UNIVERSITÀ DEL PIEMONTE ORIENTALE

School of Medicine

Department of Health Sciences

Master's degree in Medical Biotechnology

Laboratory of Human Genetics

Investigating *RFC1* Repeat Expansions in an Italian Cohort of Motor Neuron Disease Patients

Supervisor: Professor Lucia Corrado

Co-supervisor: Beatrice Piola

Candidate: Shivansh Vadhera

Matricular Number: 20054845

Academic Year 2025-2026

Table of Contents

SUMMARY

Motor Neuron Diseases (MNDs), such as Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD), are progressive neurodegenerative disorders with a complex genetic basis. While several genes are known to contribute to these conditions, many cases remain genetically unexplained. Recently, biallelic AAGGG repeat expansions in the *RFC1* gene were identified as the cause of CANVAS syndrome. The rationale for this study was driven by emerging evidence from a German patient cohort, which reported a ~2% prevalence of these pathogenic expansions among ALS patients, suggesting that *RFC1* might be a novel risk factor for motor neuron degeneration. This study aimed to investigate whether this association holds true in an Italian population to determine if *RFC1* screening should be integrated into the standard diagnostic pipeline for MNDs.

The study involved a comprehensive molecular screening of 250 unrelated individuals from the Piedmont region of Italy (237 with ALS and 7 with FTD). A central challenge in the study design was the "single band problem"—a technical pitfall where a single peak in fragment analysis could either represent a true homozygote or the failure to amplify a much larger second allele (allelic dropout). To address this, the planning included a modified PCR protocol with extended elongation times to ensure the detection of massive expansions. The methodological workflow utilized genomic DNA extraction, flanking PCR, high-resolution fragment analysis, Repeat-Primed PCR (RP-PCR), and Sanger sequencing to accurately distinguish between the benign AAAAG motif and the pathogenic AAGGG motif.

The molecular analysis revealed a 0% prevalence of biallelic pathogenic (AAGGG)n expansions within the cohort. While no patients carried the full pathogenic genotype, the study identified several individuals with very large heterozygous expansions, some exceeding 3,000–4,000 base pairs. These cases initially appeared as single bands in standard testing but were successfully characterized as the non-pathogenic AAAAG motif using the modified protocol. The most frequent allele found was a 15-repeat unit (27.4%). These results stand in clear contrast to the 2% prevalence reported in the German cohort, highlighting significant geographical variability in the genetic landscape of MND and proving that biallelic *RFC1* expansions are not a major driver of classical ALS in this Italian population.

In conclusion, this study demonstrates that biallelic *RFC1* AAGGG expansions are absent in this Italian MND cohort, suggesting that the gene's pathogenic role may be restricted to specific ethnic or geographical groups. However, a major pivot of the study is the identification of large heterozygous AAAAG expansions. The study concludes that these massive benign expansions should be further investigated as potential genetic modifiers that may influence disease susceptibility or clinical progression. While routine *RFC1* testing may not be necessary for all ALS cases, the study emphasizes the need for specialized protocols to avoid false negatives and advocates for future long-read sequencing to fully decode the impact of these large structural variations on neurodegeneration.

1.1. INTRODUCTION

1.1.2 Motor neuron diseases (MNDs): ALS and FTD

Motor neuron diseases (MNDs) encompass a diverse group of progressive neurological disorders characterized by the degeneration of upper and lower motor neurons. These neurons are responsible for transmitting signals from the brain and spinal cord to skeletal muscles, facilitating voluntary movement. The hallmark of MNDs is a gradual loss of voluntary muscle control, manifesting as weakness, spasticity, fasciculation, and eventual muscle atrophy. The spectrum of MNDs includes Amyotrophic Lateral Sclerosis (ALS), Primary Lateral Sclerosis (PLS), Progressive Muscular Atrophy (PMA), and other rare subtypes. Each subtype differs in the predominance of upper versus lower motor neuron involvement, rate of progression, and associated clinical features.

ALS is the most common and aggressive form, typically presenting with both upper and lower motor neuron degeneration, resulting in progressive paralysis (Bradley, 2009; Feldman et al., 2022; Grad et al., 2017).

The incidence of ALS is 1–2.6 cases per 100,000 persons per year, with a prevalence of approximately 6 cases per 100,000 and a lifetime risk of 1 in 350–400 individuals (La Spada & Taylor, 2010), with a male-to-female ratio of 1.35, which varies with the age of onset (Fontana et al., 2021; Logroscino et al., 2018).

ALS typically manifests in late adulthood; however, "young-onset" ALS cases (diagnosed before 45 years old) and juvenile cases (diagnosed before 25 years old) represent 10% and 1% of all cases, respectively. A recent epidemiological analysis estimated the mean age of symptom onset for typical ALS (adult-onset) at 58–63 years for sporadic ALS (sALS) and 40–60 years for familial ALS (fALS) (Masrori & Van Damme, 2020; Spencer et al., 2023).

The disease usually progresses rapidly, with many patients succumbing to respiratory failure within three to five years of symptom onset (Chia et al., 2018; Talbott et al., 2016) , although some patients demonstrate a slower disease course (Chio et al., 2009; Hardiman et al., 2017). The heterogeneity in clinical presentation and progression makes early and accurate diagnosis

challenging, particularly in cases with atypical onset or overlapping symptoms (Davies et al., 2022; Goutman et al., 2022).

85-90% of ALS cases are sporadic, with no affected family members (Masrori & Van Damme, 2020), whereas approximately 5–10% are familial (Brown & Al-Chalabi, 2017; Hardiman et al., 2017), emphasizing a strong genetic contribution.

Up to date, at least 40 genes for which mutations are implicated in the pathogenesis of ALS are known, however, four major genes: *C9orf72* (Chromosome 9 Open Reading Frame 72)*, SOD1* (Superoxide Dismutase 1)*, TARDBP* (TAR DNA-Binding Protein 43) *and FUS* (Fused in Sarcoma) account for ~48% of familial and ~5% of sporadic ALS within populations of European origin (Mejzini et al., 2019a; Zou et al., 2017).

The mutations identified are mostly nucleotide substitutions, although the (GGGGCC) hexanucleotide repeat expansion in the non-coding region of the *C9ORF72 gene* is the major contributor to ALS pathogenesis, and ALS-FTD overlap, accounting for about 40% of ALS cases with a family history of ALS and 8% of ALS without a family history (Majounie et al., 2012b). These expansions illustrate the pathogenic role of non-coding tandem repeats, which can lead to disease via mechanisms including RNA toxicity, repeat-associated non-AUG (RAN) translation, and accumulation of toxic dipeptide aggregates (Balendra & Isaacs, 2018).

The pathological effects of *C9orf72* expansions arise through several mechanisms. First, the expanded RNA transcripts form nuclear foci that sequester RNA-binding proteins, impairing normal RNA processing. Second, RAN translation produces dipeptide repeat proteins that aggregate within neurons, contributing to cellular toxicity. Third, the expansions may disrupt normal *C9orf72* gene expression, affecting cellular functions related to autophagy and vesicular trafficking.

Clinically, patients with *C9orf72* expansions may present with ALS, FTD, or a combination of both**,** with considerable phenotypic variability (Hardiman et al., 2017). This heterogeneity highlights the complexity of repeat expansion disorders and emphasizes the need for careful genetic and clinical evaluation. Moreover, the recognition of *C9orf72* expansions has encouraged researchers to examine other tandem repeat expansions as potential contributors to neurodegenerative diseases, especially in patients with atypical presentations (Smeyers et al., 2021).

Considering the three other major genes, *SOD1* encodes a 153 amino acid metalloenzyme, one of three superoxide dismutase enzymes found in humans, and provide an important antioxidant defense mechanism against free radicals (McCord & Fridovich, 1969). Mutations in SOD1 (mostly dominant missense mutations) lead to abnormal conformation and activity of the enzyme, generating toxic aggregates that damage cells (Mejzini et al., 2019b).

FUS (fused in sarcoma) is a gene encoding for an RNA-binding protein (Lagier-Tourenne & Cleveland, 2009), with mutations (mostly heterozygous missense mutations) commonly associated to a very aggressive form of early onset and juvenile ALS (other mutations cause adult-onset ALS, without cognitive or behavioral problems). It plays a role in several aspects of gene expression including transcription, pre-mRNA splicing, RNA transport, and translation regulation (Ratti & Buratti, 2016), as well as in DNA repair mechanisms (Mastrocola et al., 2013).

Finally, TDP-43 is a DNA/RNA binding protein, composed of 414 amino acids and encoded by the *TARDBP* gene. It is involved in regulating gene expression and several RNA processing steps, including pre-mRNA splicing, mRNA stability, mRNA transport, translation, and the regulation of non-coding RNAs (Tollervey et al., 2011). Dominant mutations in the *TARDBP* gene were identified as a primary cause of ALS (Gitcho et al., 2008). Frontotemporal involvement is uncommon in this type of ALS (Nijs & Van Damme, 2024).

In parallel, Frontotemporal Dementia (FTD) or frontotemporal lobar degeneration (FTLD) is a neurodegenerative condition that primarily affects the frontal and temporal lobes of the brain, leading to significant alterations in cognition, personality, and behavior (Bang et al., 2015a). FTD encompasses several clinical subtypes, including the behavioral variant, which is marked by disinhibition, apathy, or compulsive behaviors, and language variants, which primarily affect speech and comprehension(Rascovsky et al., 2011).

The estimated prevalence is about 15–22 cases per 100,000 individuals, with an incidence of 2.7–4.1/100,000. FTD accounts for 5–15% of all dementia, being the second most common dementia in the presenile age group, composed of people under the age of 65, second only to Alzheimer's disease (AD) (Bird et al., 2003). It occurs with equal frequency in both males and females (Hogan et al., 2016).

The average age of onset is usually between ages 45 and 65, (about 25% are considered late-life onset cases (Onyike & Diehl-Schmid, 2013) with 10% of FTD occurring in patients less than 45, but there have been documented cases younger than age 30 years (Graff-Radford & Woodruff, 2007; Olney et al., 2017).. The disease progresses gradually, with behavioral or language deficits appearing first, followed by cognitive decline. Mortality typically occurs within 6–8 years (Neary et al., 2005).

Several genes and chromosomal loci have been associated with FTD and especially mutations in chromosome *9* open reading frame *72 (C9orf72)*, microtubule-associated protein tau *(MAPT),* and progranulin *(PGRN)* genes account for the majority of genetically determined FTD (about 30%) and are found in 60% of familial FTD cases (Fenoglio et al., 2018; Olszewska et al., 2016). Recently, TANK-binding kinase 1 (*TBK1*) has been identified as probably the fourth most common genetic cause of FTD (Wagner et al., 2021).

Similar to ALS, *C9orf72* is the most common worldwide cause of genetic FTD, accounting for 25% of all cases (Bang et al., 2015b), and is followed by *GRN (*20%–25% of familial cases and about 10% of all cases (Wang et al., 2021)) and then *MAPT (*10–20% of familial cases and 0–3% of sporadic FTD (Benussi et al., 2015).

Notably, ALS and FTD frequently overlap both clinically and genetically, forming an ALS-FTD continuum, with up to 15% of FTD patients with motor neuron degeneration and up to 30% of ALS patients with frontotemporal dysfunction showing both conditions (Lomen-Hoerth, 2011).

This overlap suggests shared pathogenic pathways and indicates that neurodegeneration can simultaneously target motor, cognitive, and behavioral networks (Antonioni et al., 2023). The coexistence of motor and cognitive dysfunction complicates clinical management and underscores the need for comprehensive diagnostic approaches that consider both motor and cognitive domains (Strong et al., 2017). Understanding the biological mechanisms linking ALS and FTD is therefore crucial for the development of effective therapeutic strategies and for improving patient outcomes (Balendra & Isaacs, 2018).

1.2. Repeat expansions in neurodegenerative disorders

The discovery of pathogenic repeat expansions in *C9orf72* has provided a conceptual framework for understanding other tandem repeat disorders implicated in neurodegeneration. This pivotal finding emphasized that non-coding intronic repeats, beyond classical protein-coding mutations, can drive overlapping neurodegenerative syndromes, linking motor neuron dysfunction with broader multisystem neurological manifestations (DeJesus-Hernandez et al., 2011).

Some ALS-associated genes were found to be mutated in other neurodegenerative conditions. Notably, the *C9orf72* gene, where tandem repeat expansions are the most common cause of ALS and frontotemporal dementia, has also been detected in few patients with Parkinsonism, Huntington's-like disease, and Alzheimer's disease (Cooper-Knock et al., 2014).
On the other hand, although *C9orf72* is currently the only STR expansion proven to cause ALS/FTD, intermediate STR expansions in the spinocerebellar ataxia (SCA) genes, *ATXN1* (SCA1) and *ATXN2* (SCA2), have been associated with ALS risk (Elden et al., 2010a; Lattante et al., 2018).

Modest expansion of 27–33 CAG repeats in the *ATXN2* gene increase the risk of developing ALS (Elden et al., 2010a). Moreover, even if CAG repeat expansions in *HTT* represent a well-established cause of Huntington`s disease, an enrichment of *HTT* repeat expansions was also reported in ALS and FTD patients (Dewan et al., 2021; Zimmermann et al., 2025).This evidence suggests that STR expansions have shared pathophysiological mechanisms that can give rise to different clinical phenotypes (e.g., ALS or FTD) and, in certain circumstances, different underlying pathologies (e.g., HD or ALS-FTD).

1.3. *RFC1*: biological function and repeat expansions

As discussed before, some tandem repeat expansions were found to act as a common pathogenic mechanism across diverse neurological diseases, highlighting how unstable repetitive elements in both coding and non-coding regions can generate overlapping phenotypes, disrupt fundamental cellular processes, and contribute to progressive neurological decline (Ishiura et al., 2019; Paulson, 2018).

This conceptual framework naturally extends to more recently identified expansion disorders, including those involving the Replication Factor C Subunit 1 (*RFC1*) gene (OMIM #102579), further broadening the spectrum of neurodegenerative syndromes linked to repeat instability (Cortese et al., 2019; Rafehi et al., 2019).

RFC1 is an emerging gene of interest that is crucial for DNA replication and repair inside the clamp loader complex (Arbel et al., 2021).

The *RFC1* gene, located on chromosome 4p14, encodes a vital component of the replication factor C complex, which is crucial to DNA replication and repair. This complex operates as a clamp loader that enables the loading of proliferating cell nuclear antigen (PCNA) onto DNA during replication, hence assuring the processivity of DNA polymerase (Lee & Park, 2020).

The *RFC1* protein consists of many conserved domains that facilitate ATP binding and DNA interaction, rendering it essential for genomic stability maintenance (Schrecker et al., 2022).

Despite *RFC1* being widely expressed in several tissues, including the nervous system, its malfunction has predominantly been associated with neurodegenerative phenotypes rather than universal replication failure (Gaubitz et al., 2020).

In 2019, a biallelic intronic AAGGG repeat expansion in *RFC1* was identified as a causative factor for Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome (CANVAS), a multisystem neurodegenerative disorder characterized by cerebellar ataxia, sensory neuropathy, and vestibular dysfunction (Cortese et al., 2019). Prior to this discovery, the etiology of CANVAS was largely unknown, and the disorder was primarily defined based on clinical criteria. This biallelic AAGGG pentanucleotide expansion in the second intron of *RFC1* is now recognized as the most common genetic cause of CANVAS (Dominik et al., 2023). The expansion resides in the poly(A) tail of an AluSx3 element and differs in both size and motif from the reference (AAAAG)11 allele (Fig. 1).

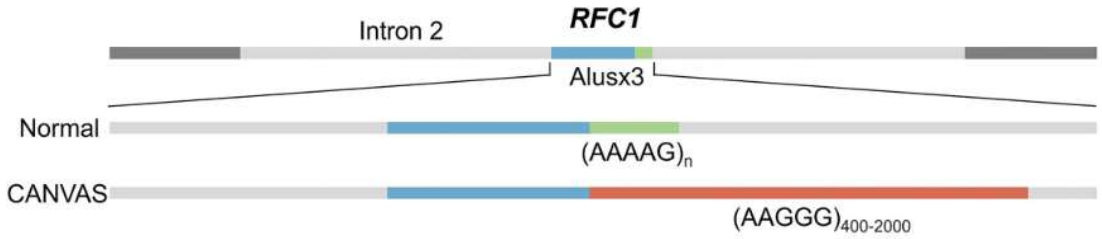

Figure 1: Localization of repeat expansion in RFC1 gene. doi.org/10.1038/s41588-019-0372-4, doi.org/10.1038/s41588-019-0387-x.

