UNIVERSITÄT ZU LÜBECK

Aus dem Institut für Neurogenetik
der Universität zu Lübeck
Direktorin: Prof. Dr. Christine Klein

"Long-read sequencing on repeat expansions"

Inauguraldissertation
zur
Erlangung der Doktorwürde
der Universität zu Lübeck

Aus der Sektion Naturwissenschaften

Vorgelegt von
Joshua Laß
aus Lübeck

Lübeck, 2025

1. Berichterstatterin: Prof. Dr. Joanne Trinh

2. Berichterstatter: Prof. Dr. Mattias Heinrich

Tag der mündlichen Prüfung: 05.03.2026

Zum Druck genehmigt. Lübeck, den 07.04.2026

Abstract

Technological advances, particularly long-read sequencing, have led to the discovery of over 25 repeat expansion disorders (REDs) in the past decade. Repeat expansions are caused by repetitions of 1 to over 1000 base pairs, classified into three main categories: microsatellites (e.g., CAG expansion in *HTT* in Huntington's disease), minisatellites (e.g., 99 bp expansion in *PLIN4*), and macrosatellites (e.g., D4Z4 in FSHD). More than one million such repeats exist in the human genome, often introduced by transposable elements such as LINEs and SINEs. Most repeats are benign, but some exceed disease-specific pathogenic thresholds, like 40 CAG repeats in the gene *HTT*. More than 80% of the REDs are associated with neurological diseases.

Repeat tract characterization can reveal interruptions caused by deletions, insertions, or mismatches. The interruptions can alter the motif, length, and pathogenic potential of repeat expansions. Such interruptions can stabilize the repeat tract during DNA replication, reduce expansion risk, influence RNA structure and toxicity, or modulate epigenetic regulation. They can often lead to milder phenotypes or delayed age at onset. Interruptions can also impact inheritance and intergenerational stability. Overall, repeat interruptions can act as genetic modifiers, shaping the stability, pathogenicity, and transmission patterns of repeat expansion disorders, making them important for diagnosis and therapy development.

This thesis hypothesizes that mosaic genetic variability occurs in repeat expansion disorders and can act as a genetic modifier.

In the first objective, the *TAF1* SVA retrotransposon in X-linked dystonia parkinsonism (XDP) was investigated using long-read sequencing technologies to determine the mosaic genetic variability, repeat length, and interruptions. The second objective was the assessment of the stability of mosaic modifiers across familial generations in XDP. In the third objective, the recently discovered repeat expansion in *FGF14* was investigated to determine if mosaic variants are present in *FGF14*-related ataxias. The last objective is to determine whether the *FGF14* repeat expansion is associated with multiple system atrophy (MSA).

In this thesis, the advantages of long-read sequencing on the investigation of repeat expansion have enabled the direct detection of genetic variability, epigenetic information, and the characterization of repeat expansions. Long-read sequencing provides specificity in detecting repeat numbers in XDP and SCA27B.

Furthermore, the technology has even enabled the discovery of novel somatic interruptions where the frequency of mosaic interruptions is associated with the stability of the repeat length in XDP. The interruptions in XDP affect the transmission of the repeat length across generations. A higher interruption frequency stabilizes the repeat length across generations, which shows a protective effect of the interruptions in repeat expansion. In SCA27B, the results led to a new categorization of the patient's affection status using the pure GAA length without interruptions. The frequency of the *FGF14* repeat expansion was slightly increased in patients with MSA compared to healthy individuals.

In conclusion, long-read sequencing has improved the detection and characterization of repeat expansions. Repeat expansions may be present in known neurological diseases as a modifier. Repeat interruptions can stabilize the repeat length and are important to consider in diagnosing repeat expansion diseases.

Zusammenfassung

Technologische Fortschritte, insbesondere die Long-Read-Sequenzierung, haben in den letzten zehn Jahren zur Entdeckung von über 25 neuen Repeat-Expansion (RE) Erkrankungen geführt. RE werden durch Wiederholungen von 1 bis über 1000 Basenpaaren verursacht und in drei Hauptkategorien eingeteilt: Mikrosatelliten (z. B. CAG-Expansion in *HTT* bei der Huntington-Krankheit), Minisatelliten (z. B. 99-bp-Expansion in *PLIN4*) und Makrosatelliten (z. B. D4Z4 bei FSHD). Im menschlichen Genom gibt es mehr als eine Million solcher RE, die häufig durch transponierbare Elemente wie LINEs und SINEs eingeführt werden. Die meisten RE sind gutartig, aber einige überschreiten krankheitsspezifische pathogene Schwellenwerte, wie z. B. 40 CAG-Wiederholungen im Gen *HTT*. Mehr als 80 % der RE-Erkrankungen stehen im Zusammenhang mit neurologischen Erkrankungen.

Die Charakterisierung der RE kann Unterbrechungen aufdecken, die durch Deletionen, Insertionen oder Fehlpaarungen verursacht werden können. Die Unterbrechungen können das Motiv, die Länge und das pathogene Potenzial von RE verändern. Solche Unterbrechungen können die RE während der DNA-Replikation stabilisieren, das Expansionsrisiko verringern, die RNA-Struktur und -Toxizität beeinflussen oder die epigenetische Regulation modulieren. Sie können zu milderen Phänotypen oder einem verzögerten Erkrankungsalter führen. Insgesamt können Unterbrechungen in RE als genetische Modifikatoren wirken, die die Stabilität, Pathogenität und Vererbung von RE-Erkrankungen beeinflussen, was sie für die Diagnose und Therapieentwicklung wichtig macht.

Die Hypothese dieser Arbeit lautet, dass bei RE-Erkrankungen eine mosaikartige genetische Variabilität auftritt, die als genetischer Modifikator wirken kann.

Im ersten Ziel wurde das *TAF1*-SVA-Retrotransposon bei X-linked Dystonia-Parkinsonism (XDP) mithilfe von Long-Read Sequenzierungstechnologien untersucht, um die mosaikartige genetische Variabilität sowie die Länge und Unterbrechungen der RE zu bestimmen. Das zweite Ziel war die Bewertung der Stabilität von den entdeckten Modifikatoren über Familiengenerationen hinweg bei XDP. Im dritten Ziel wurde die kürzlich entdeckte RE in *FGF14* untersucht, um festzustellen, ob Varianten bei *FGF14*-bedingten Ataxien vorhanden sind. Das letzte Ziel besteht darin, festzustellen, ob die *FGF14*-RE mit multipler Systematrophie (MSA) assoziiert ist.

In dieser Arbeit ermöglichten die Vorteile der Long-Read Sequenzierung bei der Untersuchung der RE den direkten Nachweis der genetischen Variabilität, epigenetischer Informationen und die Charakterisierung. Die Long-Read Sequenzierung bietet eine genaue Erkennung der Länge der RE bei XDP und SCA27B. Darüber hinaus erlaubte die Technologie sogar die Entdeckung neuartiger somatischer Unterbrechungen, bei denen die Häufigkeit von mosaikartigen Unterbrechungen mit der Stabilität der Länge der RE bei XDP assoziiert ist. Die Frequenz der Unterbrechungen in XDP beeinflussen die Übertragung der Länge der RE über Generationen hinweg. Eine höhere Häufigkeit der Unterbrechungen stabilisiert die Länge der RE über Generationen hinweg, was eine schützende Wirkung der Unterbrechungen bei RE zeigt. Bei SCA27B führten die Ergebnisse zu einer neuen Kategorisierung der Patienten unter Verwendung der reinen GAA-Länge ohne Unterbrechungen. Die Häufigkeit einer langen RE in *FGF14* war bei Patienten mit MSA im Vergleich zu gesunden Personen leicht erhöht.

Zusammenfassend lässt sich sagen, dass die Long-Read Sequenzierung die Erkennung und Charakterisierung von RE verbessert hat. RE können bei bekannten neurologischen Erkrankungen als Modifikator auftreten und Unterbrechungen der RE können die Länge der RE stabilisieren und sind bei der Diagnose von RE-Erkrankungen zu berücksichtigen.

Table of contents

Introduction

Repeat expansion disorders

Technological advances have recently increased the number of identified repeat expansion disorders (RED). As the name suggests, the REDs are caused by a repeat expansion in the human genome (Figure 1). These repetitive sequences can disrupt gene expression, form toxic RNA aggregates, or produce misfolded proteins. More than 80% of the REDs are associated with neurodegenerative and neuromuscular diseases (Depienne and Mandel 2021). Over 50 REDs have been discovered, half of them in the last ten years due to the implementation of newer technologies like long-read sequencing (Chen et al. 2025; Leitão et al. 2024). The first discovered RED was the Fragile X Syndrome (FXS) in 1991, caused by a CGG-expansion in the *FMR1* promoter, leading to an intellectual disability and autism-like features (Verkerk et al. 1991).

Repeat expansions are repetitive DNA sequences ranging from 1 to over 1000 base pairs that are repeated multiple times. Based on the repeat motif length, they can be classified into three categories. The most common repeat expansions are microsatellites, called short-tandem repeats (STR). The repeat motif is between one and nine base pairs long and is repeated multiple times. Huntington's disease (HD) is one of the most prominent REDs. The repeat motif is three bases long (CAG) and can expand up to 250 times in the *HTT* gene (Moeller et al. 2021). Minisatellites are another category of repeat expansions. They have a repeat motif ranging from 10 to 99 base pairs (bp). An example of minisatellites is the expansion in the *PLIN4* gene, which is 99 bp long and can cause autophagic vacuolar myopathy (van der Maarel and Frants 2005). The 99 bp-long repeat unit is repeated 29 to 31 times. Macrosatellites are the last category with a repeat motif over 100 base pairs. The macrosatellites are mainly located in the heterochromatin and centromere regions. An example is the repeat structure at the subtelomere of chromosome 4q (van der Maarel and Frants 2005). This structure is called “D4Z4”, and each repeat unit has a size of 3.3kb and is repeated up to 100 times. In almost all patients with facioscapulohumeral muscular dystrophy (FSHD), the D4Z4 repeat tract is contracted to 1-10 repeat units.

Over a million repeat expansions are located in the human genome (Gymrek 2017; Willems et al. 2014). They are introduced into the human genome by long interspersed nuclear elements (LINEs) or short interspersed nuclear elements (SINEs), which are transposable elements that contain repetitive elements (Grandi and An 2013; Pascarella et al. 2022). Over 90% of the repeat expansions in the human genome are benign (Kozlowski et al. 2010). The expansions must reach a specific threshold to be disease-causing, which varies in all repeat expansion disorders. In Huntington's disease, over 40 CAG repeats in the *HTT* gene are considered pathogenic (Aldous et al. 2024). In contrast, over 250 GAA repeats in the *FGF14* gene are required to cause SCA27B (D. Pellerin, Danzi, et al. 2023). Many REDs have an intermediate range of the repeat number, where individuals can develop the disease or stay healthy.

Mechanistic expansion of the repeat tract

Three mechanisms repeat expansions undertake, where they dynamically contract or expand the repeat number. The first possible mechanism is through the polymerase during DNA replication. The hairpin structure of the repeat region, especially CAG repeats, tends to promote slippage of the DNA strands during the replication process, which causes the expansion of the repeat tract (Chan et al. 2013). The second mechanism of expanding the repeat tract is through DNA repair. The gene expression of DNA repair proteins modifies many repeat expansion disorders. The genes *MSH2*, *MSH3*, and *MLH1* are downregulated in many repeat expansion disorders (Wang et al. 2025). The last mechanism to expand the repeat tract is the recombination during meiosis.

Familial anticipation

A common feature of repeat expansion disorders is the elongation of the repeat tract across the transmission from one generation to the next. Most repeat expansion disorders have an inverse correlation between the repeat number and the age at onset of the disease. The shorter the repeat length, the later the onset age (Figure 1). Due to this correlation, the repeat expansion is expanding across multiple generations, leading to an earlier age of onset. This feature is called anticipation.

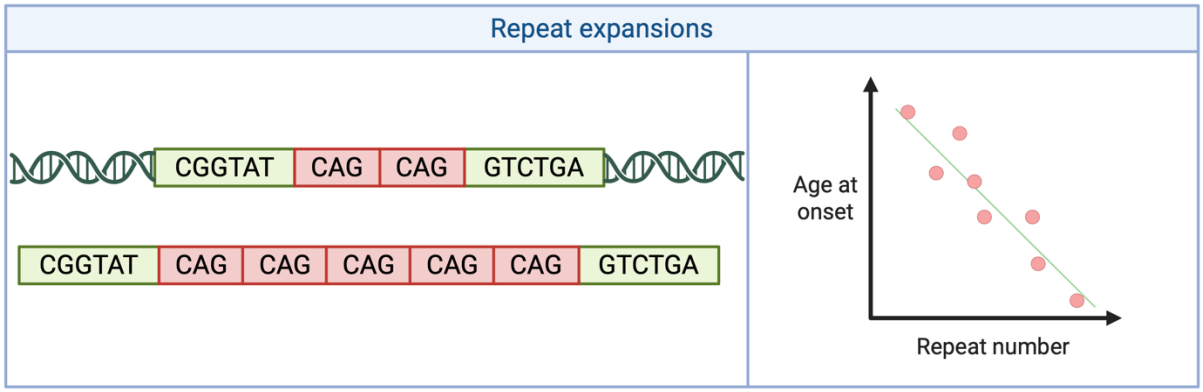

Figure 1: Repeat Expansion. (Left) Repeat expansion in a genomic context. One repeat unit has the sequence CAG. **(Right)** The correlation between the repeat number and age at onset of a repeat expansion disorder. Created in https://BioRender.com.

Pathological mechanism in repeat expansion disorders

Repeat expansions can have different functional consequences. One mechanism is epigenetic gene silencing, where repeat expansions can lead to a decreased level of the protein. This mechanism is associated with recessive or X-linked disorders. The repeat expansion can alter the chromatin conformation and availability and lead to lower gene expression. In amyotrophic lateral sclerosis (ALS), the GGGGCC expansion located in the first intron of the *c9orf72* gene is subject to epigenetic regulation via DNA methylation. On the one hand, long expansions are associated with hypermethylation of the promoter region, which leads to reduced gene expression (Liu et al. 2014). On the other hand, methylation of the repeat tract has been shown to reduce the formation of toxic RNA structures (Bauer 2016).

Another pathophysiological mechanism is RNA toxicity, where the repeat expansion undergoes transcription. This can be bidirectional, in a reverse and forward sense (Rohilla and Gagnon 2017; Groh et al. 2014). The expansion can accumulate in the transcript to RNA foci, clusters of different RNA structures and RNA-binding proteins. For example, in DM1, nuclear clumps of mRNA were detected. Splicing defects are another possibility for RNA toxicity. Repeat expansion in the mRNA can recruit additional factors, modulating splice site selection and activity. In amyotrophic lateral sclerosis (ALS), the repeat expansion in the first intron leads to G-quadruplexes forming, which interact with RNA-processing factors. This interaction affects the splicing of the *c9orf72* gene (Walsh et al. 2015).

If the repeat expansion is transcribed and has no RNA toxicity, it can still lead to protein misfolding and aggregation. The aggregation of proteins is characteristic of polyQ expansions. The repeat motif of these expansions is three bases long and codes for an amino acid. An example is Huntington's disease, where the CAG expansion leads to the translation of polyglutamine-rich proteins (Stoyas and La Spada 2018).

The last possibility is repeat-associated non-AUG (RAN) translation. Here, the translation is not dependent on a start codon and starts in the repeat expansion tracts. RAN translation can create up to six different polypeptides using all three reading frames and translating bidirectionally (Banez-Coronel and Ranum 2019; Zu et al. 2018). This mechanism is seen in more and more repeat expansion disorders.

Therapeutic approaches in repeat expansion disorders

The treatment of repeat expansion disorders is still under development, and different approaches are used. One approach is gene silencing therapy, which uses antisense oligonucleotides (ASO) or RNA interference. ASOs are synthetic strands of nucleic acids that bind to the RNA molecules with repeat tracts. The binding of the ASO leads to the degradation of the RNA or the prevention of the translation of the RNA (Bennett and Swayze 2010). Some ASOs have been developed in Huntington's disease and are already in clinical trials (Rodrigues et al. 2019; Van De Roovaart et al. 2023). Another mechanism for gene silencing therapies is RNA interference. RNA interference can be modulated by small interfering RNA (siRNA) or microRNA (miRNA). siRNA can target the mRNA, similar to ASOs, and lead to mRNA degradation (Martier et al. 2019).

Another approach is genome editing for the treatment of repeat expansion disorders. One mechanism to target and correct the repeat expansion is CRISPR/Cas9. CRISPR/Cas9 can detect the repeat expansion and cut the repeat tract in the affected genes (Alkanli et al. 2023). Until today, this approach has only been used in preclinical studies because delivering the CRISPR components in the affected tissue is still tricky.

The pathological consequences of repeat expansions are also a target for therapeutics. Small molecules can directly target the RNA with repeats or modulate the affected cellular pathway to alleviate the pathological consequences (Reddy et al. 2019). The binding of small molecules to the RNA can prevent the formation of toxic RNA structures.

Another approach is gene therapy, which performs gene replacement or regulatory element modification. In repeat expansion disorders, where the repeat expansion leads to a loss of function, healthy copies of the affected genes can be introduced by gene replacement. In DM1, functional copies of the *MBNL1* gene were introduced into cells by an AAV delivery system, but are still under investigation (Pascual-Gilabert et al. 2021). Regulated elements can be targeted to increase the expression of non-expanded alleles. In the fragile X syndrome, the repeat expansion leads to decreased expression of the *FMR1* gene. In one study, transcriptional activators were targeted to increase the expression of *FMR1* (Haenfler et al. 2018).

Another approach focuses on preventing or reversing the aggregation of toxic proteins. One mechanism is the upregulation of heat shock proteins, which are involved in the clearance of misfolded proteins (Gomez-Pastor et al. 2017).

Overall, different therapeutic strategies are investigated for repeat expansion disorders, but one key challenge for all approaches is the delivery to the affected tissue. Most repeat expansion disorders primarily affect the central nervous system. Therefore, the therapeutic agents must cross the blood-brain barrier and reach their target cells. This will enable the translation of these therapies from the preclinical to the clinical phase.

Repeat interruptions

The characterization of the repeat tract leads to the identification of interruptions in the repeat tract. These interruptions can be caused by deletions, insertions, or mismatches in the repeat sequence and can alter the repeat expansion's motif, length, and pathogenic potential. One of the most well-known repeat interruptions is found in Huntington's disease. An interruption at the 5'-end of the CAG repeat tract changes the sequence to $(CAG)_nCAA(CAG)_n$, due to converting a guanine to an adenosine (Figure 3). This slight variation impacts the pathogenicity of the repeat expansion because patients with the interruption have a delayed age at onset compared to patients with a pure CAG motif (Findlay Black et al. 2020).

Interruptions in the repeat tract can have different effects. The stabilization of the repeat expansion is one effect. Interruption can stabilize the repeat tract during DNA replication, reducing the expansion risk. Therefore, repeat interruptions are associated with milder phenotypes or delayed age at onset. The repeat interruption can also lead to new RNA secondary structures. This can decrease the RNA toxicity of the repeat expansions and alter the interaction between RNA and proteins. Another consequence of interruptions can be epigenetic modulation (Figure 2). In Fragile X-Syndrome, the interruption of the repeat tract is considered to change the promoter methylation, affecting the gene expression (Yrigollen et al. 2012).

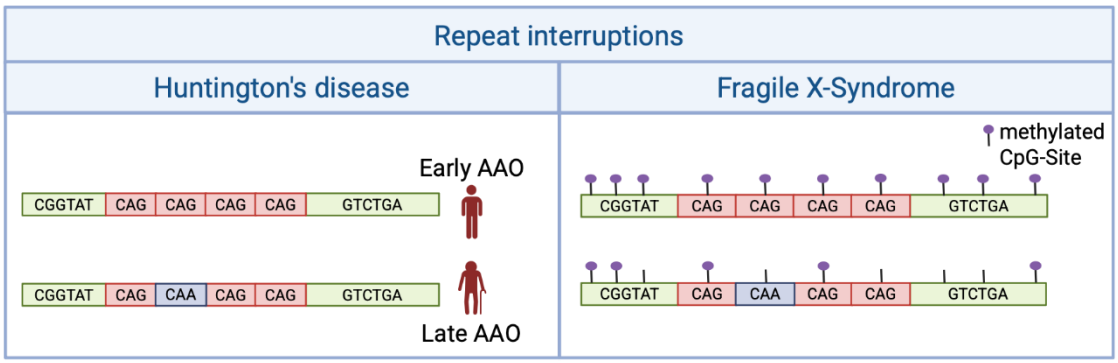

Figure 2: Repeat interruptions in repeat expansion disorders. (Huntington's disease) The repeat interruption (purple) in the repeat tract (red) leads to a delayed age at onset. (**Fragile X-Syndrome**) The repeat interruption (purple) in the repeat tract (red) leads to lower methylation frequency of the DNA. Created in https://BioRender.com. Legend: AAO: age at onset.

The interruptions can be inherited or arise somatically, often due to point mutations or errors by the polymerase. While many interruptions are located at the 5'-end of the repeat tract, i.e., in Huntington's disease, others are found throughout the repeat tract. In myotonic dystrophy type 1 (DM1), the CTG trinucleotide repeat tract can contain CCG repeats close to the 3'-end of the repeat expansion (Cumming et al. 2018). This interruption is associated with milder clinical features. The typical repeat interruption is only one repeat long, like in different spinocerebellar ataxias (SCA). In SCA1 and SCA2, the CAG repeat expansion can be interrupted in the middle of the repeat tract by a CTG sequence (Matsuyama et al. 1999). However, the repeat expansion in SCA37 is different. The pentanucleotide repeat expansion in the *DAB1* gene is interrupted by a more extended sequence in the middle of the tract.

The resulting sequence is $(ATTTT)_{60\text{-}79}(ATTTC)_{31\text{-}75}(ATTTT)_{58\text{-}90}$ (Seixas et al. 2017). The length of $(ATTTC)_n$ insertions is inversely correlated with age at onset.

Repeat interruptions also influence the inheritance patterns and intergenerational stability of repeat expansions. Repeat expansions with interruptions have a lower risk of anticipation. In Fragile X-syndrome, the methylation is affected by the interruption, and the repeat tract also has a higher stability (Yrigollen et al. 2012). Pure CGG repeats have a higher risk of expanding in maternal transmission than repeats with the AGG interruption. In Huntington's disease, paternal transmission is affected by the interruption. Usually, paternal transmission can lead to longer expansions than maternal transmission. The CAA interruption in the CAG repeats shows shorter expansions in children when they inherited the repeat expansion from their father.

In general, repeat interruptions have the potential to be genetic modifiers of repeat expansion disorders and influence repeat stability, pathogenicity, and inheritance. Understanding their structure and effects is crucial for clinical interpretation and therapeutic development.

Somatic instability

Repeat expansions are not static and can continue to change in somatic cells after birth, a process referred to as somatic instability (Figure 3). The repeat tract can be expanded or contracted in different tissues within the same individual, leading to tissue-specific repeat lengths.

The somatic instability can affect the disease severity and age at onset. More than 45 repeat expansions are associated with neurological disease (Depienne and Mandel 2021). Therefore, the somatic instability in the brain is a significant focus of current research. Somatic expansions in brain tissue can cause a higher toxicity of the repeat expansion in the disease-relevant tissue. It has also been shown that detected repeat lengths in blood can underestimate the actual repeat length in the disease-causing tissue (van Blitterswijk et al. 2013). Studies in Huntington's disease have reported that the repeat length in the blood was 42 repeats, and in the striatum, a repeat length of up to 100 repeats was detected (Handsaker et al. 2025). The striatum is the primary site of neurodegeneration in the disease. The discrepancy would lead to a misdiagnosis with a moderate disease progression. The higher expansion in the striatum is linked to earlier onset, more severe motor symptoms, and faster disease progression.

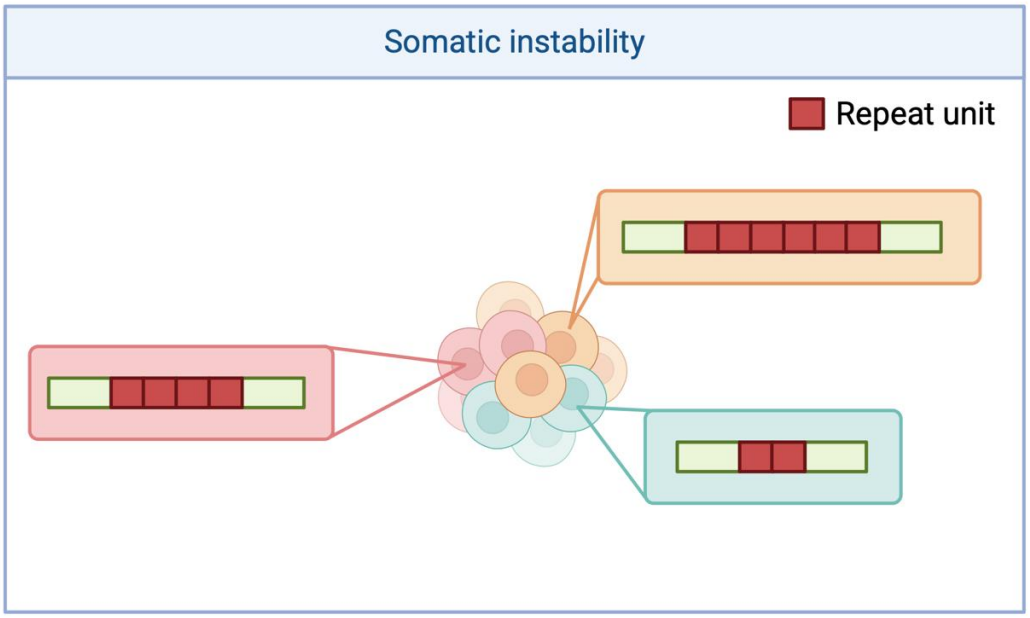

Figure 3: Somatic instability is the repeat number difference between different cells or tissues in one individual. Created in https://BioRender.com.

The somatic instability can occur due to the slippage of the DNA polymerase in the repeat tract, which would result in the insertion of more repeat units (Bzymek and Lovett 2001). Dysregulation of (mismatch repair MMR) proteins can also cause instability. The MMR-proteins can recognize the abnormal DNA structures, such as hairpins formed by the repeat tract, and ultimately start the process of removing them. The MMR-proteins MSH2, MSH3, and MLH1 are associated with several repeat expansion disorders (Rajagopal et al. 2023). In a mouse model of the Fragile X-Syndrome, the loss of the *MSH2* gene leads to a stabilized repeat number and no somatic instability (Lokanga et al. 2014). Also, the expansion of the repeat number during the transmission was prevented.

Somatic instability tends to increase with age, similar to other somatic mutations, contributing to disease progression over time. Tissues highly expressing the repeat-containing gene show greater levels of repeat expansion, suggesting a link between transcriptional activity and instability (Pellerin et al. 2025).

The somatic instability also provides a possible explanation for the phenotypic variability among patients who have identical repeat sizes in blood-derived DNA. The phenotype is then dependent on the somatic expansion in the affected tissue.

Somatic instability plays a crucial role in modulating clinical outcomes in repeat expansion disorders by influencing the repeat size in specific tissues. This has important implications for diagnostics, prognostics, and the development of therapies to stabilize repeat tracts or modify DNA repair pathways.

X-linked dystonia-parkinsonism

In the next three sections, I will briefly explain the investigated diseases in this thesis. X-linked dystonia-parkinsonism (XDP) is a neurodegenerative disorder characterized by rapidly progressive dystonia and parkinsonism (Lee et al. 2011). Also known as Lubag disease, it was first reported in 1976 (Lee et al. 1976). This rare condition is endemic to the Philippines. The initial symptoms often manifest as focal dystonia, which can progress to generalized dystonia. Parkinsonian features may develop in later stages, including resting tremors, bradykinesia, and rigidity (Lee et al. 2001).

XDP is caused by a SINE-VNTR-Alu (SVA) retrotransposon insertion in intron 32 of the *TAF1* (*TATA-binding protein-associated factor 1*) gene. The TAF1 protein is a subunit of the transcription factor IID (TFIID), which plays a crucial role in promoter recognition by RNA polymerase II (Wassarman and Sauer 2001). The insertion of the SVA retrotransposon leads to intron retention and reduced exon usage after the insertion, which causes a decreased *TAF1* expression (Domingo et al. 2015).

The *TAF1* gene is located on the X chromosome and follows an X-linked recessive inheritance pattern. Consequently, XDP predominantly affects males, with a male-to-female ratio of 100:1 (Rosales 2010). The estimated prevalence in the Philippines is around 0.32:100,000, with the highest prevalence on Panay Island (~5:100,000) (Lee et al. 2002). Due to emigration, XDP cases have been reported in the USA, UK, Canada, Germany, and Japan (Lee et al. 2011; Rosales 2010).

XDP is an adult-onset disorder with a mean age at onset (AAO) of 39.7 years in males, ranging from 12 to 64 years. Females generally develop symptoms about 13 years later on average (Evidente et al. 2004). The mean age at death is 55.6 years, resulting in a mean survival time of 16 years (Lee et al. 2011; Rosales 2010).

The SVA retrotransposon consists of five domains. At the 3' end is a poly-A signal, which is followed by a short interspersed nuclear element (SINE) region. In the middle of the SVA is a variable number tandem repeat (VNTR) region belonging to the minisatellites. The repeat motif is typically repeated five to 50 times. SVA elements are also called "mobile CpG islands" due to the VNTR region's high GC content and extensive methylation. Following the VNTR region is an Alu-like domain, and at the 5' end is a hexanucleotide domain. This domain belongs to the microsatellites, and the motif is $(CCCTCT)_n$ or $(AGAGGG)_n$. This hexanucleotide repeat domain is variable in the number of repeats, ranging from 30 to 55 (Westenberger).

The repeat number of the hexanucleotide repeat domain acts as a modifier in XDP (Westenberger et al. 2019). Higher repeat numbers are associated with decreased *TAF1* expression and an earlier AAO. This inverse correlation with the AAO also underlines anticipation in XDP, where expansions increase across generations. Notably, larger repeat expansions are more frequently observed during maternal transmission.

Brain structure abnormalities are common in XDP patients. Caudate head atrophy is observed in over 70% of cases, particularly in the Parkinsonism phase. Additionally, striatal volume loss occurs, with the rostral region being more severely affected. The extent of atrophy correlates with disease duration, and affected brain regions exhibit neuronal loss and astrogliosis.

Currently, no cure exists for XDP. Treatment focuses on symptom management of the dystonia or Parkinsonism phase. During the dystonia phase, medications such as anticholinergic agents, antihistamines, and antipsychotics may provide relief. In the Parkinsonism phase, levodopa and dopamine agonists are administered; however, these drugs can exacerbate dystonia symptoms (Jamora et al. 2011; Jankovic 2006). For severe dystonia in early disease stages, deep brain stimulation is the preferred intervention, though it does not alleviate Parkinsonian symptoms (Abejero et al. 2019).

Spinocerebellar ataxia 27 B

Spinocerebellar ataxia 27 B (SCA27B) is a recently discovered repeat expansion disorder caused by pathogenic expansion in the *FGF14* (*Fibroblast Growth Factor 14*) gene (D. Pellerin, Danzi, et al. 2023; H. Rafehi et al. 2023). The typical age at onset is between 50 and 70 years (Jacobi et al. 2015; Tezenas du Montcel et al. 2014). Unlike many other repeat expansion disorders, the inverse correlation between age at onset and repeat length is weakly detected (D. Pellerin, Danzi, et al. 2023; H. Rafehi et al. 2023). The core phenotype is a slowly progressive pancerebellar syndrome primarily characterized by gait ataxia and cerebellar oculomotor impairment (D. Pellerin, Danzi, et al. 2023; Wilke et al. 2023; David Pellerin, Heindl, et al. 2024; Wirth et al. 2023). At disease onset, most patients report gait unsteadiness, while approximately half of the patients experience episodic visual disturbances, like diplopia, blurring, and dizziness (D. Pellerin, Danzi, et al. 2023; Wilke et al. 2023; Ashton et al. 2023). As in XDP, no curative treatment is available for SCA27B. The only approved therapy, omaveloxolone, is typically used in patients with Friedreich's ataxia, another GAA repeat-associated repeat expansion disorder.

The disease follows an autosomal-dominant inheritance pattern. The mutation leads to cerebellum dysfunction that causes impaired coordination and neurological symptoms. While missense, frameshift, and nonsense mutations in *FGF14* are associated with SCA27A, only the repeat expansion is associated with SCA27B (van Swieten et al. 2003a; Dalski et al. 2005). The repeat expansion is located in the first intron of *FGF14* and has a GAA motif.

The exact pathogenic threshold of the repeat number remains under discussion. Initial studies suggest that an expansion greater than 300 GAA repeat units is pathogenic, and a repeat number between 250 and 300 has reduced penetrance (D. Pellerin, Danzi, et al. 2023; H. Rafehi et al. 2023). More recent studies have a lower threshold with 250 repeat units for pathogenicity, and the range for reduced penetrance is between 200 and 250 (Hengel et al. 2023; David Pellerin, Heindl, et al. 2024).

Repeat expansions with a pure motif over 75 repeats show high intergenerational instability, with the length of the repeat influencing the degree of instability (D. Pellerin, Danzi, et al. 2023; D. Pellerin et al. 2024). The maternal transmission has a higher tendency to expand the repeat tract. However, paternal transmission tends to contract, leading to the generation-skipping of the disease (D. Pellerin et al. 2024).

In European cohorts with unsolved adult-onset ataxia, the frequency of SCA27B ranges from 15 to 30% (D. Pellerin, Danzi, et al. 2023; H. Rafehi et al. 2023). In other regions, the frequency is around 10% (Hengel et al. 2023). Notably, the highest frequency is detected among French-Canadian patients with unsolved adult-onset ataxia, at 60% (D. Pellerin, Danzi, et al. 2023; Alshimemeri et al. 2023).

The first neuropathological examinations indicate cerebellum-specific changes in SCA27B patients (D. Pellerin, Danzi, et al. 2023; Wilke et al. 2023). The expression of *FGF14* is decreased in post-mortem cerebellum samples, suggesting a loss of function of FGF14 in SCA27B patients. A similar mechanism occurs in Friedreich's ataxia (FRDA), where GAA repeats are located in the *FXN* gene. In FRDA, the repeat expansion leads to DNA secondary structure, which inhibits transcription (Sakamoto et al. 1999; 2001).

Ongoing research is needed to clarify repeat thresholds, refine diagnostic criteria, and develop targeted therapies that address transcriptional dysregulation and neuronal dysfunction.

Multiple system atrophy

Multiple system atrophy (MSA) is an adult-onset neurodegenerative disorder affecting various systems in the brain. It was first described in 1969 by Graham and Oppenheimer (Graham and Oppenheimer 1969). MSA belongs to the group of atypical Parkinsonian syndromes and is characterized by a rapid clinical progression and an average survival of seven to nine years after diagnosis.

In contrast to Parkinson's disease patients, MSA patients have alpha-synuclein aggregation in oligodendrocytes, rather than neurons (Papp et al. 1989; Spillantini et al. 1998). This aggregation is known as glial cytoplasmic inclusions. Additionally, MSA is characterized by the loss of neuronal cells, gliosis, and widespread brain atrophy. This pathology distinguishes MSA from other synucleinopathies such as Parkinson's disease and Dementia with Lewy bodies. The condition is categorized into two subtypes: MSA-parkinsonian (MSA-P) and MSA-cerebellar (MSA-C) (Jellinger et al. 2005). The symptoms of MSA-P are similar to Parkinson's disease, with rigidity and bradykinesia. Different from Parkinson's disease, the patients don't respond to Levodopa (Wenning et al. 2000; Miki et al. 2021).

MSA-C is associated with impaired coordination, gait instability, and dysarthria. The distinction between MSA-C and other cerebellar disorders is hard.

Up to 30% of patients with sporadic adult-onset cerebellar ataxia may have MSA-C (Lin et al. 2014; Abele et al. 2002). MSA-C has partial phenotypic overlap with SCA27B (Wirth et al. 2024). Unlike other neurodegenerative disorders, genetic contributions to MSA remain poorly defined. The gene *MAPT* is associated with MSA in European ancestry, but this has not been replicated in other populations (Gu et al. 2018; Sailer et al. 2016). Recently, more risk loci in MSA were identified, implicating *GAB1*, *TENM2,* and *RABGEF1* (Chia et al. 2024).

Long-read sequencing

In the coming section, the two long-read sequencing technologies will be explained. Long-read sequencing, or third-generation sequencing, represents a powerful advancement in genomics, particularly for characterizing repeat tracts. Unlike short-read sequencing, long-read sequencing does not rely on PCR amplification and can sequence longer nucleic acid fragments. This allows the sequencing of complete repeat tracts and offers superior resolution of low-complexity regions. The two leading platforms for long-read sequencing are Oxford Nanopore Technologies (ONT) and Pacific Biosciences (PacBio) (Figure 4).

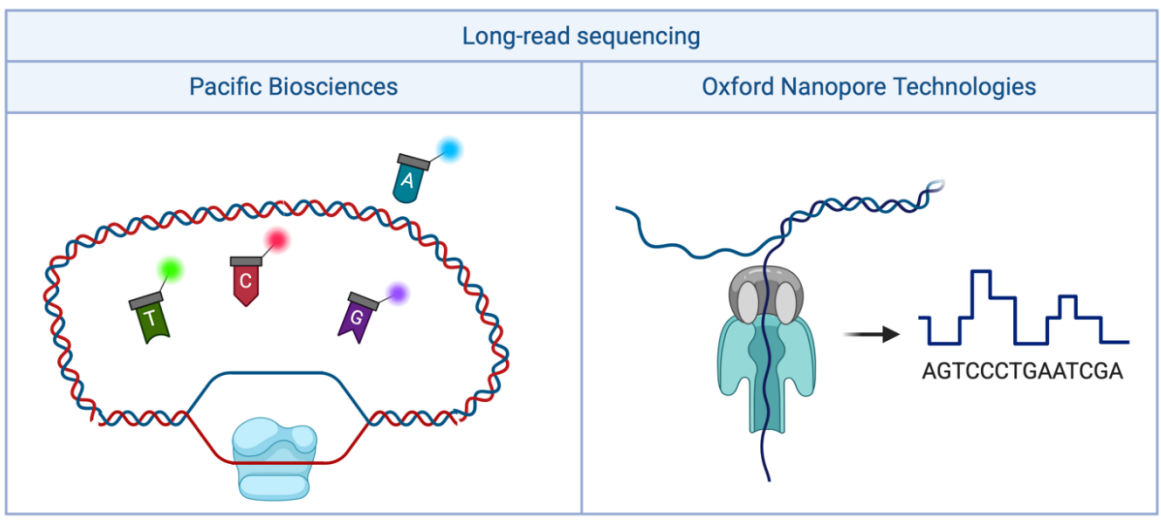

Figure 4: The two long-read sequencing approaches. Pacific Biosciences uses fluorescence-labeled nucleotides to detect the sequence. **Oxford Nanopore Technologies** uses a nanopore to detect current changes and determine the sequence. Created in https://BioRender.com.

The platform from PacBio uses single-molecule real-time (SMRT) sequencing technology. This technology utilizes Zero-Mode Waveguide (ZMW) wells that allow only a few nanometers of light

penetration, which reduces the background fluorescence (Levene et al. 2003). A DNA polymerase is attached to the bottom of each ZNW. The genomic DNA is circularized by attaching SMRTbell adapters to both ends of the dsDNA. The DNA polymerase-template complex is connected to the bottom of the ZMW. The four nucleotides are labeled with fluorophore dyes at the polyphosphate end. The light signal can detect the fluorophore dye after the polymerase incorporates a new nucleotide into the template (Pollard et al. 2018; Rhoads and Au 2015). After the detection, the fluorophore dye is cleaved off, allowing the incorporation of the next nucleotide (Eid et al. 2009). Each SMRT Cell contains millions of ZNWs, which allows high-throughput sequencing. One polymerase generates continuous reads by sequencing the circular template multiple times. The adapter sequence will be removed, resulting in various subreads of the template sequence. To create one final read, the subreads are aligned with each other. This read is then called a circular consensus sequence (CCS) read. Only CCS reads that achieve a Q-Score greater than 20 are classified as HiFi reads. These HiFi reads will then be used for the analysis. This technology can reach an accuracy of 99.9% and can directly detect base modifications. A limitation of PacBio sequencing is the maximum read length of 25 kb.

The other long-read sequencing platform is ONT. As the name indicates, this technology is based on nanopores (Eisenstein 2012). The first nanopore sequencing concept was an α-hemolysin ion channel embedded in a lipid bilayer (Kasianowicz et al. 1996). The nucleic acid can pass the lipid bilayer through the ion channel by the electric charge. The nucleic acid blocks the ion channel, which leads to a change in the electric current. This current change can be measured and used for the base calling (Clarke et al. 2009). The ONT platform uses additional motor proteins to push the nucleic acid through a pore to increase the sequencing speed. These motor proteins are adapters that bind to the 5'-end of the nucleic acid. The ONT platform has flow cells containing over 500 nanopores to sequence multiple sequences simultaneously. The current change of the nanopore can be detected via the membrane and is the raw signal.

In the process of base-calling, the raw signal is converted to a sequence of bases. The base-calling occurs in real-time when the nucleic acid passes through the pore. In the early phases of ONT, the base-calling accuracy was lower than that of PacBio and Illumina sequencing. However, the recent

updates of the base-calling algorithm allow an accuracy of 99.9% (Bogaerts et al. 2024). PacBio sequencing and ONT sequencing can detect epigenetic information. The ONT platform has no limitation on the maximum read length. The longest sequenced reads exceed 4 Mb. The integrity of the DNA sample limits the length of the reads.

ONT can perform non-targeted sequencing, like whole-genome and targeted sequencing, via PCR amplification, Cas9 enrichment, or adaptive sampling.

Long-read sequencing has improved the discovery and characterization of repeat expansion disorders. More than 20 novel repeat expansion disorders have been identified in recent years using these technologies (Depienne and Mandel 2021). The ability to sequence through entire repeat tracts, capturing repeat length, sequence interruptions, and epigenetic modifications, has been transformative for genetic diagnostics and research.

Hypothesis and Objectives

Hypothesis

Mosaic genetic variability occurs in repeat expansion disorders and can act as a genetic modifier.

Objectives

Objective 1: To investigate the *TAF1* SVA retrotransposon in X-linked dystonia parkinsonism using third-generation sequencing technologies to determine:

A) whether mosaic genetic variability is present and

B) repeat length or interruptions

Objective 2: To assess the stability of mosaic modifiers across familial generations in X-linked dystonia-parkinsonism (XDP).

Objective 3: To determine if mosaic variants are present in *FGF14*-related ataxias.

Objective 4: To determine whether the *FGF14* repeat expansion is associated with multiple system atrophy (MSA).

The result section of this thesis is based on my publications. I have written a short introduction and described my contribution for each publication.

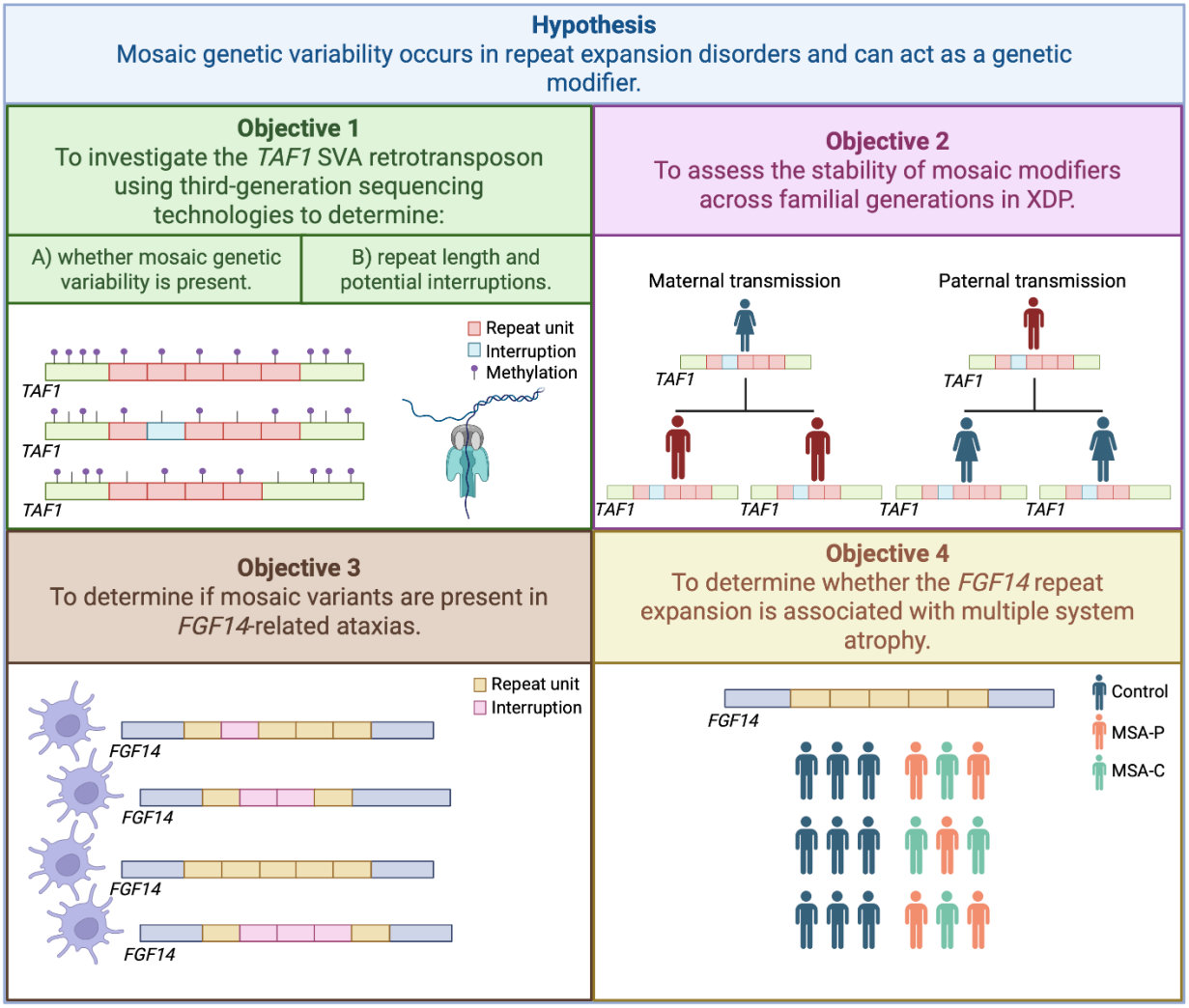

Figure 5: The hypothesis and objectives of this thesis. Created in https://BioRender.com. Legend: TAF1: Transcription factor 1; XDP: X-linked dystonia-parkinsonism; FGF14: Fibroblast growth factor 14; MSA-P: Multiple system atrophy-parkinsonian subtype; MSA-C: Multiple system atrophy-cerebellar subtype.

Results

Objective 1 A: To investigate the *TAF1* SVA retrotransposon using third-generation sequencing technologies to determine whether mosaic genetic variability is present

An insertion of an SVA retrotransposon causes X-linked dystonia-parkinsonism. The SVA retrotransposon consists of 5 domains, but only the hexanucleotide repeat domain at the 5' end is highly investigated. The repeat sequence $(AGAGGG)_n/(CCCTCT)_n$ was identified in 2017 (Bragg et al. 2017). The repeat number varies among individuals with XDP, ranging from 30 to 55. An inverse correlation has been reported between repeat number, disease onset, and severity (Bragg et al. 2017; Westenberger et al. 2019). The other four domains have not been investigated for genetic variability. The *TAF1* SVA insertion is associated with decreased *TAF1* expression. However, how the *TAF1* SVA insertion influences gene expression levels still remains unclear. The reduced expression can be rescued by excision of the SVA retrotransposon insertion (Rakovic et al. 2018; Aneichyk et al. 2018). The SVA, also known as "mobile CpG-island", due to its high GC content in the VNTR domain, may affect *TAF1* expression by changing the methylation status of the surrounding genomic regions. Twelve predicted enhancer sites are around the SVA *TAF1* insertion. The full-length *TAF1* SVA insertion and flanking regions have not been investigated until this study.

In this study, titled "Elucidating hexanucleotide repeat number and methylation within the X-linked dystonia-parkinsonism (XDP)-related SVA retrotransposon in *TAF1* with Nanopore sequencing", we wanted to show the ability of long-read sequencing on repeat expansions to investigate the *TAF1* gene. For this, we performed two different approaches with long-read sequencing by ONT. One approach was based on a long-range PCR of the repeat tract to characterize the sequence. The other approach was based on a Cas9 enrichment to obtain the epigenetic modifications of the DNA. We have analyzed blood-derived DNA from 96 patients with XDP using a long-range PCR-based approach. These 96 samples were multiplexed and loaded on two R9.4.1 flow cells from ONT and were then sequenced on a GridION. The methylation frequency was analyzed in different brain regions from one individual and in blood-derived DNA from one control.

I have created CRISPR RNAs (crRNA) for the Cas9 enrichment with the ChopChop tool. In total two products were created, one specific for the SVA insertion, and the other includes enhancer sites upstream and downstream of the SVA insertion. To obtain a good cleavage efficiency, two crRNAs were used to cut upstream of the target region and two downstream. The library was loaded on R9.4.1 flow cells and sequenced on a MinION or GridION. Susen Schaake performed the DNA extraction, PCR amplification, Cas9 enrichment, and library preparation.

After the sequencing, statistical analyses, and interpretations, I performed the bioinformatic analysis. I performed the base calling with the software Guppy, which ONT provides. The mapping of the reads to a reference sequence was performed with the software Minimap2. I called the single-nucleotide polymorphisms in the SVA *TAF1* insertion with the software Bcftools in the 96 PCR-amplified individuals. I detected the repeat length using the NCRF (Noise-canceling repeat finder) software. I performed a Spearman's correlation with the previous result of the repeat number by a long-range PCR to validate the results of the long-read sequencing. I developed a new downstream pipeline to detect the repeat tract interruptions. The pipeline counts every interruption in every position in the repeat tract of every read. I used the software Nanopolish to analyze the methylation. A non-parametric Mann-Whitney U-test assessed the differences between different tissues of a patient with XDP and a control. I performed the statistical analysis with GraphPad Prism. Finally, I wrote the first draft of the manuscript, supervised and revised by Dr. Theresa Lüth and Prof. Joanne Trinh, including all figures, tables, and supplementary materials, and I contributed to all revisions. This study has shown that Oxford Nanopore Technologies can reliably detect repeat numbers, methylation, and variation in the repeat motif.

Title

Elucidating hexanucleotide repeat number and methylation within the X-linked dystonia-parkinsonism (XDP)-related SVA retrotransposon in *TAF1* with Nanopore sequencing

Authors

Theresa Lüth[1+], **Joshua Laß[1+],** Susen Schaake[1], Inken Wohlers[2,3], Jelena Pozojevic[1], Roland Dominic Jamora[4], Raymond L. Rosales[5], Norbert Brüggemann[1,6], Gerard Saranza[7], Cid Czarina E. Diesta[8], Kathleen Schlüter[1], Ronnie Tse[1], Charles Jourdan Reyes[1], Max Brand[1], Hauke Busch[2,3], Christine Klein[1], Ana Westenberger[1], and Joanne Trinh[1]*

[1] Institute of Neurogenetics, University of Lübeck, 23538 Lübeck, Germany

[2] Medical Systems Biology Division, Lübeck Institute of Experimental Dermatology, University of Lübeck, 23538 Lübeck, Germany

[3] Institute for Cardiogenetics, University of Lübeck, 23538 Lübeck, Germany

[4] Department of Neurosciences, College of Medicine, Philippine General Hospital, University of the Philippines, Manila 1000, Philippines

[5] Department of Neurology and Psychiatry, The Hospital Neuroscience Institute, University of Santo Tomas, Manila 1008, Philippines

[6] Department of Neurology, University of Lübeck, 23538 Lübeck, Germany

[7] Section of Neurology, Department of Internal Medicine, Chong Hua Hospital, Cebu City 6000, Philippines

[8] Department of Neurosciences, Movement Disorders Clinic, Makati Medical Center, Makati City 1229, Philippines

* Correspondence author

[+] Equal contribution

Published in *Genes*, 2021, *13*(1), 126; DOI: doi.org/10.3390/genes13010126

Abstract:

Background

X-linked dystonia-parkinsonism (XDP) is an adult-onset neurodegenerative disorder characterized by progressive dystonia and parkinsonism. It is caused by a SINE-VNTR-Alu (SVA) retrotransposon insertion in the *TAF1* gene with a polymorphic $(CCCTCT)_n$ domain that acts as a genetic modifier of disease onset and expressivity.

Methods

Herein, we used Nanopore sequencing to investigate SVA genetic variability and methylation. We used blood-derived DNA from 96 XDP patients for amplicon-based deep Nanopore sequencing and validated it with fragment analysis, which was performed using fluorescence-based PCR. To detect methylation from blood- and brain-derived DNA, we used a Cas9-targeted approach.

Results

High concordance was observed for hexanucleotide repeat numbers detected with Nanopore sequencing and fragment analysis. Within the SVA locus, there was no difference in genetic variability other than variations of the repeat motif between patients. We detected high CpG methylation frequency (MF) of the SVA and flanking regions (mean MF = 0.94, SD = ±0.12). Our preliminary results suggest only subtle differences between the XDP patient and the control in predicted enhancer sites directly flanking the SVA locus.

Conclusions

Nanopore sequencing can reliably detect SVA hexanucleotide repeat numbers, methylation, and variation in the repeat motif.

Introduction

X-linked dystonia-parkinsonism (XDP) is a neurodegenerative movement disorder, and its phenomenology was first described in the literature in 1976 (L. V. Lee et al. 1976). Patients originate mainly from the Philippines or are of Filipino descent and mostly aggregate on the island of Panay. A known family history of the disease exists for ~94% of the patients. XDP originated through a founder mutation approximately 1000 years ago (Rosales 2010). The disease is characterized by dystonic movements and postures as well as parkinsonism due to an insertion of the retrotransposon SINE-VNTR-Alu (SVA) in intron 32 of the *TAF1* (TATA-binding protein-associated factor 1) gene (L. V. Lee et al. 2001; Pauly et al. 2020).

The *TAF1* SVA insertion has five domains. There is a hexanucleotide repeat domain at the 5′ end, consisting of the repeat sequence $(CCCTCT)_n$ (Bragg et al. 2017). This hexanucleotide repeat $(CCCTCT)_n$ domain varies in repeat numbers among patients, ranging from 30 to 55. The repeat number is inversely correlated with age at onset and disease severity (Bragg et al. 2017; Westenberger et al. 2019). In addition, somatic mosaicism has been observed, with a higher median number of repeats detected in the cerebellum and basal ganglia than in blood (Reyes et al. 2021). In XDP patients, seven variants have been found on the X chromosome: five single-nucleotide variants (SNVs), a 48-bp deletion, and the SVA insertion (Domingo et al. 2015; Makino et al. 2007). Within the SVA, no variants have been reported besides the $(CCCTCT)_n$ repeat polymorphism (Bragg et al. 2017).

The *TAF1* SVA insertion is also associated with decreased *TAF1* expression (Westenberger et al. 2019). The reduced *TAF1* expression observed in blood and patient-derived induced pluripotent stem cells can be rescued by excision of the retrotransposon insertion (Rakovic et al. 2019; Aneichyk et al. 2018). Thus, *TAF1* reduction is a consequence of the SVA insertion. However, the question of how the *TAF1* SVA insertion influences gene expression levels remains an enigma. Of note, two enhancers are predicted to be located upstream, and ten enhancers downstream of the *TAF1* SVA insertion. The SVA itself is highly methylated due to the high "GC" content (~60%) within the variable number tandem repeat (VNTR) region, also known as "mobile CpG-island" (Makino et al. 2007; Ewing et al. 2020). Therefore, the SVA retrotransposon insertion may affect *TAF1* expression by changing the methylation status (causing hypo- or hypermethylation) of the surrounding genomic region across several enhancer sites.

There are approximately 2700 SVA elements within the human reference genome (hg19) (Wang et al. 2005), and specific characterization of the *TAF1* SVA insertion in XDP patients has been hard to achieve with short-read sequencing technologies. *TAF1* SVA is a non-reference mobile element. Recently, mobile element insertions have been investigated in the context of the Simons Genome Diversity Project, and on average, 47 non-reference mobile element insertions are present per individual (Watkins et al. 2020). Like XDP, the insertions of SVAs have been implicated in many diseases, such as neurofibromatosis type 1 and hemophilia B (Pfaff et al. 2021). To our knowledge, the full-length *TAF1* SVA and flanking regions (>22 kb) have not been sequenced and investigated.

In this study, we establish a straightforward Nanopore sequencing workflow to investigate the genetic architecture of *TAF1* SVA by characterizing: (1) genomic variants within the SVA, (2) variations of the hexanucleotide repeat number, and (3) CpG methylation by utilizing Nanopore long-read sequencing.

Materials and Methods

Patient Demographics

The study was approved by the Ethics Committees of the University of Luebeck, Germany, and the Metropolitan Medical Center, Manila, Philippines (REF: IRB-MMC #: 10-073). To analyze genomic variants within the SVA and detect variations of the hexanucleotide repeat domain, n = 96 patients with XDP were investigated. Only male patients were included as XDP follows an X-linked recessive inheritance pattern. The mean age at onset (AAO) was 40.66 (SD = ±8.75), and the mean age at examination (AAE) and sample collection was 45.4 (SD = ±10.24) (Supplementary Table S1).

The CpG methylation was investigated in blood-derived DNA from one deceased XDP patient (L-7995) and one control (L-14529). The control was matched according to age, gender, and ethnicity. For the patient (L-7995), brain tissue samples derived from the basal ganglia (BG) and cerebellum (CRB) were also available. The patient had an AAO of 31 years. The AAE was 36 years in the patient and the control.

Single-Nucleotide Variants and Repeat Detection

DNA was extracted with the Blood and Cell Culture DNA Midi kit (Qiagen). Long-range PCR was performed to amplify the *TAF1* SVA (amplicon Size: 3.2 kb) in XDP patients, as previously described (Reyes et al. 2021), using the PrimeSTAR GXL DNA Polymerase® (Takara Bio). The primer sequences are documented in Supplementary Table S2. Subsequently, 1 µg of each patient-derived PCR product was barcoded with the Native 96 Barcoding Kit (EXP-NBD196) and multiplexed. Two libraries with the Ligation Sequencing Kit (LSK109) were generated for Nanopore sequencing on two R9.4.1 flow cells on a GridION. The input for library preparation was 200 fmol of DNA per sample.

Validation by fragment analysis to determine the repeat length of the hexanucleotide $(CCCTCT)_n$ was performed with a fluorescein amidites (FAM) labeled primer, as previously described (Bragg et al. 2017; Westenberger et al. 2019).

Methylation Detection

Cas9-targeted sequencing from Oxford Nanopore Technology (ONT) was performed to obtain the epigenetic information and enrich the target region. For the specific ligation of the sequencing adapter, the blunt ends with 5′ phosphates resulting from the Cas9 ribonucleoprotein complex, cleaving out the region of interest. The CRISPR RNAs (crRNAs) were designed with the ChopChop tool (https://chopchop.cbu.uib.no, accessed on 10 December 2021) (Montague et al. 2014). Four crRNAs were used upstream of the *TAF1* SVA insertion, and four crRNAs were used downstream (Supplementary Table S3 and Figure S1A). Two different library preparations were used for the Cas9-targeted enrichment. The first library consisted of the ~22 kb region of interest (crRNA 1, 2, 7, and 8). The second library targeted a 5.5 kb product specifically around the SVA (crRNA 3, 4, 5, and 6). The second target was 2.8 kb in size for the control without an SVA insertion. We prepared multiple libraries for the DNA derived from one patient (L-7995) or one control (L-14529). To prepare the individual libraries, two crRNAs were used to cut upstream of the target region and two downstream to enhance the efficiency of Cas9 DNA cleavage. For the blood-derived DNA of the patient with XDP, we have used five flow cells (R9.4.1) loaded with six libraries (5 × 5 µg and 1 × 1 µg). For the BG-derived DNA of the patient with XDP, we have used five flow cells (R9.4.1) loaded with seven libraries (2 × 5 µg, 3 × 3 µg, 1 × 2 µg, and 1 × 1 µg).

For the CRB-derived DNA of the patient with XDP, we have used four flow cells (R9.4.1) loaded with eight libraries (7 × 5 µg and 1 × 1 µg). For the blood-derived DNA of the healthy control, we have used five flow cells (R9.4.1) loaded with six libraries (5 × 5 µg and 1 × 1 µg).
The enriched DNA was prepared with the Nanopore Ligation Sequencing Kit (SQK-LSK109), loaded on an R9.4.1 flow cell, and sequenced with the MinION or GridION. All sequencing data obtained for methylation analysis were combined to maximize coverage depth.

Data Analysis

Base-calling was performed with Guppy version 5.0.11, and the base-calling software is available for Nanopore community members (https://community.nanoporetech.com, accessed on 10 December 2021). For the detection of the repeat length, the super accuracy model (DNA_r9.4.1_450bps_sup.cfg) and the fast model (DNA_r9.4.1_450bps_fast.cfg) were used. The corresponding configuration file names were provided as parameters for the Guppy software. The expected base-calling accuracy for the super accuracy model is 98.3% and 95.8% for the fast model. (https://community.nanoporetech.com/posts/guppy-v5-0-7-release-note, accessed on 10 December 2021). Base-calling for the methylation detection was performed with the fast model. All reads were mapped to the reference sequence with the software Minimap2 (v2.17). The coverage was determined with the software Samtools (v1.9). Variants were identified with the software Bcftools (v1.9) (https://github.com/samtools/bcftools, accessed on 10 December 2021). All reported positions by Bcftools were controlled in the VCF file to prevent false-positive results. We filtered for hemizygous allelic frequency (>90%) and good quality (Phred score Q > 20). Lastly, variants were evaluated in the Integrative Genomics Viewer (IGV) to exclude erroneously called variants within homopolymeric stretches.
The repeat length was detected using the NCRF software (Noise-canceling repeat finder) (v1.01.02) (Harris, Cechova, and Makova 2019). The median of all reads was calculated as previously described to determine the repeat length for one patient (Reyes et al. 2021). The NCRF alignment was used additionally to explore the frequency of deletions, insertions, and mismatches within the repeat motif.

