

Genome-wide association study of frontotemporal dementia identifies a C9ORF72 haplotype with a median of 12-G4C2 repeats that predisposes to pathological repeat expansions

Reus, Lianne M.; Jansen, Iris E.; van Rooij, Jeroen; van Schoor, Natasja M.; Tesi, Niccolò; Reinders, Marcel J.T.; Huisman, Martijn A.; van der Lugt, Aad; van der Lee, Sven J.; More Authors

DOI

[10.1038/s41398-021-01577-3](https://doi.org/10.1038/s41398-021-01577-3)

Publication date

2021

Document Version

Final published version

Published in

Translational Psychiatry

Citation (APA)

Reus, L. M., Jansen, I. E., van Rooij, J., van Schoor, N. M., Tesi, N., Reinders, M. J. T., Huisman, M. A., van der Lugt, A., van der Lee, S. J., & More Authors (2021). Genome-wide association study of frontotemporal dementia identifies a C9ORF72 haplotype with a median of 12-G4C2 repeats that predisposes to pathological repeat expansions. *Translational Psychiatry*, 11(1), Article 451. <https://doi.org/10.1038/s41398-021-01577-3>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

ARTICLE OPEN



Genome-wide association study of frontotemporal dementia identifies a *C9ORF72* haplotype with a median of 12-G4C2 repeats that predisposes to pathological repeat expansions

Lianne M. Reus¹✉, Iris E. Jansen^{1,2}, Merel O. Mol³, Fred van Ruissen⁴, Jeroen van Rooij³, Natasja M. van Schoor⁵, Niccolò Tesi^{1,6,7}, Marcel J. T. Reinders⁷, Martijn A. Huisman^{5,8}, Henne Holstege^{1,6}, Pieter Jelle Visser^{1,9,10}, Sterre C. M. de Boer¹, Marc Hulsman^{1,6}, Shahzad Ahmad¹¹, Najaf Amin¹¹, Andre G. Uitterlinden¹², Arfan Ikram¹¹, Cornelia M. van Duijn¹¹, Harro Seelaar³, Inez H. G. B. Ramakers⁹, Frans R. J. Verhey⁹, Aad van der Lugt^{1,3}, Jurgen A. H. R. Claassen¹⁴, Geert Jan Biessels¹⁵, Peter Paul De Deyn¹⁶, Philip Scheltens¹, Wiesje M. van der Flier^{1,5}, John C. van Swieten³, Yolande A. L. Pijnenburg¹ and Sven J. van der Lee^{1,6}

© The Author(s) 2021

Genetic factors play a major role in frontotemporal dementia (FTD). The majority of FTD cannot be genetically explained yet and it is likely that there are still FTD risk loci to be discovered. Common variants have been identified with genome-wide association studies (GWAS), but these studies have not systematically searched for rare variants. To identify rare and new common variant FTD risk loci and provide more insight into the heritability of *C9ORF72*-related FTD, we performed a GWAS consisting of 354 FTD patients (including and excluding $N = 28$ pathological repeat carriers) and 4209 control subjects. The Haplotype Reference Consortium was used as reference panel, allowing for the imputation of rare genetic variants. Two rare genetic variants nearby *C9ORF72* were strongly associated with FTD in the discovery (rs147211831: OR = 4.8, $P = 9.2 \times 10^{-9}$, rs117204439: OR = 4.9, $P = 6.0 \times 10^{-9}$) and replication analysis ($P < 1.1 \times 10^{-3}$). These variants also significantly associated with amyotrophic lateral sclerosis in a publicly available dataset. Using haplotype analyses in 1200 individuals, we showed that these variants tag a sub-haplotype of the founder haplotype of the repeat expansion that was previously found to be present in virtually all pathological *C9ORF72* G₄C₂ repeat lengths. This new risk haplotype was 10 times more likely to contain a *C9ORF72* pathological repeat length compared to founder haplotypes without one of the two risk variants (~22% versus ~2%; $P = 7.70 \times 10^{-58}$). In haplotypes without a pathologic expansion, the founder risk haplotype had a higher number of repeats (median = 12 repeats) compared to the founder haplotype without the risk variants (median = 8 repeats) ($P = 2.05 \times 10^{-260}$). In conclusion, the identified risk haplotype, which is carried by ~4% of all individuals, is a major risk factor for pathological repeat lengths of *C9ORF72* G₄C₂. These findings strongly indicate that longer *C9ORF72* repeats are unstable and more likely to convert to germline pathological *C9ORF72* repeat expansions.

Translational Psychiatry (2021)11:451 ; <https://doi.org/10.1038/s41398-021-01577-3>

INTRODUCTION

Frontotemporal dementia (FTD) is the second-most common cause of early onset dementia (3.5–15 per 100,000 in <65 years), leading to a spectrum of clinical syndromes associated with frontal and/or temporal neuronal loss [1, 2]. Clinically, FTD can be classified into the behavioral variant (bvFTD) and the language variants semantic dementia (SD) and progressive non-fluent aphasia (PNFA) [3]. FTD

is associated with motor neuron disease (FTD-MND) in 10% of all cases [4]. Currently, no treatment options are available for FTD. To identify potential treatment targets, an understanding of the underlying genetic etiology of FTD is highly needed.

Genetic factors play a major role in FTD; up to 40–50% of all FTD patients have a positive family history for dementia [5, 6]. Mutations that cause autosomal dominant FTD have been

¹Alzheimer Center Amsterdam, Department of Neurology, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, the Netherlands. ²Department of Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Amsterdam Neuroscience, Vrije University, Amsterdam, the Netherlands. ³Department of Neurology and Alzheimer Center Erasmus MC, Erasmus University Medical Center, Rotterdam, the Netherlands. ⁴Department of Clinical Genetics, Amsterdam UMC, Amsterdam, the Netherlands. ⁵Department of Epidemiology and Data Science, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam Public Health research institute, Amsterdam, the Netherlands. ⁶Section Genomics of Neurodegenerative Diseases and Aging, Department of Clinical Genetics, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, the Netherlands. ⁷Delft Bioinformatics Lab, Intelligent Systems Department, Delft University of Technology, Delft, the Netherlands. ⁸Department of Sociology, VU University, Amsterdam, the Netherlands. ⁹Department of Psychiatry & Neuropsychology, Alzheimer Center Limburg, School for Mental Health and Neuroscience, Maastricht University, Maastricht, the Netherlands. ¹⁰Department of Neurobiology, Care Sciences and Society, Division of Neurogeriatrics, Karolinska Institutet, Stockholm, Sweden. ¹¹Department of Epidemiology, Erasmus Medical Centre, Rotterdam, the Netherlands. ¹²Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. ¹³Department of Radiology & Nuclear Medicine, Erasmus University Medical Center, Rotterdam, the Netherlands. ¹⁴Radboudumc Alzheimer Center, Department of Geriatrics, Radboud University Medical Center, Nijmegen, the Netherlands. ¹⁵Department of Neurology, UMC Utrecht Brain Center, Utrecht, the Netherlands. ¹⁶Department of Neurology and Alzheimer Center Groningen, University Medical Center Groningen, Groningen, the Netherlands. ✉email: l.reus@amsterdamumc.nl