The developing association between repeat expansions in an intronic region of *RFC1* and late-onset neurological disorders indicates that the pathophysiology may be more attributable to toxic RNA or modified splicing processes rather than to the loss of protein function (Davies et al., 2022; Wu et al., 2014).

1.3.2 Molecular mechanisms of RFC1 repeat-mediated pathogenesis

Repeat expansion mutagenesis is the process by which short tandem nucleotide repeats in DNA undergo abnormal elongation, leading to numerous human diseases, particularly neurological and neuromuscular disorders. These expansions arise from several intrinsic biological mechanisms that destabilize repetitive DNA sequences.

It has been hypothesized that the emergence of the (AAGGG)n repeat motif occurred as part of an inactivation process involving G interruption of the poly(A) tail of the retrotransposon AluSx3.

Repetitive DNA motifs, particularly G-rich sequences, can form G-quadruplexes (G4s), which are secondary DNA structures stabilized by non-canonical Hoogsteen hydrogen bonds. These structures can act as transcriptional regulators. In particular, they can block the progression of RNA polymerase and interfere with mRNA translation (Fig. 2).

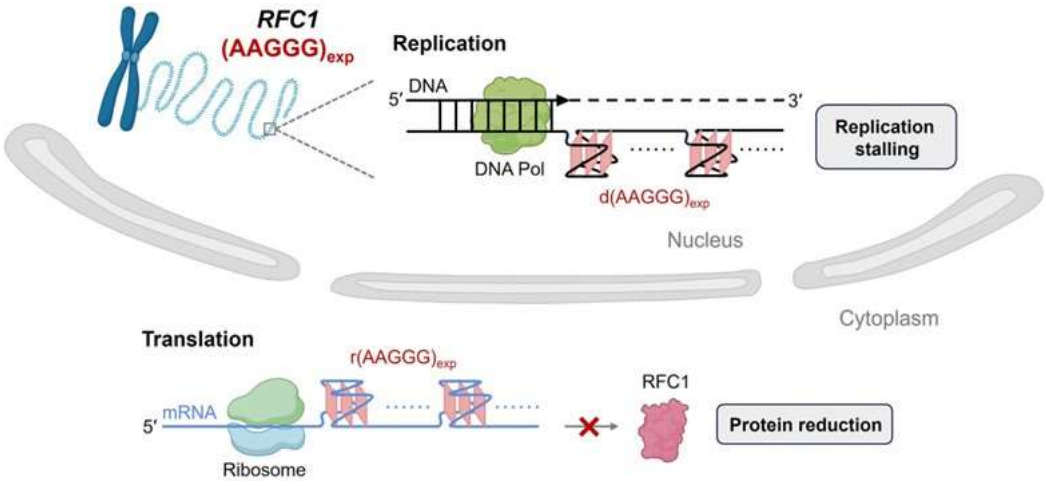

Figure 2: A proposed model for the functional consequence of G4 formation in the pathogenic RFC1 AAGGG repeats. The expanded AAGGG repeats form DNA G4 structure(s) that impede polymerase processivity and cause replication stalling, and form RNA G4 structure(s) that impair translation and reduce protein production. doi.org/10.1093/nar/gkae032.

These structures have been shown to increase the exposure of single-stranded DNA to damaging environmental agents. They may promote repeat expansion and contribute to genomic instability

during both meiotic and mitotic divisions, as well as following DNA damage. All pathogenic repeat configurations have demonstrated high G-quadruplex (G4) propensity scores, in contrast to the non-pathogenic (AAAAG)n sequence (Y. Wang et al., 2024).

1.3.3 *RFC1* repeat expansions in ALS and Motor Neuron Disorders

Along with the well-known association with CANVAS (ref), *RFC1* expansions have been associated with many neurological disorders such as Parkinsonian syndromes, sensory neuropathies, late-onset ataxia and amyotrophic lateral sclerosis (ALS) (Sullivan et al., 2021).
Recent years have witnessed an uptick in the prevalence of biallelic *RFC1* repeat expansions in patients exhibiting symptoms of primary motor neuron diseases such ALS and progressive myelitis (PMA).
This clinical extension suggests that *RFC1* may have an unrecognized role in disorders involving motor neurons, cerebellum, and sensory blurring (Calezis et al., 2024; Schaub et al., 2024).
According to these findings, *RFC1* expansions have the potential to change phenotype via modifications to repeat size, motif configuration, and other epigenetic and genetic factors. Atypical symptoms may resemble those of ALS or other forms of motor neuron disease (MND), such as involvement of upper and/or lower motor neurons, limb weakness and stiffness, fasciculations, and muscle atrophy.
Cognitive impairment and Parkinsonian features in certain individuals also point to involvement of the central nervous system.
It is possible to miss a diagnosis due to presenting variability and slow progression, especially in cases where motor abnormalities mask or diminish sensory clues. This variation emphasizes the need for genetic testing and clinical vigilance in rare or slowly developing neurological diseases (Cortese et al., 2022).
Pathogenic *RFC1* expansions have been investigated in patients with MND, including those clinically diagnosed with ALS. A first large-scale analysis of 1069 American sporadic ALS patients of white, non-Hispanic European ancestry demonstrated no pathogenic expansions, suggesting that *RFC1* does not represent a major contributor to classical ALS pathogenesis (Abramzon et al., 2021).

Conversely, Cortese et al. highlighted that *RFC1* expansions extend beyond the canonical CANVAS phenotype, raising the possibility that these repeat expansions may contribute to atypical ALS presentations, particularly in patients with overlapping motor and sensory or cerebellar features (Cortese et al., 2019).

Moreover, Schaub et al., identified biallelic *RFC1* expansions in 1,9% of patients belonging to a 107 patients German cohort presenting with ALS and primary lateral sclerosis, in some cases accompanied by mild motor impairment without the classical sensory or cerebellar manifestations typically observed in CANVAS (Schaub et al., 2024). These conflicting findings suggest that *RFC1* may contribute to a broader clinical spectrum than previously assumed, supporting the possibility that *RFC1* expansions may contribute to a rare subset of MND cases, particularly those with coexisting sensory, vestibular, or cerebellar features (Cortese et al., 2019).

Considering this, a re-evaluation of genetic risk factors in atypical ALS patients is therefore warranted, particularly given reports of higher-than-expected frequencies of *RFC1* expansions across the German MND cohorts.

Since that time, *RFC1* has been the focus of rigorous research owing to its possible association with a broader spectrum of neurodegenerative symptoms, including those that intersect with motor neuron illnesses (Colucci et al., 2022; Gisatulin et al., 2020).

Following the identification of *RFC1* repeat expansions as the genetic cause of CANVAS, researchers began to observe cases with partial or atypical clinical presentations (Magy et al., 2022). Notably, several individuals diagnosed with CANVAS or other sensory ataxias also exhibited motor neuron involvement, including evidence of degeneration in both upper and lower motor neurons (Huin et al., 2022; Reyes-Leiva et al., 2022).

These findings suggest a clinical continuum bridging classical CANVAS and ALS-like disorders, prompting a re-evaluation of diagnostic criteria for motor neuron diseases (MND), particularly in cases with atypical features such as slow progression, sensory involvement, or cerebellar signs (Huin et al., 2022).

Expanding the clinical spectrum beyond CANVAS to include MND-like phenotypes raises critical questions about the underlying pathophysiology and the broader role of *RFC1* repeat expansions in neurodegenerative processes. As a result, there is increasing interest in screening MND cohorts

for *RFC1* expansions, with the possibility that a subset of patients previously classified as ALS or other MNDs may, in fact, harbour a genetically distinct condition (Schaub et al., 2024).

1.3.4 Repeat motif variability at the *RFC1* locus

The most common wild-type allele contains a stretch of 11 (AAAAG) repeats. However, this locus is highly polymorphic, and non-pathogenic expansions of (AAAAG)n and (AAAGG)n are frequently observed.Unlike other repeat expansion disorders, *RFC1* spectrum disorder does not simply require the expansion of a wild-type motif but also substitution with a pathogenic motif, most commonly (AAGGG)n. This motif is considered disease-causing in CANVAS, particularly when the biallelic expansion exceeds ≥250 repeat units (ru). Notably, expansion of the wild-type (AAAAG)n motif to ≥450 ru has also been associated with pathogenicity.
Additionally, the RFC1 locus is highly polymorphic and, to date, at least 10 motifs have been identified, 6 of which are likely pathogenic when largely expanded(Cortese et al., 2019; Schaub et al., 2024) (Fig. 3).

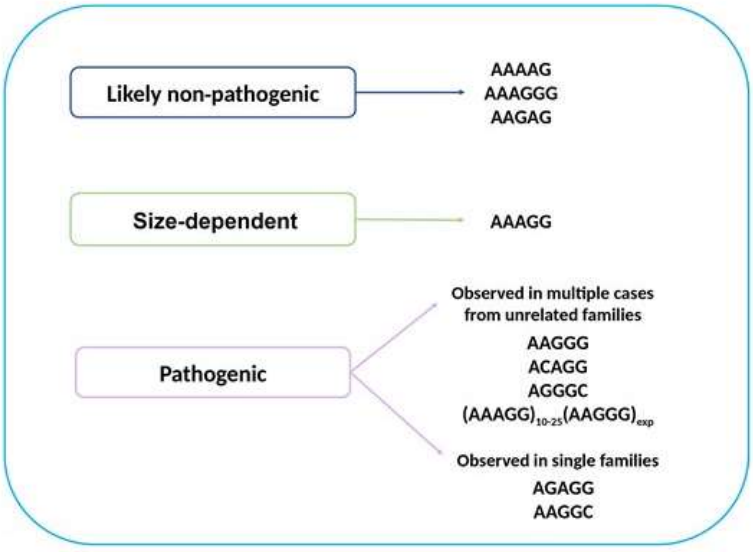

Figure 3: Normal and pathogenic significance of repeat expansion motifs at the RFC1 locus. doi.org/10.1093/brain/awad240.

Biallelic expansion of the (AAGGG)n pentanucleotide repeat, which typically replaces the reference (AAAAG)n repeat, represents the best-characterized pathogenic pattern among the

known motifs (Davies et al., 2022). It is identified in approximately 90% of CANVAS cases in European population (Dominik et al., 2023).

Although the pathogenicity of other repeat motifs, such as (ACAGG)n, (AAAGG)n, and combinations thereof, remains under investigation, these variants have also been detected in affected individuals (Watanabe et al., 2022).

In affected individuals, repeat expansions can range from several hundred to thousands of repeat units, with considerable inter-individual variability.

Important factors in clinical interpretation include the repeat motif type, its configuration, and zygosity status (homozygous, heterozygous, or compound heterozygous).

Certain repeat motifs appear to show population-specific distributions. For example, (ACAGG)n is relatively common in East Asian populations, (AAGGC)n in South Asian populations, and expanded (AAGGG)n alleles with specific configurations have been reported in Māori population (Beecroft et al., 2020; Dominik et al., 2023; Tang et al., 2025).

Conversely, motifs such as (AAGAG)n and longer (AAAAG)n repeats may represent benign variants or potential modifiers, although their clinical relevance remains uncertain (Paulson, 2018)

1.4. Diagnostic approaches and challenges

Finding strategies to identify *RFC1* repetitions and expansions of non-coding intronic repeats requires specialized molecular approaches.

When looking for this kind of repeat expansion and the specific pattern it represents, the gold standard is the repeat-primed polymerase chain reaction (RP-PCR). But RP-PCR can't tell you how big the repeat expansion is or if it's biallelic (Akçimen et al., 2019). For high-resolution investigation of repeat expansions, long-range PCR and long-read sequencing technologies like PacBio and Oxford Nanopore have been proposed as potential methods (Miyatake et al., 2022).

Despite these advances, current diagnostic approaches have important limitations, including false-negative results, limited inter-laboratory reproducibility, and challenges in distinguishing monoallelic from biallelic expansions. These technical constraints complicate clinical interpretation and hinder large-scale implementation.

Clinically, biallelic *RFC1* expansions may mimic classical motor neuron disorders such as ALS or spinal muscular atrophy (SMA), making differential diagnosis challenging. Therefore, RFC1 testing should be considered in patients with unexplained or atypical neurodegenerative

phenotypes, particularly when standard genetic panels are negative. Integrating RFC1 genotyping into broader diagnostic strategies—such as comprehensive gene panels, exome sequencing, or repeat-focused algorithms—may improve diagnostic yield, facilitate appropriate clinical management, and support the development of targeted therapeutic approaches (Cortese et al., 2020; Ebbert et al., 2018; Owusu & Savarese, 2023).

2. AIM OF THE STUDY

This thesis aimed to investigate the prevalence and molecular characteristics of *RFC1* repeat expansions in a cohort of patients with motor neuron diseases (MND) recruited from the Department of Neurology of the Maggiore della Carità hospital in Novara, Italy.

Specifically, a primary objective of this study was to evaluate the reproducibility in the Italian population of the findings reported by Schaub et al. (2024), who identified biallelic *RFC1* expansions in 1.9% of a German cohort comprising 107 patients diagnosed with amyotrophic lateral sclerosis (ALS) and primary lateral sclerosis (PLS).

Furthermore, considering the known polymorphic nature of the *RFC1* locus, we aimed to assess the motif composition of the detected expansions and evaluated their potential pathogenicity. This approach was intended to clarify the contribution of *RFC1* repeat expansions to MND and to explore their possible role in atypical or overlapping neurological phenotypes.

Ultimately, these analyses may help determine whether *RFC1* screening should be integrated into the standard diagnostic workflow for MND.

3. MATERIALS AND METHODS

3.1 Patient cohort

The study involved 250 patients, of whom 237 had been diagnosed with ALS, 7 with FTD, and none had a combined diagnosis of ALS/FTD. In addition to ALS and FTD cases, the cohort included samples classified as ataxia (n = 2), cognitive decline (n = 3), Charcot disease (n = 1) recruited from the neurology department of the Maggiore della Carità Hospital of Novara, Italy. The clinical diagnosis was assessed in accordance with established clinical standards. Before participating, all patients gave their informed consent.

3.2 Extraction of genomic DNA

We extracted genomic DNA (gDNA) from whole blood samples maintained with EDTA using salting-out and mini-prep techniques (Promega Reliaprep Mini Kit (Promega Corporation, Madison, Wisconsin, USA, Promega), according to the manufacturer's instructions. Briefly, DNA extraction began with cell lysis and continued with protein precipitation and DNA purification using ethanol based washing reagent. To assess the concentration and the purity of extracted gDNA, the NanoDrop spectrophotometer was used to check the absorbance ratios A260/A280 for protein contamination and A260/A230 for contamination from organic compounds and salts, such as phenol, guanidine, and residual ethanol. When the purity levels of the samples were higher than A260/A280 > 1.8 and A260/A230 > 2.0, we considered them suitable for use in further downstream processes. We stored the DNA at a temperature of -20°C from the time of extraction until analysis.

3.3 *RFC1* Repeat expansion analysis via PCR

The *RFC1* repeat region was amplified using flanking polymerase chain reaction (PCR), which enabled the identification and sizing of gene expansions. The reaction was carried in a total volume of 20 µL, including a 5X buffer, deoxynucleotide triphosphates (dNTPs), $MgCl_2$ (a cofactor for the DNA polymerase enzyme, essential for optimal enzyme activity, ensuring the efficiency and specificity of the reaction), forward and reverse primers that targeted the areas around the repeat

expansion, P5 primer, which is partially complementary to the genoF primer was used as a universal fluorescent tail to enable fragment analysis on a capillary electrophoresis platform, and Taq polymerase, the enzyme employed to synthesize DNA. The PCR reactions were performed using GoTaq® G2 Flexi DNA Polymerase (Promega) and set up as follows (Table 1):

REAGENTS	INTIAL CONCENTRATION	FINAL CONCENTRATION	VOLUME PER SAMPLE
Buffer 5X	5X	1X	4.0µL
dNTPs	10mM	0.6mM	1.2
$MgCl_2$	25 mM	1.25 mM	0.9 µL
Forward Primer [10]	10 pmol/µL	0.2 pmol/µL	0.4 µL
Reverse Primer [10]	10 pmol/µL	0.36 pmol/µL	0.72 µL
P5	10 pmol/µL	0.36 pmol/µL	0.72 µL
Taq polymerase	5 U/µL	0.018 U/µL	0.072 µL
Nuclease-free water	-	-	10.0 µL
Final Volume	18 µL		
	+ 2µL of DNA (concentrated 50 ng/µL)		

Table 1: PCR mixture protocol for amplification of RFC1 repeat fragment

PRIMERS	SEQUENCE
*RFC1*_genoF	5'- **ACA TAC GCA TCC CAG TTT GAG ACG** TCA AGT GAT ACT CCA GCT ACA CCG TTG C -3’
*RFC1*_genoR	5’- GTG GGA GAC AGG CCA ATCTCA G -3'
P5	5'- **ACA TAC GCA TCC CAG TTT GAG ACG** -3'

Table 2 : Oligonucleotides used for PCR. The P5 sequence represents the 5' adapter tail included in the RFC1 forward primer (bold).