For the Cas9-targeted sequencing data, methylation was called using the software Nanopolish (v0.13.2), which can detect 5′-methylcytosine (5 mC) in a CpG context. Nanopolish requires, besides the FASTQ and FAST5 files, the alignment in a BAM format as input. To counteract potential off-target effects of the CRISPR-Cas9 enrichment, the BAM file was filtered for reads with an alignment length >3 kb in the patient- or >1.5 kb in control-derived samples. Only CpG sites covered by >10 reads were included in the analysis.

Statistical Analysis

Spearman correlation was performed to assess the concordance of the detected hexanucleotide repeat number between Nanopore sequencing and fragment analysis. The median repeat number and the interquartile range detected by NCRF from the Nanopore data and the number of repeats detected with fragment analysis were used for the correlation plot.

In addition, we used the NCRF report for each sample to assess repeat motif interruptions. To determine matches and mismatches (i.e., deletions, insertions, and substitutions) between the Nanopore reads and the hexanucleotide repeat motif, NCRF uses a Smith–Waterman aligner approach and affine gap penalties (Harris, Cechova, and Makova 2019). The software reports the number of deletions, insertions, and substitutions per read. Subsequently, the mean number of repeat motif interruptions per read across all samples was calculated, as reported by NCRF, to explore accumulations of deletions, insertions, and substitutions within the SVA hexanucleotide repeat domain.

DNA methylation was compared across different tissues of a patient with XDP and a control. These differences were assessed by a non-parametric Mann–Whitney U-test, as previously described in Ewing et al. 2020 (Ewing et al. 2020).

Results

We first analyzed the sequencing data generated by PCR amplification and subsequent multiplexing on the Nanopore of the *TAF1* SVA insertion (n = 96 XDP patients). Across all individuals, we obtained a mean coverage of 17,645X (SD = ±12,392X) per barcode. The mean coverage of the samples ranged between 1690X (SD = ±190X) and 47,919X (SD = ±5074X) due to the variable sequencing efficiency of the barcodes. However, the coverage of the amplified region was even within the samples (Supplementary Figure S1B and Table S4). The mean sequence quality (Phred score) was 15.88 (SD = ±0.44), and the mean N50 was 3.38 kb (SD = ±66.42 bp) per barcoded sample.

Single-Nucleotide Polymorphisms within the SVA TAF1 insertion

SNVs located within the *TAF1* SVA insertion were called from the amplicon sequencing data of all 96 patients. After quality filtering and the final evaluation with IGV, no SNVs were detected.

Assessment of the Hexanucleotide Repeat Length

The hexanucleotide repeat number detection with long-read sequencing amplicon data resulted in a mean of 45.17 (SD = ±4.24) repeats, ranging from 35 to 57, using super accuracy base-calling (Figure 1A). Fragment analysis as an independent validation showed a mean number of 42.21 (SD = ±4.23) repeats, ranging from 33 to 54. The detected repeat numbers were highly concordant between the two methods (Spearman's $r = 0.9765$, Spearman's exploratory p-value $< 1 \times 10{-}15$, Figure 1A). However, the repeat number detected from long-read sequencing was consistently 1–4 repeat numbers higher compared to fragment analysis. Using Guppy fast base-calling for Nanopore sequencing resulted in a mean of 42.77 (SD = ±4.05) repeats, ranging from 33 to 54. Thus, we observed a higher concordance with fast base-calling between both methods (Spearman's $r = 0.9883$, Spearman's exploratory p-value $< 1 \times 10{-}15$, Figure 1B). There was an identical repeat number in n = 47 patients and a difference of ~1–2 repeats in $n = 48$ patients.

To further validate our workflow, we analyzed the previously shown negative association between the AAO and the repeat number (Bragg et al. 2017; Westenberger et al. 2019). The repeat number detected with Nanopore sequencing negatively correlated with AAO in patients with XDP (Spearman's $r < -0.80$, Spearman's exploratory p-value $< 1 \times 10{-}15$).

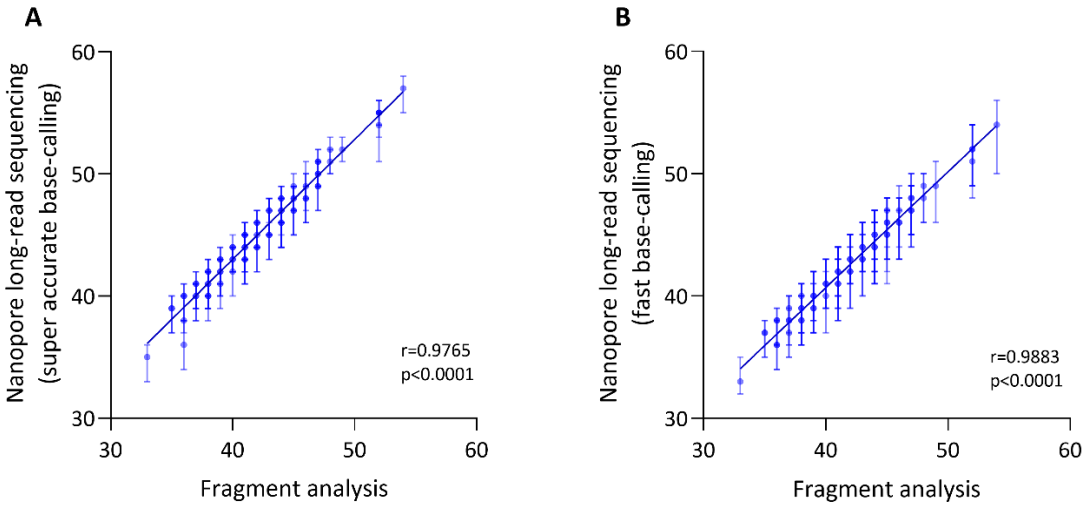

Figure 1. Repeat number detection using Nanopore long-read sequencing is highly concordant with the results from fragment analysis. Correlation between the median repeat numbers per individual of the $(CCCTCT)_n$ SVA domain, detected with fragment analysis and Nanopore sequencing using super accurate (A) or fast base-calling (B). Bars indicate the interquartile range of the detected repeat number with Nanopore sequencing. R = Spearman's rank correlation coefficient, p = Spearman's exploratory *p*-value.

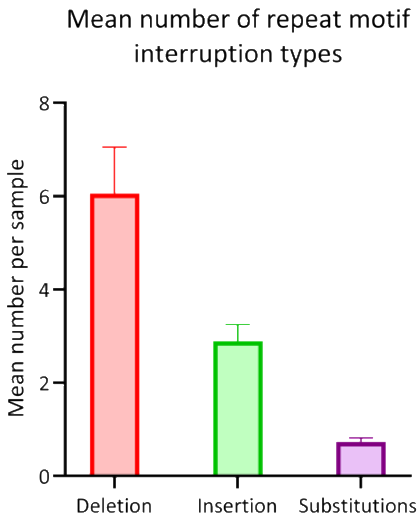

Figure 2. Occurrence of repeat motif interruptions. The bar chart shows the mean number of repeat motif interruptions per patient sample, stratified by type (i.e., deletion, insertion, substitution). The bars and whiskers represent the mean and upper limit of the standard deviation.

Next, we explored the continuity of the repeat motif. As reported by NCRF, the mean number of deletions per read within the repetitive sequence was 6.05 (SD = ±1.00), the mean number of insertions was 2.89 (SD = ±0.36), and the mean number of substitutions was 0.73 (SD = ±0.09). Thus, deletions were the most common type of interruptions detected in the hexanucleotide repeat sequence of XDP patients (Figure 2).

Methylation within the SVA and in the Flanking Regions

To assess the DNA CpG methylation of the SVA, we enriched the *TAF1* SVA insertion and flanking regions with a Cas9-targeted approach. We included blood- and brain-derived DNA from one XDP patient and blood-derived DNA from one age-matched control participant. We used two Cas9 enrichment strategies: (1) the *TAF1* SVA insertion and a short flanking region (~5.5 kb) and (2) the *TAF1* SVA insertion and a longer flanking region (~22 kb), including 12 predicted enhancer sites.

Enriching the shorter fragment resulted in an N50 of 4.5 kb for the patient-derived samples and an N50 of 2.0 kb for the control-derived sample. The mean Phred score of the reads ranged from 10.0 to 10.9, and the mean coverage ranged from 126.9X (SD = ±79.8X) to 1226.0X (SD = ±554.9X) (Supplementary Figure S1C).

Enriching the longer fragment resulted in an N50 of 4.7 kb for the patient-derived samples and an N50 of 8.8 kb for the control-derived sample. The mean Phred score of the reads ranged from 12.3 to 13.5, and the mean coverage was from 22.1X (SD = ±11.5X) to 591.0X (SD = ±1202.0X). The sequencing quality statistics were summarized in Supplementary Table S4.

Overall, the methylation levels within the SVA and in the up- and downstream flanking regions were high in the patient-derived samples (Figure 3A). However, the mean MF was lower in the brain-derived samples (BG: mean MF ± SD = 0.88 ± 0.15, CRB: mean MF ± SD = 0.90 ± 0.14) compared to the blood-derived sample (mean MF ± SD = 0.94 ± 0.12). There were n = 153 CpG sites within the SVA *TAF1* insertion (Figure 3B). Consistent with the overall methylation level across the 22 kb region, the mean CpG MF within the SVA specifically was still lower in the brain-derived samples (BG: mean MF ± SD = 0.87 ± 0.14, CRB: mean MF ± SD = 0.93 ± 0.08) compared to the blood-derived sample (mean MF ± SD = 0.96 ± 0.07) (exploratory Mann–Whitney U-test $p < 1.2 \times 10{-}6$) (Supplementary Figure S2A). In addition to patient-derived DNA, we analyzed blood-derived DNA from one healthy control (Figure 3C).

The overall MF across the SVA flanking region in the control sample was at 0.83 ± 0.17, which was lower than the patient-derived sample (MF ± SD = 0.93 ± 0.15) (exploratory Mann–Whitney U-test $p < 1 \times 10{-}15$, Supplementary Figure S2B). Despite a significant difference, the effect size was small.

There were 12 predicted enhancer sites located in the targeted region, 2 upstream and 10 downstream of the *TAF1* SVA insertion, according to the ENCODE project (reference number: wgEncodeEH000790). No CpG site was located within enhancer eight, and this predicted enhancer was excluded from the analysis. The mean MF of the enhancer sites ranged from 0.65 to 0.99 in the blood-derived sample, from 0.46 to 0.99 in the BG-derived sample, and from 0.37 to 0.99 in the CRB-derived sample (Figure 4A, Supplementary Table S5). In comparison, the mean MF of these enhancers ranged from 0.69 to 0.95 in the healthy control (Figure 4B, Supplementary Table S5). We detected significantly lower methylation of the enhancer sites within the *TAF1* SVA flanking region in the control compared to the patient-derived blood sample (exploratory Mann–Whitney U-test $p = 0.0033$, Supplementary Figure S2D). All enhancers showed lower methylation levels in the control subject, except for enhancer two. The most pronounced difference was observed at enhancer site six (mean MF patient: 0.91, mean MF control: 0.71) and nine (mean MF patient: 0.98, mean MF control: 0.70). However, the sample size is small, and the effect sizes are small, and differences remain difficult to interpret (see details in Section 4).

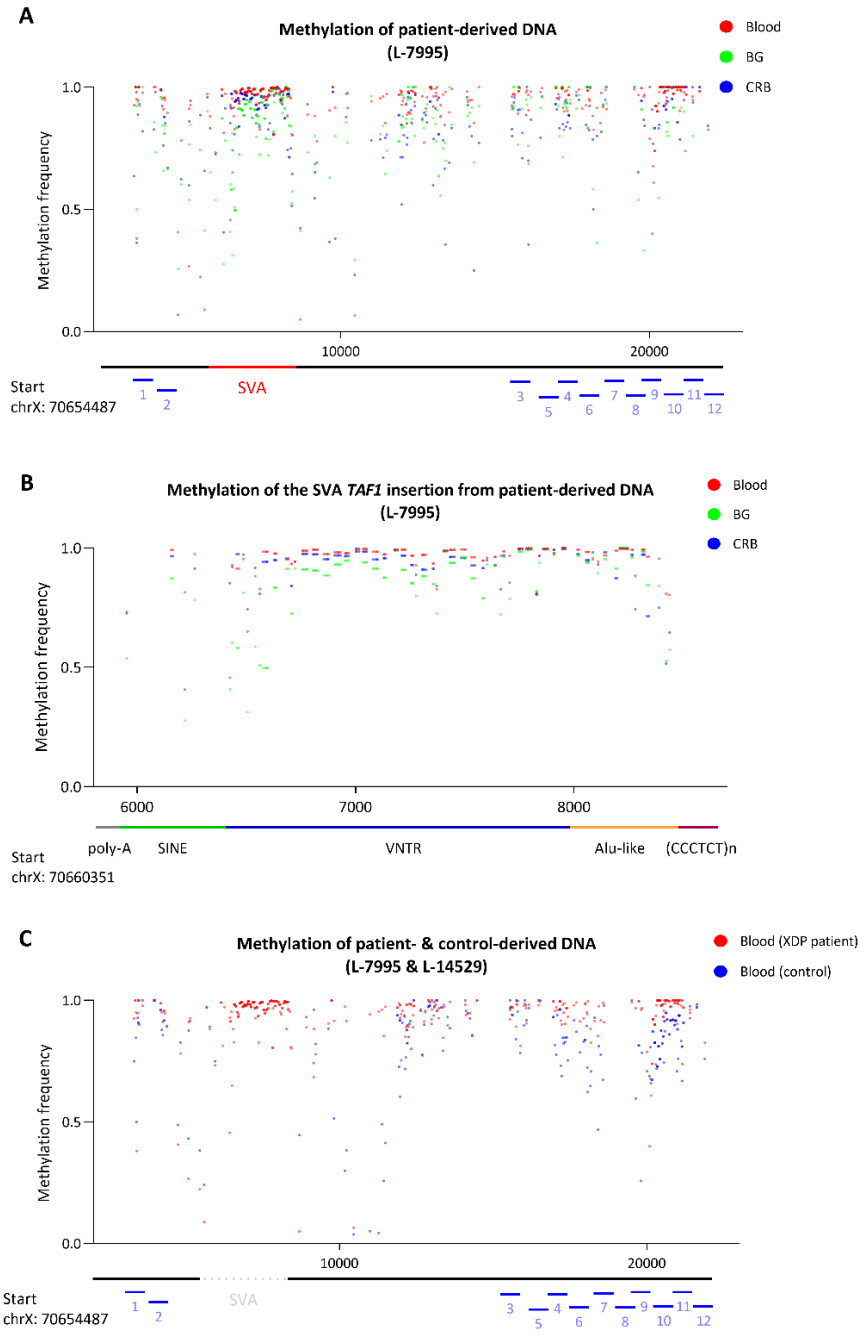

Figure 3. Methylation frequency of the *TAF1* SVA insertion and flanking regions. (**A**) Methylation frequency of two different brain tissues and blood-derived DNA from a patient with XDP. Red indicates methylation from blood-derived DNA, green from basal ganglia-derived (BG) DNA, and blue from cerebellum-derived (CRB) DNA. (**B**) Methylation frequency of *TAF1* SVA insertion with indicated SVA domains of the same patient-derived DNA samples. Red indicates methylation from blood-derived DNA, green from BG-derived DNA, and blue from CRB-derived DNA. (C) Methylation frequency of blood-derived DNA from a patient with XDP (red) and a control participant (blue). The x-axis indicates the position in the reference sequence. The bars indicate the location of predicted enhancers, the *TAF1* SVA insertion, and the insertion's subunits.

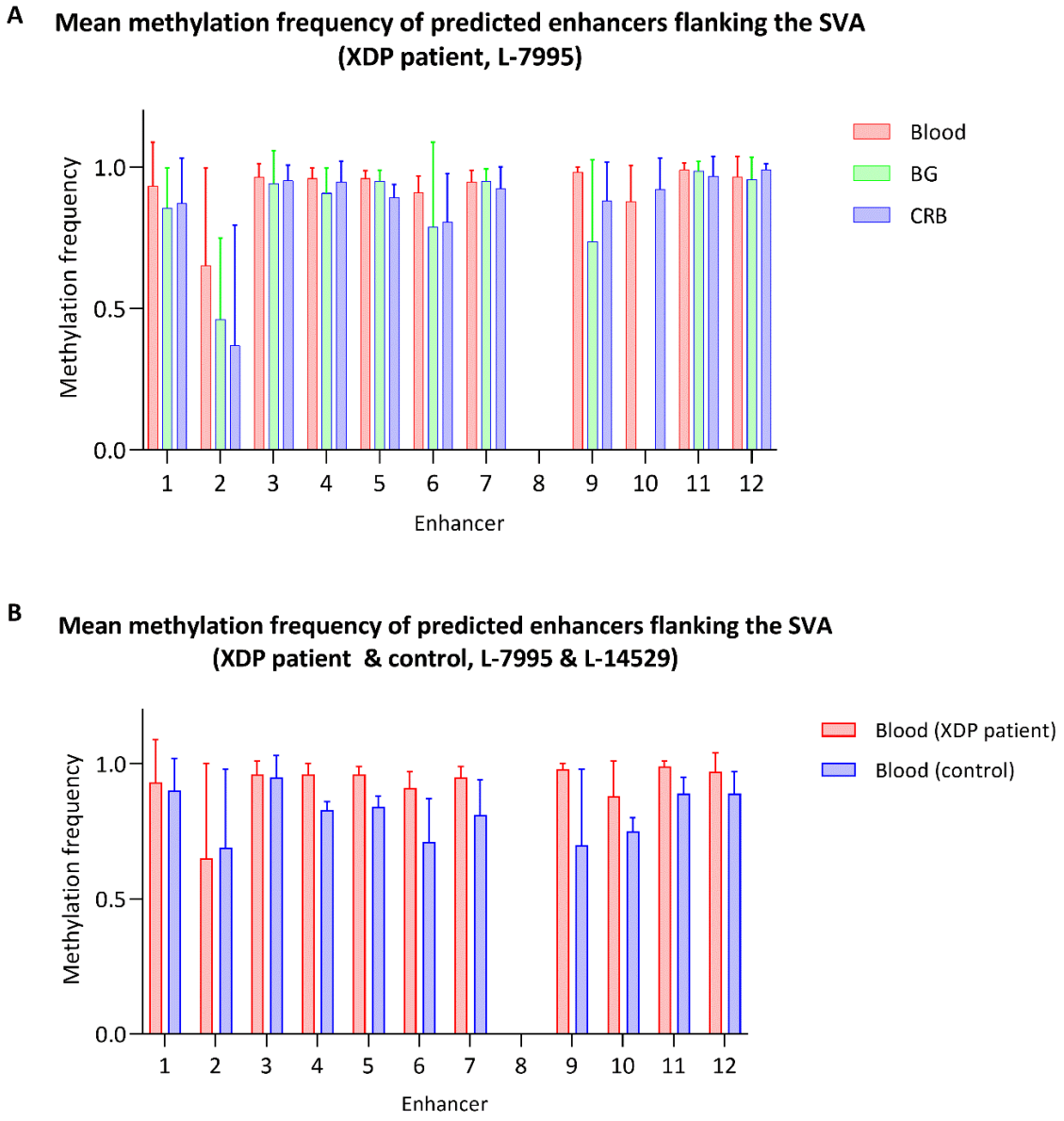

Figure 4. Methylation levels of the predicted enhancers flanking the *TAF1* SVA. The bar plot shows the CpG methylation frequency of predicted enhancer sites within the targeted region. DNA was derived from the blood, basal ganglia, and cerebellum of an XDP patient (**A**) or from the blood of a patient and a healthy control participant (**B**). The bars and whiskers represent the mean and upper limit of the standard deviation of the methylation frequency from the CpG sites within a predicted enhancer.

Discussion

In this study, we performed Nanopore single-molecule sequencing to examine the genetics and epigenetics of the full-length *TAF1* SVA insertion in patients with XDP. The novelty of our research lies in: (1) a new multiplexed workflow to quantify the *TAF1* SVA repeat number in patients that shows high concordance with fragment analysis which can be used as a cost-effective diagnostic tool in the future; (2) the exploration of novel variants within the SVA besides the repeat motif across 96 patients which has not been possible with older technologies and lastly; (3) the detection of direct CpG methylation across the full-length SVA and flanking regions up to 22 kb which incorporates 12 putative enhancer sites.

Examination of the Hexanucleotide Repeat Domain

Deep Nanopore sequencing (>5000X) of the GC-rich *TAF1* SVA insertion allowed better alignment accuracy for the low-complexity repetitive regions of the SVA than short-read sequencing approaches (De Coster et al. 2019). Our sequencing data analysis showed no evidence of any other genetic variability besides the repeat domain within the investigated locus. Thus, variability of AAO and disease severity primarily results from the hexanucleotide repeat number within the SVA (Bragg et al. 2017; Westenberger et al. 2019). As the sequence of the SVA is identical between patients except for the polymorphic repeat, this further validates the notion that the insertion of this repeat sequence in intron 32 of *TAF1* causes XDP (Nethisinghe et al. 2021).

The software tool NCRF has been specifically designed to explore repetitive sequences in noisy long-read sequencing data (Harris, Cechova, and Makova 2019). More specifically, the noise in the Nanopore signal trace is due to indel and homopolymer errors, reducing sequencing accuracy (Harris, Cechova, and Makova 2019). To decrease the noise further, we performed base-calling with the novel "super-high accuracy" model with improved read accuracy provided by the Nanopore software Guppy (v.5.0.11). Indeed, the repeat number resulting from the NCRF analysis was highly concordant with the results of the independent fragment analysis method. In concordance with the literature (Bragg et al. 2017; Westenberger et al. 2019), the repeat number detected by Nanopore sequencing was negatively associated with the AAO of patients with XDP as well, which further validates our workflow. Interestingly, the detected repeat number by NCRF was consistently 1–4 repeats longer than the number obtained from the fragment analysis.

The slightly larger repeat number detected with Nanopore sequencing could be due to deviations in the repeat pattern. We did not detect a consistently higher repeat number when we compared the results from the fast base-calling to the fragment analysis.

The higher concordance with the repeat number detected with fast base-calling could result from general repeat number detection of fragment analysis without the resolution of mismatches in the repeat motif. There was a noticeable increase in the frequency of deletions in the long-read data that require further exploration, as interruption of the repeat motif has also been reported for other repeat expansion diseases such as Friedreich's Ataxia or Huntington disease (Nethisinghe et al. 2021; Findlay Black et al. 2020). Therefore, further investigation of this issue, including single-nucleotide resolution of the repeat interruption, is required.

Methylation Status of the TAF1 SVA Insertion and Adjacent Enhancer Sites

Nanopore sequencing has been used to investigate DNA methylation in the context of other repeat expansions before (Ewing et al. 2020; Giesselmann et al. 2019). Recently, in the context of cancer, Nanopore sequencing has been used to simultaneously assess the genetics and epigenetics of transposable elements, including the CG-rich VNTR domain of SVAs, also known as "mobile CpG-island" (Ewing et al. 2020). An amplification-free Cas9-guided approach for Nanopore sequencing was introduced to maintain DNA methylation (Gilpatrick et al. 2020). This targeted approach has efficiently enriched repeat expansions causing frontotemporal dementia, amyotrophic lateral sclerosis, or fragile X syndrome (Giesselmann et al. 2019).

This study enriched the target region against the genomic background DNA, and methylation was maintained using a Cas9-guided approach. The coverage of the SVA insertion was specifically high; however, lower sequencing depth was obtained in the flanking regions. The variability in coverage could be due to the limitation of different targeting efficiencies of the crRNAs used in the Cas9 approach. We included only reads with sufficient alignment length to the reference sequence to counteract potential off-target effects.

We observed hypermethylation of the *TAF1* SVA insertion in concordance with the literature (Makino et al. 2007). Although the overall MF in all patient-derived samples was high, it was mildly reduced in the brain-derived samples. There have been speculations that neuron-specific expression of *TAF1* is reduced in patients with XDP (Makino et al. 2007).

However, recently published studies could not confirm a significant decrease in neuron-specific *TAF1* in XDP patients (Capponi et al. 2021; Valente and Bhatia 2018). In addition, it is unclear if the slight change in the methylation level in the brain-derived samples we observed could affect the expression level of *TAF1* and whether it would be relevant for the disease to develop. Furthermore, we did not specifically analyze the methylation level of neurons, which could be a perspective for future experiments.

There is the possibility that retrotransposon insertions introduce methylation changes into the flanking regions (Yates et al. 1999). There have been speculations that the hypermethylation of the SVA could also affect the methylation of adjacent cis-regulatory elements (Makino et al. 2007). The 12 predicted enhancer sites in the target region showed mostly comparable methylation levels in the blood and brain-derived samples, and only two enhancer sites showed a slight methylation decrease in the brain. Interestingly, lower methylation levels of the BG-derived sample, in particular, were also present in the SVA insertion (Figure 3B). This difference is most pronounced in the VNTR domain of the retrotransposon. Tissue-specific differences can explain the lower methylation observed in the BG and are not necessarily disease-related. Indeed, tissue-specific methylation patterns of transposable elements have been investigated with Nanopore sequencing (Ewing et al. 2020; I. Lee et al. 2020). The reduced methylation of intergenic subfamily SVAf insertion in the X-chromosome in tumor tissues is particularly interesting compared to non-tumor tissues (Ewing et al. 2020). In general, overall CpG methylation of brain tissue and peripheral tissues is highly correlated across participants. On the other hand, differences between the methylation level of the brain and peripheral tissues within a particular individual are possible, which can contribute to tissue-specific gene expression (Horvath et al. 2012; Smith et al. 2015; Braun et al. 2019).

This study aimed to establish a new workflow for direct CpG methylation detection using Nanopore sequencing. The study shows that detecting methylation across a large region is possible, including the *TAF1* SVA of 22 kb. Evidence indicates that DNA methylation can be a molecular mechanism in XDP pathogenesis. Due to the abolishment or introduction of CpG dinucleotides by disease-specific single-nucleotide changes (DSC), significant differences in methylation between XDP patients and controls at these positions have been reported, suggesting a potential effect on *TAF1* expression (Krause et al. 2020). These three reported DSCs are located ~700 kb away from the SVA. Our study focused on regions within and adjacent to the SVA insertion.

As the GC-content of the VNTR is high and the SVA is known to be hypermethylated, the change in the direct genetic environment could lead to altered methylation patterns within the SVA and flanking regions. As a pilot study, we observed overall lower methylation of the control sample compared to the XDP patient in the SVA flanking region, as well as in 10 out of the 12 predicted enhancer sites. The observed methylation differences in this study are only preliminary and should be interpreted cautiously. Another limitation of this study was the lack of brain-derived DNA from controls. Thus, more individuals and tissue types should be investigated using the Cas9-targeted approach across this 22 kb region. Still, our results show the utility of long reads in detecting the full-length SVA in *TAF1* in the context of XDP.

Conclusions

In this study, we present a straightforward and scalable long-read deep sequencing workflow to quantify the hexanucleotide repeat number of the *TAF1* SVA in patients with XDP. The high concordance of the results obtained from Nanopore sequencing with independent fragment analysis highlights the accuracy of our workflow. The long reads were also utilized to investigate variations within the SVA locus other than the repeat motif. As the sequence of the SVA locus was identical between patients besides the hexanucleotide repeat domain, our results further underline that the insertion of this repeat sequence is associated with the variability in AAO and expressivity in XDP. Lastly, an amplification-free Cas9 targeted enrichment of the SVA locus and the flanking regions allowed us to comprehensively assess the (epi-) genetics of the *TAF1* SVA locus.

Supplementary Materials

The following are available online at https://www.mdpi.com/article/10.3390/genes13010126/s1,

Author Contributions

Conceptualization, C.K., A.W. and J.T.; methodology, S.S., J.P., R.T. and C.J.R.; software, T.L. and J.L.; formal analysis, T.L., J.L., J.P., K.S., M.B., A.W. and J.T.; investigation, T.L., J.L., S.S., I.W., J.P., R.D.G.J., R.L.R., N.B., G.S., C.C.E.D., R.T., K.S., M.B., C.J.R., H.B., C.K. and A.W.; resources, R.D.G.J., R.L.R., N.B., G.S., C.C.E.D., H.B., C.K., A.W. and J.T.; data curation, S.S., J.P., R.T. and C.J.R.; writing—original draft preparation, T.L., J.L. and J.T.; writing—review and

editing, S.S., I.W., J.P., R.D.G.J., R.L.R., N.B., G.S., C.C.E.D., R.T., K.S., M.B., C.J.R., H.B., C.K. and A.W.; visualization, T.L. and J.L.; supervision, J.T.; project administration, C.K., A.W. and J.T.; funding acquisition, C.K., A.W. and J.T. All authors have read and agreed to the published version of the manuscript.

Funding

Funding has been obtained from the German Research Foundation ("ProtectMove"; FOR 2488, J.T., A.W., N.B., and C.K.; Germany's Excellence Strategy—EXC 22167-390884018, I.W. and H.B.), the Hermann and Lilly Schilling Foundation, the European Community (SysMedPD), the Canadian Institutes of Health Research (CIHR) (J.T.), Peter and Traudl Engelhorn Foundation, and the Center for XDP (CCXDP) at Massachusetts General Hospital (N.B.).

Institutional Review Board Statement

The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committees of the University of Lübeck, Germany, and the Metropolitan Medical Center, Manila, Philippines (REF: IRB-MMC #: 10-073; 31 August 2010).

Informed Consent Statement

Informed consent was obtained from all participants involved in the study.

Data Availability Statement

The data presented in this study are available on SRA (SAMN24775867-SAMN24775962, SAMN24115523-SAMN24115530). The bioinformatical commands to quantify the *TAF1* SVA $(CCCTCT)_n$ repeat length are described here: https://github.com/nanopol/xdp_sva/ (accessed on 10 December 2021).

Acknowledgments

The authors wish to thank the many patients and their families who volunteered, and the efforts of the many clinical teams involved. The authors further acknowledge computational support from the OMICS computing cluster at the University of Lübeck.

Conflicts of Interest

C.K. serves as a medical advisor for genetic testing reports to CENTOGENE GmbH in the fields of movement disorders and dementia, excluding Parkinson's disease, and is a member of the Scientific Advisory Boards of Retromer Therapeutics and Klink. A.W. provides consultancy services for research projects at CENTOGENE GmbH. NB served as a consultant for BridgeBio, Biomarin, and Centogene and is a member of the scientific advisory board of BridgeBio. He received speaker honoraria from Abbott and Zambon.

References

Aneichyk, T., W. T. Hendriks, R. Yadav, D. Shin, D. Gao, C. A. Vaine, R. L. Collins, et al. 2018. 'Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly'. *Cell* 172 (5): 897-909 e21. https://doi.org/10.1016/j.cell.2018.02.011.

Bragg, D. Cristopher, Kotchaphorn Mangkalaphiban, Christine A. Vaine, Nichita J. Kulkarni, David Shin, Rachita Yadav, Jyotsna Dhakal, et al. 2017. 'Disease Onset in X-Linked Dystonia-Parkinsonism Correlates with Expansion of a Hexameric Repeat within an SVA Retrotransposon in TAF1'. *Proceedings of the National Academy of Sciences of the United States of America* 114 (51): E11020–28. https://doi.org/10.1073/pnas.1712526114.

Braun, Patricia R., Shizhong Han, Benjamin Hing, Yasunori Nagahama, Lindsey N. Gaul, Jonathan T. Heinzman, Andrew J. Grossbach, et al. 2019. 'Genome-Wide DNA Methylation Comparison between Live Human Brain and Peripheral Tissues within Individuals'. *Translational Psychiatry* 9 (1): 47. https://doi.org/10.1038/s41398-019-0376-y.

Capponi, Simona, Nadja Stöffler, Ellen B. Penney, Karen Grütz, Sheikh Nizamuddin, Marit W. Vermunt, Bas Castelijns, et al. 2021. 'Dissection of TAF1 Neuronal Splicing and Implications for Neurodegeneration in X-Linked Dystonia-Parkinsonism'. *Brain Communications* 3 (4): fcab253. https://doi.org/10.1093/braincomms/fcab253.

De Coster, Wouter, Peter De Rijk, Arne De Roeck, Tim De Pooter, Svenn D'Hert, Mojca Strazisar, Kristel Sleegers, and Christine Van Broeckhoven. 2019. 'Structural Variants Identified by Oxford Nanopore PromethION Sequencing of the Human Genome'. *Genome Research* 29 (7): 1178–87. https://doi.org/10.1101/gr.244939.118.

Domingo, Aloysius, Ana Westenberger, Lillian V. Lee, Ingrid Brænne, Tian Liu, Inga Vater, Raymond Rosales, et al. 2015. 'New Insights into the Genetics of X-Linked Dystonia-Parkinsonism (XDP, DYT3)'. *European Journal of Human Genetics: EJHG* 23 (10): 1334–40. https://doi.org/10.1038/ejhg.2014.292.

Ewing, Adam D., Nathan Smits, Francisco J. Sanchez-Luque, Jamila Faivre, Paul M. Brennan, Sandra R. Richardson, Seth W. Cheetham, and Geoffrey J. Faulkner. 2020. 'Nanopore Sequencing Enables Comprehensive Transposable Element Epigenomic Profiling'. *Molecular Cell* 80 (5): 915-928.e5. https://doi.org/10.1016/j.molcel.2020.10.024.

Findlay Black, H., G. E. B. Wright, J. A. Collins, N. Caron, C. Kay, Q. Xia, L. Arning, et al. 2020. 'Frequency of the Loss of CAA Interruption in the HTT CAG Tract and Implications for Huntington Disease in the Reduced Penetrance Range'. *Genet Med* 22 (12): 2108–13. https://doi.org/10.1038/s41436-020-0917-z.

Giesselmann, Pay, Björn Brändl, Etienne Raimondeau, Rebecca Bowen, Christian Rohrandt, Rashmi Tandon, Helene Kretzmer, et al. 2019. 'Analysis of Short Tandem Repeat Expansions and Their Methylation State with Nanopore Sequencing'. *Nature Biotechnology* 37 (12): 1478–81. https://doi.org/10.1038/s41587-019-0293-x.

Gilpatrick, Timothy, Isac Lee, James E. Graham, Etienne Raimondeau, Rebecca Bowen, Andrew Heron, Bradley Downs, Saraswati Sukumar, Fritz J. Sedlazeck, and Winston Timp. 2020. 'Targeted Nanopore Sequencing with Cas9-Guided Adapter Ligation'. *Nature Biotechnology* 38 (4): 433–38. https://doi.org/10.1038/s41587-020-0407-5.

Harris, R. S., M. Cechova, and K. D. Makova. 2019. 'Noise-Cancelling Repeat Finder: Uncovering Tandem Repeats in Error-Prone Long-Read Sequencing Data'. *Bioinformatics* 35 (22): 4809–11. https://doi.org/10.1093/bioinformatics/btz484.

Horvath, Steve, Yafeng Zhang, Peter Langfelder, René S. Kahn, Marco P. M. Boks, Kristel van Eijk, Leonard H. van den Berg, and Roel A. Ophoff. 2012. 'Aging Effects on DNA Methylation Modules in Human Brain and Blood Tissue'. *Genome Biology* 13 (10): R97. https://doi.org/10.1186/gb-2012-13-10-r97.

Krause, Christin, Susen Schaake, Karen Grütz, Helen Sievert, Charles Jourdan Reyes, Inke R. König, Björn-Hergen Laabs, et al. 2020. 'DNA Methylation as a Potential Molecular Mechanism in X-Linked Dystonia-Parkinsonism'. *Movement Disorders: Official Journal of the Movement Disorder Society* 35 (12): 2220–29. https://doi.org/10.1002/mds.28239.

Lee, Isac, Roham Razaghi, Timothy Gilpatrick, Michael Molnar, Ariel Gershman, Norah Sadowski, Fritz J. Sedlazeck, Kasper D. Hansen, Jared T. Simpson, and Winston Timp. 2020. 'Simultaneous Profiling of Chromatin Accessibility and Methylation on Human Cell Lines with Nanopore Sequencing'. *Nature Methods* 17 (12): 1191–99. https://doi.org/10.1038/s41592-020-01000-7.

Lee, L. V., E. L. Munoz, K. T. Tan, and M. T. Reyes. 2001. 'Sex Linked Recessive Dystonia Parkinsonism of Panay, Philippines (XDP)'. *Molecular Pathology: MP* 54 (6): 362–68.

Lee, L. V., F. M. Pascasio, F. D. Fuentes, and G. H. Viterbo. 1976. 'Torsion Dystonia in Panay, Philippines'. *Advances in Neurology* 14:137–51.

Makino, Satoshi, Ryuji Kaji, Satoshi Ando, Maiko Tomizawa, Katsuhito Yasuno, Satoshi Goto, Shinnichi Matsumoto, et al. 2007. 'Reduced Neuron-Specific Expression of the TAF1 Gene Is Associated with X-Linked Dystonia-Parkinsonism'. *American Journal of Human Genetics* 80 (3): 393–406. https://doi.org/10.1086/512129.

Montague, Tessa G., José M. Cruz, James A. Gagnon, George M. Church, and Eivind Valen. 2014. 'CHOPCHOP: A CRISPR/Cas9 and TALEN Web Tool for Genome Editing'. *Nucleic Acids Research* 42 (Web Server issue): W401-407. https://doi.org/10.1093/nar/gku410.

Nethisinghe, Suran, Maheswaran Kesavan, Heather Ging, Robyn Labrum, James M. Polke, Saiful Islam, Hector Garcia-Moreno, et al. 2021. 'Interruptions of the FXN GAA Repeat Tract Delay the Age at Onset of Friedreich's Ataxia in a Location Dependent Manner'. *International Journal of Molecular Sciences* 22 (14): 7507. https://doi.org/10.3390/ijms22147507.

Pauly, Martje G., Marta Ruiz López, Ana Westenberger, Gerard Saranza, Norbert Brüggemann, Anne Weissbach, Raymond L. Rosales, et al. 2020. 'Expanding Data Collection for the MDSGene Database: X-Linked Dystonia-Parkinsonism as Use Case Example'. *Movement Disorders: Official Journal of the Movement Disorder Society* 35 (11): 1933–38. https://doi.org/10.1002/mds.28289.

Pfaff, Abigail L., Vivien J. Bubb, John P. Quinn, and Sulev Koks. 2021. 'Reference SVA Insertion Polymorphisms Are Associated with Parkinson's Disease Progression and Differential Gene Expression'. *NPJ Parkinson's Disease* 7 (1): 44. https://doi.org/10.1038/s41531-021-00189-4.

Rakovic, Aleksandar, Jonathan Ziegler, Christoph U. Mårtensson, Jannik Prasuhn, Katharina Shurkewitsch, Peter König, Henry L. Paulson, and Christine Klein. 2019. 'PINK1-Dependent Mitophagy Is Driven by the UPS and Can Occur Independently of LC3 Conversion'. *Cell Death and Differentiation* 26 (8): 1428–41. https://doi.org/10.1038/s41418-018-0219-z.

Reyes, Charles Jourdan, Björn-Hergen Laabs, Susen Schaake, Theresa Lüth, Raphaela Ardicoglu, Aleksandar Rakovic, Karen Grütz, et al. 2021. 'Brain Regional Differences in Hexanucleotide Repeat Length in X-Linked Dystonia-Parkinsonism Using Nanopore Sequencing'. *Neurology. Genetics* 7 (4): e608. https://doi.org/10.1212/NXG.0000000000000608.

Rosales, Raymond L. 2010. 'X-Linked Dystonia Parkinsonism: Clinical Phenotype, Genetics and Therapeutics'. *Journal of Movement Disorders* 3 (2): 32–38. https://doi.org/10.14802/jmd.10009.

Smith, Alicia K., Varun Kilaru, Torsten Klengel, Kristina B. Mercer, Bekh Bradley, Karen N. Conneely, Kerry J. Ressler, and Elisabeth B. Binder. 2015. 'DNA Extracted from Saliva for Methylation Studies of Psychiatric Traits: Evidence Tissue Specificity and Relatedness to Brain'. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics: The Official Publication of the International Society of Psychiatric Genetics* 168B (1): 36–44. https://doi.org/10.1002/ajmg.b.32278.

Valente, Enza Maria, and Kailash P. Bhatia. 2018. 'Solving Mendelian Mysteries: The Non-Coding Genome May Hold the Key'. *Cell* 172 (5): 889–91. https://doi.org/10.1016/j.cell.2018.02.022.

Wang, Hui, Jinchuan Xing, Deepak Grover, Dale J. Hedges, Kyudong Han, Jerilyn A. Walker, and Mark A. Batzer. 2005. 'SVA Elements: A Hominid-Specific Retroposon Family'. *Journal of Molecular Biology* 354 (4): 994–1007. https://doi.org/10.1016/j.jmb.2005.09.085.

Watkins, W. Scott, Julie E. Feusier, Jainy Thomas, Clement Goubert, Swapon Mallick, and Lynn B. Jorde. 2020. 'The Simons Genome Diversity Project: A Global Analysis of Mobile Element Diversity'. *Genome Biology and Evolution* 12 (6): 779–94. https://doi.org/10.1093/gbe/evaa086.

Westenberger, A., C. J. Reyes, G. Saranza, V. Dobricic, H. Hanssen, A. Domingo, B. H. Laabs, et al. 2019. ‘A Hexanucleotide Repeat Modifies Expressivity of X-Linked Dystonia Parkinsonism’. *Ann Neurol* 85 (6): 812–22. https://doi.org/10.1002/ana.25488.

Yates, P. A., R. W. Burman, P. Mummaneni, S. Krussel, and M. S. Turker. 1999. ‘Tandem B1 Elements Located in a Mouse Methylation Center Provide a Target for de Novo DNA Methylation’. *The Journal of Biological Chemistry* 274 (51): 36357–61. https://doi.org/10.1074/jbc.274.51.36357.

Objective 1B: To investigate the *TAF1* SVA retrotransposon using third-generation sequencing technologies to determine repeat length and potential interruptions

Long-read sequencing enables the characterization of the whole repeat tract in repeat expansion disorders. Interruptions in the repeat tract can benefit the individuals by delaying the disease onset. As we discovered in the first study, repeat interruptions occur in X-linked dystonia-parkinsonism, but where and how they are present in the repeat tract is still unclear. Many repeat interruptions are at the 5'end of the repeat expansion, but there are also examples in the middle or at the 3'end. Repeat interruptions typically affect only one repeat unit, i.e., in Huntington's disease, only the sequence of one CAG repeat is changed to CAA.

In this study, titled "Mosaic divergent repeat interruptions in XDP influence repeat stability and disease onset", we investigated how the interruptions occur in the *TAF1* SVA retrotransposon insertion and if these interruptions can influence the repeat stability and disease onset. We included 202 patients with XDP in this study. We included two brain regions from one patient with XDP to assess the somatic instability of the repeat interruptions. We performed a PCR amplification for all individuals' SVA retrotransposon in *TAF1*. I additionally enriched the SVA retrotransposon for the individual with the brain regions by a Cas9-mediated approach.

The bioinformatic pipeline was the same as in the first study, but additional quality filters were applied. The first filter was for the quality of the reads; only reads with a Phred score Q greater than twelve were included. The second filter was for the sequencing depth; only individuals with a coverage of over 1500X were included. I have identified three specific positions where single nucleotides were deleted, but not every read in one individual has these three deletions. Therefore, we calculated the frequency of every possible combination of the deletions. With these frequencies for the combinations, we designed linear regression models to investigate the influence on the repeat stability and disease onset. We also adjusted the regression model for the age of the individuals, SVA repeat number, or other modifiers like single-nucleotide polymorphisms. In this study, I performed the bioinformatic pipeline for the repeat number and interruption detection. I have identified the three specific positions at the 5'end of the repeat tract and determined the frequency of each combination of these different positions. Finally, I reviewed the first version of the manuscript and created the first figure.

This study revealed novel mosaic divergent repeat interruptions affecting both motif length and sequence (DRILS) of the canonical motif polarized within the SINE-VNTR-Alu$(AGAGGG)_n$ repeat, which involves the repeat stability.

Title

Mosaic divergent repeat interruptions in XDP influence repeat stability and disease onset

Authors

Joanne Trinh[1*], Theresa Lüth[1], Susen Schaake[1], Björn-Hergen Laabs[2], Kathleen Schlüter[1], **Joshua Laß[1]**, Jelena Pozojevic[1], Ronnie Tse[1], Inke König[2], Roland Dominic Jamora[3], Raymond L. Rosales[4], Norbert Brüggemann[1,5], Gerard Saranza[6], Cid Czarina E. Diesta[7], Frank J. Kaiser[8,9], Christel Depienne[8], Christopher E. Pearson[10,11], Ana Westenberger[1] and Christine Klein[1]

[1] Institute of Neurogenetics, University of Lübeck and University Hospital Schleswig-Holstein, Lübeck, Germany

[2] Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany

[3] Department of Neurosciences, College of Medicine—Philippine General Hospital, University of the Philippines Manila, Manila, Philippines

[4] Department of Neurology and Psychiatry, University of Santo Tomas and the CNS-Metropolitan Medical Center, Manila, Philippines

[5] Department of Neurology, University of Lübeck, Lübeck, Germany

[6] Section of Neurology, Department of Internal Medicine, Chong Hua Hospital, Cebu, Philippines

[7] Department of Neurosciences, Movement Disorders Clinic, Makati Medical Center, Makati City, Philippines

[8] Institute for Human Genetics at the University Hospital Essen, Essen, Germany

[9] Center for Rare Diseases (Essenser Zentrum für Seltene Erkrankungen—EZSE) at the University Hospital Essen, Essen, Germany

[10] Program of Genetics and Genome Biology, The Hospital for Sick Children, The Peter Gilgan Centre for Research and Learning, Toronto, Canada

[11] University of Toronto, Program of Molecular Genetics, Toronto, Canada

*Correspondence author

Published in *Brain*, 2023, *146*(3), 1075-1082; DOI: 10.1093/brain/awac160

Abstract

While many genetic causes of movement disorders have been identified, modifiers of disease expression are largely unknown. X-linked dystonia-parkinsonism (XDP) is a neurodegenerative disease caused by a SINE-VNTR-Alu(AGAGGG)$_n$ retrotransposon insertion in *TAF1*, with a polymorphic (AGAGGG)$_n$ repeat. Repeat length and variants in *MSH3* and *PMS2* explain ~65% of the variance in age at onset (AAO) in XDP. However, additional genetic modifiers, such as repeat interruptions, are conceivably at play in XDP.

Long-read nanopore sequencing of PCR amplicons from XDP patients (n= 202) was performed to assess potential repeat interruption and instability. Repeat-primed PCR and Cas9-mediated targeted enrichment confirmed the presence of identified divergent repeat motifs.

In addition to the canonical pure SINE-VNTR-Alu-5′-(AGAGGG)$_n$, we observed a mosaic of divergent repeat motifs that polarized at the beginning of the tract, where the divergent repeat interruptions varied in motif length by having one, two, or three nucleotides fewer than the hexameric motif, distinct from interruptions in other disease-associated repeats, which match the lengths of the canonical motifs. All divergent configurations occurred mosaically in two investigated brain regions (basal ganglia, cerebellum) and in blood-derived DNA from the same patient. The most common divergent interruption was AGG [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGG*(AGAGGG)$_n$], similar to the pure tract, followed by AGGG [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGGG*(AGAGGG)$_n$], at median frequencies of 0.425 (IQR: 0.42–0.43) and 0.128 (IQR: 0.12–0.13), respectively. The mosaic AGG motif was not associated with repeat number (estimate= −3.8342, P= 0.869). The mosaic pure tract frequency was associated with repeat number (estimate= 45.32, P= 0.0441) but not AAO (estimate= −41.486, P= 0.378). Importantly, the mosaic frequency of the AGGG negatively correlated with repeat number after adjusting for age at sampling (estimate= −161.09, P= 3.44× 10−5). When including the XDP-relevant *MSH3*/*PMS2* modifier single-nucleotide polymorphisms into the model, the mosaic AGGG frequency was associated with AAO (estimate= 155.1063, P= 0.047); however, the association dissipated after including the repeat number (estimate= −92.46430, P= 0.079).

We reveal novel mosaic divergent repeat interruptions affecting both motif length and sequence (DRILS) of the canonical motif polarized within the SINE-VNTR-Alu$(AGAGGG)_n$ repeat. Our study illustrates: (i) the importance of somatic mosaic genotypes; (ii) the biological plausibility of multiple modifiers (both germline and somatic) that can have additive effects on repeat instability; and (iii) that these variations may remain undetected without assessment of single molecules.

Introduction

While multiple genetic causes of movement disorders have been identified in the past decade, disease modifiers are still largely unknown for most conditions (Posey et al. 2019). Individual patients carrying the same pathogenic variant may have variable expressivity of the disease, including variable age at onset (AAO), severity, and clinical manifestations. Thus, in addition to the pathogenic variant, there are genetic modifiers influencing expressivity and onset. One emerging example of this broader concept is X-linked dystonia-parkinsonism (XDP), a neurodegenerative movement disorder endemic to the Philippines and first described in 1976 (Lee et al. 1976; Pauly et al. 2020).

XDP is one of a large, ever-growing class of >60 diseases caused by unstable tandem repeats and, in particular, a subclass of inserted repeats (Gall-Duncan et al. 2022). Adulthood-onset dystonic movements and Parkinsonism characterize XDP due to striatal volume loss as a result of an insertion of the retrotransposon SINE-VNTR-Alu (SVA) in intron 32 of the *TAF1* (*TATA-binding protein-associated factor 1*) gene (Pauly et al. 2020; Domingo et al. 2015). A hexanucleotide repeat domain within the SVA consists of the repeat sequence $(AGAGGG)_n$ (Bragg et al. 2017). This hexanucleotide repeat domain varies in numbers ranging from 30 to 55 and is a potent genetic modifier of AAO.

To date, four putative modifiers of XDP expressivity are associated with AAO and/or disease severity (Bragg et al. 2017; Laabs et al. 2021; Westenberger et al. 2019). These modifiers are the length of the hexanucleotide repeat polymorphism and modifiers of AAO related to variants in the DNA repair genes *MSH3* and *PMS2* (Laabs et al. 2021; Westenberger et al. 2019). Both types of modifiers are characteristic of repeat expansion disorders in which expanded repeats of various lengths may be transcribed.

Genetic modifiers, such as DNA repair genes, are inherited in the germline. Mosaic modifiers, which exist in every patient, necessarily originate as post-zygotic mutations but have not been studied extensively. We previously described the presence of somatic repeat length mosaicism in XDP, with a higher number of repeats detected in the cerebellum and basal ganglia compared to the same patient's blood (Reyes et al. 2021). The mosaic lengths between tissues arose as post-zygotic mutations. In this study, we identified novel mosaic repeat motif patterns that deviate from the known hexanucleotide repeat motif both in motif length and sequence. We investigated whether they act as new genetic modifiers of repeat instability and AAO of this neurodegenerative disorder.

Material and Methods

Patient demographics

The study was approved by the Ethics Committees of the University of Lübeck, Germany, and at the Metropolitan Medical Center, Manila, Philippines. To analyze genomic variants within the SVA and detect variations of the hexanucleotide repeat domain, n = 202 patients with XDP were investigated. They included different brain regions (basal ganglia, cerebellum) and blood-derived DNA from one patient with XDP. As XDP follows an X-linked recessive inheritance pattern, all patients were male. The mean AAO was 41.93 (SD = ±8.56) years, and the mean age at examination (AAE) was 47.5 (SD = ±9.84) years (Supplementary Table 1).

DNA extraction

All DNA was extracted from the Blood and Cell Culture DNA Midi kit (Qiagen). A summary of experimental procedures is described in Supplementary Table 1.

Nanopore sequencing of PCR amplicon

As previously described, long-range PCR was performed for the SVA (3.2 kb) (Reyes et al. 2021). The master mix and the amplification conditions are presented in Supplementary Tables 2 and 3, respectively (XDP-16153 F: 5′-GTTCCATTGTGTGGTTGTACCAGCGTTTGTTC-3′, XDP-19345R: 5′-CACATGAAAAGATGCCC AACATCATTAGCCATTAG-3′) (Kawarai et al. 2013).

The libraries were prepared with the Ligation Sequencing Kit (LSK109), and the samples were barcoded with the Native 96 Barcoding Kit (EXP-NBD196), using 400 ng of each patient-derived PCR product. Subsequently, the library was sequenced on a GridION (R9.4.1 flow cell), and in total, three flow cells, with one library per flow cell, were used for the sequencing of the multiplexed PCR amplicons.

Nanopore sequencing of Cas9-mediated targeted enrichment

In addition, the *TAF1* SVA was enriched by an amplification-free Cas9-mediated approach, as previously described (Reyes et al. 2021). For Cas9 enrichment, crRNAs were designed with ChopChop (https://chopchop.cbu.uib.no). Two crRNAs were used upstream of the *TAF1* SVA insertion (crRNA 1 and 2), and two crRNAs were used downstream (crRNA 3 and 4) for a 5.5 kb product specifically around the SVA (Supplementary Table 4). Blood-, basal ganglia-, and cerebellum-derived DNA from one patient was used for the Cas9-enrichment. The libraries were generated with the Ligation Sequencing Kit (LSK109), and no barcoding was performed. The sequencing was performed on the MinION (R9.4.1 flow cell). For the blood-derived DNA, four flow cells were loaded with five libraries (4 × 5 µg and 1 × 1 µg of input DNA).

For the basal ganglia-derived sample, four flow cells were loaded with five libraries (3 × 3 µg, 1 × 2 µg, and 1 × 1 µg of input DNA). Lastly, for the cerebellum-derived sample, we used three flow cells and five libraries (4 × 5 µg and 1 × 10 µg of input DNA).

Detection of repeats

Base-calling was performed with the most updated Guppy version 5.0.11. The super-accurate model (dna_r9.4.1_450bps_sup.cfg) was used to detect the repeat length. All reads were mapped to the reference sequence with the software Minimap2 (v2.17) (Supplementary Fig. 1). Samtools (v1.9) was used for coverage determination and filtering (>1500×). We filtered for Phred score Q > 12. Motif mismatch detection was achieved with 'Noise-canceling repeat finder' (NCRF) (v1.01.02) (Harris et al. 2019). The detailed commands are listed in the Supplementary material, and more information and the corresponding reference files used for the alignment are provided at: https://github.com/nanopol/xdp_sva/.

Repeat-primed PCR

Repeat-primed PCR (RP-PCR) with a FAM-tagged primer was performed for validation. The master mix and the amplification conditions are presented in Supplementary Tables 5 and 6. Supplementary Fig. 2 shows a schematic of the primer binding locations in the *TAF1* intron 32, and the hexanucleotide repeat domain of the *TAF1* SVA. For the fragment analysis, a total of 1 µl of the RP PCR products with 10.7 µl HiDi Formamide (Applied Biosystems) and 0.3 µl GeneScan™ 600 LIZ™ Dye Size Standard (Applied Biosystems) were used for capillary electrophoresis on an ABI 3500×L Genetic Analyzer (Applied Biosystems). The output was analyzed using GeneMapper software (version 4.1, Applied Biosystems).

Statistical analysis

Frequencies of deletions were estimated using NCRF, and Spearman's correlation coefficient was employed to assess the correlation and the corresponding *P*-values reported. Box plots were used to show the distribution. We aimed to adjust the mosaic frequencies for the age at sampling and designed a linear regression model predicting the mosaic AGGG frequency by age at sampling. Age impacts the mosaic AGGG frequency (Supplementary Table 7). Thus, we obtained the residuals of the regression model and used them as an adjusted predictor for repeat number or AAO. Regression models were used to assess the correlation between AAO and the frequency of the most common DRILS, AGG [5′-SINE-VNTR-Alu$(AGAGGG)_2$*AGG*$(AGAGGG)_n$], adjusted for age, three single-nucleotide polymorphisms (SNPs) in *MSH3* and *PMS2*, or for three SNPs in *MSH3* and *PMS2* and the SVA repeat number.

Data availability

The data presented in this study are available on SRA (SAMN24775867-SAMN24775962, SAMN24115523-SAMN24115530). The bioinformatic commands to quantify the *TAF1* SVA $(AGAGGG)_n$ repeat length are described at: https://github.com/nanopol/xdp_sva/.

Results

Deep sequencing of the PCR amplicon of the 3.2 kb *TAF1* SVA region in blood-derived DNA with nanopore yielded a mean coverage of 10,915X per sample (SD = ±8207) (Supplementary Fig. 1 and Table 1). We detected a prominent occurrence of deletions in every patient (n = 202), with a mean frequency of 0.97 (SD = ±0.113) at the beginning of the repeat tract, consistently present on the plus- and minus-strands in all patients (Fig. 1A and B). At the single nucleotide resolution, three deletions (deletions 1, 2, 3) were found at the 5′ end at positions 11, 14, and 17 of the repeat motif (Fig. 1C). These detected deletions lead to divergent repeat motifs that occur at the second and/or third motif of the expanded $(AGAGGG)_n$ tandem repeat (Fig. 2A).

The divergent motifs were detected as multiple combinations and at various frequencies in every patient, indicating somatic mosaicism (Fig. 2A). The most frequently detected in all of the analyzed patients in this study was the divergent repeat AGG, with the pattern $(AGAGGG)_2AGG(AGAGGG)_n$, having a median frequency 0.425 (IQR: 0.42–0.43). This occurred near equal frequency to the pure uninterrupted $(AGAGGG)_n$ tract. The second most detected divergent motif was AGGG, where the resulting repeat motif pattern was $(AGAGGG)_2AGGG(AGAGGG)_n$, with a median frequency of 0.128 (IQR: 0.12–0.13) (Fig. 2B). Other divergent repeat motifs and patterns were detected at lower frequencies (Fig. 2B). It is noteworthy that each of the divergent repeat motifs shifted the hexameric repeat frame of the repeat tract. Most change the trinucleotide repeat frame. However, the $(AGAGGG)_2AGG(AGAGGG)_n$ divergent repeat motif retained the trinucleotide frame. Curiously, the AGG motif was not associated with AAO. The significance, if any, that this was the most common form, nearly equal to the pure tract, and its retention of the trinucleotide frame, is unknown.

Further genetic validation was performed using repeat-primed PCR and Cas9-targeted enrichment of the divergent repeat motifs. Using repeat-primed PCR targeting the divergent repeat motif patterns, we observed a signal indicating divergent motifs at the beginning 5′ region of the AGAGGG repeat tract (Supplementary Fig. 2A–D).

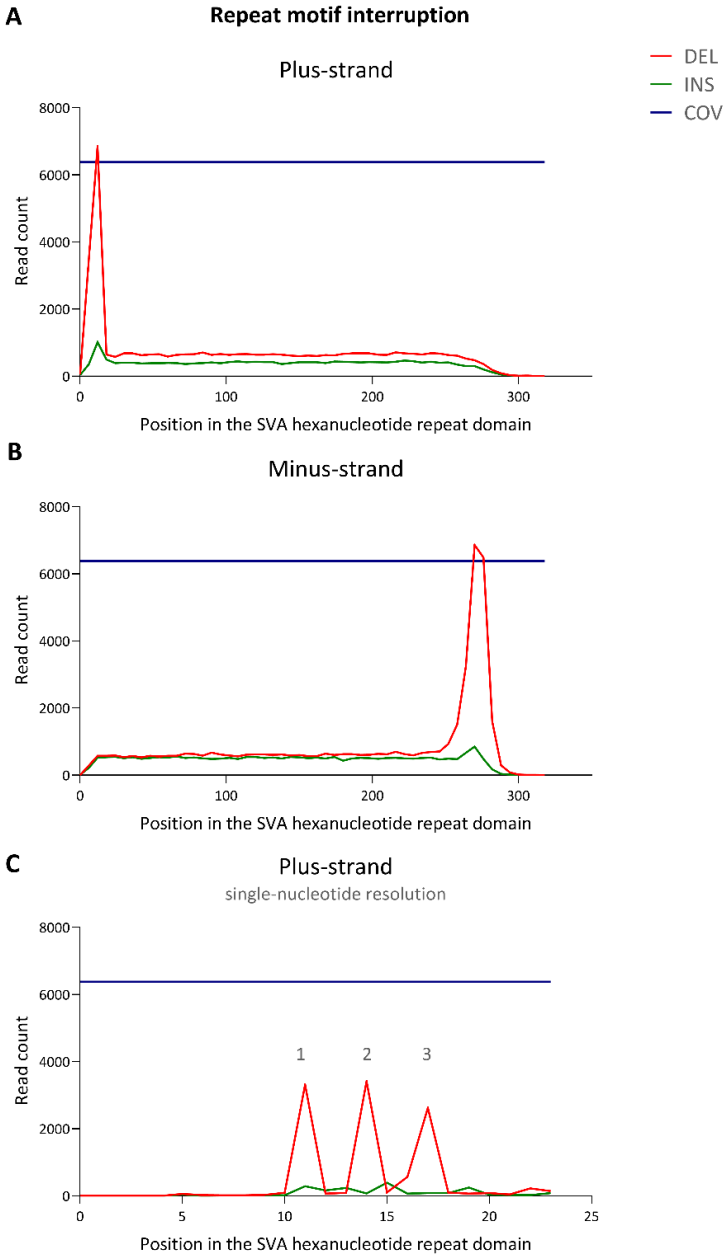

Figure 1. Detection of divergent repeat interruptions within the *TAF1* SVA hexanucleotide repeat domain. (**A**) Plus-strand reads show deletions or insertions within the SVA hexanucleotide repeat domain. (**B**) Minus-strand reads show deletions or insertions within the SVA hexanucleotide repeat domain. (**C**) Single-nucleotide resolution of the deletions detected at the 5′ end of the SVA hexanucleotide repeat domain. COV = coverage (number of reads covering the *TAF1* SVA hexanucleotide repeat domain); DEL = deletion; INS = insertion.

