

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

Joshua Laß,^{1,†} Mirja Thomsen,^{1,†} Max Borsche,^{1,2,†} Theresa Lüth,^{1,†} Julia C. Prietzsche,¹ Susen Schaaake,¹ Andona Milovanović,³ Hannah Macpherson,^{4,5} Emil K. Gustavsson,^{4,5} Paula Saffie Awad,^{6,7} Nataša Dragašević-Mišković,³ Björn-Hergen Laabs,⁸ Inke R. König,⁸ Ana Westenberger,¹ Christopher E. Pearson,⁹ Norbert Brüggemann,^{1,2} Christine Klein¹ and Joanne Trinh¹

[†]These authors contributed equally to this work.

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 identified 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 present 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, GAAGGGA, GAAGAAAGAA, GAAAAGGAAGAAGGAAGAAGGAA, GAAAAGGAAGAAGGAA, and GCAGAAGAAGAAGAA. We calculated the longest pure GAA length from the long-read data

© The Author(s) 2025. Published by Oxford University Press on behalf of The Guarantors of Brain. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

for all 40 individuals. When comparing the pure GAA tract between patients and unaffected individuals, clusters were apparent based on greater or less 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 ($p=0.016$, $R^2=0.256$). Somatic-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.

Author affiliations:

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 to: Dr. Joanne Trinh

Institute of Neurogenetics

University of Lübeck and University Hospital Schleswig-Holstein

Campus Lübeck, BMF, Building 67, Room 12

Ratzeburger Allee 160, 23538 Lübeck, Germany

E-mail: joanne.trinh@uni-luebeck.de

Running title: *FGF14* repeat length and interruptions

Keywords: repeat expansion; interruptions; SCA27B, long-read sequencing

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).^{1,2} The expansion is located in the first intron and leads to a reduction of *FGF14* expression, similar to the GAA expansion of Friedrich's ataxia.³ 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.^{1,2} Different studies suggested repeat interruptions in *FGF14* to be non-pathogenic.^{2,4-7} 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 large study used Nanopore sequencing in a

subset of individuals and found a similar frequency of GAAGGA interruptions in patients and controls, where they concluded that the interruption was non-pathogenic.⁶ 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.^{8,9} 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.^{10,11} 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).^{12,13} 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.¹⁴ 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 varies between *FGF14* studies. Even studies with large sample sizes could not consistently demonstrate such an association, which contradicts with the general understanding of repeat expansion disorders.^{15–17} Thus, in-depth genetic methods might shed new light on this relation, which is of importance 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.

Materials and 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,^{1,18} 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,¹⁹ and two patients and one unaffected member of a Chilean family recruited at a movement disorders center in Santiago, Chile.²⁰); and ii) 190 unaffected individuals from Germany.¹ Control participants were recruited within the framework of large cross-sectional or longitudinal studies independent from ataxia research dealing with Parkinson's disease or eating behavior at the Institute of Neurogenetics, Lübeck, Germany.^{21,22} All control participants underwent 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 study was approved by the local ethics committees 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 prior to inclusion in the study, which was performed according to the Declaration of Helsinki. After the collection of 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.¹ 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.¹ 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-TCCTTCTCCTTCTCCTTCTC-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.²³

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 four °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.²⁴ SAM-to-BAM conversion and BAM file handling were conducted with the Samtools software (version 1.15).²⁵ The next step was the sorting and indexing of the reads with Samtools.

The trinucleotide repeats were analyzed with the "Noise-Cancelling Repeat Finder" (NCRF, version 1.01.02).²⁶ 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. The detection of interruptions and their frequency were done with the summary output in R, as previously described.²³ 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 $\alpha = 0.013$. To compare the different methods, Bland-Altman plots were used. Additionally, correlation analysis using a linear regression model implemented with the lm-function was performed.

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 small 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 with the fragment analysis results (one of 16 individuals was outside the limits of agreement) (Fig. 2C-F).