Received: 19 March 2021 Revised: 30 July 2021 Accepted: 23 August 2021

Published online: 02 September 2021

identified in microtubule associated protein tau (*MAPT*) [7], progranulin (*GRN*) [8], and the chromosome 9 open reading frame 72 (*C9ORF72*) G₄C₂ hexanucleotide repeat expansion [9, 10]. While familial mutations account for ~30% of FTD cases, the majority of FTD is multifactorial and polygenic in nature [11]. Previous genome-wide association studies (GWAS) on FTD have identified only a handful of common genetic risk variants for FTD with small effects on developing disease [12–15]. As the majority of sporadic FTD cannot be genetically explained yet, it is likely that there are still FTD risk loci to be discovered.

Rare genetic variants (minor allele frequency (MAF) ≤ 5%) often have stronger associations with disease than common genetic variants, but reliable imputation of rare genetic variants with widely used reference panels is challenging [16]. The Haplotype Reference Consortium (HRC) allows imputation of genetic variants with a MAF up to 0.001 [17, 18]. Performing a GWAS on FTD using the HRC panel as reference panel may aid in identifying rare risk variants for FTD, thereby improving insights into the genetic etiology of FTD.

To identify rare and common variant FTD risk loci and provide more insight into the pathogenesis and heritability of *C9ORF72*-related FTD, we performed a GWAS study in a cohort of Dutch FTD patients and control subjects, using the HRC panel as reference panel.

METHODS

Study sample genome-wide association study

We performed a genome-wide association analysis in which we compared genotype data of 354 FTD patients (8%, $N = 28$ with pathological *C9ORF72* repeat length) from three cohorts with 4209 control subjects from seven cohorts. Table S1 presents a brief description of the contributing cohorts. FTD was diagnosed according to diagnostic guidelines for FTD [3, 19]. Clinical subtypes of FTD (i.e., bvFTD, SD, PNFA, and FTD-MND) were available for 311 patients from the Amsterdam Dementia Cohort (ADC). Replication analyses were performed using data from the Erasmus Medical Center and an independent sample of the LASA study, including 281 FTD patients and 618 control subjects [20, 21].

All participating studies were approved by their respective Medical Ethics Committee (Table S1). Informed consent, either from the patient or from the legal representative, was obtained from all participants.

Genotyping and imputation

The discovery cohorts were genotyped on the Illumina Genome Screening Array (GSA, GSAsharedCUSTOM_20018389_A2) v1, human genome build 37. Quality control prior to imputation has been described in depth elsewhere [22]. Briefly, genetic variants were excluded from analyses when they deviated significantly from Hardy–Weinberg equilibrium ($P < 1 \times 10^{-6}$) in the total sample of founder individuals, or had a variant call rate of <98%. Individuals with sex mismatches or an individual call rate <98% were excluded from analyses. In total, 529,668 SNPs passed QC and were submitted to the Sanger imputation server for imputation to the Haplotype Reference Consortium (HRC) reference panel (<https://imputationserver.sph.umich.edu>). We pre-phased with SHAPE-IT2 [23]. This resulted in the imputation of 39,131,578 variants [18, 24]. To identify ethnic outliers, a principal component analysis of ancestry (PCA) was performed (based on 1000Genomes clustering), using EIGENSOFT [25]. Individuals of non-European ancestry were excluded from analysis to account for population structure. Relatedness was assessed through identity by descent (IBS), and family relations up to second degree (IBS ≥ 0.3) were excluded. To account for population structure, PCs were calculated on genetic data prior to imputation. In the replication cohort, cases were genotyped on the GSA array and the controls on the Axiom-NL array from Affymetrix (Avera Institute for Human Genetics, Sioux Falls, SD) [26, 27]. Quality control was performed in the same way as described for the discovery dataset for cases and controls independently and frequencies of the variants were compared after imputation.

Genotyping across the GGGGCC *C9ORF72* repeat

Allele-specific polymerase chain reaction (PCR) was performed using 0.2 mM dNTPs (Solis Biodyne), 0.05 Units HotFirePol DNA polymerase (5 U/μl Solis Biodyne), 1x Buffer B (Solis Biodyne), 2 mM MgCl₂ (Solis Biodyne),

7% DMSO (Sigma Aldrich), 2 μM 6FAM-fluorescent labeled forward primer ((6FAM)ACTCGCTGAGGGTGAACAAG) and 2 μM reverse primer (TCGAGCTCTGAGGAGAGCC), and 100 ng of genomic DNA. A standard PCR cycling program (35 cycles) was used where the annealing temperature was set at 55 °C with a 1-min extension time for each cycle. Fragment length analysis was performed on an ABI 3730xl/3500 genetic analyzer (Applied Biosystems Inc., Foster City, CA, USA), and data was analyzed using GeneScan software (version 4/5, ABI). Chromatograms were scored for the number of alleles and the number of repeats. Samples that have large lengths and samples with two alleles of the same length show only one band in the allele-specific PCR. For these samples, repeat-primed PCR was performed (supplementary methods).

Phenome-wide association studies

We conducted phenome-wide association studies (PheWAS) on the two replicated SNPs, rs147211831, rs117204439, using the 'pewas' function of the R-package 'ieugwas' [28, 29]. Using this function, we searched traits that associate with the list of SNPs with $P < 5 \times 10^{-8}$ in all GWAS harmonized summary statistics in the MRC IEU OpenGWAS data infrastructure [29].

Haplotyping of the identified risk variants for FTD and *C9ORF72* repeat lengths

To further study the relationship between the FTD risk alleles identified in the GWAS (rs147211831-A and rs117204439-C) and *C9ORF72* repeat lengths, we phased *C9ORF72* repeat lengths to haplotypes. We re-imputed chromosome 9 using EAGLE2 for pre-phasing [30]. This resulted in phased imputed genotypes in contrast to phasing with SHAPE-IT2 (which has slightly higher imputation accuracy) [23].