Cycling Parameters used were denaturation at 96°C for 5 min, 33 cycles consisting of 30s denaturation at 96°C, 30s annealing at 64°C, and 60'' extension at 72°C, followed by a final extension at 72°C for 5 min (Table 2).

STEPS	TEMPERATURE	TIME	CYCLES
Initial Denaturation	96 °C	5 min	1
Denaturation	96 °C	30 sec	33
Annealing	64 °C	30 sec	
Extension	72 °C	1 min	
Final Extension	72 °C	5 min	1

Table 3: PCR thermal protocol.

3.4 Agarose Gel Electrophoresis

After the reaction, all PCR products were visualized on a 1.5% agarose gel. This gel was made by dissolving agarose powder in a Tris-Acetate-EDTA solution (TAE), heating it. SybrSafe DNA gel stain (Thermo Fisher Scientific) was added at a final volume of 3 µL per 100 mL of agarose solution for DNA visualization. When bound to DNA molecules and illuminated with UV light, these dye molecules fluoresce brightly. We poured the gel into a casting tray and used combs to solidify it, creating loading wells. Before loading, 3 µL of loading dye was used with 3 µL of each PCR sample. It aids in the process of loading samples into wells and tracking their migration through the gel. It contains colored dyes and a high density, which helps with run visualization and ensures the sample sinks properly into the wells. A 100 base pairs DNA ladder and a 1kb DNA ladder (Promega) were used to reference the molecular size. The electrophoresis was carried out for 30 minutes at voltages ranging from 100 to 120 V. After separation, bands were visualized under UV light. Agarose gel electrophoresis allowed assessment of amplification success and approximate fragment size.

Samples displaying a single visible band suggested the presence of one detectable allele, whereas samples with two distinct bands indicated two alleles of different sizes. However, definitive

genotype classification (homozygous versus heterozygous) was determined only after capillary electrophoresis fragment analysis, which provides higher resolution and allows discrimination between closely sized alleles that may appear as a single band on agarose gel.

3.5 Capillary Electrophoresis Fragment Analysis

To achieve high-resolution sizing of the PCR products, fluorescent fragment analysis was performed on a SeqStudio Genetic Analyzer (Applied Biosystems). This method provides a 1-base pair resolution, in contrast to gel electrophoresis, which typically offers a resolution of 10–20 base pairs. Fragment analysis was employed to accurately measure DNA fragment lengths and determine the number of repeat units of the analyzed motif, using 500 ROX dye size standards (Applied Biosystems) as molecular sizing reference.

To minimize interference from excessive background noise, PCR products were diluted based on the intensity of the bands observed during gel electrophoresis. Dilution ratios varied from 1:20 to 1:70, or in some cases, no dilution was applied. Once the PCR products were appropriately diluted, for each sample, a master mix consisting of 8.75 µL of formamide and 0.25 µL of the ROX500 size standard was prepared. 9 µL of the master mix was dispensed into each well of the optical plate, and 1 µL of the PCR product (fluorescently labeled with P5) was added to each well. The samples were incubated at 96°C for 2 minutes, quickly cooled on ice, and then loaded into the SeqStudio (Thermo Fisher Scientific). Due to instrument limitations, fragments exceeding ~800 bp could not be accurately sized by capillary electrophoresis and were therefore estimated by agarose gel migration

3.6 Data Processing and Fragment Analysis Using GeneMapper Software

GeneMapper 6 (Applied Biosystem) software was used to conduct the analysis of the data. By comparing the peaks to the ROX500 size standard (see Figure 6), the size of the PCR products was determined.

3.7 Modified PCR protocol for samples showing a single expanded allele

For samples that showed a large allele as a single band on the gel, modified PCR amplification was attempted. This modification was performed to improve amplification efficiency of large fragments, as expanded alleles may amplify less efficiently due to their increased length and potential secondary structure formation. This amplification involved modifications to the PCR thermal cycling protocol (Table 3) (increased extension time from 1‘ to 1’30”), while the PCR mixture and primer concentrations remained unchanged (Table 1).

STEPS	TEMPERATURE	TIME	CYCLES
Initial Denaturation	96 °C	5 min	1
Denaturation	96 °C	30 sec	33
Annealing	64 °C	30 sec	
Extension	72 °C	1.30 min	
Final Extension	72 °C	5 min	1

Table 3: Modified PCR thermal protocol.

3.8 Sanger Sequencing

Sanger sequencing was performed on samples exhibiting bands larger than 1000 base pairs to characterize the motif of the large-expanded alleles observed.

When two distinct bands were observed on the gel, the bands were excised from the gel and, purified using the NucleoSpin Gel and PCR Clean-up Kit (Macherey-Nagel) following the manufacturer’s protocol. The gel-purified product was quantified at the nanodrop spectrophotometer and Sanger sequenced. For samples with a single band visible on the gel, purification was carried out using EuroSAP- PCR enzymatic kit (Euroclone, Pero, MI, Italy), after which the purified product was directly sequenced. The cycling parameters for Sanger sequencing reaction are reported in Table 4.

Sequencing mix was prepared using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystem). The amplified products were then run on a SeqStudio™ Genetic Analyser (Thermofishers Scientific) and the resulting electropherograms were analysed using Chromas 2.6.6.

STEPS	TEMPERATURE	TIME	CYCLES
Initial Denaturation	96 °C	3 min	1
Denaturation	96 °C	20 sec	25
Annealing	50 °C	5 sec	
Extension	60 °C	4 min	
Resting	10 °C	---	

Table 4: PCR thermal protocol for Sanger Sequencing.

3.9 Repeat Primed PCR (RP-PCR)

Repeat-primed PCR (RP-PCR) was performed to determine the motif and the allelic configuration of two notably expanded alleles observed in one patient (ID: 13470).

To promote stronger annealing of the reverse primer to the variable (A/AA/-) 3′ terminal region of the microsatellite and its surrounding sequence, an extension of 5 to 27 nucleotides was added to the 5′ end of the reverse primer. This modification favoured the amplification of the larger allele, thereby enabling approximate sizing of the expansion in certain cases.

RP-PCR was performed using Taq polymerase (Promega), and primers targeting the most common AAAAG, AAGGG repeat motifs. Specific details of the primer design, reaction conditions, and thermocycling protocol are reported below (Tables 5, 6 and 7).

REAGENTS	INTIAL CONCENTRATION	FINAL CONCENTRATION	VOLUME PER SAMPLE
Buffer 5X	5X	1X	4.0µL
dNTPs	10mM	0.6mM	1.2
$MgCl_2$	25 mM	1.25 mM	0.9 µL
Forward Primer [10]	10 pmol/µL	0.5 pmol/µL	1.0 µL
Reverse Primer [5]	5 pmol/µL	0.25 pmol/µL	1.0 µL
Anchor [10]	10 pmol/µL	0.5 pmol/µL	1.0 µL
Taq polymerase	5 U/µL	0.018 U/µL	0.072 µL
Nuclease free water	-	-	10.0 µL
Final Volume	18 µL		
	+ 2µL of DNA (concentrated 50 ng/µL)		

Table 5: PCR Mixture protocol for RP PCR.

Temperature	**Time**	**Cycles**
96 °C	5 min	1
96 °C	35 sec	10
65 °C	35 sec	
68 °C	8 min	
96°C	35 sec	30
65°C	35 sec	
68°C	8:20 min	
72°C	10 min	1

Table 6: Thermal protocol for RP PCR.

PRIMERS	SEQUENCE
*RFC1*_RP-PCR_F	5'- /56-FAM/TCAAGTGATACTCCAGCTACACCGT -3’
*RFC1*_(AAAAG)11 _R2	5’- **CAGGAAACAGCTATGACC**AACAGAGCAAGACTCTG TTTCAAAAAGAAAAGAAAAGAAAAGAAAA -3'
*RFC1*_(AAGGG)exp _R1	5'- **CAGGAAACAGCTATGACC**AACAGAGCAAGACTCTG TTTCAAAAAAGGG AAGGGAAGGGAAGGGAA -3'
*RFC1*_(AAGGG)exp_R2	5’- **CAGGAAACAGCTATGACC**AACAGAGCAAGACTCTG TTTCAAAAAGGGAAGGGAAGGGAAGGGAA-3’
*RFC1*_(AAGGG)exp_R3	5’- **CAGGAAACAGCTATGACC**AACAGAGCAAGACTCTG TTTCAAAAGGGAAGGGAAGGGAAGGGAA-3’
*RFC1*_RP-PCR_Anchor	5'- **CAGGAAACAGCTATGACC** -3'

Table 7: Oligonucleotides used for RP-PCR. The bold sequence represents the 5' universal M13 adapter tail incorporated into the repeat-specific primers and corresponds to the RFC1_RP-PCR_Anchor primer.

The resulting PCR fragments were analyzed using a SeqStudio (Thermofishers Scientific) and data analysis was performed using GeneMapper 6 software (Applied Biosystem). Samples that exhibited a distinctive ‘sawtooth’ signal pattern, typically indicative of repeat expansions were flagged for further analysis.

4. RESULT

Given these emerging links, we aimed to explore the frequency and characteristics of *RFC1* repeat expansions in a well-defined patient population diagnosed with ALS and/or FTD.

Our study analysed a cohort of 250 unrelated individuals recruited from department of neurology clinics Mggiore della Carità hospital in Novara. All participants underwent genetic screening using flanking polymerase chain reaction (PCR), a method that enables amplification and size estimation of repeat-containing intronic regions. Additional techniques, including gel electrophoresis, fragment analysis, repeat-primed PCR (RP-PCR), and Sanger sequencing, were employed to precisely size and characterize motifs in selected cases.

A total of 250 genomic DNA samples from patients diagnosed with motor neuron degenerative disorder were analyzed to assess the presence of *RFC1* repeat expansions (Fig. 4).

4.1 Flanking PCR and gel electrophoresis-based genotyping

Initial screening was performed using standard flanking PCR targeting the intronic repeat region of *RFC1*, specifically amplifying the pentanucleotide repeat motif. Since standard flanking PCR amplifies DNA fragments up to 2–3kb, the absence of a PCR product could be indicative of a homozygous expansion, while the presence of two amplification bands suggests the presence of non-pathogenic or heterozygous alleles. The PCR products were then analyzed by agarose gel electrophoresis to confirm the success of amplification and to estimate approximately the allele size. All 250 samples showed clear and distinct bands on the agarose gel electrophoresis, indicating robust and specific amplification of the target region. No samples showed absence of *RFC1* amplification; thus, there were no patient samples that showed biallelic repeat expansions that were higher than the threshold for pathogenicity.

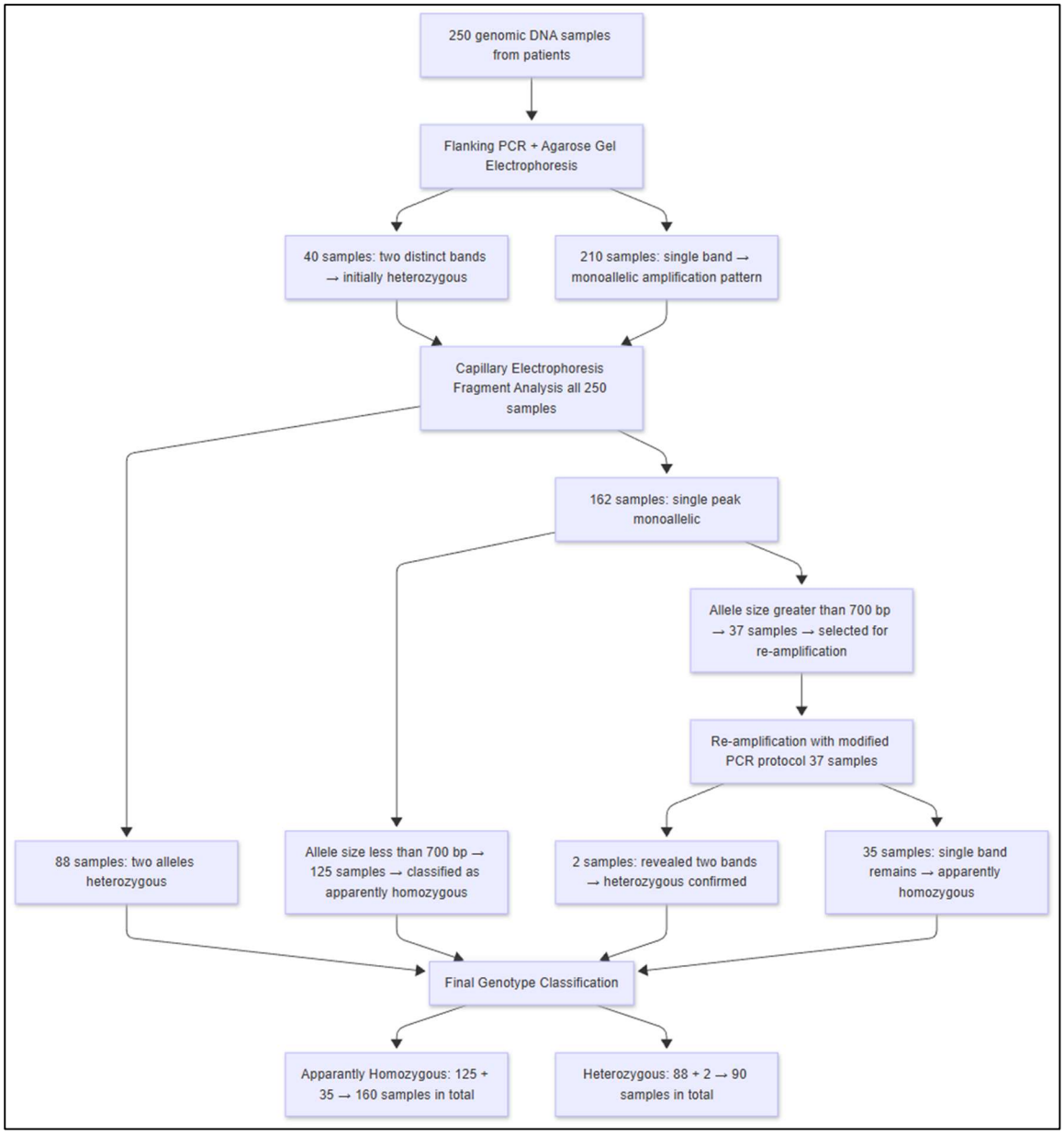

Figure 4 Experimental Workflow for the Detection and Characterization of RFC1 Repeat Expansions. The flowchart illustrates the sequential screening process applied to the cohort of 250 patients.

Notably, some samples displayed slightly higher molecular weight bands, suggestive of longer alleles, and were flagged for closer analysis in subsequent steps (Fig. 8).

Among the 250 samples analyzed by agarose gel electrophoresis following flanking PCR amplification, 210 samples (84%) displayed a single distinct band, whereas 40 samples (16%) showed two clearly distinguishable bands. Samples with two distinct bands were classified as

heterozygous, reflecting successful amplification of both alleles. Samples displaying a single band were categorized as monoallelic amplification patterns, which may represent either true homozygous or heterozygous samples with a second large unamplifiable allele.

The threshold for pathogenicity in *RFC1* for the (AAGGG)n motif is established at biallelic expansions of approximately 1,300 base pairs. The majority of samples (92%) exhibited one or two bands within a size range of approximately 350–800 base pairs (~11-98 repeat units, ru, as explained in section 4.3.), consistent with the most common non-pathogenic allele sizes observed in the cohort.