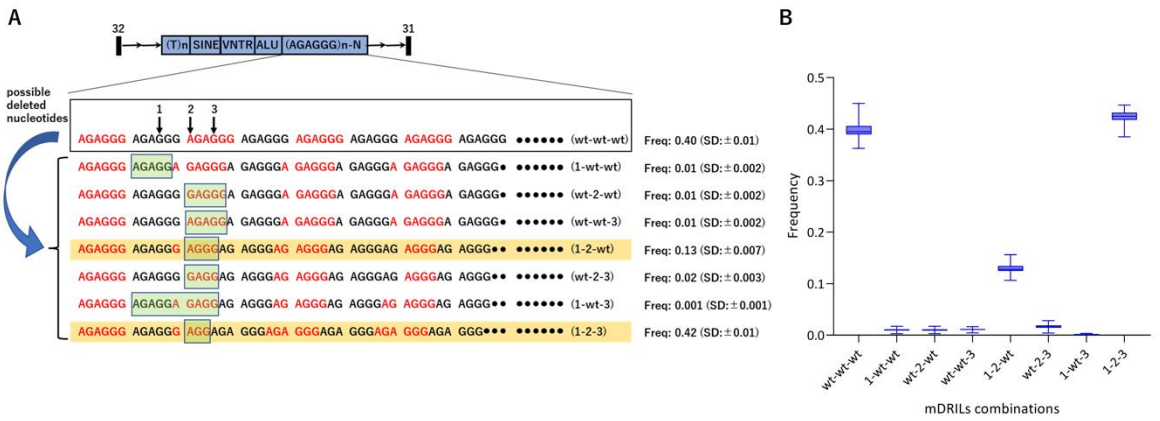

Figure 2. Overview of mDRILS combinations within the *TAF1* SVA hexanucleotide repeat motif and the corresponding detected frequencies. (**A**) Deleted nucleotides change the size and sequence of a single repeat unit of the hexameric tract. Boxed outline: change in repeat unit. Shaded highlight: common haplotypes. Black dots: indicating shifts in the repeat tract frame. The mosaic frequency of each DRIL is displayed alongside the interrupted repeat units. (**B**) The box plot shows frequencies of the combinations of deletions. The line and box represent median and interquartile range, respectively, and the whiskers represent the range.

Table 1. Residuals and coefficients of the linear model predict repeat numbers with mosaic AGGG frequency adjusted for age.

Residuals				
Min	**1Q**	**Median**	**3Q**	**Max**
-10.2224	-2.6164	-0.3089	2.2163	13.0807
Coefficients				
	Estimate	**SE**	**T value**	**Pr(>|t|)**
Intercept	41.4221	0.2804	147.70	<2e-16***
$beta_{res}$	-161.0907	37.9957	-4.24	3.44e-05***

Legend: 1Q = first quartile; 3Q = third quartile; Max = maximum; Min = minimum; res = residuals; RN = repeat number; SE = standard error; res=residuals
*Model: RN = Intercept + $beta_{res}$ * res.*

Lastly, Cas9-targeted enrichment was performed to avoid errors from PCR amplification as another validation. We found comparable frequencies of somatic divergent repeat motifs in the blood, basal ganglia, and cerebellum-derived DNA from the same patient (Fig. 3A–C). To discern repeat size and mosaic status on the same DNA fragments via long-read sequencing, we investigated the mosaicism by directly measuring how much of this somatic instability is derived from the canonical hexamer repeat tract or AGGG motif repeat tract.

We observed that there is more variability in repeat number [range: 21–53 (blood); 19–68 (basal ganglia); 19–60 (cerebellum)] when looking at the canonical hexamer repeat tract (wt-wt-wt) compared to the AGGG motif (1-2-wt) [range: 41–52 (blood); 47–84 (basal ganglia); 24–53 (cerebellum)]. Furthermore, there was a higher quartile coefficient of dispersion for the canonical repeat tract (range: 0.042–0.054) compared to the AGGG motif repeat tract (range: 0.019–0.031) (Fig. 3D and E).

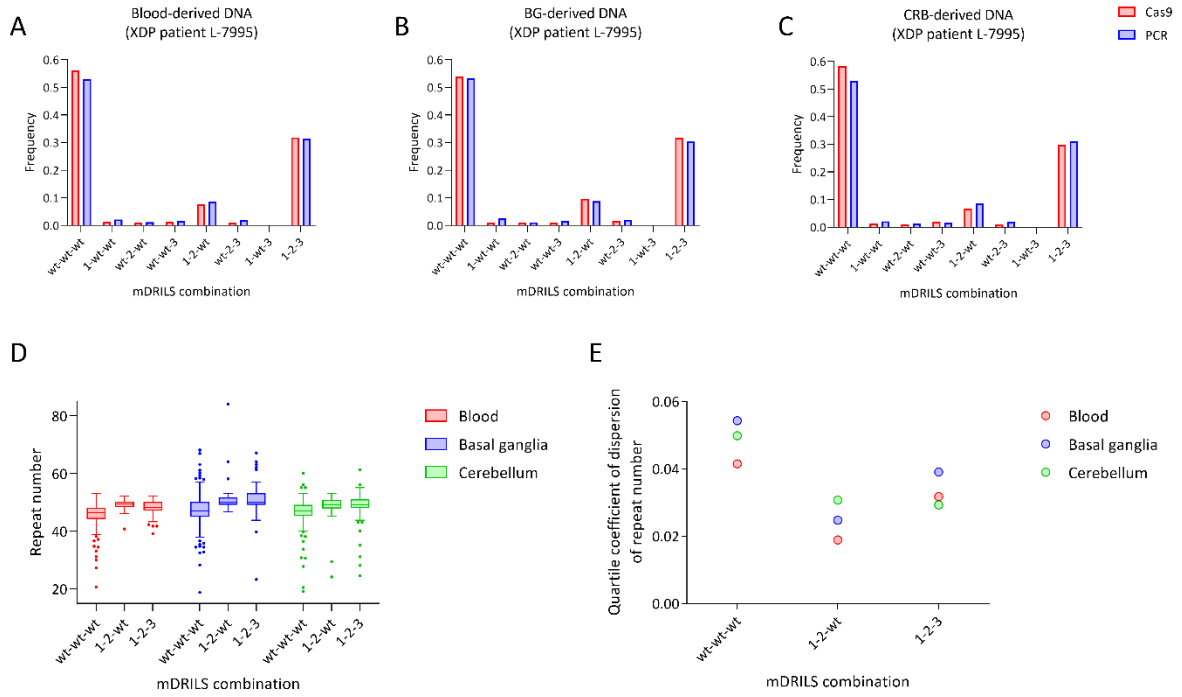

Figure 3. Comparison of mDRILS frequencies detected in blood- and brain-derived DNA, enriched for the SVA insertion by PCR or amplification-free Cas9-enrichment. The bars represent a single frequency value for each mDRILS detected from (**A**) blood-derived DNA, (**B**) basal ganglia (BG)-derived DNA, and (**C**) cerebellum (CRB)-derived DNA. (**D**) Repeat number distribution of each mDRILS, and (**E**) quartile coefficient of dispersion of each mDRILS. The bar charts show the detected frequencies of the different combinations of mDRILS. Box plot centre line represents median and box limits are upper and lower quartiles.

We focused our further analyses on the $(AGAGGG)_2AGGG(AGAGGG)_n$ and the $(AGAGGG)_2AGG(AGAGGG)_n$ repeat combinations as they were the most frequently detected. We investigated these divergent repeat motifs and their influence on repeat tract length (i.e., repeat number). The frequency of the AGGG negatively correlated with repeat tract length ($r = -0.48$, $P = 9.5 \times 10^{-13}$): the higher the frequency of AGGG, the shorter the repeat tract (Fig. 4A). This same effect was not observed for the AGG (Fig. 4B). The AGGG positively correlated with AAO ($r = 0.34$, $P = 9.5 \times 10^{-7}$), whereas the AGG did not show a correlation with AAO (Fig. 4C and D). Since somatic mosaicism may change with age, which may confound analyses, we adjusted for age at sampling. After adjusting for age at sampling, using the residuals of the regression model as an adjusted predictor for repeat number, we found that the mosaic AGGG frequency in blood DNAs was associated with repeat number (estimate = −161.09, $P = 3.44 \times 10^{-5}$) (Table 1). After adjusting for age at sampling, using the residuals of the regression model as an adjusted predictor for AAO, we found that the mosaic AGGG frequency was not associated with AAO (estimate = 138.9471, $P = 0.09$) (Table 2). When including the genetics of the XDP-relevant *MSH3*/*PMS2* SNPs into the model, the mosaic AGGG frequency was associated with AAO (estimate = 155.1063, $P = 0.047$*)*; however, the association dissipated after including the repeat number (estimate = −92.46430, $P = 0.079$) (Table 2).

We assessed the AGG repeat motif using residuals from the age at sampling in a linear model. We did not observe an association with repeat number (estimate = −3.8342, $P = 0.869$) or AAO (estimate = −1.8236, $P = 0.97$). The canonical hexamer repeat motif frequency was associated with repeat number (estimate = 45.32, $P = 0.0441$) but not AAO (estimate = −41.486, $P = 0.378$) (Supplementary Tables 8 and 9).

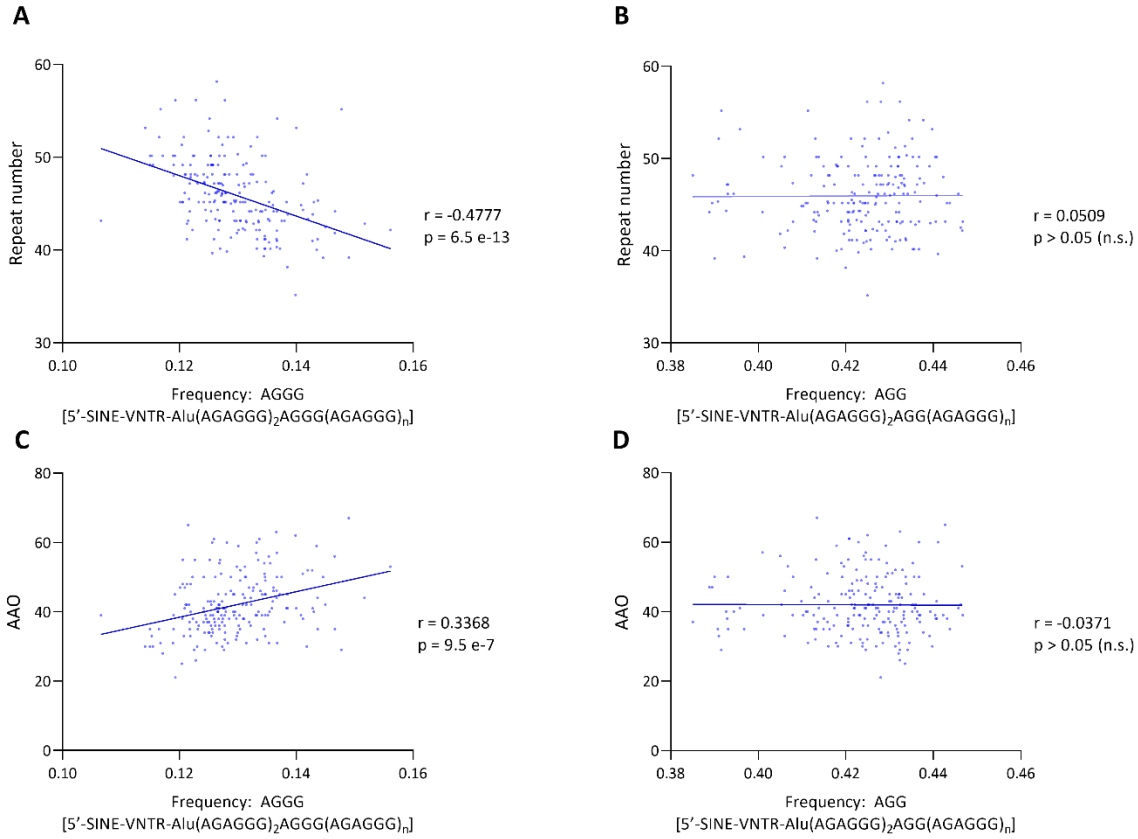

Figure 4. Relationship between mDRILs, AAO, and repeat number in patients with XDP. (**A**) The correlation between hexanucleotide repeat number and the mosaic frequency for divergent motif AGGG [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGGG*(AGAGGG)$_n$]. (**B**) The correlation between hexanucleotide repeat number and mosaic frequency for divergent motif AGG [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGG*(AGAGGG)$_n$]. (**C**) The correlation between AAO and mosaic frequency for divergent motif AGGG [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGGG*(AGAGGG)$_n$]. (**D**) The correlation between age at onset (AAO) mosaic frequency for divergent motif AGG [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGG*(AGAGGG)$_n$]. r = Spearman's rank correlation coefficient; *P* = Spearman's exploratory *P*-value.

Table 2. Residuals and coefficients of a linear model predicting age at onset with mosaic AGGG frequency.

Adjusted for age at sampling				
Residuals				
Min	**1Q**	**Median**	**3Q**	**Max**
-20.185	-5.661	-1.585	5.303	24.657
Coefficients				
	Estimate	**SE**	**t-value**	**Pr (>|t|)**
Intercept	41.9447	0.6058	69.234	$<2x10^{-16}$
betares	138.9471	82.0822	1.693	0.0921
Adjusted for age at sampling and genetic modifiers				
Residuals				
Min	**1Q**	**Median**	**3Q**	**Max**
-16.7619	-5.7232	-0.9387	5.6250	21.9343
Coefficients				
	Estimate	**SE**	**t-value**	**Pr (>|t|)**
Intercept	44.2336	1.2167	36.354	$<2x10^{-16}$
$beta_{res}$	155.1063	77.4761	2.002	0.046679
$beta_{rs245013}$	-1.9713	1.0305	-1.1913	0.057234
$beta_{rs33003}$	-1.6813	0.9739	-1.726	0.085866
$beta_{rs62456190}$	3.8217	1.0563	3.618	0.000379
Adjusted for age at sampling, genetic modifiers, and repeat number				
Residuals				
Min	**1Q**	**Median**	**3Q**	**Max**
-16.7062	-3.5218	0.1044	3.5887	15.9468
Coefficients				
	Estimate	**SE**	**t-value**	**Pr (>|t|)**
Intercept	108.70240	3.99377	27.218	$<2x10^{-16}$
$beta_{RN}$	−1.55022	0.09415	−16.465	$<2x10^{-16}$
$beta_{res}$	−92.46430	52.29976	−1.768	0.07865
$beta_{rs245013}$	−2.05898	0.66629	−3.090	0.00230
$beta_{rs33003}$	−2.06651	0.63010	−3.280	0.00123
$beta_{rs62456190}$	4.56800	0.68445	6.674	2.57×10^{-10}

Legend: 1Q = first quartile; 3Q = third quartile; Max = maximum; Min = minimum; res = residuals; RN = repeat number; SE = standard error.

Model: AAO = Intercept + $beta_{res}$ × res.

Model: AAO = Intercept + $beta_{res}$ × res + $beta_{rs245013}$ × rs245013 + $beta_{rs33003}$ × rs33003 + $beta_{rs62456190}$ × rs62456190.

Model: AAO = Intercept + $beta_{RN}$ × RN + $beta_{res}$ × res + $beta_{rs245013}$ × rs245013 + $beta_{rs33003}$ × rs33003 + $beta_{rs62456190}$ × rs62456190.

Discussion

A thorough analysis of the expanded *TAF1* SVA repeat tract in 202 XDP patients revealed novel mosaic divergent repeat interruptions affecting both motif length and sequence (DRILS) of the canonical motif polarized within the expanded repeat tract.

Repeat interruptions exist in other repeat expansion disorders (Nethisinghe et al. 2021; McFarland et al. 2013; Stolle et al. 2008; Matsuura et al. 2006). For example, interruptions with the same motif length, but varying repeat sequence from the canonical repeat motif, have been observed in fragile X syndrome (FXS), Huntington's disease, spinal cerebellar ataxia (SCA)1, SCA2, SCA3, SCA8, SCA10, SCA17, SCA31, myotonic dystrophy, and Friedreich's ataxia (Gall-Duncan et al. 2022; Pearson et al. 1998). The presence of these interruptions can modify AAO by years. Interruptions of the $(GAA)_n$ tract in Friedreich's ataxia can delay AAO by 9 years (Al-Mahdawi et al. 2018). A pure HTT $(CAG)_n$ tract can hasten disease by 13–29 years, whereas a multiply-interrupted tract can delay disease by 3–6 years (Wright et al. 2020). AGG interruptions polarized to the 5′-end of the $(CGG)_n$ of *FMR*1 are associated with FXS, fragile X-associated tremor/ataxia syndrome, and fragile X-associated primary ovarian insufficiency. The CAT interruptions of the $(CAG)_n$ of *ATXN1* in SCA1, the CAA interruptions in SCA2, the GAA of Friedreich's ataxia, and the various variant repeats at the expanded $(CTG)_n$ in *DMPK are* associated with myotonic dystrophy. In addition to affecting the AAO, interruptions can also alter clinical presentation, as in Friedreich's ataxia, myotonic dystrophy type 1 (DM1), and SCA10 (Nethisinghe et al. 2021; Matsuura et al. 2006; Wenninger et al. 2021). However, these disruptions in purity were all caused by interrupting repeat units of the same motif length. In all cases, except for DM1 and SCA8, the change of the repeat motif was a single-nucleotide replacement in the motif (i.e., CAG → CAA in the case of HTT). For DM1, the variant repeats were of various sequences, all the same number of nucleotides as the canonical motif, and were present in 8.4% of DM1 individuals with expansions (Wenninger et al. 2021). Moreover, in each of these cases, the interruptions are reported as germline repeat configurations, and the possibility that they may vary somatically as putative mosaic repeat disruptions has not been looked at. XDP divergent repeat motifs were found to be polarized at the beginning (5′ end) of the *TAF1* SVA $(AGAGGG)_n$ repeat tract in a somatic mosaic fashion, indicating a new mechanism. We postulate that these deletions stabilize the repeats, especially as a higher frequency of the AGGG divergent repeat is associated with shorter repeats. It can be inferred that the loss of divergent repeat motifs indirectly delays the AAO.

However, the association dissipates when the model includes other modifiers and repeat numbers. One possible explanation is that the mosaic deletions affect repeat stability and thus repeat number. DRILs may arise by deletions or insertions of single nucleotides from the canonical AGAGGG motif. If the mosaic divergent repeats (mDRILS) arose from deletion events, the repeat tract length will be shorter than the canonical AGAGGG repeat tract. If mDRILS arose from insertion events, the repeat tract length will be longer than the canonical hexameric repeat tract. Our analyses on n = 202 suggest this is an insertion event (Supplementary Fig. 3A and B). Notably, in the XDP hexameric repeat tract, the divergent motifs occur at the extreme 5′-end of the repeat at the first and second repeat units of the tract. This polarization is similar to polarized variant AGG and CAA repeat motifs of the *FMR1* CGG and the *HTT* CAG repeat tracts, respectively (Wright et al. 2020; Nolin et al. 2019). The polarity of the XDP mutations, coupled with their being present somatically, suggests that they may arise through a possible biological mechanism. Recently, it has been shown that FAN1 exo-nuclease pauses during the excision of CGG and CAG slip-out DNAs (Deshmukh et al. 2021). These pauses are particularly intense at the polar ends of the repeat tract, proximal to the variant interrupting CAA repeat motifs of CAG slip-outs. One might imagine a polarized error leading to sequence alterations of the repeat. Other interruption-specific pathways involve mismatch repair proteins (Rolfsmeier et al. 2000). Future studies will reveal how DRILS may arise. Our findings in XDP patients revealed mosaicism for each of the different divergent repeat configurations within a given patient sample of the repeat tract, supporting either mechanism as a dynamic process.

The DRILS in the XDP-relevant repeat may affect repeat instability by modifying the propensity to form unusual mutagenic DNA structures, as observed for the interruptions of *FMR1* (FXS) and *ATXN1* (SCA1) (Pearson et al. 1998). DRILS can also lead to pathogenic spliceoforms, translation (exonization), RAN-translation, or repeat instability (Gall-Duncan et al. 2022). As many repeat disorders (e.g., *RFC1, SCA31, SCA8*) are currently being investigated with third-generation sequencing technologies, emerging repeat unit variations like DRILS will become apparent (Gall-Duncan et al. 2022). It is possible that repeat motif variations, like DRILS, also arise in expansions of other repeats and have been missed due to the sequencing method applied. General limitations of nanopore sequencing are possible artifacts due to amplification during the long-range PCR, sequencing, or software (Harris et al. 2019).

Thus, for further confirmation, we validated the presence of divergent repeat motifs in two ways: RP-PCR and Cas9-targeted enrichment. However, somatic mosaicism can be challenging to detect, and in this case, it is evident only with Cas9 enrichment and PCR amplicon long-read sequencing. Still, replication in other cohorts and the use of different technologies are warranted.

Our study illustrates: (i) the importance of underexplored dynamic somatic mosaic genotypes (repeat tract length, motif length, and motif sequence) present in every individual (n = 202) in this case; (ii) the biological plausibility of multiple modifiers (both germline and somatic) that can have effects on repeat instability and expressivity; and (iii) that these variations may remain undetected with older technologies that do not assess single molecules. Importantly, this study sheds light on another putative modifier of XDP expressivity associated with AAO and potentially disease severity. Mosaic repeat deletions present as a novel disease mechanism that is also clinically relevant for other repeat expansion disorders and future genetic counseling with implications beyond XDP.

Funding

The study was supported by the Deutsche Forschungsgemeinschaft (DFG FOR2488; to J.T., I.R.K., N.B., A.W., and C.K.), by funding by the Collaborative Center for X-linked Dystonia-Parkinsonism at Massachusetts General Hospital (to N.B., A.W., and C.K.), the Else-Kröner Fresenius Foundation (to J.T. and N.B.), by intramural funds from the University of Lübeck (to J.T. and C.K.), a career development award from Peter Engelhorn Foundation (to J.T.) and by a career development award from the Hermann and Lilly Schilling Foundation (to C.K.). J.T. and T.L. had full access to all the data in the study and took responsibility for the integrity and accuracy of the data analysis.

Competing interests

The authors report no competing interests.

Supplementary material

Supplementary material is available at *Brain* online.

References

Al-Mahdawi, Sahar, Heather Ging, Aurelien Bayot, Francesca Cavalcanti, Valentina La Cognata, Sebastiano Cavallaro, Paola Giunti, and Mark A. Pook. 2018. 'Large Interruptions of GAA Repeat Expansion Mutations in Friedreich Ataxia Are Very Rare.' *Frontiers in Cellular Neuroscience* 12:443. https://doi.org/10.3389/fncel.2018.00443.

Bragg, D. Cristopher, Kotchaphorn Mangkalaphiban, Christine A. Vaine, Nichita J. Kulkarni, David Shin, Rachita Yadav, Jyotsna Dhakal, et al. 2017. 'Disease Onset in X-Linked Dystonia-Parkinsonism Correlates with Expansion of a Hexameric Repeat within an SVA Retrotransposon in TAF1'. *Proceedings of the National Academy of Sciences of the United States of America* 114 (51): E11020–28. https://doi.org/10.1073/pnas.1712526114.

Deshmukh, Amit Laxmikant, Marie-Christine Caron, Mohiuddin Mohiuddin, Stella Lanni, Gagan B. Panigrahi, Mahreen Khan, Worrawat Engchuan, et al. 2021. 'FAN1 Exo- Not Endo-Nuclease Pausing on Disease-Associated Slipped-DNA Repeats: A Mechanism of Repeat Instability'. *Cell Reports* 37 (10): 110078. https://doi.org/10.1016/j.celrep.2021.110078.

Domingo, Aloysius, Ana Westenberger, Lillian V. Lee, Ingrid Brænne, Tian Liu, Inga Vater, Raymond Rosales, et al. 2015. 'New Insights into the Genetics of X-Linked Dystonia-Parkinsonism (XDP, DYT3)'. *European Journal of Human Genetics: EJHG* 23 (10): 1334–40. https://doi.org/10.1038/ejhg.2014.292.

Gall-Duncan, T., N. Sato, R. K. C. Yuen, and C. E. Pearson. 2022. 'Advancing Genomic Technologies and Clinical Awareness Accelerates Discovery of Disease-Associated Tandem Repeat Sequences'. *Genome Res* 32 (1): 1–27. https://doi.org/10.1101/gr.269530.120.

Harris, R. S., M. Cechova, and K. D. Makova. 2019. 'Noise-Cancelling Repeat Finder: Uncovering Tandem Repeats in Error-Prone Long-Read Sequencing Data'. *Bioinformatics* 35 (22): 4809–11. https://doi.org/10.1093/bioinformatics/btz484.

Kawarai, Toshitaka, Paul Matthew D. Pasco, Rosalia A. Teleg, Masaki Kamada, Waka Sakai, Komei Shimozono, Makoto Mizuguchi, et al. 2013. 'Application of Long-Range Polymerase Chain Reaction in the Diagnosis of X-Linked Dystonia-Parkinsonism'. *Neurogenetics* 14 (2): 167–69. https://doi.org/10.1007/s10048-013-0357-x.

Laabs, Björn-Hergen, Christine Klein, Jelena Pozojevic, Aloysius Domingo, Norbert Brüggemann, Karen Grütz, Raymond L. Rosales, et al. 2021. 'Identifying Genetic Modifiers of Age-Associated Penetrance in X-Linked Dystonia-Parkinsonism'. *Nature Communications* 12 (1): 3216. https://doi.org/10.1038/s41467-021-23491-4.

Lee, L. V., F. M. Pascasio, F. D. Fuentes, and G. H. Viterbo. 1976. 'Torsion Dystonia in Panay, Philippines'. *Advances in Neurology* 14:137–51.

Matsuura, T., P. Fang, C. E. Pearson, P. Jayakar, T. Ashizawa, B. B. Roa, and D. L. Nelson. 2006. 'Interruptions in the Expanded ATTCT Repeat of Spinocerebellar Ataxia Type 10: Repeat Purity as a Disease Modifier?' *Am J Hum Genet* 78 (1): 125–29. https://doi.org/10.1086/498654.

McFarland, Karen N., Jilin Liu, Ivette Landrian, Rui Gao, Partha S. Sarkar, Salmo Raskin, Mariana Moscovich, et al. 2013. 'Paradoxical Effects of Repeat Interruptions on Spinocerebellar Ataxia Type 10 Expansions and Repeat Instability'. *European Journal of Human Genetics: EJHG* 21 (11): 1272–76. https://doi.org/10.1038/ejhg.2013.32.

Nethisinghe, S., M. Kesavan, H. Ging, R. Labrum, J. M. Polke, S. Islam, H. Garcia-Moreno, et al. 2021. 'Interruptions of the FXN GAA Repeat Tract Delay the Age at Onset of Friedreich's Ataxia in a Location Dependent Manner'. *Int J Mol Sci* 22 (14). https://doi.org/10.3390/ijms22147507.

Nolin, Sarah L., Anne Glicksman, Nicole Tortora, Emily Allen, James Macpherson, Montserrat Mila, Angela M. Vianna-Morgante, et al. 2019. 'Expansions and Contractions of the FMR1 CGG Repeat in 5,508 Transmissions of Normal, Intermediate, and Premutation Alleles'. *American Journal of Medical Genetics. Part A* 179 (7): 1148–56. https://doi.org/10.1002/ajmg.a.61165.

Pauly, Martje G., Marta Ruiz López, Ana Westenberger, Gerard Saranza, Norbert Brüggemann, Anne Weissbach, Raymond L. Rosales, et al. 2020. 'Expanding Data Collection for the MDSGene Database: X-Linked Dystonia-Parkinsonism as Use Case Example'. *Movement Disorders: Official Journal of the Movement Disorder Society* 35 (11): 1933–38. https://doi.org/10.1002/mds.28289.

Pearson, C. E., E. E. Eichler, D. Lorenzetti, S. F. Kramer, H. Y. Zoghbi, D. L. Nelson, and R. R. Sinden. 1998. 'Interruptions in the Triplet Repeats of SCA1 and FRAXA Reduce the Propensity and Complexity of Slipped Strand DNA (S-DNA) Formation'. *Biochemistry* 37 (8): 2701–8. https://doi.org/10.1021/bi972546c.

Posey, Jennifer E., Anne H. O'Donnell-Luria, Jessica X. Chong, Tamar Harel, Shalini N. Jhangiani, Zeynep H. Coban Akdemir, Steven Buyske, et al. 2019. 'Insights into Genetics, Human Biology and Disease Gleaned from Family Based Genomic Studies'. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 21 (4): 798–812. https://doi.org/10.1038/s41436-018-0408-7.

Reyes, Charles Jourdan, Björn-Hergen Laabs, Susen Schaake, Theresa Lüth, Raphaela Ardicoglu, Aleksandar Rakovic, Karen Grütz, et al. 2021. 'Brain Regional Differences in Hexanucleotide Repeat Length in X-Linked Dystonia-Parkinsonism Using Nanopore Sequencing'. *Neurology. Genetics* 7 (4): e608. https://doi.org/10.1212/NXG.0000000000000608.

Rolfsmeier, M. L., M. J. Dixon, and R. S. Lahue. 2000. 'Mismatch Repair Blocks Expansions of Interrupted Trinucleotide Repeats in Yeast'. *Molecular Cell* 6 (6): 1501–7. https://doi.org/10.1016/s1097-2765(00)00146-5.

Stolle, Catherine A., Edward C. Frackelton, Jennifer McCallum, Jennifer M. Farmer, Amy Tsou, Robert B. Wilson, and David R. Lynch. 2008. 'Novel, Complex Interruptions of the GAA Repeat in Small, Expanded Alleles of Two Affected Siblings with Late-Onset Friedreich Ataxia'. *Movement Disorders: Official Journal of the Movement Disorder Society* 23 (9): 1303–6. https://doi.org/10.1002/mds.22012.

Wenninger, Stephan, Sarah A. Cumming, Kristina Gutschmidt, Kees Okkersen, Aura Cecilia Jimenez-Moreno, Ferroudja Daidj, Hanns Lochmüller, et al. 2021. 'Associations Between Variant Repeat Interruptions and Clinical Outcomes in Myotonic Dystrophy Type 1'. *Neurology. Genetics* 7 (2): e572. https://doi.org/10.1212/NXG.0000000000000572.

Westenberger, A., C. J. Reyes, G. Saranza, V. Dobricic, H. Hanssen, A. Domingo, B. H. Laabs, et al. 2019. 'A Hexanucleotide Repeat Modifies Expressivity of X-Linked Dystonia Parkinsonism'. *Ann Neurol* 85 (6): 812–22. https://doi.org/10.1002/ana.25488.

Wright, G. E. B., H. F. Black, J. A. Collins, T. Gall-Duncan, N. S. Caron, C. E. Pearson, and M. R. Hayden. 2020. 'Interrupting Sequence Variants and Age of Onset in Huntington's Disease: Clinical Implications and Emerging Therapies'. *Lancet Neurol* 19 (11): 930–39. https://doi.org/10.1016/S1474-4422(20)30343-4.

Objective 2: To assess the stability of mosaic modifiers across familial generations in XDP

Anticipation is a common mechanism in repeat expansion disorders, where the repeat number increases during transmission from the parent to the children. Notably, there are differences between paternal and maternal transmission in repeat expansion disorders. In Huntington's disease, the paternal transmission has a higher expansion of the CAG repeats than the maternal transmission. In contrast, in the fragile X syndrome, the maternal transmission is higher. If interruptions are also inherited, it is still unclear.

Therefore, I investigated the stability of interruptions in the study titled "Stability of Mosaic Divergent Repeat Interruptions in X-Linked Dystonia-Parkinsonism". I additionally tested if the frequency of the most common mDRILS in XDP affects the transmission of the repeat number. I have analyzed a total of 56 families with XDP. Due to the X-linked inheritance, only father-daughter and mother-son pairs were investigated. The total number of individuals was 130; we had 17 father-daughter and 57 mother-son pairs. As in the previous studies, I performed a PCR amplification of the *TAF1* SVA retrotransposon and sequenced the library on R9.4.1 flow cells. During this study, Oxford Nanopore Technologies developed the new flow cell R10.4.1 with higher throughput and accuracy.

I used a subset of 10 randomly selected individuals to validate the R9.4.1 chemistry. The bioinformatic pipeline from the previous studies was used to detect the repeat length and interruptions. I used the new base caller Dorado instead of Guppy for the base calling of the data from the R10.4.1 flow cell. Comparing the repeat length across generations, three scenarios are possible. The repeat length can expand from the first generation to the second. The second generation can have a lower repeat length than the first generation, which we call contraction. The last scenario is that the repeat length is the same in the first and second generations, which we call retention. Only four individuals in the second generation had a contraction in repeat length. For statistical analysis, I combined the group of contractions and retentions and compared them against the group of expansions. Mann-Whitney U-tests compared these two groups regarding repeat length and interruption frequency. The cohort was separated into father-daughter and mother-son pairs to test if the maternal or paternal transmission of the repeat length was influenced by interruption frequency. I have performed the bioinformatic pipelines and the statistical analysis. Finally, I have drafted the first manuscript, including all figures, tables, and supplementary

material. Additionally, I have drafted the response letter under the supervision of Prof. Joanne Trinh. The study has shown that a higher AGGG frequency may stabilize repeats across generations.

Title

Stability of Mosaic Divergent Repeat Interruptions in X-Linked Dystonia-Parkinsonism

Authors

Joshua Laß[1], Theresa Lüth[1], Kathleen Schlüter[1], Susen Schaake[1], Björn-Hergen Laabs[2], Christoph Much[1], Roland Dominic Jamora[3], Raymond L. Rosales[4], Gerard Saranza[5], Cid Czarina E. Diesta[6], Christopher E. Pearson[7], Inke R. König[2], Norbert Brüggemann[1,8], Christine Klein[1], Ana Westenberger[1], Joanne Trinh[1]*

[1] Institute of Neurogenetics, University of Lübeck and University Hospital Schleswig-Holstein, Lübeck, Germany

[2] Institute of Medical Biometry and Statistics, University of Lübeck, Lübeck, Germany

[3] Department of Neurosciences, College of Medicine—Philippine General Hospital, University of the Philippines Manila, Manila, Philippines

[4] Department of Neurology and Psychiatry, University of Santo Tomas and the CNS-Metropolitan Medical Center, Manila, Philippines

[5] Section of Neurology, Department of Internal Medicine, Chong Hua Hospital, Cebu City, Philippines

[6] Department of Neurosciences, Movement Disorders Clinic, Makati Medical Center, Makati City, Philippines

[7] Program of Genetics and Genome Biology, The Hospital for Sick Children, The Peter Gilgan Centre for Research and Learning, Toronto, Canada

[8] Department of Neurology, University of Lübeck, Lübeck, Germany

* Correspondence author

Published in *Movement Disorders*, 2024, *39*(7), 1145-1153; DOI: 10.1002/mds.29809

Abstract

Background

X-Linked dystonia-parkinsonism (XDP) is an adult-onset neurodegenerative disorder characterized by rapidly progressive dystonia and parkinsonism. Mosaic Divergent Repeat Interruptions affecting motif Length and Sequence (mDRILS) were recently found within the *TAF1* SVA repeat tract. They were shown to associate with repeat stability and age at onset in XDP, specifically the AGGG [5′-SINE-VNTR-Alu($AGAGGG$)$_2$$AGGG$($AGAGGG$)$_n$] mDRILS.

Objective

This study aimed to investigate the stability of mDRILS frequencies and the stability of ($AGAGGG$)$_n$ repeat length during transmission in parent–offspring pairs.

Methods

Fifty-six families (n = 130) were investigated for generational transmission of repeat length and mDRILS. The mDRILS stability of 16 individuals was assessed at two sampling points, 1 year apart. DNA was sequenced with long-read technologies after long-range polymerase chain reaction amplification of the *TAF1* SVA. Repeat number and mDRILS were detected with Noise-Cancelling Repeat Finder (NCRF).

Results

When comparing the repeat domain, 51 of 65 children had either contractions or expansions of the repeat length. The AGGG frequency remained stable across generations at 0.074 (IQR: 0.069–0.078) ($z = -0.526$; $P = 0.599$). However, the median AGGG frequency in children with an expansion (0.072 [IQR: 0.066–0.076]) was lower compared with children with retention or contraction (0.080 [IQR: 0.073–0.083]) ($z = -0.007$; $P = 0.003$). In a logistic regression model, the AGGG frequency predicted the outcome of either expansion or retention/contraction when including repeat number and sex as covariates ($\beta = 80.7$; z-score = 2.63; $P = 0.0085$). The AGGG frequency varied slightly over 1 year (0.070 [IQR: 0.063–0.080] to 0.073 [IQR: 0.069–0.078]).

Conclusions

Our results show that a higher AGGG frequency may stabilize repeats across generations. This highlights the importance of further investigating mDRILS as a disease-modifying factor with generational differences.

Introduction

X-Linked dystonia-parkinsonism (XDP) is an adult-onset neurodegenerative disorder characterized by rapidly progressive dystonia and parkinsonism (Pauly et al. 2020; Lee et al. 2001). All XDP patients are of Filipino descent, given that the disease is caused by a founder mutation: a SINE-VNTR-Alu (SVA) retrotransposon insertion in intron 32 of the *TAF1* (*TATA-binding protein-associated factor 1*) gene (Pauly et al. 2020; Domingo et al. 2015). A hexanucleotide repeat domain within the *TAF1* SVA retrotransposon, $(AGAGGG)_n$, varies in length between patients and modifies the disease onset, the longer the repeat, the earlier the age at onset (AAO) (Westenberger et al. 2019; Bragg et al. 2017; Laabs et al. 2021).

In recent years, investigating repeat motif interruptions has become increasingly important. Repeat interruptions in repeat expansion disorders like Huntington's disease (HD), spinocerebellar ataxia type 10 (SCA10), myotonic dystrophy type 1 (DM1), and Friedrich's ataxia (FA) play a protective role (Ballester-Lopez et al. 2020; McFarland et al. 2013; Netravathi et al. 2009; Matsuura et al. 2006; Stolle et al. 2008; Wright et al. 2020; Morales et al. 2020; Cumming et al. 2018; Findlay Black et al. 2020). For example, interruptions at the 5′-end of the repeat motif can delay the AAO of Friedrich's ataxia by several years (Nethisinghe et al. 2021). The molecular mechanism of the disease-modifying effect is still under debate, but possible mechanisms involve repeat stabilization against expansion, interference with pathogenic splice forms, or prevention of RAN-translation (Gall-Duncan et al. 2022).

Mosaic interruptions in the hexanucleotide repeat domain in the *TAF1* SVA were previously found with long-read sequencing and were confirmed by repeat-primed polymerase chain reaction (PCR) and Cas9 enrichment (Trinh et al. 2023). As the interruptions affect the motif length and sequence, we have coined the term “mosaic Divergent Repeat Interruption affecting motif Length and Sequence” (mDRILS). In XDP, different mDRILS were identified at the 5′-end of the repeat tract, and each mDRILS is mosaic. The different mDRILS arose from insertion events after the second hexanucleotide repeat unit at the 5′-end. We discovered a positive correlation between the frequency of mDRILS and AAO: the higher the somatic frequency of the [5′-SINE-VNTR-Alu$(AGAGGG)_2AGGG(AGAGGG)_n$] mDRILS, the later the AAO in XDP patients (Trinh et al. 2023).

However, whether the mosaic interruptions and repeat length change over time and across generations warrants further investigation. We hypothesized that the maternal expansion bias across the family line negatively correlates with mDRILS frequencies. Herein, we characterize the stability of the mDRILS frequencies and the *TAF1* SVA hexanucleotide repeat length within families across generations and time.

Methods

Patient demographics

The stability of the mDRILS frequencies and repeat length across generations was investigated in the blood-derived DNA of 56 families, 17 father–daughter pairs, and 57 mother–son pairs. The total number of analyzed individuals was n = 130 (Table S1, Fig. S1). The mean AAO was 40.1 (SD: 9.0) years, and the mean age at sampling was 48.9 (SD: 16.6) years. As XDP is an X-linked recessive disorder, mothers and daughters are mostly asymptomatic carriers of the *TAF1* SVA; however, among our study participants, two mothers were affected by XDP. Most male participants, 63 of 70, were affected by XDP. The mean age of non-mutation carriers was 43.7 (SD: 17.9) years. The study was approved by the Ethics Committees of the University of Lübeck, Germany, and the Metropolitan Medical Center, Manila, Philippines.

To explore the stability of the mDRILS frequencies and repeat length across sampling collection, 16 (15 males and one female) participants were investigated at two different time points (Table S2). Nine of these men and one woman were affected by XDP. One woman who was affected by XDP was also in the study cohort. The mean AAO was 39.8 (SD: 7.0) years, and the mean age at sampling was 43.4 (SD: 7.1) years. The time between the two DNA samplings was at least 1 year.

Nanopore sequencing

Blood DNA Midi Kit (Qiagen, Hilden, Germany) was used to extract the genomic DNA from blood. The quality control for the extracted DNA and library was performed using Nanodrop (Table S1 and S2). A long-range PCR was performed to amplify the repeat domain of the *TAF1*SVA insertion, as previously described (Trinh et al. 2023). The primer sequences are 5′-GTTCCATTGTGTGGTTGTACCAGCGTTTGTTC-3′ and 5′-CACATGAAAAGATGCCCAACATCATTA GCCATTAG-3′. Depending on the repeat length, the resulting amplicons ranged between 3.0 and 3.4 kb.

The Native 96 Barcoding Kit (EXP-NBD196) was used to multiplex the PCR products, and 200 fmol of amplified PCR product was used as input. The following sequencing adapter ligation was performed with the SQK-LSK109 (ONT). The final product was then loaded on R9.4.1 flow cells on a GridION.

To validate the R9.4.1 chemistry, a subset of 10 individuals was randomly selected for amplicon sequencing using the SQK-NBD114-24 library preparation kit loaded on an R10.4.1 flow cell. Repeat number concordance between R9.4.1 and R10.4.1 for these individuals was estimated with a linear regression model in R Studio (version 3.3.0).

From the R10.4.1, the median read length was 3500 bp (IQR: 3456–3956; SD: 268), and the quality was 16.9 (IQR: 16.8–17.0; SD: 0.12). The median coverage was 38,479X (IQR: 7212–42,240; SD: 19,758). A high concordance between the R9.4.1 and R10.4.1 flow cells was detected ($P = 8.809 \times 10^{-12}$; $R = 0.998$) (Fig. S2). The interruptions of the hexanucleotide motif were detected at the same positions in data with the R9.4.1 and R10.4.1 chemistry (Fig. S3). Additionally, we performed on a single individual a Cas9 enrichment. Four CRISPR RNAs (crRNAs), two upstream and two downstream of the *TAF1* SVA insertion, were designed with the ChopChop tool (https://chopchop.cbu.uib.no, accessed on December 10, 2021) (Table S3). The resulting 5.5 kb product was used as input for the library preparation kit SQK-LSK114. The library was then loaded on an R10.4.1 flow cell and was sequenced on a GridION. The median read length was 6857 bp, and the median quality was 18.7. The coverage was 60X. The detected repeat length was identical to that of the R9.4.1 and the R10.4.1 flow cells.

Genetic data analysis

First, base-calling was performed with the super accuracy model (dna_r9.4.1_450bps_sup.cfg) of Guppy (version 6.1.1) for the R9.4.1 chemistry. For the R10.4.1 chemistry, the basecaller Dorado (version 7.2.13) with super accuracy model (dna_r10.4.1_e8.2_400bps_sup@v4.3.0) was used. The following steps were performed for both chemistries. The software Nanostat (version 1.5.0) was used to analyze the quality of the reads, and the software Filtlong (version 0.2.0) was employed for filtering. With Filtlong, we retained reads by filtering for q-score > 12 and the top 70% of reads based on the q-score. For the alignment to the reference sequence, Minimap2 (version 2.22) was used with parameters for optimal performance and accuracy for long reads on the nanopore (Li 2018). The handling of SAM/BAM files as well as the calculation of the coverage was performed with Samtools (version 1.15) (Li et al. 2009).

Using Samtools, only reads with an alignment length over 1 kb were included in the analysis. Finally, hexanucleotide repeats were detected using the Noise-Cancelling Repeat Finder (NCRF, version 1.01.02) (Harris et al. 2019). The internal NCRF filter was applied to only include reads with a maximum noise of 80% and a minimum of 10 detected repeat units. For the calculation of the repeat length, the median of all reads was used, as previously described (Lüth et al. 2022). Subsequently, the summary output was imported into R (version 3.3.0) to obtain the interruption frequency per site, as previously described (Trinh et al. 2023).

Statistical data analysis

The graphics and the statistical data analyses were performed in R (version 3.3.0). Graphical representation was created with ggplot2 (version 3.4.4, R Studio). Mann–Whitney *U*-tests were used to compare two independent groups. The analysis of the association between transmission types expansion and contraction/retention was interpreted for significance in the complete study group, based on the presence of the "a priori" hypothesis on the association between the mDRILS AGGG frequency on repeat stability. The association of mDRILS AGGG frequency on transmission type was tested for significance, and the significance level was set to $\alpha = 0.05$. All other analyses in this study were exploratory, and *P*-values were not corrected for multiple testing. Additionally, a logistic model was performed with the glm-function (formula: glm [Transmission type ~Repeat number + AGGG mDRILS frequency + Sex], family = "binomial"). Transmission type was categorized into either expansion or contraction/retention based on a difference of one repeat unit or more of the wildtype hexanucleotide motif from generation to generation. This cut-off of one repeat unit was used to label the transmission type in a previous study on XDP (Westenberger et al. 2019). We have performed fragment analysis on all patients to assess the repeat number using another method and compared the concordance of transmission type across families. Validation by fragment analysis to determine the repeat length of the hexanucleotide $(AGAGGG)_n$ was performed with a fluorescein amidite (FAM)-labeled primer, as previously described (Westenberger et al. 2019). When using the term repeat length, we specifically refer to the wildtype motif with the pure hexanucleotide repeat units.

A chi-square test was performed with the chisq.test() function in R to investigate the difference in transmission types between female and male participants in the second generation. For the validation of the run with R9.4.1 chemistry, a linear regression model was performed with the lm-function (formula: lm [R10 flow cell ~ R9 flow cell, data = Repeat number]).

Results

Transmission of SVA Hexanucleotide Interruptions and Repeat Lengths

The mDRILS frequencies and the repeat length were investigated in 17 father–daughter pairs and 57 mother–son pairs. The median quality of reads was 16.7 (IQR: 16.5–17.0; SD: 0.4), and the median read length was 3.32 kb (IQR: 3.29–3.33 kb; SD: 49.70) (Table S4).

In our study, we focused on the two most prominent mDRILS, AGGG [5′-SINE-VNTR-Alu($AGAGGG$)$_2$$AGGG$($AGAGGG$)$_n$] and AGG [5′-SINE-VNTR-Alu($AGAGGG$)$_2$$AGG$($AGAGGG$)$_n$], and on the wildtype (WT) sequence [5′-SINE-VNTR-Alu($AGAGGG$)$_n$].

The most stable mDRILS frequency was the AGGG frequency. The first generation had an average AGGG frequency of 0.074 (IQR: 0.070–0.076; SD: 0.008), and the second generation had a frequency of 0.073 (IQR: 0.069–0.079; SD: 0.007) ($z = -0.526$; $P = 0.599$) (Fig. 1B). The AGG frequency was slightly increased during the transmission from 0.439 (IQR: 0.413–0.454; SD: 0.033) in the first generation to 0.449 (IQR: 0.427–0.460; SD: 0.025) in the second generation ($z = 1.818$; $P = 0.069$) (Fig. 1C). The WT frequency showed a different pattern to the mDRILS frequencies and decreased from 0.431 (IQR: 0.420–0.447; SD: 0.022) to 0.427 (IQR: 0.416–0.439; SD: 0.019) during transmission ($z = -1.675$; $P = 0.094$) (Fig. 1D). Overall, differences across generations were marginally different.

Comparing the repeat length across generations, a change in the repeat length in 51 of 65 children was detected. In the second generation, 38 individuals had a higher repeat length than the first generation (expansion). Only four individuals of the second generation had a lower repeat length (contraction) (Fig. 2, Table S5). Thirty individuals of the second generation had no change in the repeat number (retention). The rates of transmission types were different between mother–son pairs and father–daughter pairs (Fig. 2). An expansion was detected in 40 of 50 second-generation males (80%) and only 7 of 15 females (47%). Conversely, retention was identified in 9 of 50 second-generation males (18%) and 5 of 15 females (33%).

A contraction was detected in 1 of 50 males (2%) and 3 of 15 females (20%) in the second generation. A chi-square test of the transmission types showed a significant difference between female and male participants of the second generation (chi-square statistic = 6.07; $P = 0.048$). In total, the median repeat length of the second generation was increased to 46 (IQR: 44–48; SD: 4) compared with the first generation, with 45 (IQR: 43–47; SD: 4) ($z = 1.616$; $P = 0.106$) (Fig. 1A).

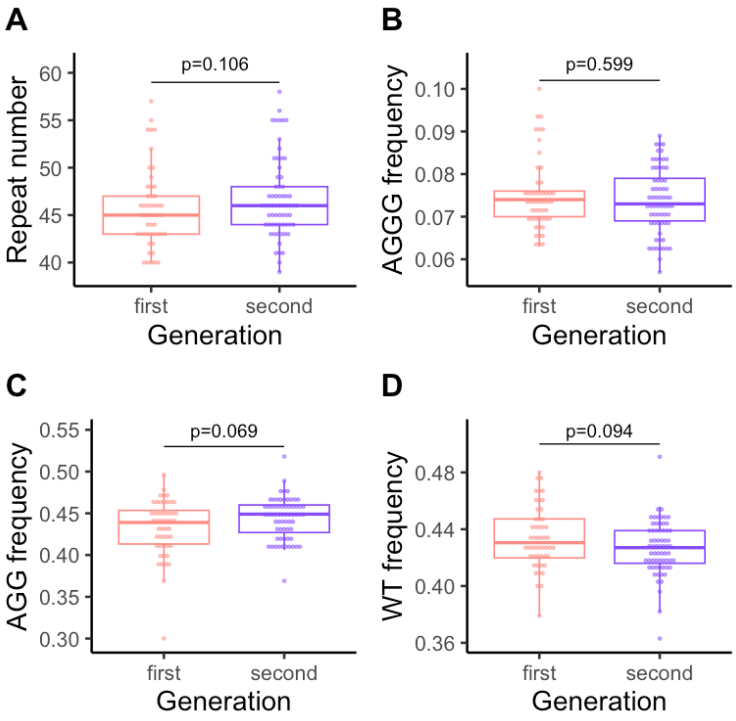

Figure 1. Repeat number and mosaic Divergent Repeat Interruptions affecting motif Length and Sequence (mDRILS) frequencies of the first and second generations. (**A**) Repeat number based on the wildtype motif with the pure hexanucleotide repeat units; (**B**) AGGG frequency; (**C**) AGG frequency; (**D**) WT frequency. Mann–Whitney *U* tests were performed for statistical comparison. AGG = [5′-SINE-VNTR-Alu($AGAGGG$)$_2$$AGG$($AGAGGG$)$_n$], AGGG = [5′-SINE-VNTR-Alu($AGAGGG$)$_2$$AGGG$($AGAGGG$)$_n$]; WT = SINE-VNTR-Alu-5′-($AGAGGG$)$_n$

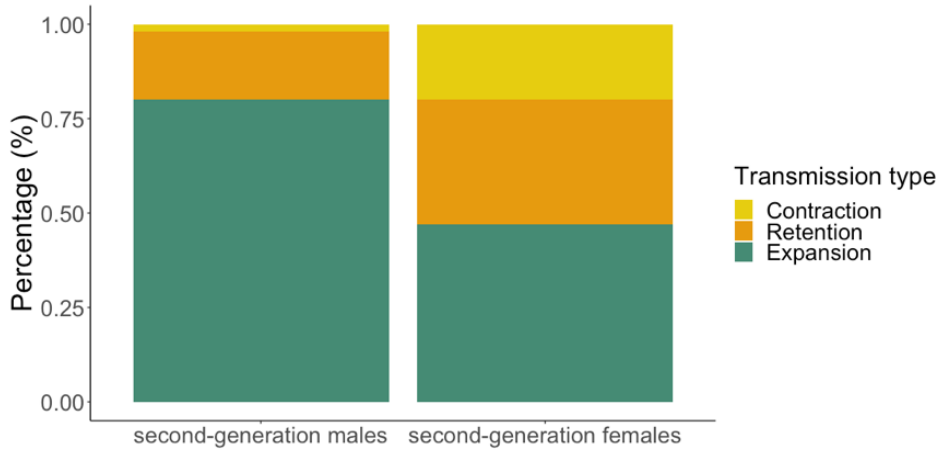

Figure 2. The percentage of transmission types in mother–son and father–daughter pairs. Contraction: Individuals of the second generation had a lower repeat length than those of the first generation. Expansion: Individuals of the second generation had a higher repeat length than those of the first generation. Retention: the second generation had no change in the repeat number compared with the first generation.

Consequently, the mDRILS frequencies and repeat length showed differences between the expansion and contraction/retention groups of the second generation (Table S5). In the second generation, the expansion group had a longer repeat length (median 46; IQR: 44.5–50.5; SD: 4.2) compared with the retention or contraction groups (median 44.5; IQR: 43–46; SD: 4) ($z = -2.612$; $P = 0.009$) (Fig. 3A).

Individuals of the second generation with an expansion had a lower AGGG frequency with 0.072 (IQR: 0.066–0.076; SD: 0.007) compared with the group of retentions or contractions with 0.080 (IQR: 0.073–0.083; SD: 0.007) ($z = -0.007$; $P = 0.003$) (Fig. 3B). Similar to the repeat length, the AGG mDRILS frequency was higher in the group of individuals with an expansion (0.452 [IQR: 0.438–0.462; SD: 0.022]) ($z = -2.432$; $P = 0.015$) (Fig. 3C). The WT frequency in individuals with an expansion (0.422 [IQR: 0.414–0.433; SD: 0.017]) is lower compared with individuals with a retention or contraction (0.438 [IQR: 0.425–0.445; SD: 0.022]) ($z = 2.20$; $P = 0.028$) (Fig. 3D).

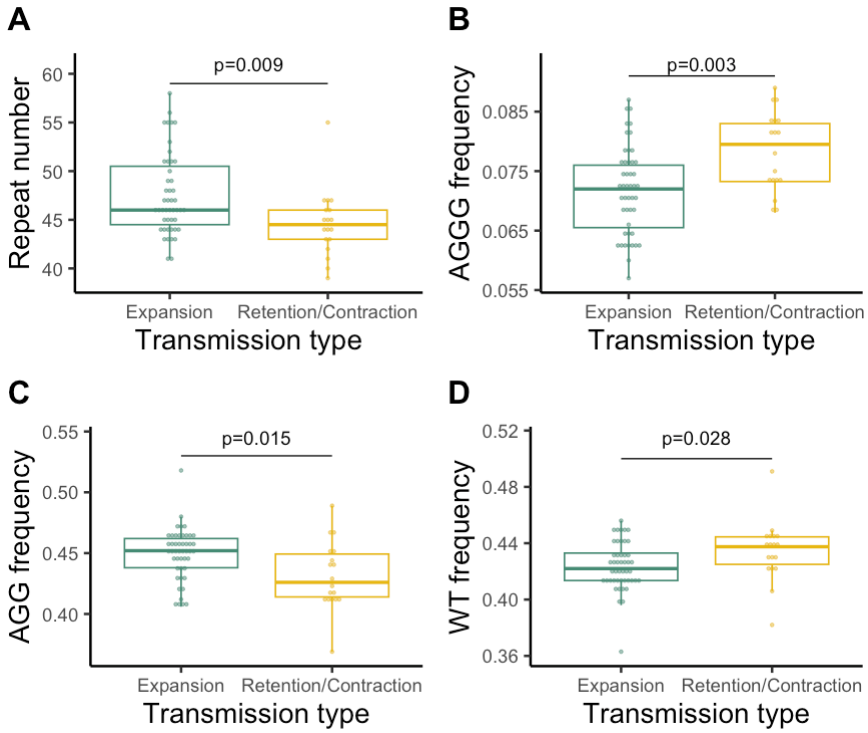

Figure 3. Repeat number and mosaic Divergent Repeat Interruptions affecting motif Length and Sequence (mDRILS) frequencies between transmission groups of the second generation. Transmission type was grouped into either expansion or contraction/retention. (**A**) Repeat number based on the wildtype motif with the pure hexanucleotide repeat units; (**B**) AGGG frequency; (**C**) AGG frequency; (**D**) WT frequency. Mann–Whitney *U* tests were performed for statistical comparison. AGG = [5′-SINE-VNTR-Alu($AGAGGG$)$_2$$AGG$($AGAGGG$)$_n$], AGGG = [5′-SINE-VNTR-Alu($AGAGGG$)$_2$$AGGG$($AGAGGG$)$_n$]; WT = SINE-VNTR-Alu-5′-($AGAGGG$)$_n$.

Next, we investigated differences between the sexes. In total, only small changes between females and males were detected. For the AGGG frequency, no differences were detected between females (0.074 [IQR: 0.070–0.081; SD: 0.007]) and males (0.073 [IQR: 0.068–0.073; SD: 0.009]) ($z = 0.948$; $P = 0.343$) (Fig. 4D). On a trend level, the AGG frequency was higher in males with 0.452 (IQR: 0.434–0.462; SD: 0.029) compared with females with 0.435 (IQR: 0.415–0.452; SD: 0.029) ($z = -1.873$; $P = 0.061$) (Fig. 4G). The WT frequency was slightly lower in males (0.425 [IQR: 0.415–0.435; SD: 0.022]) ($z = 1.896$; $P = 0.058$) (Fig. 4J). The repeat length was higher in males at 46 (IQR: 44–48.25; SD: 4.33) compared with females at 45 (IQR: 43–47; SD: 4) ($z = -0.960$; $P = 0.337$) (Fig. 4A). More significant sex effects were detected in each generation separately.

In the first generation, males had a higher AGGG frequency ($z = -4.941$; $P = 7.769 \times 10^{-7}$); and in the second generation, males had lower AGGG frequency with 0.072 (IQR: 0.067–0.074; SD: 0.006) compared with females with 0.083 (IQR: 0.083–0.087; SD: 0.002) ($z = 5.352$; $P = 8.714 \times 10^{-8}$) (Fig. 4E,F). Males of the first generation had a lower AGG frequency than females ($z = 4.135$; $P = 3.552 \times 10^{-5}$), but in the second generation, females had a lower AGG frequency ($z = -5.042$; $P = 4.608 \times 10^{-7}$) than males (Fig. 4H,I). As for the AGG frequency, the WT frequency was lower in males of the first generation ($z = 3.291$; $P = 0.001$) and higher in males of the second generation ($z = -4.685$; $P = 2.806 \times 10^{-6}$) than in females (Fig. 4K,L). Similar patterns were found in the repeat length. In the first generation, females had a higher repeat length with 45.5 (IQR: 43–48; SD: 4) compared with males with 43 (IQR: 42.25–44.5; SD: 2) ($z = 2.807$; $P = 0.005$) (Fig. 4B). In the second generation, males had a higher repeat length ($z = -2.612$; $P = 0.009$) (Fig. 4C).

In a regression model, the dependencies of repeat number, AGGG frequency, and sex for the transmission type were analyzed. The regression showed that the AGGG frequency has the strongest association with the transmission type ($\beta = 80.7$; $z = 2.63$; $P = 0.009$). The repeat number and the sex had no association with the transmission type (Table S6).

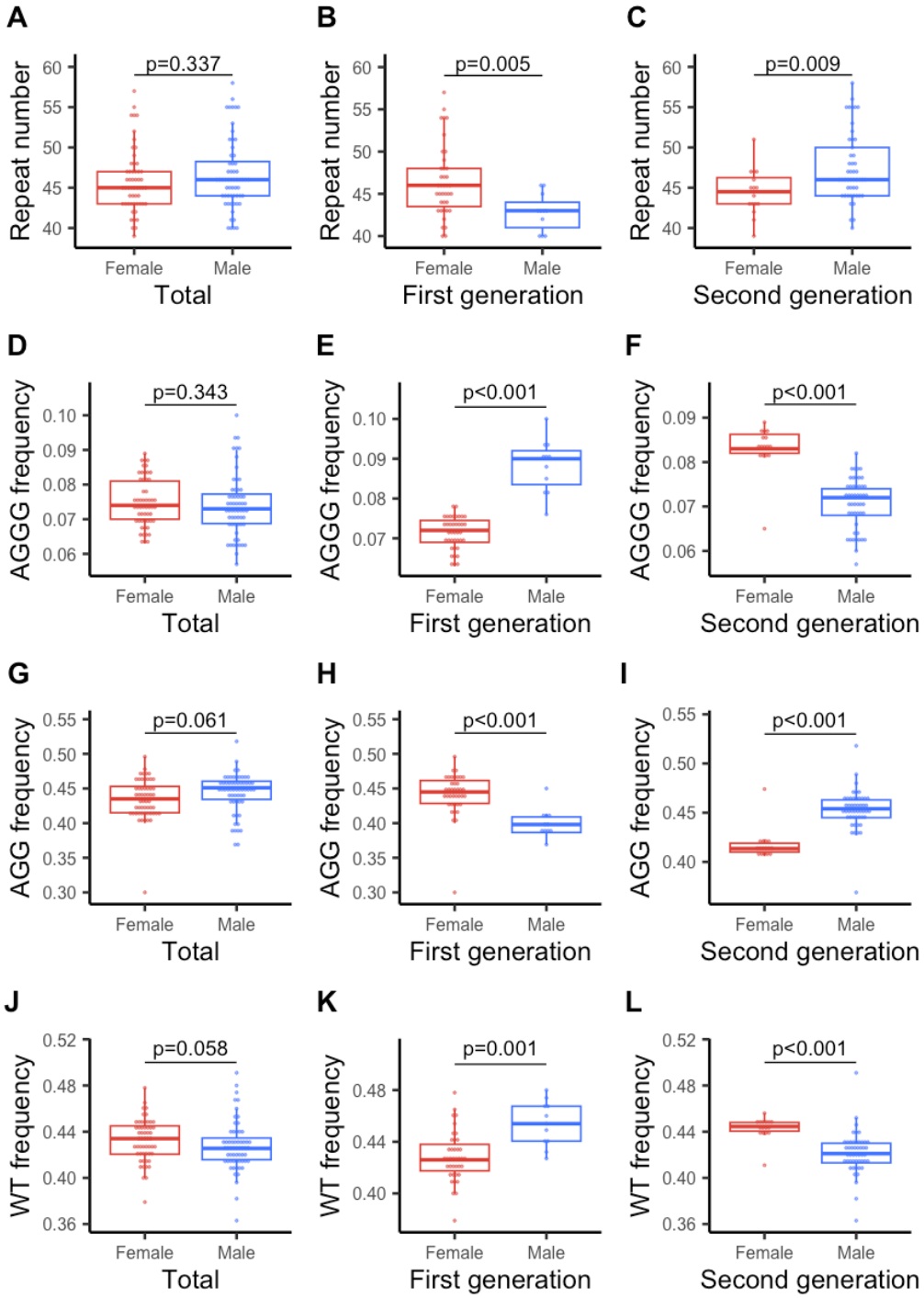

Figure 4. Repeat number and mosaic Divergent Repeat Interruptions affecting motif Length and Sequence (mDRILS) frequency differences between females and males within each generation. (**A**) Repeat number based on the wildtype motif with the pure hexanucleotide repeat units; (**B**) AGGG frequency; (**C**) AGG frequency; (**D**) WT frequency. Mann–Whitney *U* tests were performed for statistical comparison. G1: first generation; G2: second generation. AGG = [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGG*(AGAGGG)$_n$], AGGG = [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGGG*(AGAGGG)$_n$]; WT = SINE-VNTR-Alu-5′-(AGAGGG)$_n$.