In accordance with previous studies, we found that the founder haplotype could be simplified to just one variant, rs3849942 (founder SNP = T) (Fig. S1) [31, 32]. Therefore, we were able to construct three SNP-haplotypes covering the *C9ORF72* gene. In short, we classified all haplotypes into the ancestral (non-founder) haplotype (rs3849942-C) and the founder haplotype (rs3849942-T). Subsequently, we split these haplotypes on having at least one risk allele (rs147211831-A and/or rs117204439-C) or no risk alleles. This resulted in four haplotype groups: ancestral non-risk, ancestral risk, founder non-risk, and founder risk haplotypes. We then mapped *C9ORF72* lengths to these haplotypes using a Bayesian classifier as described in Fig. S2 and the supplementary methods. The distribution of the *C9ORF72* lengths in the training dataset of these ancestral and founder haplotypes is presented in Fig. S3.

Statistical analysis

Association analysis on FTD patients versus controls was performed using PLINK version 2.0 [33, 34]. We used the Firth fallback option to fit logistic regression models, adjusting for population stratification (PC1-5). This model automatically uses Firth regression if the model does not converge (e.g., mainly for rare variants). SNPs with a low imputation quality ($R^2 < 0.3$) and a MAF < 0.5% were excluded. In total 8,813,788 variants were analyzed. To examine whether genome-wide significant loci ($P < 5 \times 10^{-8}$) were driven by pathological *C9ORF72* repeat carriers, analyses were repeated excluding patients who carried a pathological *C9ORF72* repeat length or did not have *C9ORF72* lengths available (discovery $N = 275$ FTD patients/239 controls; replication $N = 198$ FTD patients/618 controls). Additional analyses were performed including age and sex as covariates. We performed a meta-analysis on genome-wide significant loci ($P < 5 \times 10^{-8}$) using fixed-effects model with the rmeta package [35]. Last, we stratified analyses by clinical subgroups of FTD.

Additional statistical analyses were performed using R studio (version 4.0.3, Bunny-Wunnies freak out, R Development Core team 2010). To examine the association between haplotype and *C9ORF72* repeat length, we compared *C9ORF72* repeat expansion carriership using the proportion test and *C9ORF72* repeat lengths (excluding *C9ORF72* repeat expansion carriers) between haplotype groups (i.e., ancestral non-risk, ancestral risk, founder non-risk, and founder risk haplotypes), using Kruskal–Wallis test.

RESULTS

An overview of sample characteristics is shown in Table S2. The discovery FTD sample included less females, was younger compared to the controls and included ~8% ($N = 28/354$) pathological *C9ORF72*

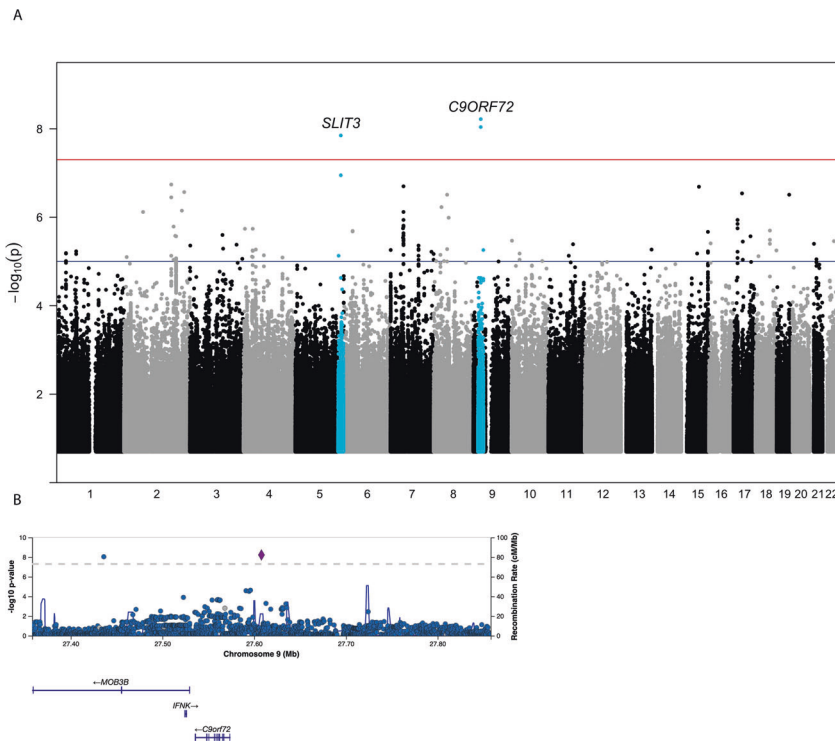


Fig. 1 Manhattan plot and regional plot of the discovery analysis on the genome-wide association with frontotemporal dementia. **A** Manhattan plot. The discovery analysis included $N = 354$ FTD patients and $N = 4209$ controls. The genome-wide significance threshold ($p < 5 \times 10^{-8}$) has been highlighted in red and the suggestive significance threshold ($P < 1 \times 10^{-5}$) is depicted in blue. For each genome-wide significant locus, loci are named by the closest located gene. **B** Regional plot for the *C9ORF72* locus on chromosome 9. The genetic variant depicted in purple represents the strongest associated variant. Abbreviation(s) FTD: frontotemporal dementia, *C9ORF72*: chromosome 9 open reading frame 72, *SLIT3*: Slit Guidance Ligand 3.

repeat length carriers (9%; $N = 25/281$ in replication). Of the FTD patients with clinical subtyping available, $N = 194$ were classified as bvFTD, $N = 74$ as SD, $N = 25$ as PFNA, $N = 18$ as FTD-MND and $N = 43$ were unclassified.

Association with FTD

Variants in two genomic loci were significantly associated with FTD ($p < 5 \times 10^{-8}$) (Fig. 1A and Table 1). No genomic loci were significant in the analysis excluding pathological *C9ORF72* repeat carriers. There was no genomic inflation in the GWAS ($\lambda = 0.009$) (Fig. S4).

A single intronic variant in a locus on chromosome 5 (rs76679949), located on the *SLIT3* (Slit Guidance Ligand 3) gene, was associated with a 3.7 times increased risk of FTD (MAF-cases = 4.4%; MAF-controls = 1.9%; $P = 1.4 \times 10^{-8}$) (Fig. S5). There were no additional variants in linkage with this variant to support the association and we were not able to replicate the association in the replication dataset ($P = 0.18$; OR = 1.58). We consider this locus a false positive finding and did not investigate it further.