1 **The pure GAA tract predicts disease manifestation and age at onset**

2 Next, we assessed the pure GAA tract length, without interruptions, in patients and unaffected
 3 individuals. The range of the pure GAA tract was 11 to 486 repeat units across these individuals.
 4 For individuals with interruptions, the longest GAA tract was used for further analyses. The
 5 comparison between patients and unaffected individuals resulted in the identification of four
 6 distinct clusters (Fig. 3A and 4A). Recent literature has identified SCA27B patients with a lower
 7 repeat number than 250.^{6,27} We observed a gap between 66 to 226 pure GAA repeats (Fig. 4A).
 8 Therefore, we used a threshold of 200 pure GAA repeat units, albeit in a suggestive manner. The
 9 first group consisted of affected patients with a pure GAA length of >200 repeat units and
 10 diagnosed with SCA27B (affected, Fig. 3A and 4A indicated in red). In the second group were
 11 unaffected individuals with a pure GAA tract >200, indicating pre-manifesting carrier status (pre-
 12 manifesting, Fig. 3A and 4A indicated in yellow). Regarding AAO, the median AAO of SCA27B
 13 patients is reported to be 60 years (range 21-87 years)²⁸, which we used as a supposed threshold
 14 for the categorization of affected and pre-manifesting individuals in Figure 3A. Of note, these six
 15 individuals had an age at examination of 6, 35, 50, 51, 53, and 80 years, respectively. The third
 16 group consisted of unaffected individuals with a pure GAA length <200 and no diagnosis
 17 (unaffected, Fig. 3A and 4A indicated in blue). The last group consisted of patients with a pure
 18 GAA tract <200 repeat units (other ataxia, Fig. 3A and 4A indicated in green). Next, we tested the
 19 relationship between the pure GAA lengths and AAO (Fig. 3B). There was an association between
 20 GAA repeat length and AAO ($p=0.016$, $R^2=0.256$) (Fig. 4B) in patients with a pure GAA tract
 21 >250 repeat units.

23 **Novel repeat interruptions were found through long-read** 24 **sequencing**

25 Seven different repeat interruptions were detected (Table 1,2). Among these, the most frequently
 26 observed interruption motif was (GAAGGA)_n, characterized by a single adenosine to guanine
 27 conversion (A-to-G). This interruption motif was found in twelve individuals (six affected and six

unaffected). The (GAAGGA)_n interruption motif was repeated two to 158 times and was located at the 5' end of the repeat tract, (GAA)₁₋₂₀(GAAGGA)₂₋₁₅₈ (GAA)_n.

Another variant repeat motif was (GAA)₃GAAAAGAAGAAGGAAGAAGGAA(GAA)_n. This motif was identified in two unaffected individuals and included one deletion and two insertions. A similar motif, (GAA)₃GAAAAGAAGAAGGAA(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(GAAGAAAGAA)₃(GAA)_n, which resulted from an insertion of an adenosine. The other motif was (GAA)₄(GAAGAG)₂(GAA)_n. Both motifs were detected in patients.

The most complex motif was (GAA)₂₂(GCAGAAGAAGAA)₃(GCAGAA)(GCAGAAGAAGAAGAA)₂₄(GCAGAA)₁₂(GCAGCAGAA)₃(GCAGAA)₁₃(GCAGCAGAAGCAGAA)₄(GCAGAAGAAGAAGAA)₆(GAA)₂₂GAG(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 reported *FGF14* SCA27B patients.⁴

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.

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.²⁰ 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 (GAAGGA)_n had a mosaic frequency ranging from 0.37 to 0.93. Patients with (GAAGGA)_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 GAAAGAAGAAGGAAGAAGGAA motif and the GAAAGAAGAAGGAA motif were similar, at 0.82 and 0.84, respectively. The shortest interruption motif, GAG, had a slightly lower frequency of 0.70 compared to the other motifs. In contrast to the other motifs, the most complex motif had the overall 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.

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 the short and long allele 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 nor interruption frequencies.

Discussion

Following our discovery of mDRILS in X-linked dystonia-parkinsonism,¹⁴ 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 both 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.

For the analysis of *FGF14* intronic repeat expansions, we propose a new concept to investigate and interpret their potential pathogenicity and 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,⁴ 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. This may further significantly broaden the group of *FGF14* GAA repeat expansion carriers who will likely manifest disease but are currently not considered to be at risk, according to present published recommendations.

Otherwise, our data implies 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.

1 As many studies screened in unspecified ataxia cohorts for repeat expansions in *FGF14* to debunk
2 the clinical phenotype of SCA27B without considering interruptions, our data demonstrates that
3 14% (4/28) of our ataxia patients with *FGF14* repeat expansions had rather non-pathogenic
4 expansions if the presence of mDRILS is taken into account. Consideration of the pure GAA repeat
5 tract and mDRILS is required, especially if the phenotype is not entirely consistent (e.g., early-
6 onset disease, additional movement disorder features) with the typical clinical picture of SCA27B
7 (i.e., pure late-onset cerebellar ataxia +/- episodic features). Moreover, the accurate detection of
8 pure GAA repeats enabled by advanced sequencing methods such as Nanopore and PacBio
9 sequencing might allow more robust genotype-phenotype relations, not only regarding the age at
10 onset but also in the context of disease severity and progression. We suggest that it is feasible to
11 utilize long-read sequencing in *FGF14* repeat expansion disorders in future clinical practice.

