of C9ORF72, due to reduced expression driven by a hexanucleotide repeat expansion in the gene promoter, is the cause of chromosome 9p linked ALS/FTLD. CORF72 is highly conserved in vertebrates. In worms (e.g. c. elegans), the corresponding gene exists of the C-terminal half only, while it is completely absent in insects (e.g. drosophila). It is an orphan gene with unknown function. Based on functions of genes with similar expression patterns in human and mouse, a function in protein ubiquitination for proteasomal degradation and/or a function in RNA processing are predicted.

Table 2: The frequency of the repeat expansion in familial patients was 44% in ALS, 71% in ALS/FTLD and 10% in FTLD patients

			repeat expansion carriers	
		Sample size	Number	Frequency
FTLD	Total	267	14	0.05
	Familial	89	9	0.10
	Sporadic or unknown	178	5	0.03
FTLD-ALS	Total	23	6	0.26
	Familial	7	5	0.71
	Sporadic or unknown	16	1	0.06
ALS	Total	101	13	0.13
	Familial	16	7	0.44
	Sporadic or unknown	85	6	0.07

3. Repeat primer-based PCR detection assay of repeat expansion

In a first series of PCR cycles, a fluorescently labeled forward primer F is used in combination with repeat-anchor primer R1. The R1 primer anneals at random positions in the repeat sequence of the genomic template. In subsequent series of PCR cycles, these fragments are further amplified using forward primer F in combination with anchor primer R2. Amplification products are separated on an Applied Biosystems ABI 3730 and analyzed. For each repeat unit present in the genomic template DNA, a differently sized PCR amplification fragment is

observed. Expanded repeats are characterized by a high number of different fragments. See Figure 1.

PCR Primers:

F: 5'-TCCTCACTCACCCACTCG-3'

R1: 5'-CGTACGCATCCCAGTTTGAGAGGGGCCGGGGCCGGGGCCGGGGC-3'

R2: 5'-CGTACGCATCCCAGTTTGAGA-3'

PCR Protocol

DNA (100ng/µl)	1µl	98°C 10'
Buffer 2	2µl	97°C 35"
dNTP (10mM) 1µl		10 cycli - 53°C - 35"
F primer (10µM)	0.66µl	68°C 2'
R1 primer (10µM)	0.066µl	97°C 35"
R2 primer (10µM)	0.66µl	25 cycli 53°C 35"
Betaine (3M)	13.86µl	68°C 2' (+20" for each successive) cycle)
Enzyme**	0.75µl	68°C 10'
		4°C forever

** Expand Long Template PCR system (Roche)

Examples of the analysis of genomic templates in neurologically healthy control individuals without repeat expansion are shown in Figure 2.

Example of the analysis of genomic templates in an ALS patient with repeat expansion is depicted in Figure 3.

4. A pan-European study of the C9orf72 repeat associated with FTLD and ALS: geographic prevalence, genomic instability and intermediate repeats

In this example, we determined the geographic prevalence of the pathological G_4C_2 expansion in a sizable pan-European cohort of FTLD and ALS patients (N=988), ascertained within a newly founded European Early-Onset Dementia (EOD) consortium. We obtained an overall mutation frequency of 8.81% in patients with a possible or probable diagnosis of FTLD and/or ALS. In the combined European and Flanders-Belgian cohorts (N=1489), we calculated

frequencies in familial patients with FTLD of 9.51%, with ALS of 38% and with FTLD plus ALS (FTLD-ALS) of 56.67%. In both cohorts we observed in G_4C_2 expansion carriers, a significant higher frequency ($p < 0.0001$) of short insertion/deletions in a low complexity sequence (LCS) adjacent to the G_4C_2 repeat and extending it by extra GC-rich repeat units, perhaps making it more prone to pathological expansion. Using a repeat-primed PCR assay that excludes the LCS and an allele genotyping assay, we identified G_4C_2 alleles with intermediate lengths of 8 to 25 repeat units. In genetic association studies in the Flanders-Belgian patient/control cohort, we obtained significant allelic association with the G_4C_2 intermediate repeats (OR=1.21, 95% CI (1.00-1.46); p=0.047), with the principal effect size in homozygous ALS patients (OR=2.32, 95% CI (1.15-4.68); p=0.019). These data indicate that G_4C_2 intermediate repeats might act as a pre-mutation and/or a susceptibility factor for disease.

4.1 The European EOD Consortium

The European cohort used to determine the contribution and distribution of the *C9orf72* repeat expansion mutation to FTLD and ALS in Europe, was collected through the European Early-Onset Dementia (EOD) consortium (Table 3). The European EOD consortium brings together epidemiological, clinical and biological data with biomaterials of EOD patients throughout Europe to stimulate high-profile translational dementia research. The consortium currently holds 40 partner groups from 16 countries: Belgium, France, the Netherlands, UK, Germany, Austria, Portugal, Spain, Italy, Sweden, Finland, Denmark, Czech Republic, Bulgaria, Greece, and Turkey (Figure 7). In this study, the European cohort included 1060 patients that were ascertained within the European EOD consortium of which 988 had a possible or probable diagnosis of FTLD (n=744), ALS (n=184) or FTLD-ALS (n=60) (Table 4). In an additional 72 other patients, clinical presentation showed indications of FTLD or ALS together with symptomatology of other neurodegenerative brain diseases such as Alzheimer or Parkinson disease. A pathological diagnosis on autopsied brain was obtained for 60 patients comprising FTLD-TDP (n=23), FTLD-UPS (n=1), FTLD-MND-TDP (n=15) and MND-TDP (n=21) pathological diagnoses. Information on family history of disease was available for 816 (82,59%) of the 988 European cohort patients of which 303 had a positive family history of disease and 513 had no indications of affected relatives and were considered sporadic patients (Table 5). Average onset age and onset age range were comparable between the FTLD and FTLD-ALS groups. In the ALS patient group average onset age and range were about 5 years earlier.

Table 3: European cohort - Patients per country overall and per clinical subgroup

Country	**Partner Group**	**FTLD**	**FTLD-ALS**	**ALS**	**Uncertain**	**Total**
	European EOD Consortium	**744 (33)** [4]	**60 (17)**	**184 (20)**	**72**	**1060 (60)**
Italy n = 402	Brescia, Brescia University	192	11	26		229
	Brescia, IRCCS	47	1		21	69
	Florence, FLO	84	5		6	95

Country	Partner Group	FTLD	FTLD-ALS	ALS	Uncertain	Total
	Verona, UNIVR	1 (1)				1
	Antwerp, DMG DSF [2]	4		4		8
Germany n = 283	Tubingen, HIH	38	8	82		128
	Munich, TUM	129	2			131
	Munich, LMU	2 (1)				2
	Bonn	20		1		21
	Antwerp, DMG DSF [2]			1		1
Portugal n = 193	Lisbon	77	3 (1)		34	114
	Coimbra	68	3	2	6	79
Spain n = 61	Barcelona, Brain bank	13 (13)	9 (9)	16 (16)		38
	Barcelona	16	6		0	22
	Antwerp, DMG DSF	1				1
Sweden n = 47	Stockholm, KI	37 (3)	2	2	4	45
	Antwerp, DMG DSF [2]	1		1		2
Bulgaria n = 31	Sofia, MU		3	28		31
Belgium n=26 [3]	Liège, CRC-ULG-CHU	1				1
	Antwerp, DMG DSF [2]	8		16	1	25
Czech Republic n = 13	Prague, TH	5 (5)	5 (5)	3 (3)		13
Austria n = 3	Vienna, MUV		2 (2)	1 (1)		3
Denmark n = 1	Antwerp, DMG DSF [2]			1		1

Note: [1] The European EOD consortium was launched in August 2011 and is coordinated by Christine Van Broeckhoven and Julie van der Zee, Department of Molecular Genetics, VIB, Antwerp, Belgium. Patients with clinical diagnosis of FTLD or ALS, or indications of FTLD or ALS symptoms are listed. The 1060 patients included [2]13 additional patients from other European countries that had been referred for clinical genetic testing to the Diagnostic Service Facility (DSF) in the VIB Department of Molecular Genetics, DMG DSF and [3] 26 patients from Wallonia, the French speaking part of Belgium. [4]Number of patients with autopsy confirmed neuropathology diagnosis is indicated between brackets behind the total and subgroup number of patients contributed per country.

Table 4: Descriptive characteristics of the European and Flanders-Belgian cohorts and clinical subgroups

Clinical diagnosis cohorts	Total n	Familial n (%)	Disease onset ± SD (range)	Pathology n (%)
European Cohort				
Total	988	303 (30.64)	61.4 ± 10.9 (20-88)	60 (6.07)
FTLD	744	246 (33.06)	62.8 ± 9.0 (28-88)	22 (2.95)
FTLD-ALS	60	23 (38.33)	61.1 ± 9.8 (31-83)	18 (8.60)
ALS	184	34 (18.48)	56.0 ± 15.1 (20-85)	20 (10.87)
Other [(1)]	72	40 (55.56)	60.0 ± 12.1 (35-95)	N.A.
Flanders-Belgian Cohort				
Total	501	124 (24.75)	62.0 ± 10.3 (27-85)	29 (5.79)
FTLD	337	101 (29.97)	63.0 ± 9.7 (29-85)	21 (6.23)
FTLD-ALS	23	7 (30.43)	62.7 ± 10.0 (39-75)	3 (13.04)
ALS	141	16 (11.35)	59.6 ± 11.4 (27-79)	5 (3.55)
Combined European and Flanders-Belgian cohorts				
Total	1489	427 (28.68)	61.6 ± 10.7 (20-88)	89 (5.98)
FTLD	1081	347 (32.10)	62.9 ± 9.2 (28-88)	43 (3.98)
FTLD-ALS	83	30 (36.14)	61.5 ± 9.9 (31-83)	21 (25.30)
ALS	325	50 (15.38)	57.6 ± 13.7 (20-85)	25 (7.69)

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

5

4.2 *C9orf72* pathological G_4C_2 expansions in the European cohort

To determine the impact and distribution of the pathological G_4C_2 expansion across Europe, G_4C_2 repeat lengths were determined by repeat-primed PCR (forward RP-PCR, figure 8) followed by fluorescent fragment length analysis. This assay is able to discriminate between the presence (minimal 60 repeat units) and absence (maximal 25 repeat units) of a pathological G_4C_2 expansion. A pathological G_4C_2 expansion was observed in 8.81% (87/988) of the European cohort (Table 5). Per clinical subgroup, 6.18% of FTLD, 33.33% of FTLD-ALS and 11.41% of ALS patients carried a pathological G_4C_2 expansion (Table 5). In familial patients, the overall frequency increased to 14.52% with subgroup frequencies of 8.54% (21/246) in FTLD, 47.83% (11/23) in FTLD-ALS and 35.29% (12/34) in ALS patients. In sporadic patients, the overall frequency decreased to 5.07% with subgroup frequencies of 3.64% (13/357) in FTLD, 18.52% (5/27) in FTLD-ALS and 6.20% (8/129) in ALS patients. All pathological G_4C_2 expansion carriers in the European cohort carried at least one T-allele of the single-nucleotide-polymorphism (SNP) rs2814707, the top SNP in the ALS genome-wide association studies and tagging the chromosome 9p21 risk haplotype (Laaksovirta et al (2010); Van Es MA et al (2009)). To evaluate the distribution of the pathological G_4C_2 expansion across the partnering countries we calculated, overall and per clinical phenotype, mutation frequencies per country contributing to the European cohort (Table 6). In most countries the pathological G_4C_2 expansion mutation showed a comparable overall frequency ranging from 5.87 to 8.75%. The Bulgarian patients showed an intermediate overall frequency of 12.90% driven by the predominance of ALS patients in this cohort. However, a marked overall enrichment was observed in the Spanish (21.31%) and Swedish (25.58%) patient cohorts. Looking at clinical subgroups, 30.00% was observed in the FTLD and 26.67% in the FTLD-ALS clinical subgroups from Spain, and 21.05% in the FTLD patients from Sweden.

Table 5: Frequencies of the *C9orf72* pathological G_4C_2 expansion in the European and Flanders-Belgian cohorts and clinical subtype groups

	FTLD		**FTLD-ALS**		**ALS**		**Total**	
European Cohort								
Total	46/744	6.18%	20/60	33.33%	21/184	11.41%	87/988	8.81%
Familial	21/246	8.54%	11/23	47.83%	12/34	35.29%	44/303	14.52%
Sporadic	13/357	3.64%	5/27	18.52%	8/129	6.20%	26/513	5.07%
Other [(1)]	12/141	8.51%	4/10	0.40%	1/21	4.76%	17/172	9.88%
Flanders-Belgian Cohort								
Total	21/337	6.23%	7/23	30.43%	13/141	9.22%	41/501	8.18%
Familial	12/101	11.88%	6/7	85.71%	7/16	43.75%	25/124	20.16%
Sporadic	9/236	3.81%	1/16	6.25%	6/125	4.80%	16/377	4.75%
Combined European and Flanders-Belgian cohorts								
Total	67/1081	6.20%	27/83	32.53%	34/325	10.46%	128/1489	8.60%

	FTLD		FTLD-ALS		ALS		Total	
Familial	33/347	9.51%	17/30	56.67%	19/50	38.00%	69/427	16.16
Sporadic	22/593	3.71%	6/43	13.95%	14/254	5.51%	42/890	4.72

Note: [1]Other includes patients of which there was no family history data available in the European cohort.