A smaller subset of samples (8%) demonstrated at least one allele between 800–1300 base pairs (~100-200 ru), representing relatively larger alleles but still below currently recognized pathogenic thresholds.

One sample (ID: 13470) exhibited two distinct bands at approximately 1100 base pairs and 3000 base pairs (~140 and 540 ru, respectively) on agarose gel electrophoresis. The ~3000 base pair fragment represented the largest allele detected during the initial screening of the cohort. No other samples demonstrated band sizes exceeding 1300 base pairs (~200 ru) at this stage.

Subsequent investigations focused exclusively on samples showing at least one band greater than 700 base pairs on agarose gel, as well as those presenting a single band but raising suspicion of a potentially unamplifiable second allele. In total, 37 samples met these criteria. All samples showing a single band in this context were classified as apparently homozygous, acknowledging the possibility that they could in fact be heterozygous with one unamplifiable expanded allele.

These 37 samples were subjected to re-amplification using a modified PCR protocol with extended elongation time. Following this approach, two samples (ID: 13819 and ID: 11052), initially appearing as single-band cases, revealed the presence of a second large allele. Specifically, sample ID: 13819 exhibited two distinct bands at approximately 1000 base pairs and 3000 base pairs (~140 and 540 ru, respectively), while sample ID: 11052 showed bands at approximately 700 base pairs and 4000 base pairs (~80 and 740 ru, respectively). These findings demonstrate that modified PCR conditions enabled the detection of previously unamplified expanded alleles in a subset of samples initially classified as apparently homozygous.

4.2 Capillary Electrophoresis Fragment Analysis and Genotyping

Following agarose gel electrophoresis, all 250 samples were analyzed using fluorescent capillary electrophoresis fragment analysis to refine allele sizing and confirm genotypes. In the majority of cases, fragment sizes were concordant between agarose gel and capillary electrophoresis, thereby validating the reliability of the initial screening approach. Representative gel images are shown in Figure 5 and 6.

Out of 250 samples, 88 samples demonstrated two distinct alleles on fragment analysis, including 40 samples that initially showed two bands on agarose gel and an additional 48 samples in which the alleles were closely sized and appeared as a single band on gel. These 88 samples were consistently classified as heterozygous, as confirmed by both methods.

The remaining 162 samples displayed a single band on agarose gel and a single corresponding peak on capillary electrophoresis, indicating monoallelic amplification. These samples were further stratified based on allele size. Samples showing a single allele below 700 base pairs (~80 ru) were classified as apparently homozygous, as these fragment sizes fall well within the established non-pathogenic range and did not require additional investigation.

Among these 160 monoallelic cases, 37 samples exhibited a single band greater than 700 base pairs and were therefore selected for re-amplification. Following re-amplification, 2 samples revealed a second allele and were reclassified as heterozygous**,** while the remaining 35 samples retained a single band and were designated as apparently homozygous. In these cases, it was not possible to determine whether the genotype represented a true homozygous or heterozygous with a second unamplifiable expanded allele.

Alleles exceeding 700 base pairs could not be reliably resolved by the capillary electrophoresis fragment analyser due to its upper detection limit. For such cases, allele size estimation was based on electrophoretic migration on agarose gel using 100 bp and 1 kb DNA ladders as molecular standards, and these values are therefore reported as approximate.

Fragment analysis was essential for resolving closely sized heterozygous alleles that occasionally appeared as a single thick band on agarose gel. In several instances, capillary electrophoresis

revealed two distinct but closely spaced peaks, thereby correcting the initial gel-based interpretation.

An example of gel electrophoresis was reported below (Fig.5).

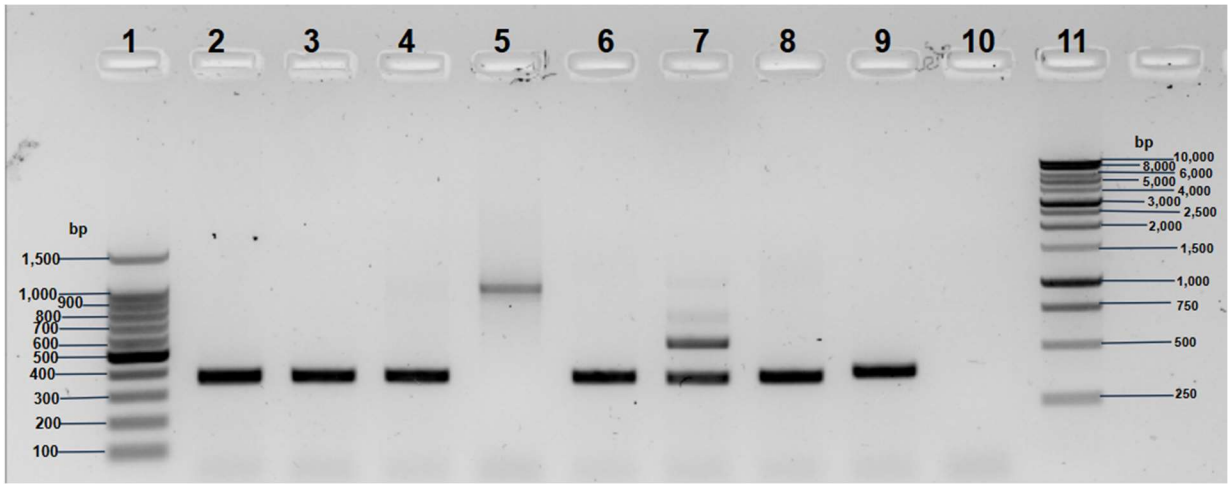

Figure 5 Example of 1.5% agarose gel electrophoresis of PCR products for the RFC1 gene. Lanes 1 and 11: DNA ladders (Promega); lane 10: negative control. Single-band samples (~300–1000 bp) indicate monoallelic amplification, while lane 7 shows two bands (~300 bp and ~700 bp), representing a heterozygous genotype.

Lanes 1 and 11 contain the 100 bp and 1 kb DNA ladders, respectively, used as molecular size markers. Lanes 2, 3, 4, 6, 8, and 9 show samples with a single amplification band of approximately 300 bp. Lane 5 displays a single allele of approximately 1000 bp. Lane 7 shows two distinct bands at ~300 bp and ~700 bp, consistent with a heterozygous genotype carrying two different allele sizes. Lane 10 serves as the negative control and shows no amplification, confirming the absence of contamination. As shown in Figure 5, samples exhibiting two distinct bands on the agarose gel, lane 7th with bands around 300 bp and 700 bp, were classified as heterozygous, indicating the presence of two alleles of different sizes.

This classification was confirmed by capillary electrophoresis fragment analysis, which showed two distinct peaks corresponding to the different alleles (Fig. 6d). In some cases, such as the sample in lane 2, the gel displayed a single thick band, which could be misinterpreted as a homozygous pattern. However, capillary electrophoresis fragment analysis revealed two peaks very close to

each other (Fig. 6c), indicating a heterozygous genotype with alleles of similar but not identical size.

In contrast, the samples in lane 3rd and 4th, which also appeared as single bands on the gel, were confirmed to be apparently homozygous, as fragment analysis showed a single well-defined peak, confirming the presence of two alleles of identical size (Fig. 6a and 6b). Remains the possibility of a second unamplifiable allele.

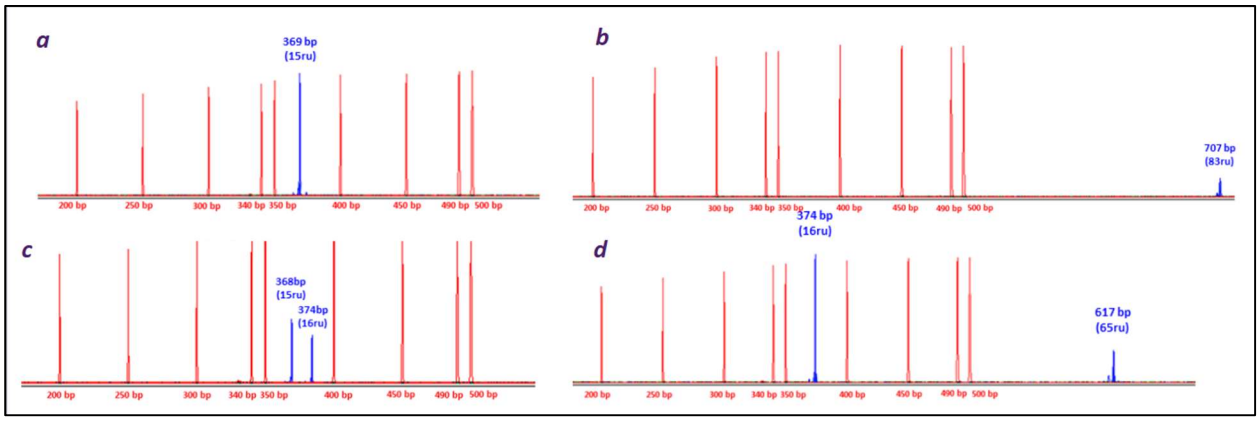

Figure 6: Fragment analysis of RFC1 repeat alleles. Representative electropherograms showing fragment size distribution and repeat unit (ru) number. (a) Apparently homozygous normal allele at 369 bp (15ru). (b) Apparently homozygous with one allele at 707bp (c) Heterozygous with two closely sized normal alleles at 368 bp (15ru) and 374 bp (16ru). (d) Heterozygous genotype with one normal allele at 374 bp (16ru) and one expanded allele at 617 bp (65ru). Blue peaks represent sample fragments; red peaks correspond to the internal size standard.

Moving back to figure 5, the sample in lane 5th, with a band around 1000 base pairs, exceeded the upper detection limit of the fragment analyser. Therefore, the size of the allele was estimated based on agarose gel electrophoresis to correspond to approximatively 140 repeat units. As only one band was visible on the gel, this sample was classified as uncertain in terms of zygosity since it remains the possibility of a second, larger unamplifiable allele. Overall, genotyping was determined through combined evaluation of agarose gel electrophoresis and capillary electrophoresis results; however, samples classified as apparently homozygous remain genetically unresolved without further expansion-specific testing. Based on the finalized genotyping derived from combined agarose gel electrophoresis and capillary fragment analysis, allele sizes were subsequently converted into repeat numbers to enable quantitative assessment of repeat distribution across the cohort. This repeat unit calculation allowed systematic evaluation of allele

frequency, identification of unusually large expansions, and characterization of the overall repeat size spectrum observed in this cohort.

This combined approach of gel electrophoresis and fragment analysis proved essential for accurate genotyping, especially in distinguishing closely spaced heterozygous alleles from true homozygous patterns and identifying cases that require further analysis.

4.3. Repeat units calculation and allele distribution

In this study, the distribution of *RFC1* repeat alleles was analysed in the 250 samples mentioned. Each sample was assessed for two alleles, and the repeat numbers were determined using the following formula:

$$\text{Repeat number} = \frac{\text{Product size (base pair)} - 293}{5}$$

Here, 293 base pairs correspond to the non-repeat flanking region, which is the constant portion of the amplicon. By subtracting 293 base pairs from the product size (in bp), the non-variable part of the sequence is eliminated. The remaining value is then divided by 5 (the length of each repeat unit) to obtain the total number of repeat units (ru).

An example of data is shown in Figure 7.

	Sample ID	**Length_ Allele_1**	**Length _Allele _2**	**Repeats _Allele_ 1**	**Repeats _Allele_ 2**	**Length/r epeats_ Allele_1 (accordi ng to the gel)**	**Length/r epeats_ Allele_2 (accordi ng to the gel)**	**Genotype**
1.	**13593**	364.09	369.09	14	15			heterozygous
2.	**13592**	374	374	16	16			Apparently homozygous
3.	**13569**	368.99	368.99	15	15			Apparently homozygous
4.	**13567**	374.01	374.01	16	16			Apparently homozygous

5.	**13566**	363.2	363.2	14	14			Apparently homozygous
6.	**13564**	369.05	369.05	15	15			Apparently homozygous
7.	**13505**	369.03	850	15	110		850 / 110	heterozygous
8.	**13504**	369.05	369.05	15	15			Apparently homozygous
9.	**13503**	364.12	364.12	14	14			Apparently homozygous
10	**13502**	369.05	369.05	15	15			Apparently homozygous
11	**13501**	373.9	373.9	16	16			Apparently homozygous
12	**13495**	369.04	369.04	15	15			Apparently homozygous
13	**13493**	369.04	369.04	15	15			Apparently homozygous
14	**13492**	369.1	373.99	15	16			heterozygous
15	**13491**	900	1000	120	140	900 / 120	1000 / 140	heterozygous
16	**13490**	359.05	359.05	13	13			Apparently homozygous
17	**13489**	369.05	374.02	15	16			heterozygous
18	**13488**	364.06	369.02	14	15			heterozygous
19	**13485**	364	369.02	14	15			Apparently homozygous
20	**13482**	900		130		900 / 130		Apparently homozygous
21	**13481**	364.14	378.96	14	17			heterozygous
22	**13474**	369.02	373.99	15	16			heterozygous
23	**13470**	1100	3000	160	540	1100 / 160	3000 / 540	heterozygous
24	**13432**	359.12	369.05	13	15			heterozygous
25	**13357**	364.03	364.03	14	14			Apparently homozygous
26	**13346**	364.08	364.08	14	14			Apparently homozygous
27	**13294**	364.09	364.09	14	14			Apparently homozygous

28	**13293**	364.08	369.05	14	15			heterozygous
29	**13292**	359.07	359.07	13	13			Apparently homozygous
30	**13291**	1300		200		1300 / 200		Apparently homozygous

Figure 7: Example of allele sizing, repeat counts, and genotype classification for RFC1 PCR samples.

The first column lists the unique sample IDs. Columns two and three display the lengths of Allele 1 and Allele 2 as determined by fragment analysis. Columns four and five show the corresponding repeat counts, calculated using the formula: (Length in bp−293)/5, where 293 base pairs represent the flanking regions. Columns six and seven show allele lengths and repeat counts estimated from agarose gel electrophoresis only for cases where fragment analysis could not detect large expansions. Red-highlighted entries indicate data obtained from gel estimation rather than fragment analysis. The final column classifies each sample as heterozygous (two clearly distinct alleles), or apparently homozygous (only one allele visible or the second allele not amplifiable).

The table above presents example data from a subset of *RFC1* samples to illustrate how allele sizes, repeat counts, and genotype classifications were determined. It serves as an illustrative example of the various allele configurations observed in the study cohort, including typical, expanded, and ambiguous genotypes.

Allele size distribution analysis demonstrated that allele sizes of the *RFC1* repeat ranged from 11 to 740 units, exhibiting a highly skewed distribution and the majority of alleles were concentrated within the lower repeat unit range. The majority of alleles (approximately 50% of all observed alleles) were concentrated within a narrow range of 13–16 repeat units, indicating strong conservation at this locus in the cohort. Among these, the 15-repeat units allele was the most frequent, occurring in 27.4% of genotypes, while alleles with 14 repeat units and 16 repeat units accounted for 11% and 8% of alleles, respectively. Alleles beyond 16 repeat units showed a marked decline in frequency, with the 17–21 repeat units range occurring at less than 3%, marking the transition toward the expanded tail of the distribution.

Alleles exceeding 100 repeat units were uncommon, representing less than 5.6% of the total allelic pool. Accurate characterization of alleles above 200 repeats (~1% of samples) required methodological adjustments due to limitations in fragment resolution. Capillary electrophoresis allowed reliable sizing up to 94 repeats, whereas longer alleles showed reduced signal intensity and decreased resolution of individual repeat units. For alleles exceeding the approximate 700 bp

detection limit of CE, repeat numbers were estimated based on migration patterns in agarose gel electrophoresis relative to 100 bp and 1 kb DNA ladders.

The overall distribution demonstrates that large expansions are typically heterozygous, with one expanded allele per individual. Three samples exhibited extreme expansions, representing the most pronounced instabilities within the cohort. Samples 13470 and 13819 showed amplicons of approximately 3,000 bp, corresponding to ~540 ru, and sample 11052 displayed the largest expansion at ~4,000 bp, corresponding to ~740 ru. All of these samples were heterozygous, carrying one expanded allele alongside a normal allele. In contrast, sample 13291 exhibited only a single large band of approximately 1,300 bp (~200 ru), suggesting the presence of a single detectable allele. For subsequent analyses, we focused exclusively on samples carrying at least one allele with >200 repeat units**,** ensuring that all large expansions were captured for detailed investigation.