12 Notably, this situation in SCA27B extends beyond our previous findings in X-linked dystonia-
13 parkinsonism, where the pathogenic insertion in the *TAF1* gene is the clear-cut cause in all affected
14 individuals, and mDRILS in the hexanucleotide repeats within this insertion act as a modifier of
15 AAO only. Intriguingly, in addition to the penetrance-determining length of the pure GAA tract in
16 SCA27B, we can assume a similar AAO-modifying effect by mDRILS in *FGF14* as well, with a
17 higher mDRILS frequency resulting in shorter uninterrupted GAA tracts, which, in turn, are
18 associated with a later AAO (Fig. 3B). As a relationship between repeat length and age at onset is
19 well-established in the field of repeat expansion disorders and, therefore, is expected to occur
20 likewise in *FGF14*-related disease, our study shows the importance of taking into account
21 mDRILS to better detect phenotypic correlations and impact on repeat stability.¹⁵ The mDRILS in
22 both the XDP-relevant and *FGF14* repeats may affect repeat instability by modifying the
23 propensity to form unusual mutagenic DNA structures, as observed with the interruptions of *FMR1*
24 (FXS) and *ATXN1* (SCA1).²⁹ As with other disease-associated repeat expansions, including the
25 GAA expansions causing Friedreich's ataxia,^{10,30} SCA27B-associated *FGF14* repeat expansions
26 arise on specific haplotypes and can have interruptions in the repeats, suggesting the contribution
27 of *cis*-elements both flanking and within the unstable repeat, respectively.^{1,11,31,32} There is a
28 possibility that a similar mechanism and consequence of interrupting mDRILS play a role in
29 Friedreich's ataxia. However, long-read sequencing on repeat interruption motifs in Friedreich's
30 ataxia has not yet been performed. Evidence suggests an association of specific repeat interruptions
31 with specific haplotypes (flanking *cis*-elements) that may predispose to parent-to-offspring

(germline) repeat expansions.^{1,11,31,32} 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.^{11,31,32}

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 and a larger number of individuals with interruptions are needed to thoroughly examine this. 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,^{33,34} *in vitro* studies of the interrupted versus pure repeats of the SCA1/*ATXN1*, SCA10/*ATXN10*, and FXS/*FMR1* loci,³⁵⁻³⁷ 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.^{11,38,39}

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.⁴⁰⁻⁴⁴ The formation of these structures highlights the pathogenicity of the repeat sequence and can alter how the repeats are metabolized (replication/repair),⁴³ which proteins can be bound and reveal novel manners by which they may be therapeutically targeted, for example, these pathogenic structures can be bound by the ligands TMPyP4 and BRACO-19.^{40,43} Somatic incurred mDRILS of the pathogenic *FGF14* repeat may also lead to pathogenic variations of spliceforms, translation (exonization), repeat-associated non-AUG (RAN) translation, ribosomal frameshifting, ribosomal pausing, transcriptional slippage in the repeat, or repeat instability.^{14,45-50}

The assessment of the *FGF14* mDRILS across time and generations would be beneficial to elucidate the potential impact of somatic changes in the mDRILS frequency on disease progression or severity. Longitudinal monitoring of patients with SCA27B and their mDRILS frequency was not performed in this study. However, our group previously investigated the mDRILS frequency

1 in families affected by XDP, providing evidence that a higher frequency may stabilize the repeats
2 across generations.⁵¹ Thus, future studies should aim for an analogous exploration of *FGF14*
3 mDRILS across time within individual patients and across generations.

4 Overall, all methods (RP-PCR, Sanger sequencing, Nanopore sequencing, and PacBio sequencing)
5 were concordant and detected the same interruptions. We confidently identified five out of seven
6 different repeat interruption motifs within *FGF14*. Nanopore sequencing analysis performed thus
7 far in *FGF14*-related disease could demonstrate that interruptions are present in both unaffected
8 and ataxia patients but have not yet explored 1) the interruptions' position and 2) their impact on
9 repeat expansion pathogenicity at the individual level.⁶

10 Limitations of our study include the relatively small sample size compared to other studies.
11 Moreover, we cannot independently support our estimate of pathogenicity based on advanced
12 genetic methods, as we do not investigate *FGF14* expression, and further biomarkers robustly
13 distinguishing SCA27B from other ataxias are not available to date. Thus, the development of
14 such, preferably accessible biomarkers would be highly beneficial to the field. Future studies with
15 larger sample size, longitudinal examinations, and more in-depth clinical data are needed to
16 investigate not only the influence of pure GAA repeat number on age at onset but also a potential
17 influence on disease severity and progression. Monitoring the mDRILS and their impact on disease
18 progression and severity over a lifetime would be valuable. Our study remains primarily
19 observational and descriptive, necessitating further refinement of mechanistic novelty. Finally, in
20 the context of mosaicism, additional biomaterials such as fibroblasts, induced pluripotent stem
21 cells, or post-mortem brain tissue should be investigated for mDRILS. Future studies should clarify
22 whether the observed mDRILS pattern represents a universal modifier across repeat expansion
23 disorders or a disease-specific phenomenon.

24 In conclusion, this study highlights the importance of an in-depth, multi-method assessment of
25 repeat tract purity in repeat expansion disorders. The correlation of mosaic interruptions with the
26 repeat stability and indirectly patient status suggests a protective effect of mDRILS. Long-read
27 sequencing can uncover the length of the pure GAA motif along with repeat interruptions.
28 However, it is warranted that the assessment of repeat interruptions, the pure GAA tract, and
29 mDRILS be extended to other repeat expansion disorders and that their potential protective
30 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, it also highlights 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 that support the findings of this study are available from the corresponding author, upon reasonable request.