Longitudinal Investigation of the SVA Hexanucleotide Repeats

The stability of mDRILS frequencies and repeat length across time was investigated in 15 male and one female participants. Most individuals had 1 year between the two DNA samplings. The median quality of reads in the dataset was 16.3 (IQR: 16.1–16.6; SD: 0.4), and the median read length was 3.35 kb (IQR: 3.33–3.37 kb; SD: 67.28). The AGGG and AGG frequency variation was observed across the two time points. At the first time point, the AGGG frequency was slightly lower with 0.070 (IQR: 0.063–0.080; SD: 0.010) compared with the second time point with 0.073 (IQR: 0.069–0.078; SD: 0.008) ($z = 0.340$; $P = 0.734$) (Fig. 5B). The AGG frequency changed from 0.409 (IQR: 0.398–0.428; SD: 0.022) to 0.425 (IQR: 0.411–0.436; SD: 0.018) between the two time points ($z = 1.585$; $P = 0.113$) (Fig. 5C). The WT frequency was 0.459 (IQR: 0.444–0.466; SD: 0.021) at the first time point and 0.449 (IQR: 0.436–0.458; SD: 0.015) at the second ($z = -1.375$; $P = 0.169$) (Fig. 5D).

Ten individuals had identical repeat lengths at both time points (mean time = 1.6 years [SD: 0.84]), and three individuals had an increase by one repeat (mean time = 2 years [SD: 1.73]). Lastly, three individuals had a decrease by one repeat between the two sampling time points (mean time = 1 year). The median repeat length was 46 for both time point 1 (IQR: 44.75–47.50; SD: 2.71) and time point 2 (IQR: 44.75–48.00; SD: 2.48) ($z = 0.019$; $P = 0.985$) (Fig. 5A).

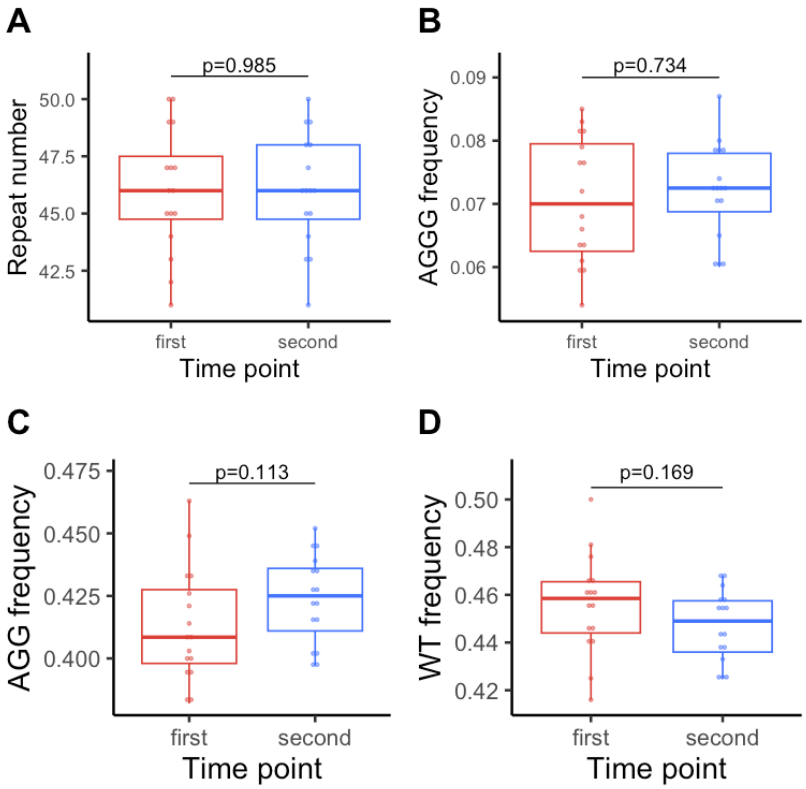

Figure 5. Repeat number and mosaic Divergent Repeat Interruptions affecting motif Length and Sequence (mDRILS) frequency differences across two-time points. (**A**) Repeat number based on the wildtype motif with the pure hexanucleotide repeat units, (**B**) AGGG frequency, (**C**) AGG frequency, (**D**) WT frequency. Mann–Whitney *U* tests were performed for statistical analysis. AGG = [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGG*(AGAGGG)$_n$]; AGGG = [5′-SINE-VNTR-Alu(AGAGGG)$_2$*AGGG*(AGAGGG)$_n$]; WT = SINE-VNTR-Alu-5′-(AGAGGG)$_n$.

Discussion

In this study, we analyzed the stability of the mDRILS frequencies and repeat length during transmission from parents to their children. Additionally, we assessed the stability longitudinally at two-time points in a subgroup of individuals.

Using Nanopore long-read sequencing, we show that the repeat length differed across generations. Maternal transmission of the wildtype hexanucleotide motif had a higher tendency than paternal transmission. These results using newer technology are consistent with previous results produced with repeat-primed PCR (Westenberger et al. 2019). We have consistently shown that long-read sequencing is a reliable detector of repeat expansions (Lüth et al. 2022).

Children with a lower AGGG frequency are more likely to have an expanded repeat number compared to their parents. Combined with the previously identified inverse correlation between repeat length and AAO, this indicates an earlier AAO over generations in the families with a lower AGGG frequency. A higher frequency of the AGGG mDRILS may stabilize repeats across generations. This stabilization effect of the repeat length by the AGGG mDRILS has previously been seen in individuals of a cross-sectional cohort (Trinh et al. 2023). This may suggest that specific mosaic interruptions can aid in future genetic counseling of patients.

The sex bias of the repeat length and mDRILS frequencies can be explained by the preferential inheritance of expansions in sons passed down by the mother (Westenberger et al. 2019). A longer repeat length and lower AGGG frequency were detected in mother–son pairs compared to father–daughter pairs.

Of the 16 patients assessed within a short ~1-year period, more individuals (n = 10) have a stable repeat length. Six individuals have a small one repeat unit change across the same time span. In contrast, the mDRILS frequencies are unstable in every individual. Of note, we evaluated only two time points within a short timeframe, which may imply that more drastic changes can be observed over a longer period. The variation of the AGGG frequency could not be correlated with repeat length changes due to the short period. As evidence of showing that there are drastic changes across time, a higher AGGG frequency is associated with retention or even contraction of the repeat length in families. This leads to the speculation that father–daughter pairs over time and generations would have a delayed onset (Trinh et al. 2023).

It is important to note the limitations of our study which include (1) mainly availability of families with only two generations, (2) a modest sample size of 130 individuals, (3) issues of potential confounding factors of the time sampling, and (4) recall bias usually results in earlier reporting of signs/symptoms in subsequent generations due to higher awareness and/or better medical care. Thus, replication in a larger cohort and more time between the two sampling time points should be warranted to draw further conclusions. Furthermore, studies across two generations (i.e., from grandparents to grandchildren) in families affected with XDP would be beneficial, given the inheritance in an X-linked recessive manner. In addition, blood-derived DNA does not directly correlate with brain-derived somatic variation. Previously, the repeat length somatic instability in brain regions has been detected (Reyes et al. 2021; Wheeler and Dion 2021). Thus, it is plausible that a similar mechanism can be seen across different brain regions for mDRILS.

Of note, a technical limitation of the study is the accuracy of the ONT R9.4.1 chemistry, which is around 99%.

The accuracy is lower in repetitive regions like the hexanucleotide domain of the SVA retrotransposon. The newer ONT R10.4.1 chemistry can reach an accuracy of 99.9%, comparable with Illumina sequencing. In this study, the results with the ONT R9.4.1 chemistry were validated in a subset of 10 individuals with the ONT R10.4.1. The repeat number showed a high concordance between the two chemistries, and the interruptions were identified at the same positions. Nanopore sequencing may not have the accuracy of Illumina or Sanger analysis; however, fragment analysis cannot detect mosaicism and may underestimate the repeat number (Bonnet et al. 2023). In total, 47 families are entirely concordant for both methods regarding transmission type (expansion or contraction/retention) (Table S4). Although the two methods highly correlate ($r = 0.985$; $P = 2.2 \times 10^{-16}$, Fig. S4), five families did not show the same transmission type, and at least one offspring did not correspond in four families. As a sensitivity analysis, we used only the concordant families (n = 47 families) to test our a priori hypothesis and found similar results ($z = -2.652$; $P = 0.008$). This sensitivity analysis provides evidence for the robustness of this finding.

However, mDRILS are not exactly somatic for each individual, as the interruptions are observed across generations, only at varying frequencies. We speculate that mDRILS actually arose from insertion events, given that previous data show that stratification of mDRIL reads gives different read lengths. However, the exact mechanism remains to be elucidated. Single-molecule DNA sequencing across various brain regions is warranted for a more refined approach to detecting somatic instability.

In conclusion, our data show variable mDRILS frequencies and stable repeat length across generations and time. Thus, our results highlight the importance of (1) further investigating mDRILS as a disease-modifying factor potentially impacting offspring and (2) investigating other repeat expansion disorders with long-read sequencing to detect disease-modifying factors.

Acknowledgment

A special thanks to the families and patients who participated in this study.

Financial Disclosures

This study was supported by the German Research Foundation ("ProtectMove"; FOR 2488, I.R.K., N.B., C.K., A.W., and J.T.), the European Community (SysMedPD), the Canadian Institutes of Health Research (CIHR) (J.T.), Peter and Traudl Engelhorn Foundation, and the Center for XDP (CCXDP) at Massachusetts General Hospital (N.B.). CK serves as a medical advisor to Centogene and Retromer Therapeutics and received speaking honoraria from Desitin. The remaining authors report no disclosures.

Data Availability Statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Supplementary material

Supplementary material is available at *Movement Disorders* online.

References

Ballester-Lopez, Alfonsina, Emma Koehorst, Miriam Almendrote, Alicia Martínez-Piñeiro, Giuseppe Lucente, Ian Linares-Pardo, Judit Núñez-Manchón, et al. 2020. 'A DM1 Family with Interruptions Associated with Atypical Symptoms and Late Onset but Not with a Milder Phenotype'. *Human Mutation* 41 (2): 420–31. https://doi.org/10.1002/humu.23932.

Bonnet, Céline, David Pellerin, Virginie Roth, Guillemette Clément, Marion Wandzel, Laëtitia Lambert, Solène Frismand, et al. 2023. 'Optimized Testing Strategy for the Diagnosis of GAA-FGF14 Ataxia/Spinocerebellar Ataxia 27B'. *Scientific Reports* 13 (1): 9737. https://doi.org/10.1038/s41598-023-36654-8.

Bragg, D. Cristopher, Kotchaphorn Mangkalaphiban, Christine A. Vaine, Nichita J. Kulkarni, David Shin, Rachita Yadav, Jyotsna Dhakal, et al. 2017. 'Disease Onset in X-Linked Dystonia-Parkinsonism Correlates with Expansion of a Hexameric Repeat within an SVA Retrotransposon in TAF1'. *Proceedings of the National Academy of Sciences of the United States of America* 114 (51): E11020–28. https://doi.org/10.1073/pnas.1712526114.

Cumming, Sarah A., Mark J. Hamilton, Yvonne Robb, Helen Gregory, Catherine McWilliam, Anneli Cooper, Berit Adam, et al. 2018. 'De Novo Repeat Interruptions Are Associated with Reduced Somatic Instability and Mild or Absent Clinical Features in Myotonic Dystrophy Type 1'. *European Journal of Human Genetics: EJHG* 26 (11): 1635–47. https://doi.org/10.1038/s41431-018-0156-9.

Domingo, Aloysius, Ana Westenberger, Lillian V. Lee, Ingrid Brænne, Tian Liu, Inga Vater, Raymond Rosales, et al. 2015. 'New Insights into the Genetics of X-Linked Dystonia-Parkinsonism (XDP, DYT3)'. *European Journal of Human Genetics: EJHG* 23 (10): 1334–40. https://doi.org/10.1038/ejhg.2014.292.

Findlay Black, H., G. E. B. Wright, J. A. Collins, N. Caron, C. Kay, Q. Xia, L. Arning, et al. 2020. 'Frequency of the Loss of CAA Interruption in the HTT CAG Tract and Implications for Huntington Disease in the Reduced Penetrance Range'. *Genet Med* 22 (12): 2108–13. https://doi.org/10.1038/s41436-020-0917-z.

Gall-Duncan, T., N. Sato, R. K. C. Yuen, and C. E. Pearson. 2022. 'Advancing Genomic Technologies and Clinical Awareness Accelerates Discovery of Disease-Associated Tandem Repeat Sequences'. *Genome Res* 32 (1): 1–27. https://doi.org/10.1101/gr.269530.120.

Harris, R. S., M. Cechova, and K. D. Makova. 2019. ‘Noise-Cancelling Repeat Finder: Uncovering Tandem Repeats in Error-Prone Long-Read Sequencing Data’. *Bioinformatics* 35 (22): 4809–11. https://doi.org/10.1093/bioinformatics/btz484.

Laabs, Björn-Hergen, Christine Klein, Jelena Pozojevic, Aloysius Domingo, Norbert Brüggemann, Karen Grütz, Raymond L. Rosales, et al. 2021. ‘Identifying Genetic Modifiers of Age-Associated Penetrance in X-Linked Dystonia-Parkinsonism’. *Nature Communications* 12 (1): 3216. https://doi.org/10.1038/s41467-021-23491-4.

Lee, L. V., E. L. Munoz, K. T. Tan, and M. T. Reyes. 2001. ‘Sex Linked Recessive Dystonia Parkinsonism of Panay, Philippines (XDP)’. *Molecular Pathology: MP* 54 (6): 362–68.

Li, H. 2018. ‘Minimap2: Pairwise Alignment for Nucleotide Sequences’. *Bioinformatics* 34 (18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. ‘The Sequence Alignment/Map Format and SAMtools’. *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Lüth, Theresa, Joshua Laβ, Susen Schaake, Inken Wohlers, Jelena Pozojevic, Roland Dominic G. Jamora, Raymond L. Rosales, et al. 2022. ‘Elucidating Hexanucleotide Repeat Number and Methylation within the X-Linked Dystonia-Parkinsonism (XDP)-Related SVA Retrotransposon in TAF1 with Nanopore Sequencing’. *Genes* 13 (1): 126. https://doi.org/10.3390/genes13010126.

Matsuura, T., P. Fang, C. E. Pearson, P. Jayakar, T. Ashizawa, B. B. Roa, and D. L. Nelson. 2006. ‘Interruptions in the Expanded ATTCT Repeat of Spinocerebellar Ataxia Type 10: Repeat Purity as a Disease Modifier?’ *Am J Hum Genet* 78 (1): 125–29. https://doi.org/10.1086/498654.

McFarland, Karen N., Jilin Liu, Ivette Landrian, Rui Gao, Partha S. Sarkar, Salmo Raskin, Mariana Moscovich, et al. 2013. ‘Paradoxical Effects of Repeat Interruptions on Spinocerebellar Ataxia Type 10 Expansions and Repeat Instability’. *European Journal of Human Genetics: EJHG* 21 (11): 1272–76. https://doi.org/10.1038/ejhg.2013.32.

Morales, Fernando, Melissa Vásquez, Eyleen Corrales, Rebeca Vindas-Smith, Carolina Santamaría-Ulloa, Baili Zhang, Mario Sirito, Marcos R. Estecio, Ralf Krahe, and Darren G. Monckton. 2020. ‘Longitudinal Increases in Somatic Mosaicism of the Expanded CTG Repeat in Myotonic Dystrophy Type 1 Are Associated with Variation in Age-at-Onset’. *Human Molecular Genetics* 29 (15): 2496–2507. https://doi.org/10.1093/hmg/ddaa123.

Nethisinghe, S., M. Kesavan, H. Ging, R. Labrum, J. M. Polke, S. Islam, H. Garcia-Moreno, et al. 2021. ‘Interruptions of the FXN GAA Repeat Tract Delay the Age at Onset of Friedreich’s Ataxia in a Location Dependent Manner’. *Int J Mol Sci* 22 (14). https://doi.org/10.3390/ijms22147507.

Netravathi, M., Pramod Kumar Pal, Meera Purushottam, Kandavel Thennarasu, Mitali Mukherjee, and Sanjeev Jain. 2009. ‘Spinocerebellar Ataxias Types 1, 2 and 3: Age Adjusted Clinical Severity of Disease at Presentation Correlates with Size of CAG Repeat Lengths’. *Journal of the Neurological Sciences* 277 (1–2): 83–86. https://doi.org/10.1016/j.jns.2008.10.016.

Pauly, Martje G., Marta Ruiz López, Ana Westenberger, Gerard Saranza, Norbert Brüggemann, Anne Weissbach, Raymond L. Rosales, et al. 2020. ‘Expanding Data Collection for the MDSGene Database: X-Linked Dystonia-Parkinsonism as Use Case Example’. *Movement Disorders: Official Journal of the Movement Disorder Society* 35 (11): 1933–38. https://doi.org/10.1002/mds.28289.

Reyes, Charles Jourdan, Björn-Hergen Laabs, Susen Schaake, Theresa Lüth, Raphaela Ardicoglu, Aleksandar Rakovic, Karen Grütz, et al. 2021. ‘Brain Regional Differences in Hexanucleotide Repeat Length in X-Linked Dystonia-Parkinsonism Using Nanopore Sequencing’. *Neurology. Genetics* 7 (4): e608. https://doi.org/10.1212/NXG.0000000000000608.

Stolle, Catherine A., Edward C. Frackelton, Jennifer McCallum, Jennifer M. Farmer, Amy Tsou, Robert B. Wilson, and David R. Lynch. 2008. ‘Novel, Complex Interruptions of the GAA Repeat in Small, Expanded Alleles of Two Affected Siblings with Late-Onset Friedreich Ataxia’. *Movement Disorders: Official Journal of the Movement Disorder Society* 23 (9): 1303–6. https://doi.org/10.1002/mds.22012.

Trinh, J., T. Luth, S. Schaake, B. H. Laabs, K. Schluter, J. Labeta, J. Pozojevic, et al. 2023. ‘Mosaic Divergent Repeat Interruptions in XDP Influence Repeat Stability and Disease Onset’. *Brain* 146 (3): 1075–82. https://doi.org/10.1093/brain/awac160.

Westenberger, A., C. J. Reyes, G. Saranza, V. Dobricic, H. Hanssen, A. Domingo, B. H. Laabs, et al. 2019. ‘A Hexanucleotide Repeat Modifies Expressivity of X-Linked Dystonia Parkinsonism’. *Ann Neurol* 85 (6): 812–22. https://doi.org/10.1002/ana.25488.

Wheeler, Vanessa C., and Vincent Dion. 2021. 'Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models'. *Journal of Huntington's Disease* 10 (1): 123–48. https://doi.org/10.3233/JHD-200426.

Wright, G. E. B., H. F. Black, J. A. Collins, T. Gall-Duncan, N. S. Caron, C. E. Pearson, and M. R. Hayden. 2020. 'Interrupting Sequence Variants and Age of Onset in Huntington's Disease: Clinical Implications and Emerging Therapies'. *Lancet Neurol* 19 (11): 930–39. https://doi.org/10.1016/S1474-4422(20)30343-4.

Objective 3: To determine if mosaic variants are present in other repeat expansion disorders using SCA27B as an example

The pathogenic threshold in recently identified repeat expansion disorders is highly discussed. In SCA27B, the first studies suggested a threshold of 300 GAA repeat units for pathogenicity, and a repeat number between 250 and 300 for reduced penetrance (D. Pellerin, Danzi, et al. 2023; H. Rafehi et al. 2023). More recent studies have proposed a lower threshold of 250 repeat units for pathogenicity, and a range between 200 and 250 for reduced penetrance (Hengel et al. 2023; David Pellerin, Heindl, et al. 2024). These studies have used a long-range PCR to determine the repeat length in the *FGF14* gene and have also identified some healthy individuals with a repeat length over 250. In this study, titled "*FGF14* repeat length and mosaic interruptions: modifiers of spinocerebellar ataxia 27B?", we wanted to distinguish the repeat motifs, mosaicism, and number of repeat interruptions present in *FGF14*-related ataxia patients and unaffected individuals.

We performed advanced genetic methods on 40 individuals with a repeat number of over 250 repeats. The 40 individuals were selected from a previously screened cohort of 304 individuals with late-onset ataxia and 190 unaffected individuals. The screening was done by determining the repeat length with a long-range PCR. The PCR product of the 40 individuals with a high repeat number was then used as input for the long-read sequencing by ONT to analyze the repeat tract. Additionally, a repeat-primed PCR was performed on these individuals to test for possible interruptions of the GAA repeats. Sanger sequencing on all individuals and long-read sequencing by Pacific Biosciences on a subset of 16 individuals were performed to confirm the identified interruptions by long-read sequencing. Mirja Thomsen performed the long-range PCR, repeat-primed PCR, and Sanger sequencing. Susen Schaake performed the long-read sequencing using ONT. The long-read sequencing by PacBio was performed in London by our collaborators Hannah MacPherson and Emil K. Gustavsson. I performed the bioinformatic analysis for the two long-read sequencing approaches.

I have adapted the pipeline from the hexanucleotide repeat motif in XDP to the trinucleotide motif in SCA27B. We have identified 27 patients with late-onset ataxia and 13 unaffected individuals with a repeat length over 250 repeats. Using Bland-Altman analyses, I validated the repeat number detection using different approaches, such as long-range PCR, ONT, and PacBio. Half of the 40 individuals, 20, had interruptions in the GAA repeat tract.

Therefore, I calculated the longest pure GAA tract of all individuals. Based on the pure GAA tract, we established a new categorization of SCA27B. A linear regression model was applied to test the association between the pure GAA length and the disease onset.

The interruption frequency was calculated for every individual with an interruption. A Bland-Altman analysis determined the concordance of the interruption frequencies between the two long-read sequencing approaches. I performed Mann-Whitney U-tests to investigate the difference in interruption frequencies between affected and unaffected individuals.

Finally, I drafted the first manuscript, figures, and tables. Additionally, I drafted the rebuttal letters for manuscript revision under the supervision of Prof. Joanne Trinh. This study demonstrated that long-read sequencing is required to detect complex repeat interruptions accurately and that the interruptions enabled a new categorization based on remaining pure GAA repeats.

Title

FGF14 repeat length and mosaic interruptions: modifiers of spinocerebellar ataxia 27B?

Authors

Joshua Laß MSc[1†], Mirja Thomsen MSc[1†], Max Borsche MD[1,2†], Theresa Lüth PhD[1†], Julia C. Prietzsche BSc[1], Susen Schaake BSc[1], Andona Milovanović MD[3], Hannah Macpherson MSc[4,5], Emil K. Gustavsson PhD[4,5], Paula Saffie Awad MD[6,7], Nataša Dragašević-Mišković MD[3], Björn-Hergen Laabs PhD[8], Inke R. König PhD[8], Ana Westenberger PhD[1], Christopher E. Pearson PhD[9], Norbert Brüggemann MD[1,2], Christine Klein MD[1], Joanne Trinh PhD[1]*

[†]These authors contributed equally to this work.

[1] Institute of Neurogenetics, University of Lübeck and University Hospital Schleswig-Holstein, 23538 Lübeck, Germany

[2] Department of Neurology, University of Lübeck and University Hospital Schleswig-Holstein, 23538 Lübeck, Germany

[3] Neurology Clinic, University Clinical Center of Serbia, Belgrade, 11000, Serbia

[4] Department of Genetics and Genomic Medicine, UCL GOS Institute of Child Health, London, WC1E 6BT, UK

[5] Department of Neurodegenerative Diseases, Institute of Neurology, UCL, London, WC1E 6BT, UK

[6] Programa de Pós-Graduação em Ciências Médicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, 90035003, Brazil

[7] Servicio de Neurología, Clínica Santa María, Santiago, 7520349, Chile

[8] Institute of Medical Biometry and Statistics, University of Lübeck, 23538 Lübeck, Germany

[9] The Hospital for Sick Children, Genetics & Genome Biology; Department of Molecular Genetics, University of Toronto, Toronto, ON M5G 0A4, Canada

*Correspondence author

Published in *Brain*, 2025, awaf183; DOI: 10.1093/brain/awaf183. Online ahead of print

Abstract

Deep intronic *FGF14* repeat expansions have been identified as a frequent genetic cause of late-onset cerebellar ataxias, explaining up to 30% of patients. Interruptions between repeats have previously been determined to impact the penetrance in other repeat expansion disorders. Repeat interruptions within *FGF14* have yet to be characterized in detail.

We utilized long-range PCR, Sanger sequencing, repeat-primed PCR, Nanopore, and PacBio sequencing to distinguish the repeat motifs, mosaicism, and number of repeat interruptions in *FGF14*-related ataxia patients and unaffected individuals.

A total of 304 patients with late-onset ataxia and 190 unaffected individuals were previously screened for repeat expansions in the *FGF14* gene by long-range PCR, identifying 37 individuals with expanded repeat lengths (≥250 repeats). These, along with three newly identified expansion carriers, were included in the present study, and advanced genetic methods were applied to investigate the repeat composition in 27 patients and 13 unaffected individuals. The expansions, based on Nanopore data, ranged from 236 to 486 repeats (SD = 60), with 20 individuals showing repeat interruptions, including complex motifs such as *GAG*, GAA*GGA*, GAAGAA*A*GAA, GAAA*A*GAAGAA*G*GAAGAA*G*GAA, GAAA*A*GAAGAA*G*GAA, and *GCA*GAAGAAGAAGAA. We calculated the longest pure GAA length from the long-read data for all 40 individuals. When comparing the pure GAA tract between patients and unaffected individuals, clusters were apparent based on greater or fewer than 200 repeats. Five ataxia patients with interruptions still had a remaining pure GAA expansion <200. We observed an association of the pure GAA length with age at onset (p=0.016, R^2=0.256). Somatically-incurred mosaic divergent repeat interruptions were discovered that affect motif length and sequence (mDRILS), which varied in number and mosaicism (frequency: 0.37-0.93). The mDRILS correlated with pure GAA length (p=0.022, R^2=0.334), with a higher mosaic frequency of interruptions in unaffected individuals compared to patients (unaffected: 0.90; patients: 0.67; p=0.009).

We demonstrate that i) long-read sequencing is required to detect complex repeat interruptions accurately; ii) repeat interruptions in *FGF14* are mosaic, have various lengths and start positions in the repeat tract, and can thereby be annotated as mDRILS, which iii) enabled us to establish a categorization based on remaining pure GAA repeats quantifying the impact of mDRILS on pathogenicity or age at onset, dependent on the interruption length and position, with high accuracy. Finally, we iv) provide evidence that mosaicism stabilizes pure GAA repeats in interrupted *FGF14* repeat expansions.

Introduction

Genetic cerebellar ataxias are commonly associated with tandem repeat expansions. The recently identified deep intronic repeat expansions of $(GAA)_n$ in the *FGF14* (Fibroblast Growth Factor) gene cause spinocerebellar ataxia 27B (SCA27B) (H. Rafehi et al. 2023; D. Pellerin, Danzi, et al. 2023). The expansion is located in the first intron and reduces *FGF14* expression, similar to the GAA expansion of Friedrich's ataxia (Ohshima et al. 1998). In SCA27B, a threshold of ≥250 repeats is considered disease-causing, whereby expansions between 250 and 300 repeats are likely pathogenic with reduced penetrance, and expansions with ≥300 repeats are fully penetrant (H. Rafehi et al. 2023; D. Pellerin, Danzi, et al. 2023). Different studies suggested repeat interruptions in *FGF14* to be non-pathogenic (D. Pellerin, Danzi, et al. 2023; D. Pellerin, Iruzubieta, et al. 2023; Ouyang et al. 2024; Mohren et al. 2024; Ando et al. 2024). However, evidence for interruption non-pathogenicity relates mainly to family studies thus far, and more in-depth analysis of the specific repeat expansion sequence, motif, and interruption length has yet to be performed. One extensive study used Nanopore sequencing in a subset of individuals and found a similar frequency of GAA*GGA* interruptions in patients and controls, where they concluded that the interruption was non-pathogenic (Mohren et al. 2024). However, interruptions have not been further characterized in terms of interruption length, position, and mosaicism. Finally, the role of shorter repeat interruptions and the impact of their position within the repeat has yet to be deciphered in detail.
Long-read sequencing has revealed an increased appreciation of the number of loci with expanded repeats, the sequence motifs, and their purity (Gall-Duncan et al. 2022; Rajan-Babu et al. 2024).

While many repeat expansion disorders are now characterized, with known *cis*-elements flanking the unstable repeat, including *FGF14* repeat tract purity, modifiers of disease expression are largely unknown (Cleary and Pearson 2003; D. Pellerin et al. 2024). The hexanucleotide repeat relevant for X-linked dystonia-parkinsonism (XDP) consists of the hexanucleotide $(AGAGGG)_n$ sequence repeated 30 to 55 times and is a strong genetic modifier of age at onset (AAO) (Aneichyk et al. 2018; Rakovic et al. 2018). A novel mosaic repeat motif pattern that deviates from the known hexanucleotide repeat motif, both in motif length and sequence (mDRILS), modifies repeat stability in XDP (Trinh et al. 2023). This genetic association in XDP demonstrates the importance of somatic mosaic genotypes and the biological plausibility of multiple germline and somatic modifiers, which may collectively contribute to repeat instability. These variations may remain undetected without assessment of single molecules. Data on the correlation between repeat length and age at onset vary between *FGF14* studies. Even studies with large sample sizes could not consistently demonstrate such an association, which contradicts the general understanding of repeat expansion disorders (Paulson 2018; Wilke et al. 2023; Iruzubieta et al. 2023). Thus, in-depth genetic methods might shed new light on this relation, which is important for clinical care and patient counseling.

Our study aimed to delineate the repeat tract sequence, mosaicism, and number of repeat interruptions in *FGF14*-related late-onset cerebellar ataxia patients and unaffected individuals (Fig. 1). Our findings show that: i) long-read sequencing is required to detect complex repeat interruptions accurately; ii) repeat interruptions in *FGF14* are mosaic, have various lengths and start positions in the repeat tract, and can thereby be annotated as mDRILS, which iii) enabled us to establish a categorization based on remaining pure GAA repeats quantifying the impact of mDRILS on pathogenicity or age at onset, dependent on the interruption length and position, with high accuracy. Finally, we iv) provide evidence that mosaicism stabilizes pure GAA repeats in interrupted *FGF14* repeat expansions.

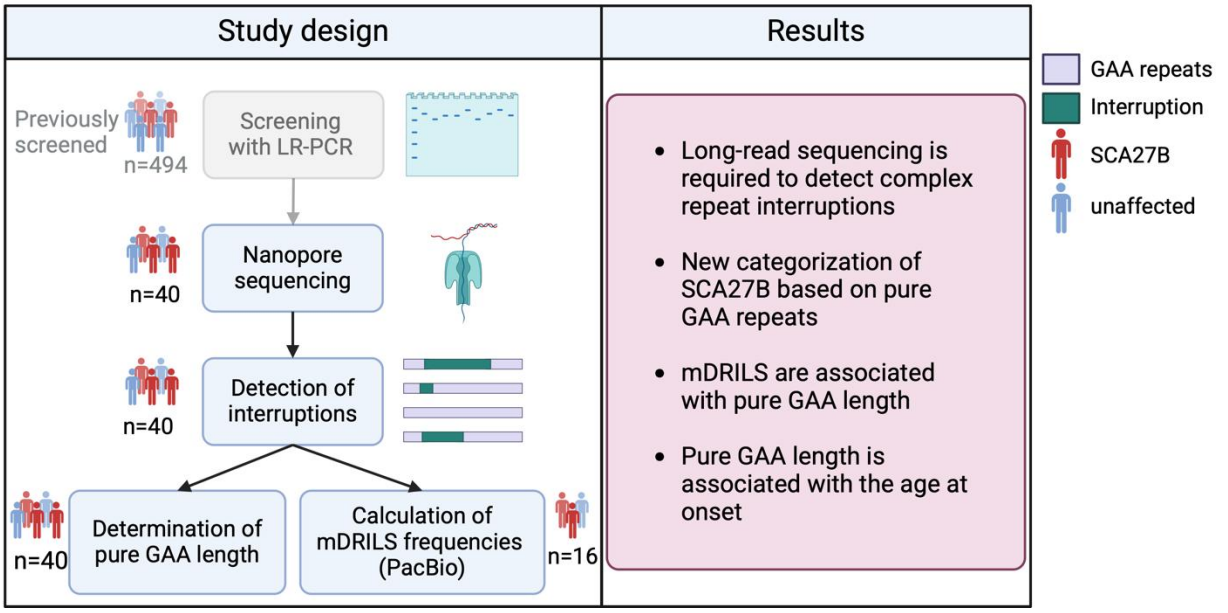

Figure 1. Overview of the study design. Initial screening of *FGF14* repeat expansions (≥250 repeats) was previously performed with long-range PCR (H. Rafehi et al. 2023; Borsche et al. 2024; Milovanovic et al. 2024; Saffie Awad et al. 2023). In the present study, individuals with detected expansions were further analyzed using long-read sequencing to detect interruptions, pure GAA tract length, and somatic mosaicism. Legend: LR-PCR = long-range PCR, PacBio = PacBio sequencing, mDRILS = mosaic divergent repeat interruption affecting length and sequence. Created in BioRender. Laß, J. (2025) https://BioRender.com/a67g089.

Methods

Participant recruitment

The individuals analyzed in greater detail in this study (n=40) were selected from previous (n=37) and ongoing (n=3) screening efforts for *FGF14* repeat expansions conducted at the Institute of Neurogenetics, Lübeck, Germany (Fig. 1). These screening efforts included i) patients with ataxia of unknown genetic cause (134 patients from Germany recruited at the tertiary referral centers for ataxia and vertigo at the Department of Neurology, University of Lübeck, Germany (H. Rafehi et al. 2023; Borsche et al. 2024), 167 patients from Serbia recruited at the Department for Neurodegenerative Diseases and Movement Disorders, Neurology Clinic at the University Clinical Center of Serbia, Belgrade, Serbia (Milovanovic et al. 2024), and two patients and one unaffected member of a Chilean family recruited at a movement disorders center in Santiago, Chile (Saffie Awad et al. 2023).); and ii) 190 unaffected individuals from Germany (H. Rafehi et al. 2023).

Control participants were recruited within the framework of extensive cross-sectional or longitudinal studies independent from ataxia research dealing with Parkinson's disease or eating behavior at the Institute of Neurogenetics, Lübeck, Germany (Kasten et al. 2013; Krause et al. 2019). All control participants underwent a structured neurological examination and were free of signs of neurological disease at the time of blood taking, which was set as age at examination.
The inclusion criterion for ataxia patients was the presence of progressive cerebellar ataxia of unknown cause. Patients with secondary forms of ataxia (lesional, toxic, inflammatory, or paraneoplastic) and known repeat-expansion SCAs (SCA1, 2, 3, 6, and 17) were excluded. The local ethics committees approved the study at the University of Lübeck, Germany, the Medical Faculty of the University of Belgrade, Serbia, and CETRAM (Centro de Estudios de Transtornos del Movimiento), Chile. All patients gave written informed consent before inclusion in the study, which was performed according to the Declaration of Helsinki. After collecting peripheral blood samples, genomic DNA was extracted using the QIAamp DNA Blood Mini Kit (Qiagen) according to the manufacturer's instructions.

Long-range PCR

Long-range PCR was performed to amplify the GAA repeat region of *FGF14* using flanking primers FGF14-F (5'-CAGTTCCTGCCCACATAGAGC-3') and FGF14-R (5'-AGCAATCGTCAGTCAGTGTAAGC-3'), as previously described (H. Rafehi et al. 2023). The predicted product is 315 bp long and includes 50 GAA repeats based on the hg38 reference. A 25 µL PCR reaction was set up using the Platinum SuperFi II Master Mix (Thermo Fisher Scientific), 5% DMSO, 0.5 µM forward and reverse primers, and 100 ng of genomic DNA, with amplification performed using a touchdown PCR protocol (Supplementary Table 1). PCR products were visualized using agarose gel electrophoresis. For fragment analysis, an M13F-tail (CACGACGTTGTAAAACGAC) was attached to the forward primer, and a third FAM-labeled primer (FAM-M13F) was added to analyze products by capillary electrophoresis on a Genetic Analyzer 3500XL (Applied Biosystems). Fragment sizes were determined using GeneMapper software (Applied Biosystems) with a GeneScan™ 1200 LIZ™ Size Standard. For samples with repeat lengths greater than 250, corresponding to products with a size greater than ~900 bp, repeat-primed PCR (RP-PCR), Sanger sequencing, Nanopore, and PacBio sequencing were conducted.

Repeat-primed PCR (RP-PCR)

RP-PCR was used to analyze the repeat composition and test for possible interruptions of the $(GAA)_n$ repeat, following a protocol and primers adapted from a previous publication (H. Rafehi et al. 2023). The design included a locus-specific primer (FGF14-R), a repeat-containing primer with an M13F-tail designed to amplify the GAA/TCC or GAAGGA/CTTCCT motif (FGF14-RP-GAA: 5'-M13F-CTTCTTCTTCTTCTTCTTCTT-3'; FGF14-RP-GAAGGA: 5'-M13F-TCCTTCTCCTTCTCCTTC-3'), and a FAM-labeled M13-primer (FAM-M13F). Cycling conditions are displayed in Supplementary Table 2. The products were analyzed on a Genetic Analyzer 3500XL, producing a ladder-like pattern indicative of repeat expansions of the respective motif.

Nanopore sequencing

A Nanopore sequencing workflow was used to analyze the repeat expansion in *FGF14*. The approach involved PCR amplification of the repeat tract, utilizing the long-range PCR products (see above) as an input.

To sequence all individuals on a single R10.4.1 flow cell, the long-range PCR products were multiplexed using the Native Barcoding Kit NBD114-96 (ONT). A 200 fmol input of the amplified PCR product was used. Library preparation was performed with the SQK-LSK114 kit (ONT). The final library was subsequently loaded onto an R10.4.1 flow cell (FLO-MIN114) and sequenced on a GridION platform as previously described (Lüth et al. 2022).

Sanger sequencing

Sanger sequencing was performed on the long-range PCR products to confirm their specificity and the repeat motif sequence. Sequencing reactions were prepared using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) with primers FGF14-F and FGF14-R. The sequencing conditions included 25 cycles of 96 °C for ten seconds, 55 °C for five seconds, and 60 °C for three minutes. Sequencing products were purified using sodium acetate precipitation and analyzed on a Genetic Analyzer 3500XL (Applied Biosystems).

Pacbio sequencing

For each sample, a 1.3X cleanup was performed using PacBio SMRTbell beads. Quality control checks were conducted using the Qubit 1x dsDNA HS kit (Thermo Fisher) for concentration and the Femto Pulse with the Genomic DNA 165kb Kit (Agilent) for fragment analysis. Samples were then processed using an adapted version of the standard PacBio multiplexed amplicon library protocol (102-359-000) and the SMRTbell Prep Kit 3.0. A final concentration of 78.38 fmol of each sample was added, and reaction volumes were divided by six for End Repair and A-tailing and by 5.4 for Adapter Ligation. During ligation, samples were barcoded with unique SMRTbell adapters, and an additional step of incubation at 65 °C for ten minutes was added before the four °C hold. Samples were pooled before a 1.2X Pronex bead clean and elution in 40 µL. Nuclease treatment was performed in a total volume of 50 µL at 37 °C for 15 min and then held at 4 °C. A second 1.2X Pronex bead cleanup was carried out with final elution in 15 µL. The final library underwent quality assessment using the Qubit and Femto Pulse. Sequencing was carried out as a standard amplicon library on the Sequel IIe using binding kit 3.2. The library was loaded at 70 pM with a movie time of 30 hours.

Bioinformatic analysis

For Nanopore sequencing, base-calling was performed using Dorado's (version 7.2.13) super accuracy model (dna_r10.4.1_e8.2_400bps_sup@v4.3.0). Read quality was analyzed with the Nanostat software (version 1.5.0). For PacBio sequencing, the BAM files were demultiplexed with the software Lima (version 2.9.0) and converted to FASTQ files using Samtools (version 1.15). For both Nanopore and PacBio sequencing, Minimap2 (version 2.22) was used to align the reads to the reference sequence with parameters for long Nanopore sequencing reads (Li 2018). SAM-to-BAM conversion and BAM file handling were conducted with the Samtools software (version 1.15) (Li et al. 2009). The next step was sorting and indexing the reads with Samtools.

The trinucleotide repeats were analyzed with the "Noise-Cancelling Repeat Finder" (NCRF, version 1.01.02) (Harris et al. 2019). Only reads with a maximum noise of 80% and a minimum of 15 detected repeat units were included in the analysis. A minimum threshold of 200 repeats was set to filter for the long allele. A maximum threshold of 100 repeats was applied to assign reads to the short allele. The repeat length was determined with the median repeat length of all reads.

As previously described, interruptions and their frequency were detected with the summary output in R (Lüth et al. 2022). The interruption frequency was then calculated by dividing the number of reads with interruption by the total number of reads for that individual. The scripts and reference file are provided at: https://github.com/joshua21997/FGF14-repeat-expansion.

Statistical analysis

The graphical representation and statistical analyses were performed in R (version 4.3.0) and Biorender. Visualization was done with the ggplot2 package (version 3.4.4). Mann-Whitney U-tests were performed for pairwise comparisons between patients and unaffected individuals, with the significance level set to $\alpha = 0.05$ for the Mann-Whitney U-tests. The adjusted significance level for multiple testing based on Bonferroni is $\alpha = 0.013$. Bland-Altman plots were used to compare the different methods. Additionally, correlation analysis was performed using a linear regression model implemented with the lm-function.

Results

Long-read sequencing can robustly detect FGF14 repeat length

Nanopore sequencing on the above-described 40 individuals with *FGF14* repeat expansions revealed a read length of the long allele at a median of 1018 bp (IQR:899-1112 bp), and the median q-score was 14.6 (IQR:14.2-15.4). The detected repeat number with Nanopore sequencing ranged from 236 to 486, and the median repeat length was 309 (IQR:276-373) (Supplementary Table 3). The fragment analysis detected comparable repeat lengths for the long allele (median repeat length: 324, IQR:288-376) (Fig. 2A). The Bland-Altman analysis between Nanopore sequencing and fragment analysis showed a slight bias, and three of 40 individuals were out of the limits of agreement (Fig. 2B).

PacBio sequencing was performed on 16/40 individuals with the *FGF14* repeat expansion as a second validation of the repeat length. The median q-score of the PacBio sequencing was 38.7 (IQR:37.7-41.5), and the median read length was 1162 bp (IQR:1108-2310 bp). The median repeat number of the long allele with PacBio sequencing was 318 (IQR:273-342).

Comparing the repeat tract length, the PacBio sequencing results were concordant with the Nanopore sequencing results and the fragment analysis results (one of 16 individuals was outside the limits of agreement) (Fig. 2C-F).

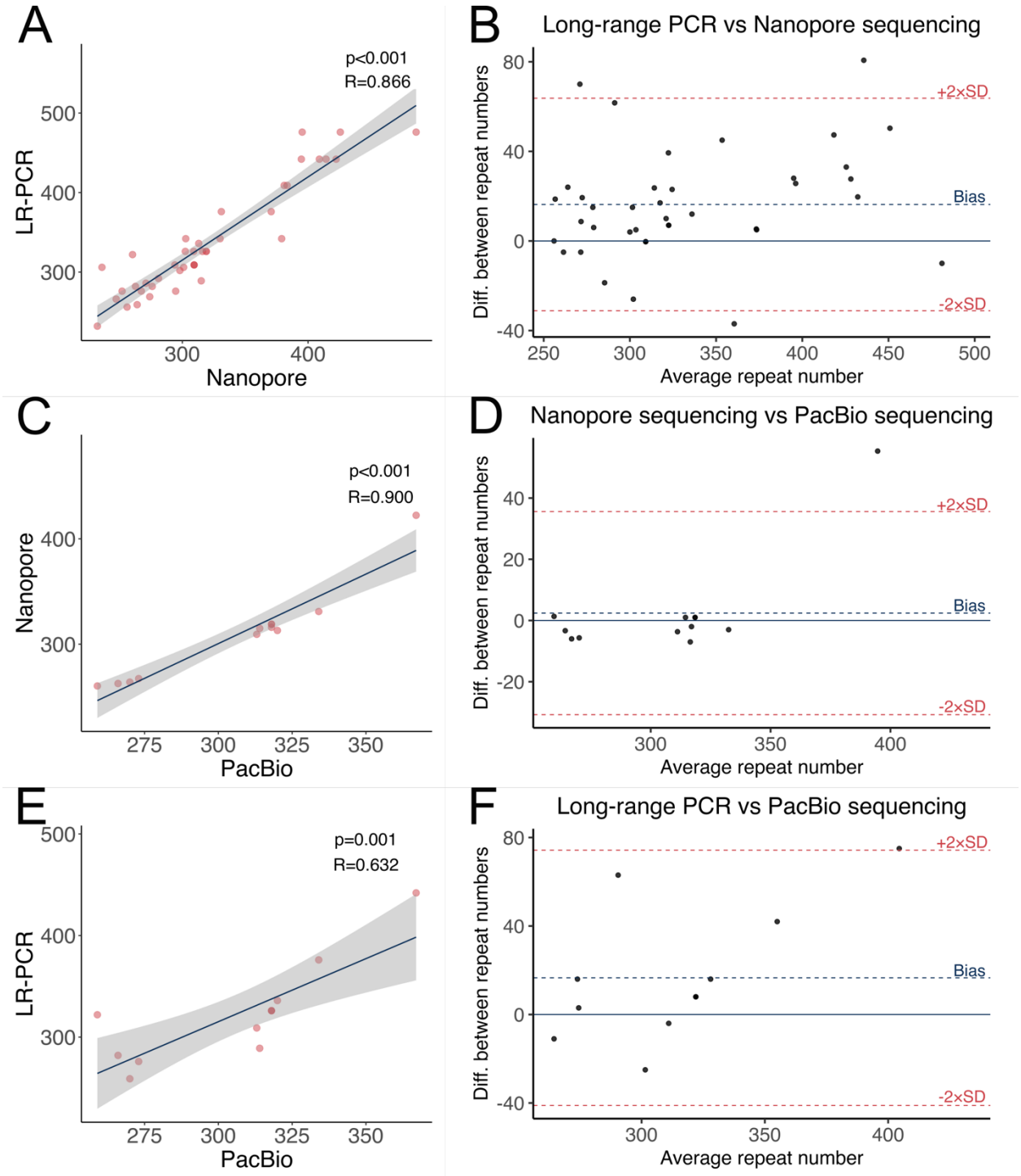

Figure 2. Comparison of the repeat length between different methods. (A) Correlation between LR-PCR and Nanopore. **(B)** Bland Altman Plots between LR-PCR and Nanopore **(C)** Correlation between Nanopore and PacBio. **(D)** Bland-Altman Plots between Nanopore and PacBio. **(E)** Correlation between LR-PCR and PacBio. **(F)** Bland-Altman Plots between LR-PCR and PacBio. A linear regression model was used for statistical analysis. Legend: LR-PCR = long-range PCR, Nanopore = Oxford Nanopore Technology sequencing, PacBio = PacBio sequencing.

The pure GAA tract predicts disease manifestation and age at onset

Next, we assessed the pure GAA tract length, without interruptions, in patients and unaffected individuals. The range of the pure GAA tract was 11 to 486 repeat units across these individuals. The longest GAA tract was used for further analyses for individuals with interruptions. The comparison between patients and unaffected individuals resulted in the identification of four distinct clusters (Fig. 3A and 4A). Recent literature has identified SCA27B patients with fewer repeat numbers than 250 (Mohren et al. 2024; Hengel et al. 2023). We observed a gap between 66 and 226 pure GAA repeats (Fig. 4A). Therefore, we used a threshold of 200 pure GAA repeat units, albeit in a suggestive manner. The first group consisted of affected patients with a pure GAA length of >200 repeat units and diagnosed with SCA27B (affected, Fig. 3A and 4A indicated in red). In the second group were unaffected individuals with a pure GAA tract >200, indicating pre-manifesting carrier status (pre-manifesting, Fig. 3A, and 4A indicated in yellow). Regarding AAO, the median AAO of SCA27B patients is reported to be 60 years (range 21-87 years) (Pellerin et al. 1993), which we used as a supposed threshold for categorizing affected and pre-manifesting individuals in Figure 3A. Of note, these six individuals had an age at examination of 6, 35, 50, 51, 53, and 80 years, respectively. The third group consisted of unaffected individuals with a pure GAA length <200 and no diagnosis (unaffected, Fig. 3A and 4A indicated in blue). The last group consisted of patients with a pure GAA tract <200 repeat units (other ataxia, Fig. 3A, and 4A indicated in green). Next, we tested the relationship between the pure GAA lengths and AAO (Fig. 3B). There was an association between GAA repeat length and AAO (p=0.016, R^2=0.256) (Fig. 4B) in patients with a pure GAA tract >250 repeat units.

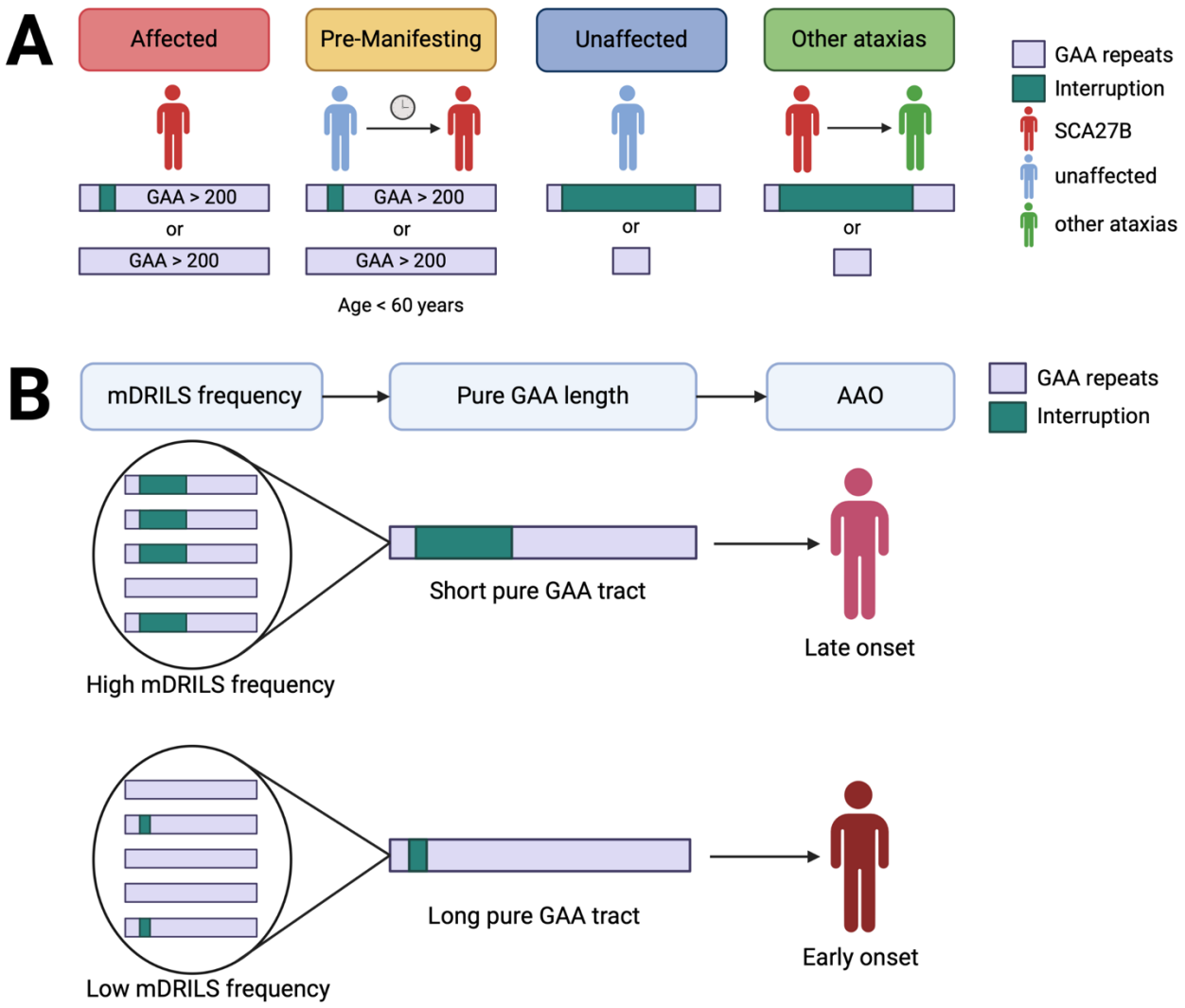

Figure 3. Schematic illustration of: (A) New categorization of individuals based on pure GAA length. Affected individuals with SCA27B, represented in red, have either >200 pure GAA repeats (i.e., without interruption) or >200 pure GAA repeats despite an interruption. Pre-manifesting carriers with an earlier age (<60 years of age) are highlighted in yellow. Unaffected individuals have long repeat interruptions and short pure GAA (<200 repeats) tracts. Lastly, misdiagnoses of SCA27B ataxia can occur in case of a long repeat interruption resulting in uninterrupted GAA repeats of < 200. **(B) Relationship between mDRILS frequency, pure GAA length, and age at onset.** mDRILS are modifiers of repeat stability and are more frequent in late-onset SCA27B than early-onset SCA27B. Legend: mDRILS = mosaic divergent repeat interruptions affecting length and sequence, AAO = age at onset. Created in BioRender. Klein, C. (2025) https://BioRender.com/e06b901.

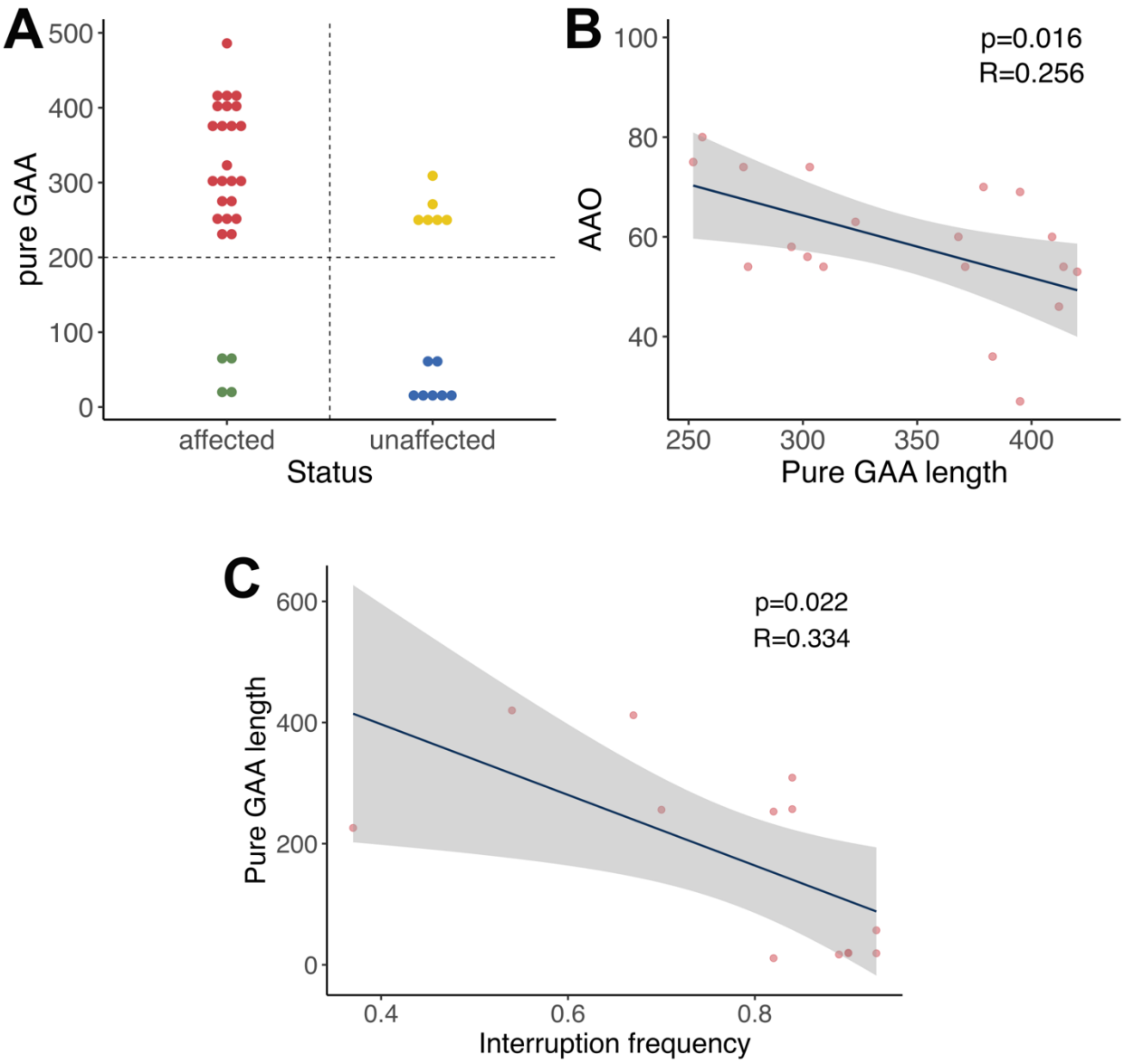

Figure 4. Analysis of the pure GAA length was determined by Nanopore sequencing. (A) Differences between affected and unaffected repeat expansion carriers in pure GAA length. Colour coding as described in Figure 3A. (i.e., red = affected, yellow = pre-manifesting, blue = unaffected, green = other ataxias) **(B)** Correlation between age at onset (AAO) and pure GAA length. **(C)** Correlation between interruption frequency and pure GAA length. A linear regression model was used for statistical analysis. The adjusted significance level is $\alpha = 0.013$.

Novel repeat interruptions were found through long-read sequencing

Seven different repeat interruptions were detected (Table 1,2). Among these, the most frequently observed interruption motif was (GAA*GGA*)$_n$, characterized by a single adenosine to guanine conversion (A-to-*G*). This interruption motif was found in twelve individuals (six affected and six unaffected). The (GAA*GGA*)$_n$ interruption motif was repeated two to 158 times and was located at the 5'end of the repeat tract, (GAA)$_{1\text{-}20}$(GAA*GGA*)$_{2\text{-}158}$ (GAA)$_n$.

Another variant repeat motif was (GAA)$_3$GAAA*A*GAAGAA*G*GAAGAA*G*GAA(GAA)$_n$. This motif was identified in two unaffected individuals, including one deletion and two insertions. A similar motif, (GAA)$_3$*G*AAA*A*GAAGAA*G*GAA(GAA)$_n$, was identified in another unaffected. The shortest interruption motif observed was (GAA)$_7$(*GAG*)$_3$(GAA)$_n$, detected in one patient.

Two motifs were exclusively detected by Nanopore sequencing. One motif was (GAA)$_5$(GAAGAA*A*GAA)$_3$(GAA)$_n$, which resulted from an insertion of an adenosine. The other motif was (GAA)$_4$(GAA*GAG*)$_2$(GAA)$_n$. Both motifs were detected in patients.

The most complex motif was (GAA)$_{22}$(*GCA*GAAGAAGAA)$_3$(*GCA*GAA)(*GCA*GAAGAAGAAGAA)$_{24}$ (*GCA*GAA)$_{12}$(*GCAGCA*GAA)$_3$(*GCA*GAA)$_{13}$(*GCAGCA*GAA*GCA*GAA)$_4$(*GCA*GAAGAAGAA GAA)$_6$ (GAA)$_{22}$*GAG*(GAA)$_n$. This motif was observed in two affected members and one unaffected member of one family. However, the affected members of this family have a phenotype that differs from that of reported *FGF14* SCA27B patients (D. Pellerin, Iruzubieta, et al. 2023).

Mosaic divergent repeat interruptions are associated with the pure GAA stability

Given the detection of mDRILS in XDP, we investigated the mosaicism of *FGF14* repeat interruptions in the long-read data. We calculated the mosaic frequency of the repeat motifs for each individual. The mosaicism obtained from Nanopore sequencing was similar to that from PacBio sequencing (p=0.017, R^2=0.364) (Fig. 5A, B). We utilized PacBio sequencing data for mosaic frequency calculations due to its higher q-score. Notably, the repeat interruption mosaicism was negatively associated with pure GAA length (p=0.022, R^2=0.334) (Fig. 4C). The higher the mosaic frequency, the shorter the pure GAA length.

Table 1: Overview of the unaffected individuals with an expanded *FGF14* repeat expansion.

ID	Status	Previously Published (PMID)	RN (ONT)	RN (LR-PCR)	RN (PacBio)	Interruption frequency (PacBio)	Pure GAA	Interruption in the long allele (ONT)	Methods confirming interruption
L-3013	Unaffected	36493768	316	326	318	0.90	20	GAA*GGA*	PacBio, Sanger
L-3329	Unaffected	36493768	271	286	276	-	271	-	Sanger
L-3344	Unaffected	36493768	319	326	318	0.90	19	GAA*GGA*	PacBio, Sanger
L-3479	Unaffected	36493768	260	322	259	0.93	57	GAA*GGA*	PacBio, Sanger
L-3501	Unaffected	36493768	319	326	318	0.93	19	GAA*GGA*	PacBio, Sanger
L-3656	Unaffected	36493768	294	309	-	-	11	GAA*GGA*	-
L-3657	Unaffected	36493768	309	326	-	-	309	-	Sanger
L-8761	Unaffected	36493768	264	259	270	0.84	257	*GAAAA*GAAGAA*GGAA*	PacBio, Sanger
L-8782	Unaffected	36493768	256	256	-	-	256	-	-
L-8934	Unaffected	36493768	315	289	314	0.89	17	GAA*GGA*	PacBio, Sanger
L-10408	Unaffected	36493768	263	282	266	0.82	253	*GAAAA*GAAGAA*GGAA*GAA*GGAA*	PacBio, Sanger
L-10410	Unaffected	36493768	281	292	-	-	243	*GAAAA*GAAGAA*GGAA*GAA*GGAA*	Sanger
L-14996	Unaffected	37246629	309	309	-	-	66	*GCA*GAAGAAGAAGAA	-

Legend: RN = repeat number, ONT = Oxford Nanopore Technology sequencing, LR-PCR = long-range PCR, PacBio = PacBio sequencing, Sanger = Sanger sequencing. Repeat motif variations are indicated by the underlined and italicized nucleotides.