4.4 Sanger Sequencing of selected samples

To characterize the repeat motif composition of large alleles, selected samples were subjected to Sanger sequencing.

Large expanded alleles were analyzed in samples ID: 13470, ID: 13819, ID: 11052, and ID: 13291. Sequencing was performed on both forward and reverse strands to ensure sequence accuracy.

For samples ID: 13470, ID: 13819, and ID: 11052, the target bands were excised from agarose gel and purified prior to sequencing. In contrast, sample ID: 13291 was sequenced directly after ExoSAP treatment of the PCR product.

Agarose gel electrophoresis of these samples is reported in figure 8.

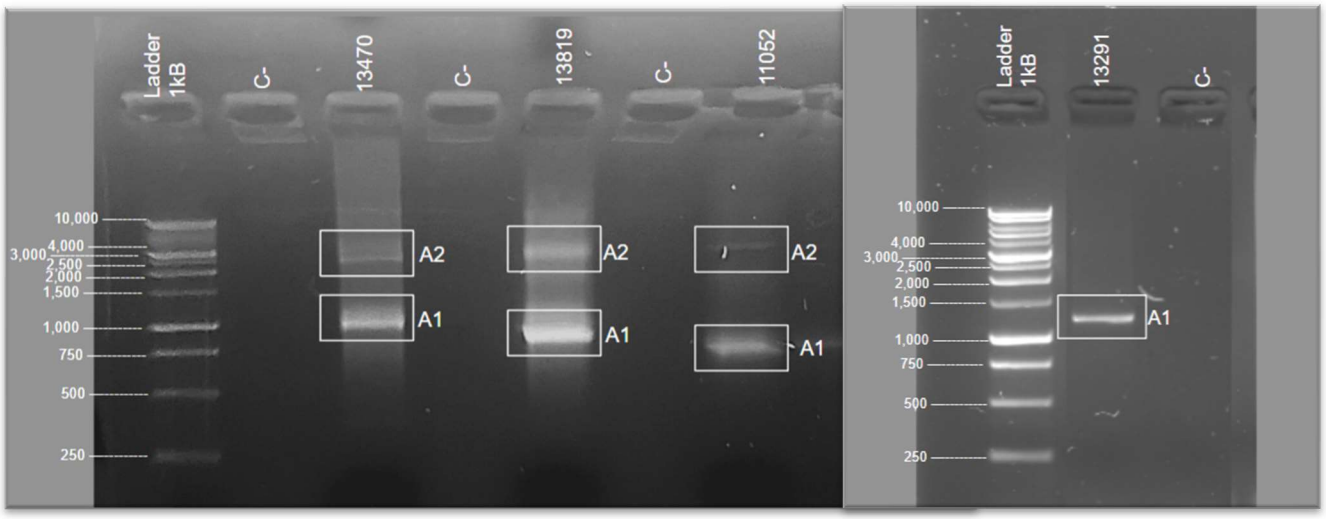

Figure 8: Agarose gel electrophoresis of RFC1 PCR products in samples with large-expanded alleles. Electrophoresis was performed on a 1.5% agarose gel to resolve high molecular weight fragments. 1 kb DNA ladder used as a molecular size marker (Promega). Samples 13470, 13819, and 11052 each display two distinct bands, corresponding to the smaller allele (A1) and the larger expanded allele (A2), confirming a heterozygous expansion state in these individuals. Sample 13291 presents as an apparently homozygous single high molecular weight band. C- represents the negative control, showing no amplification as expected.

As can be observed from the gel reported above, samples ID: 13470, ID: 13819, and ID: 11052 exhibit two distinct bands, consistent with a heterozygous expansion. Sample ID: 13291 presents a single high molecular weight band, indicating the presence of a single amplifiable allele.

These PCR products were subjected to Sanger Sequencing to further characterize the repeat motif of the amplified alleles.

Given the highly polymorphic structure of this locus, the primary goal was to determine whether the observed expanded alleles consisted of pathogenic motifs, particularly the AAGGG repeat, or of non-pathogenic/interrupted repeat motif, such as AAAAG or AAAGG.

According to our analysis the observed motifs for each allele are reported below:

- Sample ID:13470

Sample ID: 13470 exhibited two distinct bands on gel electrophoresis, approximately 1100 base pairs and 3000 base pairs in size. Sanger sequencing of the lower band revealed a pure AAAAG motif, with at least 105 consecutive AAAAG repeats observed in the forward read. Reverse sequencing confirmed the same motif (Fig. 9a and 9b).

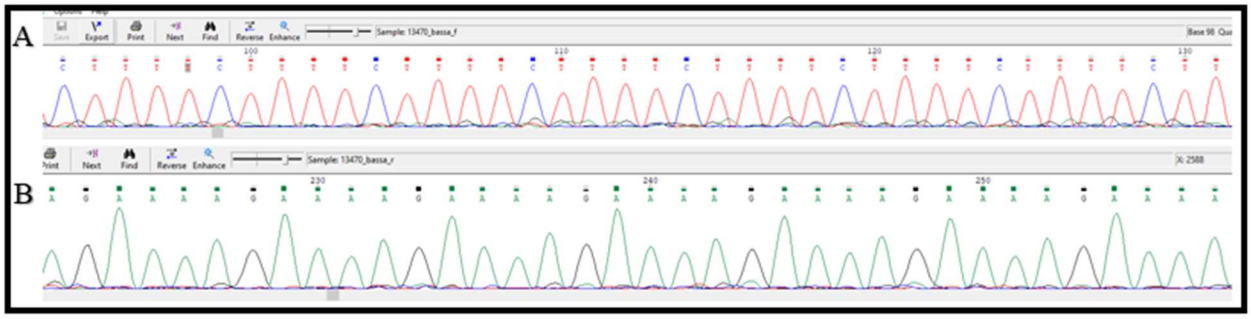

Figure 9: Sanger sequencing chromatograms of Sample ID: 13470. (A) Allele 1 – forward read. (B) Allele 1 – reverse read.

The larger allele (~3000 base pairs) was not successfully sequenced; both forward and reverse reads failed due to poor signal quality and noisy chromatograms. This failure is likely attributable to the extensive repeat length and secondary structure formation within the expanded region, as well as the presence of non-specific amplification products that interfered with sequencing quality.

- Sample ID:13819

Similar to ID: 13470, sample ID: 13819 after amplification with modified PCR protocol, presented two bands on agarose gel electrophoresis: of approximately 1000 base pairs and 3000 base pairs.

Sanger sequencing of the smaller band confirmed the presence of at least 105 AAAAG repeats in a pure configuration (Fig. 10a and 10b).

Unlike sample ID: 13470, the larger allele in this case yielded readable sequencing data, and the motif was confirmed to be pure AAAAG (Fig. 10c and 10d).

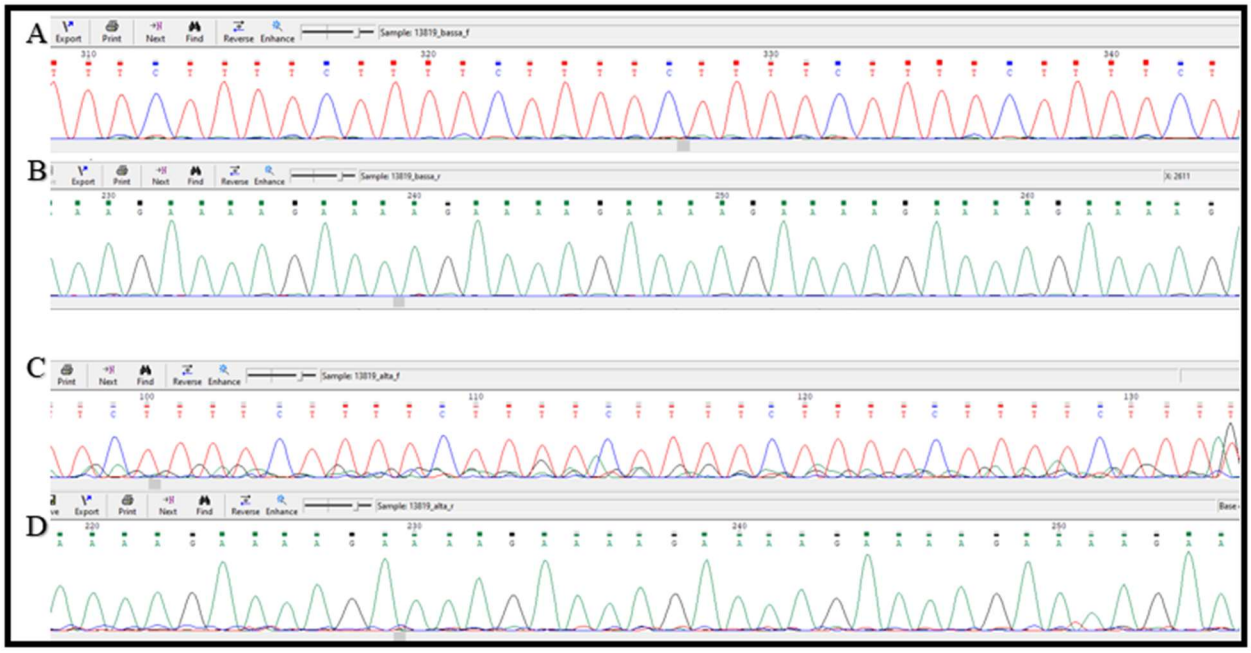

Figure 10: Sanger sequencing chromatograms of Sample ID: 13819. (A) Allele 1- Forward read. (B) Allele 1- Reverse read. (C) Allele 2 – Forward read. (D) Allele 2 – Reverse read.

- Sample ID:11052

Sample ID: 11052 exhibited two distinct bands on agarose gel electrophoresis, approximately of 700 base pairs and 4000 base pairs in size. Sanger sequencing of the lower band revealed a pure AAAAG motif, with at least 82 consecutive AAAAG repeats observed in the forward read. Reverse sequencing confirmed the same motif (Fig. 11a and 11b). The larger allele (~4000 base pairs) was not successfully sequenced.

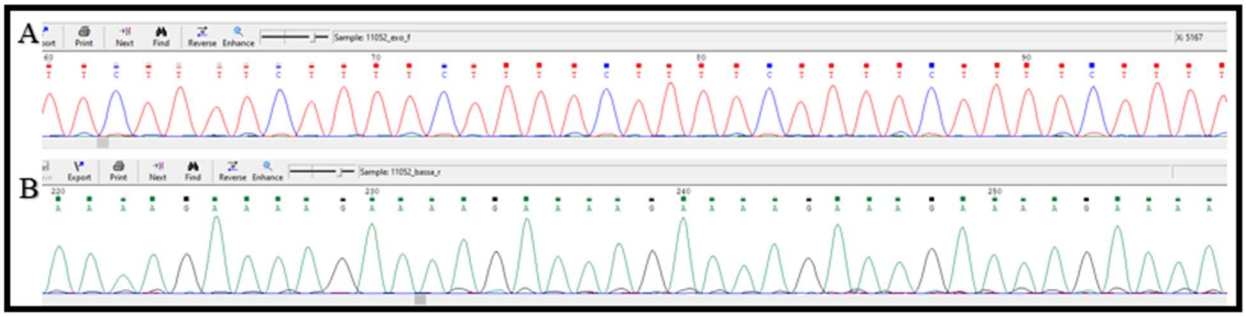

Figure 11: Sanger sequencing chromatograms of Sample ID: 11052. (A) Allele 1 – forward read. (B) Allele 1 – reverse read.

- Sample ID:13291

Sample ID: 13291 displayed a single band; thus, we proceeded with ExoSap purification of the PCR product prior to sequencing.

The reverse read showed an initial stretch of 8 AAAAG repeats, followed by over 50 AAAGG repeats. The forward read confirmed this structure, with approximately 3–4 AAAAG repeats leading into a region of ~80 AAAGG repeats, which is marked by a red line below the chromatogram (Fig. 12).

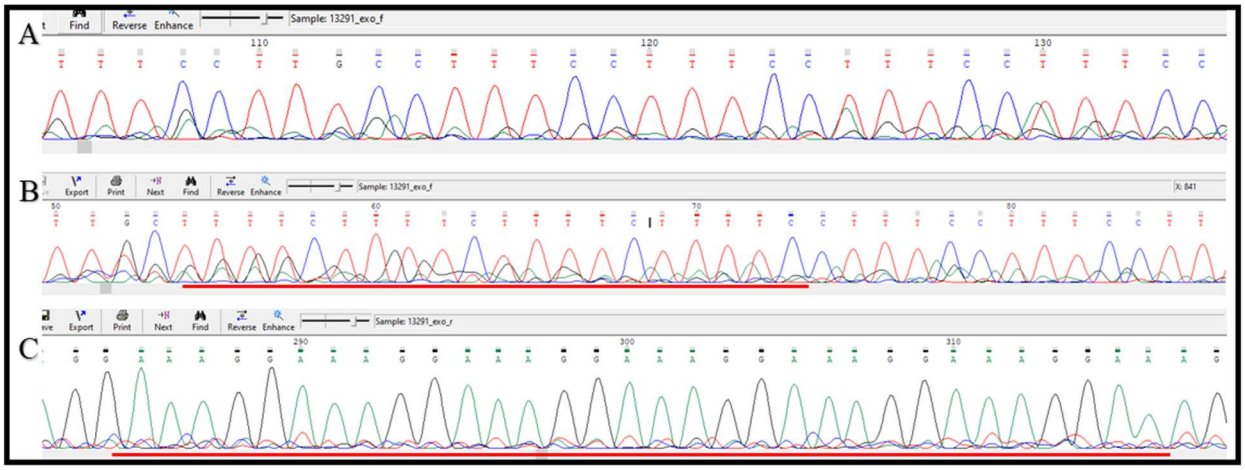

Figure 12: Sanger sequencing chromatograms of Sample ID: 13291. (A–B) Allele 1 – forward reads showing the repeat motif structure. (C) Allele 1 – reverse read confirming the sequence configuration. The red line indicates the transition between repeat motifs.

In summary, Sanger sequencing successfully characterized the repeat motif composition of most *RFC1* alleles, confirming the presence of predominantly AAAAG repeats in lower bands and mixed motifs in specific samples (Sample Id:13291). However, extremely large alleles (~3,000–4,000 bp) were uninterpretable by Sanger sequencing, likely due to a combination of factors including their excessive length, secondary structure formation, and the presence of non-specific or co-migrating PCR products, which resulted in noisy chromatograms and poor base calling. Consequently, these alleles were excluded from direct sequence analysis, and their repeat sizes were instead estimated based on gel electrophoresis migration. Despite these limitations, Sanger sequencing provided reliable motif information for all interpretable alleles, enabling accurate

classification of repeat structure and supporting downstream analysis of pathogenic versus non-pathogenic expansions.

The data obtained from Sanger Sequencing of these samples are summarized in table 8:

Sample ID	Allele type	Sizing (base pairs by gel)	Observed motif(s)	Approx./minimum repeat count observed**
13470	Allele 1	~1100 bp	AAAAG (pure)	At least 105 ru (visible from sanger sequencing)
	Allele 2	~3000 bp	Not determined by sanger sequencing	~540 ru (by gel)
13819	Allele 1	~1000 bp	AAAAG (pure)	At least 105 ru (visible from sanger sequencing)
	Allele 2	~3000 bp	AAAAG (pure)	At least 110 ru (visible from sanger sequencing)
11052	Allele 1	~750 bp	AAAAG (pure)	At least 82 ru (visible from sanger sequencing)
	Allele 2	~4000 bp	Not determined by sanger sequencing	~740 ru (by gel)
13291	Allele 1	~1300 bp	AAAAG + AAAGG (mixed)	At least 80 ru (visible from sanger sequencing)

Table 8: Table reporting the allele motif sequence according to Sanger sequencing.