Acknowledgements

We express our deepest gratitude to the families and patients who have participated 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 and recognize her comments on the manuscript. We would also like to thank Frauke Hinrichs for her technical assistance. The thumbnail image for the online table of contents was in part created in BioRender. Klein, C. (2025) <https://BioRender.com/e06b901>.

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 by Bial. The remaining authors report no disclosures.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Rafehi H, Read J, Szmulewicz DJ, et al. An intronic GAA repeat expansion in FGF14 causes the autosomal-dominant adult-onset ataxia SCA50/ATX-FGF14. *Am J Hum Genet.* 2023;110(1):105-119. doi:10.1016/j.ajhg.2022.11.015
2. Pellerin D, Danzi MC, Wilke C, et al. Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia. *N Engl J Med.* 2023;388(2):128-141. doi:10.1056/NEJMoa2207406
3. Ohshima K, Montermini L, Wells RD, Pandolfo M. 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.* 1998;273(23):14588-14595. doi:10.1074/jbc.273.23.14588
4. Pellerin D, Iruzubieta P, Tekgul S, et al. 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.* 2023;38(8):1575-1577. doi:10.1002/mds.29552
5. Ouyang R, Wan L, Pellerin D, et al. The genetic landscape and phenotypic spectrum of GAA-FGF14 ataxia in China: a large cohort study. *EBioMedicine.* 2024;102:105077. doi:10.1016/j.ebiom.2024.105077

6. Mohren L, Erdlenbruch F, Leita E, et al. Identification and characterisation of pathogenic and non-pathogenic FGF14 repeat expansions. *Nat Commun.* 2024;15(1):7665. doi:10.1038/s41467-024-52148-1
7. Ando M, Higuchi Y, Yuan J, et al. Clinical variability associated with intronic FGF14 GAA repeat expansion in Japan. *Ann Clin Transl Neurol.* 2024;11(1):96-104. doi:10.1002/acn3.51936
8. Gall-Duncan T, Sato N, Yuen RKC, Pearson CE. Advancing genomic technologies and clinical awareness accelerates discovery of disease-associated tandem repeat sequences. *Genome Res.* 2022;32(1):1-27. doi:10.1101/gr.269530.120
9. Rajan-Babu IS, Dolzhenko E, Eberle MA, Friedman JM. Sequence composition changes in short tandem repeats: heterogeneity, detection, mechanisms and clinical implications. *Nat Rev Genet.* 2024;25(7):476-499. doi:10.1038/s41576-024-00696-z
10. Cleary JD, Pearson CE. The contribution of cis-elements to disease-associated repeat instability: clinical and experimental evidence. *Cytogenet Genome Res.* 2003;100(1-4):25-55. doi:10.1159/000072837
11. Pellerin D, Del Gobbo GF, Couse M, et al. A common flanking variant is associated with enhanced stability of the FGF14-SCA27B repeat locus. *Nat Genet.* 2024;56(7):1366-1370. doi:10.1038/s41588-024-01808-5
12. Aneichyk T, Hendriks WT, Yadav R, et al. Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell.* 2018;172(5):897-909 e21. doi:10.1016/j.cell.2018.02.011
13. Rakovic A, Domingo A, Grutz K, et al. Genome editing in induced pluripotent stem cells rescues TAF1 levels in X-linked dystonia-parkinsonism. *Mov Disord.* 2018;33(7):1108-1118. doi:10.1002/mds.27441
14. Trinh J, Luth T, Schaake S, et al. Mosaic divergent repeat interruptions in XDP influence repeat stability and disease onset. *Brain.* 2023;146(3):1075-1082. doi:10.1093/brain/awac160
15. Paulson H. Repeat expansion diseases. *Handb Clin Neurol.* 2018;147:105-123. doi:10.1016/B978-0-444-63233-3.00009-9