The second locus on chromosome 9 contained two genetic variants that were significantly associated with FTD risk (rs117204439 and rs147211831). These two variants are located on both sides of the *C9ORF72* gene (Fig. 1B). In our data the two variants were in partial linkage ($R^2 = 0.041$; $D' = 0.52$) and 48% (43/90) of the carriers of rs117204439-C also carried the rs147211831-A allele. The most significant SNP, rs117204439, associated with a ~4.9 times increased risk of FTD (risk allele = C; MAF-cases = 3.9%; MAF-controls = 1.5%; $P = 6.0 \times 10^{-9}$). The second variant, rs147211831, associated with a ~4.8 times increased risk of FTD (risk allele = A; MAF-cases = 3.2%; MAF-controls = 0.9%; $P = 9.2 \times 10^{-9}$). When including the allele status of rs117204439 as covariate in the logistic regression model on FTD, the association of rs147211831 with FTD remained significant ($P = 4.8 \times 10^{-3}$;

OR = 2.7). This was also the case for the association of rs147211831 with FTD, corrected for rs117204439 ($P = 2.6 \times 10^{-3}$, OR = 2.8). These observed residual associations after adjusting for the other SNP suggests that signals were driven by their shared haplotype rather than by the specific SNP. Analyses stratified by clinical subtypes of FTD showed that associations were strongest in bvFTD (OR = 5.3–5.5) and FTD-MND (OR = 13.1–16.9) (Fig. S6). Results for rs147211831 and rs117204439 were similar when repeating the analyses correcting for age and sex (Table S3).

In our independent replication datasets, both variants near *C9ORF72* significantly associated with increased risk for FTD (rs117204439 MAF-cases = 3.6%; MAF-controls = 1.5%; $P = 2.0 \times 10^{-3}$; OR = 3.2, rs147211831 MAF-cases = 3.1%; MAF-controls = 1.1%; OR = 3.95, $P = 1.1 \times 10^{-3}$) (Table 1). Associations were strongest in bvFTD (OR = 3.9–4.1) and FTD-MND (OR = 6.2–6.3) (Table S4). Meta-analysis on the discovery and replication data showed similar results (rs117204439; $P = 5.6 \times 10^{-11}$; OR = 4.22, rs147211831; $P = 5.2 \times 10^{-11}$; OR = 4.62) (Table S5). After excluding pathological *C9ORF72* repeat length carriers and FTD patients without *C9ORF72* data ($N = 28/52$ carriers/unknown in discovery, $N = 25/58$ in replication) the association was no longer significant in both the discovery (rs117204439; $P = 0.05$; OR = 2.12, rs147211831 $P = 0.15$; OR = 1.87) and the replication cohorts (rs117204439 $P = 0.60$; OR = 1.32, rs147211831 $P = 0.09$; OR = 2.33). It is unlikely that results were driven by a single Dutch family, as the haplotype is relatively common and family relations up to second degree were excluded from analyses.

FTD risk alleles associate with amyotrophic lateral sclerosis in PheWAS

Both of the identified risk alleles for FTD showed an association with amyotrophic lateral sclerosis (ALS) ($N_{\text{GWAS}} = 12,663$ ALS

Table 1. FTD summary statistics discovery and replication analysis for SNPs exceeding genome-wide significance ($P < 5e-08$) in the discovery analysis.

SNP	Chr	Bp	Closest Gene	ref/alt	% alt	Discovery			Replication			Discovery excluding C9ORF72 carriers ^a			Replication excluding C9ORF72 carriers ^a						
						OR	Beta	SE	P	OR	Beta	SE	P	OR	Beta	SE	P	OR	Beta	SE	P
rs76679949	5	168220640	SLIT3	G/C	0.044/ 0.019	3.75	1.32	0.23	1.40×10^{-8}	1.58	0.46	0.34	0.18	NA	NA	NA	NA	NA	NA	NA	
rs147211831	9	27436084	C9ORF72	C/A	0.032/ 0.009	4.96	1.60	0.28	9.18×10^{-9}	3.95	1.37	0.42	1.05×10^{-3}	1.87	0.62	0.44	0.15	2.33	0.85	0.51	0.09
rs117204439	9	27607973	C9ORF72	T/C	0.039/ 0.015	4.85	1.58	0.27	5.96×10^{-9}	3.21	1.17	0.38	2.04×10^{-3}	2.12	0.75	0.39	0.05	1.32	0.27	0.52	0.60

^aFor these analyses we excluded FTD patients with pathological C9ORF72 repeat lengths and FTD patients without C9ORF72 data available ($N = 28/52$ carriers/unknown in discovery, $N = 25/58$ in replication). FTD: frontotemporal dementia, SNPs: single nucleotide polymorphisms, Chr: chromosome, Bp: base pair location, ref: reference allele, alt: alternative allele, OR: odds ratio, SE: standard error, NA: not applicable, C9ORF72 chromosome 9 open reading frame 72, SLIT3 Slit Guidance Ligand 3.

patients/53,439 controls) [36]. Risk allele rs147211831-A was associated with a ~1.9-fold increased risk of ALS ($P = 2.3 \times 10^{-20}$) and rs117204439-C with a 1.6-fold increased risk ($P = 3.1 \times 10^{-14}$) (Table S6). No other traits showed significant associations with the variants.

Screening of risk SNP carriers for C9ORF72 repeat expansions

The ADC also includes subjects diagnosed with other types of dementia and mild cognitive impairment ($N = 2543$). From these samples, we selected 58 non-related carriers of the FTD risk alleles rs117204439-C and rs147211831-A of European ancestry. We found that four of these 58 risk allele carriers had a pathological C9ORF72 repeat expansion. The diagnoses of the patients were diverse including vascular dementia, a psychiatric diagnosis, mild cognitive impairment, and a postponed diagnosis.

C9ORF72 risk alleles associate with intermediate repeat length in haplotype analysis

C9ORF72 repeat lengths were measured in a total of 1578 subjects from the ADC cohort, of whom 1327 had SNP-array data available. We excluded 104 individuals with a non-European ancestry and 23 individuals that were related ($IBS > 0.2$) to each other, leaving $N = 1200$ individuals for the haplotype analysis (Table S7). We attempted phasing C9ORF72 repeat lengths to haplotypes in all $N = 1200$ participants (see Methods section and Figs. S1 and S2).