*This table summarizes the results of Sanger sequencing for selected RFC1 alleles that showed large bands during earlier analyses. Each row corresponds to an individual allele and includes the following columns: The Sample ID column lists the unique identifier for each patient sample analyzed. Allele Type specifies whether the sequenced fragment was the A2 (larger allele), A1 (smaller allele), or a single visible band, helping distinguish heterozygous from homozygous or ambiguous genotypes. Sequencing quality reflects the clarity of the chromatogram: "Good" indicates reliable reads, "Moderate (noisy)" shows background interference but interpretable results, and "Unreadable" or "Poor" means the sequence could not be interpreted. Observed motif(s): reports the repeat structure within the allele—either a pure motif like AAAAG or a mixed configuration such as AAAAG + AAAGG. ****Approximate Repeat Count Observed**: Reports the estimated number of repeat units detected. **This number reflects only the repeat tract visible in the sequencing chromatogram and does not represent the full allele length**, which may extend beyond the read length limit of the capillary electrophoresis system. In cases where sequencing failed, the*

repeat size was inferred indirectly based on band size observed in gel electrophoresis. Lastly, the Interpretation column provides a classification of each allele as pathogenic, non-pathogenic, or inconclusive, based on the motif type and repeat length, in accordance with current scientific literature.

4.5. Repeat-Primed PCR (RP-PCR)

To further characterize the *RFC1* repeat expansion in sample ID: 13470, we performed repeat-primed PCR (RP-PCR) using motif-specific primers for both the common AAAAG and the pathogenic AAGGG repeat motifs. The aim was to clarify the nature of the expansion and determine whether the pathogenic motif was present.
For this analysis, we used four reverse primers targeting the motifs: one for AAAAG and three for AAGGG, as described in section 3.9 of the Materials and Methods.

Initial PCR and genotyping of sample ID:13470 revealed two bands at approximately 1000 and 3000 base pairs, suggesting the presence of large repeat expansions, potentially reaching up to 500 repeats.

When RP-PCR was performed using AAAAG-specific primers, we observed a typical sawtooth pattern (Fig. 13a), with a peak corresponding to an allele comprising 136 AAAAG repeats. This characteristic pattern, with multiple regularly spaced peaks, results from slippage during PCR amplification of repetitive sequences and is typically observed as output of RP-PCR.

In contrast, RP-PCR with AAGGG-specific primers (Fig. 13b–d) showed atypical electropherogram traces, with no regular sawtooth pattern. These results likely indicate the absence of a pure AAGGG expansion, or alternatively, the presence of an interrupted or alternate motif that does not produce the expected RP-PCR profile.

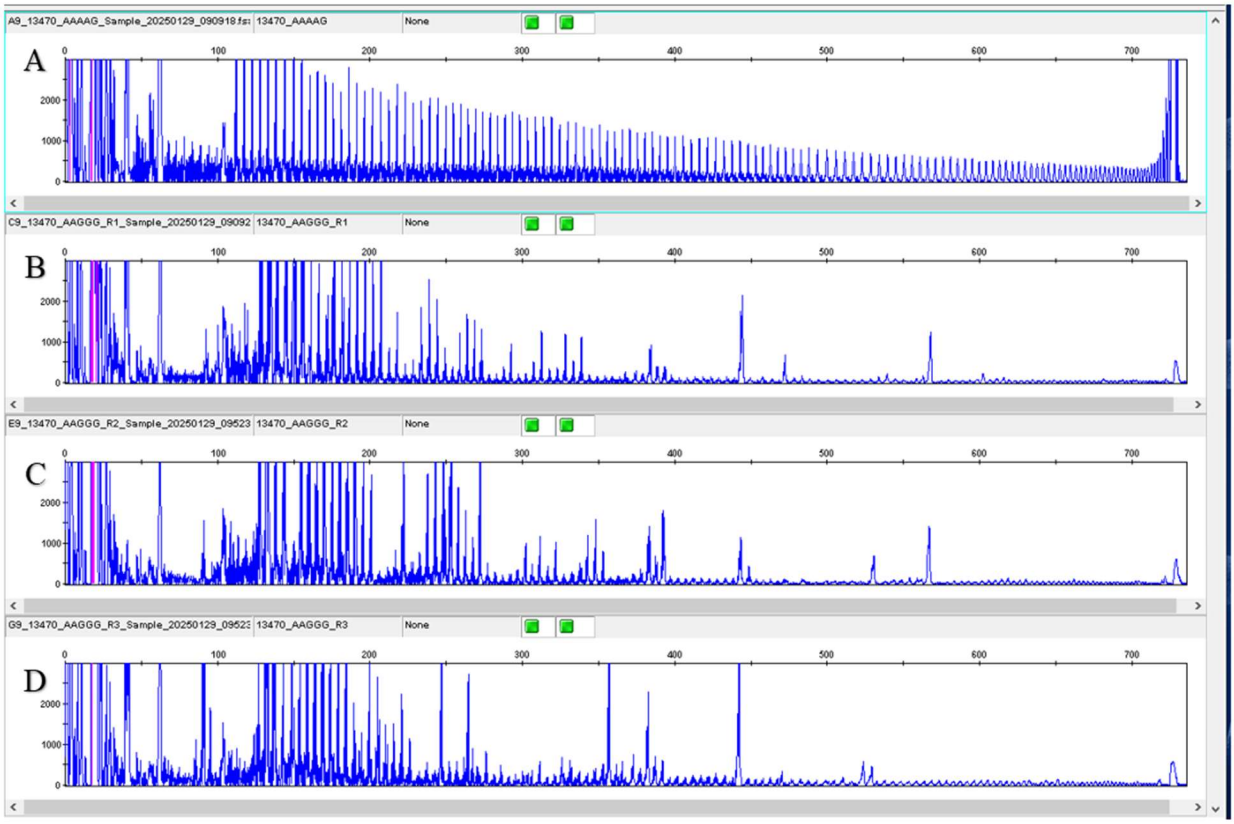

Figure 13: Electropherograms of RP-PCR reactions targeting the RFC1 locus for sample ID:13470.

A: RP-PCR targeting the AAAAG motif shows a characteristic sawtooth pattern with a peak corresponding to an allele of 136 repeats, indicative of a pure AAAAG expansion. B–D: RP-PCR reactions targeting the AAGGG motif using primers adapted to three SNPs (as described in Materials and Methods) show atypical profiles, suggestive of absence or interruption of the AAGGG repeat motif.

In summary, targeting the AAAAG motif, we observed a typical sawtooth pattern, indicative of a pure AAAAG repeat expansion, with a peak at 136 repeats. This is consistent with an *RFC1* expansion; however, this motif and the allele length are considered to be non-pathogenic.

In contrast, targeting the AAGGG motif in the presence of specific SNPs revealed no characteristic sawtooth pattern. The irregular electropherogram traces suggest the absence of a pure or uninterrupted AAGGG expansion, possibly pointing to the absence of a AAGGG motif, or the presence of an interrupted repeat.

5. DISCUSSION

Tandem repeat expansions are a fundamental class of genetic mechanisms underlying neurodegenerative diseases. Pathogenic expansions disrupt cellular homeostasis through a triad of mechanisms: toxic gain-of-function at the RNA level, repeat-associated non-AUG (RAN) translation into aggregate-prone dipeptide repeat proteins, and impaired DNA repair pathways (Ellerby, 2019; La Spada & Taylor, 2010). In Motor Neuron Disease (MND), the most prominent example is the hexanucleotide GGGGCC expansion in *C9orf72*, which stands as the most frequent genetic cause of both Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD) (DeJesus-Hernandez et al., 2011; Majounie et al., 2012a). Furthermore, intermediate-length polyglutamine repeats in genes such as *ATXN1* and *ATXN2* have been identified as significant genetic modifiers, influencing susceptibility and disease progression even when falling below traditional "pathogenic" thresholds (Elden et al., 2010b; Sproviero et al., 2017).

The discovery of biallelic intronic AAGGG repeat expansions in the *RFC1* gene as the causative factor of CANVAS (Cerebellar Ataxia, Neuropathy, and Vestibular Areflexia Syndrome) has significantly broadened the spectrum of neurological disorders associated with this locus (Cortese et al., 2019). Given that CANVAS often involves motor neuron signs, and some patients with *RFC1* expansions present with a phenotype mimicking Primary Lateral Sclerosis (PLS) or ALS (Huin et al., 2022b), it became imperative to investigate whether *RFC1* contributes to the genetic etiology of classical MND. Two previous studies investigated *RFC1* repeat expansions in ALS with conflicting results. A large-scale analysis of over 1,000 sporadic US ALS patients found no evidence of these expansions in ALS (Abramzon et al., 2021). More recently, biallelic *RFC1* expansions were found in approximately 2% prevalence of a German MND cohort (Schaub et al., 2024). These conflicting results are not so uncommon in the genetics of ALS and other disorders. There are well-documented cases of marked geographic heterogeneity in the distribution of pathogenic genetic variants, including the hexanucleotide (GGGGCC)n repeat expansion in C9orf72 associated with ALS and frontotemporal dementia.

In ALS, the frequency of the *C9orf72* expansion shows a clear geographic gradient. In populations of Northern European ancestry (e.g., Finland, the United Kingdom, Scandinavia, and North America with predominantly European descent), it represents the most common genetic cause of familial ALS (up to 30–40%) and accounts for a substantial fraction of sporadic cases (approximately 5–10%). In Southern Europe, reported frequencies are generally lower. In East Asian populations (China, Japan, Korea), the expansion is rare, typically <1–2% in sporadic ALS and uncommon even in familial forms. This pattern is largely explained by a founder effect. Overall, geographic variability in mutation frequency is a well-established phenomenon and can reflect founder effects, genetic drift, natural selection, population structure, and demographic history.

A primary objective of this study was to evaluate the reproducibility of the findings by Schaub et al., our investigation of 250 Italian patients yielded no biallelic pathogenic (AAGGG)exp expansions, although our study is sufficiently powered to rule out a prevalence of 2%.

In fact, if the true prevalence were 2%, observing zero cases in a sample of 250 individuals would be extremely unlikely. Conversely, our data is consistent with the recent large-scale international studies (Abramzon et al., 2021).

A closer clinical inspection of the positive German cases reveals they presented with "atypical" features, such as unusually slow progression, sensory involvement, or concurrent cerebellar ataxia, suggesting that *RFC1* may underlie a narrow subset of "ALS-mimics" rather than classical, rapidly progressive ALS and the reported expansions may thus represent a rare "phenocopy" rather than a consistent contributing factor. Consequently, our findings reinforce the conclusion that *RFC1* does not play a significant role in the etiology of typical ALS/FTD in the European population, highlighting the need for fine clinical characterization when considering *RFC1* screening as a diagnostic tool.

A major technical focus of our study shifted toward the systematic resolution of samples that appeared as "single bands" during initial PCR and agarose gel electrophoresis. Approximately 64% of our cohort displayed an apparently homozygous pattern. This presents a significant diagnostic hurdle known as allelic dropout.

In repeat expansion disorders, very large or G-rich expansions (like the AAGGG motif) often fail to amplify during standard flanking PCR due to the formation of stable secondary structures, such as G-quadruplexes, which inhibit DNA polymerase (Liu et al., 2024). Consequently, a patient who is compound heterozygous, carrying one small allele alongside an unamplifiable pathogenic expansion, may be misidentified as homozygous for the shorter allele due to allelic dropout.

To mitigate this, we employed a modified PCR protocol with extended elongation times, which successfully revealed "hidden" second alleles in samples like ID:13819 and ID:11052. However, we must acknowledge the limitation that standard PCR and capillary electrophoresis fragment analysis cannot definitively rule out a second expansion in all single-band cases. This ambiguity remains a critical point of concern in clinical genetics, suggesting that any "homozygous" MND patient with atypical features should ideally undergo Repeat-Primed PCR (RP-PCR) or long-read sequencing to confirm their status (Owusu & Savarese, 2023).

One of the most significant findings in our cohort was the presence of very large alleles, exceeding 3,000–4,000 base pairs, which, once characterized, were found to be non-pathogenic. We identified several cases of large AAAAG expansions, including one with more than 500 repeat units (ID: 13470). Furthermore, among the samples appearing homozygous by standard PCR and fragment analysis, we cannot rule out the possibility of large, non-amplifiable alleles. Although highly investigative, identifying and characterizing these large alleles is crucial.

In most diseases with tandem repeat expansions, such as Huntington's disease or Fragile X syndrome, the size of the expansion is the primary predictor of pathogenicity. Current knowledge indicates that for *RFC1* expansions, the motif sequence is the ultimate determinant. The identification of large heterozygous expansions suggests a broader genetic implication. The presence of such massive alleles, even with a nonpathogenic motif, could indicate significant genomic instability at the *RFC1* locus. Therefore, the potential role of these large AAAAG expansions as genetic modifiers cannot be completely ruled out. A recent study reported a higher frequency of RFC1 expansions in Idiopathic Peripheral Neuropathy (Tang Z et al., 2025).

As intermediate *ATXN2* repeats increase ALS risk without causing ataxia, expanded non-pathogenic *RFC1* repeats may subtly influence local chromatin architecture or RNA metabolism

(Elden et al., 2010b). This structural instability could create a cumulative "burden" on cellular DNA repair mechanisms, potentially modifying the onset or severity of MND in patients who already carry other primary mutations (e.g., in *SOD1* or *TARDBP*) (Nijs & Van Damme, 2024). By characterizing these large heterozygous alleles, our study suggests that the *RFC1* locus in MND patients may harbor a unique degree of structural instability that warrants further investigation as a potential risk modifier, moving beyond a simple binary assessment of pathogenicity.

The reliance on Sanger sequencing for motif characterization posed significant challenges, as the highly repetitive nature of the *RFC1* locus often led to signal degradation and 'stuttering,' particularly in alleles exceeding 3000 bp. Consequently, if a patient carried a rare pathogenic motif in a compound heterozygous state with an unamplifiable allele, our current workflow might not have detected it. This highlights that while Sanger sequencing is the gold standard for small variants, it is poorly suited for the massive, unstable expansions characteristic of *RFC1*-related. Furthermore, our workflow prioritized the two most common motifs (AAAAG and AAGGG). Recent studies have identified a "motif forest" including ACAGG and AGGGC motifs, which predominate in Asian populations and carry pathogenic potential (Tsuchiya et al., 2020; Izumi et al., 2024). The modified PCR protocol successfully resolved heterozygous expansions that were initially missed. However, the inability to sequence the full length of the 3000 bp allele in sample 13470 highlights the persistent challenge of using Sanger sequencing for massive tandem repeats (Liu et al., 2024).

Our study may have been limited by the inability to detect these rarer motifs or cryptic interruptions within the repeats. Future research should prioritize Long-Read Sequencing (LRS) (Oxford Nanopore or PacBio). LRS can span the entire repeat region, providing base-calling accuracy that resolves both the exact size and the internal motif structure, including any stabilizing interruptions that might influence pathogenicity (Ebbert et al., 2018).

The rarity of pathogenic AAGGG expansions in our Italian cohort likely reflects the ancestral background of the Northern Italian population. Allele distributions at the *RFC1* locus are known to vary significantly due to founder effects. While AAGGG expansions are relatively common in Western European CANVAS cohorts, they are exceptionally rare in other regions. The high frequency of the 13–16 repeat allele in our study represents the standard "ancestral" configuration

for this region (Cortese et al., 2019; Rafehi et al., 2019). We did not further investigate these small alleles with external references because they fall far below the pathogenic threshold; however, multi-ethnic studies remain essential to ensure that regional variations in *RFC1* motifs are not leading to under-diagnosis in non-European populations.

Despite these limitations, our study contributes important evidence to the ongoing debate regarding the role of *RFC1* in MNDs. By demonstrating the absence of pathogenic AAGGG expansions in a large, well-characterized cohort, we provide support for the conclusion that *RFC1* is unlikely to be a significant risk factor for classical ALS/FTD in the Italian population.
This has direct clinical implications: while *RFC1* testing should be considered in patients with atypical presentations, it may not be necessary in routine genetic screening of typical ALS cases. This targeted approach could streamline genetic diagnostics, reduce unnecessary testing, and provide clearer guidance to patients and families during counselling.

Looking ahead, future research should pursue three key directions.
First, the adoption of long-read sequencing will be critical to resolve large or complex expansions, enabling a definitive assessment of *RFC1* alleles beyond the limits of conventional PCR. Second, studies should explore the functional consequences of non-pathogenic expansions, particularly large AAAAG repeats, to determine whether they act as modifiers of neurodegeneration. Third, multi-center and multi-ethnic cohorts will be essential to clarify the true prevalence and phenotypic spectrum of *RFC1* expansions across different populations.
Integrating clinical, genetic, and biomarker data will further refine genotype–phenotype correlations and improve our understanding of *RFC1* in neurodegenerative disease.