16. Wilke C, Pellerin D, Mengel D, et al. GAA-FGF14 ataxia (SCA27B): phenotypic profile, natural history progression and 4-aminopyridine treatment response. *Brain*. 2023;146(10):4144-4157. doi:10.1093/brain/awad157
17. Iruzubieta P, Pellerin D, Bergareche A, et al. Frequency and phenotypic spectrum of spinocerebellar ataxia 27B and other genetic ataxias in a Spanish cohort of late-onset cerebellar ataxia. *Eur J Neurol*. 2023;30(12):3828-3833. doi:10.1111/ene.16039
18. Borsche M, Thomsen M, Szmulewicz DJ, et al. Bilateral vestibulopathy in RFC1-positive CANVAS is distinctly different compared to FGF14-linked spinocerebellar ataxia 27B. *J Neurol*. 2024;271(2):1023-1027. doi:10.1007/s00415-023-12050-0
19. Milovanovic A, Dragasevic-Miskovic N, Thomsen M, et al. RFC1 and FGF14 Repeat Expansions in Serbian Patients with Cerebellar Ataxia. *Mov Disord Clin Pr*. 2024;11(6):626-633. doi:10.1002/mdc3.14020
20. Saffie Awad P, Lohmann K, Hirmas Y, et al. Shaking Up Ataxia: FGF14 and RFC1 Repeat Expansions in Affected and Unaffected Members of a Chilean Family. *Mov Disord*. 2023;38(6):1107-1109. doi:10.1002/mds.29390
21. Kasten M, Hagenah J, Graf J, et al. Cohort Profile: a population-based cohort to study non-motor symptoms in parkinsonism (EPIPARK). *Int J Epidemiol*. 2013;42(1):128-128k. doi:10.1093/ije/dys202
22. Krause C, Sievert H, Geißler C, et al. Critical evaluation of the DNA-methylation markers ABCG1 and SREBF1 for Type 2 diabetes stratification. *Epigenomics*. 2019;11(8):885-897. doi:10.2217/epi-2018-0159
23. Luth T, Labeta J, Schaake S, et al. Elucidating Hexanucleotide Repeat Number and Methylation within the X-Linked Dystonia-Parkinsonism (XDP)-Related SVA Retrotransposon in TAF1 with Nanopore Sequencing. *Genes Basel*. 2022;13(1). doi:10.3390/genes13010126
24. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34(18):3094-3100. doi:10.1093/bioinformatics/bty191
25. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-2079. doi:10.1093/bioinformatics/btp352

26. Harris RS, Cechova M, Makova KD. Noise-cancelling repeat finder: uncovering tandem repeats in error-prone long-read sequencing data. *Bioinformatics*. 2019;35(22):4809-4811. doi:10.1093/bioinformatics/btz484
27. Hengel H, Pellerin D, Wilke C, et al. As Frequent as Polyglutamine Spinocerebellar Ataxias: SCA27B in a Large German Autosomal Dominant Ataxia Cohort. *Mov Disord*. 2023;38(8):1557-1558. doi:10.1002/mds.29559
28. Pellerin D, Danzi M, Renaud M, et al. GAA-FGF14-Related Ataxia. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, eds. *GeneReviews((R))*.; 1993.
29. Yousuf A, Ahmed N, Qurashi A. Non-canonical DNA/RNA structures associated with the pathogenesis of Fragile X-associated tremor/ataxia syndrome and Fragile X syndrome. *Front Genet*. 2022;13:866021. doi:10.3389/fgene.2022.866021
30. Al-Mahdawi S, Ging H, Bayot A, et al. Large Interruptions of GAA Repeat Expansion Mutations in Friedreich Ataxia Are Very Rare. *Front Cell Neurosci*. 2018;12:443. doi:10.3389/fncel.2018.00443
31. De T, Sharma P, Upilli B, et al. Spinocerebellar ataxia type 27B (SCA27B) in India: insights from a large cohort study suggest ancient origin. *Neurogenetics*. 2024;25(4):393-403. doi:10.1007/s10048-024-00770-y
32. Miyatake S, Doi H, Yaguchi H, et al. Complete nanopore repeat sequencing of SCA27B (GAA-FGF14 ataxia) in Japanese. *J Neurol Neurosurg Psychiatry*. 2024;95(12):1187-1195. doi:10.1136/jnnp-2024-333541
33. Libby RT, Hagerman KA, Pineda VV, et al. CTCF cis-regulates trinucleotide repeat instability in an epigenetic manner: a novel basis for mutational hot spot determination. *PLoS Genet*. 2008;4(11):e1000257. doi:10.1371/journal.pgen.1000257
34. Libby RT, Monckton DG, Fu YH, et al. Genomic context drives SCA7 CAG repeat instability, while expressed SCA7 cDNAs are intergenerationally and somatically stable in transgenic mice. *Hum Mol Genet*. 2003;12(1):41-50. doi:10.1093/hmg/ddg006
35. Mulvihill DJ, Nichol Edamura K, Hagerman KA, Pearson CE, Wang YH. Effect of CAT or AGG interruptions and CpG methylation on nucleosome assembly upon trinucleotide repeats