We were able to reliably assign C9ORF72 lengths to the haplotype for 2352/2400 haplotypes (98%). These include 1743 (74.1%) ancestral non-risk haplotypes, 14 (0.6%) ancestral risk haplotypes, 535 (22.7%) founder non-risk haplotypes, and 60 (2.6%) founder risk haplotypes (Fig. 2).

Of all pathological C9ORF72 repeat lengths, 96.8% ($N = 30/31$) were mapped to the founder haplotype and one was mapped to the ancestral haplotype. Of the 31 repeat expansion haplotypes, 13 (41.9%) were the founder risk haplotype, 12 (38.7%) the founder non-risk haplotype, and 6 (19.4%, 5 founder and 1 ancestral) could not be assigned to a haplotype (Fig. 2).

As the founder risk haplotype was much less prevalent, the C9ORF72 repeat expansion was ~10-times more likely to be on a founder risk haplotype compared to the founder non-risk haplotype. In total, 21.7% of the founder risk haplotypes had pathological C9ORF72 repeat lengths ($N = 13/60$) compared to the founder non-risk haplotypes (2.2%, $N = 12/535$) ($P = 7.70 \times 10^{-58}$) (Table S8). Next, we compared the distribution of C9ORF72 repeats in the four haplotypes, excluding haplotypes with a pathological C9ORF72 repeat length (>30 repeat elements). Founder risk haplotypes had a median of 12 repeat elements, which was significantly higher than the founder non-risk haplotypes (median = 8, $P = 1.03 \times 10^{-38}$), ancestral risk haplotypes (median = 2, $P = 2.35 \times 10^{-8}$) and ancestral non-risk haplotypes (median = 2, $P = 3.33 \times 10^{-245}$) (Table S8 and Fig. 3).

DISCUSSION

Our findings show that the two variants rs117204439 and rs147211831 tag a C9ORF72 haplotype that is carried by ~4% of the population. This founder risk haplotype greatly increases the risk for a pathological C9ORF72 repeat length, which has been associated with FTD and the related motor neuron disorder ALS. Pathological lengths were ~10-times more likely to be present on this founder risk haplotype than on the founder haplotype without the risk variants. Haplotype analyses showed that the well-known founder haplotype with at least one risk allele had a median of 12 repeats compared to a median of 8 for the founder haplotype without risk alleles. The results of this study imply that an increased number of C9ORF72 repeat units increases the risk of conversion from a non-pathological repeat length to a pathological repeat length during parent-offspring transmissions.

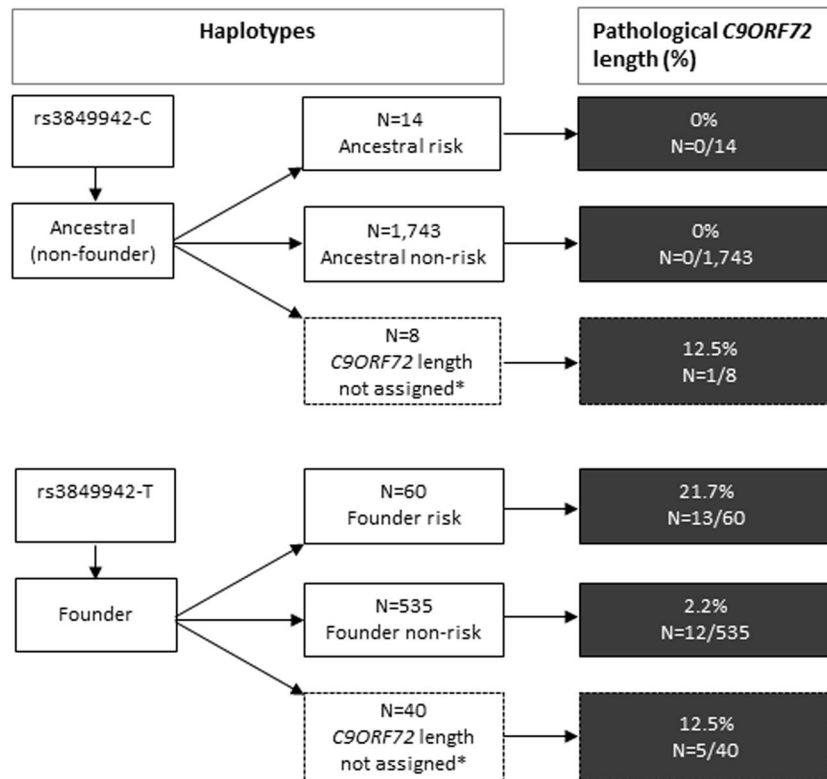


Fig. 2 Pathological *C9ORF72* repeat lengths in *C9ORF72* haplotypes, including ancestral non-risk haplotypes, ancestral risk haplotypes, and founder non-risk haplotypes and founder risk haplotypes. Haplotypes could be mapped to *C9ORF72* repeat lengths for $N = 1743$ ancestral (non-founder) non-risk haplotypes, $N = 14$ ancestral risk haplotypes, $N = 535$ founder non-risk haplotypes, and $N = 60$ founder risk haplotypes. The founder haplotype is defined by the presence of rs3849942-C (tags ancestral allele) or rs3849942-T (tags founder allele). Risk status is defined by the presence of at least one risk allele, including rs117204439-C or rs147211831-A. *The subset of haplotypes that could not be assigned to *C9ORF72* lengths had a short allele with a low probability (<0.8) for the ancestral allele (rs3849942-C) and *C9ORF72* lengths differed by more than 3 repeats. Abbreviation(s) *C9ORF72*: chromosome 9 open reading frame 72.