In conclusion, this study provides substantial evidence that *RFC1* expansions do not contribute significantly to MNDs in the Italian cohort examined, though rare or atypical cases in other populations may still be influenced by pathogenic alleles.
The identification of large but non-pathogenic expansions highlights the complexity of interpreting repeat variation and underscores the need for advanced genomic technologies to refine classification. By situating our findings within the broader landscape of ALS genetics, we emphasize that *RFC1*-related disease represents a distinct but occasionally overlapping spectrum,

rather than a core component of ALS pathogenesis. Gaining a deeper understanding of this spectrum will be critical not only for accurate diagnosis and counselling but also for shaping future therapeutic strategies in repeat expansion disorders.

REFERENCES

1. Abramzon, Y., Dewan, R., Cortese, A., Resnick, S., Ferrucci, L., Houlden, H., & Traynor, B. J. (2021). Investigating RFC1 expansions in sporadic amyotrophic lateral sclerosis. *Journal of the Neurological Sciences*, *430*, 118061. https://doi.org/10.1016/j.jns.2021.118061
2. Akçimen, F., Ross, J. P., Bourassa, C. V., Liao, C., Rochefort, D., Gama, M. T. D., Dicaire, M.-J., Barsottini, O. G., Brais, B., Pedroso, J. L., Dion, P. A., & Rouleau, G. A. (2019). Investigation of the RFC1 Repeat Expansion in a Canadian and a Brazilian Ataxia Cohort: Identification of Novel Conformations. *Frontiers in Genetics*, *10*, 1219. https://doi.org/10.3389/fgene.2019.01219
3. Antonioni, A., Raho, E. M., Lopriore, P., Pace, A. P., Latino, R. R., Assogna, M., Mancuso, M., Gragnaniello, D., Granieri, E., Pugliatti, M., Di Lorenzo, F., & Koch, G. (2023). Frontotemporal Dementia, Where Do We Stand? A Narrative Review. *International Journal of Molecular Sciences*, *24*(14), 11732. https://doi.org/10.3390/ijms241411732
4. Arbel, M., Choudhary, K., Tfilin, O., & Kupiec, M. (2021). PCNA Loaders and Unloaders—One Ring That Rules Them All. *Genes*, *12*(11), 1812. https://doi.org/10.3390/genes12111812
5. Balendra, R., & Isaacs, A. M. (2018). C9orf72-mediated ALS and FTD: Multiple pathways to disease. *Nature Reviews. Neurology*, *14*(9), 544–558. https://doi.org/10.1038/s41582-018-0047-2
6. Bang, J., Spina, S., & Miller, B. L. (2015a). Frontotemporal dementia. *Lancet*, *386*(10004), 1672–1682. https://doi.org/10.1016/S0140-6736(15)00461-4
7. Bang, J., Spina, S., & Miller, B. L. (2015b). Frontotemporal dementia. *The Lancet*, *386*(10004), 1672–1682. https://doi.org/10.1016/S0140-6736(15)00461-4
8. Beecroft, S. J., Cortese, A., Sullivan, R., Yau, W. Y., Dyer, Z., Wu, T. Y., Mulroy, E., Pelosi, L., Rodrigues, M., Taylor, R., Mossman, S., Leadbetter, R., Cleland, J., Anderson, T., Ravenscroft, G., Laing, N. G., Houlden, H., Reilly, M. M., & Roxburgh, R. H. (2020). A Māori specific RFC1 pathogenic repeat configuration in CANVAS, likely due to a founder allele. *Brain*, *143*(9), 2673–2680. https://doi.org/10.1093/brain/awaa203

9. Benussi, A., Padovani, A., & Borroni, B. (2015). Phenotypic Heterogeneity of Monogenic Frontotemporal Dementia. *Frontiers in Aging Neuroscience*, *7*. https://doi.org/10.3389/fnagi.2015.00171
10. Bird, T., Knopman, D., VanSwieten, J., Rosso, S., Feldman, H., Tanabe, H., Graff-Raford, N., Geschwind, D., Verpillat, P., & Hutton, M. (2003). Epidemiology and genetics of frontotemporal dementia/Pick's disease. *Annals of Neurology*, *54*(S5), S29–S31. https://doi.org/10.1002/ana.10572
11. Bradley, W. G. (2009). Updates on amyotrophic lateral sclerosis: Improving patient care. *Annals of Neurology*, *65*(S1), S1–S2. https://doi.org/10.1002/ana.21546
12. Brown, R. H., & Al-Chalabi, A. (2017). Amyotrophic Lateral Sclerosis. *The New England Journal of Medicine*, *377*(2), 162–172. https://doi.org/10.1056/NEJMra1603471
13. Calezis, C., Bonello-Palot, N., Verschueren, A., Azulay, J.-P., Fortanier, E., Grapperon, A.-M., Kouton, L., Gallard, J., Salort-Campana, E., Attarian, S., & Delmont, E. (2024). Electrophysiological features of the peripheral neuropathy in patients with pathologic biallelic RFC1 repeat expansions. *Muscle & Nerve*, *70*(5), 1046–1052. https://doi.org/10.1002/mus.28257
14. Chia, R., Chiò, A., & Traynor, B. J. (2018). Novel genes associated with amyotrophic lateral sclerosis: Diagnostic and clinical implications. *The Lancet Neurology*, *17*(1), 94–102. https://doi.org/10.1016/S1474-4422(17)30401-5
15. Chio, A., Logroscino, G., Hardiman, O., Swingler, R., Mitchell, D., Beghi, E., & Traynor, B. G. (2009). Prognostic factors in ALS: A critical review. *Amyotrophic Lateral Sclerosis : Official Publication of the World Federation of Neurology Research Group on Motor Neuron Diseases*, *10*(5–6), 310–323. https://doi.org/10.3109/17482960802566824
16. Colucci, F., Di Bella, D., Pisciotta, C., Sarto, E., Gualandi, F., Neri, M., Ferlini, A., Contaldi, E., Pugliatti, M., Pareyson, D., & Sensi, M. (2022). Beyond canvas: Behavioral onset of rfc1-expansion disease in an Italian family-causal or casual? *Neurological Sciences: Official Journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*, *43*(8), 5095–5098. https://doi.org/10.1007/s10072-022-06137-1
17. Cooper-Knock, J., Shaw, P. J., & Kirby, J. (2014). The widening spectrum of C9ORF72-related disease; genotype/phenotype correlations and potential modifiers of clinical

phenotype. *Acta Neuropathologica*, *127*(3), 333–345. https://doi.org/10.1007/s00401-014-1251-9

18. Cortese, A., Simone, R., Sullivan, R., Vandrovcova, J., Tariq, H., Yau, W. Y., Humphrey, J., Jaunmuktane, Z., Sivakumar, P., Polke, J., Ilyas, M., Tribollet, E., Tomaselli, P. J., Devigili, G., Callegari, I., Versino, M., Salpietro, V., Efthymiou, S., Kaski, D., … Houlden, H. (2019). Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nature Genetics*, *51*(4), 649–658. https://doi.org/10.1038/s41588-019-0372-4
19. Cortese, A., Tozza, S., Yau, W. Y., Rossi, S., Beecroft, S. J., Jaunmuktane, Z., Dyer, Z., Ravenscroft, G., Lamont, P. J., Mossman, S., Chancellor, A., Maisonobe, T., Pereon, Y., Cauquil, C., Colnaghi, S., Mallucci, G., Curro, R., Tomaselli, P. J., Thomas-Black, G., … Reilly, M. M. (2020). Cerebellar ataxia, neuropathy, vestibular areflexia syndrome due to RFC1 repeat expansion. *Brain*, *143*(2), 480–490. https://doi.org/10.1093/brain/awz418
20. Davies, K., Szmulewicz, D. J., Corben, L. A., Delatycki, M., & Lockhart, P. J. (2022). RFC1-Related Disease. *Neurology: Genetics*, *8*(5), e200016. https://doi.org/10.1212/NXG.0000000000200016
21. DeJesus-Hernandez, M., Mackenzie, I. R., Boeve, B. F., Boxer, A. L., Baker, M., Rutherford, N. J., Nicholson, A. M., Finch, N. A., Flynn, H., Adamson, J., Kouri, N., Wojtas, A., Sengdy, P., Hsiung, G.-Y. R., Karydas, A., Seeley, W. W., Josephs, K. A., Coppola, G., Geschwind, D. H., … Rademakers, R. (2011). Expanded GGGGCC Hexanucleotide Repeat in Noncoding Region of C9ORF72 Causes Chromosome 9p-Linked FTD and ALS. *Neuron*, *72*(2), 245–256. https://doi.org/10.1016/j.neuron.2011.09.011
22. Dewan, R., Chia, R., Ding, J., Hickman, R. A., Stein, T. D., Abramzon, Y., Ahmed, S., Sabir, M. S., Portley, M. K., Tucci, A., Ibáñez, K., Shankaracharya, F. N. U., Keagle, P., Rossi, G., Caroppo, P., Tagliavini, F., Waldo, M. L., Johansson, P. M., Nilsson, C. F., … Vaughan, J. (2021). Pathogenic Huntingtin Repeat Expansions in Patients with Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron*, *109*(3), 448-460.e4. https://doi.org/10.1016/j.neuron.2020.11.005
23. Dominik, N., Magri, S., Currò, R., Abati, E., Facchini, S., Corbetta, M., Macpherson, H., Di Bella, D., Sarto, E., Stevanovski, I., Chintalaphani, S. R., Akcimen, F., Manini, A.,

Vegezzi, E., Quartesan, I., Montgomery, K.-A., Pirota, V., Crespan, E., Perini, C., … Cortese, A. (2023). Normal and pathogenic variation of RFC1 repeat expansions: Implications for clinical diagnosis. *Brain*, *146*(12), 5060–5069. https://doi.org/10.1093/brain/awad240

24. Ebbert, M. T. W., Farrugia, S. L., Sens, J. P., Jansen-West, K., Gendron, T. F., Prudencio, M., McLaughlin, I. J., Bowman, B., Seetin, M., DeJesus-Hernandez, M., Jackson, J., Brown, P. H., Dickson, D. W., van Blitterswijk, M., Rademakers, R., Petrucelli, L., & Fryer, J. D. (2018). Long-read sequencing across the C9orf72 'GGGGCC' repeat expansion: Implications for clinical use and genetic discovery efforts in human disease. *Molecular Neurodegeneration*, *13*(1), 46. https://doi.org/10.1186/s13024-018-0274-4
25. Elden, A. C., Kim, H.-J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M. M., Padmanabhan, A., Clay-Falcone, D., McCluskey, L., Elman, L., Juhr, D., Gruber, P. J., Rüb, U., Auburger, G., Trojanowski, J. Q., … Gitler, A. D. (2010a). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*, *466*(7310), 1069–1075. https://doi.org/10.1038/nature09320
26. Elden, A. C., Kim, H.-J., Hart, M. P., Chen-Plotkin, A. S., Johnson, B. S., Fang, X., Armakola, M., Geser, F., Greene, R., Lu, M. M., Padmanabhan, A., Clay-Falcone, D., McCluskey, L., Elman, L., Juhr, D., Gruber, P. J., Rüb, U., Auburger, G., Trojanowski, J. Q., … Gitler, A. D. (2010b). Ataxin-2 intermediate-length polyglutamine expansions are associated with increased risk for ALS. *Nature*, *466*(7310), 1069–1075. https://doi.org/10.1038/nature09320
27. Ellerby, L. M. (2019). Repeat Expansion Disorders: Mechanisms and Therapeutics. *Neurotherapeutics*, *16*(4), 924–927. https://doi.org/10.1007/s13311-019-00823-3
28. Feldman, E. L., Goutman, S. A., Petri, S., Mazzini, L., Savelieff, M. G., Shaw, P. J., & Sobue, G. (2022). Amyotrophic lateral sclerosis. *The Lancet*, *400*(10360), 1363–1380. https://doi.org/10.1016/S0140-6736(22)01272-7
29. Fenoglio, C., Scarpini, E., Serpente, M., & Galimberti, D. (2018). Role of Genetics and Epigenetics in the Pathogenesis of Alzheimer's Disease and Frontotemporal Dementia. *Journal of Alzheimer's Disease*, *62*(3), 913–932. https://doi.org/10.3233/JAD-170702

30. Fontana, A., Marin, B., Luna, J., Beghi, E., Logroscino, G., Boumédiene, F., Preux, P.-M., Couratier, P., & Copetti, M. (2021). Time-trend evolution and determinants of sex ratio in Amyotrophic Lateral Sclerosis: A dose–response meta-analysis. *Journal of Neurology*, *268*(8), 2973–2984. https://doi.org/10.1007/s00415-021-10464-2
31. Gaubitz, C., Liu, X., Magrino, J., Stone, N. P., Landeck, J., Hedglin, M., & Kelch, B. A. (2020). Structure of the human clamp loader reveals an autoinhibited conformation of a substrate-bound AAA+ switch. *Proceedings of the National Academy of Sciences*, *117*(38), 23571–23580. (world). https://doi.org/10.1073/pnas.2007437117
32. Gisatulin, M., Dobricic, V., Zühlke, C., Hellenbroich, Y., Tadic, V., Münchau, A., Isenhardt, K., Bürk, K., Bahlo, M., Lockhart, P. J., Lohmann, K., Helmchen, C., & Brüggemann, N. (2020). Clinical spectrum of the pentanucleotide repeat expansion in the RFC1 gene in ataxia syndromes. *Neurology*, *95*(21), e2912–e2923. https://doi.org/10.1212/WNL.0000000000010744
33. Gitcho, M. A., Baloh, R. H., Chakraverty, S., Mayo, K., Norton, J. B., Levitch, D., Hatanpaa, K. J., White, C. L., 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. *Annals of Neurology*, *63*(4), 535–538. https://doi.org/10.1002/ana.21344
34. Goutman, S. A., Hardiman, O., Al-Chalabi, A., Chió, A., Savelieff, M. G., Kiernan, M. C., & Feldman, E. L. (2022). Recent advances in the diagnosis and prognosis of ALS. *The Lancet. Neurology*, *21*(5), 480–493. https://doi.org/10.1016/S1474-4422(21)00465-8
35. Grad, L. I., Rouleau, G. A., Ravits, J., & Cashman, N. R. (2017). Clinical Spectrum of Amyotrophic Lateral Sclerosis (ALS). *Cold Spring Harbor Perspectives in Medicine*, *7*(8), a024117. https://doi.org/10.1101/cshperspect.a024117
36. Graff-Radford, N., & Woodruff, B. (2007). Frontotemporal Dementia. *Seminars in Neurology*, *27*(1), 048–057. https://doi.org/10.1055/s-2006-956755
37. Hardiman, O., Al-Chalabi, A., Chio, A., Corr, E. M., Logroscino, G., Robberecht, W., Shaw, P. J., Simmons, Z., & van den Berg, L. H. (2017). Amyotrophic lateral sclerosis. *Nature Reviews Disease Primers*, *3*(1), 17071. https://doi.org/10.1038/nrdp.2017.71
38. Hogan, D. B., Jetté, N., Fiest, K. M., Roberts, J. I., Pearson, D., Smith, E. E., Roach, P., Kirk, A., Pringsheim, T., & Maxwell, C. J. (2016). The Prevalence and Incidence of

Frontotemporal Dementia: A Systematic Review. *Canadian Journal of Neurological Sciences / Journal Canadien Des Sciences Neurologiques*, *43*(S1), S96–S109. https://doi.org/10.1017/cjn.2016.25