Table 2: Overview of the patients with an expanded *FGF14* repeat expansion.

ID	Status	Previously Published (PMID)	RN (ONT)	RN (LR-PCR)	RN (PacBio)	Interruption frequency (PacBio)	Pure GAA	Interruption in the long allele (ONT)	Methods confirming interruption
L-14575	Patient	36493768, 37861706	426	476	366	0.54	420	GAA*GGA*	PacBio, Sanger
L-14630	Patient	36493768	313	336	320	0.84	309	GAA*GGA*	PacBio, Sanger
L-14846	Patient	38487929	371	376	-	-	371	-	Sanger
L-14853	Patient	38487929	298	302	-	-	20	GAA*GGA*	Sanger
L-14894	Patient	38487929	302	326	-	-	302	-	Sanger
L-14904	Patient	38487929	247	266	-	-	247	-	Sanger
L-14911	Patient	38487929	395	476	-	-	395	-	Sanger
L-14995	Patient	37246629	309	309	313	0.46	66	*GCA*GAAGAAGAAGAA	PacBio, Sanger
L-14997	Patient	37246629	309	309	-	-	66	*GCA*GAAGAAGAAGAA	Sanger
L-15166	Patient	36493768	331	376	334	0.37	226	GAA*GGA*	PacBio
L-15713	Patient	36493768, 37861706	303	342	-	-	303	-	Sanger
L-15739	Patient	36493768, 37861706	252	276	-	-	252	-	Sanger
L-15754	Patient	36493768	274	269	-	-	274	-	Sanger
L-15764	Patient	36493768, 37861706	330	342	342	-	323	GAA*GAG*	Not confirmed
L-15891	Patient	36493768	301	306	-	-	20	GAA*GGA*	Sanger
L-17665	Patient	36493768	422	442	367	0.67	412	GAA*GGA*	PacBio, Sanger
L-17672	Patient	36493768, 37861706	414	442	-	-	414	-	Sanger
L-18362	Patient	36493768	486	476	-	-	486	-	Sanger
L-18384	Patient	36493768, 37861706	409	442	-	-	409	-	Sanger
L-20363	Patient	36493768, 37861706	381	409	390	-	368	GAA*GAAA*GAA	Not confirmed
L-20409	Patient	-	379	342	-	-	379	-	Sanger
L-20618	Patient	38487929	395	442	-	-	395	-	Sanger
L-20635	Patient	38487929	236	306	-	-	236	-	Sanger
L-20648	Patient	38487929	383	409	-	-	383	-	-
L-20656	Patient	38487929	276	282	-	-	276	-	Sanger
L-22657	Patient	-	295	276	-	-	295	-	Sanger
L-22867	Patient	-	267	276	273	0.7	256	*GAG*	PacBio, Sanger

Legend: RN = repeat number, ONT = Oxford Nanopore Technology sequencing, LR-PCR = long-range PCR, PacBio = PacBio sequencing, Sanger = Sanger sequencing. Repeat motif variations are indicated by the underlined and italicized nucleotides.

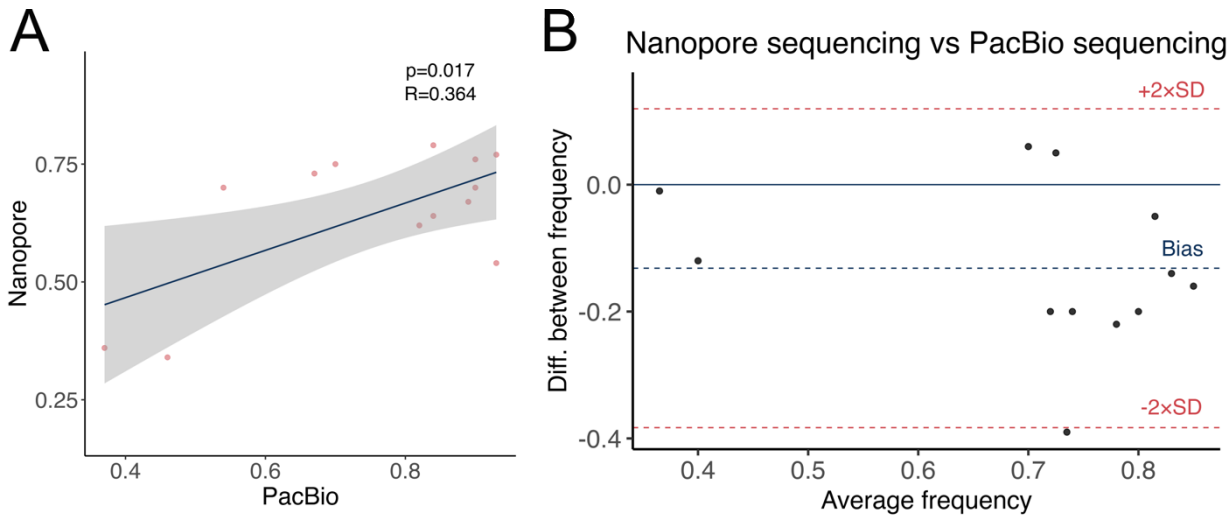

Figure 5. Comparison of the interruption frequency of all motifs between Nanopore and PacBio sequencing. (A) Correlation between Nanopore and PacBio sequencing. **(B)** Bland-Altman plot between Nanopore and PacBio sequencing. A linear regression model was used for statistical analysis.

To compare the interruption frequency between groups of unaffected individuals and patients, we only included patients with the SCA27B phenotype; thus, a previously published Chilean family with a different phenotype was excluded (Saffie Awad et al. 2023). Unaffected individuals had higher interruption frequencies (0.90, IQR:0.89-0.90) compared to patients (0.67, IQR:0.54-0.70) (p=0.009, Z=-2.627) (Fig. 6A). The most common interruption motif (GAA*GGA*)$_n$ had a mosaic frequency ranging from 0.37 to 0.93. Patients with (GAA*GGA*)$_n$ exhibited a lower median frequency of 0.61 (IQR:0.50-0.71) compared to unaffected individuals, who had a median frequency of 0.90 (IQR:0.90-0.93) (p=0.014, Z=-2.470) (Fig. 6B). The mosaic frequencies of the GAAA*A*GAAGAA*G*GAAGAA*G*GAA motif and the GAAA*A*GAAGAA*G*GAA motif were similar, at 0.82 and 0.84, respectively. The shortest interruption motif, *GAG,* had a slightly lower frequency of 0.70 than the other motifs. In contrast to the other motifs, the most complex motif had the lowest frequency with 0.46.

Further investigation of repeat interruptions was performed to rule out DNA contamination from the non-expanded short allele. The median repeat number on the short allele was 18 (IQR:17-19). Repeat interruptions on the short allele were seen in three individuals (two patients and one unaffected) (Supplementary Table 4). However, the interruption motifs in the short allele did not overlap with the expanded allele patterns.

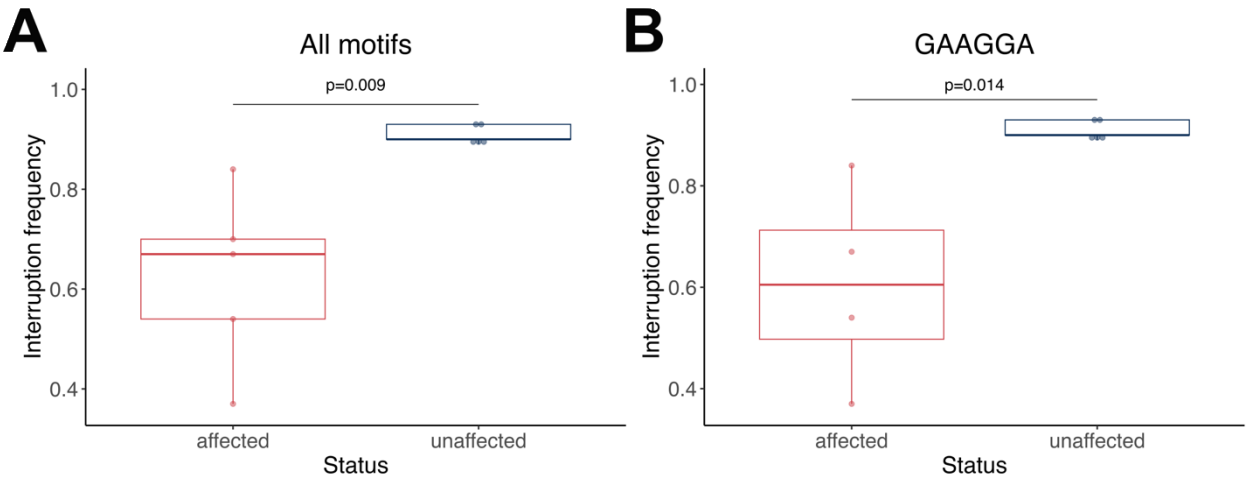

Figure 6. Analysis of the interruption frequency. Differences between affected and unaffected individuals in interruption frequency of all motifs **(A)** and the (GAA*GGA*)$_n$ motif **(B)**. Mann-Whitney U tests were performed for statistical analysis. The adjusted significance level is α= 0.013.

While some interrupting sequences compromise the stability or toxicity of GAA repeats, others may stabilize the repeat or exert more severe damaging effects. Thus, we investigated the 5'-flanking regions of the short and long alleles in individuals with an interruption. All 20 individuals had a CTTTCTGT motif in the 5'-flanking region of the repeat motif on the long allele. For the short allele, three different motifs were identified. In ten of the 20 individuals with an interruption, the motif CTTTCTGTGTAGTCATAGTACCCC was detected. The motif CTTTCTGTG was identified in nine individuals. The third motif, CTTTCTGAG, was found in one individual. The 5'-flanking region of the short or long allele did not correlate with complex motifs or interruption frequencies.

Discussion

Following our discovery of mDRILS in X-linked dystonia-parkinsonism (Trinh et al. 2023), we now observe that mDRILS can also act as modifiers of penetrance and age of onset (AAO) in a much more common condition, SCA27B. In XDP, the *TAF1* insertion is only present in patients, whereas the FGF14 repeat expansion is observed in SCA27B patients and unaffected individuals. In contrast to XDP, unaffected individuals have a higher frequency of mDRILS. We consider mDRILS an important finding of translational relevance, as it suggests a novel, shared mechanism across different repeat expansion disorders that has the potential to predict disease manifestation in individual repeat expansion carriers in a personalized fashion and represents a somatic phenomenon that may be amenable to environmental and lifestyle changes. To analyze *FGF14* intronic repeat expansions, we propose a new concept to investigate and interpret their potential pathogenicity. We challenge the previously held, more simplistic view of repeat interruptions, abolishing the pathogenic effect of the expanded repeat in general. Specifically, we propose four different scenarios, all related to the length of the pure GAA repeat tract: first, people with ataxia and an uninterrupted *FGF14* intronic repeat length of >200 repeats have SCA27B, despite the presence of an interrupted repeat, typically at the 5' end of the expansion. All the patients in this group were examined by an experienced ataxia/movement disorder specialist and had a clinical syndrome consistent with the established SCA27B phenotype. Second, people with the same molecular constellation and age <60 years are considered to be in their premanifesting phase of SCA27B. Third, people with repeat interruptions resulting in a pure GAA repeat tract of <200 will not go on to develop SCA27B. Fourth, in patients previously presumed with SCA27B and carrying a large repeat interruption with <200 uninterrupted GAA repeats, the diagnosis needs to be revisited and likely revised to another type of ataxia (Fig. 3A). An example is the Chilean family with two members previously diagnosed with SCA27B but with a phenotype that differs from reported *FGF14* SCA27B patients (D. Pellerin, Iruzubieta, et al. 2023), where we have identified a pure GAA tract with 65 repeat units. The threshold of 200 uninterrupted GAA repeats is a suggestion based on the data from the present study, albeit limited by our relatively low sample size. Therefore, a larger sample size is needed to validate and refine the real threshold of pure GAA repeats. Of further note, we here only included individuals with ≥250 repeats using long-range PCR, which was the starting point of our initial study based on the state of the literature at that point in time. However, our present data suggest that individuals carrying repeat expansions in the range of 200-250 repeats warrant further investigation, as this group may also harbor uninterrupted GAA repeat stretches of ~200 repeats in length.

According to the published recommendations, this may further significantly broaden the group of *FGF14* GAA repeat expansion carriers who will likely manifest disease but are currently not considered at risk.

Otherwise, our data imply that not all interruptions can be considered non-pathogenic, as, for example, five ataxia patients with interruptions still had a remaining pure GAA expansion >200. As many studies screened in unspecified ataxia cohorts for repeat expansions in *FGF14* to debunk the clinical phenotype of SCA27B without considering interruptions, our data demonstrate that 14% (4/28) of our ataxia patients with *FGF14* repeat expansions had rather non-pathogenic expansions if the presence of mDRILS is taken into account. Consideration of the pure GAA repeat tract and mDRILS is required, especially if the phenotype is not entirely consistent (e.g., early-onset disease, additional movement disorder features) with the typical clinical picture of SCA27B (i.e., pure late-onset cerebellar ataxia +/- episodic features). Moreover, the accurate detection of pure GAA repeats enabled by advanced sequencing methods such as Nanopore and PacBio sequencing might allow more robust genotype-phenotype relations, regarding the age at onset and in the context of disease severity and progression. Utilizing long-read sequencing in *FGF14* repeat expansion disorders in future clinical practice is feasible.

Notably, this situation in SCA27B extends beyond our previous findings in X-linked dystonia-parkinsonism, where the pathogenic insertion in the *TAF1* gene is the clear-cut cause in all affected individuals, and mDRILS in the hexanucleotide repeats within this insertion act as a modifier of AAO only. Intriguingly, in addition to the penetrance-determining length of the pure GAA tract in SCA27B, we can assume a similar AAO-modifying effect by mDRILS in *FGF14* as well, with a higher mDRILS frequency resulting in shorter uninterrupted GAA tracts, which, in turn, are associated with a later AAO (Fig. 3B). As a relationship between repeat length and age at onset is well-established in the field of repeat expansion disorders and, therefore, is expected to occur likewise in *FGF14*-related disease, our study shows the importance of taking into account mDRILS better to detect phenotypic correlations and impact on repeat stability (Paulson 2018). The mDRILS in the XDP-relevant and *FGF14* repeats may affect repeat instability by modifying the propensity to form unusual mutagenic DNA structures, as observed with the interruptions of *FMR1* (FXS) and *ATXN1* (SCA1) (Yousuf et al. 2022).

Like other disease-associated repeat expansions, including the GAA expansions causing Friedreich's ataxia (Al-Mahdawi et al. 2018; Cleary and Pearson 2003), SCA27B-associated *FGF14* repeat expansions arise on specific haplotypes. They can have interruptions in the

repeats, suggesting the contribution of *cis*-elements both flanking and within the unstable repeat, respectively (H. Rafehi et al. 2023; D. Pellerin et al. 2024; De et al. 2024; Miyatake et al. 2024). There is a possibility that a similar mechanism and consequence of interrupting mDRILS play a role in Friedreich's ataxia. However, long-read sequencing on repeat interruption motifs in Friedrich's ataxia has not yet been performed. Evidence suggests an association of specific repeat interruptions with specific haplotypes (flanking *cis*-elements) that may predispose to parent-to-offspring (germline) repeat expansions (H. Rafehi et al. 2023; D. Pellerin et al. 2024; De et al. 2024; Miyatake et al. 2024). Possible mechanistic paths by which such flanking sequence changes may act, which may involve altered predisposition to variable chromatic impact and/or unusual nucleic acid structure formation (D. Pellerin et al. 2024; De et al. 2024; Miyatake et al. 2024).

Importantly, the repeat sequence variations in *FGF14* we report herein arise somatically in post-natal tissues. We did not observe an association of the motif within the 5'-flanking region of the short allele and interruption motifs. Though our sample size is small, more individuals with interruptions are needed to examine this thoroughly. Experimental *in vivo* elucidation of the mechanism by which an intra-repeat or a flanking *cis*-element may act to modulate repeat instability, germline, or somatic instability can be quite complex. Animal studies of the SCA7 CAG tract (Libby et al. 2008; 2003), *in vitro* studies of the interrupted versus pure repeats of the SCA1/*ATXN1*, SCA10/*ATXN10*, and FXS/*FMR1* loci (Mulvihill et al. 2005; Volle and Delaney 2013; Hagerman et al. 2009), and recent analysis of post-mortem tissues of the DM1/*DMPK*, *C9orf72,* and *FGF14* locus suggest chromatin compaction differences may mediate the function of *cis*-elements upon repeat instability (D. Pellerin et al. 2024; Barbé et al. 2017; Mirceta et al. 2024).

mDRILS-mediated alterations of the secondary and tertiary structures assumed by the repeat-containing mRNA may alter RNA-foci formation and can alter the proteins that may be bound and possibly sequestered from their normal functions. It is notable that the cerebellar ataxia, neuropathy, and vestibular areflexia syndrome (CANVAS)-associated $(AAGGG)_n$•$(CCCTT)_n$, but not the non-pathogenic expanded repeat motifs, including $(AAAAG)_n$•$(CTTTT)_n$, in *RFC1* can assume both triplex and quadruplex structures (Abdi et al. 2023; Hisey et al. 2024; Wang et al. 2024; Singh et al. 2024; Kudo et al. 2024). The formation of these structures highlights the pathogenicity of the repeat sequence.

It can alter how the repeats are metabolized (replication/repair) (Singh et al. 2024), which proteins can be bound to, and reveal novel ways they may be therapeutically targeted. For example, the ligands TMPyP4 and BRACO-19 (Abdi et al. 2023; Singh et al. 2024) can bind these pathogenic structures. Somatically incurred mDRILS of the pathogenic *FGF14* repeat may also lead to pathogenic variations of spliceoforms, translation (exonization), repeat-associated non-AUG (RAN) translation, ribosomal frameshifting, ribosomal pausing, transcriptional slippage in the repeat, or repeat instability (Trinh et al. 2023; Stochmanski et al. 2012; Gaspar et al. 2000; Toulouse et al. 2005; Aviner et al. 2024; Stein et al. 2022; Parsons et al. 1998).

The assessment of the *FGF14* mDRILS across time and generations would be beneficial in elucidating the potential impact of somatic changes in the mDRILS frequency on disease progression or severity. This study did not perform longitudinal monitoring of patients with SCA27B and their mDRILS frequency. However, our group previously investigated the mDRILS frequency in families affected by XDP, proving that a higher frequency may stabilize the repeats across generations (Laß et al. 2024). Thus, future studies should aim for an analogous exploration of *FGF14* mDRILS across time within individual patients and generations.

Overall, all methods (RP-PCR, Sanger sequencing, Nanopore sequencing, and PacBio sequencing) were concordant and detected the same interruptions. We confidently identified five out of seven different repeat interruption motifs within *FGF14.* Nanopore sequencing analysis performed thus far in *FGF14*-related disease could demonstrate that interruptions are present in unaffected and ataxia patients. However, it has not yet explored 1) the interruptions' position and 2) their impact on repeat expansion pathogenicity at the individual level (Mohren et al. 2024).

Limitations of our study include the relatively small sample size compared to other studies. Moreover, we cannot independently support our estimate of pathogenicity based on advanced genetic methods, as we do not investigate *FGF14* expression, and further biomarkers robustly distinguishing SCA27B from other ataxias are not available to date. Thus, developing such, preferably accessible biomarkers would benefit the field. Future studies with larger sample sizes, longitudinal examinations, and more in-depth clinical data are needed to investigate the influence of pure GAA repeat number on age at onset and a potential influence on disease severity and progression. It would be valuable to monitor the mDRILS and their impact on disease progression and severity over a lifetime. Our study remains primarily observational and descriptive, necessitating further refinement of mechanistic novelty.

Finally, in the context of mosaicism, additional biomaterials such as fibroblasts, induced pluripotent stem cells, or post-mortem brain tissue should be investigated for mDRILS. Future studies should clarify whether the observed mDRILS pattern represents a universal modifier across repeat expansion disorders or a disease-specific phenomenon.

In conclusion, this study highlights the importance of an in-depth, multi-method assessment of repeat tract purity in repeat expansion disorders. The correlation of mosaic interruptions with the repeat stability and indirectly patient status suggests a protective effect of mDRILS. Long-read sequencing can uncover the length of the pure GAA motif along with repeat interruptions. However, it is warranted that the assessment of repeat interruptions, the pure GAA tract, and mDRILS be extended to other repeat expansion disorders and that their potential protective mechanism(s) be elucidated thoroughly. While our findings refine both the diagnosis of (*FGF14*-related) ataxia and the prognosis in non-manifesting or not-yet-manifesting repeat expansion carriers, they also highlight the necessity of including repeat interruption analysis not only in the research setting but also in diagnostic testing for SCA27B to avoid false-positive or false-negative testing results and interpretation thereof.

Data availability

The data supporting this study's findings are available from the corresponding author upon reasonable request.

Supplementary material

Supplementary material is available at *Brain* online.

Acknowledgment

We express our deepest gratitude to the families and patients participating in this study. We acknowledge Prof. Katja Lohmann for providing long-range and repeat-primed PCR analysis data of previously published and a small set of unpublished FGF14 expansion carriers. We recognize her comments on the manuscript. We would also like to thank Frauke Hinrichs for her technical assistance.

Funding

This study was supported by the German Research Foundation (FOR2488 to J.T., C.K., Heisenberg grant, J.T., BR4328.2-1, GRK1957, N.B.), the Else Kröner Fresenius Foundation (EKFS, J.T.), and the Canadian Institutes of Health Research (FRN-148910 and FRN-173282

to C.E.P.). C.E.P. holds a Tier 1 Canada Research Chair in Disease-Associated Genome Instability.

Competing interests

C.K. serves as a medical advisor to Centogene, Takeda, and Retromer Therapeutics and received speaking honoraria from Desitin and Bial. A.W. serves as an advisor for medical writing to CENTOGENE GmbH. N.B. received honoraria from Abbott, Abbvie, Biogen, Biomarin, Bridgebio, Centogene, Esteve, Ipsen, Merz, Teva, and Zambon. M.B. receives honoraria from Bial. The remaining authors report no disclosures.

References

Abdi, Mohammad Hossein, Bita Zamiri, Gholamreza Pazuki, Soroush Sardari, and Christopher E. Pearson. 2023. 'Pathogenic CANVAS-Causing but Not Nonpathogenic RFC1 DNA/RNA Repeat Motifs Form Quadruplex or Triplex Structures.' *The Journal of Biological Chemistry* 299 (10): 105202. https://doi.org/10.1016/j.jbc.2023.105202.

Al-Mahdawi, Sahar, Heather Ging, Aurelien Bayot, Francesca Cavalcanti, Valentina La Cognata, Sebastiano Cavallaro, Paola Giunti, and Mark A. Pook. 2018. 'Large Interruptions of GAA Repeat Expansion Mutations in Friedreich Ataxia Are Very Rare.' *Frontiers in Cellular Neuroscience* 12:443. https://doi.org/10.3389/fncel.2018.00443.

Ando, M., Y. Higuchi, J. Yuan, A. Yoshimura, F. Kojima, Y. Yamanishi, Y. Aso, et al. 2024. 'Clinical Variability Associated with Intronic FGF14 GAA Repeat Expansion in Japan'. *Ann Clin Transl Neurol* 11 (1): 96–104. https://doi.org/10.1002/acn3.51936.

Aneichyk, T., W. T. Hendriks, R. Yadav, D. Shin, D. Gao, C. A. Vaine, R. L. Collins, et al. 2018. 'Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly'. *Cell* 172 (5): 897-909 e21. https://doi.org/10.1016/j.cell.2018.02.011.

Aviner, R., T. T. Lee, V. B. Masto, K. H. Li, R. Andino, and J. Frydman. 2024. 'Polyglutamine-Mediated Ribotoxicity Disrupts Proteostasis and Stress Responses in Huntington's Disease'. *Nat Cell Biol* 26 (6): 892–902. https://doi.org/10.1038/s41556-024-01414-x.

Barbé, Lise, Stella Lanni, Arturo López-Castel, Silvie Franck, Claudia Spits, Kathelijn Keymolen, Sara Seneca, et al. 2017. 'CpG Methylation, a Parent-of-Origin Effect for Maternal-Biased Transmission of Congenital Myotonic Dystrophy.' *American Journal of Human Genetics* 100 (3): 488–505. https://doi.org/10.1016/j.ajhg.2017.01.033.

Borsche, M., M. Thomsen, D. J. Szmulewicz, B. Lubbers, F. Hinrichs, P. J. Lockhart, K. Lohmann, C. Helmchen, and N. Bruggemann. 2024. 'Bilateral Vestibulopathy in RFC1-Positive CANVAS Is Distinctly Different Compared to FGF14-Linked Spinocerebellar Ataxia 27B'. *J Neurol* 271 (2): 1023–27. https://doi.org/10.1007/s00415-023-12050-0.

Cleary, J. D., and C. E. Pearson. 2003. 'The Contribution of Cis-Elements to Disease-Associated Repeat Instability: Clinical and Experimental Evidence'. *Cytogenet Genome Res* 100 (1–4): 25–55. https://doi.org/10.1159/000072837.

De, Tiyasha, Pooja Sharma, Bharathram Upilli, A. Vivekanand, Shreya Bari, Akhilesh Kumar Sonakar, Achal Kumar Srivastava, and Mohammed Faruq. 2024. 'Spinocerebellar Ataxia Type 27B (SCA27B) in India: Insights from a Large Cohort Study Suggest Ancient Origin.' *Neurogenetics* 25 (4): 393–403. https://doi.org/10.1007/s10048-024-00770-y.

Gall-Duncan, T., N. Sato, R. K. C. Yuen, and C. E. Pearson. 2022. 'Advancing Genomic Technologies and Clinical Awareness Accelerates Discovery of Disease-Associated Tandem Repeat Sequences'. *Genome Res* 32 (1): 1–27. https://doi.org/10.1101/gr.269530.120.

Gaspar, C., M. Jannatipour, P. Dion, J. Laganiere, J. Sequeiros, B. Brais, and G. A. Rouleau. 2000. 'CAG Tract of MJD-1 May Be Prone to Frameshifts Causing Polyalanine Accumulation'. *Hum Mol Genet* 9 (13): 1957–66. https://doi.org/10.1093/hmg/9.13.1957.

Hagerman, Katharine A., Haihe Ruan, Kerrie Nichol Edamura, Tohru Matsuura, Christopher E. Pearson, and Yuh-Hwa Wang. 2009. 'The ATTCT Repeats of Spinocerebellar Ataxia Type 10 Display Strong Nucleosome Assembly Which Is Enhanced by Repeat Interruptions.' *Gene* 434 (1–2): 29–34. https://doi.org/10.1016/j.gene.2008.12.011.

Harris, R. S., M. Cechova, and K. D. Makova. 2019. 'Noise-Cancelling Repeat Finder: Uncovering Tandem Repeats in Error-Prone Long-Read Sequencing Data'. *Bioinformatics* 35 (22): 4809–11. https://doi.org/10.1093/bioinformatics/btz484.

Hengel, H., D. Pellerin, C. Wilke, Z. Fleszar, B. Brais, T. Haack, A. Traschutz, L. Schols, and M. Synofzik. 2023. 'As Frequent as Polyglutamine Spinocerebellar Ataxias: SCA27B in a Large German Autosomal Dominant Ataxia Cohort'. *Mov Disord* 38 (8): 1557–58. https://doi.org/10.1002/mds.29559.

Hisey, Julia A., Elina A. Radchenko, Nicholas H. Mandel, Ryan J. McGinty, Gabriel Matos-Rodrigues, Anastasia Rastokina, Chiara Masnovo, et al. 2024. 'Pathogenic CANVAS (AAGGG)n Repeats Stall DNA Replication Due to the Formation of Alternative DNA Structures.' *Nucleic Acids Research* 52 (8): 4361–74. https://doi.org/10.1093/nar/gkae124.

Iruzubieta, P., D. Pellerin, A. Bergareche, I. Albajar, E. Mondragon, A. Vinagre, R. Fernandez-Torron, et al. 2023. 'Frequency and Phenotypic Spectrum of Spinocerebellar Ataxia 27B and Other Genetic Ataxias in a Spanish Cohort of Late-Onset Cerebellar Ataxia'. *Eur J Neurol* 30 (12): 3828–33. https://doi.org/10.1111/ene.16039.

Kasten, Meike, Johann Hagenah, Julia Graf, Anne Lorwin, Eva-Juliane Vollstedt, Elke Peters, Alexander Katalinic, Heiner Raspe, and Christine Klein. 2013. 'Cohort Profile: A Population-Based Cohort to Study Non-Motor Symptoms in Parkinsonism (EPIPARK).' *International Journal of Epidemiology* 42 (1): 128–128k. https://doi.org/10.1093/ije/dys202.

Krause, Christin, Helen Sievert, Cathleen Geißler, Martina Grohs, Alexander T. El Gammal, Stefan Wolter, Olena Ohlei, et al. 2019. 'Critical Evaluation of the DNA-Methylation Markers ABCG1 and SREBF1 for Type 2 Diabetes Stratification.' *Epigenomics* 11 (8): 885–97. https://doi.org/10.2217/epi-2018-0159.

Kudo, Kenta, Karin Hori, Sefan Asamitsu, Kohei Maeda, Yukari Aida, Mei Hokimoto, Kazuya Matsuo, Yasushi Yabuki, and Norifumi Shioda. 2024. 'Structural Polymorphism of the Nucleic Acids in Pentanucleotide Repeats Associated with the Neurological Disorder CANVAS.' *The Journal of Biological Chemistry* 300 (4): 107138. https://doi.org/10.1016/j.jbc.2024.107138.

Laß, Joshua, Theresa Lüth, Kathleen Schlüter, Susen Schaake, Björn-Hergen Laabs, Christoph Much, Roland Dominic Jamora, et al. 2024. 'Stability of Mosaic Divergent Repeat Interruptions in X-Linked Dystonia-Parkinsonism'. *Movement Disorders: Official Journal of the Movement Disorder Society* 39 (7): 1145–53. https://doi.org/10.1002/mds.29809.

Li, H. 2018. 'Minimap2: Pairwise Alignment for Nucleotide Sequences'. *Bioinformatics* 34 (18): 3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and Subgroup Genome Project Data Processing. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.

Libby, Randell T., Katharine A. Hagerman, Victor V. Pineda, Rachel Lau, Diane H. Cho, Sandy L. Baccam, Michelle M. Axford, et al. 2008. 'CTCF Cis-Regulates Trinucleotide Repeat Instability in an Epigenetic Manner: A Novel Basis for Mutational Hot Spot Determination.' *PLoS Genetics* 4 (11): e1000257. https://doi.org/10.1371/journal.pgen.1000257.

Libby, Randell T., Darren G. Monckton, Ying-Hui Fu, Refugio A. Martinez, John P. McAbney, R. Lau, David D. Einum, et al. 2003. 'Genomic Context Drives SCA7 CAG Repeat Instability, While Expressed SCA7 cDNAs Are Intergenerationally and Somatically Stable in Transgenic Mice.' *Human Molecular Genetics* 12 (1): 41–50. https://doi.org/10.1093/hmg/ddg006.

Luth, T., J. Labeta, S. Schaake, I. Wohlers, J. Pozojevic, R. D. G. Jamora, R. L. Rosales, et al. 2022. 'Elucidating Hexanucleotide Repeat Number and Methylation within the X-Linked Dystonia-Parkinsonism (XDP)-Related SVA Retrotransposon in TAF1 with Nanopore Sequencing'. *Genes (Basel)* 13 (1). https://doi.org/10.3390/genes13010126.

Milovanovic, A., N. Dragasevic-Miskovic, M. Thomsen, M. Borsche, F. Hinrichs, A. Westenberger, C. Klein, et al. 2024. 'RFC1 and FGF14 Repeat Expansions in Serbian Patients with Cerebellar Ataxia'. *Mov Disord Clin Pract* 11 (6): 626–33. https://doi.org/10.1002/mdc3.14020.

Mirceta, Mila, Monika H. M. Schmidt, Natalie Shum, Tanya K. Prasolava, Bryanna Meikle, Stella Lanni, Mohiuddin Mohiuddin, et al. 2024. 'C9orf72 Repeat Expansion Creates the Unstable Folate-Sensitive Fragile Site FRA9A.' *NAR Molecular Medicine* 1 (4): ugae019. https://doi.org/10.1093/narmme/ugae019.

Miyatake, Satoko, Hiroshi Doi, Hiroaki Yaguchi, Eriko Koshimizu, Naoki Kihara, Tomoyasu Matsubara, Yasuko Mori, et al. 2024. 'Complete Nanopore Repeat Sequencing of SCA27B (GAA-FGF14 Ataxia) in Japanese.' *Journal of Neurology, Neurosurgery, and Psychiatry* 95 (12): 1187–95. https://doi.org/10.1136/jnnp-2024-333541.

Mohren, L., F. Erdlenbruch, E. Leitao, F. Kilpert, G. S. Hones, S. Kaya, C. Schroder, et al. 2024. 'Identification and Characterisation of Pathogenic and Non-Pathogenic FGF14 Repeat Expansions'. *Nat Commun* 15 (1): 7665. https://doi.org/10.1038/s41467-024-52148-1.

Mulvihill, David J., Kerrie Nichol Edamura, Katharine A. Hagerman, Christopher E. Pearson, and Yuh-Hwa Wang. 2005. 'Effect of CAT or AGG Interruptions and CpG Methylation on Nucleosome Assembly upon Trinucleotide Repeats on Spinocerebellar Ataxia, Type 1 and Fragile X Syndrome.' *The Journal of Biological Chemistry* 280 (6): 4498–4503. https://doi.org/10.1074/jbc.M413239200.

Ohshima, K., L. Montermini, R. D. Wells, and M. Pandolfo. 1998. 'Inhibitory Effects of Expanded GAA.TTC Triplet Repeats from Intron I of the Friedreich Ataxia Gene on Transcription and Replication in Vivo'. *J Biol Chem* 273 (23): 14588–95. https://doi.org/10.1074/jbc.273.23.14588.

Ouyang, R., L. Wan, D. Pellerin, Z. Long, J. Hu, Q. Jiang, C. Wang, et al. 2024. 'The Genetic Landscape and Phenotypic Spectrum of GAA-FGF14 Ataxia in China: A Large Cohort Study'. *EBioMedicine* 102 (April):105077. https://doi.org/10.1016/j.ebiom.2024.105077.

Parsons, M. A., R. R. Sinden, and M. G. Izban. 1998. 'Transcriptional Properties of RNA Polymerase II within Triplet Repeat-Containing DNA from the Human Myotonic Dystrophy and Fragile X Loci'. *J Biol Chem* 273 (41): 26998–8. https://doi.org/10.1074/jbc.273.41.26998.

Paulson, H. 2018. 'Repeat Expansion Diseases'. *Handb Clin Neurol* 147:105–23. https://doi.org/10.1016/B978-0-444-63233-3.00009-9.

Pellerin, D., M. C. Danzi, C. Wilke, M. Renaud, S. Fazal, M. J. Dicaire, C. K. Scriba, et al. 2023. 'Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia'. *N Engl J Med* 388 (2): 128–41. https://doi.org/10.1056/NEJMoa2207406.

Pellerin, D., M. Danzi, M. Renaud, H. Houlden, M. Synofzik, S. Zuchner, and B. Brais. 1993. 'GAA-FGF14-Related Ataxia'. In *GeneReviews((R))*, edited by M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, and A. Amemiya. Seattle (WA).

Pellerin, D., G. F. Del Gobbo, M. Couse, E. Dolzhenko, S. K. Nageshwaran, W. A. Cheung, I. R. L. Xu, et al. 2024. 'A Common Flanking Variant Is Associated with Enhanced Stability of the FGF14-SCA27B Repeat Locus'. *Nat Genet* 56 (7): 1366–70. https://doi.org/10.1038/s41588-024-01808-5.

Pellerin, D., P. Iruzubieta, S. Tekgul, M. C. Danzi, C. Ashton, M. J. Dicaire, M. Wandzel, et al. 2023. 'Non-GAA Repeat Expansions in FGF14 Are Likely Not Pathogenic-Reply to: "Shaking Up Ataxia: FGF14 and RFC1 Repeat Expansions in Affected and Unaffected Members of a Chilean Family"'. *Mov Disord* 38 (8): 1575–77. https://doi.org/10.1002/mds.29552.

Rafehi, H., J. Read, D. J. Szmulewicz, K. C. Davies, P. Snell, L. G. Fearnley, L. Scott, et al. 2023. 'An Intronic GAA Repeat Expansion in FGF14 Causes the Autosomal-Dominant Adult-Onset Ataxia SCA50/ATX-FGF14'. *Am J Hum Genet* 110 (1): 105–19. https://doi.org/10.1016/j.ajhg.2022.11.015.

Rajan-Babu, Indhu-Shree, Egor Dolzhenko, Michael A. Eberle, and Jan M. Friedman. 2024. 'Sequence Composition Changes in Short Tandem Repeats: Heterogeneity, Detection, Mechanisms and Clinical Implications.' *Nature Reviews. Genetics* 25 (7): 476–99. https://doi.org/10.1038/s41576-024-00696-z.

Rakovic, A., A. Domingo, K. Grutz, L. Kulikovskaja, P. Capetian, S. A. Cowley, I. Lenz, et al. 2018. 'Genome Editing in Induced Pluripotent Stem Cells Rescues TAF1 Levels in X-Linked Dystonia-Parkinsonism'. *Mov Disord* 33 (7): 1108–18. https://doi.org/10.1002/mds.27441.

Saffie Awad, P., K. Lohmann, Y. Hirmas, F. Hinrichs, M. Thomsen, M. Kauffman, T. Luth, et al. 2023. 'Shaking Up Ataxia: FGF14 and RFC1 Repeat Expansions in Affected and Unaffected Members of a Chilean Family'. *Mov Disord* 38 (6): 1107–9. https://doi.org/10.1002/mds.29390.

Singh, Krishna, Sakshi Shukla, Uma Shankar, Neha Jain, Rishav Nag, Kumari Aditi Pramod, and Amit Kumar. 2024. 'Elucidating the Pathobiology of Cerebellar Ataxia with Neuropathy and Vestibular Areflexia Syndrome (CANVAS) with Its Expanded RNA Structure Formation and Proteinopathy.' *Scientific Reports* 14 (1): 28054. https://doi.org/10.1038/s41598-024-78947-6.

Stein, K. C., F. Morales-Polanco, J. van der Lienden, T. K. Rainbolt, and J. Frydman. 2022. 'Ageing Exacerbates Ribosome Pausing to Disrupt Cotranslational Proteostasis'. *Nature* 601 (7894): 637–42. https://doi.org/10.1038/s41586-021-04295-4.

Stochmanski, S. J., M. Therrien, J. Laganiere, D. Rochefort, S. Laurent, L. Karemera, R. Gaudet, et al. 2012. 'Expanded ATXN3 Frameshifting Events Are Toxic in Drosophila and Mammalian Neuron Models'. *Hum Mol Genet* 21 (10): 2211–18. https://doi.org/10.1093/hmg/dds036.

Toulouse, A., F. Au-Yeung, C. Gaspar, J. Roussel, P. Dion, and G. A. Rouleau. 2005. 'Ribosomal Frameshifting on MJD-1 Transcripts with Long CAG Tracts'. *Hum Mol Genet* 14 (18): 2649–60. https://doi.org/10.1093/hmg/ddi299.

Trinh, J., T. Luth, S. Schaake, B. H. Laabs, K. Schluter, J. Labeta, J. Pozojevic, et al. 2023. 'Mosaic Divergent Repeat Interruptions in XDP Influence Repeat Stability and Disease Onset'. *Brain* 146 (3): 1075–82. https://doi.org/10.1093/brain/awac160.

Volle, Catherine B., and Sarah Delaney. 2013. 'AGG/CCT Interruptions Affect Nucleosome Formation and Positioning of Healthy-Length CGG/CCG Triplet Repeats.' *BMC Biochemistry* 14 (November):33. https://doi.org/10.1186/1471-2091-14-33.

Wang, Yang, Junyan Wang, Zhenzhen Yan, Jianing Hou, Liqi Wan, Yingquan Yang, Yu Liu, Jie Yi, Pei Guo, and Da Han. 2024. 'Structural Investigation of Pathogenic RFC1 AAGGG Pentanucleotide Repeats Reveals a Role of G-Quadruplex in Dysregulated Gene Expression in CANVAS.' *Nucleic Acids Research* 52 (5): 2698–2710. https://doi.org/10.1093/nar/gkae032.

Wilke, C., D. Pellerin, D. Mengel, A. Traschutz, M. C. Danzi, M. J. Dicaire, M. Neumann, et al. 2023. 'GAA-FGF14 Ataxia (SCA27B): Phenotypic Profile, Natural History Progression and 4-Aminopyridine Treatment Response'. *Brain* 146 (10): 4144–57. https://doi.org/10.1093/brain/awad157.

Yousuf, A., N. Ahmed, and A. Qurashi. 2022. 'Non-Canonical DNA/RNA Structures Associated with the Pathogenesis of Fragile X-Associated Tremor/Ataxia Syndrome and Fragile X Syndrome'. *Front Genet* 13:866021. https://doi.org/10.3389/fgene.2022.866021.

Objective 4: To determine whether the *FGF14* repeat expansion is associated with multiple system atrophy

Multiple system atrophy (MSA) is a rare adult-onset neurodegenerative disorder that has a partial phenotypic overlap with SCA27B. The following two studies investigated the frequency of repeat expansions in the *FGF14* gene in MSA patients.

In the first study, "Intronic *FGF14* GAA repeat expansions impact progression and survival in multiple system atrophy", 657 patients with MSA and 1,003 controls were screened for the expansion. The screening was performed with long-range PCR, and long-read sequencing by ONT confirmed individuals with expansions. The groups were compared using Fisher's test for categorical variables and Kruskal–Wallis and ANOVA tests for continuous variables. The Spearman rank correlation was used to test the concordance between long-range PCR and long-read sequencing. I performed the bioinformatic analysis of the long-read sequencing data in this study.

In the second study, "Genetic testing for SCA27B in Korean Multiple System Atrophy", 199 patients with MSA in Korea were screened for repeat expansion in the *FGF14* gene. In contrast to the first study, whole-genome sequencing with short-read sequencing was performed for the screening. The software ExpansionHunter was used to determine the repeat number of the short-read sequencing data. Long-range PCR was performed on all patients and additionally on 196 control participants. The detected repeat number between ExpansionHunter and long-range PCR was not consistent. Therefore, long-read sequencing by ONT was performed on a subset of 11 individuals. I performed the analysis of the long-read sequencing data. I applied linear regression models to see how the different methods correlate with the repeat number detection. The long-read sequencing data also enabled the detection of interruptions. Finally, I co-wrote the first draft of the manuscript and supported the revision stages. Both studies have shown that expanded repeat expansions in *FGF14* had a higher frequency in MSA patients than in unaffected individuals. The highest frequency was detected in MSA patients with more severe symptoms. Therefore, screening for *FGF14* repeat expansion should be considered for multiple system atrophy patients with rapid loss of mobility for better treatment.

Title

Intronic *FGF14* GAA repeat expansions impact progression and survival in multiple system atrophy

Authors

Viorica Chelban[1,2]*, David Pellerin[1,3,4], Nirosen Vijiaratnam[5,6], Hamin Lee[1], Yen Yee Goh[1], Lauren Brown[1], Sara Sambin[7,8], Danielle Seilhean[9], Stephane Lehericy[7,10], Pablo Iruzubieta[1,4,11,12], Rahema Mohammad[1], Eleanor Self[1], Annarita Scardamaglia[1], Cameron Lee[1], Miriama Ostrozovicova[1], Marie-Josée Dicaire[4], Christine Girges[13], Emil K. Gustavsson[14,15], David Murphy[13], Toby Curless[13,16], **Joshua Laß[17],** Joanne Trinh[17], Timothy Rittman[18,19], James B. Rowe[18,19,20], Marios Hadjivassiliou[21], Neil Archibald[22], Matt C. Danzi[3], Catherine Ashton[4,23], Virginie Roth[24], Marion Wandzel[24], Warren A. Cheung[25], Djordje O. Gveric[26], Bart De Vil[27,28,29], Jordan Follett[30], P. Nigel Leigh[31], Lukas Beichert[32,33], Tomi Pastinen[34], Céline Bonnet[24,35], Mathilde Renaud[34,35,36], Wassilios G. Meissner[37,38,39,40], Anne Sieben[27,28,41], David Crosiers[28,29], Patrick Cras[28,29], Stephan Zuchner[3], Jean-Christophe Corvol[7,8], Matthew J. Farrer[30], Matthis Synofzik[32,33], Bernard Brais[4,42], Tom Warner[13,16], Huw R. Morris[13], Zane Jaunmuktane[13,16,43], Tom Foltynie[13] and Henry Houlden[1]

[1] Neuromuscular Disease Department, UCL Queen Square Institute of Neurology, University College London, London, WC1N 3BG, UK

[2] Neurobiology and Medical Genetics Laboratory, "Nicolae Testemitanu" State University of Medicine and Pharmacy, Chisinau, MD-2004, Republic of Moldova

[3] Dr. John T. Macdonald Foundation Department of Human Genetics and John P. Hussman Institute for Human Genomics, University of Miami Miller School of Medicine, Miami, FL33136 Florida, USA

[4] Department of Neurology and Neurosurgery, Montreal Neurological Hospital and Institute, McGill University, Montreal, QC H3A 2B4, Canada

[5] Clinical Research Centre for Movement Disorders and Gait, Kingston Centre, Parkinson's Foundation Centre of Excellence, Monash Health, Cheltenham, VIC 3192, Australia

[6] School of Clinical Sciences, Department of Medicine, Monash University, Clayton, VIC 3168, Australia

[7] Sorbonne Université, Institut du Cerveau – Paris Brain Institute – ICM, Inserm, CNRS, 75651 Paris Cedex 13, Paris, France

[8] Assistance Publique Hôpitaux de Paris, Department of Neurology, CIC Neurosciences, Hôpital Pitié-Salpêtrière, 75651 Paris Cedex 13, Paris, France

[9] Département de Neuropathologie, Institut de Neurologie, DMU Neurosciences, Groupe Hospitalier Pitié-Salpêtrière, 47-83 bd de l'Hôpital, 75651 Paris Cedex 13, France
[10] Assistance Publique Hôpitaux de Paris, Department of Neuroradiology, Hôpital Pitié- Salpêtrière, 75013 Paris, France
[11] Department of Neurology, Donostia University Hospital, Biogipuzkoa Health Research Institute, Donostia-San Sebastián, 20014, Spain
[12] CIBERNED Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas-Instituto de Salud Carlos III (CIBER-CIBERNED-ISCIII), Madrid, 28029, Spain.
[13] Department Clinical and Movement Neuroscience, UCL Queen Square Institute of Neurology, University College London, London, WC1N 3BG, UK
[14] Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London; London, WC1N 3BG, UK,
[15] UK Dementia Research Institute at The University of Cambridge, Cambridge, CB2 0AH, UK
[16] Queen Square Brain Bank for Neurological Disorders, UCL Queen Square Institute of Neurology, London, WC1N 1PJ, UK
[17] Institute of Neurogenetics, University of Lübeck, 23538, Lübeck, Germany
[18] Department of Clinical Neurosciences, Cambridge University and Cambridge University Hospitals NHS Trust and MRC Cognition and Brain Sciences Unit, CB2 0SZ, Cambridge, UK
[19] Cambridge University Hospitals NHS Trust, Hills road, CB2 0QQ, Cambridge, UK
[20] MRC Cognition and Brain Sciences Unit, Chaucher road, CB2 7EF, Cambridge, UK
[21] Academic Department of Neurosciences, Sheffield Teaching Hospitals NHS Trust, Royal Hallamshire Hospital, Sheffield, S10 2JF, UK
[22] Department of Neurology, South Tees NHS Foundation Trust, Middlesbrough, TS4 3BW, UK
[23] Department of Neurology, Royal Perth Hospital, Perth, WA 6000, Australia
[24] Laboratoire de Génétique, CHRU de Nancy, 54511, Nancy France
[25] Genomic Medicine Center, Children's Mercy Kansas City, Kansas City, MO 64108, USA
[26] Department of Brain Sciences, Imperial College London, London, W12 0NN, UK
[27] Neuropathology lab, IBB-NeuroBiobank BB190113, Born Bunge Institute, Antwerp, 2610, Belgium
[28] Translational Neurosciences, Born-Bunge Institute, Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, 2610, Belgium
[29] Department of Neurology, Antwerp University Hospital, Edegem, 2650, Belgium
[30] McKnight Brain Institute, Department of Neurology, University of Florida, FL 32610, Florida, USA
[31] Department of Neuroscience, Brighton and Sussex Medical School, Brighton, BN1 9PX, UK
[32] German Center for Neurodegenerative Diseases (DZNE), Tübingen, 72076, Germany
[33] Division of Translational Genomics of Neurodegenerative Diseases, Hertie-Institute for Clinical Brain Research and Center of Neurology, University of Tübingen, Tübingen, 72076, Germany
[34] Service de Génétique Clinique, CHRU de Nancy, Nancy, 54511, France

[35] INSERM-U1256 NGERE, Université de Lorraine, 54500, Nancy, France
[36] Service de Neurologie, CHRU de Nancy, Nancy, 54511, France
[37] Service de Neurologie - Maladies Neurodégénératives, IMNc, CRMR AMS, CHU de Bordeaux, Bordeaux, 33076, France
[38] Univ. Bordeaux, CNRS, IMN, UMR 5293, F-33000 Bordeaux, France
[39] Department of Medicine, University of Otago, Christchurch, 9016, New Zealand
[40] New Zealand Brain Research Institute, Christchurch, 8011, New Zealand
[41] Department of Pathology, Antwerp University Hospital - UZA, Antwerp, 2650, Belgium
[42] Department of Human Genetics, McGill University, Montreal, QC H3A 0G4, Canada
[43] Division of Neuropathology, National Hospital for Neurology and Neurosurgery, University College London NHS Foundation Trust, London, UK, WC1N 3BG
* Correspondence author

Published in *Brain,* 2025, awaf134. DOI: 10.1093/brain/awaf134, Online ahead of print

Abstract

Background

Partial phenotypic overlap has been suggested between multiple system atrophy (MSA) and spinocerebellar ataxia 27B, the autosomal dominant ataxia caused by an intronic GAA•TTC repeat expansion in *FGF14*. This study investigated the frequency of *FGF14* GAA•TTC repeat expansion in clinically diagnosed and pathologically confirmed multiple system atrophy cases.

Methods

We screened 657 multiple system atrophy cases (193 clinically diagnosed and 464 pathologically confirmed) and 1,003 controls. The *FGF14* repeat locus was genotyped using long-range PCR and bidirectional repeat-primed PCRs, and expansions were confirmed with targeted long-read Oxford Nanopore Technologies sequencing.

Results

We identified 19 multiple system atrophy cases carrying an *FGF14* $(GAA)_{\geq 250}$ expansion (2.89%, n=19/657), a significantly higher frequency than in controls (1.40%, n=12/1,003) (p=0.04). Long-read Oxford Nanopore Technologies sequencing confirmed repeat sizes and polymorphisms detected by PCR, with high concordance (Pearson's r=0.99, p<0.0001). Seven multiple system atrophy patients had a pathogenic *FGF14* $(GAA)_{\geq 300}$ expansion (five pathologically confirmed and two clinically diagnosed), and 12 had an intermediate $(GAA)_{250\text{-}299}$ expansion (six pathologically confirmed and six clinically diagnosed). A similar proportion of cerebellar-predominant and parkinsonism-predominant multiple system atrophy cases had *FGF14* expansions. Multiple system atrophy patients carrying an *FGF14* $(GAA)_{\geq 250}$ expansion exhibited severe gait ataxia, autonomic dysfunction, and Parkinsonism in keeping with an MSA phenotype, with a faster progression to falls (p=0.03) and regular wheelchair use (p=0.02) compared to the multiple system atrophy cases without *FGF14* GAA expansion. The length of the GAA•TTC repeat expansion inversely correlated with survival in multiple system atrophy patients (r = −0.67; p=0.02), but not with age of onset.

Conclusions

Therefore, screening for *FGF14* GAA•TTC repeat expansion should be considered for multiple system atrophy patients with rapid loss of mobility and for complete diagnostic accuracy at inclusion in disease-modifying multiple system atrophy drug trials.

Introduction

Multiple system atrophy (MSA) is a rare adult-onset neurodegenerative disorder whose aetiology remains unknown. A variable combination of progressive Parkinsonism, cerebellar ataxia, and dysautonomia characterizes it. However, a predominance of either Parkinsonism (MSA-P) or cerebellar impairment (MSA-C) usually occurs (Goh et al. 2023). The pathological hallmark of MSA is the accumulation of alpha-synuclein, which aggregates in mature oligodendrocytes to form glial cytoplasmic inclusions, defining it as a synucleinopathy alongside Parkinson's disease (PD) and dementia with Lewy bodies (Papp et al. 1989; Spillantini et al. 1998). Diagnosis of MSA can be challenging, with about 80% of clinically diagnosed cases meeting pathological criteria at autopsy (Gilman et al. 2008; Koga et al. 2015; Miki et al. 2021). Clinical heterogeneity at presentation and during progression partly explains the diagnostic challenges, though contributing factors, including genetic causes and modifiers, remain poorly understood. Although family history was an exclusion criterion in previous MSA diagnostic criteria (Gilman et al. 2008), they have been revised to be inclusive (Wenning et al. 2022). Multiplex families have been described with mixed MSA and Lewy body disease presentations linked to rare *COQ2* variants in Japanese families. However, data supporting increased risk for MSA with common *COQ2* variants remains equivocal (Multiple-System Atrophy Research Collaboration 2013; Jeon et al. 2014). Recently, more risk loci in MSA were identified, implicating *GAB1*, *lnc-LRRC49-3*, *TENM2,* and *RABGEF1* in Europeans (Chia et al. 2024) and *PLA2G4C* in a meta-analysis of Asian and Caucasian MSA patients (Nakahara et al. 2023), respectively.

Spinocerebellar ataxia 27B (SCA27B), caused by an intronic GAA•TTC repeat expansion in the Fibroblast Growth Factor 14 gene (*FGF14)* (Pellerin et al. 2023; Rafehi et al. 2023), has emerged as a significant contributor to previously undiagnosed cases of late-onset cerebellar ataxia (D. Pellerin, Danzi, et al. 2023; Wirth et al. 2023; Méreaux et al. 2024). Repeat expansions of at least 250 GAA•TTC repeats ($(GAA)_{\geq 250}$) are considered pathogenic, albeit with incomplete penetrance for expansions of 250 to 299 repeat units (Pellerin et al. 2023; Rafehi et al. 2023; Méreaux et al. 2024). Symptom onset typically occurs in the fifth to seventh decade (Pellerin et al. 2024). While episodic ataxia and cerebellar ocular motor disturbances, such as downbeat nystagmus (DBN), are hallmark features of SCA27B, non-cerebellar manifestations, including Parkinsonism and dysautonomia, have been reported with variable frequency across different cohorts (Pellerin et al. 2023; Wirth et al. 2023; Wilke et al. 2023; Wirth et al. 2024).

The observation of features of neurogenic bladder and pyramidal signs in some patients, in addition to late-onset ataxia, also suggests partial phenotypic overlap between SCA27B and MSA (Wirth et al. 2024). Thus, we investigated the frequency of *FGF14* GAA•TTC repeat expansion in clinically diagnosed MSA patients and pathologically confirmed MSA cases and explored the diagnostic and prognostic implications.

Materials and Methods

Patients with MSA included in this study were recruited with informed consent under ethics-approved research protocols. The study was approved by the institutional review boards of the University College London, London (UCLH: 04/N034), Montreal Neurological Hospital, Montreal (MPE-CUSM-15-915), the Center for Neurology, Tübingen (598/2011BO1), and the Children's Mercy Kansas City (Study #11120514).

Subjects

The screening flowchart is illustrated in Figure 1. Pathologically confirmed MSA patients were recruited as part of an international collaboration of movement disorders centers, brain banks, and the cohort of patients from the Neuroprotection and Natural History in Parkinson Plus Syndromes (NNIPPS) study (Supplementary data 2) and GIE-Neuro CEB (BB-0033-00011) (Bensimon et al. 2009).

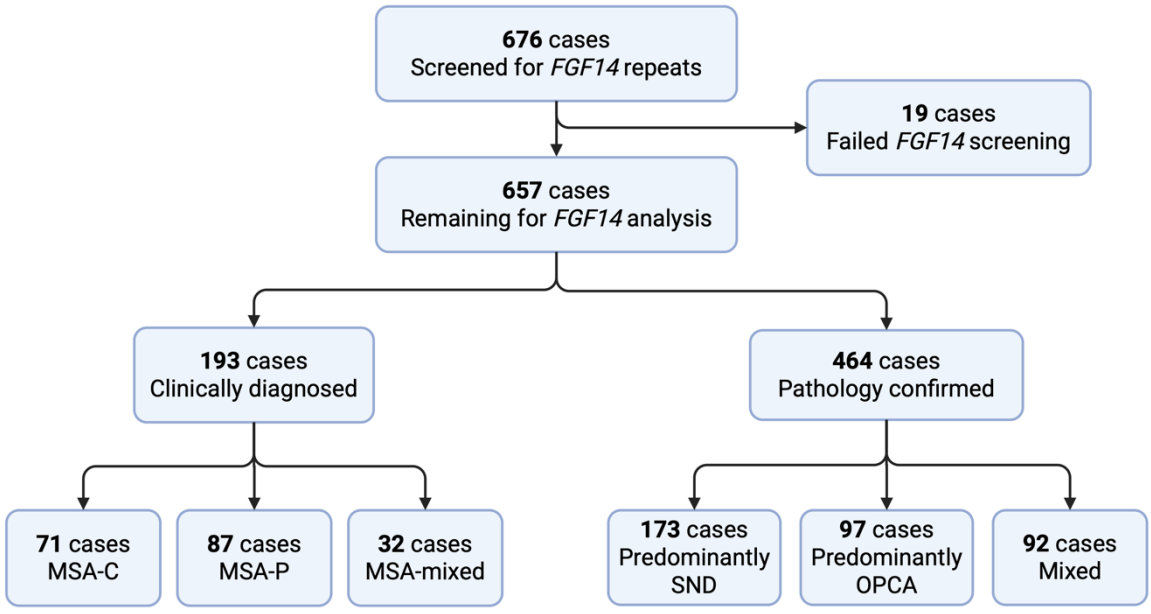

Figure 1. Study flowchart diagram. Legend: MSA-C = MSA cerebellar subtype, MSA-P = MSA parkinsonian subtype, SND = striatonigral degeneration, OPCA = olivopontocerebellar atrophy.

The definite neuropathological diagnosis of MSA was established based on the MSA working group criteria (Trojanowski et al. 2007). Definite MSA patients were further characterized as striatonigral degeneration (SND) or olivopontocerebellar atrophy (OPCA) predominant, based on autopsy findings.

Clinically diagnosed MSA patients were recruited from the MSA specialist clinic at UCL between 2012 and 2021 and fulfilled the diagnosis of probable MSA as per the 2008 MSA diagnostic criteria (Gilman et al. 2008). The 2022 MSA diagnostic criteria were not applied as patients were diagnosed before the publication (Wenning et al. 2022). The 2022 MSA diagnostic criteria were applied retrospectively to all $(GAA)_{\geq 300}$-*FGF14*-positive MSA cases, all meeting the most updated criteria. In addition, a lack of family history of cerebellar ataxia, disease onset after 30 years of age, and negative genetic testing for common repeat expansions causing ataxia (expansions in *ATXN1, ATXN2, ATXN3, ATXN7, CACNA1A, TBP, ATN1, FXN, RFC1*, and *FMR1*) was confirmed in all clinically diagnosed cases. Clinically diagnosed MSA patients were assigned "probable" or "possible" cerebellar (MSA-C), Parkinsonian (MSA-P), or mixed (MSA-mixed) subtypes. There was no biased selection for any MSA clinical subtype at inclusion in this analysis. Participants reported their sex, race, and ethnic group.

Phenotyping was performed through a review of medical records and neuroimaging performed as part of standard clinical care (with variable acquisition protocol), when possible, patient re-evaluation using a standardized data sheet for all MSAs with an *FGF14* expansion of at least 250 repeats.

We analyzed disease milestones from a subset of patients who took part in a longitudinal MSA study as part of the Progressive Supranuclear Palsy-Corticobasal Syndrome-Multiple System Atrophy' study (PROSPECT-M-UK) with the *FGF14* repeat expansion measure. The PROSPECT-M-UK study protocol was previously described elsewhere (Jabbari et al. 2020). Information on symptoms at onset and MSA clinical subtype was available for 138 MSA cases (83 male and 55 female) with non-expanded *FGF14*. Longitudinal data on disease milestones were available for 96 cases.

In addition, we genotyped the *FGF14* repeat locus in 1,003 non-ataxic control individuals. The control cohort included 276 individuals of European origin from the Montreal Neurological Hospital, Montreal, QC, Canada, 202 individuals of European ancestry from the University of Tübingen, Tübingen, Germany, and 525 individuals of overwhelmingly European origin from the Children's Mercy Research Institute's Genomic Answers for Kids program.

DNA extraction

DNA was extracted from blood for clinically diagnosed MSA cases, while DNA was extracted from brain tissue for pathologically confirmed cases. Depending on availability, DNA from 30 mg of frozen cerebellar or frontal cortex was extracted using QIAamp DNA Mini Kit (Qiagen, Venlo, Netherlands) as per protocol and was stored at -80 °C until use. A subset of MSA patients (*n*=21) had matched DNA extracted from both blood and brain tissue.