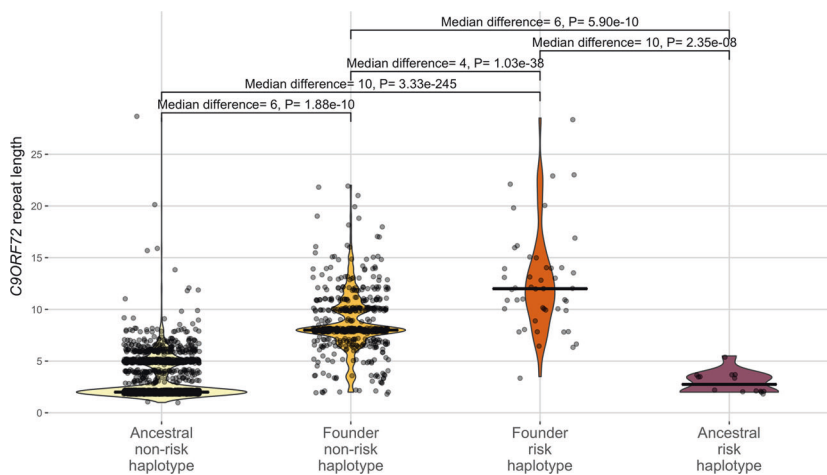


Fig. 3 *C9ORF72* repeat length (excluding haplotypes with >30 repeats) in *C9ORF72* haplotypes, including ancestral non-risk haplotypes, ancestral risk haplotypes, founder non-risk haplotypes and founder risk haplotypes. The founder haplotype is defined by the presence of rs3849942-C (tags ancestral allele) or rs3849942-T (tags founder allele). Risk status is defined by the presence of at least one risk allele, including rs117204439-C or rs147211831-A. Abbreviation(s) *C9ORF72*: chromosome 9 open reading frame 72.

Common variants at the *C9ORF72* locus at chromosome 9p21 have been identified previously as a genetic risk region for FTD and ALS [14, 15, 37–39]. Most, but not all [40], studies showed that association signals within the 9p21 region were driven by

carriers of the pathological repeat length of the G_4C_2 repeat in the *C9ORF72* gene [9, 10, 41]. These variants tag a so-called ‘Finnish founder haplotype’ of ~ 200 kb [31, 42]. This haplotype has a common founder and likely originated in Northern Europe

and spread from there to other regions [31, 32]. Haplotype analyses of carriers of pathological *C9ORF72* repeat lengths showed that nearly all carriers share (a part of) this haplotype. Therefore, the leading hypothesis is that pathological *C9ORF72* repeat lengths have been introduced on this haplotype into the population on multiple events due to a permissive allele [43–45]. This is a form of mutation in which repeat lengths expand within tissues [46] and during parent-offspring transmission [47], thereby predisposing to pathological repeat lengths [48, 49]. The founder haplotype had ~8 repeat units, compared to 2–4 units in the ancestral haplotype. Probably, this is the permissive allele that is associated with repeat instability. Still, it is debated whether the 8-unit repeat is more prone to repeat expansions as the inheritance of <30 repeats was found to be stable over generations [50]. On the other hand, it has been shown that pathological *C9ORF72* repeat lengths frequently vary over generations [43, 44, 51]. With our study, we add to this knowledge that a sub-haplotype of the founder haplotype with a median of 12 repeat units explains the majority of the pathological repeat lengths. This makes it plausible that the longer the G₄C₂ *C9ORF72* repeat is, the more likely it is that a de novo pathological expansion occurs during meiosis. Still, these expansion events must be extremely rare as the haplotype we identified is carried by only 4% of the Dutch population and by ~1–3% of all populations of European ancestry [25].

The molecular mechanisms underlying *C9ORF72* repeat instability involve DNA damage, since *C9ORF72* repeats have shown to interfere with DNA replication via abnormal nucleic acid structures (e.g., the formation of G-quadruplex structures, hairpins, and R-loops) [49, 52, 53]. *C9ORF72* repeats can form abnormal nucleic acid structures with as few as four repeats and repeat instability increases with longer *C9ORF72* repeats [49]. This may explain why the founder risk haplotype (with intermediate repeats) and the founder non-risk haplotype (with a lower range of repeats) both predispose to de novo pathological repeats, but differ in the proportion of pathological *C9ORF72* repeat lengths (~21.7% and ~2.2%, respectively) [31]. Further longitudinal research in multiple generations of carriers of the identified haplotype is required to confirm the higher conversion rate to longer *C9ORF72* repeat lengths in carriers of the founder risk haplotype compared to carriers of the founder non-risk haplotype. This type of study is also required to examine whether the risk haplotype serves as a pre-mutation or as predisposing allele for further stepwise mutation. Moreover, future studies should further investigate the possibility that the *C9ORF72* region contains additional genetic and epigenetic variants conferring risk to FTD.

While the identified SNPs tagging the founder risk haplotype cannot replace the *C9ORF72* repeat length assessments itself, a potential implementation of our findings is the use of the risk SNPs as pre-screener for the presence of a pathological *C9ORF72* repeat length in large population samples with array genotype data available. We were able to identify four previously undiscovered repeat expansion carriers that had another diagnosis than FTD or ALS. This underlines the diverse clinical presentation of subjects carrying the pathological *C9ORF72* repeat expansion.

Several limitations should be taken into account. While this study provides relevant insights into the genetic architecture of FTD in populations of European ancestry, further studies are required to examine the genetic architecture of FTD in other populations – particularly because *C9ORF72* repeat lengths differ across ethnic populations [54]. Finally, we could not map all phased *C9ORF72* haplotypes to *C9ORF72* repeat lengths. Ideally, we would have used long read sequencing data to confirm the phases of these haplotypes. Nonetheless, because findings for the unmapped haplotypes are in line with the rest of our results (Table S9) we do not think that this has influenced the results.

To conclude, we identified two risk SNPs for FTD that tag a 12-repeat sub-haplotype of the 8-repeat founder haplotype, which predisposes to *C9ORF72* pathological repeat lengths. We hypothesize that the longer repeat length makes the *C9ORF72* repeat more unstable and thus more susceptible to pathological expansion. To further understand the dynamic relationship between risk founder haplotypes (with increased repeat instability) and expansions of the *C9ORF72* repeat, it is essential that our efforts will be extended using functional follow-up studies and studies over generations.

CODE AVAILABILITY

Codes used to generate results are available upon request.