39. Huin, V., Coarelli, G., Guemy, C., Boluda, S., Debs, R., Mochel, F., Stojkovic, T., Grabli, D., Maisonobe, T., Gaymard, B., Lenglet, T., Tard, C., Davion, J. B., Sablonnière, B., Monin, M. L., Ewenczyk, C., Viala, K., Charles, P., Le Ber, I., … Durr, A. (2022). Motor neuron pathology in CANVAS due to RFC1 expansions. *Brain*, *145*(6), 2121–2132. https://doi.org/10.1093/brain/awab449
40. Ishiura, H., Shibata, S., Yoshimura, J., Suzuki, Y., Qu, W., Doi, K., Almansour, M. A., Kikuchi, J. K., Taira, M., Mitsui, J., Takahashi, Y., Ichikawa, Y., Mano, T., Iwata, A., Harigaya, Y., Matsukawa, M. K., Matsukawa, T., Tanaka, M., Shirota, Y., … Tsuji, S. (2019). Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. *Nature Genetics*, *51*(8), 1222–1232. https://doi.org/10.1038/s41588-019-0458-z
41. La Spada, A. R., & Taylor, J. P. (2010). Repeat expansion disease: Progress and puzzles in disease pathogenesis. *Nature Reviews. Genetics*, *11*(4), 247–258. https://doi.org/10.1038/nrg2748
42. Lagier-Tourenne, C., & Cleveland, D. W. (2009). Rethinking ALS: The FUS about TDP-43. *Cell*, *136*(6), 1001–1004. https://doi.org/10.1016/j.cell.2009.03.006
43. Lattante, S., Pomponi, M. G., Conte, A., Marangi, G., Bisogni, G., Patanella, A. K., Meleo, E., Lunetta, C., Riva, N., Mosca, L., Carrera, P., Bee, M., Zollino, M., & Sabatelli, M. (2018). ATXN1 intermediate-length polyglutamine expansions are associated with amyotrophic lateral sclerosis. *Neurobiology of Aging*, *64*, 157.e1-157.e5. https://doi.org/10.1016/j.neurobiolaging.2017.11.011
44. Lee, K., & Park, S. H. (2020). Eukaryotic clamp loaders and unloaders in the maintenance of genome stability. *Experimental & Molecular Medicine*, *52*(12), 1948–1958. https://doi.org/10.1038/s12276-020-00533-3
45. Logroscino, G., Piccininni, M., Marin, B., Nichols, E., Abd-Allah, F., Abdelalim, A., Alahdab, F., Asgedom, S. W., Awasthi, A., Chaiah, Y., Daryani, A., Do, H. P., Dubey, M., Elbaz, A., Eskandarieh, S., Farhadi, F., Farzadfar, F., Fereshtehnejad, S.-M., Fernandes, E., … Murray, C. J. L. (2018). Global, regional, and national burden of motor neuron

diseases 1990–2016: A systematic analysis for the Global Burden of Disease Study 2016. *The Lancet Neurology*, *17*(12), 1083–1097. https://doi.org/10.1016/S1474-4422(18)30404-6

46. Lomen-Hoerth, C. (2011). Clinical Phenomenology and Neuroimaging Correlates in ALS-FTD. *Journal of Molecular Neuroscience*, *45*(3), 656–662. https://doi.org/10.1007/s12031-011-9636-x
47. Magy, L., Chazelas, P., Richard, L., Deschamps, N., Frachet, S., Vallat, J.-M., Magdelaine, C., Favreau, F., Bessaguet, F., Lia, A.-S., & Duchesne, M. (2022). Early Diagnosis in Cerebellar Ataxia, Neuropathy, Vestibular Areflexia Syndrome (CANVAS) by Focusing on Major Clinical Clues: Beyond Ataxia and Vestibular Impairment. *Biomedicines*, *10*(8), 2046. https://doi.org/10.3390/biomedicines10082046
48. Majounie, E., Renton, A. E., Mok, K., Dopper, E. G., Waite, A., Rollinson, S., Chiò, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., Swieten, J. C. van, Abramzon, Y., Johnson, J. O., Sendtner, M., Pamphlett, R., Orrell, R. W., Mead, S., Sidle, K. C., Houlden, H., … Traynor, B. J. (2012a). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *The Lancet Neurology*, *11*(4), 323–330. https://doi.org/10.1016/S1474-4422(12)70043-1
49. Majounie, E., Renton, A. E., Mok, K., Dopper, E. G., Waite, A., Rollinson, S., Chiò, A., Restagno, G., Nicolaou, N., Simon-Sanchez, J., Van Swieten, J. C., Abramzon, Y., Johnson, J. O., Sendtner, M., Pamphlett, R., Orrell, R. W., Mead, S., Sidle, K. C., Houlden, H., … Traynor, B. J. (2012b). Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: A cross-sectional study. *The Lancet Neurology*, *11*(4), 323–330. https://doi.org/10.1016/S1474-4422(12)70043-1
50. Masrori, P., & Van Damme, P. (2020). Amyotrophic lateral sclerosis: A clinical review. *European Journal of Neurology*, *27*(10), 1918–1929. https://doi.org/10.1111/ene.14393
51. Mastrocola, A. S., Kim, S. H., Trinh, A. T., Rodenkirch, L. A., & Tibbetts, R. S. (2013). The RNA-binding Protein Fused in Sarcoma (FUS) Functions Downstream of Poly(ADP-ribose) Polymerase (PARP) in Response to DNA Damage. *Journal of Biological Chemistry*, *288*(34), 24731–24741. https://doi.org/10.1074/jbc.M113.497974

52. McCord, J. M., & Fridovich, I. (1969). Superoxide dismutase. An enzymic function for erythrocuprein (hemocuprein). *The Journal of Biological Chemistry*, *244*(22), 6049–6055.
53. Mejzini, R., Flynn, L. L., Pitout, I. L., Fletcher, S., Wilton, S. D., & Akkari, P. A. (2019a). ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? *Frontiers in Neuroscience*, *13*, 1310. https://doi.org/10.3389/fnins.2019.01310
54. Mejzini, R., Flynn, L. L., Pitout, I. L., Fletcher, S., Wilton, S. D., & Akkari, P. A. (2019b). ALS Genetics, Mechanisms, and Therapeutics: Where Are We Now? *Frontiers in Neuroscience*, *13*, 1310. https://doi.org/10.3389/fnins.2019.01310
55. Miyatake, S., Koshimizu, E., Fujita, A., Doi, H., Okubo, M., Wada, T., Hamanaka, K., Ueda, N., Kishida, H., Minase, G., Matsuno, A., Kodaira, M., Ogata, K., Kato, R., Sugiyama, A., Sasaki, A., Miyama, T., Satoh, M., Uchiyama, Y., … Matsumoto, N. (2022). Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. *Npj Genomic Medicine*, *7*(1), 62. https://doi.org/10.1038/s41525-022-00331-y
56. Neary, D., Snowden, J., & Mann, D. (2005). Frontotemporal dementia. *The Lancet Neurology*, *4*(11), 771–780. https://doi.org/10.1016/S1474-4422(05)70223-4
57. Nijs, M., & Van Damme, P. (2024). The genetics of amyotrophic lateral sclerosis. *Current Opinion in Neurology*, *37*(5), 560–569. https://doi.org/10.1097/WCO.0000000000001294
58. Olney, N. T., Spina, S., & Miller, B. L. (2017). Frontotemporal Dementia. *Neurologic Clinics*, *35*(2), 339–374. https://doi.org/10.1016/j.ncl.2017.01.008
59. Olszewska, D. A., Lonergan, R., Fallon, E. M., & Lynch, T. (2016). Genetics of Frontotemporal Dementia. *Current Neurology and Neuroscience Reports*, *16*(12), 107. https://doi.org/10.1007/s11910-016-0707-9
60. Onyike, C. U., & Diehl-Schmid, J. (2013). The epidemiology of frontotemporal dementia. *International Review of Psychiatry*, *25*(2), 130–137. https://doi.org/10.3109/09540261.2013.776523
61. Owusu, R., & Savarese, M. (2023). Long-read sequencing improves diagnostic rate in neuromuscular disorders. *Acta Myologica*, *42*(4), 123–128. https://doi.org/10.36185/2532-1900-394
62. Paulson, H. (2018). Repeat expansion diseases. *Handbook of Clinical Neurology*, *147*, 105–123. https://doi.org/10.1016/B978-0-444-63233-3.00009-9

63. Rafehi, H., Szmulewicz, D. J., Bennett, M. F., Sobreira, N. L. M., Pope, K., Smith, K. R., Gillies, G., Diakumis, P., Dolzhenko, E., Eberle, M. A., Barcina, M. G., Breen, D. P., Chancellor, A. M., Cremer, P. D., Delatycki, M. B., Fogel, B. L., Hackett, A., Halmagyi, G. M., Kapetanovic, S., … Lockhart, P. J. (2019). Bioinformatics-Based Identification of Expanded Repeats: A Non-reference Intronic Pentamer Expansion in RFC1 Causes CANVAS. *American Journal of Human Genetics*, *105*(1), 151–165. https://doi.org/10.1016/j.ajhg.2019.05.016
64. Rascovsky, K., Hodges, J. R., Knopman, D., Mendez, M. F., Kramer, J. H., Neuhaus, J., van Swieten, J. C., Seelaar, H., Dopper, E. G. P., Onyike, C. U., Hillis, A. E., Josephs, K. A., Boeve, B. F., Kertesz, A., Seeley, W. W., Rankin, K. P., Johnson, J. K., Gorno-Tempini, M.-L., Rosen, H., … Miller, B. L. (2011). Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain: A Journal of Neurology*, *134*(Pt 9), 2456–2477. https://doi.org/10.1093/brain/awr179
65. Ratti, A., & Buratti, E. (2016). Physiological functions and pathobiology of TDP-43 and FUS/TLS proteins. *Journal of Neurochemistry*, *138*, 95–111. https://doi.org/10.1111/jnc.13625
66. Reyes-Leiva, D., Aldecoa, I., Gelpi, E., & Rojas-García, R. (2022). Motor neuron involvement expands the neuropathological phenotype of late-onset ataxia in RFC1 mutation (CANVAS). *Brain Pathology*, *32*(4), e13051. https://doi.org/10.1111/bpa.13051
67. Schaub, A., Erdmann, H., Scholz, V., Timmer, M., Cordts, I., Günther, R., Reilich, P., Abicht, A., & Schöberl, F. (2024). Analysis and occurrence of biallelic pathogenic repeat expansions in RFC1 in a German cohort of patients with a main clinical phenotype of motor neuron disease. *Journal of Neurology*, *271*(9), 5804–5812. https://doi.org/10.1007/s00415-024-12519-6
68. Schrecker, M., Castaneda, J. C., Devbhandari, S., Kumar, C., Remus, D., & Hite, R. K. (2022). Multistep loading of a DNA sliding clamp onto DNA by replication factor C. *eLife*, *11*, e78253. https://doi.org/10.7554/eLife.78253
69. Smeyers, J., Banchi, E.-G., & Latouche, M. (2021). C9ORF72: What It Is, What It Does, and Why It Matters. *Frontiers in Cellular Neuroscience*, *15*. https://doi.org/10.3389/fncel.2021.661447

70. Spencer, P. S., Palmer, V. S., Kisby, G. E., Lagrange, E., Horowitz, B. Z., Valdes Angues, R., Reis, J., Vernoux, J.-P., Raoul, C., & Camu, W. (2023). Early-onset, conjugal, twin-discordant, and clusters of sporadic ALS: Pathway to discovery of etiology via lifetime exposome research. *Frontiers in Neuroscience*, *17*, 1005096. https://doi.org/10.3389/fnins.2023.1005096
71. Sproviero, W., Shatunov, A., Stahl, D., Shoai, M., van Rheenen, W., Jones, A. R., Al-Sarraj, S., Andersen, P. M., Bonini, N. M., Conforti, F. L., Van Damme, P., Daoud, H., Del Mar Amador, M., Fogh, I., Forzan, M., Gaastra, B., Gellera, C., Gitler, A. D., Hardy, J., … Al-Chalabi, A. (2017). ATXN2 trinucleotide repeat length correlates with risk of ALS. *Neurobiology of Aging*, *51*, 178.e1-178.e9. https://doi.org/10.1016/j.neurobiolaging.2016.11.010
72. Strong, M. J., Abrahams, S., Goldstein, L. H., Woolley, S., Mclaughlin, P., Snowden, J., Mioshi, E., Roberts-South, A., Benatar, M., HortobáGyi, T., Rosenfeld, J., Silani, V., Ince, P. G., & Turner, M. R. (2017). Amyotrophic lateral sclerosis - frontotemporal spectrum disorder (ALS-FTSD): Revised diagnostic criteria. *Amyotrophic Lateral Sclerosis & Frontotemporal Degeneration*, *18*(3–4), 153–174. https://doi.org/10.1080/21678421.2016.1267768
73. Sullivan, R., Yau, W. Y., Chelban, V., Rossi, S., Dominik, N., O'Connor, E., Hardy, J., Wood, N., Cortese, A., & Houlden, H. (2021). RFC1-related ataxia is a mimic of early multiple system atrophy. *Journal of Neurology, Neurosurgery, and Psychiatry*, *92*(4), 444–446. https://doi.org/10.1136/jnnp-2020-325092
74. Talbott, E. O., Malek, A. M., & Lacomis, D. (2016). The epidemiology of amyotrophic lateral sclerosis. In *Handbook of Clinical Neurology* (Vol. 138, pp. 225–238). Elsevier. https://doi.org/10.1016/B978-0-12-802973-2.00013-6
75. Tang, Z., Ovunc, S. S., Mehinovic, E., Thomas, S., Ulibarri, J., Li, Z., Baldridge, D., Cruchaga, C., Johnson, M., Milbrandt, J., Callaghan, B., PNRR Study Group, Höke, A., Todd, P. K., & Jin, S. C. (2025). Heterozygous and Homozygous RFC1 AAGGG Repeat Expansions are Common in Idiopathic Peripheral Neuropathy. *medRxiv: The Preprint Server for Health Sciences*, 2025.04.18.25325809. https://doi.org/10.1101/2025.04.18.25325809

76. Tollervey, J. R., Curk, T., Rogelj, B., Briese, M., Cereda, M., Kayikci, M., König, J., Hortobágyi, T., Nishimura, A. L., Župunski, V., Patani, R., Chandran, S., Rot, G., Zupan, B., Shaw, C. E., & Ule, J. (2011). Characterizing the RNA targets and position-dependent splicing regulation by TDP-43. *Nature Neuroscience*, *14*(4), 452–458. https://doi.org/10.1038/nn.2778
77. Wagner, M., Lorenz, G., Volk, A. E., Brunet, T., Edbauer, D., Berutti, R., Zhao, C., Anderl-Straub, S., Bertram, L., Danek, A., Deschauer, M., Dill, V., Fassbender, K., Fliessbach, K., Götze, K. S., Jahn, H., Kornhuber, J., Landwehrmeyer, B., Lauer, M., … Winkelmann, J. (2021). Clinico-genetic findings in 509 frontotemporal dementia patients. *Molecular Psychiatry*, *26*(10), 5824–5832. https://doi.org/10.1038/s41380-021-01271-2
78. Wang, X., Zeng, P., Fang, Y., Zhang, T., & Tian, Q. (2021). Progranulin in neurodegenerative dementia. *Journal of Neurochemistry*, *158*(2), 119–137. https://doi.org/10.1111/jnc.15378
79. Watanabe, K., Nakashima, M., Wakatsuki, R., Bunai, T., Ouchi, Y., Nakamura, T., Miyajima, H., & Saitsu, H. (2022). Cognitive Impairment in a Complex Family With AAGGG and ACAGG Repeat Expansions in RFC1 Detected by ExpansionHunter Denovo. *Neurology: Genetics*, *8*(3), e682. https://doi.org/10.1212/NXG.0000000000000682
80. Wu, T. Y., Taylor, J. M., Kilfoyle, D. H., Smith, A. D., McGuinness, B. J., Simpson, M. P., Walker, E. B., Bergin, P. S., Cleland, J. C., Hutchinson, D. O., Anderson, N. E., Snow, B. J., Anderson, T. J., Paermentier, L. A. F., Cutfield, N. J., Chancellor, A. M., Mossman, S. S., & Roxburgh, R. H. (2014). Autonomic dysfunction is a major feature of cerebellar ataxia, neuropathy, vestibular areflexia 'CANVAS' syndrome. *Brain*, *137*(10), 2649–2656. https://doi.org/10.1093/brain/awu196
81. Zimmermann, M., Mengel, D., Raupach, K., Haack, T., Neumann, M., & Synofzik, M. (2025). Frequency and neuropathology of HTT repeat expansions in FTD/ALS: Co-existence rather than causation. *Journal of Neurology*, *272*(1), 58. https://doi.org/10.1007/s00415-024-12822-2
82. Zou, Z.-Y., Zhou, Z.-R., Che, C.-H., Liu, C.-Y., He, R.-L., & Huang, H.-P. (2017). Genetic epidemiology of amyotrophic lateral sclerosis: A systematic review and meta-analysis.

Journal of Neurology, Neurosurgery & Psychiatry, *88*(7), 540–549. https://doi.org/10.1136/jnnp-2016-315018