Genetic screening for FGF14 repeat expansions

The *FGF14* repeat locus was genotyped by long-range polymerase chain reaction (LR-PCR). Repeat sizes were measured by capillary electrophoresis or, when not possible, agarose gel electrophoresis, as described previously (Bonnet et al. 2023). Patients with an allele of at least 200 triplets underwent bidirectional repeat-primed PCRs (RP-PCR) targeting the 5'-end and the 3'-end of the locus to ascertain the presence of a GAA•TTC repeat expansion and interruptions (Bonnet et al. 2023). RP-PCR products were analyzed on an ABI 3730*xl* or ABI 3130*xl* DNA Analyzer using the GeneScan 1200 Liz Dye Size Standard. Results were analyzed using the GeneMapper software. We measured the length of uninterrupted GAA•TTC repeats, excluding polymorphisms and interruptions. Only uninterrupted GAA•TTC-pure expansions of at least 250 repeats were included in the downstream analysis.

Alleles of at least 300 GAA•TTC repeats ($(GAA)_{>300}$) were considered pathogenic, while alleles of 250-299 GAA•TTC repeats ($(GAA)_{250\text{-}299}$) were considered likely pathogenic with incomplete penetrance, based on previously published data (Fig. 2) (Pellerin et al. 2023; Rafehi et al. 2023; Méreaux et al. 2024).

Sanger sequencing of PCR amplification products was performed in cases with an allele of at least 200 triplets at the Centre d'expertise et de services Génome Québec using the ABI3730*xl* DNA Analyzer (Applied Biosystems) and in the Laboratoire de Génétique du Centre Hospitalier Régional Universitaire de Nancy using the ABI3130*xl* DNA Analyzer (Applied Biosystems), as described previously (Bonnet et al. 2023).

The control cohorts from Montreal and Tübingen were genotyped using a standardized PCR-based protocol (Bonnet et al. 2023). In contrast, as reported previously, the Children's Mercy Research Institute control cohort was genotyped by long-read PacBio HiFi sequencing (D. Pellerin et al. 2024).

Long-read sequencing of cases with FGF14 GAA expansion

Cases with *FGF14* expansions of at least 250 repeats were further analyzed with long-read sequencing using Oxford Nanopore Technologies (ONT). Long-range PCR (LR-PCR) amplicons were normalized to 150 ng/µl and then multiplexed using native barcoding expansion PCR-free library preparation kits and the SQK-LSK110 sequencing kit as per manufacturer's instructions (Oxford Nanopore Technologies), multiplexed and sequenced on the PromethION platform using the R10.4.1 flow cell (Oxford Nanopore Technologies). Each run included a negative control. Reads were base-called and demultiplexed using Guppy. Sequences were aligned to the GRCh38 reference human genome.

The "Noise-Cancelling Repeat Finder" (NCRF, version 1.01.02) was used to analyze the *FGF14* trinucleotide repeats from long-read sequencing data (Harris et al. 2019). A threshold of 200 repeats was set to filter for the long allele, and only reads with a maximum noise of 80% were included in the analysis. As previously reported, the repeat length was determined with the median repeat length of all reads (Lüth et al. 2022).

Statistical Analysis

Categorical variables are reported as numbers and percentages. Continuous variables are presented as the mean ± SD (and median/range for the expanded cases as small numbers). Groups were compared using Fisher's test for categorical variables and Kruskal–Wallis, and ANOVA tests for continuous variables, with a two-tailed type I error of 0.05. We used Spearman's rank correlations to assess the relationship between variables. Bonferroni corrections were applied to adjust for multiple comparisons. All statistical analyses were carried out with STATA 17.0 software.

Results

Demographic features of MSA subjects

We screened 657 MSA cases for the *FGF14* GAA•TTC repeat expansion, of which 464 were pathologically confirmed MSA and 193 were clinically diagnosed. The demographics of the study participants are summarized in Supplementary Table 1. In the clinically diagnosed MSA cases, 45.8% (*n*=87/190) had MSA-P, 37.4% (71/190) had MSA-C, and 16.8% (*n*=32/190) had MSA-mixed subtype. In the pathologically confirmed cohort, 45.3% (*n*=150/331) had predominantly SND subtype, 26.9% (89/331) had predominantly OPCA subtype, and 27.8% (*n*=92/331) had a mixed phenotype (similar level of SND and OPCA involvement). Most cases (93.5% overall, of which 89% were clinically diagnosed and 95% pathologically confirmed MSA) were of European origin.

There were no statistically significant differences in sex proportion and the MSA subtype (MSA-P/C or mixed) between the clinically diagnosed and the pathologically confirmed MSA cohorts. Information on age at onset, clinical subtype, and survival was available in 521 of 657 cases, and on pathology subtype (SND/OPCA/mixed) in 331 of 464 pathologically confirmed cases. A significantly younger age at onset (58.0 ± 9.5 vs 59.9 ± 8.6 years, p=0.02) and shorter survival (7.6 ± 3.4 vs 8.2 ± 3.1 years, p=0.02) were noted in the pathologically confirmed MSA cohort compared to the clinically diagnosed MSA group (Supplementary Table 1).

People with cerebellar-predominant phenotype had an earlier age of onset in both the clinical (57.9 ± 8.6 years, adjusted p=0.05) and pathologically confirmed cohort (56.7 ± 8.7 years, adjusted p=0.05) (Supplementary Table 2). The age of onset in the clinical cohort was 60.7 ± 8.7 years for MSA-P and 62.3± 7.8 years for mixed MSA. In the pathologically confirmed cohort, the age of onset in SND cases was 59.3 ± 9.7 years and 56.5 ± 9.7 years in mixed cases. No disease duration differences were noted between clinical subtypes in the clinically diagnosed or pathology-confirmed MSA cohorts.

Distribution of FGF14 GAA•TTC repeat expansion in MSA versus healthy controls

Using a combination of long-range polymerase chain reaction (PCR) and bidirectional repeat-primed PCRs in the whole cohort (n=657 MSA cases, 1,314 chromosomes), we identified 19 MSA cases carrying an *FGF14* $(GAA)_{\geq 250}$ expansion (2.9%, 19/657). Of these, 7 MSA cases (1.1%; 7/657) carried an *FGF14* $(GAA)_{\geq 300}$ expansion (Supplementary Fig. 1), and 12 MSA cases (1.8%; 12/657) had an *FGF14* $(GAA)_{250\text{-}299}$ expansion (Supplementary Fig. 2, Table 1). We found no cases carrying biallelic expansions. We also detected 10 non-GAA-pure expansions. RP-PCR accurately detected polymorphisms and interruptions within these expanded alleles, and the results were consistent with long-read ONT sequencing. The frequency of $(GAA)_{\geq 250}$ expansions in the pathologically confirmed MSA cohort (n=464) was 2.4% (11/464). Specifically, *FGF14* $(GAA)_{\geq 300}$ expansions were identified in five pathologically confirmed MSA cases, corresponding to a frequency of 1.1% (5/464). The frequency of $(GAA)_{250\text{-}299}$ expansions in this group was 1.3% (6/464). In the clinically diagnosed cohort (n=193), the frequency of $(GAA)_{\geq 250}$ expansions was 4.2% (8/193), including 1.0% (2/193) for $(GAA)_{\geq 300}$ expansions and 3.1% (6/193) for $(GAA)_{250\text{-}300}$ expansions. The frequency of $(GAA)_{250\text{-}299}$ and $(GAA)_{\geq 300}$ expansions in controls was 1.2% (12/1,003) and 0.2% (2/1,003), respectively. One of the controls carrying a $(GAA)_{\geq 300}$ expansion was aged 55 at the time of DNA sampling and did not exhibit ataxia, while age at sampling was not available for the second control.

The frequency of $(GAA)_{\geq 250}$ expansions was significantly higher in the combined MSA cohorts (2.9%; 19/657) compared to controls (1.4%; 14/1,003) (odds ratio, 2.10 [95% CI, 0.99 to 4.57]; p=0.04). A significantly higher frequency of $(GAA)_{\geq 250}$ expansions was observed in the clinically diagnosed MSA cohort (4.2%, 8/193 vs. 1.40%, 14/1,003; odds ratio, 3.05 [95% CI, 1.09 to 7.92]; p=0.01) but not in the pathologically confirmed MSA cohort (2.4%, 11/464 vs 1.4%, 14/1,003; odds ratio, 1.71 [95% CI, 0.70 to 4.10]; p=0.20).

Table 1. Allelic frequency of *FGF14* GAA expansions in multiple system atrophy

	All MSA screened (n=657, alleles=1314)		Pathologically confirmed MSA (n=464, alleles=928)		Clinically diagnosed MSA (n=193, alleles=386)	
	$GAA_{\geq 300}$	$GAA_{250\text{-}299}$	$GAA_{\geq 300}$	$GAA_{250\text{-}299}$	$GAA_{\geq 300}$	$GAA_{250\text{-}299}$
MSA cases	7 (1.06%)	12 (1.83%)	5 (1.07%)	6 (1.29%)	2 (1.04%)	6 (3.11%)
Controls (*n*=1,003)	2 (0.19%)	12 (1.20%)	2 (0.19%)	12 (1.20%)	2 (0.19%)	12 (1.20%)
P value	0.033	0.30	0.036	1	0.12	0.056

Legend: n=number, AF=allele frequency. Statistical significance set at $p<0.05$ using Fisher's exact test.

We then compared the frequency of $(GAA)_{\geq 300}$ and $(GAA)_{250\text{-}299}$ expansions in cases and controls. We found that the frequency of $(GAA)_{\geq 300}$ expansions was significantly greater in pathologically confirmed MSA (1.1%, 5/464) compared to controls (0.2%, 2/1,003) (odds ratio, 5.44 [95% CI, 0.89 to 57.32]; p=0.036), while it did not significantly differ in clinically diagnosed cases (1.0%, 2/193 vs. 0.2%, 2/1,003; odds ratio, 5.23 [95% CI, 0.38 to 72.71]; p=0.12). The frequency of $(GAA)_{250\text{-}299}$ expansions was similar between controls and pathologically confirmed MSA cases (1.3%, 6/464 vs. 1.2%, 12/1,003; odds ratio, 1.08 [95% CI, 0.33 to 3.14]; p=1) and clinically diagnosed MSA cases (3.1%, 6/193 vs. 1.2%, 12/1,003; odds ratio, 2.65 [95% CI, 0.80 to 7.73]; p=0.056).

We observed a wide variation of repeat sizes (Fig. 3A, Supplementary Fig. 3). In the pathologically confirmed MSA cases, the median size of $(GAA)_{\geq 300}$ alleles and $(GAA)_{250\text{-}299}$ alleles was 311 repeat units (range, 302 to 346) and 279 repeat units (range, 267 to 293), respectively. In the two clinically diagnosed cases, the sizes of the expanded $(GAA)_{\geq 300}$ alleles were 337 and 353 repeat units, and the median size of the intermediate $(GAA)_{250\text{-}299}$ alleles in the clinically diagnosed cases was 260 repeat units (range, 252 to 291). There was no statistically significant difference in repeat sizes in the two MSA diagnostic categories.

In our cohort, 12.9% of cases (85/657) had interruptions in the GAA repeats. Eight MSA cases (1.4%, 8/657), of which five were pathologically confirmed, and two were clinically diagnosed, carried the $(GAAGGA)_n$ expansion (Supplementary Fig. 4). In comparison, two cases had a different $[(GAA)_n(GCA)_m]_z$ expansion.

These non-GAA-pure expansions were previously shown to be non-pathogenic conformation expansions for ataxia in individuals of European origin (D. Pellerin, Danzi, et al. 2023; Hengel et al. 2023; D. Pellerin, Iruzubieta, et al. 2023).

Long-read sequencing repeat expansion sizing concurs with PCR estimates

To further assess and confirm the repeat expansions in *FGF14* GAA•TTC, we performed ONT sequencing in 25 cases. These included the cases with *FGF14* $(GAA)_{\geq 250}$, cases with non-GAA pure expansions, and cases with complex GAA motif conformations detected by LR and RP-PCR. We found minimal difference in the number of GAA•TTC repeats between PCR-based and ONT estimates (Pearson's r=0.99, p<0.0001, Fig. 3B).

Comparison between brain and blood-extracted DNA

We obtained matched blood and brain-derived DNA from 21 pathologically confirmed MSA cases, none carrying an *FGF14* pathogenic repeat expansion. Data on brain tissue type were available in 14 cases; in nine cases, DNA was extracted from cerebellum tissue, and in five cases, from frontal cortex tissue. *FGF14* GAA•TTC allele sizes in matched frontal cortex tissue and blood-derived DNA correlated strongly (Pearson's r=0.9, p<0.0001), irrespective of the repeat size of the largest allele (Fig. 3C). Two cases with GAA•TTC >100 repeats had a more than one repeat difference between blood and cerebellar brain tissue, $(GAA)_{137}$ and $(GAA)_{119}$ compared to $(GAA)_{148}$ and $(GAA)_{124}$ in the cerebellum (Supplementary Fig. 5). Alleles shorter than 100 GAA•TTC repeats were of equal length in both brain and blood tissues.

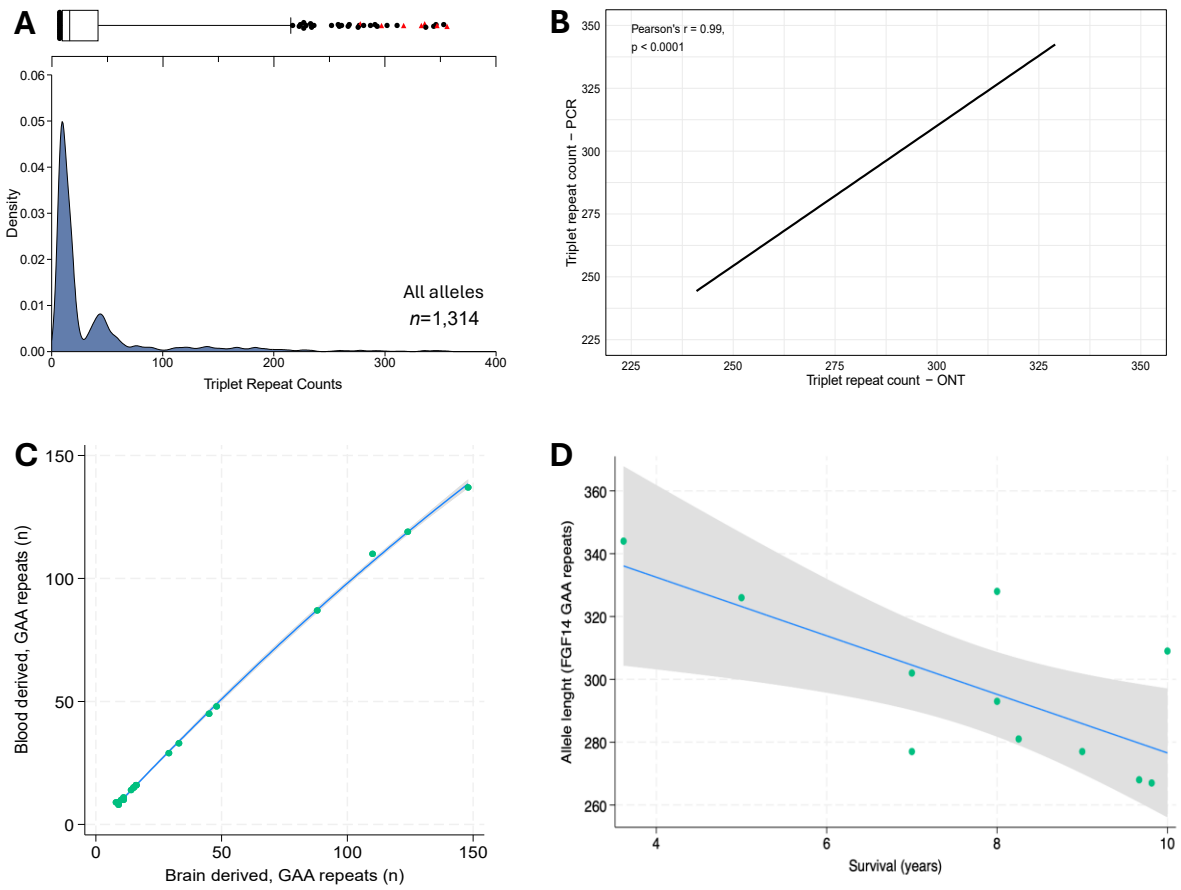

Figure 3. *FGF14* allelic distribution in MSA patients and correlation with survival. (A) *FGF14* allelic distribution in all MSA patients included in this study. Allele distribution of the *FGF14* repeat locus in 657 MSA cases (1314 chromosomes). Among the patients with GAA-*FGF14*–positive MSA, seven were heterozygous for a $(GAA)_{\geq 300}$ expansions, and 12 were heterozygous for a $(GAA)_{250\text{-}299}$ expansion. The density plot shows allele-size frequencies, with higher densities indicating greater frequencies. The box-and-whisker plots show the allelic distribution in patients. The box indicates the 25th percentile (first quartile), the median, and the 75th percentile (third quartile), and the whiskers indicate the 2.5th and 97.5th percentiles. Black dots represent outliers. Red triangles represent expanded alleles of non-GAA-pure repeats, and the red line marks the threshold of $(GAA)_{300}$ repeat units. **(B)** Nanopore repeat size estimates concur with PCR estimates. Comparison of repeat size estimates by long-range-PCR (LR-PCR) and Oxford Nanopore Technologies (ONT) adaptive sequencing targeting 14 individuals carrying a repeat expansion. **(C)** Correlation of repeat size in DNA from brain and blood matched samples. Correlation between the size of the *FGF14* GAA•TTC repeat measured in matched samples from individuals with blood-extracted and brain-extracted DNA (Pearson's r=0.9, p<0.0001). **(D)** Correlation between allele size and survival. Statistically significant negative correlation between the size of the *FGF14* GAA•TTC repeat expansion and survival (calculated from disease onset until death) in patients with MSA (Pearson's $r = -0.67$; $p=0.02$).

In-depth genotype-phenotype correlation of FGF14 (GAA)$_{\geq 300}$ patients with MSA

Seven patients diagnosed with MSA, of whom five (71.4%, 5/7) had pathological confirmation of diagnosis, carried a (GAA)$_{\geq 300}$ expansion. In four of the five definite MSA cases, the neuropathological subtype was striatonigral degeneration (SND), while the other patient had mixed pathology. The two clinically diagnosed MSA patients with an *FGF14* (GAA)$_{\geq 300}$ expansion had mixed MSA-P and C, and MSA-C predominant clinical phenotypes, respectively. The clinical characteristics of these patients are summarized in Supplementary Table 3 and Supplementary Data 1. The median age of onset in the *FGF14* (GAA)$_{\geq 300}$ MSA cases was 58.6 (range, 49-69) years. We found no statistically significant correlation between onset age and the repeat expansion size (7 patients; Spearman's r=−0.3; p=0.3). The first symptom experienced by patients at clinical onset was gait ataxia (40.0%, 2/5), Parkinsonism (40.0%, 2/5), and autonomic failure (20.0%, 1/5), with data missing for two cases (Table 2). None of the patients experienced episodic symptoms at disease onset (including episodic worsening of gait impairments or dizziness). The median disease duration from onset until death was 7.5 (range, 3.6-12) years in *FGF14* (GAA)$_{\geq 300}$ cases, 8.3 (range, 4-11) years in *FGF14* (GAA)$_{250\text{-}299}$ cases, and 7.1 (range 3-12) years in *FGF14* GAA-negative MSA cases with no statistical difference between the groups. At the time of MSA diagnosis, the cardinal clinical features in the seven *FGF14* (GAA)$_{\geq 300}$ cases consisted of a combination of autonomic failure plus Parkinsonism and autonomic failure plus cerebellar ataxia (data missing in one case). However, within the first year of symptom onset, five out of six patients had progressed to exhibit the complete triad of autonomic failure with Parkinsonism and cerebellar ataxia.

Detailed clinical description from onset until death or last seen alive was available in 6 of the 7 cases. Cerebellar features predominantly included gait ataxia (100%, 6/6), limb ataxia (66.7%, 4/6), dysarthria (83.3%, 5/6) with onset in the first two years of disease, dysphagia (66.7%, 4/6), and cerebellar ocular motor signs (50.0%, 3/6). No information was available specifically on down-beat nystagmus, diplopia, oscillopsia, or vertigo in any of these cases. Non-cerebellar characteristics were consistent with MSA phenotype and included autonomic failure (100% of cases) consisting of neurogenic bladder dysfunction (100%, 6/6), orthostatic hypotension (66.7%, 4/6), erectile dysfunction (100% of male cases, 3/3), gastro-intestinal symptoms (100%, 5/5) and sialorrhea or sweating abnormality (60.0%, 3/5). Extrapyramidal features included bradykinesia (83.3%, 5/6) and mild rest tremor (66.7%, 4/6). Levodopa treatment was initiated in 83.3% (5/6) of patients, and minimal or no response was reported in all cases. Additional features included postural instability with retropulsion, polyminimyoclonus, cervical dystonia, and depression.

Despite the patients' advanced age, cognitive impairment (based on clinical evaluation) was uncommon (20.0%, 1/5 patients). The sensation was normal in all patients. Disease progression was rapid in most cases. Walking aids were used in the first 5 years from onset (40.0%, 2/5), progressing to using a wheelchair shortly after that in three of the five patients for whom data were available.

Median age at death was 63 years (range, 55–77), with no statistically significant correlation between age of death and the repeat expansion length (6 patients; Spearman's r=−0.48; p=0.2).

Table 2. Clinical characteristics of MSA patients according to their *FGF14* GAA repeat expansion status

Clinical Characteristics	*FGF14* $GAA_{<250}$	*FGF14* $GAA_{\geq 250}$	*P* value (adj. p value)	*FGF14* $GAA_{\geq 300}$	*FGF14* $GAA_{250-299}$	*P* value (adj. p value)
Diagnostic certainty for MSA						
Clinically Diagnosed, % (n)	29 (185/638)	57.9 (11/19)	0.21(1)	28.6 (2/7)	50 (6/12)	0.63 (1)
Pathologically confirmed, % (n)	71 (453/638)	42.1 (8/19)	0.21(1)	71.4 (5/7)	50 (6/12)	0.63 (1)
MSA clinical subtype						
MSA-C, % (n)	32.1 (170/530)	27.8 (5/18)	0.70 (1)	14.3 (1/7)	36.4 (4/11)	0.59 (1)
MSA-P, % (n)	49.2 (261/530)	50 (9/18)	0.95 (1)	85.7 (6/7)	27.2 (3/11)	0.05 (0.75)
MSA-Mixed, % (n)	18.7 (99/530)	22.2 (4/18)	0.75 (1)	0 (0/7)	36.4 (4/11)	0.11 (1)
MSA pathology subtype						
MSA-predominantly SND, % (n)	47.6 (168/353)	60 (6/10)	0.52 (1)	80 (4/5)	40 (2/5)	0.52 (1)
MSA-predominantly OPCA, % (n)	27.5 (97/353)	0 (0/10)	0.06 (0.96)	0 (0/5)	0 (0/5)	1(1)
MSA-mixed, % (n)	24.9 (88/353)	40 (4/10)	0.28 (1)	20 (1/5)	60 (3/5)	0.52 (1)
First symptom at onset						
Gait ataxia, % (n)	31.9 (43/135)	35.3 (6/17)	0.78 (1)	40 (2/5)	33.3 (4/12)	1(1)
Parkinsonism, % (n)	34.8 (47/135)	41.2 (7/17)	0.61(1)	40 (2/5)	41.7 (5/12)	1(1)
Erectile disfunction (in men), % (n)	11.1 (9/81)	25 (2/8)	0.26 (1)	33.3 (1/3)	20 (1/5)	1(1)
Neurogenic bladder, % (n)	11.1 (15/135)	23.5 (4/17)	0.23 (1)	0 (0/5)	33.3 (4/12)	0.26 (1)
Orthostatic hypotension, % (n)	13.3 (18/135)	0 (0/17)	0.22 (1)	0 (0/5)	0 (0/12)	NA

Legend: adj. p value = adjusted p value, n = number of cases with available information, N/A = not available, MSA-C = MSA cerebellar subtype, MSA-P = MSA parkinsonian subtype, SND = striatonigral degeneration, OPCA = olivopontocerebellar atrophy, IQR = interquartile range. Statistical significance set at p<0.05.

In-depth genotype-phenotype correlation of FGF14 (GAA)$_{250-299}$ patients with MSA

Twelve patients diagnosed with MSA carried an *FGF14* (GAA)$_{250–299}$ expansion, half of whom (6/12) had a pathologically confirmed diagnosis (Supplementary Table 3, Supplementary Data 1). Most of these cases were of European descent (10/12, 83.33%), and none of them had episodic symptoms at disease onset. The *FGF14* (GAA)$_{250-299}$ MSA cases were phenotypically similar to the *FGF14* (GAA)$_{\geq 300}$ cases (Table 3, Supplementary Table 3). Half of the patients had a predominant MSA-C phenotype (6/12, 50%), four patients had an MSA-P predominant phenotype (33.3%,4/12), and two patients had a mixed MSA phenotype (16.7%, 2/12). Five clinically diagnosed MSA cases (n=6) had an MSA-C predominant phenotype. The age of onset in the *FGF14* (GAA)$_{250–299}$ MSA cases was, on average, 55.7 (±7.7) years. There was no correlation between age at onset and the size of the repeat expansion in patients with (GAA)$_{250-299}$ (n=12, Pearson's r=−0.3; p=0.3).

Table 3. Clinical features, progression, and *FGF14* GAA repeat expansion status in MSA

Clinical Characteristics	***FGF14* $GAA_{<250}$**	***FGF14* $GAA_{\geq 250}$**	***P* value (adj. p value)**	***FGF14* $GAA_{\geq 300}$**	***FGF14* $GAA_{250-299}$**	***P* value (adj. p value)**
Age of onset, years	58.7 (37.4–80, ±9.3)	56.8 (42–73, ±8.1)	0.38 (1)	58.6 (49–70, ±9.0)	55.7 (42–73, ±7.7)	0.47 (1)
Age of death, years	66.1 [60–72]	63 [59–71.4]	0.46 (1)	65.6 [8–75]	63.0 [60–69.3]	0.98 (1)
Disease duration, years	7.1 [5.1–9.5]	8 [7–9.7]	0.47 (1)	7.5 [5–10]	8.3 [7–9.7]	0.72 (1)
Cardinal clinical features when MSA diagnosis was first suspected, % (n)						
Parkinsonism	27.1 (26/96)	11.7 (2/17)	0.23 (1)	0 (0/6)	9.1 (1/11)	1.00 (1)
Parkinsonism and cerebellar ataxia	1.04 (1/96)	11. 8 (2/17)	0.05 (0.7)	0 (0/6)	18 .2 (2/11)	0.52 (1)
Parkinsonism and autonomic failure	15.6 (15/96)	32.5 (4/17)	0.48 (1)	16.7 (1/6)	9.1 (1/11)	1.00 (1)
Cerebellar ataxia and autonomic failure	21.9 (21/96)	0 (0/17)	0.04 (0.06)	0 (0/6)	0 (0/11)	N/A
Parkinsonism, cerebellar ataxia and autonomic failure	9.4 (9/96)	11. 8 (2/17)	0.67 (1)	16.7 (1/6)	9.1 (1/11)	1.00 (1)
Clinical features at the time of last examination						
Cerebellar syndrome, % (n)						
Gait ataxia	98.8 (83/84)	100 (17/17)	1.00 (1)	100 (6/6)	100 (11/11)	N/A
Limb ataxia (upper or lower),	98.8 (83/84)	68.8 (11/16)	0.000 (0.001)	66.7 (4/6)	70 (7/10)	1.00 (1)
Dysarthria	94.1 (80/85)	93.3 (14/15)	1.00 (1)	83.3 (5/6)	100 (9/9)	0.40 (1)
Dysphagia	71.4 (60/84)	81.3 (13/16)	0.55 (1)	66.7 (4/6)	90 (9/10)	0.51 (1)
Cerebellar ocular motor signs	84.7 (72/85)	73.3 (11/15)	0.28 (1)	33.3 (2/6)	100 (9/9)	0.01(0.028)
Postural tremor	85.7 (72/84)	43.8 (7/16)	0.001(0.01)	40 (2/5)	45.5 (5/11)	1.00 (1)

Autonomic features, % (n)						
Autonomic dysfunction (any)	98.8 (84/85)	100 (17/17)	1.00 (1)	100 (6/6)	100 (11/11)	N/A
Neurogenic bladder	95.3 (81/85)	100 (17/17)	1.00 (1)	100 (6/6)	100 (11/11)	N/A
Gastro-intestinal features	83.5 (71/85)	80 (12/15)	0.72 (1)	100 (5/5)	70 (7/10)	0.50 (1)
Orthostatic hypotension	71.8 (61/85)	81.3 (13/16)	0.55 (1)	66.7 (4/6)	90 (9/10)	0.51(1)
Erectile dysfunction in men	95.4 (41/43)	100 (8/8)	1.00 (1)	100 (3/3)	100 (5/5)	N/A
Parkinsonian syndrome, % (n)						
Bradykinesia	95.2 (80/84)	87.5 (14/16)	0.25 (1)	80 (4/5)	90.9 (10/11)	1.00 (1)
Postural instability	88.1 (74/84)	64.3 (9/14)	0.03(0.047)	20 (1/5)	88.9 (8/9)	0.36 (1)
Rest tremor	50 (42/84)	30 (3/10)	0.32 (1)	60 (3/5)	0 (0/5)	0.16 (1)
Levodopa trial	75.5 (114/151)	77.8 (14/18)	1.00 (1)	83.3 (5/6)	75 (9/12)	1.00 (1)
Positive levodopa response	7.1 (8/113)	7.7 (1/13)	1.00 (1)	0 (0/6)	11.1 (1/9)	1.00 (1)
Minimal/partial levodopa response	50.4 (57/113)	30.8 (4/13)	0.12 (1)	33.3 (2/6)	22.2 (2/9)	1.00 (1)
Negative levodopa response	42.5 (48/113)	61.5 (8/13)	0.19 (1)	33.3 (2/6)	77.7 (7/9)	0.31(1)
Progression						
Falls in the first year from onset, % (n)	35.7 (20/56)	66.7 (10/15)	0.03 (0.047)	50 (2/4)	72.7 (8/11)	0.24 (1)
Disease duration to first falls, years	2.71 [0.52–3.88], n=27	0.5 [0.5–3], n=14	0.01(0.028)	1.0 [0–1], (n=3)	0.5 [0.5–3], (n=11)	1.00 (1)
Use of walking aid in the first 5 years from onset, %, (n)	77.4 (41/53)	76.9 (10/13)	1.00 (1)	75 (3/4)	77.7 (7/9)	0.47 (1)
Disease duration to regular use of walking aid, years	3.4 [2.1–4.8], n=34	3.45 [1.0–6.0], n=11	0.2 (1)	5 [2–7], n=3	3.5 [0.75–5.5], n=8	0.54 (1)
Regular use of wheelchair in the first 5 years from onset, % (n)	60.5 (23/38)	68.75 (11/16)	0.336 (1)	40 (2/5)	100 (8/8)	0.04 (0.06)
Disease duration to regular use of wheelchair, years	7 [3.4-7], n=15	5 [4.5-6], n=8	0.02 (0.039)	5.5 [4-7], n=2	5 [5-5], n=6	0.787 (1)

Legend: adj. p value = adjusted p value, *n* = number of cases with available information, N/A = not available, IQR = interquartile range. Values are presented as mean (range ± SD), median [IQR], or % (n).

The first symptom experienced by patients at clinical onset was gait ataxia (33.3%, 4/12), autonomic failure (25%, 3/12), Parkinsonism with autonomic failure (25%, 3/12), and Parkinsonism without autonomic failure (16.7%, 2/12). Dysphagia and dysarthria were frequent features in this subgroup, with 83.3% (5/6) of patients experiencing them in the first five years from onset. Over half of the patients had upper limb ataxia and dysdiadochokinesia (58.3%, 7/12). Autonomic dysfunction was present in all patients, either as the first symptom (66.6%, 8/12) or soon after the disease onset. It comprised orthostatic hypotension (90%, 9/10), neurogenic bladder (100%,11/11), and erectile dysfunction (100%, 5/5 male patients). Other symptoms experienced and signs noted in this group included rapid eye movement sleep behavior disorder (77.7%, 7/9), stridor (44.4%, 5/9), broken pursuit (66.6%, 6/9), and horizontal gaze-evoked nystagmus (22.2%, 2/9). Ten patients were treated with levodopa. Apart from three who reported mild clinical benefits, all the others did not experience any benefit from levodopa therapy. Sensation was normal in all patients, and cognitive function remained preserved, even in advanced disease stages. Apart from case C1, all clinically diagnosed patients had died by the time of this analysis, and the median disease survival in this group was 8.3 years (range, 4-11 years). Postural instability with retro-pulsion was more frequent in MSA cases without the *FGF14* expansion (adjusted p=0.047).

Impact of the FGF14 expansion on the MSA phenotype

We compared three groups defined by the size of the longest allele: <250 repeats (n=638), 250–299 repeats (n=12), and ≥300 repeats (n=7) (Table 2). There was no statistically significant difference in sex distribution, age at onset, predominant symptoms at onset, or disease duration between the three groups (Table 2).

A subgroup of the MSA cases was prospectively recruited into a natural history study (n=96). We assessed the clinical progression and disease milestones in MSA patients with *FGF14* $(GAA)_{\geq 250}$ and those with *FGF14* $(GAA)_{<250}$. *Signific*antly, none of the *FGF14* $(GAA)_{\geq 250}$ cases presented with cerebellar ataxia with autonomic failure alone (in the absence of Parkinsonism) compared to the non-expanded group (21.9%) (p=0.04, adjusted p=0.06). Furthermore, a higher proportion of cases with dual pathology compared to *FGF14*-negative MSA cases presented with a mixed phenotype of Parkinsonism and cerebellar ataxia (11.8%) compared to MSA patients without the *FGF14* expansion (1.04%) (p=0.05, adjusted p=0.06). There were no significant differences in the age of onset, disease duration, or symptom of onset between the two groups.

In the 15 *FGF14* $(GAA)_{\geq 250}$ MSA patients with information on the time of onset of falls, ten (66.6%) experienced frequent falls within the first year, a significantly higher proportion than that of the *FGF14* $(GAA)_{<250}$ MSA cases (37.7%, 20/56) (p=0.03, adjusted p=0.047). The median time from onset to falls was significantly shorter in *FGF14* $(GAA)_{\geq 250}$ MSA (0.5 years, n=14) compared to the *FGF14* $(GAA)_{<250}$ MSA cases (2.71 years, n=27) (p=0.01, adjusted p=0.028) (Table 3). Similarly, a shorter time to regular wheelchair use was observed in MSA patients with *FGF14* expansions (median 5 years, range 2-8 years) compared with MSA cases without *FGF14* expansions (median 7 years, range 4-11 years) (p=0.02, adjusted p=0.039). At the time of the last examination, a statistically significantly higher number of patients presented with moderate or severe dysphagia in the MSA cases with *FGF14* $(GAA)_{\geq 250}$ (81.3%) compared to MSA patients without the *FGF14* expansion (74.1%) (p<0.001).

The frequency of the *FGF14* $(GAA)_{\geq 250}$ repeat expansion was 2.2% (8/249) among patients with MSA-P, 2.4% (4/166) among patients with MSA-C, and 4.8% (5/104) among patients with MSA-mixed phenotype. The frequency was 4.7% (6/126) among patients with rapidly progressive disease (patients who died within five years from onset). None of the patients with disease onset before age 45 carried an *FGF14* GAA•TTC repeat expansion. In the entire cohort, an earlier age of onset was associated with a shorter survival (Pearson's r=-0.39, p<0.001). Furthermore, in the pathologically confirmed MSA cases carrying an *FGF14* $(GAA)_{\geq 250}$ repeat expansion, we found a negative correlation between the *FGF14* GAA•TTC repeat expansion size and disease survival (n=11, Pearson's r= −0.67; p=0.02) (Fig. 3D, Supplementary Table 4).

Neuroimaging features

Brain MRI was available for two cases carrying an *FGF14* $(GAA)_{\geq 300}$ expansion (Fig. 4) and nine cases carrying an *FGF14* $(GAA)_{250\text{-}299}$ expansion (Supplementary Table 3). All eleven cases had MRI findings consistent with an MSA diagnosis. Atrophy of cerebellar hemispheres and/or vermis assessed on visual inspection was reported in six (54.5%), brainstem atrophy in five (45.4%), and putamen atrophy in three (27.2%). Hyperintensity of the middle cerebellar peduncle was reported in three (27.2%), and a "hot-cross bun sign" in two patients (18.1%). None of the cases had a disproportionate cerebellar atrophy at the time of the MRI reports. Two patients had a DAT scan, showing bilaterally reduced basal ganglia uptake.

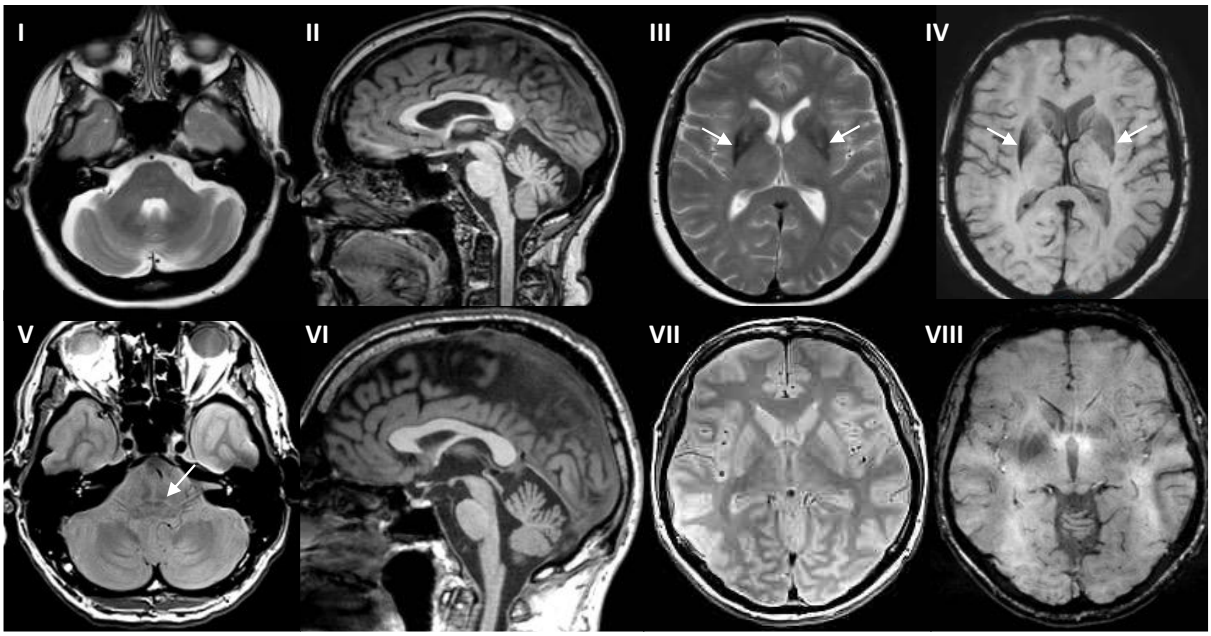

Figure 4. MRI features in *FGF14* (GAA)$_{\geq 300}$ patients with MSA. The top panels show MRI features in case of C1 at age 59, one year from symptoms onset, and the bottom panels show MRI features in P3 at age 55, five years from symptoms onset. Panels I and III: axial T2-weighted images, II and VI: sagittal T1-weighted images, V and VII: axial proton-density images, IV and VIII: axial susceptibility-weighted images. Case C1 had a clinically established MSA-P diagnosis and *FGF14* (GAA)$_{353}$ repeat expansion. Case P3 had a neuropathologically established MSA-mixed diagnosis and *FGF14* (GAA)$_{326}$ repeat expansion. "Hot-cross bun" signs (panels I and V) and cerebellar atrophy (panels II and VI) are present in both cases, more pronounced in case P3. The hypointensity of the putamen in C1 is seen in panels III and IV.

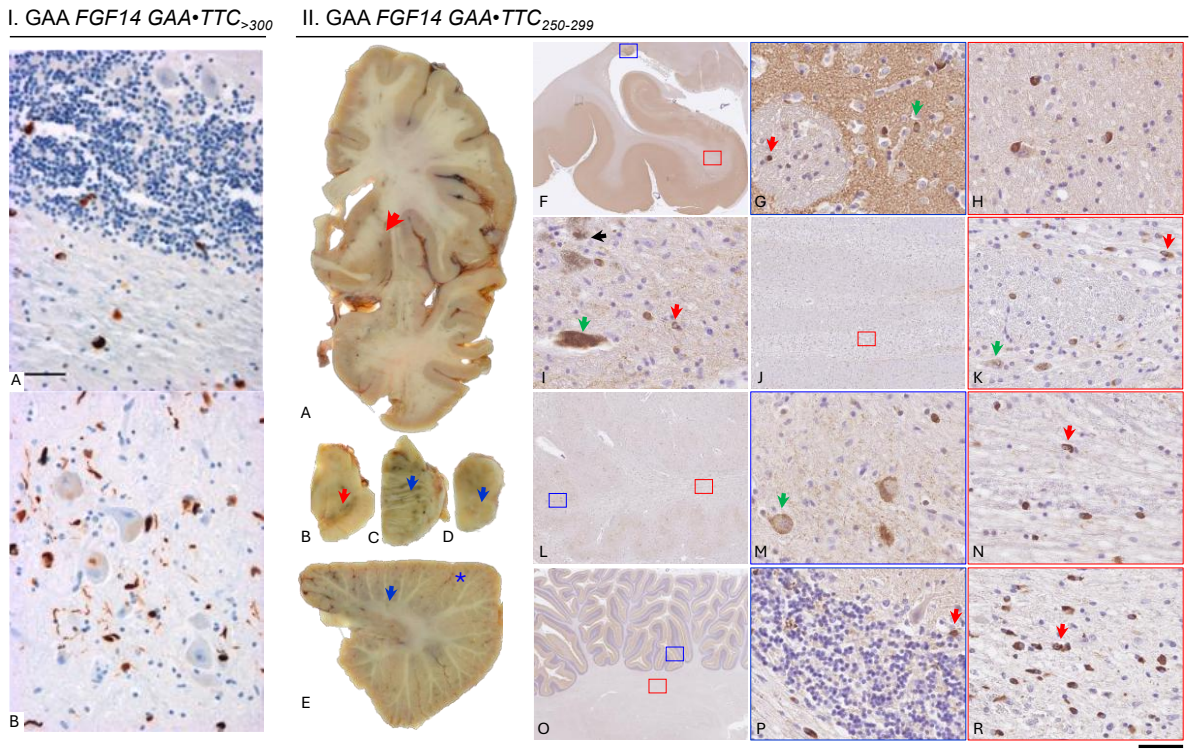

Figure 5. Pathological findings in MSA cases with *FGF14* GAA repeat expansions. (I) Findings in *FGF14* (GAA)≥300 positive patients with MSA. Alpha-synuclein immunohistochemistry findings in *FGF14* (GAA)≥300 positive patients with MSA show numerous intra-oligodendroglial inclusions in the cerebellum (**A**) and medulla oblongata (**B**). **(II) *FGF14* (GAA)250-299 patients with MSA.** Case P7 (281 GAA repeats): (**A**) Severe putaminal atrophy, typical of MSA, is seen on the coronal section (red arrow). (**B**) Substantia nigra in the midbrain shows prominent pallor (red arrow). (**C**) The height of the pontine base is preserved (blue arrow), and (**D**) the inferior olivary nucleus is visible (blue arrow). (**E**) In the cerebellum at the level of the dentate nucleus, there is no evidence of significant white matter atrophy (blue arrow), the cerebellar cortex shows no apparent atrophy (blue asterisk), and the dentate nucleus is unremarkable. The macroscopic appearances are typical of MSA-SND. Case P10 (277 GAA repeats): (**F**) Hippocampus shows no atrophy. (**G**) In the tail of the caudate nucleus (blue square in F), there are glial cytoplasmic inclusions in the grey matter and within striatopallidal fibers (red arrow), and neuronal cytoplasmic inclusions in the grey matter (magenta arrow). (**H**) There are occasional glial cytoplasmic inclusions in the white matter of the parahippocampal gyrus (red square in F) and adjacent gyri. (**I**) In the substantia nigra, there is a prominent depletion of pigmented neurons, with neuromelanin deposition freely in the neuropil (black arrow); occasional residual pigmented neurons contain diffuse cytoplasmic α-synuclein aggregates, and there are occasional glial cytoplasmic inclusions across the midbrain (red arrow). (**J**) The pontine base has a prominent atrophy of the transverse fibers and pontine base nuclei (red square in J). (**K**) There are numerous neuronal cytoplasmic and intranuclear inclusions (magenta arrow) in the pontine nuclei and glial cytoplasmic inclusions (red arrow) in the nuclei and transverse fibers. (**L, M, and N**) In the medulla, there are frequent diffuse neuronal cytoplasmic inclusions in the inferior olivary nucleus (blue square in L and magenta arrow in M) and glial cytoplasmic inclusions in the white matter (red square in L and red arrow in N). (**O, P, and R**) In the cerebellum, there is a moderate depletion of Purkinje cells but good preservation of the granule cells, with occasional glial cytoplasmic inclusions in the cortex (red arrow in P) and numerous glial cytoplasmic inclusions in the cerebellar white matter (red arrow in R). The histological appearances are typical of MSA with equal SND and OPCA involvement. Scale bar: (I) 40 μm in A and B, (II) 3mm in F and O; 50μm in G, H, I, K, M, N, P, and R; 400μm in J; 0.7mm in L**.**

Neuropathological examination of MSA cases with FGF14 (GAA)$_{\geq 250}$

Two cases with *FGF14* (GAA)$_{\geq 300}$ and five cases with *FGF14* (GAA)$_{250\text{-}299}$ had the neuropathology examination repeated after the *FGF14* expansion was identified (Fig. 5). The two cases with *FGF14* (GAA)$_{\geq 300}$ included patient P3 and P4 (Supplementary data 1). The neuropathological features observed in these seven cases fall within the expected spectrum of MSA pathology. Macroscopic examination in patient P4 performed at eight years from disease onset (Fig. 5) showed severe atrophy of the cerebellum, predominating in the superior vermis, dentate nucleus, and middle cerebellar peduncles. The inferior olivary nucleus was macroscopically normal. The pyramidal tracts appeared brownish. The pallidum, thalamus, and sub-thalamic nuclei were normal in contrast to the severe atrophy of the putamen. Microscopically, the putamen, substantia nigra, and pontine nuclei showed a massive neuronal depopulation with astrocytic gliosis. Abundant oligodendroglial cytoplasmic inclusions are positive for alpha-synuclein and were found in the striatum, pons, cerebellum, bulbar olive, transverse, and perpendicular pontine fibers. The cerebellum showed a moderate loss of Purkinje cells with numerous empty baskets and torpedoes.

The glomeruli in the granular layer were rarefied. The cerebellar white matter was atrophic and contained numerous alpha-synuclein-positive oligodendroglial inclusions.

Neuropathological examination at five years from onset in patient P3 showed no macroscopic abnormalities, in particular no cerebellar atrophy, no pallor of the substantia nigra, and no abnormal staining of the putamen. Microscopic examination of the cerebellum revealed no white matter atrophy or loss of Purkinje cells. The only abnormality was a pallor in the hilum of the dentate nucleus. The putamen had no notable neuronal loss, gliosis, or pigmentary deposits, but numerous rounded intra-oligodendroglial inclusions were labeled with anti-alpha synuclein and anti-ubiquitin antibodies. In the cortical areas, particularly in the frontal regions, there were countless alpha-synuclein-positive inclusions in the gray and white matter. In the substantia nigra there was a neuronal loss but no Lewy bodies. Alpha-synuclein immunohistochemistry showed numerous oligodendroglial and neuronal inclusions, which were also associated with dystrophic neurites. Alpha-synuclein-labeled inclusions were abundant in pontine nuclei, medulla oblongata, and cerebellum. None of the areas examined showed polyglutamine inclusion (polyQ, 1C2 antibody), as is common in spinocerebellar ataxias.

In the three cases carrying an *FGF14* $(GAA)_{250\text{-}299}$ expansion and olivopontocerebellar atrophy (P6, P9, P10), macroscopic and microscopic examination showed cerebellar white matter atrophy and moderate depletion of Purkinje cells, with good preservation of the granule cell layer and mild gliosis in the molecular layer. In the two MSA cases with striatonigral degeneration dominant pattern (P7, P8), there was no macroscopic evidence of cerebellar cortical or white matter atrophy in one case. However, microscopically, glial cytoplasmic inclusions were evident in the cerebellar white matter (Fig. 5). The dentate nucleus showed no significant atrophy, and none of the cases with intermediate expansions showed an unusual cerebellar atrophy pattern, exceeding the extent expected to be present in MSA. The pathology observed in these cases was a classic MSA pathology without any pathological evidence to suggest cerebellar cortical atrophy beyond what is typically seen in MSA. Quantitative measurements were not applicable since there was no evident additional cortical cerebellar atrophy.

Discussion

Our study investigated the frequency and potential contribution of *FGF14* GAA•TTC repeat expansions to the MSA phenotype by screening a large cohort of clinically diagnosed and pathologically confirmed MSA patients. We found seven MSA patients carrying an *FGF14* $(GAA)_{\geq 300}$ repeat expansion, five of whom were pathologically confirmed as fulfilling the gold standard neuropathological diagnosis of MSA. All five cases with pathologically confirmed *FGF14* $(GAA)_{\geq 300}$ had typical alpha-synuclein positive GCIs and no atypical features for MSA on autopsy. None of these five patients reported a family history of MSA or other forms of ataxia. The two clinically diagnosed MSA cases with *FGF14* $(GAA)_{\geq 300}$ had typical clinical and radiological features fulfilling both the second consensus statement on the diagnosis criteria and the retrospectively applied Movement Disorder Society (MDS) criteria for an MSA diagnosis and were followed up longitudinally (Gilman et al. 2008; Wenning et al. 2022). These findings underscore the importance of genetic screening in this patient population, as it may have diagnostic and clinical implications. The discovery of *FGF14* GAA•TTC repeat expansions in a subset of pathologically confirmed MSA patients in our study introduces significant challenges for accurate and complete clinical diagnosis of these two conditions (MSA and SCA27B) when they co-exist.

We found *FGF14* $(GAA)_{\geq 300}$ expansions in 1.1% (7/657) of the total MSA cohort and $(GAA)_{250\text{-}299}$ expansions in 1.8% (12/657) of the cohort. Two healthy controls (2/1,003, one aged just 55) also carried *FGF14* $(GAA)_{\geq 300}$ expansions. Interestingly, most of the *FGF14* $(GAA)_{\geq 300}$ expansions uncovered in this study were relatively small (maximum size was 353 GAA repeats) compared to expansions in SCA27B patients, which may reach more extended sizes.

Including our study, a total of 1,843 MSA patients have been tested for the GAA•TTC expansions in *FGF14* (Supplementary Table 5). The published reports predominantly screened blood-extracted DNA from clinically diagnosed MSA-C cases, and almost all used the 2008 second consensus statement on MSA diagnosis criteria (Gilman et al. 2008). Our study is the first to analyze a pathologically confirmed MSA cohort. Unlike most MSA cohorts screened for *FGF14* GAA•TTC expansions that included predominantly MSA-C cases, our cohort screened all MSA phenotypes. We found a similar proportion of *FGF14* expansions in both MSA-C and MSA-P, highlighting the importance of genetic testing in both these presentations when a clinical suspicion arises, irrespective of the predominant MSA clinical subtype.

In our cohort, patients with MSA carrying an *FGF14* $(GAA)_{\geq 250}$ repeat expansion exhibited a combination of autonomic failure, cerebellar ataxia, and extrapyramidal symptoms, aligning with the classical MSA phenotype. However, we show that the *FGF14* $(GAA)_{\geq 250}$ has an impact on the functional status of MSA patients leading to a significantly higher proportion of cases presenting with falls in the first year from onset in the *FGF14* GAA-positive MSA patients compared to *FGF14* GAA-negative MSA cases (p=0.03), a significantly reduced time to falls (p=0.03) and shorter time to wheelchair use (p=0.02). In comparison, many SCA27B patients remain ambulant after 15 years of disease (Wilke et al. 2023). Shared characteristics with SCA27B include adult-onset cerebellar ataxia and, occasionally, a variable combination of mild pyramidal features, neurogenic bladder, autonomic dysfunction, and, sometimes, Parkinsonism. However, unlike SCA27B, most MSA patients with *FGF14* $(GAA)_{\geq 250}$ repeat expansion reported here experienced early and severe gait ataxia and significant autonomic dysfunction, including neurogenic bladder and orthostatic hypotension from the disease onset. Autonomic dysfunction may occur in the later stages of SCA27B (Wilke et al. 2023). Indeed, bladder symptoms, an early and highly disabling MSA feature, are usually reported in less than half the SCA27B patients, described explicitly as urinary urgency or frequency (David Pellerin, Wilke, et al. 2024).

With the introduction of new MSA diagnostic criteria, the bladder features for either clinically probable or established MSA should include unexplained urinary retention and/or urinary incontinence, which should help differentiate the urinary profiles of the two conditions (Wenning et al. 2022). Dysarthria is absent in a substantial proportion of patients with SCA27B, even after a long disease duration (Wilke et al. 2023). In our cohort of *FGF14* GAA-positive MSA cases, dysarthria was present from early stages and was severe, frequently associated with dysphagia. Furthermore, downbeat nystagmus and oscillopsia are frequent in SCA27B (D. Pellerin, Danzi, et al. 2023; Shirai et al. 2023). However, in our study, horizontal nystagmus and interrupted saccades were the only ocular motor signs reported in the *FGF14* GAA-positive MSA cases, most likely related to incomplete data reporting, as these features in most cases were reviewed post-mortem from clinical notes. The subsequent disease course and progression of the *FGF14* GAA-positive MSA cases were consistent with a predominantly fast-progressing MSA and different from that of SCA27B. All cases included in this study fulfilled either the gold standard, neuropathological diagnosis of MSA, or the highest diagnostic certainty of clinical diagnosis of MSA.

Furthermore, we have no evidence that *FGF14* GAA•TTC repeat expansions contribute directly to MSA pathology or act as a causative factor for MSA. Instead, they likely represent a co-occurring, distinct condition that may modulate the clinical course of MSA, potentially contributing to a shorter survival, as we have shown in our cohort. This interpretation is further supported by the neuropathological findings of seven reported SCA27B cases, none showing evidence of alpha-synuclein pathology or features consistent with MSA. The clinical diagnosis for patients with both MSA pathology and an *FGF14* expansion remained compatible with MSA diagnosis, as their presentation aligns more closely with the characteristic features of MSA rather than the typical phenotype of SCA27B. SCA27B is defined by a late-onset, slowly progressive, and relatively pure pancerebellar syndrome, often with episodic symptoms at onset and cerebellar ocular motor signs, including downbeat nystagmus. A few studies have shown that *FGF14* expansions in the incompletely penetrant range (250–299 GAA repeats) may co-occur with other pathogenic variants associated with hereditary ataxias (Wirth et al. 2023; Méreaux et al. 2024). In those cases, the clinical phenotype usually reflected the co-occurring variant with earlier onset and a more severe or complex presentation that deviated from the classic SCA27B phenotype. Similarly, in patients with both MSA pathology and *FGF14* expansions, the phenotype was dominated by MSA, supporting MSA as the primary clinical diagnosis in these cases.

However, future studies comparing MSA patients with other neurodegenerative conditions, such as Parkinson's disease and dementia with Lewy bodies, could help explore whether *FGF14* expansions are uniquely associated with MSA or part of a broader neurodegenerative spectrum.

Recently, cerebellar ataxia with neuropathy and vestibular areflexia syndrome (CANVAS) linked to biallelic AAGGG expansions in *RFC1* has attracted attention as a differential diagnosis of MSA, as well as of SCA27B (David Pellerin, Wilke, et al. 2024; Sullivan et al. 2021). However, no *RFC1* expansions have been identified in pathology-confirmed MSA cases (Sullivan et al. 2020). Despite the phenotypic overlap between MSA, *RFC1*-related disease, and *FGF14* GAA-related disease, certain features may help differentiate these disorders. Chronic cough, a prevalent feature in *RFC1*-related disease, was uncommon in our cohort (Cortese et al. 2020; Gisatulin et al. 2020; Traschütz et al. 2021). While motor neuropathy is typically absent or minimal in *RFC1*-positive patients, it may co-occur with sensory neuropathy in some SCA27B patients, but may be lacking in our MSA cases (David Pellerin, Wilke, et al. 2024; Cortese et al. 2020; Currò et al. 2021). Our patients did not exhibit any vestibular abnormalities, unlike many patients with *RFC1*-related disease.

Importantly, the size of the GAA•TTC repeat expansion was inversely correlated to survival in MSA patients. Future studies should include a detailed assessment of cerebellar involvement severity and related complications such as dysphagia, aspiration pneumonia, and increased frequency and early falls in MSA cases with *FGF14* expansions and their impact on survival.

Our findings suggest that *FGF14* expansions may modulate the disease course in MSA, highlighting the need for genetic testing in these patients. No correlation was found with age of onset. In SCA27B, the size of the GAA•TTC repeat expansion shows a weak negative correlation with age at onset and has not been associated with disease severity or progression (D. Pellerin, Danzi, et al. 2023; David Pellerin, Danzi, et al. 2024; Méreaux et al. 2024; Wilke et al. 2023). Significant phenotypic variability among patients carrying expansions of similar sizes has been reported in SCA27B, suggesting that additional factors influence disease expressivity (Méreaux et al. 2024; David Pellerin, Danzi, et al. 2024; Wilke et al. 2023).

The phenotypic similarities between MSA and SCA27B complicate the clinical diagnosis, especially in the early stages of the disease. This study supports the integration of genetic testing for *FGF14* expansions in the diagnostic workup of MSA, particularly for patients with early-onset cerebellar ataxia, Parkinsonism, and autonomic dysfunction.

This approach could improve diagnostic accuracy and facilitate more personalized management strategies for patients with MSA. Screening for *FGF14* expansions is essential for patient stratification in MSA trials, as these expansions can impact progression and survival in MSA.
Several cohorts of MSA-C or sporadic late-onset cerebellar ataxia (SLOCA), including MSA-C, have been screened for *FGF14* GAA•TTC repeat expansions (Matsushima et al. 2024; Wirth et al. 2024; Ouyang et al. 2024; Satolli et al. 2024). Only one study had previously identified *FGF14* $GAA_{\geq 300}$ repeat expansions in two patients initially diagnosed clinically as possible MSA (Wirth et al. 2024). Based on the genetic findings and re-evaluation of their phenotype, a final diagnosis of SCA27B was reached for these cases. In a Chinese cohort of 527 MSA-C cases, four patients were found to have an intermediate expansion ranging in size from 264 to 275 repeat units (Ouyang et al. 2024). Similar findings were reported in two European cohorts and two Japanese cohorts (Matsushima et al. 2024; Wirth et al. 2024; Satolli et al. 2024; Ando et al. 2024). No detailed clinical description of the cases with intermediate expansion was presented in the European cohort. Studies on East-Asian populations did not find high frequencies of *FGF14* GAA expansions, with only one case having an intermediate expansion identified in two Japanese cohorts (Ando et al. 2024). The patient was classified as SCA27B upon re-evaluation. SCA27B patients progressed more slowly than probable MSA-C patients in all the reported cohorts. These previous studies highlight the importance of *FGF14* screening in rare cases of SCA27B presenting early on as an MSA-mimic.
Unlike previous reports of *FGF14* screening in MSA, most of the cases in our study (70%) had a pathologically confirmed diagnosis of MSA. Neuropathological studies in SCA27B have demonstrated that neuronal loss is predominant in the cerebellar cortex, especially in the vermis, compared to the hemispheres (D. Pellerin, Danzi, et al. 2023; Wilke et al. 2023). The neuropathological findings in SCA27B, described in seven cases to date, are primarily restricted to the cerebellum (D. Pellerin, Danzi, et al. 2023; Wilke et al. 2023; Abou Chaar et al. 2024). All cases displayed cerebellar cortical atrophy, more pronounced in the vermis than the hemispheres, with widespread loss of Purkinje neurons (most severe in the anterior vermis), mild granule-cell loss, and gliosis of the molecular layer. ‘Empty baskets’ were observed in less chronically affected regions. No substantial atrophy was observed in the dentate nuclei or cerebellar peduncles. Importantly, none of the cases exhibited intranuclear or cytoplasmic p62-positive inclusions, polyglutamine immunoreactivity, or alpha-synuclein or tau pathology in the cerebellum. Additionally, the substantia nigra showed no evidence of atrophy or neuronal loss. In some cases, mild neuronal loss and gliosis were noted in the inferior olivary and vestibular nuclei.

A recent study that examined the somatic instability of the *FGF14* GAA•TTC repeat in peripheral tissues matched with post-mortem brains and uncovered a tendency for the repeat to expand in cerebellar tissue somatically (Pellerin et al. 2025). Our study tested all five pathologically confirmed MSA cases with *FGF14* $(GAA)_{>300}$ expansions on DNA extracted from the cerebellum. This may account for the higher frequency of positive cases in this MSA cohort. Most of our MSA cases with an *FGF14* expansion had a variable combination of striatonigral degeneration and olivopontocerebellar involvement. This aligns with the clinical presentation of these patients, who often exhibited a mix of Parkinsonism and cerebellar ataxia. Our study did not find significant differences in alpha-synuclein pathology on neuropathological examination among MSA patients with or without *FGF14* GAA•TTC expansions, suggesting that these genetic mutations may contribute to the MSA phenotype through mechanisms independent of alpha-synuclein aggregation. Previous autopsy of SCA27B cases did not find significant alpha-synuclein pathology in the brain (D. Pellerin, Danzi, et al. 2023; Wilke et al. 2023). Further research is needed to elucidate how *FGF14* GAA•TTC expansion influences neurodegenerative processes and their interaction with synucleinopathies. To explore whether *FGF14* expansions lead to unique neuropathological alterations, beyond classical MSA features, particularly in the cerebellum, future studies would require advanced quantitative and molecular techniques for further morphometric analyses.