Table 6: *C9orf72* pathological G4C2 expansion carriers calculated per partnering European EOD country overall and per clinical subgroup

Country	**Total patients n[1]**	**Overall carriers (%)[2]**	**FTLD carriers (%)[2]**	**FTLD-ALS carriers (%)[2]**	**ALS carriers (%)[2]**
Total	1489	128/1489 (8.60)	67/1081 (6.20)	27/83 (32.53)	34/325 (10.46)
Belgium[3]	526	46/526 (8.75)	22/346 (6.36)	7/23 (30.43)	17/157 (10.83)
Italy	375	22/375 (5.87)	17/328 (5.18)	4/17 (23.53)	1/30 (3.33)
Germany	283	18/283 (6.36)	4/189 (2.12)	3/10 (30.00)	11/84 (13.10)
Portugal	153	11/153 (7.19)	7/145 (4.83)	3/6	1/2
Spain	61	13/61 (21.31)	9/30 (30.00)	4/15 (26.67)	N.A.
Sweden	43	11/43 (25.58)	8/38 (21.05)	1/2	2/3
Bulgaria	31	4/31 (12.90)	N.A.	3/3	1/28 (3.57)
Czech Republic	13	1/13 (7.69)	N.A.	1/5	N.A.

Note: [1]Total number of patients with a possible or probable diagnosis of disease; [2]Carrier percentages is calculated only for countries that contributed 10 or more patients to the European cohort and for subgroups with minimal 10 patients (Table S.1). [3]The Belgium sample includes apart from the Flanders-Belgian cohort an additional 25 patients ascertained in Wallonia-Belgium. N.A. Not applicable, no carriers were identified in these subgroups.

4.3 *C9orf72*-associated phenotype of the mutation carriers in the European cohort

Of the 87 carriers of a pathological G_4C_2 expansion, 46 had a clinical diagnosis of FTLD, 20 of FTLD-ALS and 21 of ALS (Table 7). The average onset age was 58.4 ± 8.0 years with an onset age range of 40 to 75 years and was comparable in the FTLD, FTLD-ALS and ALS subgroups (Table 7). The average duration of disease in 33 deceased carriers was 4.4 ± 3.3

years (range 1 to 14) but differed in the clinical subgroups with shortest survival for ALS carriers with on average 2.8 ± 1.6 years (n=12, range 1 to 6) and longest for FTLD patients with 6.2 ± 4.9 years (n=9, range 1 to 14 years). The survival of FTLD-ALS carriers was in between both other subgroups with 4.6 ± 3.5 years (n=12, range 1 to 14).

Table 7: Descriptive characteristics of G_4C_2 expansion mutation carriers in the European clinical subgroups

	European cohort			**Flanders -Belgian cohort**		
Clinical subgroup	**Total n**	**Average onset age years ± SD (range)**	**Average duration[(2)] years ± SD (range) (n)**	**Total n**	**Average onset age years ± SD (range)**	**Average duration years ± SD (range) (n)**
Total	87	58.4 ± 8.0 (40-75)	4.4 ± 3.3 (1-14) (n=33)	41	55.5 ± 8.9 (38-71)	4.9 ± 4.2 (1-17) (n=19)
FTLD	46	57.9 ± 7.6 (44-75)	6.2 ± 4.9 (1-14) (n=9)	21	55.3 ± 8.4 (42-71)	6.9 ± 4.7 (3-17) (n=9)
FTLD-ALS	20	57.4 ± 8.3 (40-71)	4.6 ± 3.5 (1-14) (n=12)	7	57.9 ± 9.7 (43-69)	2.9 ± 0.7 (2-4) (n=6)
ALS	21	60.5 ± 8.5 (43-74)	2.8 ± 1.6 (1-6) (n=12)	21	54.5 ± 9.9 (38-64)	3.3 ± 4.5 (1-10) (n=4)

Note: [(2)]Average duration of disease was calculated for n pathological G_4C_2 expansion carriers with information on age at death.

4.4 Genomic complexity in the *C9orf72* region adjacent to the G_4C_2 repeat

In the genomic sequence, the *C9orf72* G_4C_2 repeat is contiguous with a GC-rich, low complexity sequence (LCS), located upstream of exon 1a of the longest *C9orf72* isoform (Figure 8). We successfully sequenced the GC-rich LCS in 432 patients of the Flanders-Belgian cohort and in 752 matched control individuals by repeat-primed PCR sequencing (RP-PCR for sequencing, figure 8). We observed short deletions of 5 to 23 base pairs (bp) and insertions of 1 to 12 bp in a total of 25 individuals of which 13 patient carriers of a pathological G_4C_2 expansion (13/39 = 33.33%), 8 patient non-carriers (8/393 = 2.04%) and 4 control persons (4/752 = 0.53%) (Table 8). These variable insertion/deletions (indels) were significantly more frequently observed in carriers of a pathological G_4C_2 expansion compared to the group of non-carrier patients (OR 24.06, 95% 95% (9.16-63.23); p<0.0001) and control persons (OR 93.50, 95% CI(28.53-306.39); p<0.0001). Also, remarkably is that 10 of 13 (76.92%) patient carriers presented with the same 10-bp CCCCGACCAC deletion which was not observed in patient non-carriers and control persons (Table 8). The 10-bp deletion is contiguous with the G_4C_2 repeat and joins two GC-rich sequences and extends the GC-rich motif of the G_4C_2 repeat though with imperfect repeats (Figure 8). In this context, it is striking that deletion of the ACCAC motif between both GC-rich regions (Figure 8) was seen in 11 of the 39 patients carrying an expanded G4C2 repeat (28%), only once in the non-carrier patients

(1/393 = 0.25%) and once in control individuals (1/752=0.13%) (Table 8). To replicate these findings, we successfully sequenced the LCS in 68 patient carriers and 122 patient non-carriers from the European cohort. We observed a comparable high frequency of indels in the patient carriers (19/68 = 27.94%) and no indels in the non-carriers (0/122, <0.82%) (OR 46.92, 95% CI (6.11-360.14; p<0.0001) (Table 8). Ten of 19 (52.63%) patient carriers presented the same 10-bp deletion and in 14 of 19 (73.68%) the ACCAC sequence was deleted (Table 8). All carriers of the 10-bp deletion had at least one rs2814707 T-allele and the 10-bp deletion was co-transmitted with the pathological G_4C_2 expansion in available relatives of 2 pathological G_4C_2 expansion carriers of the Flanders-Belgian cohort, suggesting that the 10-bp LCS deletion is probably located on the disease haplotype.

Table 8: Low Complexity Sequence (LCS) indels in carriers of a pathological G_4C_2 expansion and in non-carriers

Diagnosis	Number	Indel [1]	Genomic position [2]
Flanders-Belgian cohort			
Patients expansion carriers (13/39)			
FTLD	7	del CCCCG**ACCAC**	27563512-27563521
FTLD-ALS	1	del CGGGCCCGCCCCG**ACCAC**GCCCC	27563504-27563526
FTLD-ALS	2	del CCCCG**ACCAC**	27563512-27563521
ALS	1	del CCCCG**ACCAC**	27563512-27563521
ALS	1	del CGGGCCCGCCCC	27563504-27563515
ALS	1	ins CGGGCCCGCCCC	27563516
Patients without expansion (8/393)			
FTLD	1	del **ACCAC**	27563517-27563521
FTLD	3	del GGCCCGCCCCG	27563506-27563516
FTLD-ALS	1	del GGCCCGCCCCCGGGCCCGCCCCG	27563494-27563516
ALS	1	del GGCCCGCCCCG	27563506-27563516
ALS	1	ins CGGGACCGCCCC	27563516
ALS	1	ins G	27563492
Controls (4/752)			
control	3	del GGCCCGCCCCG	27563506-27563516
control	1	del GCCCGCCCCG**ACCAC**G	27563507-27563522
European cohort			
Patients expansion carriers (19/68)			

FTLD	6	del CCCCG**ACCAC**	27563512-27563521
FTLD	1	del **ACCAC**	27563517-27563521
FTLD	1	del CCCCCGGGCCCGCCCCG**ACCAC**	27563500-27563521
FTLD	1	ins C	27563493
FTLD	1	ins CGGGCCCGCCCC	27563516
FTLD	1	del CGGGCCCGCCCC	27563504-27563515
FTLD-ALS	1	del CCCCG**ACCAC**	27563512-27563521
FTLD-ALS	1	del GCCCGCCCCCGGGCCCGCCCCG**ACCAC**G	27563495-27563522
FTLD-ALS	1	del GGCCCGCCCCG	27563506-27563516
ALS	3	del CCCCG**ACCAC**	27563512-27563521
ALS	1	del GCCCCCGGGCCCGCCCCG**ACCAC**GCCCC	27563498-27563526
ALS	1	del CGGGCCCGCCCC	27563504-27563515

Note:[(1)] The nucleotide sequence ACCAC is most frequently deleted in the LCS adjacent to the G4C2 repeat in C9orf72 (indicated in bold)[(2)] genomic location is based on NCBIbuild37 - hg19

The LCS adjacent to the G_4C_2 repeat is comprised in the PCR fragments produced by the forward RP-PCR assay that we used to identify pathological G_4C_2 expansion carriers (Figure 8). To eliminate possible influences of LCS variability when assaying the G_4C_2 for the presence of pathological expansions or for sizing the repeat for normal alleles, we developed a reverse RP-PCR assay on the sense strand that eliminates the LCS from the PCR amplicon (reverse RP-PCR, Figure 8). This RP-PCR assay confirmed the presence of the pathological G_4C_2 expansion in patients of the Flanders-Belgian as well as the European cohort.

4.5 Sizing of normal *C9orf72* G_4C_2 repeat lengths

We used the reverse RP-PCR assay to more precisely determine the length of the longer normal G_4C_2 alleles. In addition, we used a STR genotyping assay (STR-PCR, Figure 8) optimized for high GC-content, to size the normal alleles in non-carriers and the wild-type allele in carriers of a G_4C_2 expansion mutation. When we compared the allele lengths obtained with the two assays in the Flanders-Belgian cohort, we obtained a 99% match. The remaining 1%, we could match when we adjusted normal allele size in subjects in whom we had observed a LCS indel by sequencing. This implies that the STR-PCR alleles can be used for correct sizing of normal repeat alleles. The observed lengths of the G_4C_2 repeat ranged from 3 to 25 units in the Flanders-Belgian cohort (Figure 9). In the European cohort, the repeat sizes ranged between 3 and 22 units. In our study, we observed on the shortest allele 2 perfect G_4C_2 repeat units and 1 imperfect repeat G_4C_1 in contrast to the reference sequence which shows 3 perfect G_4C_2 repeat units next to the imperfect G_4C_1 repeat (NCBIbuild37 - hg19) (Figure 8).

4.6 Genetic association with *C9orf72* intermediate G4C2 alleles

Visual comparison of the frequency distribution of normal repeat alleles in Flanders-Belgian patients (n=419) and control individuals (n=1005) (Figure 9), suggested 3 groups of alleles with a comparable pattern of frequency distribution (3-4 units (50.8% in patients versus 53.1% in controls), 5-6-7 units (23.0% in patients versus 24.4% in controls) and 8-25 units (26.3% in patients versus 22.5% in controls). We did not detect significant allelic association between groups 1 and 2. Consequently, we calculated genetic association with one group of short (3 to 7 repeat units) and a second group of intermediate (8 to 25 repeat units) alleles and disease. We obtained significant allelic association in the overall cohort (OR = 1.21; 95% CI 1.00 - 1.46, p=0.047) and genotypic association in homozygous carriers of intermediate repeat alleles (OR = 1.65; 95% CI 1.01 – 2.70, p=0.047) (Table 9). The risk effect was derived mainly from the ALS subgroup in which we calculated significant allelic (OR = 1.39; 95% CI 1.03 – 1.87, p=0.032) and genotypic association in homozygous carriers of intermediate repeat alleles (OR = 2.32; 95% CI 1.15 – 4.68, p=0.019). In the FTLD subgroup a slight increase in frequency of intermediate alleles (25.0%) was observed compared to control individuals (22.5%) (Table 9).

Table 9: Allelic and genotypic association of *C9orf72* G_4C_2 intermediate repeats in the Flanders-Belgian patient/control cohort

Phenotype	**Genotype**	**Controls**		**Patients**		**Logistic regression**		**EU patients**	
		N	**%**	**N**	**%**	**OR (95%CI)**	**p-value**	**N**	**%**
Overall	S	1559	77.5	620	73.7	1.21 (1.00-1.46)	**0.047**	1303	75.9
	I	453	22.5	221	**26.3**			413	24.1
	S/S	597	59.4	230	54.9	ref		493	57.5
	S/I	363	36.1	159	37.9	1.13 (0.89-1.44)	0.322	317	36.9
	I/I	45	4.5	30	7.2	1.65 (1.01-2.70)	**0.047**	48	5.6
FTLD	S	1559	77.5	423	75.0	1.14 (0.92-1.42)	0.236	997	74.6
	I	453	22.5	141	**25.0**			339	25.4
	S/S	597	59.4	158	56.0	ref		371	55.5
	S/I	363	36.1	107	37.9	1.11 (0.84-1.47)	0.471	255	38.2
	I/I	45	4.5	17	6.0	1.41 (0.78-2.55)	0.253	42	6.3
ALS	S	1559	77.5	174	71.0	1.39 (1.03-1.87)	**0.032**	237	78.5
	I	453	22.5	71	**29.0**			65	21.5
	S/S	597	59.4	64	52.9	ref		91	60.3
	S/I	363	36.1	45	37.2	1.16 (0.77-1.75)	0.465	55	36.4
	I/I	45	4.5	12	**9.9**	2.32 (1.15-4.68)	**0.019**	5	3.3

Note: OR = odds ratio, 95% CI = 95% confidence interval. Patients with a known mutation in an FTLD or ALS gene (including pathological C9orf72 G4C2 expansion carriers) were excluded from all association analyses. p-values were calculated using logistic regression analysis and were corrected for age and gender. S = short repeat alleles with < 8 repeat units; I = Intermediate repeat alleles with ≥ 8 repeat units.