- on spinocerebellar ataxia, type 1 and fragile X syndrome. *J Biol Chem.* 2005;280(6):4498-4503. doi:10.1074/jbc.M413239200
36. Volle CB, Delaney S. AGG/CCT interruptions affect nucleosome formation and positioning of healthy-length CGG/CCG triplet repeats. *BMC Biochem.* 2013;14:33. doi:10.1186/1471-2091-14-33
37. Hagerman KA, Ruan H, Edamura KN, Matsuura T, Pearson CE, Wang YH. The ATTCT repeats of spinocerebellar ataxia type 10 display strong nucleosome assembly which is enhanced by repeat interruptions. *Gene.* 2009;434(1-2):29-34. doi:10.1016/j.gene.2008.12.011
38. Barbé L, Lanni S, López-Castel A, et al. CpG Methylation, a Parent-of-Origin Effect for Maternal-Biased Transmission of Congenital Myotonic Dystrophy. *Am J Hum Genet.* 2017;100(3):488-505. doi:10.1016/j.ajhg.2017.01.033
39. Mirceta M, Schmidt MHM, Shum N, et al. C9orf72 repeat expansion creates the unstable folate-sensitive fragile site FRA9A. *NAR Mol Med.* 2024;1(4):ugae019. doi:10.1093/narmme/ugae019
40. Abdi MH, Zamiri B, Pazuki G, Sardari S, Pearson CE. Pathogenic CANVAS-causing but not nonpathogenic RFC1 DNA/RNA repeat motifs form quadruplex or triplex structures. *J Biol Chem.* 2023;299(10):105202. doi:10.1016/j.jbc.2023.105202
41. Hisey JA, Radchenko EA, Mandel NH, et al. Pathogenic CANVAS (AAGGG)_n repeats stall DNA replication due to the formation of alternative DNA structures. *Nucleic Acids Res.* 2024;52(8):4361-4374. doi:10.1093/nar/gkae124
42. Wang Y, Wang J, Yan Z, et al. Structural investigation of pathogenic RFC1 AAGGG pentanucleotide repeats reveals a role of G-quadruplex in dysregulated gene expression in CANVAS. *Nucleic Acids Res.* 2024;52(5):2698-2710. doi:10.1093/nar/gkae032
43. Singh K, Shukla S, Shankar U, et al. Elucidating the pathobiology of Cerebellar Ataxia with Neuropathy and Vestibular Areflexia Syndrome (CANVAS) with its expanded RNA structure formation and proteinopathy. *Sci Rep.* 2024;14(1):28054. doi:10.1038/s41598-024-78947-6

44. Kudo K, Hori K, Asamitsu S, et al. Structural polymorphism of the nucleic acids in pentanucleotide repeats associated with the neurological disorder CANVAS. *J Biol Chem.* 2024;300(4):107138. doi:10.1016/j.jbc.2024.107138
45. Stochmanski SJ, Therrien M, Laganieri J, et al. Expanded ATXN3 frameshifting events are toxic in *Drosophila* and mammalian neuron models. *Hum Mol Genet.* 2012;21(10):2211-2218. doi:10.1093/hmg/dds036
46. Gaspar C, Jannatipour M, Dion P, et al. CAG tract of MJD-1 may be prone to frameshifts causing polyalanine accumulation. *Hum Mol Genet.* 2000;9(13):1957-1966. doi:10.1093/hmg/9.13.1957
47. Toulouse A, Au-Yeung F, Gaspar C, Roussel J, Dion P, Rouleau GA. Ribosomal frameshifting on MJD-1 transcripts with long CAG tracts. *Hum Mol Genet.* 2005;14(18):2649-2660. doi:10.1093/hmg/ddi299
48. Aviner R, Lee TT, Mastro VB, Li KH, Andino R, Frydman J. Polyglutamine-mediated ribotoxicity disrupts proteostasis and stress responses in Huntington's disease. *Nat Cell Biol.* 2024;26(6):892-902. doi:10.1038/s41556-024-01414-x
49. Stein KC, Morales-Polanco F, van der Lienden J, Rainbolt TK, Frydman J. Ageing exacerbates ribosome pausing to disrupt cotranslational proteostasis. *Nature.* 2022;601(7894):637-642. doi:10.1038/s41586-021-04295-4
50. Parsons MA, Sinden RR, Izban MG. Transcriptional properties of RNA polymerase II within triplet repeat-containing DNA from the human myotonic dystrophy and fragile X loci. *J Biol Chem.* 1998;273(41):26998-27008. doi:10.1074/jbc.273.41.26998
51. Laß J, Lüth T, Schlüter K, et al. Stability of Mosaic Divergent Repeat Interruptions in X-Linked Dystonia-Parkinsonism. *Mov Disord Off J Mov Disord Soc.* 2024;39(7):1145-1153. doi:10.1002/mds.29809

Figure legends

Figure 1 Overview of the study design. Initial screening of *FGF14* repeat expansions (≥ 250 repeats) was previously performed with long-range PCR.^{1,17-19} 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>.

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.

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 compared to early-onset SCA27B. Legend: mDRILS = mosaic divergent repeat interruptions affection 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, determined by Nanopore sequencing. (A) Differences between affected and unaffected repeat expansion carriers in pure GAA length.

Shape coding: triangle = affected, diamond = pre-manifesting, circle = unaffected, square = 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 $\alpha = 0.013$.

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.