This study has several limitations. As the cohort is primarily pathologically based, the ascertainment of age of onset and clinical milestones relied on patient medical record reassessment. Secondly, our study cohorts were predominantly of European ancestry, which may limit the generalizability of the findings to other populations. The relatively small number of cases that carry an *FGF14* GAA•TTC expansion limits the statistical power of some analyses and the ability to draw definitive conclusions about the phenotypic differences between subgroups. Longitudinal studies are needed to better understand the natural history of MSA in patients with *FGF14* expansions and to evaluate the potential therapeutic implications of these genetic findings.

In conclusion, this study highlights the potential clinical and genetic overlap between MSA and SCA27B. These expansions in a subset of MSA neuropathologically confirmed patients underscore the need for genetic testing in diagnosing patients with clinical suspicion of MSA or in MSA patients with atypical and fast-progressing trajectories. These findings have important implications for improving diagnostic accuracy, drug trials, and understanding MSA's molecular underpinnings, ultimately contributing to better clinical management and therapeutic development for this complex neurodegenerative disorder.

Supplementary Materials

Supplementary material is available at *Brain* online.

Funding

VC was supported by the Association of British Neurologists' Academic Clinical Training Research Fellowship (grant ABN 540868) and the Guarantors of Brain (grant 565908). VC and HH received grants from the MSA Trust, MSA Coalition, MANX MSA, King Baudouin Foundation**,** the National Institute for Health Research (NIHR), University College London Hospitals Biomedical Research Centre, Michael J Fox Foundation (MJFF), Fidelity Trust, Rosetrees Trust, Ataxia UK, Alzheimer's Research UK (ARUK), NIH NeuroBioBank, and MRC Brainbank Network. This work was supported by an award from The Multiple System Atrophy Coalition (Core G). DP holds a Fellowship award from the Canadian Institutes of Health Research. JT received support from the German Research Foundation (DFG). PI was supported by a European Academy of Neurology Clinical Fellowship and a Fellowship from the Movement Disorders Group of the Spanish Society of Neurology. MJF was supported through the Lee and Lauren Fixel Chair for Parkinson's disease research. JBR is supported by the Medical Research Council (MC_UU_00030/14; MR/T033371/1), NIHR Cambridge Biomedical Research Centre (NIHR203312), and the Cambridge Centre for Parkinson-plus. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. M.S., C.W., and L.B. were supported by the Clinician Scientist program "PRECISE.net" funded by Else Kröner-Fresenius-Stiftung. For the purpose of open access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

The funders had no role in the conduct of this study.

Data Availability Statement

Individual deidentified data may be shared with any qualified investigator at a reasonable request (a data transfer agreement may be required to specify conditions of use required to protect anonymity and remain consistent with participant consent).

Acknowledgments

The authors would like to thank the participants and their families for their essential help with this work. We thank Dr. Paul Lockhart for a kind gift of amplicons with *FGF14* $(GAA)_n$ expansions in the pathologic range. We also thank the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida for contributing to long-read completion. Tissue samples and associated clinical and neuropathological data were supplied by the Queen Square Brain Bank, IBB-NeuroBiobank, CIBERNED Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas-Instituto de Salud Carlos III, the Multiple Sclerosis Society Tissue Bank (funded by the Multiple Sclerosis Society of Great Britain and Northern Ireland, registered charity 207495). We also thank Dr Gilbert Bensimon, Dr Christine Payan, Professor Ammar Al-Chalabi, and the NNIPPS Consortium for access to the NNIPPS data. We also thank the PROSPECT-M UK study for access to the sample and data from clinically diagnosed MSA patients.

Conflicts of Interest

JBR has undertaken remunerated consultancy or advisory board roles for Astronautx, Astex, Asceneuron, Alector, CumulusNeuro, Curasen, Eisai, Prevail, and SV Health, and has received academic grant income from AstraZeneca, Lilly, GSJ, and Janssen as partners in Dementias Platform UK. MS has received consultancy honoraria from Ionis, UCB, Prevail, Orphazyme, Biogen, Servier, Reata, GenOrph, AviadoBio, Biohaven, Zevra, Lilly, and Solaxa, all unrelated to the present manuscript. WGM has received consultancy fees from Lundbeck, Takeda, Inhibikase, GE, and Koneksa. HRM is employed by UCL. In the last 12 months, he reports paid consultancy from Roche, Aprinoia, AI Therapeutics, and Amylyx; lecture fees/honoraria - BMJ, Kyowa Kirin, Movement Disorders Society. Research Grants from Parkinson's UK, Cure Parkinson's Trust, PSP Association, Medical Research Council, and Michael J Fox Foundation. Dr Morris is a co-applicant on a patent application related to C9ORF72 - Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140). TF has served on Advisory Boards for Peptron, Voyager Therapeutics, Handl therapeutics, Gain therapeutics, Living Cell Technologies, Abbvie, Bluerock, Bayer & Bial. He has received honoraria for talks sponsored by Bayer, Bial, Profile Pharma, Boston Scientific & Novo Nordisk.

All other authors report no competing interests.

References

Abou Chaar, Widad, Anirudh N. Eranki, Hannah A. Stevens, Sonya L. Watson, Darice Y. Wong, Veronica S. Avila, Megan Delfeld, et al. 2024. 'Clinical, Radiological and Pathological Features of a Large American Cohort of Spinocerebellar Ataxia (SCA27B)'. *Annals of Neurology* 96 (6): 1092–1103. https://doi.org/10.1002/ana.27060.

Ando, M., Y. Higuchi, J. Yuan, A. Yoshimura, F. Kojima, Y. Yamanishi, Y. Aso, et al. 2024. 'Clinical Variability Associated with Intronic FGF14 GAA Repeat Expansion in Japan'. *Ann Clin Transl Neurol* 11 (1): 96–104. https://doi.org/10.1002/acn3.51936.

Bensimon, Gilbert, Albert Ludolph, Yves Agid, Marie Vidailhet, Christine Payan, P. Nigel Leigh, and NNIPPS Study Group. 2009. 'Riluzole Treatment, Survival and Diagnostic Criteria in Parkinson plus Disorders: The NNIPPS Study'. *Brain: A Journal of Neurology* 132 (Pt 1): 156–71. https://doi.org/10.1093/brain/awn291.

Bonnet, Céline, David Pellerin, Virginie Roth, Guillemette Clément, Marion Wandzel, Laëtitia Lambert, Solène Frismand, et al. 2023. 'Optimized Testing Strategy for the Diagnosis of GAA-FGF14 Ataxia/Spinocerebellar Ataxia 27B'. *Scientific Reports* 13 (1): 9737. https://doi.org/10.1038/s41598-023-36654-8.

Chia, Ruth, Anindita Ray, Zalak Shah, Jinhui Ding, Paola Ruffo, Masashi Fujita, Vilas Menon, et al. 2024. 'Genome Sequence Analyses Identify Novel Risk Loci for Multiple System Atrophy'. *Neuron* 112 (13): 2142-2156.e5. https://doi.org/10.1016/j.neuron.2024.04.002.

Cortese, Andrea, Stefano Tozza, Wai Yan Yau, Salvatore Rossi, Sarah J. Beecroft, Zane Jaunmuktane, Zoe Dyer, et al. 2020. 'Cerebellar Ataxia, Neuropathy, Vestibular Areflexia Syndrome Due to RFC1 Repeat Expansion'. *Brain: A Journal of Neurology* 143 (2): 480–90. https://doi.org/10.1093/brain/awz418.

Currò, Riccardo, Alessandro Salvalaggio, Stefano Tozza, Chiara Gemelli, Natalia Dominik, Valentina Galassi Deforie, Francesca Magrinelli, et al. 2021. 'RFC1 Expansions Are a Common Cause of Idiopathic Sensory Neuropathy'. *Brain: A Journal of Neurology* 144 (5): 1542–50. https://doi.org/10.1093/brain/awab072.

Gilman, S., G. K. Wenning, P. A. Low, D. J. Brooks, C. J. Mathias, J. Q. Trojanowski, N. W. Wood, et al. 2008. 'Second Consensus Statement on the Diagnosis of Multiple System Atrophy'. *Neurology* 71 (9): 670–76. https://doi.org/10.1212/01.wnl.0000324625.00404.15.

Gisatulin, Maria, Valerija Dobricic, Christine Zühlke, Yorck Hellenbroich, Vera Tadic, Alexander Münchau, Klaus Isenhardt, et al. 2020. 'Clinical Spectrum of the Pentanucleotide Repeat Expansion in the RFC1 Gene in Ataxia Syndromes'. *Neurology* 95 (21): e2912–23. https://doi.org/10.1212/WNL.0000000000010744.

Goh, Yee Yen, Emma Saunders, Samantha Pavey, Emma Rushton, Niall Quinn, Henry Houlden, and Viorica Chelban. 2023. 'Multiple System Atrophy'. *Practical Neurology* 23 (3): 208–21. https://doi.org/10.1136/pn-2020-002797.

Harris, R. S., M. Cechova, and K. D. Makova. 2019. 'Noise-Cancelling Repeat Finder: Uncovering Tandem Repeats in Error-Prone Long-Read Sequencing Data'. *Bioinformatics* 35 (22): 4809–11. https://doi.org/10.1093/bioinformatics/btz484.

Hengel, H., D. Pellerin, C. Wilke, Z. Fleszar, B. Brais, T. Haack, A. Traschutz, L. Schols, and M. Synofzik. 2023. 'As Frequent as Polyglutamine Spinocerebellar Ataxias: SCA27B in a Large German Autosomal Dominant Ataxia Cohort'. *Mov Disord* 38 (8): 1557–58. https://doi.org/10.1002/mds.29559.

Jabbari, Edwin, Negin Holland, Viorica Chelban, P. Simon Jones, Ruth Lamb, Charlotte Rawlinson, Tong Guo, et al. 2020. 'Diagnosis Across the Spectrum of Progressive Supranuclear Palsy and Corticobasal Syndrome'. *JAMA Neurology* 77 (3): 377–87. https://doi.org/10.1001/jamaneurol.2019.4347.

Jeon, Beom S., Matt J. Farrer, Stephanie F. Bortnick, and Korean Canadian Alliance on Parkinson's Disease and Related Disorders. 2014. 'Mutant COQ2 in Multiple-System Atrophy'. *The New England Journal of Medicine* 371 (1): 80. https://doi.org/10.1056/NEJMc1311763.

Koga, Shunsuke, Naoya Aoki, Ryan J. Uitti, Jay A. van Gerpen, William P. Cheshire, Keith A. Josephs, Zbigniew K. Wszolek, J. William Langston, and Dennis W. Dickson. 2015. 'When DLB, PD, and PSP Masquerade as MSA: An Autopsy Study of 134 Patients'. *Neurology* 85 (5): 404–12. https://doi.org/10.1212/WNL.0000000000001807.

Lüth, Theresa, Joshua Laβ, Susen Schaake, Inken Wohlers, Jelena Pozojevic, Roland Dominic G. Jamora, Raymond L. Rosales, et al. 2022. 'Elucidating Hexanucleotide Repeat Number and Methylation within the X-Linked Dystonia-Parkinsonism (XDP)-Related SVA Retrotransposon in TAF1 with Nanopore Sequencing'. *Genes* 13 (1): 126. https://doi.org/10.3390/genes13010126.

Matsushima, Masaaki, Hiroaki Yaguchi, Eriko Koshimizu, Akihiko Kudo, Shinichi Shirai, Takeshi Matsuoka, Shigehisa Ura, et al. 2024. 'FGF14 GAA Repeat Expansion and ZFHX3 GGC Repeat Expansion in Clinically Diagnosed Multiple System Atrophy Patients'. *Journal of Neurology* 271 (6): 3643–47. https://doi.org/10.1007/s00415-024-12308-1.

Méreaux, Jean-Loup, Claire-Sophie Davoine, David Pellerin, Giulia Coarelli, Marie Coutelier, Claire Ewenczyk, Marie-Lorraine Monin, et al. 2024. 'Clinical and Genetic Keys to Cerebellar Ataxia Due to FGF14 GAA Expansions'. *EBioMedicine* 99 (January):104931. https://doi.org/10.1016/j.ebiom.2023.104931.

Miki, Yasuo, Eiki Tsushima, Sandrine C. Foti, Kate M. Strand, Yasmine T. Asi, Adam Kenji Yamamoto, Conceição Bettencourt, et al. 2021. 'Identification of Multiple System Atrophy Mimicking Parkinson's Disease or Progressive Supranuclear Palsy'. *Brain: A Journal of Neurology* 144 (4): 1138–51. https://doi.org/10.1093/brain/awab017.

Multiple-System Atrophy Research Collaboration. 2013. 'Mutations in COQ2 in Familial and Sporadic Multiple-System Atrophy'. *The New England Journal of Medicine* 369 (3): 233–44. https://doi.org/10.1056/NEJMoa1212115.

Nakahara, Yasuo, Jun Mitsui, Hidetoshi Date, Kristine Joyce Porto, Yasuhiro Hayashi, Atsushi Yamashita, Yoshio Kusakabe, et al. 2023. 'Genome-Wide Association Study Identifies a New Susceptibility Locus in PLA2G4C for Multiple System Atrophy'. *medRxiv: The Preprint Server for Health Sciences*, May, 2023.05.02.23289328. https://doi.org/10.1101/2023.05.02.23289328.

Ouyang, R., L. Wan, D. Pellerin, Z. Long, J. Hu, Q. Jiang, C. Wang, et al. 2024. 'The Genetic Landscape and Phenotypic Spectrum of GAA-FGF14 Ataxia in China: A Large Cohort Study'. *EBioMedicine* 102 (April):105077. https://doi.org/10.1016/j.ebiom.2024.105077.

Papp, M. I., J. E. Kahn, and P. L. Lantos. 1989. 'Glial Cytoplasmic Inclusions in the CNS of Patients with Multiple System Atrophy (Striatonigral Degeneration, Olivopontocerebellar Atrophy and Shy-Drager Syndrome)'. *Journal of the Neurological Sciences* 94 (1–3): 79–100. https://doi.org/10.1016/0022-510x(89)90219-0.

Pellerin, D., M. C. Danzi, C. Wilke, M. Renaud, S. Fazal, M. J. Dicaire, C. K. Scriba, et al. 2023. 'Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia'. *N Engl J Med* 388 (2): 128–41. https://doi.org/10.1056/NEJMoa2207406.

Pellerin, D., G. F. Del Gobbo, M. Couse, E. Dolzhenko, S. K. Nageshwaran, W. A. Cheung, I. R. L. Xu, et al. 2024. 'A Common Flanking Variant Is Associated with Enhanced Stability of the FGF14-SCA27B Repeat Locus'. *Nat Genet* 56 (7): 1366–70. https://doi.org/10.1038/s41588-024-01808-5.

Pellerin, D., P. Iruzubieta, S. Tekgul, M. C. Danzi, C. Ashton, M. J. Dicaire, M. Wandzel, et al. 2023. 'Non-GAA Repeat Expansions in FGF14 Are Likely Not Pathogenic-Reply to: "Shaking Up Ataxia: FGF14 and RFC1 Repeat Expansions in Affected and Unaffected Members of a Chilean Family"'. *Mov Disord* 38 (8): 1575–77. https://doi.org/10.1002/mds.29552.

Pellerin, David, Matt C. Danzi, Mathilde Renaud, Henry Houlden, Matthis Synofzik, Stephan Zuchner, and Bernard Brais. 2024. 'Spinocerebellar Ataxia 27B: A Novel, Frequent and Potentially Treatable Ataxia'. *Clinical and Translational Medicine* 14 (1): e1504. https://doi.org/10.1002/ctm2.1504.

Pellerin, David, Jean-Loup Méreaux, Susana Boluda, Matt C. Danzi, Marie-Josée Dicaire, Claire-Sophie Davoine, David Genis, et al. 2025. 'Somatic Instability of the FGF14-SCA27B GAA•TTC Repeat Reveals a Marked Expansion Bias in the Cerebellum'. *Brain: A Journal of Neurology* 148 (4): 1258–70. https://doi.org/10.1093/brain/awae312.

Pellerin, David, Carlo Wilke, Andreas Traschütz, Sara Nagy, Riccardo Currò, Marie-Josée Dicaire, Hector Garcia-Moreno, et al. 2024. 'Intronic FGF14 GAA Repeat Expansions Are a Common Cause of Ataxia Syndromes with Neuropathy and Bilateral Vestibulopathy'. *Journal of Neurology, Neurosurgery, and Psychiatry* 95 (2): 175–79. https://doi.org/10.1136/jnnp-2023-331490.

Rafehi, H., J. Read, D. J. Szmulewicz, K. C. Davies, P. Snell, L. G. Fearnley, L. Scott, et al. 2023. 'An Intronic GAA Repeat Expansion in FGF14 Causes the Autosomal-Dominant Adult-Onset Ataxia SCA50/ATX-FGF14'. *Am J Hum Genet* 110 (1): 105–19. https://doi.org/10.1016/j.ajhg.2022.11.015.

Satolli, Sara, Salvatore Rossi, Elisa Vegezzi, David Pellerin, Maria Laura Manca, Melissa Barghigiani, Carla Battisti, et al. 2024. 'Correction to: Spinocerebellar Ataxia 27B: A Frequent and Slowly Progressive Autosomal-Dominant Cerebellar Ataxia-Experience from an Italian Cohort'. *Journal of Neurology* 271 (12): 7650–51. https://doi.org/10.1007/s00415-024-12629-1.

Shirai, Shinichi, Keiichi Mizushima, Keishi Fujiwara, Eriko Koshimizu, Masaaki Matsushima, Satoko Miyatake, Ikuko Iwata, Hiroaki Yaguchi, Naomichi Matsumoto, and Ichiro Yabe. 2023. 'Case Series: Downbeat Nystagmus in SCA27B'. *Journal of the Neurological Sciences* 454 (November):120849. https://doi.org/10.1016/j.jns.2023.120849.

Spillantini, M. G., R. A. Crowther, R. Jakes, N. J. Cairns, P. L. Lantos, and M. Goedert. 1998. 'Filamentous Alpha-Synuclein Inclusions Link Multiple System Atrophy with Parkinson's Disease and Dementia with Lewy Bodies'. *Neuroscience Letters* 251 (3): 205–8. https://doi.org/10.1016/s0304-3940(98)00504-7.

Sullivan, Roisin, Wai Yan Yau, Viorica Chelban, Salvatore Rossi, Natalia Dominik, Emer O'Connor, John Hardy, Nicholas Wood, Andrea Cortese, and Henry Houlden. 2021. 'RFC1-Related Ataxia Is a Mimic of Early Multiple System Atrophy'. *Journal of Neurology, Neurosurgery, and Psychiatry* 92 (4): 444–46. https://doi.org/10.1136/jnnp-2020-325092.

Sullivan, Roisin, Wai Yan Yau, Viorica Chelban, Salvatore Rossi, E. O'Connor, Nicholas W. Wood, Andrea Cortese, and Henry Houlden. 2020. 'RFC1 Intronic Repeat Expansions Absent in Pathologically

Confirmed Multiple Systems Atrophy'. *Movement Disorders: Official Journal of the Movement Disorder Society* 35 (7): 1277–79. https://doi.org/10.1002/mds.28074.

Traschütz, Andreas, Andrea Cortese, Selina Reich, Natalia Dominik, Jennifer Faber, Heike Jacobi, Annette M. Hartmann, et al. 2021. 'Natural History, Phenotypic Spectrum, and Discriminative Features of Multisystemic RFC1 Disease'. *Neurology* 96 (9): e1369–82. https://doi.org/10.1212/WNL.0000000000011528.

Trojanowski, J. Q., T. Revesz, and Neuropathology Working Group on MSA. 2007. 'Proposed Neuropathological Criteria for the Post Mortem Diagnosis of Multiple System Atrophy'. *Neuropathology and Applied Neurobiology* 33 (6): 615–20. https://doi.org/10.1111/j.1365-2990.2007.00907.x.

Wenning, Gregor K., Iva Stankovic, Luca Vignatelli, Alessandra Fanciulli, Giovanna Calandra-Buonaura, Klaus Seppi, Jose-Alberto Palma, et al. 2022. 'The Movement Disorder Society Criteria for the Diagnosis of Multiple System Atrophy'. *Movement Disorders: Official Journal of the Movement Disorder Society* 37 (6): 1131–48. https://doi.org/10.1002/mds.29005.

Wilke, C., D. Pellerin, D. Mengel, A. Traschutz, M. C. Danzi, M. J. Dicaire, M. Neumann, et al. 2023. 'GAA-FGF14 Ataxia (SCA27B): Phenotypic Profile, Natural History Progression and 4-Aminopyridine Treatment Response'. *Brain* 146 (10): 4144–57. https://doi.org/10.1093/brain/awad157.

Wirth, Thomas, Céline Bonnet, Clarisse Delvallée, David Pellerin, Thomas Bogdan, Guillemette Clément, Audrey Schalk, et al. 2024. 'Does Spinocerebellar Ataxia 27B Mimic Cerebellar Multiple System Atrophy?' *Journal of Neurology* 271 (4): 2078–85. https://doi.org/10.1007/s00415-024-12182-x.

Wirth, Thomas, Guillemette Clément, Clarisse Delvallée, Céline Bonnet, Thomas Bogdan, Andra Iosif, Audrey Schalk, et al. 2023. 'Natural History and Phenotypic Spectrum of GAA-FGF14 Sporadic Late-Onset Cerebellar Ataxia (SCA27B)'. *Movement Disorders: Official Journal of the Movement Disorder Society* 38 (10): 1950–56. https://doi.org/10.1002/mds.29560.

Title

Genetic Testing for SCA27B in Korean Multiple System Atrophy

Authors

Joshua Laß[1], Michele Berselli[2], Doug Rioux[2], Susen Schaake[1], Jordan Follett[3], Jonathan E. Bravo[3], Alexander D. Veit[2], William Ronchetti[2], Sarah B. Reiff[2], Matthew J. Huentelman[4], Dana Vuzman[2,5], Pamela Bower[6], Vikram Khurana[7], Joanne Trinh[1], Beomseok Jeon[8], Han-Joon Kim[8]* and Matthew J. Farrer[3]*

[1] Institute of Neurogenetics, University of Lübeck, Ratzeburger Allee 160, 23538 Lübeck, Germany

[2] Department of Biomedical Informatics, Harvard Medical School, 10 Shattuck Street, Boston, MA 02115, USA

[3] McKnight Brain Institute, Department of Neurology, College of Medicine, University of Florida, 1149 Newell Drive, Gainesville, FL 32611, USA

[4] Neurogenomics Division, Translational Genomics Research Institute, 445 N. Fifth Street, Phoenix, AZ 85004, USA

[5] Division of Genetics, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, MA 02115, USA

[6] Mission MSA (formerly MSA Coalition), 1660 International Drive, Suite 600, McLean, VA 22102, USA

[7] Division of Movement Disorders, Department of Neurology, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, USA

[8] Department of Neurology, Seoul National University Hospital, Seoul National University College of Medicine, 101, Daehak-Ro, Jongno-Gu, Seoul 03080, Korea.

* Contributed equally and corresponding author

Published in *Brain*, 2025, awaf263, DOI: 10.1093/brain/awaf263, Online ahead of print

Letter to the editor

Recently, Chelban et al. reported that FGF14 $GAA_{\geq 250}$ repeat expansions impact progression and survival in multiple system atrophy (MSA) (Chelban et al. 2025). They identified 2.89% (n=19/657) in MSA versus 1.40% (n=12/1003) in controls. These 19 cases were reported to have faster progression and a repeat size that was inversely correlated with survival. To extend and validate these results, we comprehensively examined all genetic variability in the *FGF14* locus in MSA in an ongoing longitudinal study from South Korea (Kim et al. 2011; Shin et al. 2022). Patients were enrolled in the Seoul National University Hospital (SNUH) outpatient clinic setting from 2012 to 2021. Patients with 'probable MSA', according to the 2nd consensus criteria, who provided written informed consent, were eligible to participate. The study was approved by the institutional review boards (IRB) at SNUH (H-1601-048-733) and the University of Florida (#IRB202000632) to enable whole genome sequencing. Our cohort included 199 individuals (91 male, 108 female) with a mean age at examination of 58.2 ± 8.3 (median 58; range 34-79), of which 122 patients had MSA-C and 77 MSA-P (Supplementary table 1) (Shin et al. 2022). The minimum mean sequencing depth for all samples was 35X (Supplementary methods). All data have been shared with the MSA Coalition Collaborative Core Network (https://missionmsa.org/), and additional anonymized genomic data have also been made available from healthy Korean volunteers participating in the Korean Genome Project (KGP, controls n=1,048) (Jeon et al. 2020). All non-synonymous single-nucleotide variants (SNVs) and small insertions and deletions (indels) in *FGF14* were investigated. None were consistent with SCA27A, and no variants have a known pathogenic or likely pathogenic annotation in Clinvar (Landrum et al. 2018). All genomic variability observed in *FGF14* in MSA was also examined in jointly called WGS data from the Korean Genome Project (KGP controls) (Jeon et al. 2020). One MSA-C patient has a rare coding variant observed in exon 5 c.707C>T (p.Ala326Val; NC_000013.11:g.101722883G>A; MAF KGPcontrols= 0.0019, gnomADEastAsia=0.0002), whereas a total of ten patients had rare indel variants of uncertain significance (VUS) within the 3'UTR (Supplementary table 2). We interrogated all MSA genomes for *FGF14* 5'-flanking $(GAA)_n$ repeats with the EH algorithm (Dolzhenko et al. 2019). The $(GAA)_n$ triplet repeat number was 17.8 ± 16.2 SD in Korean MSA patients versus 17.9 ± 17.2 SD in KGP controls. The most frequent $(GAA)_n$ repeat number per allele was 10 (27.4%). Comparable results and minor allele frequencies (MAF) were predicted for all motifs in cases and controls (Supplementary Figure 1). However, we sought to validate allele sizes based on EH results with PCR-based amplicon length analysis.

FGF14 $(GAA)_n$ locus PCR was subsequently performed in 199 MSA patients and 196 ethnically matched control participants (Figure 1). The overall distributions in control individuals (mean=37 ± SD 57 bp) and MSA (mean=44 ± SD 68 bp) were not significantly different (p>0.07). The majority of allele sizes in both groups were <50 bp, including 89.2% (350/392) of controls and 86% (296/344) of MSA. However, intermediate-sized $FGF14_{>150}$ alleles were found in 13.4% (46/344) of MSA (mean=195 ± SD 80 bp) and 6.9% (27/392) of controls (mean=224 ± SD 64 bp). However, the threshold for intermediate repeat expansion is >250, and alleles with fewer GAA repeats should be considered in the normal range. As amplicon PCR with gel sizing and EH results were inconsistent, we used linear regression to assess their relationship (p=0.18, R^2=0.07). A subset of expanded alleles (n=11) was also interrogated using long-read sequencing. The median quality of reads (phred score) was q=13.4, and the median read length (including both alleles) was 0.934 kb. The mean read coverage was 63,411X. The results of the repeat length were more consistent with amplicon PCR gel sizing results (p=$0.75x10^{-4}$, R^2=0.82). It was possible to detect a complex repeat interruption pattern in 10/11 patients with the long expansion $(GCAGAAGAAGAAGAA)_n(GCAGAA)_n(GAA)_n(GAG)$ (Supplementary table 2). These patients were comparable to the cohort overall, and none had atypical clinical features (Supplementary Table 1). In contrast to the PCR, long-read sequencing did not correlate with EH for $(GAA)_n$ expansion size (p=0.052, R^2=0.29).

In summary, *FGF14* haploinsufficiency and loss-of-function contribute to spinocerebellar ataxia type 27A (SCA27A) and SCA27B (van Swieten et al. 2003b; David Pellerin et al. 2023; Haloom Rafehi et al. 2023). Nevertheless, there is no evidence for pathogenic genetic variability within the *FGF14* locus in this Korean sample of MSA patients. Our results exclude *FGF14* $(GAA)_{>300}$ pathogenic repeat expansions and rare *FGF14* coding variants. However, we do observe a higher frequency of intermediate FGF14 $(GAA)_{>150}$ expanded alleles in patients (13.4%; 46/344) compared to controls (6.9%; 27/392). While intermediate-sized $FGF14_{<250}$ repeats are not considered pathogenic, and this sample size is too small for significance, further reporting across a range of repeat sizes with meta-analysis may reveal some relationship with MSA susceptibility. Amplicon PCR gel sizing and long-read sequencing were highly correlated, but the former is not a precise tool for analyzing exact repeat lengths or interruptions (Supplementary Table 2). Unfortunately, EH could not accurately annotate long, complex expanded repeats. Thus, we recommend amplicon PCR analysis for future genetic testing to highlight potential $(GAA)_{>250}$ repeat expansions, with subsequent confirmation by long-read sequencing.

Supplementary material

Supplementary material is available at *Brain* online.

Acknowledgments

We thank all patients and families who participated in this study. We thank Dr. Paul Lockhart for a kind gift of amplicons with *FGF14* $(GAA)_n$ expansions in the pathologic range. We thank previous members of the Laboratory of Neurogenetics and Neuroscience for their technical support, and the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida. This work was made possible by an Award from Mission Multiple System Atrophy. GCAP is supported by the Blavatnik Clinical Pilot Award to Drs. Peter Park and Shamil Sunyaev, The Chan Zuckerberg Initiative EOSS5 grant 2022-309591 (an advised fund of the Silicon Valley Community Foundation) to Dana Vuzman, The Department of Biomedical Informatics and The Park Lab at Harvard Medical School. Matthew J. Farrer was supported through the Clinical & Translational Genomics Institute and a Lee and Lauren Fixel Chair for Parkinson's disease research.

References

Chelban, Viorica, David Pellerin, Nirosen Vijiaratnam, Hamin Lee, Yen Yee Goh, Lauren Brown, Sara Sambin, et al. 2025. 'Intronic FGF14 GAA Repeat Expansions Impact Progression and Survival in Multiple System Atrophy'. *Brain*, April, awaf134. https://doi.org/10.1093/brain/awaf134.

Dolzhenko, Egor, Viraj Deshpande, Felix Schlesinger, Peter Krusche, Roman Petrovski, Sai Chen, Dorothea Emig-Agius, et al. 2019. 'ExpansionHunter: A Sequence-Graph-Based Tool to Analyze Variation in Short Tandem Repeat Regions'. *Bioinformatics (Oxford, England)* 35 (22): 4754–56. https://doi.org/10.1093/bioinformatics/btz431.

Jeon, Sungwon, Youngjune Bhak, Yeonsong Choi, Yeonsu Jeon, Seunghoon Kim, Jaeyoung Jang, Jinho Jang, et al. 2020. 'Korean Genome Project: 1094 Korean Personal Genomes with Clinical Information'. *Science Advances* 6 (22): eaaz7835. https://doi.org/10.1126/sciadv.aaz7835.

Kim, Han-Joon, Beom S. Jeon, Jee-Young Lee, and Ji Young Yun. 2011. 'Survival of Korean Patients with Multiple System Atrophy'. *Movement Disorders: Official Journal of the Movement Disorder Society* 26 (5): 909–12. https://doi.org/10.1002/mds.23580.

Landrum, Melissa J., Jennifer M. Lee, Mark Benson, Garth R. Brown, Chen Chao, Shanmuga Chitipiralla, Baoshan Gu, et al. 2018. 'ClinVar: Improving Access to Variant Interpretations and Supporting Evidence'. *Nucleic Acids Research* 46 (D1): D1062–67. https://doi.org/10.1093/nar/gkx1153.

Pellerin, David, Matt C. Danzi, Carlo Wilke, Mathilde Renaud, Sarah Fazal, Marie-Josée Dicaire, Carolin K. Scriba, et al. 2023. 'Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia'. *New England Journal of Medicine* 388 (2): 128–41. https://doi.org/10.1056/NEJMoa2207406.

Rafehi, Haloom, Justin Read, David J. Szmulewicz, Kayli C. Davies, Penny Snell, Liam G. Fearnley, Liam Scott, et al. 2023. 'An Intronic GAA Repeat Expansion in FGF14 Causes the Autosomal-Dominant Adult-Onset Ataxia SCA50/ATX-FGF14'. *The American Journal of Human Genetics* 110 (1): 105–19. https://doi.org/10.1016/J.AJHG.2022.11.015.

Shin, Jung Hwan, Han-Joon Kim, Chan Young Lee, Hee Jin Chang, Kyung Ah Woo, and Beomseok Jeon. 2022. 'Laboratory Prognostic Factors for the Long-Term Survival of Multiple System Atrophy'. *NPJ Parkinson's Disease* 8 (1): 141. https://doi.org/10.1038/s41531-022-00413-9.

Swieten, John C. van, Esther Brusse, Bianca M. de Graaf, Elmar Krieger, Raoul van de Graaf, Inge de Koning, Anneke Maat-Kievit, et al. 2003. 'A Mutation in the Fibroblast Growth Factor 14 Gene Is Associated with Autosomal Dominant Cerebellar Ataxia [Corrected]'. *American Journal of Human Genetics* 72 (1): 191–99. https://doi.org/10.1086/345488.

Discussion

In this thesis, the advantages of long-read sequencing on the investigation of repeat expansion have enabled the direct detection of genetic variability, epigenetic information, and the characterization of repeat expansions. Long-read sequencing provides specificity in detecting repeat numbers in XDP and SCA27B (Lüth et al. 2022; Laß, Thomsen, et al. 2025) (Figure 6). Furthermore, the technology has even enabled the discovery of novel somatic interruptions (Trinh et al. 2023; Laß, Thomsen, et al. 2025) where the frequency of mosaic interruptions is associated with the stability of the repeat length (Laß et al. 2024). In SCA27B, my results led to a new categorization of the patient's affection status using the pure GAA length without interruptions (Laß, Thomsen, et al. 2025). I have also contributed to two studies showing phenotypic pleiotropy of the *FGF14* repeat expansion, observing the expansion in patients with MSA (Chelban et al. 2025; Laß, Berselli, et al. 2025).

Repeat length detection with long-read sequencing

In the first study (Lüth et al. 2022), there is high concordance between long-read sequencing and fragment analysis. Interestingly, we consistently detected one to four more repeats compared to the fragment analysis in the *TAF1* hexanucleotide repeat expansion, which the interruption at the beginning of the repeat tract can explain. Besides *TAF1* and *FGF14*, a study has shown accurate detection for SCA-related repeat expansions using Oxford Nanopore Technologies (De Boer et al. 2025). They have designed a gene panel for long-read sequencing and could accurately determine the repeat length at over 20 loci. Similar to our study, long-read sequencing is suggested as a better tool for diagnosing repeat expansions in neurological diseases compared to short-reads and other PCR-based methods (De Boer et al. 2025).

A limitation of the detection of the repeat length is the software. In our studies, we used the software Noise-canceling repeat finder (NCRF), which is optimized software for long-read sequencing data. The software STRique, which is more commonly used, did not detect the repeat length concordant with the fragment analysis results as NCRF. Additionally, NCRF enables the direct detection of interruptions in every read, which is impossible with STRique. A limitation of our results is that no other available software can detect mosaic interruptions in the repeat tract. Long reads spanning across the repeat region allow the identification of the exact location of the repeat expansion and length simultaneously.

For example, long-read sequencing discovered a GGC repeat expansion in the *NOTCH2NLC* gene in neuronal intranuclear inclusion disease and determined the repeat length (Sone et al. 2019). In 2019, long-read sequencing detected TTTCA and TTTTA repeat expansions in the intron of *SAMD12* and determined the repeat length in families with familial cortical myoclonic tremor with epilepsy (Zeng et al. 2019). In this study, short-read sequencing could not identify the causative mutation within this locus. The results underline the potential of long-read sequencing as a diagnostic tool.

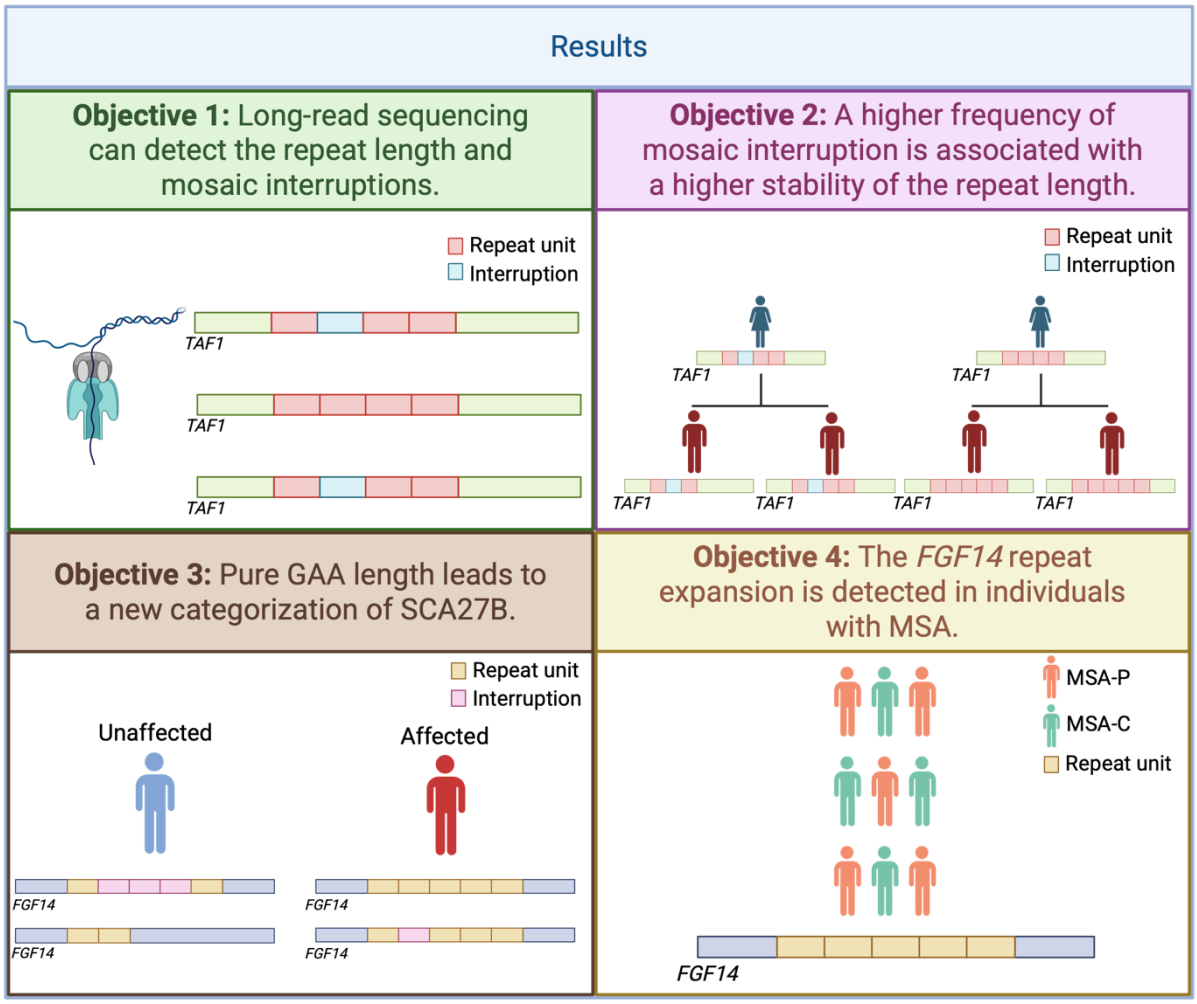

Figure 6: Summary of the results. ***Objective 1:*** I have shown that long-read sequencing by Oxford Nanopore technologies can accurately detect the repeat length and has enabled the identification of mosaic interruption in the *TAF1* repeat expansion (Lüth et al. 2022; Trinh et al. 2023). ***Objective 2:*** In a follow-up study, I have analyzed the repeat number's stability across generations and found that a higher frequency of the mosaic interruption is associated with a higher stability of the repeat length (Laß et al. 2024). ***Objective 3:*** I am supposing a new categorization of SCA27B based on the pure repeat length in *FGF14*. Individuals with a pure repeat length longer than 200 are affected with SCA27B, and individuals with a shorter pure repeat length are unaffected (Laß, Thomsen, et al. 2025). ***Objective 4:*** The repeat expansion in *FGF14* was also detected in individuals with MSA in two independent cohorts with a low frequency (Chelban et al. 2025), created in https://BioRender.com. Legend: TAF1: Transcription factor 1; XDP: X-linked dystonia-parkinsonism; FGF14: Fibroblast growth factor 14; MSA-P: Multiple system atrophy-parkinsonian subtype; MSA-C: Multiple system atrophy-cerebellar subtype.

General limitations of long-read sequencing include difficulties with larger expansions, like in c9orf72. Only a few studies have successfully sequenced the full-length repeat expansion in *c9orf72* with low coverage (DeJesus-Hernandez et al. 2021; Ebbert et al. 2018; Udine et al. 2024). The repeat length can be determined with low coverage, but a complete characterization of the repeat tract is impossible.

The advantage of Oxford Nanopore Technologies is that methylation can be detected simultaneously. In XDP, we performed a Cas9-targeted approach to investigate the frequency of methylation in the SVA insertion and flanking regions. It is suggested that retrotransposons can introduce methylation changes into the flanking regions (Yates et al. 1999). We observed hypermethylation of the *TAF1* SVA insertion in concordance with the literature (Makino et al. 2007). However, only two enhancer sites in the flanking regions showed a slight increase in methylation frequency in patient-derived DNA. The brain-derived DNA was mildly reduced compared to the blood-derived DNA in the patient. It is unclear if this subtle change could affect the expression level of *TAF1.* A limitation of the study is the lack of brain-derived DNA from a control to compare the methylation frequency of the enhancer site in the brain tissue. Therefore, it is unclear if the decreased methylation frequency in the brain-derived DNA of the patient is due to the insertion of the SVA or the brain tissue. Still, these results show the utility of long-read sequencing in investigating the methylation frequency of SVA insertion in *TAF1*.

Characterization of the repeat tract by long-read sequencing

Previous methods for repeat length detection, like fragment analysis or long-range PCR, cannot detect interruptions in the repeat tract due to a lack of base pair information. Additionally, triplet repeat-primed PCR or MolbII digestion must be performed to detect repeat interruptions. In contrast, long-read sequencing enables the direct characterization of the repeat tract and repeat length determination in one experiment. In this thesis, two examples of repeat interruptions were discovered. In XDP, a specific combination of interruptions at the beginning of the 5'-repeat tract was detected, and in the *FGF14* gene, different interruption motifs starting at the 5'-end were detected.

Polarization of repeat interruptions to the 5' end is also seen in other repeat expansion disorders. The AGG interruption in the *FMR1* gene and the CAA interruption in the *HTT* gene are detected at the 5'-end of the repeat tract (Wright et al. 2020; Nolin et al. 2019).

A similar location of many interruptions suggests they may arise through a biological mechanism, i.e., during the mismatch repair mechanism or by an exo-nuclease.

Repeat interruptions, in general, have been observed in several repeat expansion disorders and may be a protective factor for the diseases. They can influence methylation, stability of the repeat lengths, and RNA structures. A recent study has shown that the CGG repeat tracts typically fold into a stem-loop hairpin structure that can undergo a conformational rearrangement (Shen et al. 2024). The repeat interruption leads to the formation of an internal loop, which reduces the slippage of the polymerase in the repeat tract. This is an example of the protective effect of repeat interruptions.

Another example is SCA2, where the CAG repeat expansion in the *ATXN2* gene can be interrupted by CAA. The CAA interruption can stabilize the repeat tract by preventing the formation of the hairpin. A recent study has shown that the position of the interruption can increase or decrease the stability of the CAG tract (Lyasota et al. 2024).

In DM1, 3-5% of the families with DM1 have an interruption in the CTG repeat tract (Rajagopal et al. 2023). These interruptions stabilize the repeat tract and reduce the somatic instability of the repeat number (Cumming et al. 2018). A slower disease progression with a lack of typical clinical features in DM1 can result from the interruptions (Braida et al. 2010).

The interruptions of the repeat expansions in *TAF1* and *FGF14* have been found in a somatic mosaic fashion, which has not yet been seen in other repeat expansion disorders. The known repeat interruptions are reported as germline configurations, and the possibility that they may vary somatically as putative mosaic repeat disruptions has not been looked at. Long-read sequencing will improve the characterization of repeat expansions and the detection of interruptions.

At the time of our first study, the accuracy of the long-read sequencing by Oxford Nanopore technologies was not as high as today (Lüth et al. 2022). Oxford Nanopore Technologies offers different base-calling models, a fast base-calling model, and a super-accuracy base-calling model to address customer needs. We could only detect the interruptions with the super-accuracy base-calling model due to its higher q-score of the bases. The determined repeat length with the fast base-calling model had a high concordance with the fragment analysis. Still, detecting the interruptions at the beginning of the repeat tract was unsuccessful. Therefore, the accuracy of the base calling is a limitation when investigating interruptions in the repeat tract. Today, long-read sequencing by Oxford Nanopore Technologies can reach an accuracy of 99.9%. We have validated the results of our first study with improved accuracy and identified the interruptions at the same position and with the same frequency (Laß et al. 2024).

Repeat interruptions are an essential finding of translational relevance, as they suggest a shared mechanism across different repeat expansion disorders that can potentially enhance the prediction of disease manifestation in individual repeat expansion carriers.

Repeat interruptions act as a modifier

The results in this thesis show that repeat interruption can act as a modifier in repeat expansion disorders. The interruptions in XDP affect the transmission of the repeat length across generations (Figure 6). A high mDRILS frequency can stabilize the repeat length across generations, which shows a protective effect of the interruptions in repeat expansion. Anticipation (longer repeat lengths across familial transmission) would typically lead to an earlier age at onset of XDP over multiple generations, but mDRILS behave differently.

As mDRILS is novel, investigations across families have been limited. Still, examples of non-somatic repeat interruptions have been observed. In the Fragile X-Syndrome, interruptions are associated with increased stability of the repeat expansion across generations. The interruptions on the maternal allele stabilize the repeat tract across generations, and alleles without interruptions are more likely to expand the repeat number (Domniz et al. 2018). In XDP, we observed that the paternal allele has a higher mDRILS frequency, and the higher frequency is likely to stabilize the repeat length across generations.

Additionally, interruptions in the repeat tract can modify the age at onset by years. In Friedreich's ataxia, Interruptions of the $(GAA)_n$ tract can delay age at onset by 9 years (Al-Mahdawi et al. 2018). In Huntington's disease, a pure HTT $(CAG)_n$ tract without interruption can hasten disease by 13–29 years (Wright et al. 2020). In contrast, a multiply-interrupted tract can delay disease by 3–6 years in Huntington's disease. In XDP, we observed that mDRILS indirectly delays the age at onset (Trinh et al. 2023). A higher mDRILS frequency is associated with a shorter repeat length, which is linked to a later onset of symptoms.

Besides affecting the age at onset and stabilizing the repeat number, repeat interruptions can alter the clinical phenotype. In DM1, the phenotype is milder in patients with an interruption than in patients with a pure repeat tract (Rasmussen et al. 2022; Peric et al. 2021). But interruptions can also lead to a more severe phenotype. This is seen in SCA10, where the interruptions lead to an earlier age at onset, higher frequency of epileptic seizures, and worse ataxia severity (Hasan et al. 2025; Matsuura et al. 2006; McFarland et al. 2013).

In this thesis, the repeat length was investigated longitudinally for a maximum of five years. In most patients, the repeat length was stable in this time frame, or there was a small one-repeat unit change. In contrast, the mDRILS frequency was very unstable. The variation of the AGGG frequency could not be correlated with repeat length changes due to the short period. Additionaly, the impact on the disease duration and severity cannot be displayed in this short time frame. Therefore, a larger time frame is necessary to investigate these effects. However, the changes in the mDRILS frequency lead to the speculation that an impact on the repeat length is possible over a longer period.

Repeat interruptions and pathogenicity of the repeat expansion in *FGF14*

In SCA27B, the detection of the interruptions leads to a new concept of the pathogenic threshold of the repeat expansion in the *FGF14* gene. The new concept is based on the pure length of the GAA tract. This tract is the longest pure GAA tract, depending on the position of the interruption.

In SCA27B, the pure GAA length was used to determine a new pathogenic threshold. The new proposed threshold is 200 pure GAA repeats, independent of the length or position of interruptions. Therefore, four scenarios are possible: first, individuals with over 200 pure GAA repeat units and an ataxia diagnosis are individuals with SCA27B. Second, individuals with the same genetic consistency and age below 60 are considered premanifesting carriers of SCA27B. Third, unaffected individuals with a pure GAA length under 200 will not develop SCA27B. Fourth, previously diagnosed individuals with SCA27B with a pure GAA length under 200 need to be revisited and likely revised to another form of ataxia. In the study, Chilean families were previously diagnosed with SCA27B but differed from the reported phenotype of SCA27B patients. I have shown that these individuals carry a long interruption in the repeat tract, which results in a pure GAA length of 65. A movement disorder specialist will now revisit them to determine the correct type of ataxia. This example shows the importance of characterizing the repeat tract for a better diagnosis, ultimately leading to better treatment for these individuals. This concept is based on the analysis of 40 individuals. A larger cohort is needed to validate and refine the threshold length.

In the literature, interruptions in the repeat expansion of the *FGF14* gene were considered to be non-pathogenic. Still, in this thesis, the data imply that short interruptions can also be detected in individuals with SCA27B (Mohren et al. 2024). Individuals with a short interruption at the 3'- or 5'-end of the repeat expansion, and a pure GAA length over 200, have SCA27B. This shows the importance of characterizing the repeat tract with long-read sequencing.
The length of the pure repeat tract was also investigated in Spinocerebellar Ataxia (SCA) in the novel gene *THAP11*. A study identified more interruptions in healthy individuals than those with SCA (Tan et al. 2023). Like the repeat expansion in *FGF14*, the affected individuals had a longer pure CAG tract than the unaffected individuals. It is suggested that the toxicity of the translated protein is dependent on the length of the pure CAG tract.
The pure repeat length was also investigated in Huntington's disease (Wright et al. 2019). The CAA repeat interruption in the CAG repeat tract codes for the same amino acid, glutamine, but is still associated with an earlier age at onset. The somatic instability is also higher in individuals with no interruption (Wright et al. 2019). The pure CAG repeat number has a higher effect on the age at onset than the polyglutamine length, unaffected by the interruption.

Repeat expansions in other neurological disorders

In the last two studies, the repeat expansion in the *FGF14* gene was investigated in individuals with MSA (Laß, Berselli, et al. 2025; Chelban et al. 2025). The frequency of the expansion over 250 repeats was higher in patients compared to healthy individuals (Figure 6). Interestingly, no difference in the frequency was detected between the two subtypes, MSA-C and MSA-P. The MSA-C subtype is more similar to SCA27B and was expected to have a higher expansion frequency in the *FGF14* gene. The frequency of the repeat expansion over 250 was relatively low, with 2.9% in the individuals with MSA. The individuals carrying a repeat expansion greater than 250 had more severe symptoms than patients without a *FGF14* repeat expansion. For example, these individuals have a higher tendency to fall in the first year after disease onset and have a shorter time to the use of a wheelchair. These examples show that the repeat expansion can be a potential modifier in MSA.
In the other study, a Korean cohort was screened for the *FGF14* repeat expansion in individuals with MSA. 13,4% of the individuals with MSA had an expansion of over 150 repeats. These results show that intermediate repeat expansions can also be detected in other neurological disorders. MSA is one of the first examples of a disorder that is not itself a repeat expansion disorder but can still be modified by repeat expansions.

A limitation of these two studies is the relatively low number of individuals carrying a repeat expansion in *FGF14*. These results show the importance of screening for repeat expansion in other neurological disorders, especially the expansion in the *FGF14* gene. Parkinson's disease has similar symptoms to the MSA-P subtype, which may indicate that the expansion in the *FGF14* gene can potentially act as a modifier in PD. Further research is needed to investigate the influence of the FGF14 repeat expansion on neurodegenerative processes.

Outlook

In this thesis, the interruptions in the repeat tract were detected, characterized, and tested for associations with the repeat stability and age at onset. Still, the functional consequences of the interruption were not investigated. One possible consequence of the observed repeat interruptions could be their effect on the formation of DNA structures. The *TAF1* repeat expansion in XDP, the high guanine content, suggests the potential for G-quadruplex formation. The detected interruptions may alter the spacing between the guanine bases, thereby interfering with forming such structures. Understanding how these interruptions influence DNA structures could provide new insights into the regulation of gene expression or the pathogenic mechanism underlying XDP.

The detection of mosaic interruptions in XDP and SCA27B highlights the potential for similar patterns in other repeat expansion disorders. Friedreich's ataxia is a promising candidate with the same repeat motif with GAA as SCA27B. Given this similar sequence, it is plausible that the repeat expansion in the *FXN* gene may also harbor mosaic interruptions.

The detection of interruptions in SCA27B has demonstrated its relevance for improving genetic diagnosis in repeat expansion carriers. This highlights the importance of accurately determining the pure repeat length in other repeat expansions. Determining the position and motif composition of interruptions will be essential for refining diagnostic interpretation, improving genotype-phenotype correlations, and ultimately informing prognosis and the development of therapeutic strategies in repeat expansion disorders.

Long-read sequencing platforms have opened new possibilities for detecting repeat expansion disorders in a diagnostic setting. Pacific Biosciences offers an enrichment kit for a selected set of genes based on CRISPR/Cas9-mediated targeting of specific loci. The number of diseases covered is limited by the design of the kit. In contrast, Oxford Nanopore Technologies has introduced adaptive sampling, enabling targeted enrichment without amplification or enrichment kits. During adaptive sampling, a segment of a DNA strand is sequenced and immediately mapped to a reference genome. If the read does not align to a predefined target region, it is rejected, allowing the sequencer to move on to the next DNA strand. Reads that match the predefined region are sequenced fully. This process enables efficient enrichment of genomic regions directly from native DNA. This approach is promising for developing diagnostic panels covering a broad spectrum of repeat expansion disorders. Not only can it accommodate all known repeat expansion disorders, but it also allows the inclusion of newly discovered repeat expansions.

By leveraging adaptive sampling, long-read sequencing could offer a flexible and amplification-free diagnostic tool for repeat expansion disorders, improving detection sensitivity and enabling faster and more comprehensive diagnoses.

References

Abdi, Mohammad Hossein, Bita Zamiri, Gholamreza Pazuki, Soroush Sardari, and Christopher E. Pearson. 2023. 'Pathogenic CANVAS-Causing but Not Nonpathogenic RFC1 DNA/RNA Repeat Motifs Form Quadruplex or Triplex Structures.' *The Journal of Biological Chemistry* (United States) 299 (10): 105202. https://doi.org/10.1016/j.jbc.2023.105202.

Abejero, Joshua Emmanuel E., Roland Dominic G. Jamora, Theodor S. Vesagas, et al. 2019. 'Long-Term Outcomes of Pallidal Deep Brain Stimulation in X-Linked Dystonia Parkinsonism (XDP): Up to 84 Months Follow-up and Review of Literature'. *Parkinsonism & Related Disorders* 60 (March): 81–86. https://doi.org/10.1016/j.parkreldis.2018.09.022.

Abele, M., K. Bürk, L. Schöls, et al. 2002. 'The Aetiology of Sporadic Adult-Onset Ataxia'. *Brain: A Journal of Neurology* 125 (5): 961–68. https://doi.org/10.1093/brain/awf107.

Abou Chaar, Widad, Anirudh N. Eranki, Hannah A. Stevens, et al. 2024. 'Clinical, Radiological and Pathological Features of a Large American Cohort of Spinocerebellar Ataxia (SCA27B)'. *Annals of Neurology* 96 (6): 1092–103. https://doi.org/10.1002/ana.27060.

Aldous, Sarah G., Edward J. Smith, Christian Landles, et al. 2024. 'A CAG Repeat Threshold for Therapeutics Targeting Somatic Instability in Huntington's Disease'. *Brain: A Journal of Neurology* 147 (5): 1784–98. https://doi.org/10.1093/brain/awae063.

Alkanli, Suleyman Serdar, Nevra Alkanli, Arzu Ay, and Isil Albeniz. 2023. 'CRISPR/Cas9 Mediated Therapeutic Approach in Huntington's Disease'. *Molecular Neurobiology* 60 (3): 1486–98. https://doi.org/10.1007/s12035-022-03150-5.

Al-Mahdawi, Sahar, Heather Ging, Aurelien Bayot, et al. 2018. 'Large Interruptions of GAA Repeat Expansion Mutations in Friedreich Ataxia Are Very Rare.' *Frontiers in Cellular Neuroscience* (Switzerland) 12: 443. https://doi.org/10.3389/fncel.2018.00443.

Alshimemeri, Sohaila, Danah Abo Alsamh, Lily Zhou, et al. 2023. 'Demographics and Clinical Characteristics of Autosomal Dominant Spinocerebellar Ataxia in Canada'. *Movement Disorders Clinical Practice* 10 (3): 440–51. https://doi.org/10.1002/mdc3.13666.

Ando, M., Y. Higuchi, J. Yuan, et al. 2024. 'Clinical Variability Associated with Intronic FGF14 GAA Repeat Expansion in Japan'. *Ann Clin Transl Neurol* 11 (1): 96–104. https://doi.org/10.1002/acn3.51936.

Aneichyk, T., W. T. Hendriks, R. Yadav, et al. 2018. 'Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly'. *Cell* 172 (5): 897-909 e21. https://doi.org/10.1016/j.cell.2018.02.011.

Ashton, Catherine, Elisabetta Indelicato, David Pellerin, et al. 2023. 'Spinocerebellar Ataxia 27B: Episodic Symptoms and Acetazolamide Response in 34 Patients'. *Brain Communications* 5 (5): fcad239. https://doi.org/10.1093/braincomms/fcad239.

Aviner, R., T. T. Lee, V. B. Masto, K. H. Li, R. Andino, and J. Frydman. 2024. 'Polyglutamine-Mediated Ribotoxicity Disrupts Proteostasis and Stress Responses in Huntington's Disease'. *Nat Cell Biol* 26 (6): 892–902. https://doi.org/10.1038/s41556-024-01414-x.

Ballester-Lopez, Alfonsina, Emma Koehorst, Miriam Almendrote, et al. 2020. 'A DM1 Family with Interruptions Associated with Atypical Symptoms and Late Onset but Not with a Milder Phenotype'. *Human Mutation* 41 (2): 420–31. https://doi.org/10.1002/humu.23932.

Banez-Coronel, Monica, and Laura P. W. Ranum. 2019. 'Repeat-Associated Non-AUG (RAN) Translation: Insights from Pathology'. *Laboratory Investigation; a Journal of Technical Methods and Pathology* 99 (7): 929–42. https://doi.org/10.1038/s41374-019-0241-x.

Barbé, Lise, Stella Lanni, Arturo López-Castel, et al. 2017. 'CpG Methylation, a Parent-of-Origin Effect for Maternal-Biased Transmission of Congenital Myotonic Dystrophy.' *American Journal of Human Genetics* (United States) 100 (3): 488–505. https://doi.org/10.1016/j.ajhg.2017.01.033.

Bauer, Peter O. 2016. 'Methylation of C9orf72 Expansion Reduces RNA Foci Formation and Dipeptide-Repeat Proteins Expression in Cells'. *Neuroscience Letters* 612 (January): 204–9. https://doi.org/10.1016/j.neulet.2015.12.018.

Bennett, C. Frank, and Eric E. Swayze. 2010. 'RNA Targeting Therapeutics: Molecular Mechanisms of Antisense Oligonucleotides as a Therapeutic Platform'. *Annual Review of Pharmacology and Toxicology* 50: 259–93. https://doi.org/10.1146/annurev.pharmtox.010909.105654.

Bensimon, Gilbert, Albert Ludolph, Yves Agid, et al. 2009. ‘Riluzole Treatment, Survival and Diagnostic Criteria in Parkinson plus Disorders: The NNIPPS Study’. *Brain: A Journal of Neurology* 132 (Pt 1): 156–71. https://doi.org/10.1093/brain/awn291.

Blitterswijk, Marka van, Mariely DeJesus-Hernandez, Ellis Niemantsverdriet, et al. 2013. ‘Association between Repeat Sizes and Clinical and Pathological Characteristics in Carriers of C9ORF72 Repeat Expansions (Xpansize-72): A Cross-Sectional Cohort Study’. *The Lancet. Neurology* 12 (10): 978–88. https://doi.org/10.1016/S1474-4422(13)70210-2.

Bogaerts, Bert, An Van den Bossche, Bavo Verhaegen, et al. 2024. ‘Closing the Gap: Oxford Nanopore Technologies R10 Sequencing Allows Comparable Results to Illumina Sequencing for SNP-Based Outbreak Investigation of Bacterial Pathogens’. *Journal of Clinical Microbiology* 62 (5): e0157623. https://doi.org/10.1128/jcm.01576-23.

Bonnet, Céline, David Pellerin, Virginie Roth, et al. 2023. ‘Optimized Testing Strategy for the Diagnosis of GAA-FGF14 Ataxia/Spinocerebellar Ataxia 27B’. *Scientific Reports* 13 (1): 9737. https://doi.org/10.1038/s41598-023-36654-8.

Borsche, M., M. Thomsen, D. J. Szmulewicz, et al. 2024. ‘Bilateral Vestibulopathy in RFC1-Positive CANVAS Is Distinctly Different Compared to FGF14-Linked Spinocerebellar Ataxia 27B’. *J Neurol* 271 (2): 1023–27. https://doi.org/10.1007/s00415-023-12050-0.

Bragg, D. Cristopher, Kotchaphorn Mangkalaphiban, Christine A. Vaine, et al. 2017. ‘Disease Onset in X-Linked Dystonia-Parkinsonism Correlates with Expansion of a Hexameric Repeat within an SVA Retrotransposon in TAF1’. *Proceedings of the National Academy of Sciences of the United States of America* 114 (51): E11020–28. https://doi.org/10.1073/pnas.1712526114.

Braida, Claudia, Rhoda K. A. Stefanatos, Berit Adam, et al. 2010. ‘Variant CCG and GGC Repeats within the CTG Expansion Dramatically Modify Mutational Dynamics and Likely Contribute toward Unusual Symptoms in Some Myotonic Dystrophy Type 1 Patients’. *Human Molecular Genetics* 19 (8): 1399–412. https://doi.org/10.1093/hmg/ddq015.

Bzymek, M., and S. T. Lovett. 2001. ‘Instability of Repetitive DNA Sequences: The Role of Replication in Multiple Mechanisms’. *Proceedings of the National Academy of Sciences of the United States of America* 98 (15): 8319–25. https://doi.org/10.1073/pnas.111008398.