4.7 Intermediate repeat length and 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=0.001), located on the risk haplotype. Most of the risk association we could attribute to the carriers of a pathological G_4C_2 expansion, but in the remaining cohort without carriers a residual association signal remained for homozygous carriers of the T-allele (OR = 1.88; 95% CI 1.17 – 3.03; p = 0.009) (Table 10). We used 3 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 successfully sequencing all exons in 493 patients and by exon-based quantitative analysis of putative deletions/duplications in 413 patients. Except for 3 missense mutations – p.T49R, p.T66S and p.T375I, without a clear *in silico* predicted deleterious effect, no other potentially pathogenic mutations were identified (Table 11). Next, we calculated genetic association after excluding individuals homozygous for intermediate repeat alleles (n=122), and showed that the residual association completely disappeared (p=0.11) (Table 10). Accordingly, when we compared the distribution of normal repeat lengths between control persons homozygous for the rare T-allele and control persons homozygous for the common C-allele (Figure 10), we observed a highly significant allelic association between the presence of an intermediate repeat allele and the T-allele (p<0.0001) (Figure 10).

Table 10: Association with rs2814707: effect with and without intermediate repeat carriers.

Phenotype	**Genotype**	**Controls**		**Patients**		**Logistic regression**	
		N	**(%)**	**N**	**(%)**	**OR (95% CI)**	**p-value**
All patients and controls							
Total	CC	597	58.2	228	55.7	ref	
	CT	382	37.2	147	35.9	0.99 (0.78-1.27)	0.965
	TT	46	4.4	34	8.3	1.88 (1.17-3.03)	**0.009**
FTLD	CC	597	58.2	155	56.6	ref	
	CT	382	37.2	98	35.8	0.98 (0.73-1.30)	0.864
	TT	46	4.4	21	7.7	1.77 (1.02-3.06)	**0.043**
ALS	CC	597	58.2	65	53.7	ref	
	CT	382	37.2	44	36.4	1.05 (0.70-1.58)	0.823
	TT	46	4.4	12	9.9	2.31 (1.14-4.66)	**0.020**

Phenotype	Genotype	Controls		Patients		Logistic regression	
		N	(%)	N	(%)	OR (95% CI)	p-value
Excluding *C9orf72* homozygous intermediate repeat length carriers							
Total	CC	570	60.8	224	59.9	ref	
	CT	358	38.2	143	38.2	1.02 (0.79-1.31)	0.906
	TT	10	1.1	7	1.9	2.33 (0.83-6.56)	0.110
FTLD	CC	570	60.8	154	60.4	ref	
	CT	358	38.2	96	37.6	0.99 (0.74-1.32)	0.946
	TT	10	1.1	5	2	2.40 (0.77-7.50)	0.133
ALS	CC	570	60.8	70	58.8	ref	
	CT	358	38.2	47	39.5	1.07 (0.72-1.59)	0.736
	TT	10	1.1	2	1.7	2.23 (0.46-10.86)	0.321

Note: OR = odds ratio, 95% CI = 95% confidence interval. ORs were calculated for all patients, and after exclusion of patients and control individuals carrying two C9orf72 intermediate repeat alleles. Patients with a known mutation in an FTLD or ALS gene (including pathological C9orf72 G_4C_2 expansion carriers) were excluded from all association analyses. p-values were calculated using logistic regression analysis and were corrected for age and gender.

Table 11: Predictions *C9orf72* missense mutations

Variant	SIFT[1,2]		SNAP		SNPs&GO		MutPred	PMut	
	90%[3]	100%[3]	effect	RI[4]	effect	RI[4]	g-score[5]	Output[6]	R[4]
p.T49R	0.47	0.33	non-neutral	0	neutral	9	0.216	0.8493	6
p.T66S	0.11	0.07	non-neutral	1	neutral	9	0.682	0.1107	7
p.T250I	0.01	0.01	non-neutral	2	neutral	8	0.566	0.7668	5

[1]SIFT prediction was performed using automated SIFT homologue retrieval (from UniProt-TrEMBL 39.6 database), [2]Scores <0.05 are predicted to affect protein function while scores ≥0.05 are predicted to be tolerated, [3]For each set of entered homologues test sequences 90% or 100% identical sequences were removed, [4]RI=reliability index, ranging between 0 and 9, 0 being the least reliable, [5]The MutPred g-score implicates the probability that the amino acid substitution is deleterious/disease-associated. Mutations with a score above 0.75 are considered as confidently predicted to be deleterious,[6]PMut outputs >0.5 are predicted to be pathological. Pathological predictions of variants are indicated in bold.

MATERIALS AND METHODS

4.8 Study populations

European cohort

The European cohort was collected after an initial call for participation in a Pan-European study of the frequency across Europe of the *C9orf72* G_4C_2 expansion mutation from within the European Early-Onset Dementia (EOD) consortium (Table 3). Table 3 describes the number of patients per country and per clinical subgroup contributed by European EOD consortium members to the European cohort. We received DNA and clinical and demographic information on the 1060 patients as well as pathology data of 60 patients obtained at autopsy. The 1060 patients also included 26 patients from Wallonia, the French speaking part of Belgium, and 13 patients from different European countries that were referred for clinical genetic testing of causal genes to the Diagnostic Service Facility in our Department of Molecular Genetics, DMG DSF). Patients had been diagnosed according to established clinical diagnostic criteria i.e. the Neary criteria for FTLD and the El Escorial criteria for ALS (Neary D et al (1998); Brooks BR et al (2000)). For the neuropathology diagnosis the Mackenzie consensus criteria were followed (Mackenzie IR et al (2010)).

Flanders-Belgian cohort

Flanders-Belgian FTLD 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. Index patients were evaluated using a standard clinical diagnostic protocol, including detailed recording of clinical and family history, neurological examination, neuropsychological testing, biochemical analyses and neuroimaging. The ALS patients were recruited through the neuromuscular reference centers of the Antwerp University Hospital and the University Hospitals Leuven Gasthuisberg. Additional patients were included who had initially been referred to the DMG DSF for clinical genetic testing. Patients were diagnosed with behavioral variant FTD (bvFTD), progressive non-fluent aphasia (PNFA) or semantic dementia (SD) according to consensus diagnostic criteria for FTLD (Brooks BR et al (2000)); with definite, probable or laboratory supported probable ALS according to the revised El Escorial criteria (Mackenzie IR et al (2010)) or with concomitant FTLD-ALS. Postmortem analysis confirmed clinical diagnosis in 21 FTLD, 5 ALS and 3 FTLD-ALS patients. For the genetic studies, index patients were re-contacted by trained research nurses who performed extensive genealogical studies and investigated family history of the disease in the extended pedigrees with the help of medical records. Mutation analyses of FTLD (*C9orF72*, *GRN*, *MAPT*, *VCP* and *CHMP2B*) and ALS (*C9orF72*, *SOD1*, *TARDBP*, *FUS*, *ATXN2* and *VCP*) genes identified pathogenic mutations in 60 (16.7%) of the FTLD and 17 (12.1%) of the ALS patients. Age-matched community controls were recruited from the same geographical region of Flanders-Belgian as the patients. All were submitted to a medical survey inquiring about personal as well as family medical history of neurodegenerative or psychiatric disease, followed by a Mini Mental State Examination test (MMSE). Only participants free of the aforementioned diseases and with a MMSE score > 26 were included in the control population (n=1083).

Ethical assurance

All participants or their legal guard gave written informed consent for participation in the clinical and genetic studies, and for brain autopsy if appropriate. The clinical study protocol and the informed consent forms for patient ascertainment were approved by the local Medical Ethics Committees of the collaborating centers. The genetic and pathological study protocols and informed consent forms were approved by the Medical Ethics Committee of the University Hospital of Antwerp, Belgium.

4.9 *C9orf72* G_4C_2 genetic PCR assays for genotyping

To determine the frequency of the *C9orf72* pathological G_4C_2 expansion in the European cohort, the forward repeat-primed PCR (forward RP-PCR; Figure 8) assay was used as described [2]. In the reverse repeat-primed PCR assay (reverse RP-PCR; Figure 8), we used a locus-specific fluorescently labeled reverse primer (5'-AGTACTCGCTGAGGGTGAAC-3'), a first forward primer (F1, 5'-CGTACGCATCCCAGTTTGAGAGCCCCGGCCCCGGCCCCGG CCCC-3') consisting of the reference amount of GGGGCC repeat and an anchor and a second forward primer (F2, 5'-CGTACGCATCCCAGTTTGAGA-3') consisting of the anchor (reverse RP-PCR; Figure 2). We used the Expand Long Template PCR System (Roche Diagnostics) amplification protocol with an annealing temperature of 55°C. The resulting PCR products were separated and analyzed on an ABI 3730 automated sequencer (Applied Biosystems) and analyzed using an in-house developed TraCl genotyping software (http://www.vibgeneticservicefacility.be). For accurate sizing of the normal G_4C_2 repeat length, we developed a genotyping assay using a fluorescently labeled primer pair flanking the repeat (5'-CCAGCTTCGGTCAGAGAAAT-3' and 5'-CAGGTGTGGGTTTAGGAGGT-3') (Figure 8). The PCR amplification protocol was optimized for alleles with intermediate repeat length because of higher GC content. We used the KAPA HiFi HotStart DNA Polymerase with dNTPs kit (Kapa Biosystems) with the KAPA HiFi GC buffer, 1M betaine, 50ng genomic DNA, and 0.3µM of primers. The PCR cycling profile was as follows: denaturation at 95°C for 5 minutes, 33 cycles at 98°C for 20 seconds, 59°C for 15 seconds, and 72°C for 1 minute; and a final extension at 72°C for 5 minutes. The resulting PCR products were size separated and analyzed on an ABI 3730 automated sequencer (Applied Biosystems) with GENESCAN LIZ600 as a size standard (Applied Biosystems) and genotypes were assigned using an in-house developed TraCl genotyping software (http://www.vibgeneticservicefacility.be).

4.10 Sequencing of the *C9orf72* GC-rich LCS adjacent of the G_4C_2 repeat

We used the product of an alternative forward repeat-primed PCR (RP-PCR for sequencing; Figure 2) amplified with the Expand Long Template PCR System (Roche Diagnostics) amplification protocol under previously described conditions [2] with a gene-specific forward primer (5'-ATGCGTCGAGCTCTGAGGAG-3') and two reverse primers (R1, 5'-CGTACGCATCCCAGTTTGAGAGGGGCCGGGGCCGGGGCCGGGGC-3' and R2, 5'-CGTACGCATCCCAGTTTGAGA-3') (Figure 8). PCR amplicons were purified with 10U exonuclease I (USB Corporation) and 2U shrimp alkaline phosphatase (USB Corporation). Purified amplicons were sequenced with the forward primer using the BigDye Terminator Cycle Sequencing kit v3.1 (Applied Biosystems) and analyzed on an ABI 3730 DNA Analyzer (Applied Biosystems). Co-segregation of variations in the flanking repeat with the presence of a pathological G_4C_2 expansion was analyzed in two available families.

4.11 *C9orf72* exon sequencing

Primers were designed for all 11 exons and exon-intron boundaries and 5'UTR of both major isoforms of *C9orf72* using Primer 3 (Rosen S. et al (2000)). Standard PCRs on 20ng genomic DNA amplified exons and exon-intron boundaries with optimized conditions. PCR amplicons were purified using ExoSAP-IT® (USB Corporation, Cleveland, Ohio), sequenced in both directions using BigDye® Terminator Cycle Sequencing kit v3.1 (Applied Biosystems) and analyzed on an ABI3730xl DNA Analyzer (Applied Biosystems). Sequences were analyzed using NovoSNP software (Weckx S et al (2005). Exons containing rare mutations were also PCR sequenced in genomic DNA of 400 control individuals using the same primers and conditions as for the mutation analysis.

4.12 *C9orf72* exon-based dosage analysis

We screened the Flanders-Belgian cohort for *C9orf72* exonic deletions or duplications using the Multiplex Amplicon Quantification (MAQ) technique (Kumps C et al (2010). consisting of a multiplex PCR amplification of fluorescently labeled test and reference amplicons, followed by fragment analysis on an ABI 3730 DNA analyzer (Applied Biosystems). Thirteen test amplicons located in all exons of *C9orf72* and 16 reference amplicons located at randomly selected genomic positions outside known copy number variants (CNVs) were simultaneously PCR-amplified on 20 ng genomic DNA. Peak areas of the test amplicons were normalized to these of the reference amplicons. Comparison of normalized peak areas between a patient and control individuals resulted in a dosage quotient (DQ) for each test amplicon, calculated by the MAQ software package (http://www.vibgeneticservicefacility.be/MAQ.htm). DQ values below 0.75 or above 1.25 were considered indicative of a heterozygous deletion or duplication respectively.