REFERENCES

- Seelaar H, Rohrer JD, Pijnenburg YA, Fox NC, van Swieten JC. Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review. *J Neurol Neurosurg Psychiatry*. 2011;82:476–86.
- Harvey RJ, Skelton-Robinson M, Rossor MN. The prevalence and causes of dementia in people under the age of 65 years. *J Neurol Neurosurg Psychiatry*. 2003;74:1206–9.
- Rascovsky K, Hodges JR, Knopman D, Mendez MF, Kramer JH, Neuhaus J, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*. 2011;134:2456–77.
- Neary D, Snowden JS, Mann DM, Northen B, Goulding PJ, Macdermott N. Frontal lobe dementia and motor neuron disease. *J Neurol Neurosurg Psychiatry*. 1990;53:23–32.
- Greaves CV, Rohrer JD. An update on genetic frontotemporal dementia. *J Neurol*. 2019;266:2075–86.
- Rohrer JD, Guerreiro R, Vandrovцова J, Uphill J, Reiman D, Beck J, et al. The heritability and genetics of frontotemporal lobar degeneration. *Neurology*. 2009;73:1451–8.
- Hutton M, Lendon CL, Rizzo P, Baker M, Froelich S, Houlden H, et al. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*. 1998;393:702–5.
- Baker M, Mackenzie IR, Pickering-Brown SM, Gass J, Rademakers R, Lindholm C, et al. Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17. *Nature*. 2006;442:916–9.
- DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, Rutherford NJ, et al. Expanded GGGGCC hexanucleotide repeat in noncoding region of *C9ORF72* causes chromosome 9p-linked FTD and ALS. *Neuron*. 2011;72:245–56.
- Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, Gibbs JR, et al. A hexanucleotide repeat expansion in *C9ORF72* is the cause of chromosome 9p21-linked ALS-FTD. *Neuron*. 2011;72:257–68.
- Seelaar H, Kamphorst W, Rosso SM, Azmani A, Masdjedi R, de Koning I, et al. Distinct genetic forms of frontotemporal dementia. *Neurology*. 2008;71:1220–6.
- Pottier C, Ren Y, Perkerson RB, Baker M, Jenkins GD, van Blitterswijk M, et al. Genome-wide analyses as part of the international FTLT-DTP whole-genome sequencing consortium reveals novel disease risk factors and increases support for immune dysfunction in FTLT. *Acta Neuropathol*. 2019;137:879–99.
- Pottier C, Zhou X, Perkerson RB, Baker M, Jenkins GD, Serie DJ, et al. Potential genetic modifiers of disease risk and age at onset in patients with frontotemporal lobar degeneration and GRN mutations: a genome-wide association study. *Lancet Neurol*. 2018;17:548–58.
- Diekstra FP, Van Deerlin VM, van Swieten JC, Al-Chalabi A, Ludolph AC, Weishaupt JH, et al. *C9orf72* and *UNC13A* are shared risk loci for amyotrophic lateral sclerosis and frontotemporal dementia: a genome-wide meta-analysis. *Ann Neurol*. 2014;76:120–33.
- Van Deerlin VM, Sleiman PM, Martinez-Lage M, Chen-Plotkin A, Wang LS, Graff-Radford NR, et al. Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions. *Nat Genet*. 2010;42:234–9.
- Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS discovery: biology, function, and translation. *Am J Hum Genet*. 2017;101:5–22.
- Iglesias AI, van der Lee SJ, Bonnemaier P, Höhn R, Nag A, Gharahkhani P, et al. Haplotype reference consortium panel: practical implications of imputations with large reference panels. *Hum Mutat*. 2017;38:1025–32.
- McCarthy S, Das S, Kretzschmar W, Delaneau O, Wood AR, Teumer A, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279–83.

19. Neary D, Snowden JS, Gustafson L, Passant U, Stuss D, Black S, et al. Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. *Neurology*. 1998;51:1546–54.
20. Hoogendijk EO, Deeg DJH, Poppelaars J, van der Horst M, Broese van Groenou MI, Comijs HC, et al. The Longitudinal Aging Study Amsterdam: cohort update 2016 and major findings. *Eur J Epidemiol*. 2016;31:927–45.
21. Hoogendijk EO, Deeg D, de Breijl S, Klokgieters SS, Kok A, Stringa N, et al. The Longitudinal Aging Study Amsterdam: cohort update 2019 and additional data collections. *Eur J Epidemiol*. 2020;35:61–74.
22. Tesi N, van der Lee SJ, Hulsman M, Jansen IE, Stringa N, van Schoor N, et al. Centenarian controls increase variant effect sizes by an average twofold in an extreme case-extreme control analysis of Alzheimer's disease. *Eur J Hum Genet*. 2019;27:244–53.
23. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013;10:5–6.
24. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284–7.
25. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature*. 2015;526:68–74.
26. Ehli EA, Abdellaoui A, Fedko IO, Grieser C, Nohzadeh-Malakshah S, Willemsen G, et al. A method to customize population-specific arrays for genome-wide association testing. *Eur J Hum Genet*. 2017;25:267–70.
27. Stringa N, Milaneschi Y, van Schoor NM, Suanet B, van der Lee S, Holstege H, et al. Genetic liability for depression, social factors and their interaction effect in depressive symptoms and depression over time in older adults. *Am J Geriatr Psychiatry*. 2020;28:844–55.
28. Elsworth B, Lyon M, Alexander T, Liu Y, Matthews P, Hallett J, et al. The MRC IEU OpenGWAS data infrastructure. *bioRxiv*. <https://doi.org/10.1101/2020.08.10.244293> (2020).
29. Hemani G, Zheng J, Elsworth B, Wade KH, Haberland V, Baird D, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7:7.
30. Loh PR, Danecek P, Palamara PF, Fuchsberger C, A Reshef Y, K Finucane H, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet*. 2016;48:1443–8.
31. Smith BN, Newhouse S, Shatunov A, Vance C, Topp S, Johnson L, et al. The C9ORF72 expansion mutation is a common cause of ALS+/-FTD in Europe and has a single founder. *Eur J Hum Genet*. 2013;21:102–8.
32. Mok K, Traynor BJ, Schymick J, Tienari PJ, Laaksovirta H, Peuralinna T, et al. Chromosome 9 ALS and FTD locus is probably derived from a single founder. *Neurobiol Aging*. 2012;33:209 e203–208.
33. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007;81:559–75.
34. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
35. Lumley T. *Lumley MT Package 'rmeta'* (2018).
36. Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide analyses identify KIF5A as a novel ALS gene. *Neuron*. 2018;97:1268–83.
37. Laaksovirta H, Peuralinna T, Schymick JC, Scholz SW, Lai SL, Myllykangas L, et al. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol*. 2010;9:978–85.
38. Shatunov A, Mok K, Newhouse S, Weale ME, Smith B, Vance C, et al. Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study. *Lancet Neurol*. 2010;9:986–94.
39. van Es MA, Veldink JH, Saris CG, Blauw HM, van Vught PW, Birve A, et al. Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis. *Nat Genet*. 2009;41:1083–7.
40. Jones AR, Woolacott I, Shatunov A, Cooper-Knock J, Buchman V, Sproviero W, et al. Residual association at C9orf72 suggests an alternative amyotrophic lateral sclerosis-causing hexanucleotide repeat. *Neurobiol Aging*. 2013;34:2234 e2231–2237.
41. Gijssels I, Van Langenhove T, van der Zee J, Sleegers K, Philtjens S, Kleinberger G, et al. A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study. *Lancet Neurol*. 2012;11:54–65.
42. Majounie E, Renton AE, Mok K, Dopper EG, Waite A, Rollinson S, et al. Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol*. 2012;11:323–30.
43. Beck J, Poulter M, Hensman D, Rohrer JD, Mahoney CJ, Adamson G, et al. Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population. *Am J Hum Genet*. 2013;92:345–53.
44. Gijssels I, Van Mossevelde S, van der Zee J, Sieben A, Engelborghs S, De Bleecker J, et al. The C9orf72 repeat size correlates with onset age of disease, DNA methylation and transcriptional downregulation of the promoter. *Mol Psychiatry*. 2016;21:1112–24.
45. Shamim U, Ambawat S, Singh J, Thomas A, Pradeep-Chandra-Reddy C, Suroliya V, et al. C9orf72 hexanucleotide repeat expansion in Indian patients with ALS: a common founder and its geographical predilection. *Neurobiol Aging*. 2020;88:156.e151–156.e159.
46. McGoldrick P, Zhang M, van Blitterswijk M, Sato C, Moreno D, Xiao S, et al. Unaffected mosaic C9orf72 case: RNA foci, dipeptide proteins, but upregulated C9orf72 expression. *Neurology*. 2018;90:e323–e331.
47. Xi Z, van Blitterswijk M, Zhang M, McGoldrick P, McLean JR, Yunusova Y, et al. Jump from pre-mutation to pathologic expansion in C9orf72. *Am J Hum Genet*. 2015;96:962–70.
48. Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: mechanisms of dynamic mutations. *Nat Rev Genet*. 2005;6:729–42.
49. Thys RG, Wang YH. DNA replication dynamics of the GGGGCC repeat of the C9orf72 Gene. *J Biol Chem*. 2015;290:28953–62.
50. van der Zee J, Gijssels I, Dillen L, Van Langenhove T, Theuns J, Engelborghs S, et al. A Pan-European study of the C9orf72 repeat associated with FTL: geographic prevalence, genomic instability, and intermediate repeats. *Hum Mutat*. 2013;34:363–73.
51. Kaivola K, Kiviharju A, Jansson L, Rantalainen V, Eriksson JG, Strandberg TE, et al. C9orf72 hexanucleotide repeat length in older population: normal variation and effects on cognition. *Neurobiol Aging*. 2019;84:242.e247–242.e212.
52. Konopka A, Atkin JD. The emerging role of DNA damage in the pathogenesis of the C9orf72 repeat expansion in amyotrophic lateral sclerosis. *Int J Mol Sci*. 2018;19:3137.
53. Walker C, Herranz-Martin S, Karyka E, Liao C, Lewis K, Elsayed W, et al. C9orf72 expansion disrupts ATM-mediated chromosomal break repair. *Nat Neurosci*. 2017;20:1225–35.
54. Ng ASL, Tan EK. Intermediate C9orf72 alleles in neurological disorders: does size really matter? *J Med Genet*. 2017;54:591–7.