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

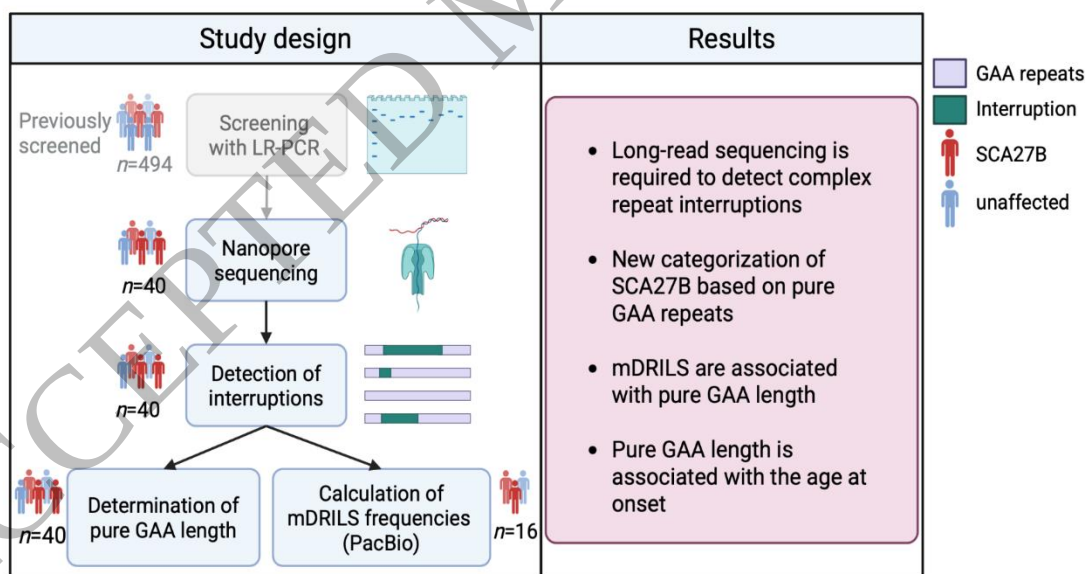


Figure 1
209x108 mm (x DPI)