4.13 Genetic association studies

We calculated association with disease of *C9orf72* intermediate G_4C_2 alleles (sized between 8 and 25 repeat units) with disease. Allelic and genotypic frequencies were compared between patients and control individuals. Odds ratios (OR) with 95% confidence interval (CI) were calculated in a logistic regression model, adjusted for age and gender. Analyses were performed in SPSS 16.0 (SPSS Inc., Chicago, IL). We had previously genotyped SNP rs2814707 tagging the chromosome 9p21 risk haplotype in the Flanders-Belgian patient and control cohorts. We calculated ORs with 95% CIs for rs2814707 genotypes in the different clinical subgroups, relative to the most frequent rs2814707 genotype CC in a logistic regression model, stratifying the patient and control groups by presence or absence of *C9orf72* intermediate repeats. Analyses were performed in SPSS 16.0 (SPSS Inc., Chicago, IL). In addition, we studied the correlation between the minor risk T-allele and intermediate repeat length. In the European cohort, rs2814707 was genotyped by direct PCR-based Sanger sequencing.

4.14 Onset age and survival analysis

In the patients without a pathological G_4C_2 repeat of the Flanders-Belgian cohort and of the European cohort, we calculated correlation between age at onset and repeat length using a non-parametric Kruskal-Wallis test. We further compared age at onset between short repeat length (3 – 7) and intermediate repeat (8 – 25) length in a Kaplan-Meier survival analysis. Also

the modifying effect of the wild-type allele of pathological G_4C_2 expansion carriers on onset age was assessed.

References

Arai, T., Hasegawa, M., Akiyama, H., Ikeda, K., Nonaka, T., Mori, H., Mann, D., Tsuchiya, K., Yoshida, M., Hashizume, Y. & Oda, T. (2006). TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Biochem Biophys Res Commun, 351, 3, 602-611

Baker, M., Mackenzie, I.R., Pickering-Brown, S.M., Gass, J., Rademakers, R., Lindholm, C., Snowden, J., Adamson, J., Sadovnick, A.D., Rollinson, S., Cannon, A., Dwosh, E., Neary, D., Melquist, S., Richardson, A., Dickson, D., Berger, Z., Eriksen, J., Robinson, T., Zehr, C., Dickey, C.A., Crook, R., McGowan, E., Mann, D., Boeve, B., Feldman, H. & Hutton, M. (2006). Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. Nature, 442, 7105, 916-919

Boxer, A.L., Mackenzie, I.R., Boeve, B.F., Baker, M., Seeley, W.W., Crook, R., Feldman, H., Hsiung, G.Y., Rutherford, N., Laluz, V., Whitwell, J., Foti, D., McDade, E., Molano, J., Karydas, A., Wojtas, A., Goldman, J., Mirsky, J., Sengdy, P., Dearmond, S., Miller, B.L. & Rademakers, R. (2010). Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family. J Neurol Neurosurg Psychiatry

Brooks, B.R., Miller, R.G., Swash, M. & Munsat, T.L. (2000). El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis. Amyotroph Lateral Scler Other Motor Neuron Disord, 1, 5, 293-299

Cooper, G.M., Zerr, T., Kidd, J.M., Eichler, E.E. & Nickerson, D.A. (2008). Systematic assessment of copy number variant detection via genome-wide SNP genotyping. Nat Genet, 40, 10, 1199-1203

Cruts, M., Gijselinck, I., van der Zee, J., Engelborghs, S., Wils, H., Pirici, D., Rademakers, R., Vandenberghe, R., Dermaut, B., Martin, J.J., van Duijn, C., Peeters, K., Sciot, R., Santens, P., de Pooter, T., Mattheijssens, M., Van den Broeck, M., Cuijt, I., Vennekens, K., De Deyn, P.P., Kumar-Singh, S. & Van Broeckhoven, C. (2006). Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. Nature, 442, 7105, 920-924

Drmanac, R., Sparks, A.B., Callow, M.J., Halpern, A.L., Burns, N.L., Kermani, B.G., Carnevali, P., Nazarenko, I., Nilsen, G.B., Yeung, G., Dahl, F., Fernandez, A., Staker, B., Pant, K.P., Baccash, J., Borcherding, A.P., Brownley, A., Cedeno, R., Chen, L., Chernikoff, D., Cheung, A., Chirita, R., Curson, B., Ebert, J.C., Hacker, C.R., Hartlage, R., Hauser, B., Huang, S., Jiang, Y., Karpinchyk, V., Koenig, M., Kong, C., Landers, T., Le, C., Liu, J., McBride, C.E., Morenzoni, M., Morey, R.E., Mutch, K., Perazich, H., Perry, K., Peters, B.A., Peterson, J., Pethiyagoda, C.L., Pothuraju, K., Richter, C., Rosenbaum, A.M., Roy, S., Shafto, J., Sharanhovich, U., Shannon, K.W., Sheppy, C.G., Sun, M., Thakuria, J.V., Tran, A., Vu, D., Zaranek, A.W., Wu, X., Drmanac, S., Oliphant, A.R., Banyai, W.C., Martin, B., Ballinger, D.G., Church, G.M. & Reid, C.A. (2010). Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science, 327, 5961, 78-81

Gijselinck, I., Engelborghs, S., Maes, G., Cuijt, I., Peeters, K., Mattheijssens, M., Joris, G., Cras, P., Martin, J.J., De Deyn, P.P., Kumar-Singh, S., Van Broeckhoven, C. & Cruts, M. (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, 5, 606-616

Gitcho, M.A., Baloh, R.H., Chakraverty, S., Mayo, K., Norton, J.B., Levitch, D., Hatanpaa, K.J., White, C.L., III, Bigio, E.H., Caselli, R., Baker, M., Al Lozi, M.T., Morris, J.C., Pestronk, A., Rademakers, R., Goate, A.M. & Cairns, N.J. (2008). TDP-43 A315T mutation in familial motor neuron disease. Ann Neurol, 63, 4, 535-538

Greenway, M.J., Andersen, P.M., Russ, C., Ennis, S., Cashman, S., Donaghy, C., Patterson, V., Swingler, R., Kieran, D., Prehn, J., Morrison, K.E., Green, A., Acharya, K.R., Brown, R.H., Jr. & Hardiman, O. (2006). ANG mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. Nat Genet, 38, 4, 411-413

Hutton, M., Lendon, C.L., Rizzu, P., Baker, M., Froelich, S., Houlden, H., Pickering-Brown, S., Chakraverty, S., Isaacs, A., Grover, A., Hackett, J., Adamson, J., Lincoln, S., Dickson, D., Davies, P., Petersen, R.C., Stevens, M., de Graaff, E., Wauters, E., van Baren, J., Hillebrand, M., Joosse, M., Kwon, J.M., Nowotny, P., Che, L.K., Norton, J., Morris, J.C., Reed, L.A., Trojanowski, J., Basun, H., Lannfelt, L., Neystat, M., Fahn, S., Dark, F., Tannenberg, T., Dodd, P.R., Hayward, N., Kwok, J.B., Schofield, P.R., Andreadis, A., Snowden, J., Craufurd, D., Neary, D., Owen, F., Oostra, B.A., Hardy, J., Goate, A., van Swieten, J., Mann, D., Lynch, T. & Heutink, P. (1998). Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature, 393, 6686, 702-705

Johnson, J.O., Mandrioli, J., Benatar, M., Abramzon, Y., Van Deerlin, V.M., Trojanowski, J.Q., Gibbs, J.R., Brunetti, M., Gronka, S., Wuu, J., Ding, J., McCluskey, L., Martinez-Lage, M., Falcone, D., Hernandez, D.G., Arepalli, S., Chong, S., Schymick, J.C., Rothstein, J., Landi, F., Wang, Y.D., Calvo, A., Mora, G., Sabatelli, M., Monsurro, M.R., Battistini, S., Salvi, F., Spataro, R., Sola, P., Borghero, G., Galassi, G., Scholz, S.W., Taylor, J.P., Restagno, G., Chio, A. & Traynor, B.J. (2010). Exome sequencing reveals VCP mutations as a cause of familial ALS. Neuron, 68, 5, 857-864

Kabashi, E., Valdmanis, P.N., Dion, P., Spiegelman, D., McConkey, B.J., Vande, V.C., Bouchard, J.P., Lacomblez, L., Pochigaeva, K., Salachas, F., Pradat, P.F., Camu, W., Meininger, V., Dupre, N. & Rouleau, G.A. (2008). TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis. Nat Genet, 40, 5, 572-574

Kumps C. et al (2010) Multiplex Amplicon Quantification (MAQ), a fast and efficient method for the simultaneous detection of copy number alterations in neuroblastoma. BMC Genomics 11: 298.

Kovacs, G.G., Murrell, J.R., Horvath, S., Haraszti, L., Majtenyi, K., Molnar, M.J., Budka, H., Ghetti, B. & Spina, S. (2009). TARDBP variation associated with frontotemporal dementia, supranuclear gaze palsy, and chorea. Mov Disord, 24, 12, 1843-1847

Krueger, K.A., Tsuji, S., Fukuda, Y., Takahashi, Y., Goto, J., Mitsui, J., Ishiura, H., Dalton, J.C., Miller, M.B., Day, J.W. & Ranum, L.P. (2009). SNP haplotype mapping in a small ALS family. PLoS One, 4, 5, e5687

Kwiatkowski, T.J., Jr., Bosco, D.A., Leclerc, A.L., Tamrazian, E., Vanderburg, C.R., Russ, C., Davis, A., Gilchrist, J., Kasarskis, E.J., Munsat, T., Valdmanis, P., Rouleau, G.A., Hosler, B.A., Cortelli, P., de Jong, P.J., Yoshinaga, Y., Haines, J.L., Pericak-Vance, M.A., Yan, J., Ticozzi, N., Siddique, T., McKenna-Yasek, D., Sapp, P.C., Horvitz, H.R., Landers, J.E. &

Brown, R.H., Jr. (2009). Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis. Science, 323, 5918, 1205-1208

Laaksovirta, H., Peuralinna, T., Schymick, J.C., Scholz, S.W., Lai, S.L., Myllykangas, L., Sulkava, R., Jansson, L., Hernandez, D.G., Gibbs, J.R., Nalls, M.A., Heckerman, D., Tienari, P.J. & Traynor, B.J. (2010). Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. Lancet Neurol, 9, 10, 978-985

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., Martinaud, O., Lacomblez, L., Vercelletto, M., Salachas, F., Sellal, F., Didic, M., Thomas-Anterion, C., Puel, M., Michel, B.F., Besse, C., Duyckaerts, C., Meininger, V., Campion, D., Dubois, B. & Brice, A. (2009). Chromosome 9p-linked families with frontotemporal dementia associated with motor neuron disease. Neurology, 72, 19, 1669-1676

Lillo, P. & Hodges, J.R. (2009). Frontotemporal dementia and motor neurone disease: overlapping clinic-pathological disorders. J Clin Neurosci, 16, 9, 1131-1135

Lomen-Hoerth, C., Murphy, J., Langmore, S., Kramer, J.H., Olney, R.K. & Miller, B. (2003). Are amyotrophic lateral sclerosis patients cognitively normal? Neurology, 60, 7, 1094-1097

Luty, A.A., Kwok, J.B., Dobson-Stone, C., Loy, C.T., Coupland, K.G., Karlstrom, H., Sobow, T., Tchorzewska, J., Maruszak, A., Barcikowska, M., Panegyres, P.K., Zekanowski, C., Brooks, W.S., Williams, K.L., Blair, I.P., Mather, K.A., Sachdev, P.S., Halliday, G.M. & Schofield, P.R. (2010). Sigma nonopioid intracellular receptor 1 mutations cause frontotemporal lobar degeneration-motor neuron disease. Ann Neurol, 68, 5, 639-649

Luty, A.A., Kwok, J.B., Thompson, E.M., Blumbergs, P., Brooks, W.S., Loy, C.T., Dobson-Stone, C., Panegyres, P.K., Hecker, J., Nicholson, G.A., Halliday, G.M. & Schofield, P.R. (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, Revesz T, Rozemuller AJ, Kumar-Singh S, Akiyama H, Baborie A, Spina S, Dickson DW, Trojanowski JQ, Mann DM (2010) Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update. Acta Neuropathol 119: 1-4.

Momeni, P., Schymick, J., Jain, S., Cookson, M.R., Cairns, N.J., Greggio, E., Greenway, M.J., Berger, S., Pickering-Brown, S., Chio, A., Fung, H.C., Holtzman, D.M., Huey, E.D., Wassermann, E.M., Adamson, J., Hutton, M.L., Rogaeva, E., George-Hyslop, P., Rothstein, J.D., Hardiman, O., Grafman, J., Singleton, A., Hardy, J. & Traynor, B.J. (2006). Analysis of IFT74 as a candidate gene for chromosome 9p-linked ALS-FTD. BMC Neurol, 6, 44

Morita, M., Al Chalabi, A., Andersen, P.M., Hosler, B., Sapp, P., Englund, E., Mitchell, J.E., Habgood, J.J., de Belleroche, J., Xi, J., Jongjaroenprasert, W., Horvitz, H.R., Gunnarsson, L.G. & Brown, R.H., Jr. (2006). A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia. Neurology, 66, 6, 839-844

Neary, D., Snowden, J.S., Gustafson, L., Passant, U., Stuss, D., Black, S., Freedman, M., Kertesz, A., Robert, P.H., Albert, M., Boone, K., Miller, B.L., Cummings, J. & Benson, D.F.