Chan, Nelson L. S., Jinzhen Guo, Tianyi Zhang, et al. 2013. ‘Coordinated Processing of 3’ Slipped (CAG)n/(CTG)n Hairpins by DNA Polymerases β and δ Preferentially Induces Repeat Expansions’. *The Journal of Biological Chemistry* 288 (21): 15015–22. https://doi.org/10.1074/jbc.M113.464370.

Chelban, Viorica, David Pellerin, Nirosen Vijiaratnam, et al. 2025. ‘Intronic *FGF14* GAA Repeat Expansions Impact Progression and Survival in Multiple System Atrophy’. *Brain*, ahead of print, April 16. https://doi.org/10.1093/brain/awaf134.

Chen, Zhongbo, Huw R. Morris, James Polke, et al. 2025. ‘Repeat Expansion Disorders’. *Practical Neurology* 25 (3): 204–16. https://doi.org/10.1136/pn-2023-003938.

Chia, Ruth, Anindita Ray, Zalak Shah, et al. 2024. ‘Genome Sequence Analyses Identify Novel Risk Loci for Multiple System Atrophy’. *Neuron* 112 (13): 2142-2156.e5. https://doi.org/10.1016/j.neuron.2024.04.002.

Clarke, James, Hai-Chen Wu, Lakmal Jayasinghe, Alpesh Patel, Stuart Reid, and Hagan Bayley. 2009. ‘Continuous Base Identification for Single-Molecule Nanopore DNA Sequencing’. *Nature Nanotechnology* 4 (4): 265–70. https://doi.org/10.1038/nnano.2009.12.

Cleary, J. D., and C. E. Pearson. 2003. ‘The Contribution of Cis-Elements to Disease-Associated Repeat Instability: Clinical and Experimental Evidence’. *Cytogenet Genome Res* 100 (1–4): 25–55. https://doi.org/10.1159/000072837.

Cortese, Andrea, Stefano Tozza, Wai Yan Yau, et al. 2020. ‘Cerebellar Ataxia, Neuropathy, Vestibular Areflexia Syndrome Due to RFC1 Repeat Expansion’. *Brain: A Journal of Neurology* 143 (2): 480–90. https://doi.org/10.1093/brain/awz418.

Cumming, Sarah A., Mark J. Hamilton, Yvonne Robb, et al. 2018. ‘De Novo Repeat Interruptions Are Associated with Reduced Somatic Instability and Mild or Absent Clinical Features in Myotonic Dystrophy Type 1’. *European Journal of Human Genetics: EJHG* 26 (11): 1635–47. https://doi.org/10.1038/s41431-018-0156-9.

Currò, Riccardo, Alessandro Salvalaggio, Stefano Tozza, et al. 2021. ‘RFC1 Expansions Are a Common Cause of Idiopathic Sensory Neuropathy’. *Brain: A Journal of Neurology* 144 (5): 1542–50. https://doi.org/10.1093/brain/awab072.

Dalski, Andreas, Jassemien Atici, Friedmar R. Kreuz, Yorck Hellenbroich, Eberhard Schwinger, and Christine Zühlke. 2005. 'Mutation Analysis in the Fibroblast Growth Factor 14 Gene: Frameshift Mutation and Polymorphisms in Patients with Inherited Ataxias'. *European Journal of Human Genetics: EJHG* 13 (1): 118–20. https://doi.org/10.1038/sj.ejhg.5201286.

De Boer, Eddy N., Arjen J. Scheper, Dennis Hendriksen, et al. 2025. 'Nanopore Long-Read Sequencing as a First-Tier Diagnostic Test to Detect Repeat Expansions in Neurological Disorders'. *International Journal of Molecular Sciences* 26 (7): 2850. https://doi.org/10.3390/ijms26072850.

De, Tiyasha, Pooja Sharma, Bharathram Upilli, et al. 2024. 'Spinocerebellar Ataxia Type 27B (SCA27B) in India: Insights from a Large Cohort Study Suggest Ancient Origin.' *Neurogenetics* (United States) 25 (4): 393–403. https://doi.org/10.1007/s10048-024-00770-y.

DeJesus-Hernandez, Mariely, Ross A Aleff, Jazmyne L Jackson, et al. 2021. 'Long-Read Targeted Sequencing Uncovers Clinicopathological Associations for*C9orf72*-Linked Diseases'. *Brain* 144 (4): 1082–88. https://doi.org/10.1093/brain/awab006.

Depienne, Christel, and Jean-Louis Mandel. 2021. '30 Years of Repeat Expansion Disorders: What Have We Learned and What Are the Remaining Challenges?' *American Journal of Human Genetics* 108 (5): 764–85. https://doi.org/10.1016/j.ajhg.2021.03.011.

Deshmukh, Amit Laxmikant, Marie-Christine Caron, Mohiuddin Mohiuddin, et al. 2021. 'FAN1 Exo- Not Endo-Nuclease Pausing on Disease-Associated Slipped-DNA Repeats: A Mechanism of Repeat Instability'. *Cell Reports* 37 (10): 110078. https://doi.org/10.1016/j.celrep.2021.110078.

Dolzhenko, Egor, Viraj Deshpande, Felix Schlesinger, et al. 2019. 'ExpansionHunter: A Sequence-Graph-Based Tool to Analyze Variation in Short Tandem Repeat Regions'. *Bioinformatics (Oxford, England)* 35 (22): 4754–56. https://doi.org/10.1093/bioinformatics/btz431.

Domingo, Aloysius, Ana Westenberger, Lillian V. Lee, et al. 2015. 'New Insights into the Genetics of X-Linked Dystonia-Parkinsonism (XDP, DYT3)'. *European Journal of Human Genetics: EJHG* 23 (10): 1334–40. https://doi.org/10.1038/ejhg.2014.292.

Domniz, Noam, Liat Ries-Levavi, Yoram Cohen, et al. 2018. 'Absence of AGG Interruptions Is a Risk Factor for Full Mutation Expansion Among Israeli FMR1 Premutation Carriers'. *Frontiers in Genetics* 9: 606. https://doi.org/10.3389/fgene.2018.00606.

Ebbert, Mark T. W., Stefan L. Farrugia, Jonathon P. Sens, et al. 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). https://doi.org/10.1186/s13024-018-0274-4.

Eid, John, Adrian Fehr, Jeremy Gray, et al. 2009. 'Real-Time DNA Sequencing from Single Polymerase Molecules'. *Science (New York, N.Y.)* 323 (5910): 133–38. https://doi.org/10.1126/science.1162986.

Eisenstein, Michael. 2012. 'Oxford Nanopore Announcement Sets Sequencing Sector Abuzz'. *Nature Biotechnology* 30 (4): 295–96. https://doi.org/10.1038/nbt0412-295.

Evidente, Virgilio Gerald H., Dagmar Nolte, Stephan Niemann, et al. 2004. 'Phenotypic and Molecular Analyses of X-Linked Dystonia-Parkinsonism ("lubag") in Women'. *Archives of Neurology* 61 (12): 1956–59. https://doi.org/10.1001/archneur.61.12.1956.

Findlay Black, H., G. E. B. Wright, J. A. Collins, et al. 2020. 'Frequency of the Loss of CAA Interruption in the HTT CAG Tract and Implications for Huntington Disease in the Reduced Penetrance Range'. *Genet Med* 22 (12): 2108–13. https://doi.org/10.1038/s41436-020-0917-z.

Gall-Duncan, T., N. Sato, R. K. C. Yuen, and C. E. Pearson. 2022. 'Advancing Genomic Technologies and Clinical Awareness Accelerates Discovery of Disease-Associated Tandem Repeat Sequences'. *Genome Res* 32 (1): 1–27. https://doi.org/10.1101/gr.269530.120.

Gaspar, C., M. Jannatipour, P. Dion, et al. 2000. 'CAG Tract of MJD-1 May Be Prone to Frameshifts Causing Polyalanine Accumulation'. *Hum Mol Genet* 9 (13): 1957–66. https://doi.org/10.1093/hmg/9.13.1957.

Gilman, S., G. K. Wenning, P. A. Low, et al. 2008. 'Second Consensus Statement on the Diagnosis of Multiple System Atrophy'. *Neurology* 71 (9): 670–76. https://doi.org/10.1212/01.wnl.0000324625.00404.15.

Gisatulin, Maria, Valerija Dobricic, Christine Zühlke, et al. 2020. 'Clinical Spectrum of the Pentanucleotide Repeat Expansion in the RFC1 Gene in Ataxia Syndromes'. *Neurology* 95 (21): e2912–23. https://doi.org/10.1212/WNL.0000000000010744.
Goh, Yee Yen, Emma Saunders, Samantha Pavey, et al. 2023. 'Multiple System Atrophy'. *Practical Neurology* 23 (3): 208–21. https://doi.org/10.1136/pn-2020-002797.
Gomez-Pastor, Rocio, Eileen T. Burchfiel, Daniel W. Neef, et al. 2017. 'Abnormal Degradation of the Neuronal Stress-Protective Transcription Factor HSF1 in Huntington's Disease'. *Nature Communications* 8 (February): 14405. https://doi.org/10.1038/ncomms14405.
Graham, J. G., and D. R. Oppenheimer. 1969. 'Orthostatic Hypotension and Nicotine Sensitivity in a Case of Multiple System Atrophy'. *Journal of Neurology, Neurosurgery, and Psychiatry* 32 (1): 28–34. https://doi.org/10.1136/jnnp.32.1.28.
Grandi, Fiorella C., and Wenfeng An. 2013. 'Non-LTR Retrotransposons and Microsatellites: Partners in Genomic Variation'. *Mobile Genetic Elements* 3 (4): e25674. https://doi.org/10.4161/mge.25674.
Groh, Matthias, Lara Marques Silva, and Natalia Gromak. 2014. 'Mechanisms of Transcriptional Dysregulation in Repeat Expansion Disorders'. *Biochemical Society Transactions* 42 (4): 1123–28. https://doi.org/10.1042/BST20140049.
Gu, XiaoJing, YongPing Chen, QingQing Zhou, et al. 2018. 'Analysis of GWAS-Linked Variants in Multiple System Atrophy'. *Neurobiology of Aging* 67 (July): 201.e1-201.e4. https://doi.org/10.1016/j.neurobiolaging.2018.03.018.
Gymrek, Melissa. 2017. 'A Genomic View of Short Tandem Repeats'. *Current Opinion in Genetics & Development* 44 (June): 9–16. https://doi.org/10.1016/j.gde.2017.01.012.
Haenfler, Jill M., Geena Skariah, Caitlin M. Rodriguez, et al. 2018. 'Targeted Reactivation of FMR1 Transcription in Fragile X Syndrome Embryonic Stem Cells'. *Frontiers in Molecular Neuroscience* 11: 282. https://doi.org/10.3389/fnmol.2018.00282.
Hagerman, Katharine A., Haihe Ruan, Kerrie Nichol Edamura, Tohru Matsuura, Christopher E. Pearson, and Yuh-Hwa Wang. 2009. 'The ATTCT Repeats of Spinocerebellar Ataxia Type 10 Display Strong Nucleosome Assembly Which Is Enhanced by Repeat Interruptions.' *Gene* (Netherlands) 434 (1–2): 29–34. https://doi.org/10.1016/j.gene.2008.12.011.
Handsaker, Robert E., Seva Kashin, Nora M. Reed, et al. 2025. 'Long Somatic DNA-Repeat Expansion Drives Neurodegeneration in Huntington's Disease'. *Cell* 188 (3): 623-639.e19. https://doi.org/10.1016/j.cell.2024.11.038.
Harris, R. S., M. Cechova, and K. D. Makova. 2019. 'Noise-Cancelling Repeat Finder: Uncovering Tandem Repeats in Error-Prone Long-Read Sequencing Data'. *Bioinformatics* 35 (22): 4809–11. https://doi.org/10.1093/bioinformatics/btz484.
Hasan, Ali, Gabriel Vasata Furtado, Elaine Miglorini, et al. 2025. 'The Impact of Interrupted ATXN10 Expansions on Clinical Findings of Spinocerebellar Ataxia Type 10'. *Journal of Neurology* 272 (4). https://doi.org/10.1007/s00415-025-13003-5.
Hengel, H., D. Pellerin, C. Wilke, et al. 2023. 'As Frequent as Polyglutamine Spinocerebellar Ataxias: SCA27B in a Large German Autosomal Dominant Ataxia Cohort'. *Mov Disord* 38 (8): 1557–58. https://doi.org/10.1002/mds.29559.
Hisey, Julia A., Elina A. Radchenko, Nicholas H. Mandel, et al. 2024. 'Pathogenic CANVAS (AAGGG)n Repeats Stall DNA Replication Due to the Formation of Alternative DNA Structures.' *Nucleic Acids Research* (England) 52 (8): 4361–74. https://doi.org/10.1093/nar/gkae124.
Iruzubieta, P., D. Pellerin, A. Bergareche, et al. 2023. 'Frequency and Phenotypic Spectrum of Spinocerebellar Ataxia 27B and Other Genetic Ataxias in a Spanish Cohort of Late-Onset Cerebellar Ataxia'. *Eur J Neurol* 30 (12): 3828–33. https://doi.org/10.1111/ene.16039.
Jabbari, Edwin, Negin Holland, Viorica Chelban, et al. 2020. 'Diagnosis Across the Spectrum of Progressive Supranuclear Palsy and Corticobasal Syndrome'. *JAMA Neurology* 77 (3): 377–87. https://doi.org/10.1001/jamaneurol.2019.4347.
Jacobi, Heike, Sophie Tezenas du Montcel, Peter Bauer, et al. 2015. 'Long-Term Disease Progression in Spinocerebellar Ataxia Types 1, 2, 3, and 6: A Longitudinal Cohort Study'. *The Lancet. Neurology* 14 (11): 1101–8. https://doi.org/10.1016/S1474-4422(15)00202-1.
Jamora, Roland Dominic G., Cid Czarina E. Diesta, Paul Matthew D. Pasco, and Lillian V. Lee. 2011. 'Oral Pharmacological Treatment of X-Linked Dystonia Parkinsonism: Successes and

Failures'. *The International Journal of Neuroscience* 121 Suppl 1: 18–21. https://doi.org/10.3109/00207454.2010.544433.
Jankovic, Joseph. 2006. 'Treatment of Dystonia'. *The Lancet. Neurology* 5 (10): 864–72. https://doi.org/10.1016/S1474-4422(06)70574-9.
Jellinger, Kurt A., Klaus Seppi, and Gregor K. Wenning. 2005. 'Grading of Neuropathology in Multiple System Atrophy: Proposal for a Novel Scale'. *Movement Disorders: Official Journal of the Movement Disorder Society* 20 Suppl 12 (August): S29-36. https://doi.org/10.1002/mds.20537.
Jeon, Beom S., Matt J. Farrer, Stephanie F. Bortnick, and Korean Canadian Alliance on Parkinson's Disease and Related Disorders. 2014. 'Mutant COQ2 in Multiple-System Atrophy'. *The New England Journal of Medicine* 371 (1): 80. https://doi.org/10.1056/NEJMc1311763.
Jeon, Sungwon, Youngjune Bhak, Yeonsong Choi, et al. 2020. 'Korean Genome Project: 1094 Korean Personal Genomes with Clinical Information'. *Science Advances* 6 (22): eaaz7835. https://doi.org/10.1126/sciadv.aaz7835.
Kasianowicz, J. J., E. Brandin, D. Branton, and D. W. Deamer. 1996. 'Characterization of Individual Polynucleotide Molecules Using a Membrane Channel'. *Proceedings of the National Academy of Sciences of the United States of America* 93 (24): 13770–73. https://doi.org/10.1073/pnas.93.24.13770.
Kasten, Meike, Johann Hagenah, Julia Graf, et al. 2013. 'Cohort Profile: A Population-Based Cohort to Study Non-Motor Symptoms in Parkinsonism (EPIPARK).' *International Journal of Epidemiology* (England) 42 (1): 128–128k. https://doi.org/10.1093/ije/dys202.
Kawarai, Toshitaka, Paul Matthew D. Pasco, Rosalia A. Teleg, et al. 2013. 'Application of Long-Range Polymerase Chain Reaction in the Diagnosis of X-Linked Dystonia-Parkinsonism'. *Neurogenetics* 14 (2): 167–69. https://doi.org/10.1007/s10048-013-0357-x.
Kim, Han-Joon, Beom S. Jeon, Jee-Young Lee, and Ji Young Yun. 2011. 'Survival of Korean Patients with Multiple System Atrophy'. *Movement Disorders: Official Journal of the Movement Disorder Society* 26 (5): 909–12. https://doi.org/10.1002/mds.23580.
Koga, Shunsuke, Naoya Aoki, Ryan J. Uitti, et al. 2015. 'When DLB, PD, and PSP Masquerade as MSA: An Autopsy Study of 134 Patients'. *Neurology* 85 (5): 404–12. https://doi.org/10.1212/WNL.0000000000001807.
Kozlowski, Piotr, Mateusz de Mezer, and Wlodzimierz J. Krzyzosiak. 2010. 'Trinucleotide Repeats in Human Genome and Exome'. *Nucleic Acids Research* 38 (12): 4027–39. https://doi.org/10.1093/nar/gkq127.
Krause, Christin, Helen Sievert, Cathleen Geißler, et al. 2019. 'Critical Evaluation of the DNA-Methylation Markers ABCG1 and SREBF1 for Type 2 Diabetes Stratification.' *Epigenomics* (England) 11 (8): 885–97. https://doi.org/10.2217/epi-2018-0159.
Kudo, Kenta, Karin Hori, Sefan Asamitsu, et al. 2024. 'Structural Polymorphism of the Nucleic Acids in Pentanucleotide Repeats Associated with the Neurological Disorder CANVAS.' *The Journal of Biological Chemistry* (United States) 300 (4): 107138. https://doi.org/10.1016/j.jbc.2024.107138.
Laabs, Björn-Hergen, Christine Klein, Jelena Pozojevic, et al. 2021. 'Identifying Genetic Modifiers of Age-Associated Penetrance in X-Linked Dystonia-Parkinsonism'. *Nature Communications* 12 (1): 3216. https://doi.org/10.1038/s41467-021-23491-4.
Landrum, Melissa J., Jennifer M. Lee, Mark Benson, et al. 2018. 'ClinVar: Improving Access to Variant Interpretations and Supporting Evidence'. *Nucleic Acids Research* 46 (D1): D1062–67. https://doi.org/10.1093/nar/gkx1153.
Laß, Joshua, Michele Berselli, Doug Rioux, et al. 2025. 'Genetic Testing for SCA27B in Korean Multiple System Atrophy'. *Brain: A Journal of Neurology*, July 17, awaf263. https://doi.org/10.1093/brain/awaf263.
Laß, Joshua, Theresa Lüth, Kathleen Schlüter, et al. 2024. 'Stability of Mosaic Divergent Repeat Interruptions in X-Linked Dystonia-Parkinsonism'. *Movement Disorders: Official Journal of the Movement Disorder Society* 39 (7): 1145–53. https://doi.org/10.1002/mds.29809.
Laß, Joshua, Mirja Thomsen, Max Borsche, et al. 2025. '*FGF14* Repeat Length and Mosaic Interruptions: Modifiers of Spinocerebellar Ataxia 27B?' *Brain*, ahead of print, May 17. https://doi.org/10.1093/brain/awaf183.

Lee, L. V., E. L. Munoz, K. T. Tan, and M. T. Reyes. 2001. 'Sex Linked Recessive Dystonia Parkinsonism of Panay, Philippines (XDP)'. *Molecular Pathology: MP* 54 (6): 362–68.
Lee, L. V., F. M. Pascasio, F. D. Fuentes, and G. H. Viterbo. 1976. 'Torsion Dystonia in Panay, Philippines'. *Advances in Neurology* 14: 137–51.
Lee, Lillian V., Elma Maranon, Cynthia Demaisip, et al. 2002. 'The Natural History of Sex-Linked Recessive Dystonia Parkinsonism of Panay, Philippines (XDP)'. *Parkinsonism & Related Disorders* 9 (1): 29–38. https://doi.org/10.1016/s1353-8020(02)00042-1.
Lee, Lillian V., Corazon Rivera, Rosalia A. Teleg, et al. 2011. 'The Unique Phenomenology of Sex-Linked Dystonia Parkinsonism (XDP, DYT3, "Lubag")'. *The International Journal of Neuroscience* 121 Suppl 1: 3–11. https://doi.org/10.3109/00207454.2010.526728.
Leitão, E., C. Schröder, and C. Depienne. 2024. 'Identification and Characterization of Repeat Expansions in Neurological Disorders: Methodologies, Tools, and Strategies'. *Revue Neurologique* 180 (5): 383–92. https://doi.org/10.1016/j.neurol.2024.03.005.
Levene, M. J., J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead, and W. W. Webb. 2003. 'Zero-Mode Waveguides for Single-Molecule Analysis at High Concentrations'. *Science (New York, N.Y.)* 299 (5607): 682–86. https://doi.org/10.1126/science.1079700.
Li, H. 2018. 'Minimap2: Pairwise Alignment for Nucleotide Sequences'. *Bioinformatics* 34 (18): 3094–100. https://doi.org/10.1093/bioinformatics/bty191.
Li, H., B. Handsaker, A. Wysoker, et al. 2009. 'The Sequence Alignment/Map Format and SAMtools'. *Bioinformatics* 25 (16): 2078–79. https://doi.org/10.1093/bioinformatics/btp352.
Libby, Randell T., Katharine A. Hagerman, Victor V. Pineda, et al. 2008. 'CTCF Cis-Regulates Trinucleotide Repeat Instability in an Epigenetic Manner: A Novel Basis for Mutational Hot Spot Determination.' *PLoS Genetics* (United States) 4 (11): e1000257. https://doi.org/10.1371/journal.pgen.1000257.
Libby, Randell T., Darren G. Monckton, Ying-Hui Fu, et al. 2003. 'Genomic Context Drives SCA7 CAG Repeat Instability, While Expressed SCA7 cDNAs Are Intergenerationally and Somatically Stable in Transgenic Mice.' *Human Molecular Genetics* (England) 12 (1): 41–50. https://doi.org/10.1093/hmg/ddg006.
Lin, David J., Katherine L. Hermann, and Jeremy D. Schmahmann. 2014. 'Multiple System Atrophy of the Cerebellar Type: Clinical State of the Art'. *Movement Disorders: Official Journal of the Movement Disorder Society* 29 (3): 294–304. https://doi.org/10.1002/mds.25847.
Liu, Elaine Y., Jenny Russ, Kathryn Wu, et al. 2014. 'C9orf72 Hypermethylation Protects against Repeat Expansion-Associated Pathology in ALS/FTD'. *Acta Neuropathologica* 128 (4): 525–41. https://doi.org/10.1007/s00401-014-1286-y.
Lokanga, Rachel Adihe, Xiao-Nan Zhao, and Karen Usdin. 2014. 'The Mismatch Repair Protein MSH2 Is Rate Limiting for Repeat Expansion in a Fragile X Premutation Mouse Model'. *Human Mutation* 35 (1): 129–36. https://doi.org/10.1002/humu.22464.
Lüth, T., J. Laß, S. Schaake, et al. 2022. 'Elucidating Hexanucleotide Repeat Number and Methylation within the X-Linked Dystonia-Parkinsonism (XDP)-Related SVA Retrotransposon in TAF1 with Nanopore Sequencing'. *Genes (Basel)* 13 (1). https://doi.org/10.3390/genes13010126.
Lyasota, Oksana, Anna Dorohova, Jose Luis Hernandez-Caceres, et al. 2024. 'Stability of the CAG Tract in the ATXN2 Gene Depends on the Localization of CAA Interruptions'. *Biomedicines* 12 (8): 1648. https://doi.org/10.3390/biomedicines12081648.
Maarel, Silvère M. van der, and Rune R. Frants. 2005. 'The D4Z4 Repeat-Mediated Pathogenesis of Facioscapulohumeral Muscular Dystrophy'. *American Journal of Human Genetics* 76 (3): 375–86. https://doi.org/10.1086/428361.
Makino, Satoshi, Ryuji Kaji, Satoshi Ando, et al. 2007. 'Reduced Neuron-Specific Expression of the TAF1 Gene Is Associated with X-Linked Dystonia-Parkinsonism'. *American Journal of Human Genetics* 80 (3): 393–406. https://doi.org/10.1086/512129.
Martier, Raygene, Jolanda M. Liefhebber, Ana García-Osta, et al. 2019. 'Targeting RNA-Mediated Toxicity in C9orf72 ALS and/or FTD by RNAi-Based Gene Therapy'. *Molecular Therapy. Nucleic Acids* 16 (June): 26–37. https://doi.org/10.1016/j.omtn.2019.02.001.
Matsushima, Masaaki, Hiroaki Yaguchi, Eriko Koshimizu, et al. 2024. 'FGF14 GAA Repeat Expansion and ZFHX3 GGC Repeat Expansion in Clinically Diagnosed Multiple System Atrophy Patients'. *Journal of Neurology* 271 (6): 3643–47. https://doi.org/10.1007/s00415-024-12308-1.

Matsuura, T., P. Fang, C. E. Pearson, et al. 2006. 'Interruptions in the Expanded ATTCT Repeat of Spinocerebellar Ataxia Type 10: Repeat Purity as a Disease Modifier?' *Am J Hum Genet* 78 (1): 125–29. https://doi.org/10.1086/498654.

Matsuyama, Z., Y. Izumi, M. Kameyama, H. Kawakami, and S. Nakamura. 1999. 'The Effect of CAT Trinucleotide Interruptions on the Age at Onset of Spinocerebellar Ataxia Type 1 (SCA1)'. *Journal of Medical Genetics* 36 (7): 546–48.

McFarland, Karen N., Jilin Liu, Ivette Landrian, et al. 2013. 'Paradoxical Effects of Repeat Interruptions on Spinocerebellar Ataxia Type 10 Expansions and Repeat Instability'. *European Journal of Human Genetics: EJHG* 21 (11): 1272–76. https://doi.org/10.1038/ejhg.2013.32.

Méreaux, Jean-Loup, Claire-Sophie Davoine, David Pellerin, et al. 2024. 'Clinical and Genetic Keys to Cerebellar Ataxia Due to FGF14 GAA Expansions'. *EBioMedicine* 99 (January): 104931. https://doi.org/10.1016/j.ebiom.2023.104931.

Miki, Yasuo, Eiki Tsushima, Sandrine C. Foti, et al. 2021. 'Identification of Multiple System Atrophy Mimicking Parkinson's Disease or Progressive Supranuclear Palsy'. *Brain: A Journal of Neurology* 144 (4): 1138–51. https://doi.org/10.1093/brain/awab017.

Milovanovic, A., N. Dragasevic-Miskovic, M. Thomsen, et al. 2024. 'RFC1 and FGF14 Repeat Expansions in Serbian Patients with Cerebellar Ataxia'. *Mov Disord Clin Pract* 11 (6): 626–33. https://doi.org/10.1002/mdc3.14020.

Mirceta, Mila, Monika H. M. Schmidt, Natalie Shum, et al. 2024. 'C9orf72 Repeat Expansion Creates the Unstable Folate-Sensitive Fragile Site FRA9A.' *NAR Molecular Medicine* (England) 1 (4): ugae019. https://doi.org/10.1093/narmme/ugae019.

Miyatake, Satoko, Hiroshi Doi, Hiroaki Yaguchi, et al. 2024. 'Complete Nanopore Repeat Sequencing of SCA27B (GAA-FGF14 Ataxia) in Japanese.' *Journal of Neurology, Neurosurgery, and Psychiatry* (England) 95 (12): 1187–95. https://doi.org/10.1136/jnnp-2024-333541.

Moeller, Ashley A., Marcia V. Felker, Jennifer A. Brault, Laura C. Duncan, Rizwan Hamid, and Meredith R. Golomb. 2021. 'Patients With Extreme Early Onset Juvenile Huntington Disease Can Have Delays in Diagnosis: A Case Report and Literature Review'. *Child Neurology Open* 8: 2329048X211036137. https://doi.org/10.1177/2329048X211036137.

Mohren, L., F. Erdlenbruch, E. Leitao, et al. 2024. 'Identification and Characterisation of Pathogenic and Non-Pathogenic FGF14 Repeat Expansions'. *Nat Commun* 15 (1): 7665. https://doi.org/10.1038/s41467-024-52148-1.

Morales, Fernando, Melissa Vásquez, Eyleen Corrales, et al. 2020. 'Longitudinal Increases in Somatic Mosaicism of the Expanded CTG Repeat in Myotonic Dystrophy Type 1 Are Associated with Variation in Age-at-Onset'. *Human Molecular Genetics* 29 (15): 2496–507. https://doi.org/10.1093/hmg/ddaa123.

Multiple-System Atrophy Research Collaboration. 2013. 'Mutations in COQ2 in Familial and Sporadic Multiple-System Atrophy'. *The New England Journal of Medicine* 369 (3): 233–44. https://doi.org/10.1056/NEJMoa1212115.

Mulvihill, David J., Kerrie Nichol Edamura, Katharine A. Hagerman, Christopher E. Pearson, and Yuh-Hwa Wang. 2005. 'Effect of CAT or AGG Interruptions and CpG Methylation on Nucleosome Assembly upon Trinucleotide Repeats on Spinocerebellar Ataxia, Type 1 and Fragile X Syndrome.' *The Journal of Biological Chemistry* (United States) 280 (6): 4498–503. https://doi.org/10.1074/jbc.M413239200.

Nakahara, Yasuo, Jun Mitsui, Hidetoshi Date, et al. 2023. 'Genome-Wide Association Study Identifies a New Susceptibility Locus in PLA2G4C for Multiple System Atrophy'. *medRxiv: The Preprint Server for Health Sciences*, May 2, 2023.05.02.23289328. https://doi.org/10.1101/2023.05.02.23289328.

Nethisinghe, S., M. Kesavan, H. Ging, et al. 2021. 'Interruptions of the FXN GAA Repeat Tract Delay the Age at Onset of Friedreich's Ataxia in a Location Dependent Manner'. *Int J Mol Sci* 22 (14). https://doi.org/10.3390/ijms22147507.

Netravathi, M., Pramod Kumar Pal, Meera Purushottam, Kandavel Thennarasu, Mitali Mukherjee, and Sanjeev Jain. 2009. 'Spinocerebellar Ataxias Types 1, 2 and 3: Age Adjusted Clinical Severity of Disease at Presentation Correlates with Size of CAG Repeat Lengths'. *Journal of the Neurological Sciences* 277 (1–2): 83–86. https://doi.org/10.1016/j.jns.2008.10.016.

Nolin, Sarah L., Anne Glicksman, Nicole Tortora, et al. 2019. 'Expansions and Contractions of the FMR1 CGG Repeat in 5,508 Transmissions of Normal, Intermediate, and Premutation

Alleles'. *American Journal of Medical Genetics. Part A* 179 (7): 1148–56. https://doi.org/10.1002/ajmg.a.61165.
Ohshima, K., L. Montermini, R. D. Wells, and M. Pandolfo. 1998. 'Inhibitory Effects of Expanded GAA.TTC Triplet Repeats from Intron I of the Friedreich Ataxia Gene on Transcription and Replication in Vivo'. *J Biol Chem* 273 (23): 14588–95. https://doi.org/10.1074/jbc.273.23.14588.
Ouyang, R., L. Wan, D. Pellerin, et al. 2024. 'The Genetic Landscape and Phenotypic Spectrum of GAA-FGF14 Ataxia in China: A Large Cohort Study'. *EBioMedicine* 102 (April): 105077. https://doi.org/10.1016/j.ebiom.2024.105077.
Papp, M. I., J. E. Kahn, and P. L. Lantos. 1989. 'Glial Cytoplasmic Inclusions in the CNS of Patients with Multiple System Atrophy (Striatonigral Degeneration, Olivopontocerebellar Atrophy and Shy-Drager Syndrome)'. *Journal of the Neurological Sciences* 94 (1–3): 79–100. https://doi.org/10.1016/0022-510x(89)90219-0.
Parsons, M. A., R. R. Sinden, and M. G. Izban. 1998. 'Transcriptional Properties of RNA Polymerase II within Triplet Repeat-Containing DNA from the Human Myotonic Dystrophy and Fragile X Loci'. *J Biol Chem* 273 (41): 26998–7008. https://doi.org/10.1074/jbc.273.41.26998.
Pascarella, Giovanni, Chung Chau Hon, Kosuke Hashimoto, et al. 2022. 'Recombination of Repeat Elements Generates Somatic Complexity in Human Genomes'. *Cell* 185 (16): 3025-3040.e6. https://doi.org/10.1016/j.cell.2022.06.032.
Pascual-Gilabert, Marta, Arturo López-Castel, and Ruben Artero. 2021. 'Myotonic Dystrophy Type 1 Drug Development: A Pipeline toward the Market'. *Drug Discovery Today* 26 (7): 1765–72. https://doi.org/10.1016/j.drudis.2021.03.024.
Paulson, H. 2018. 'Repeat Expansion Diseases'. *Handb Clin Neurol* 147: 105–23. https://doi.org/10.1016/B978-0-444-63233-3.00009-9.
Pauly, Martje G., Marta Ruiz López, Ana Westenberger, et al. 2020. 'Expanding Data Collection for the MDSGene Database: X-Linked Dystonia-Parkinsonism as Use Case Example'. *Movement Disorders: Official Journal of the Movement Disorder Society* 35 (11): 1933–38. https://doi.org/10.1002/mds.28289.
Pearson, C. E., E. E. Eichler, D. Lorenzetti, et al. 1998. 'Interruptions in the Triplet Repeats of SCA1 and FRAXA Reduce the Propensity and Complexity of Slipped Strand DNA (S-DNA) Formation'. *Biochemistry* 37 (8): 2701–8. https://doi.org/10.1021/bi972546c.
Pellerin, D., M. C. Danzi, C. Wilke, et al. 2023. 'Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia'. *N Engl J Med* 388 (2): 128–41. https://doi.org/10.1056/NEJMoa2207406.
Pellerin, D., M. Danzi, M. Renaud, et al. 1993. 'GAA-FGF14-Related Ataxia'. In *GeneReviews((R))*, edited by M. P. Adam, J. Feldman, G. M. Mirzaa, R. A. Pagon, S. E. Wallace, and A. Amemiya. Seattle (WA).
Pellerin, D., G. F. Del Gobbo, M. Couse, et al. 2024. 'A Common Flanking Variant Is Associated with Enhanced Stability of the FGF14-SCA27B Repeat Locus'. *Nat Genet* 56 (7): 1366–70. https://doi.org/10.1038/s41588-024-01808-5.
Pellerin, D., P. Iruzubieta, S. Tekgul, et al. 2023. 'Non-GAA Repeat Expansions in FGF14 Are Likely Not Pathogenic-Reply to: "Shaking Up Ataxia: FGF14 and RFC1 Repeat Expansions in Affected and Unaffected Members of a Chilean Family"'. *Mov Disord* 38 (8): 1575–77. https://doi.org/10.1002/mds.29552.
Pellerin, David, Matt C. Danzi, Mathilde Renaud, et al. 2024. 'Spinocerebellar Ataxia 27B: A Novel, Frequent and Potentially Treatable Ataxia'. *Clinical and Translational Medicine* 14 (1): e1504. https://doi.org/10.1002/ctm2.1504.
Pellerin, David, Matt C. Danzi, Carlo Wilke, et al. 2023. 'Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia'. *New England Journal of Medicine* 388 (2): 128–41. https://doi.org/10.1056/NEJMoa2207406.
Pellerin, David, Felix Heindl, Carlo Wilke, et al. 2024. 'GAA-FGF14 Disease: Defining Its Frequency, Molecular Basis, and 4-Aminopyridine Response in a Large Downbeat Nystagmus Cohort'. *EBioMedicine* 102 (April): 105076. https://doi.org/10.1016/j.ebiom.2024.105076.
Pellerin, David, Jean-Loup Méreaux, Susana Boluda, et al. 2025. 'Somatic Instability of the FGF14-SCA27B GAA•TTC Repeat Reveals a Marked Expansion Bias in the Cerebellum'. *Brain: A Journal of Neurology* 148 (4): 1258–70. https://doi.org/10.1093/brain/awae312.

Pellerin, David, Carlo Wilke, Andreas Traschütz, et al. 2024. 'Intronic FGF14 GAA Repeat Expansions Are a Common Cause of Ataxia Syndromes with Neuropathy and Bilateral Vestibulopathy'. *Journal of Neurology, Neurosurgery, and Psychiatry* 95 (2): 175–79. https://doi.org/10.1136/jnnp-2023-331490.

Peric, Stojan, Jovan Pesovic, Dusanka Savic-Pavicevic, Vidosava Rakocevic Stojanovic, and Giovanni Meola. 2021. 'Molecular and Clinical Implications of Variant Repeats in Myotonic Dystrophy Type 1'. *International Journal of Molecular Sciences* 23 (1): 354. https://doi.org/10.3390/ijms23010354.

Pollard, Martin O., Deepti Gurdasani, Alexander J. Mentzer, Tarryn Porter, and Manjinder S. Sandhu. 2018. 'Long Reads: Their Purpose and Place'. *Human Molecular Genetics* 27 (R2): R234–41. https://doi.org/10.1093/hmg/ddy177.

Posey, Jennifer E., Anne H. O'Donnell-Luria, Jessica X. Chong, et al. 2019. 'Insights into Genetics, Human Biology and Disease Gleaned from Family Based Genomic Studies'. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 21 (4): 798–812. https://doi.org/10.1038/s41436-018-0408-7.

Rafehi, H., J. Read, D. J. Szmulewicz, et al. 2023. 'An Intronic GAA Repeat Expansion in FGF14 Causes the Autosomal-Dominant Adult-Onset Ataxia SCA50/ATX-FGF14'. *Am J Hum Genet* 110 (1): 105–19. https://doi.org/10.1016/j.ajhg.2022.11.015.

Rafehi, Haloom, Justin Read, David J. Szmulewicz, et al. 2023. 'An Intronic GAA Repeat Expansion in FGF14 Causes the Autosomal-Dominant Adult-Onset Ataxia SCA50/ATX-FGF14'. *The American Journal of Human Genetics* 110 (1): 105–19. https://doi.org/10.1016/J.AJHG.2022.11.015.

Rajagopal, Sangeerthana, Jasmine Donaldson, Michael Flower, Davina J. Hensman Moss, and Sarah J. Tabrizi. 2023. 'Genetic Modifiers of Repeat Expansion Disorders'. *Emerging Topics in Life Sciences* 7 (3): 325–37. https://doi.org/10.1042/ETLS20230015.

Rajan-Babu, Indhu-Shree, Egor Dolzhenko, Michael A. Eberle, and Jan M. Friedman. 2024. 'Sequence Composition Changes in Short Tandem Repeats: Heterogeneity, Detection, Mechanisms and Clinical Implications.' *Nature Reviews. Genetics* (England) 25 (7): 476–99. https://doi.org/10.1038/s41576-024-00696-z.

Rakovic, A., A. Domingo, K. Grutz, et al. 2018. 'Genome Editing in Induced Pluripotent Stem Cells Rescues TAF1 Levels in X-Linked Dystonia-Parkinsonism'. *Mov Disord* 33 (7): 1108–18. https://doi.org/10.1002/mds.27441.

Rasmussen, Astrid, Mathis Hildonen, John Vissing, Morten Duno, Zeynep Tümer, and Ulf Birkedal. 2022. 'High Resolution Analysis of DMPK Hypermethylation and Repeat Interruptions in Myotonic Dystrophy Type 1'. *Genes* 13 (6): 970. https://doi.org/10.3390/genes13060970.

Reyes, Charles Jourdan, Björn-Hergen Laabs, Susen Schaake, et al. 2021. 'Brain Regional Differences in Hexanucleotide Repeat Length in X-Linked Dystonia-Parkinsonism Using Nanopore Sequencing'. *Neurology. Genetics* 7 (4): e608. https://doi.org/10.1212/NXG.0000000000000608.

Rhoads, Anthony, and Kin Fai Au. 2015. 'PacBio Sequencing and Its Applications'. *Genomics, Proteomics & Bioinformatics* 13 (5): 278–89. https://doi.org/10.1016/j.gpb.2015.08.002.

Rodrigues, Filipe B., Lori Quinn, and Edward J. Wild. 2019. 'Huntington's Disease Clinical Trials Corner: January 2019'. *Journal of Huntington's Disease* 8 (1): 115–25. https://doi.org/10.3233/jhd-190001.

Rohilla, Kushal J., and Keith T. Gagnon. 2017. 'RNA Biology of Disease-Associated Microsatellite Repeat Expansions'. *Acta Neuropathologica Communications* 5 (1): 63. https://doi.org/10.1186/s40478-017-0468-y.

Rolfsmeier, M. L., M. J. Dixon, and R. S. Lahue. 2000. 'Mismatch Repair Blocks Expansions of Interrupted Trinucleotide Repeats in Yeast'. *Molecular Cell* 6 (6): 1501–7. https://doi.org/10.1016/s1097-2765(00)00146-5.

Rosales, Raymond L. 2010. 'X-Linked Dystonia Parkinsonism: Clinical Phenotype, Genetics and Therapeutics'. *Journal of Movement Disorders* 3 (2): 32–38. https://doi.org/10.14802/jmd.10009.

Saffie Awad, P., K. Lohmann, Y. Hirmas, et al. 2023. 'Shaking Up Ataxia: FGF14 and RFC1 Repeat Expansions in Affected and Unaffected Members of a Chilean Family'. *Mov Disord* 38 (6): 1107–9. https://doi.org/10.1002/mds.29390.

Sailer, Anna, Sonja W. Scholz, Michael A. Nalls, et al. 2016. 'A Genome-Wide Association Study in Multiple System Atrophy'. *Neurology* 87 (15): 1591–98. https://doi.org/10.1212/WNL.0000000000003221.

Sakamoto, N., P. D. Chastain, P. Parniewski, et al. 1999. 'Sticky DNA: Self-Association Properties of Long GAA.TTC Repeats in R.R.Y Triplex Structures from Friedreich's Ataxia'. *Molecular Cell* 3 (4): 465–75. https://doi.org/10.1016/s1097-2765(00)80474-8.

Sakamoto, N., J. E. Larson, R. R. Iyer, L. Montermini, M. Pandolfo, and R. D. Wells. 2001. 'GGA*TCC-Interrupted Triplets in Long GAA*TTC Repeats Inhibit the Formation of Triplex and Sticky DNA Structures, Alleviate Transcription Inhibition, and Reduce Genetic Instabilities'. *The Journal of Biological Chemistry* 276 (29): 27178–87. https://doi.org/10.1074/jbc.M101852200.

Satolli, Sara, Salvatore Rossi, Elisa Vegezzi, et al. 2024. 'Correction to: Spinocerebellar Ataxia 27B: A Frequent and Slowly Progressive Autosomal-Dominant Cerebellar Ataxia-Experience from an Italian Cohort'. *Journal of Neurology* 271 (12): 7650–51. https://doi.org/10.1007/s00415-024-12629-1.

Seixas, Ana I., Joana R. Loureiro, Cristina Costa, et al. 2017. 'A Pentanucleotide ATTTC Repeat Insertion in the Non-Coding Region of DAB1, Mapping to SCA37, Causes Spinocerebellar Ataxia'. *American Journal of Human Genetics* 101 (1): 87–103. https://doi.org/10.1016/j.ajhg.2017.06.007.

Shen, Yang-I., Kai-Chun Cheng, Yu-Jie Wei, and I.-Ren Lee. 2024. 'Structural Dynamics Role of AGG Interruptions in Inhibition CGG Repeat Expansion Associated with Fragile X Syndrome'. *ACS Chemical Neuroscience* 15 (2): 230–35. https://doi.org/10.1021/acschemneuro.3c00712.

Shin, Jung Hwan, Han-Joon Kim, Chan Young Lee, Hee Jin Chang, Kyung Ah Woo, and Beomseok Jeon. 2022. 'Laboratory Prognostic Factors for the Long-Term Survival of Multiple System Atrophy'. *NPJ Parkinson's Disease* 8 (1): 141. https://doi.org/10.1038/s41531-022-00413-9.

Shirai, Shinichi, Keiichi Mizushima, Keishi Fujiwara, et al. 2023. 'Case Series: Downbeat Nystagmus in SCA27B'. *Journal of the Neurological Sciences* 454 (November): 120849. https://doi.org/10.1016/j.jns.2023.120849.

Singh, Krishna, Sakshi Shukla, Uma Shankar, et al. 2024. 'Elucidating the Pathobiology of Cerebellar Ataxia with Neuropathy and Vestibular Areflexia Syndrome (CANVAS) with Its Expanded RNA Structure Formation and Proteinopathy.' *Scientific Reports* (England) 14 (1): 28054. https://doi.org/10.1038/s41598-024-78947-6.

Sone, Jun, Satomi Mitsuhashi, Atsushi Fujita, et al. 2019. 'Long-Read Sequencing Identifies GGC Repeat Expansions in NOTCH2NLC Associated with Neuronal Intranuclear Inclusion Disease'. *Nature Genetics* 51 (8): 1215–21. https://doi.org/10.1038/s41588-019-0459-y.

Spillantini, M. G., R. A. Crowther, R. Jakes, N. J. Cairns, P. L. Lantos, and M. Goedert. 1998. 'Filamentous Alpha-Synuclein Inclusions Link Multiple System Atrophy with Parkinson's Disease and Dementia with Lewy Bodies'. *Neuroscience Letters* 251 (3): 205–8. https://doi.org/10.1016/s0304-3940(98)00504-7.

Stein, K. C., F. Morales-Polanco, J. van der Lienden, T. K. Rainbolt, and J. Frydman. 2022. 'Ageing Exacerbates Ribosome Pausing to Disrupt Cotranslational Proteostasis'. *Nature* 601 (7894): 637–42. https://doi.org/10.1038/s41586-021-04295-4.

Stochmanski, S. J., M. Therrien, J. Laganiere, et al. 2012. 'Expanded ATXN3 Frameshifting Events Are Toxic in Drosophila and Mammalian Neuron Models'. *Hum Mol Genet* 21 (10): 2211–18. https://doi.org/10.1093/hmg/dds036.

Stolle, Catherine A., Edward C. Frackelton, Jennifer McCallum, et al. 2008. 'Novel, Complex Interruptions of the GAA Repeat in Small, Expanded Alleles of Two Affected Siblings with Late-Onset Friedreich Ataxia'. *Movement Disorders: Official Journal of the Movement Disorder Society* 23 (9): 1303–6. https://doi.org/10.1002/mds.22012.

Stoyas, Colleen A., and Albert R. La Spada. 2018. 'The CAG-Polyglutamine Repeat Diseases: A Clinical, Molecular, Genetic, and Pathophysiologic Nosology'. *Handbook of Clinical Neurology* 147: 143–70. https://doi.org/10.1016/B978-0-444-63233-3.00011-7.

Sullivan, Roisin, Wai Yan Yau, Viorica Chelban, et al. 2020. 'RFC1 Intronic Repeat Expansions Absent in Pathologically Confirmed Multiple Systems Atrophy'. *Movement Disorders:*

Official Journal of the Movement Disorder Society 35 (7): 1277–79. https://doi.org/10.1002/mds.28074.
Sullivan, Roisin, Wai Yan Yau, Viorica Chelban, et al. 2021. 'RFC1-Related Ataxia Is a Mimic of Early Multiple System Atrophy'. *Journal of Neurology, Neurosurgery, and Psychiatry* 92 (4): 444–46. https://doi.org/10.1136/jnnp-2020-325092.
Swieten, John C. van, Esther Brusse, Bianca M. de Graaf, et al. 2003a. 'A Mutation in the Fibroblast Growth Factor 14 Gene Is Associated with Autosomal Dominant Cerebellar Ataxia [Corrected]'. *American Journal of Human Genetics* 72 (1): 191–99. https://doi.org/10.1086/345488.
Swieten, John C. van, Esther Brusse, Bianca M. de Graaf, et al. 2003b. 'A Mutation in the Fibroblast Growth Factor 14 Gene Is Associated with Autosomal Dominant Cerebellar Ataxia [Corrected]'. *American Journal of Human Genetics* 72 (1): 191–99. https://doi.org/10.1086/345488.
Tan, Dandan, Cuijie Wei, Zhao Chen, et al. 2023. 'CAG Repeat Expansion in THAP11 Is Associated with a Novel Spinocerebellar Ataxia'. *Movement Disorders: Official Journal of the Movement Disorder Society* 38 (7): 1282–93. https://doi.org/10.1002/mds.29412.
Tezenas du Montcel, Sophie, Alexandra Durr, Maria Rakowicz, et al. 2014. 'Prediction of the Age at Onset in Spinocerebellar Ataxia Type 1, 2, 3 and 6'. *Journal of Medical Genetics* 51 (7): 479–86. https://doi.org/10.1136/jmedgenet-2013-102200.
Toulouse, A., F. Au-Yeung, C. Gaspar, J. Roussel, P. Dion, and G. A. Rouleau. 2005. 'Ribosomal Frameshifting on MJD-1 Transcripts with Long CAG Tracts'. *Hum Mol Genet* 14 (18): 2649–60. https://doi.org/10.1093/hmg/ddi299.
Traschütz, Andreas, Andrea Cortese, Selina Reich, et al. 2021. 'Natural History, Phenotypic Spectrum, and Discriminative Features of Multisystemic RFC1 Disease'. *Neurology* 96 (9): e1369–82. https://doi.org/10.1212/WNL.0000000000011528.
Trinh, J., T. Luth, S. Schaake, et al. 2023. 'Mosaic Divergent Repeat Interruptions in XDP Influence Repeat Stability and Disease Onset'. *Brain* 146 (3): 1075–82. https://doi.org/10.1093/brain/awac160.
Trojanowski, J. Q., T. Revesz, and Neuropathology Working Group on MSA. 2007. 'Proposed Neuropathological Criteria for the Post Mortem Diagnosis of Multiple System Atrophy'. *Neuropathology and Applied Neurobiology* 33 (6): 615–20. https://doi.org/10.1111/j.1365-2990.2007.00907.x.
Udine, Evan, NiCole A. Finch, Mariely DeJesus-Hernandez, et al. 2024. 'Targeted Long-Read Sequencing to Quantify Methylation of the C9orf72 Repeat Expansion'. *Molecular Neurodegeneration* 19 (1). https://doi.org/10.1186/s13024-024-00790-0.
Van De Roovaart, Hannah J., Nguyen Nguyen, and Timothy D. Veenstra. 2023. 'Huntington's Disease Drug Development: A Phase 3 Pipeline Analysis'. *Pharmaceuticals* 16 (11): 1513. https://doi.org/10.3390/ph16111513.
Verkerk, A. J., M. Pieretti, J. S. Sutcliffe, et al. 1991. 'Identification of a Gene (FMR-1) Containing a CGG Repeat Coincident with a Breakpoint Cluster Region Exhibiting Length Variation in Fragile X Syndrome'. *Cell* 65 (5): 905–14. https://doi.org/10.1016/0092-8674(91)90397-h.
Volle, Catherine B., and Sarah Delaney. 2013. 'AGG/CCT Interruptions Affect Nucleosome Formation and Positioning of Healthy-Length CGG/CCG Triplet Repeats.' *BMC Biochemistry* (England) 14 (November): 33. https://doi.org/10.1186/1471-2091-14-33.
Walsh, Matthew J., Johnathan Cooper-Knock, Jennifer E. Dodd, et al. 2015. 'Invited Review: Decoding the Pathophysiological Mechanisms That Underlie RNA Dysregulation in Neurodegenerative Disorders: A Review of the Current State of the Art'. *Neuropathology and Applied Neurobiology* 41 (2): 109–34. https://doi.org/10.1111/nan.12187.
Wang, Nan, Shasha Zhang, Peter Langfelder, et al. 2025. 'Distinct Mismatch-Repair Complex Genes Set Neuronal CAG-Repeat Expansion Rate to Drive Selective Pathogenesis in HD Mice'. *Cell* 188 (6): 1524-1544.e22. https://doi.org/10.1016/j.cell.2025.01.031.
Wang, Yang, Junyan Wang, Zhenzhen Yan, et al. 2024. 'Structural Investigation of Pathogenic RFC1 AAGGG Pentanucleotide Repeats Reveals a Role of G-Quadruplex in Dysregulated Gene Expression in CANVAS.' *Nucleic Acids Research* (England) 52 (5): 2698–710. https://doi.org/10.1093/nar/gkae032.

Wassarman, D. A., and F. Sauer. 2001. 'TAF(II)250: A Transcription Toolbox'. *Journal of Cell Science* 114 (Pt 16): 2895–902. https://doi.org/10.1242/jcs.114.16.2895.
Wenning, G. K., Y. Ben-Shlomo, A. Hughes, S. E. Daniel, A. Lees, and N. P. Quinn. 2000. 'What Clinical Features Are Most Useful to Distinguish Definite Multiple System Atrophy from Parkinson's Disease?' *Journal of Neurology, Neurosurgery, and Psychiatry* 68 (4): 434–40. https://doi.org/10.1136/jnnp.68.4.434.
Wenning, Gregor K., Iva Stankovic, Luca Vignatelli, et al. 2022. 'The Movement Disorder Society Criteria for the Diagnosis of Multiple System Atrophy'. *Movement Disorders: Official Journal of the Movement Disorder Society* 37 (6): 1131–48. https://doi.org/10.1002/mds.29005.
Wenninger, Stephan, Sarah A. Cumming, Kristina Gutschmidt, et al. 2021. 'Associations Between Variant Repeat Interruptions and Clinical Outcomes in Myotonic Dystrophy Type 1'. *Neurology. Genetics* 7 (2): e572. https://doi.org/10.1212/NXG.0000000000000572.
Westenberger, A., C. J. Reyes, G. Saranza, et al. 2019. 'A Hexanucleotide Repeat Modifies Expressivity of X-Linked Dystonia Parkinsonism'. *Ann Neurol* 85 (6): 812–22. https://doi.org/10.1002/ana.25488.
Wheeler, Vanessa C., and Vincent Dion. 2021. 'Modifiers of CAG/CTG Repeat Instability: Insights from Mammalian Models'. *Journal of Huntington's Disease* 10 (1): 123–48. https://doi.org/10.3233/JHD-200426.
Wilke, C., D. Pellerin, D. Mengel, et al. 2023. 'GAA-FGF14 Ataxia (SCA27B): Phenotypic Profile, Natural History Progression and 4-Aminopyridine Treatment Response'. *Brain* 146 (10): 4144–57. https://doi.org/10.1093/brain/awad157.
Willems, Thomas, Melissa Gymrek, Gareth Highnam, 1000 Genomes Project Consortium, David Mittelman, and Yaniv Erlich. 2014. 'The Landscape of Human STR Variation'. *Genome Research* 24 (11): 1894–904. https://doi.org/10.1101/gr.177774.114.
Wirth, Thomas, Céline Bonnet, Clarisse Delvallée, et al. 2024. 'Does Spinocerebellar Ataxia 27B Mimic Cerebellar Multiple System Atrophy?' *Journal of Neurology* (Germany) 271 (4): 2078–85. https://doi.org/10.1007/s00415-024-12182-x.
Wirth, Thomas, Guillemette Clément, Clarisse Delvallée, et al. 2023. 'Natural History and Phenotypic Spectrum of GAA-FGF14 Sporadic Late-Onset Cerebellar Ataxia (SCA27B)'. *Movement Disorders: Official Journal of the Movement Disorder Society* 38 (10): 1950–56. https://doi.org/10.1002/mds.29560.
Wright, G. E. B., H. F. Black, J. A. Collins, et al. 2020. 'Interrupting Sequence Variants and Age of Onset in Huntington's Disease: Clinical Implications and Emerging Therapies'. *Lancet Neurol* 19 (11): 930–39. https://doi.org/10.1016/S1474-4422(20)30343-4.
Wright, Galen E. B., Jennifer A. Collins, Chris Kay, et al. 2019. 'Length of Uninterrupted CAG, Independent of Polyglutamine Size, Results in Increased Somatic Instability, Hastening Onset of Huntington Disease'. *American Journal of Human Genetics* 104 (6): 1116–26. https://doi.org/10.1016/j.ajhg.2019.04.007.
Yates, P. A., R. W. Burman, P. Mummaneni, S. Krussel, and M. S. Turker. 1999. 'Tandem B1 Elements Located in a Mouse Methylation Center Provide a Target for de Novo DNA Methylation'. *The Journal of Biological Chemistry* 274 (51): 36357–61. https://doi.org/10.1074/jbc.274.51.36357.
Yousuf, A., N. Ahmed, and A. Qurashi. 2022. 'Non-Canonical DNA/RNA Structures Associated with the Pathogenesis of Fragile X-Associated Tremor/Ataxia Syndrome and Fragile X Syndrome'. *Front Genet* 13: 866021. https://doi.org/10.3389/fgene.2022.866021.
Yrigollen, Carolyn M., Blythe Durbin-Johnson, Louise Gane, et al. 2012. 'AGG Interruptions within the Maternal FMR1 Gene Reduce the Risk of Offspring with Fragile X Syndrome'. *Genetics in Medicine: Official Journal of the American College of Medical Genetics* 14 (8): 729–36. https://doi.org/10.1038/gim.2012.34.
Zeng, Sheng, Mei-Yun Zhang, Xue-Jing Wang, et al. 2019. 'Long-Read Sequencing Identified Intronic Repeat Expansions in SAMD12 from Chinese Pedigrees Affected with Familial Cortical Myoclonic Tremor with Epilepsy'. *Journal of Medical Genetics* 56 (4): 265–70. https://doi.org/10.1136/jmedgenet-2018-105484.
Zu, Tao, Amrutha Pattamatta, and Laura P. W. Ranum. 2018. 'Repeat-Associated Non-ATG Translation in Neurological Diseases'. *Cold Spring Harbor Perspectives in Biology* 10 (12):a033019. https://doi.org/10.1101/cshperspect.a033019.

Appendix

Table of the online links of the publication and supplementary material.

Objective	Address to publication	Address to supplementary material
1A) Nanopore and repeat length and methylation	https://www.mdpi.com/2073-4425/13/1/126	https://www.mdpi.com/article/10.3390/genes13010126/s1
1B) Interruptions in XDP	https://academic.oup.com/brain/article/146/3/1075/6575062?login=true	https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/brain/146/3/10.1093_brain_awac160/1/awac160_supplementary_data.pdf?Expires=1758181011&Signature=yrI9ejJ8wMy3b9HPO--W3airtygmaAEqnW~zoF24NUIhLaB1KqVoyW~EWwWIp-qxJSjVee~uMo8~GZRKDyTI0y0nZc6MSuD31NjSWLMkxWu9feJlUEUqzqR35TtL2ftEOJ2vRT99tLKtaLZZCX-P9aSC1BxEUJ2lntpnn5zRUSfz3JBd3KuuTm3NcoDxmAQD~a9KxT9ygzzVi52ZS2Owgr1j8ucuQ3~W8OnHER1CI2~t2FyEfRcDTzca350yDToRSJWnDDpFFDJmXsAYosVoDQ~BEtTuXBvrLKDIRtkdxGyFF-Lja5atckfyKjQqrS~BSqV4iT7FdcFj9PM0dYKFRg__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA
2) Stability of modifiers in XDP	https://movementdisorders.onlinelibrary.wiley.com/doi/10.1002/mds.29809	https://movementdisorders.onlinelibrary.wiley.com/action/downloadSupplement?doi=10.1002%2Fmds.29809&file=mds29809-sup-0001-supinfo.docx
3) FGF14 in SCA27B	https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awaf183/8133976?login=true	https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/brain/PAP/10.1093_brain_awaf183/2/awaf183_supplementary_data.pdf?Expires=1758180949&Signature=4vMnnb8yNXo2aZ56~2RpjrDIdsFILAYlFUK-svRQ5LK46oIrkRAOd3FQAwfwQcF5-8PDCKHget5Qpmsjx8hXu4gq5LOFSxK1x1opkqLKaU1TBJu78Iay~E7l92GVr~jLXudTVRZWLU7l6oDmRDwwUynQgkGWm3q~tAP-kRYuCRg8Wm0EnoP4RTtUtigMRcqrxic8HCPKLyiukQ~lYIt34QYvTpbNXbMSke9HVhh-viZs1P-eoxgltAWFwtDTr9NpDqF7LvH7fcWM4zmm5m0hBsjWF7aECsc3V~H6MkFwn-Kp7ihG-1uMJxXGq9plLKGXctCQCTjTBnAchvOCs6KVKg__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA
4A) FGF14 in MSA	https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awaf134/8114733	https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/brain/PAP/10.1093_brain_awaf134/2/awaf134_supplementary_data.pdf?Expires=1758181130&Signature=KwNO5c4qWCFYKtPfFO~bEMuGUs9NRR8SDko6v6gPa8LRz78~4I6qVSz2NFiq6TwQzTHH46rSi4JL1rHmuJyOBW0ZlKIR5P7e1sZd~EwaJCY8eQEKZWaDqnEt3AFZFXUBTc97lxKagpbJfCKm2GybhoFTme9ddx-~lchzmsBC7BVVs7cIUYp99WJ22FAfXtpCLL1bbwvmIvBR4xBKNGc1Y~VR5SWyz1hFWKCH0xTfIHxtwpNXRk-aQU7tHJSDofzd2ZOYkPC3SlBtpXSAGgwS0kkQZNg~LT0k~zuCH6eXs~BYicE1F5TkUu7Jba8wkCsnYTcDhcwXeugjOPByqGOLvA__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA

4B) FGF14 in MSA (Korean Cohort)	https://academic.oup.com/brain/advance-article/doi/10.1093/brain/awaf263/8203709?login=false	https://oup.silverchair-cdn.com/oup/backfile/Content_public/Journal/brain/PAP/10.1093_brain_awaf263/1/awaf263_supplementary_data.pdf?Expires=1761286812&Signature=p6Kgt4gLp~l7u~0-xxxWH-3NkKMo57A5B7tRPkMT6fQq3dwSNkPBaYI~jxSdjREBLkAfpIPYUSShoh-taDI9upVgsi2J5sPqI93CjZuJNGBylN4Z1p~LRVoBu7WsVyCdZgRI0OTo2EjgX0k0niJPcL-k6P14pvtPyffi~3jH4mkL5v488Cg2ZRy0wNDZWAyUQbwYVr4qJZnEl~0FOY7kPMEfEd8PEIM~Q2A3bu2jbYADcOSw-1g8irbQu7uO4onXvG-1lOyWzCk7v75h51SiuI5u43aG4gbob65W4rOZwX8Y-jbzcxkPHP-sAbG1FPDfsID~XFS3Hauzp~VCBNg~oQ__&Key-Pair-Id=APKAIE5G5CRDK6RD3PGA