ACKNOWLEDGEMENTS

Research of the Alzheimer Center Amsterdam is part of the neurodegeneration research program of Amsterdam Neuroscience. The Alzheimer Center Amsterdam is supported by Stichting Alzheimer Nederland and Stichting VUmc funds. This research is supported by ABOARD, which is a public-private partnership receiving funding from ZonMW (#73305095007) and Health-Holland, Topsector Life Sciences & Health (PPP-allowance; #LSHM20106). More than 30 partners participate in ABOARD. ABOARD also receives funding from Edwin Bouw Fonds and Gieskes-Strijbisfonds. The clinical database structure was developed with funding from Stichting Dioraphte. Genotyping of the Dutch case-control samples was performed in the context of EADB (European Alzheimer DNA biobank) funded by the JPco-fuND FP-829-029 (ZonMW projectnumber 733051061). The Longitudinal Aging Study Amsterdam is supported by a grant from the Netherlands Ministry of Health Welfare and Sports, Directorate of Long-Term Care. The work described in this study was carried out in the context of the Parelsoer Institute (PSI). PSI was part of and funded by the Dutch Federation of University Medical Centers and has received initial funding from the Dutch Government (from 2007 to 2011). Since 2020, this work was carried out in the context of Parelsoer clinical biobanks at Health-RI (<https://www.health-ri.nl/initiatives/parelsoer>). The 100-plus study work was supported by Stichting Alzheimer Nederland (WE09.2014-03), Stichting Dioraphte, horstingstuit foundation, Memorabel (ZonMW projectnumber 733050814), and Stichting VUmc Fonds. Genotyping of the 100-Plus Study was performed in the context of EADB (European Alzheimer DNA biobank) funded by the JPco-fuND FP-829-029 (ZonMW projectnumber 733051061). The European Medical Information Framework for AD (EMIF-AD) PreclinAD study and EMIF 90+ study have been funded by the Innovative Medicines Initiative Joint Undertaking under EMIF grant agreement #115372. We thank and acknowledge the International FTD-Genomics Consortium (IFGC). We thank Eleonora Maria Vromen from Amsterdam UMC, Alzheimer Center for proofreading our manuscript.

AUTHOR CONTRIBUTIONS

Writing: LMR, SJvdL, and YALP. Analysis: LMR, SJvdL, IEJ, and NT. Computation: MJTR. Reviewing and editing: LMR, SJvdL, WMvdf, YALP, SCMDb, JAHRC, GJB, and PPDd. Revision of the manuscript: all. Contribution to sample collection discovery data: SJvdL, IEJ, FvR, NMvS, MJTR, MAH, HH, PJV, WMvdf, YALP, NT, MH, PS, SA, NA, AGU,

MAI, CvD, JcVs, JvR, MOM, HS, IR, FV, AvdL, JAHRC, GJB, and PpD. Contribution to sample collection replication data: JcVs, JvR, MOM, and HS. Secured funding: WvdF and YALP. Supervision study: SvdL and YALP.

CONFLICT OF INTEREST

YALP received a personal fellowship from the Dutch brain foundation. All other authors report no disclosures or conflicts of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41398-021-01577-3>.

Correspondence and requests for materials should be addressed to LianneM. Reus

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021