(1998). Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology, 51, 6, 1546-1554

Neumann, M., Sampathu, D.M., Kwong, L.K., Truax, A.C., Micsenyi, M.C., Chou, T.T., Bruce, J., Schuck, T., Grossman, M., Clark, C.M., McCluskey, L.F., Miller, B.L., Masliah, E., Mackenzie, I.R., Feldman, H., Feiden, W., Kretzschmar, H.A., Trojanowski, J.Q. & Lee, V.M. (2006). Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. Science, 314, 5796, 130-133

Pearson, J.P., Williams, N.M., Majounie, E., Waite, A., Stott, J., Newsway, V., Murray, A., Hernandez, D., Guerreiro, R., Singleton, A.B., Neal, J. & Morris, H.R. (2011). Familial frontotemporal dementia with amyotrophic lateral sclerosis and a shared haplotype on chromosome 9p. J Neurol, 258, 4, 647-655

Ravasi, T., Suzuki, H., Cannistraci, C.V., Katayama, S., Bajic, V.B., Tan, K., Akalin, A., Schmeier, S., Kanamori-Katayama, M., Bertin, N., Carninci, P., Daub, C.O., Forrest, A.R., Gough, J., Grimmond, S., Han, J.H., Hashimoto, T., Hide, W., Hofmann, O., Kamburov, A., Kaur, M., Kawaji, H., Kubosaki, A., Lassmann, T., van, N.E., MacPherson, C.R., Ogawa, C., Radovanovic, A., Schwartz, A., Teasdale, R.D., Tegner, J., Lenhard, B., Teichmann, S.A., Arakawa, T., Ninomiya, N., Murakami, K., Tagami, M., Fukuda, S., Imamura, K., Kai, C., Ishihara, R., Kitazume, Y., Kawai, J., Hume, D.A., Ideker, T. & Hayashizaki, Y. (2010). An atlas of combinatorial transcriptional regulation in mouse and man. Cell, 140, 5, 744-752

Ringholz, G.M., Appel, S.H., Bradshaw, M., Cooke, N.A., Mosnik, D.M. & Schulz, P.E. (2005). Prevalence and patterns of cognitive impairment in sporadic ALS. Neurology, 65, 4, 586-590

Roach, J.C., Glusman, G., Smit, A.F., Huff, C.D., Hubley, R., Shannon, P.T., Rowen, L., Pant, K.P., Goodman, N., Bamshad, M., Shendure, J., Drmanac, R., Jorde, L.B., Hood, L. & Galas, D.J. (2010). Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing. Science

Rollinson, S., Mead, S., Snowden, J., Richardson, A., Rohrer, J., Halliwell, N., Usher, S., Neary, D., Mann, D., Hardy, J. & Pickering-Brown, S. (2011). Frontotemporal lobar degeneration genome wide association study replication confirms a risk locus shared with amyotrophic lateral sclerosis. Neurobiol Aging, 32, 4, 758-7

Rosen, D.R., Siddique, T., Patterson, D., Figlewicz, D.A., Sapp, P., Hentati, A., Donaldson, D., Goto, J., O'Regan, J.P., Deng, H.X. & . (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. Nature, 362, 6415, 59-62

Rosen S, Skaletsky H (2000) prmer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 132: 365-386.

Rosso, S.M., Donker, K.L., Baks, T., Joosse, M., de, K., I, Pijnenburg, Y., de Jong, D., Dooijes, D., Kamphorst, W., Ravid, R., Niermeijer, M.F., Verheij, F., Kremer, H.P., Scheltens, P., van Duijn, C.M., Heutink, P. & van Swieten, J.C. (2003). Frontotemporal dementia in The Netherlands: patient characteristics and prevalence estimates from a population-based study. Brain, 126, Pt 9, 2016-2022

Rowland, L.P. & Shneider, N.A. (2001). Amyotrophic lateral sclerosis. N Engl J Med, 344, 22, 1688-1700

Sampathu, D.M., Neumann, M., Kwong, L.K., Chou, T.T., Micsenyi, M., Truax, A., Bruce, J., Grossman, M., Trojanowski, J.Q. & Lee, V.M. (2006). Pathological heterogeneity of frontotemporal lobar degeneration with ubiquitin-positive inclusions delineated by ubiquitin immunohistochemistry and novel monoclonal antibodies. Am J Pathol, 169, 4, 1343-1352

Sela, N., Mersch, B., Gal-Mark, N., Lev-Maor, G., Hotz-Wagenblatt, A. & Ast, G. (2007). Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome. Genome Biol, 8, 6, R127

Shatunov, A., Mok, K., Newhouse, S., Weale, M.E., Smith, B., Vance, C., Johnson, L., Veldink, J.H., van Es, M.A., van den Berg, L.H., Robberecht, W., Van, D.P., Hardiman, O., Farmer, A.E., Lewis, C.M., Butler, A.W., Abel, O., Andersen, P.M., Fogh, I., Silani, V., Chio, A., Traynor, B.J., Melki, J., Meininger, V., Landers, J.E., McGuffin, P., Glass, J.D., Pall, H., Leigh, P.N., Hardy, J., Brown, R.H., Jr., Powell, J.F., Orrell, R.W., Morrison, K.E., Shaw, P.J., Shaw, C.E. & Al-Chalabi, A. (2010). Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. Lancet Neurol, 9, 10, 986-994

Skibinski, G., Parkinson, N.J., Brown, J.M., Chakrabarti, L., Lloyd, S.L., Hummerich, H., Nielsen, J.E., Hodges, J.R., Spillantini, M.G., Thusgaard, T., Brandner, S., Brun, A., Rossor, M.N., Gade, A., Johannsen, P., Sorensen, S.A., Gydesen, S., Fisher, E.M. & Collinge, J. (2005). Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia. Nat Genet, 37, 8, 806-808

Sreedharan, J., Blair, I.P., Tripathi, V.B., Hu, X., Vance, C., Rogelj, B., Ackerley, S., Durnall, J.C., Williams, K.L., Buratti, E., Baralle, F., de Belleroche, J., Mitchell, J.D., Leigh, P.N., Al Chalabi, A., Miller, C.C., Nicholson, G. & Shaw, C.E. (2008). TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. Science, 319, 5870, 1668-1672

Valdmanis, P.N., Dupre, N., Bouchard, J.P., Camu, W., Salachas, F., Meininger, V., Strong, M. & Rouleau, G.A. (2007). Three families with amyotrophic lateral sclerosis and frontotemporal dementia with evidence of linkage to chromosome 9p. Arch Neurol, 64, 2, 240-245

Van Deerlin, V.M., Leverenz, J.B., Bekris, L.M., Bird, T.D., Yuan, W., Elman, L.B., Clay, D., Wood, E.M., Chen-Plotkin, A.S., Martinez-Lage, M., Steinbart, E., McCluskey, L., Grossman, M., Neumann, M., Wu, I.L., Yang, W.S., Kalb, R., Galasko, D.R., Montine, T.J., Trojanowski, J.Q., Lee, V.M., Schellenberg, G.D. & Yu, C.E. (2008). TARDBP mutations in amyotrophic lateral sclerosis with TDP-43 neuropathology: a genetic and histopathological analysis. Lancet Neurol, 7, 5, 409-416

Van Deerlin, V.M., Sleiman, P.M., Martinez-Lage, M., Chen-Plotkin, A., Wang, L.S., Graff-Radford, N.R., Dickson, D.W., Rademakers, R., Boeve, B.F., Grossman, M., Arnold, S.E., Mann, D.M., Pickering-Brown, S.M., Seelaar, H., Heutink, P., van Swieten, J.C., Murrell, J.R., Ghetti, B., Spina, S., Grafman, J., Hodges, J., Spillantini, M.G., Gilman, S., Lieberman, A.P., Kaye, J.A., Woltjer, R.L., Bigio, E.H., Mesulam, M., Al-Sarraj, S., Troakes, C., Rosenberg, R.N., White, C.L., III, Ferrer, I., Llado, A., Neumann, M., Kretzschmar, H.A., Hulette, C.M., Welsh-Bohmer, K.A., Miller, B.L., Alzualde, A., de Munain, A.L., McKee, A.C., Gearing, M., Levey, A.I., Lah, J.J., Hardy, J., Rohrer, J.D., Lashley, T., Mackenzie, I.R., Feldman, H.H., Hamilton, R.L., Dekosky, S.T., van der Zee, J., Kumar-Singh, S., Van, B.C., Mayeux, R., Vonsattel, J.P., Troncoso, J.C., Kril, J.J., Kwok, J.B., Halliday, G.M., Bird, T.D., Ince, P.G., Shaw, P.J., Cairns, N.J., Morris, J.C., McLean, C.A., DeCarli, C., Ellis,

W.G., Freeman, S.H., Frosch, M.P., Growdon, J.H., Perl, D.P., Sano, M., Bennett, D.A., Schneider, J.A., Beach, T.G., Reiman, E.M., Woodruff, B.K., Cummings, J., Vinters, H.V., Miller, C.A., Chui, H.C., Alafuzoff, I., Hartikainen, P., Seilhean, D., Galasko, D., Masliah, E., Cotman, C.W., Tunon, M.T., Martinez, M.C., Munoz, D.G., Carroll, S.L., Marson, D., Riederer, P.F., Bogdanovic, N., Schellenberg, G.D., Hakonarson, H., Trojanowski, J.Q. & Lee, V.M. (2010). Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. Nat Genet, 42, 3, 234-239

van Es, M.A., Veldink, J.H., Saris, C.G., Blauw, H.M., van Vught, P.W., Birve, A., Lemmens, R., Schelhaas, H.J., Groen, E.J., Huisman, M.H., Van Der Kooi, A.J., De, V.M., Dahlberg, C., Estrada, K., Rivadeneira, F., Hofman, A., Zwarts, M.J., van Doormaal, P.T., Rujescu, D., Strengman, E., Giegling, I., Muglia, P., Tomik, B., Slowik, A., Uitterlinden, A.G., Hendrich, C., Waibel, S., Meyer, T., Ludolph, A.C., Glass, J.D., Purcell, S., Cichon, S., Nothen, M.M., Wichmann, H.E., Schreiber, S., Vermeulen, S.H., Kiemeney, L.A., Wokke, J.H., Cronin, S., McLaughlin, R.L., Hardiman, O., Fumoto, K., Pasterkamp, R.J., Meininger, V., Melki, J., Leigh, P.N., Shaw, C.E., Landers, J.E., Al-Chalabi, A., Brown, R.H., Jr., Robberecht, W., Andersen, P.M., Ophoff, R.A. & van den Berg, L.H. (2009). Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. Nat Genet, 41, 10, 1083-1087

Van Langenhove, T., van der Zee, J., Sleegers, K., Engelborghs, S., Vandenberghe, R., Gijselinck, I., Van den Broeck, M., Mattheijssens, M., Peeters, K., De Deyn, P.P., Cruts, M. & Van, B.C. (2010). Genetic contribution of FUS to frontotemporal lobar degeneration. Neurology, 74, 5, 366-371

Vance, C., Al Chalabi, A., Ruddy, D., Smith, B.N., Hu, X., Sreedharan, J., Siddique, T., Schelhaas, H.J., Kusters, B., Troost, D., Baas, F., de, J., V & Shaw, C.E. (2006). Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3. Brain, 129, Pt 4, 868-876

Vance, C., Rogelj, B., Hortobagyi, T., De Vos, K.J., Nishimura, A.L., Sreedharan, J., Hu, X., Smith, B., Ruddy, D., Wright, P., Ganesalingam, J., Williams, K.L., Tripathi, V., Al Saraj, S., Al Chalabi, A., Leigh, P.N., Blair, I.P., Nicholson, G., de Belleroche, J., Gallo, J.M., Miller, C.C. & Shaw, C.E. (2009). Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. Science, 323, 5918, 1208-1211

Watts, G.D., Wymer, J., Kovach, M.J., Mehta, S.G., Mumm, S., Darvish, D., Pestronk, A., Whyte, M.P. & Kimonis, V.E. (2004). Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. Nat Genet, 36, 4, 377-381

Weckx S et al (2005) novoSNP, a novel computational tool for sequence variation discovery. Genome Res. 15: 436-442.

Yan J, Slifer S, Siddique N, Chen W, Yong S, Erdong L, Haines JL, Pericak-Vance M, Siddique T. 2008. Fine-Mapping and Candidate Gene Sequencing of the Chromosome 9p Locus of ALS/FTD.

Yokoseki, A., Shiga, A., Tan, C.F., Tagawa, A., Kaneko, H., Koyama, A., Eguchi, H., Tsujino, A., Ikeuchi, T., Kakita, A., Okamoto, K., Nishizawa, M., Takahashi, H. & Onodera, O. (2008). TDP-43 mutation in familial amyotrophic lateral sclerosis. Ann Neurol, 63, 4, 538-542

Claims

1. A method for diagnosing ALS, FTLD, ALS and FTLD or susceptibility to ALS, FTLD, ALS and FTLD comprising detecting the presence of a hexanucleotide repeat expansion of 5'-GGGGCC-3' in SEQ ID NO: 1 in a sample comprising nucleic acids from a subject, and diagnosing the subject as having or being susceptible to ALS, FTLD, ALS and FTLD when said hexanucleotide repeat expansion is present.
2. A method according to claim 1 wherein said hexanucleotide repeat expansion comprises more than 60 repeats.
3. A method according to claim 1 wherein said hexanucleotide expansion comprises more than 80 repeats.
4. A method according to any one of claims 1, 2 or 3 wherein said diagnosis is based on a PCR method.
5. A method according to claim 4 wherein said diagnosis is based on a repeat primer-based PCR detection assay.
6. A method according to any one of claims 1, 2 or 3 wherein the presence of the hexanucleotide repeat expansion leads to a reduced presence of SEQ ID NO: 2 and 3 in the cells of a patient.
7. A method according to claim 6 wherein the reduced presence of SEQ ID NO: 2 and 3 is determined with an antibody with a specificity for SEQ ID NO: 2.
8. A transgenic animal comprising a hexanucleotide repeat expansion of more than 60 repeats of 5'-GGGGCC-3' in the orthologous sequence of SEQ ID NO: 1.
9. A transgenic animal which is a hemizygous knockout for the orthologous gene of C9ORF72.
10. A diagnostic kit having primers or antibodies for carrying out the diagnostic methods of any one of claims 1 to 7.
11. A nucleic acid sequence depicted in SEQ ID NO: 1 comprising a hexanucleotide repeat expansion of 5'-GGGGCC-3' consisting of between 30 and 200 repeats in exon 1A.

Figure 1

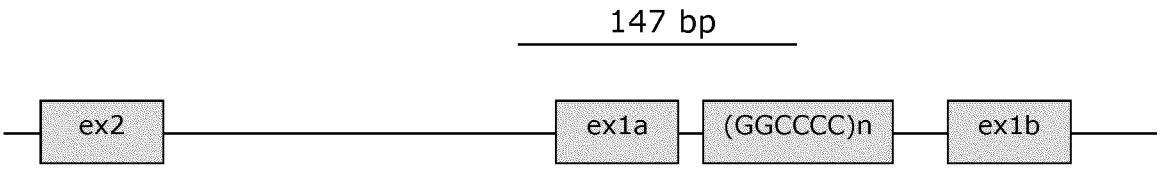


Figure 2

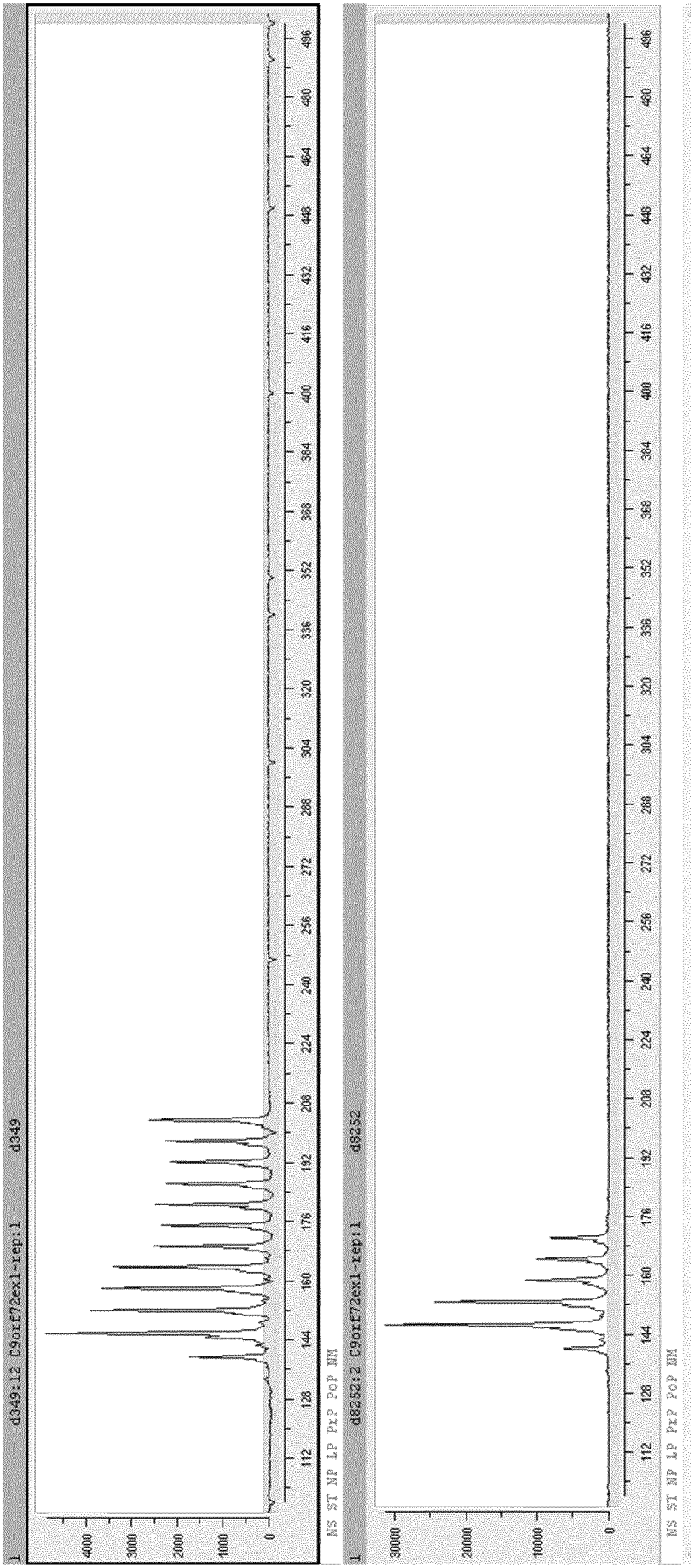

d349:12 C9orf72ex1-rep:1
d349
MS ST NP LP PrP PoP MM
d8252:2 C9orf72ex1-rep:1
d8252
MS ST NP LP PrP PoP MM

Figure 3

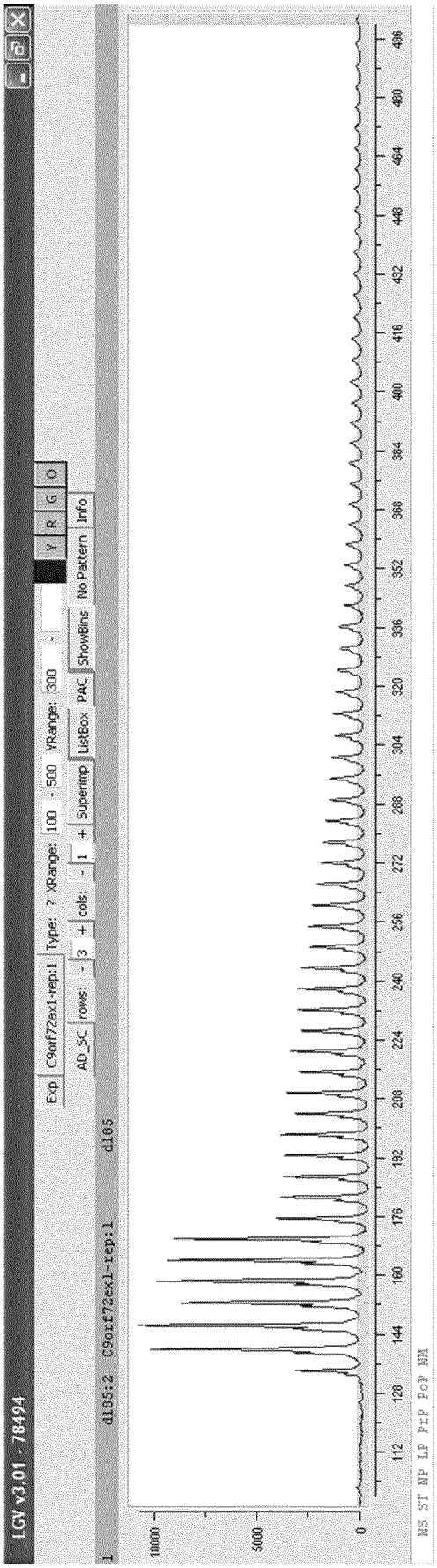

LGV v3.01 - 78494
Exp
C9orf72ex1-rep:1
Type:
? XRange:
100
500
YRange:
300
Y
R
G
O
AD_SC
rows:
3
cols:
1
Superimp
ListBox
PAC
ShowBins
No Pattern
Info
1
d185:2 C9orf72ex1-rep:1
d185
10000
5000
0
112
128
144
160
176
192
208
224
240
256
272
288
304
320
336
352
368
384
400
416
432
448
464
480
496
NS ST NP LP P+P PoP MM

Figure 4

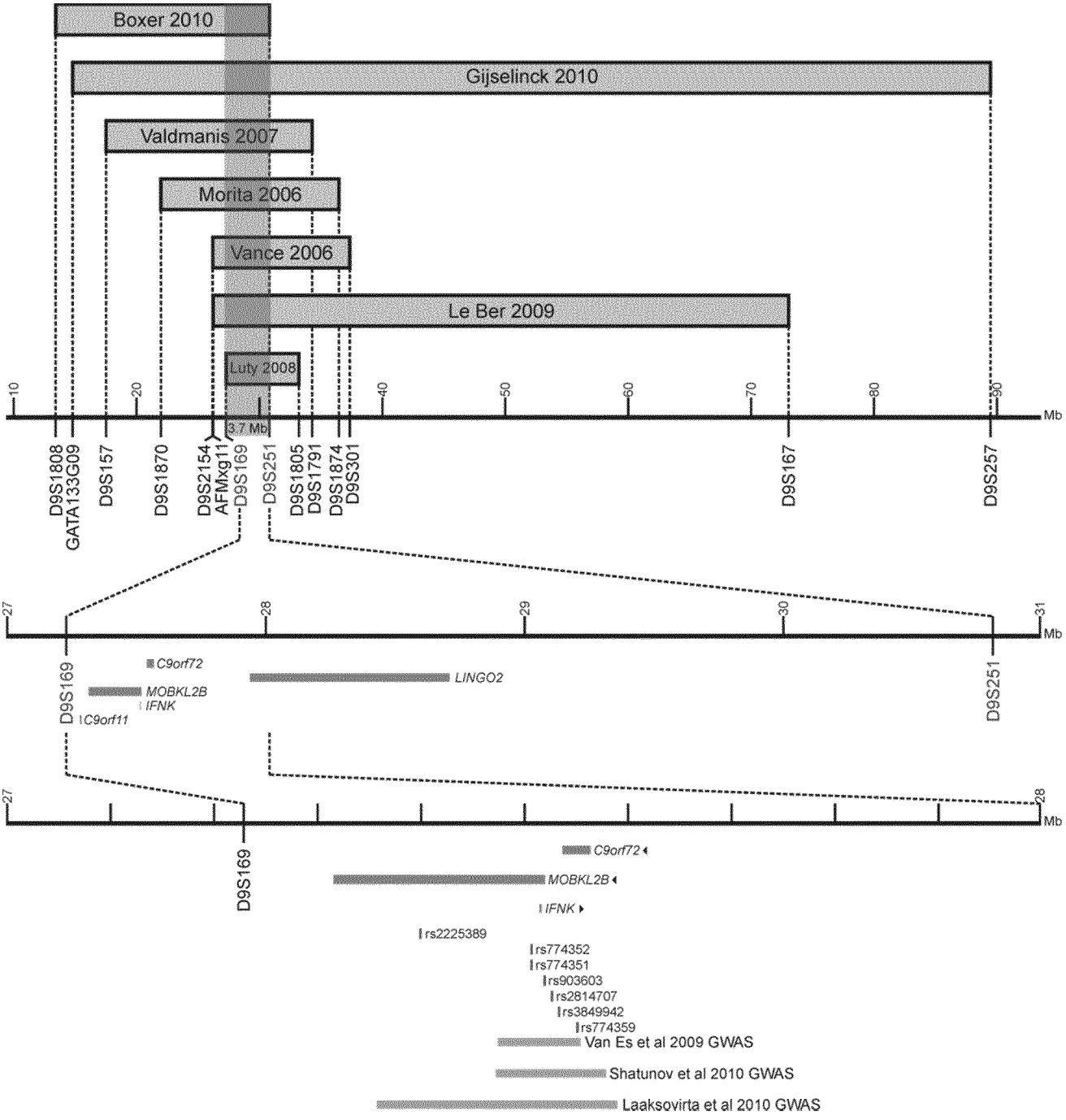

Boxer 2010
Gijselinck 2010
Valdmanis 2007
Morita 2006
Vance 2006
Le Ber 2009
Luty 2008
3.7 Mb
Mb
D9S1808
GATA133G09
D9S157
D9S1870
D9S2154
AFMxg11
D9S169
D9S251
D9S1805
D9S1791
D9S1874
D9S301
D9S167
D9S257
C9orf72
MOBKL2B
IFNK
C9orf11
LINGO2
rs2225389
rs774352
rs774351
rs903603
rs2814707
rs3849942
rs774359
Van Es et al 2009 GWAS
Shatunov et al 2010 GWAS
Laaksovirta et al 2010 GWAS

Figure 5

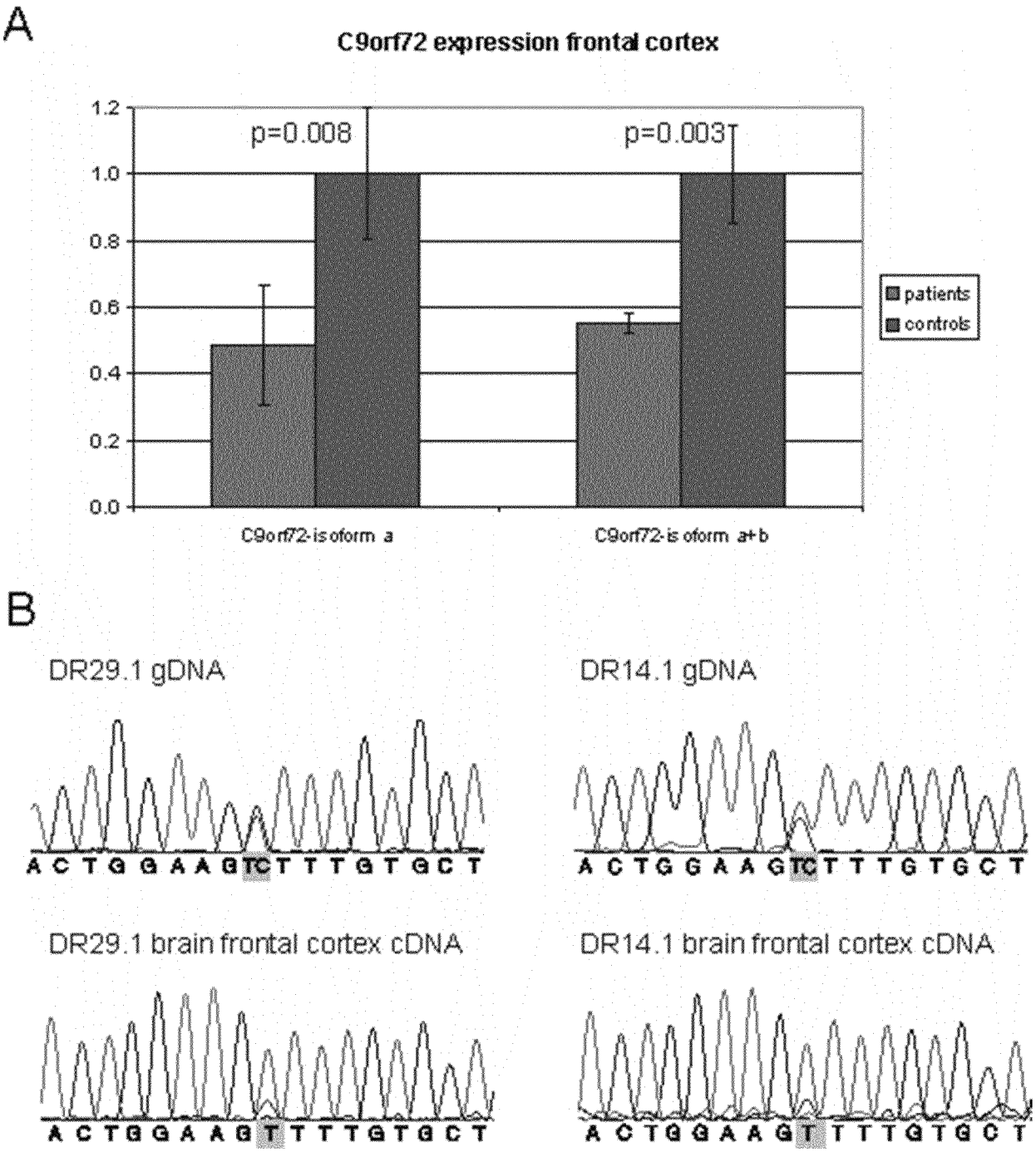

A
C9orf72 expression frontal cortex
p=0.008
p=0.003
1.2
1.0
0.8
0.6
0.4
0.2
0.0
patients
controls
C9orf72-isoform a
C9orf72-isoform a+b
B
DR29.1 gDNA
DR14.1 gDNA
DR29.1 brain frontal cortex cDNA
DR14.1 brain frontal cortex cDNA

Figure 6

```
1 ccagaagctg atgctgccat gcttcctata cagcctgcag aaccatgagt caattaaacc
       61 tcttttcttt gtaaattacc cagtctcaag tatttcttta tagcaatgca agaatggact
      121 aatacagaaa attgttactg agaagaaggg cattgctata aagatacctg aaaatgtaga
      181 agtgactttg gaaccggcta acaggcagaa gttgaaacat tttagagggc tcagaagaag
      241 acagaaagat gagagaaagt ttggaactcg ctaggaactt gttgagtggt tgtaaccaaa
      301 atactgatag tgatatagac agtgaagtcc aggctgagga ggtctcagat ggaaatgaga
      361 aatttattgg gaatgagtaa aggtcaggtt tgctatgctt tagcaaagag cttagctgca
      421 ttgttcctct gttctaggga tctgtgaaat cttagactta agaatgatga tttagggtat
      481 ctggcagaag aaatttctaa gcagcagagt gttcaagaag taacctagct gcttctaata
      541 gcctatgctc ataggcatga gcacagaaat gacctgaaat tggaacttac acttaaaagg
      601 gaagcagagc ataaaagttt gtaaattttg cagcctggcc atgtggtagt aaagaaaagc
      661 tcgttctcag gagaggaagt caagcaggct gcataaattt gcataactaa aaggaaggca
      721 agggctgata accaaaacaa tggggagaaa gactcatagg actaacaggc attttatttt
      781 attttatttt tattttatta ttattatact ttaagtttta gggtacatgt gcacaatgtg
      841 caggttagtt gcatatgtat acatgtgcca tgctggtgtg ctgcacccat taactcgtca
      901 tttagcatta ggtatatctc ctaatgctat ccctcccccc tcccccaccc cacaacagtc
      961 cccagagtgt gatgttcccc ttcctgtgtc catgtgttct cattgttcaa ttcccaccta
     1021 tgagtgagaa catgtggtgt ttggtttttt gaccttgcaa tagtttactg agaatgacga
     1081 tttccaattt catccatgtc cctacaaagg acatgaactc atcatttttt atggctgcat
     1141 agtattccat ggtgtatatg tgccacattt tcttaatcca gtctatcact gttggacatt
     1201 tgggttggtt ccaagtcttt gctattgtga atagtgccac aataaacata gtgtgcatgt
     1261 gtctttatag cagcaggatt tatagtcctt tgggtatata cccagtgatg ggatggctgg
     1321 gtcaaatggt atttctagtt ctagatccct gaggaatcgc cacactgact tccacaatgg
     1381 ttgaactagt ttacagtccc accaacagtg taaaagtgtt cctaataggc attttaggct
     1441 ttcatggtgg tccctctcat cacaggcccc gaggcctagg aggactgaat catttcctgg
     1501 gccaggccta gggcccctgc tccctcttac agccttggga ctctgctccc tgaatcccag
     1561 ctgctcaaag gggcccaggt actgttacag taggtagcta atcaggcatg agtggggtaa
     1621 gagagaagtc cccaccaccc accaggaatg tcaggcaacc atcagatgat ggtcaggcag
     1681 ttgtcatact gcctctctaa aatagtaatt ggttgcagcc agcaccaggg agaggcaact
     1741 tctcaataga tagaaacacc tgaaattggt aactgggcgc ttccaataag atctcaggaa
     1801 ctgagagagt gggcttaaca tgcacattaa gaggcaaaat ggtgaagtat gacctttggg
     1861 ggcattccac cggaaaaggg aagaaagcct caggtaagca tgtatacaac tccagtaaac
     1921 acactgcaca cgctcacctt ccaagtgcaa gcagggcacc atgcatgcgg caagctcacc
     1981 cttagggaag gaccaaggga aaggggcaca agatgtcaga agtaggccag tgtataagat
     2041 cctaggttca aggtcaaaca gggcacttga cctccaaggt gcccacttgg gcctcttcca
     2101 aatgtacttt cctttcattc ctgttctaaa gctttttaat aaacttttac tcctgctctg
     2161 aaacttgtcg cagtctcttt ttctgcctta tgcctcttgg tcaaattctt tcttctgagg
     2221 aggcaagaat tgaggttgct gcagacccac atggatttgc agctggtaac tcagataact
     2281 ttcaccagta agaatacagt tcaggctgct gcttcacagg gtgccaggca taagccttgg
     2341 tggcttccat aagctgtgaa gccggcgggc gcacataatg caagagttga ggcttaagaa
     2401 gctctgccta gattttagag gatgtatgaa aaagcctgga tgtccagaca gaagcctgtt
     2461 actggggtgg aatcctcatg gagaacatct actagggaag caaggagaag aaatgtgggg
     2521 ttgcagcccc cacagagagt cccctggggc actgcctagc agagctatga caagacagcc
     2581 accgtcctcc agaccccaga atggtagatc caccaacaac ttgcaccctg cagcctggaa
     2641 aagctgcaag cactcaatgc tagcccatga gagcagctgt gggagatgaa ccctggaaaa
     2701 ccacaggggt ggttctgccc aaggttttgg gagcccactc attgcatcag tgttccctgg
     2761 gtgtgagtca aaggagatta tttcagagct ttaacattta atgactgccc ggctggcttt
     2821 cagacttgca atggggccct atagcctctt tcttttggca gatttctccc tttcggaatg
     2881 gcagtatctg cccaatgcct ataccccccat tgtatctttg aagcaattac cttgtttttg
     2941 attttacagg ttcataggta gaagggacta gcttcgtctc aggtgagact tgggactttg
     3001 gacttttgaa tgaatgctgg atcgagttaa gactttgggg aactgttggt aaggcacgac
```

3061 agtattttgc aatatgagaa ggacattaga tttgggaggg gccagagttg gaataacatg
3121 gtttggatct ctgtccccac ccaaatctca tgttcaactg taatccccag tgttggaggt
3181 tgggcctggt gggaggtgag tggattatgg ggtggcttct aatggttttg tacagtcccc
3241 tcttggtact atatagtgag ttctgacaag atctagttgt ttaaacgtat gtagcacctc
3301 ccatttctct cttcccccag ttcctgccat gtgaagtctg gggtctccct atgccttcca
3361 tcatgatttt aagttcccta tggcctgccc agaagctgat ccagccatgc ttcttgtaca
3421 gcctgcagaa ctgtgagcca ttaaactttt ctttataaat tacccagttt cagttatttc
3481 tttatagcag tgtaagaatg gactaacaca attattaacg ctagtcctca tgttgtacat
3541 taaatctcta gatgtattag acgtaactgc aactttgtac cctaccctac aatttttctt
3601 ccccccaagc cccccaacca agggtctact ctgtttctat aaattcagtt gttttttaat
3661 tccacgtata agtgaagtac aactcagtgt agaaacttgg taaatgctag ctacttgtta
3721 taagctgtca gtcaaaataa aaatacagag atgaatctct aaattaagtg atttatttgg
3781 gaagaaagaa ttgcaattag ggcatacatg tagatcagat ggtcttcggt atatccacac
3841 aacaaagaaa agggggaggt tttgttaaaa aagagaaatg ttacatagtg ctctttgaga
3901 aaattcattg gcactattaa ggatctgagg agctggtgag tttcaactgg tgagtgatgg
3961 tggtagataa aattagagct gcagcaggtc attttagcaa ctattagata aaactggtct
4021 caggtcacaa cgggcagttg cagcagctgg acttggagag aattacactg tgggagcagt
4081 gtcatttgtc ctaagtgctt ttctaccccc taccccact attttagttg ggtataaaaa
4141 gaatgaccca atttgtatga tcaactttca caaagcatag aacagtagga aaagggtctg
4201 tttctgcaga aggtgtagac gttgagagcc attttgtgta tttattcctc cctttcttcc
4261 tcggtgaatg attaaaacgt tctgtgtgat ttttagtgat gaaaaagatt aaatgctact
4321 cactgtagta agtgccatct cacacttgca gatcaaaagg cacacagttt aaaaaacctt
4381 tgtttttta cacatctgag tggtgtaaat gctactcatc tgtagtaagt ggaatctata
4441 cacctgcaga ccaaaagacg caaggtttca aaaatctttg tgttttttac acatcaaaca
4501 gaatggtacg tttttcaaaa gttaaaaaaa aacaactcat ccacatattg caactagcaa
4561 aaatgacatt ccccagtgtg aaaatcatgc ttgagagaat tcttacatgt aaaggcaaaa
4621 ttgcgatgac tttgcagggg accgtgggat tcccgcccgc agtgccggag ctgtccccta
4681 ccagggtttg cagtggagtt ttgaatgcac ttaacagtgt cttacggtaa aaacaaaatt
4741 tcatccacca attatgtgtt gagcgcccac tgcctaccaa gcacaaacaa aaccattcaa
4801 aaccacgaaa tcgtcttcac tttctccaga tccagcagcc tcccctatta aggttcgcac
4861 acgctattgc gccaacgctc ctccagagcg ggtcttaaga taaaagaaca ggacaagttg
4921 ccccgcccca tttcgctagc ctcgtgagaa aacgtcatcg cacatagaaa acagacagac
4981 gtaacctacg gtgtcccgct aggaaagaga ggtgcgtcaa acagcgacaa gttccgccca
5041 cgtaaaagat gacgcttggt gtgtcagccg tccctgctgc ccggttgctt ctcttttggg
5101 ggcggggtct agcaagagca ggtgtgggtt taggaggtgt gtgtttttgt ttttcccacc
5161 ctctctcccc actacttgct ctcacagtac tcgctgaggg tgaacaagaa aagacctgat
5221 aaagattaac cagaagaaaa caaggaggga aacaaccgca gcctgtagca agctctggaa
5281 ctcaggagtc gcgcgcta**gg ggccggggcc ggggccgggg c**gtggtcggg gcgggcccgg
5341 gggcgggccc ggggcggggc tGCGGTTGCG GTGCCTGCGC CCGCGGCGGC GGAGGCGCAG
5401 GCGGTGGCGA GTGGgtgagt gaggaggcgg catcctggcg ggtggctgtt
tggggttcgg

Figure 7

Figure 8

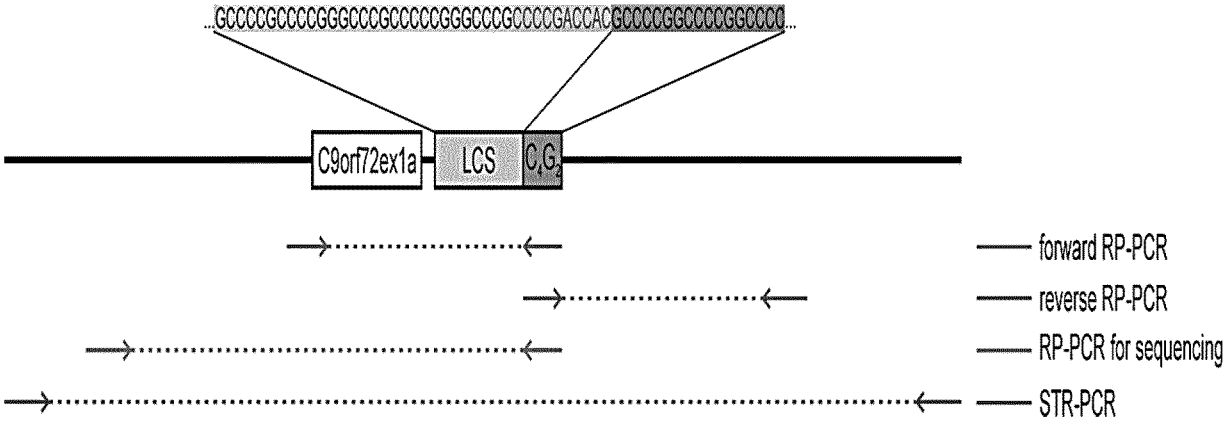


Figure 9

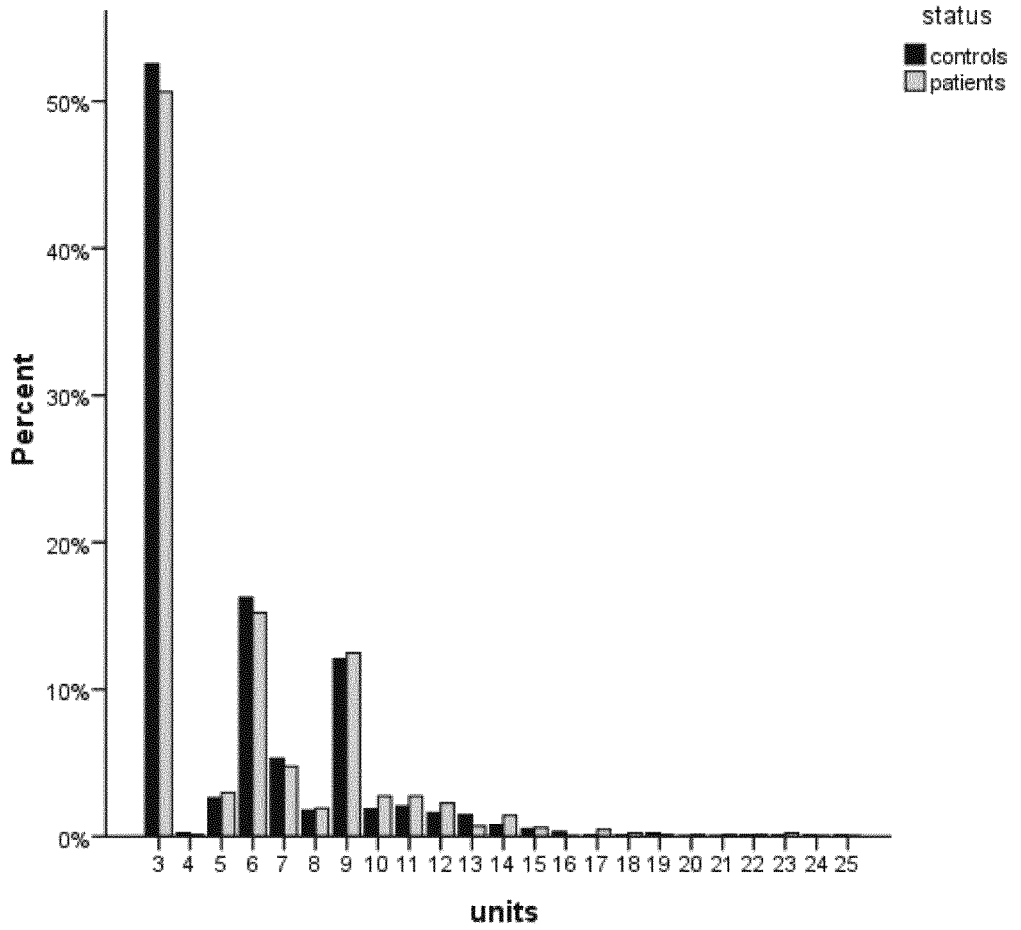
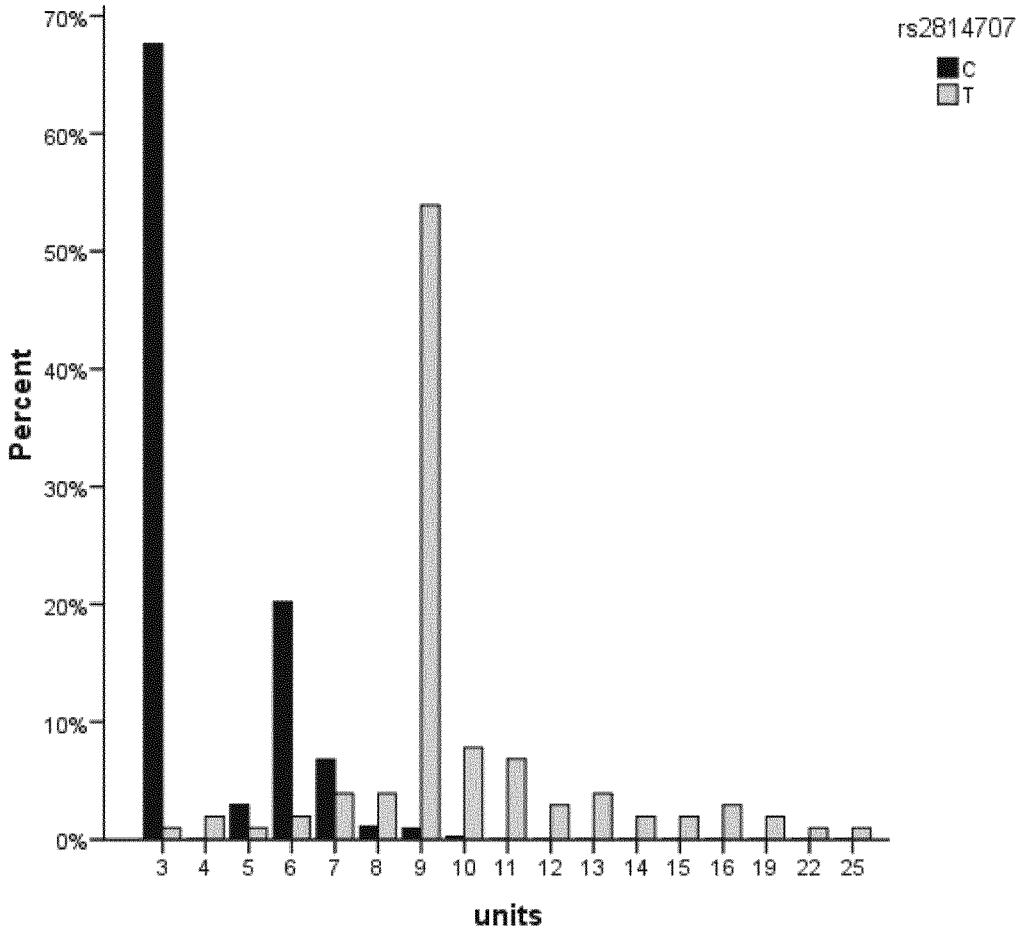


Figure 10


rs2814707
C
T
Percent
70%
60%
50%
40%
30%
20%
10%
0%
3 4 5 6 7 8 9 10 11 12 13 14 15 16 19 22 25
units

INTERNATIONAL SEARCH REPORT

International application No.
PCT/EP2012/068456

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application and necessary to the claimed invention, the international search was carried out on the basis of:

   a. (means)

   [ ] on paper

   [X] in electronic form

   b. (time)

   [X] in the international application as filed

   [ ] together with the international application in electronic form

   [ ] subsequently to this Authority for the purpose of search

2. [ ] In addition, in the case that more than one version or copy of a sequence listing and/or table relating thereto has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that in the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No
PCT/EP2012/068456

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12Q1/68 A01K67/027 G01N33/53
ADD.

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
EPO-Internal, BIOSIS, Sequence Search, EMBASE, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2008/095261 A1 (POWMRI LTD [AU]; SCHOFIELD PETER [AU]; KWOK JOHN [AU]; LUTY AGNES [AU]) 14 August 2008 (2008-08-14) page 75, line 19 - page 76, line 7 claims 1-10 ----- -/--	10

[X] Further documents are listed in the continuation of Box C.

[X] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
20 November 2012	30/11/2012

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Ulbrecht, Matthias

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	SARA ROLLINSON ET AL: "Frontotemporal lobar degeneration genome wide association study replication confirms a risk locus shared with amyotrophic lateral sclerosis", NEUROBIOLOGY OF AGING, TARRYTOWN, NY, US, vol. 32, no. 4, 9 December 2010 (2010-12-09), pages 758.e1-758.e7, XP028158073, ISSN: 0197-4580, DOI: 10.1016/J.NEUROBIOLAGING.2010.12.005 [retrieved on 2010-12-15] page 758.e1, left-hand column, paragraph 2	1-11
A	MICHAEL A VAN ES ET AL: "Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis", NATURE GENETICS, vol. 41, no. 10, 1 October 2009 (2009-10-01), pages 1083-1087, XP55044719, ISSN: 1061-4036, DOI: 10.1038/ng.442 figure 2	1-11
A	LAAKSOVIRTA H ET AL: "Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study", LANCET NEUROLOGY, LANCET PUBLISHING GROUP, LONDON, GB, vol. 9, no. 10, 1 October 2010 (2010-10-01), pages 978-985, XP027599044, ISSN: 1474-4422, DOI: 10.1016/S1474-4422(10)70184-8 [retrieved on 2010-10-01] figure 3	1-11
X,P	MARIELY DEJESUS-HERNANDEZ ET AL: "Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region ofCauses Chromosome 9p-Linked FTD and ALS", NEURON, CELL PRESS, US, vol. 72, no. 2, 21 September 2011 (2011-09-21), pages 245-256, XP028322560, ISSN: 0896-6273, DOI: 10.1016/J.NEURON.2011.09.011 [retrieved on 2011-09-17] the whole document	1-11

-/--

1

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X,P	ALANE RENTON ET AL: "A Hexanucleotide Repeat Expansion inIs the Cause of Chromosome 9p21-Linked ALS-FTD", NEURON, CELL PRESS, US, vol. 72, no. 2, 21 September 2011 (2011-09-21), pages 257-268, XP028322561, ISSN: 0896-6273, DOI: 10.1016/J.NEURON.2011.09.010 [retrieved on 2011-09-17] the whole document	1-11
X,P	ILSE GIJSELINCK ET AL: "A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study", THE LANCET NEUROLOGY, vol. 11, no. 1, 8 December 2011 (2011-12-08), pages 54-65, XP55044218, ISSN: 1474-4422, DOI: 10.1016/S1474-4422(11)70261-7 the whole document	1-11
T	Cheah B.C.: "Neurophysiological biomarkers in Amyotrophic Lateral Sclerosis", PhD thesis, August 2011 (2011-08), XP007921297, Retrieved from the Internet: URL:unsworks.unsw.edu.au [retrieved on 2012-11-19] page 32, last paragraph page 189, paragraph 3	1-11

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/EP2012/068456

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2008095261 A1	14-08-2008	AU 2008213742 A1	14-08-2008
		CA 2677339 A1	14-08-2008
		EP 2115167 A1	11-11-2009
		JP 2010517540 A	27-05-2010
		US 2010028356 A1	04-02-2010
		WO 2008095261 A1	14-08-2008