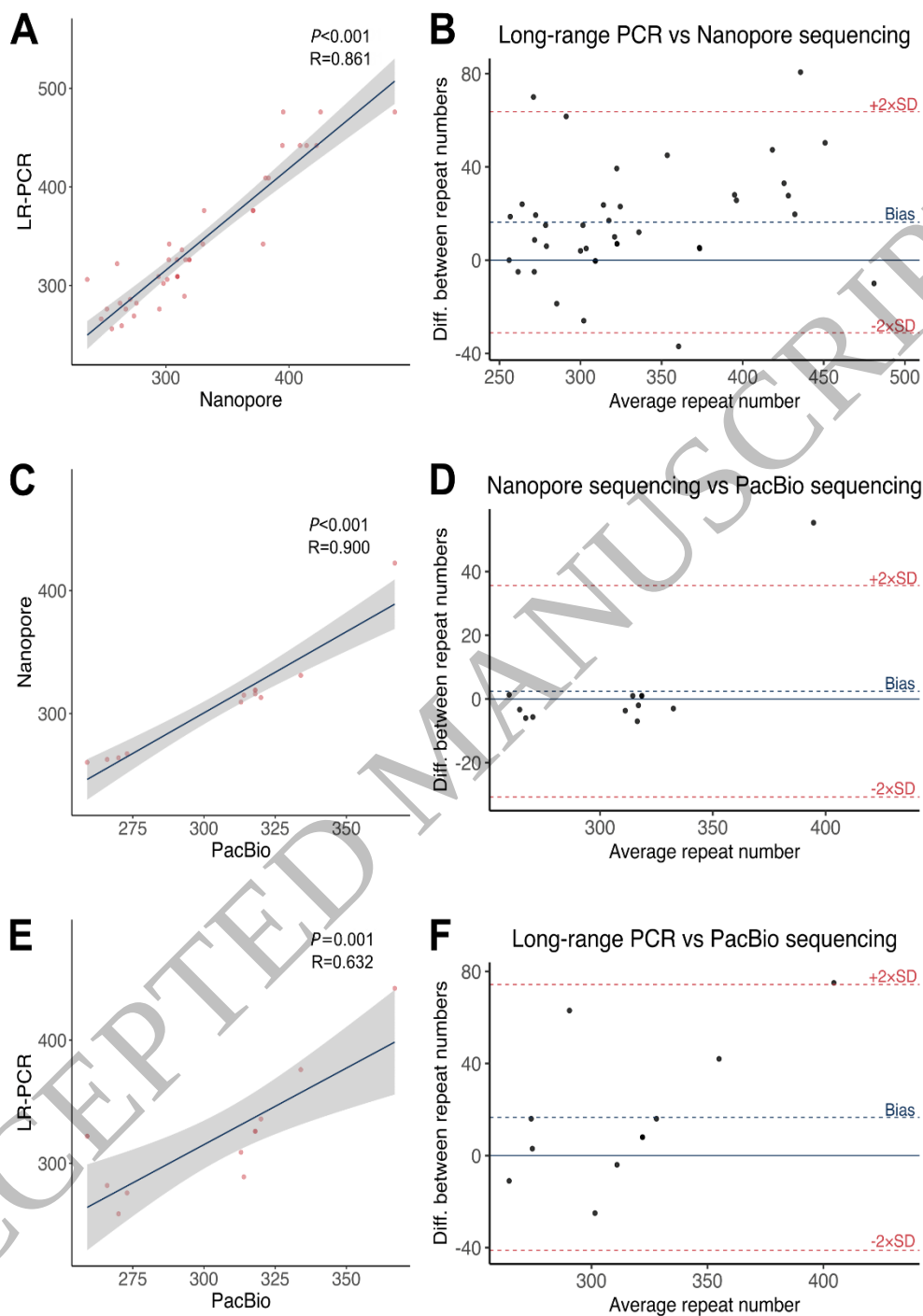
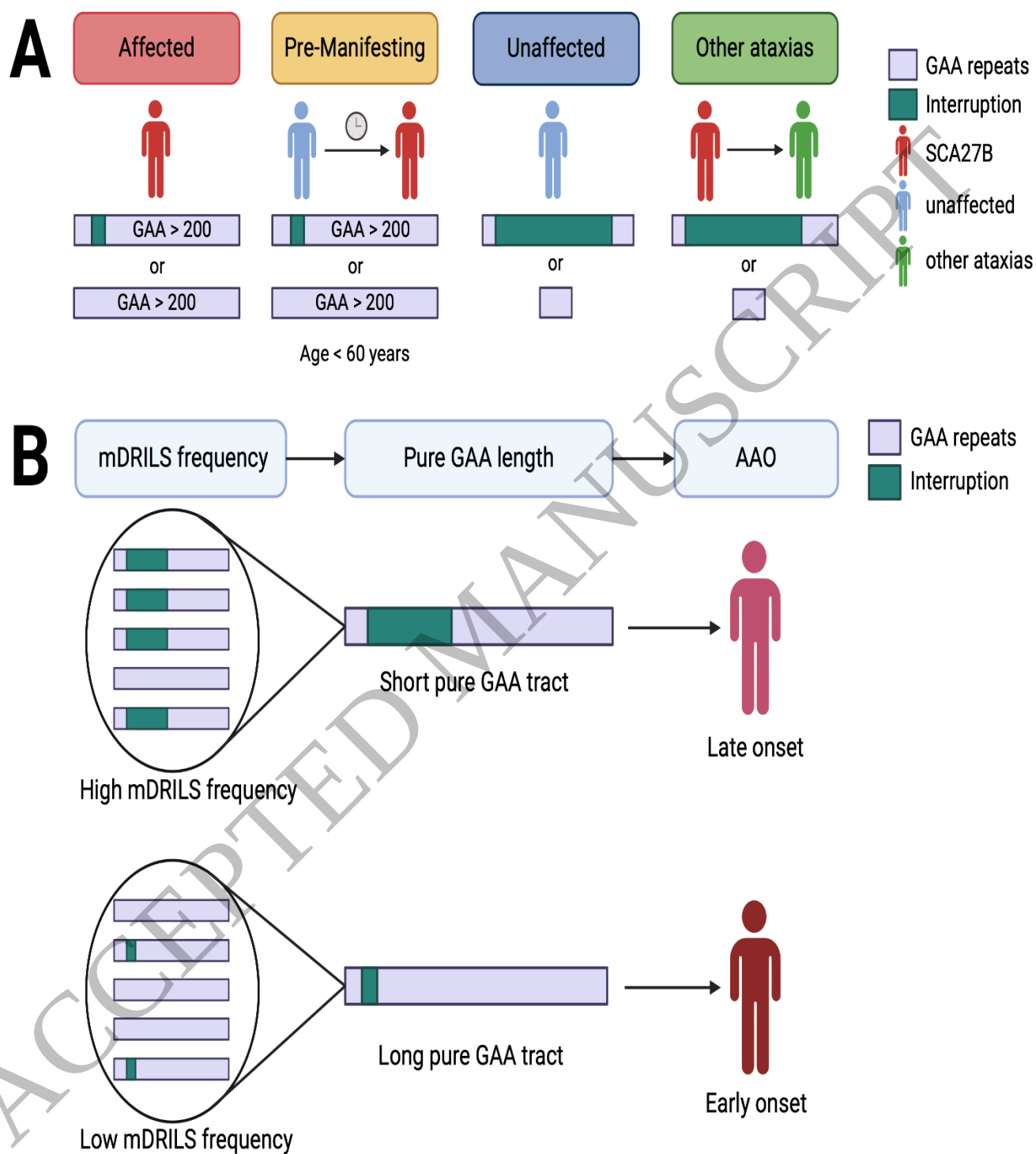


Figure 2
199x265 mm (x DPI)



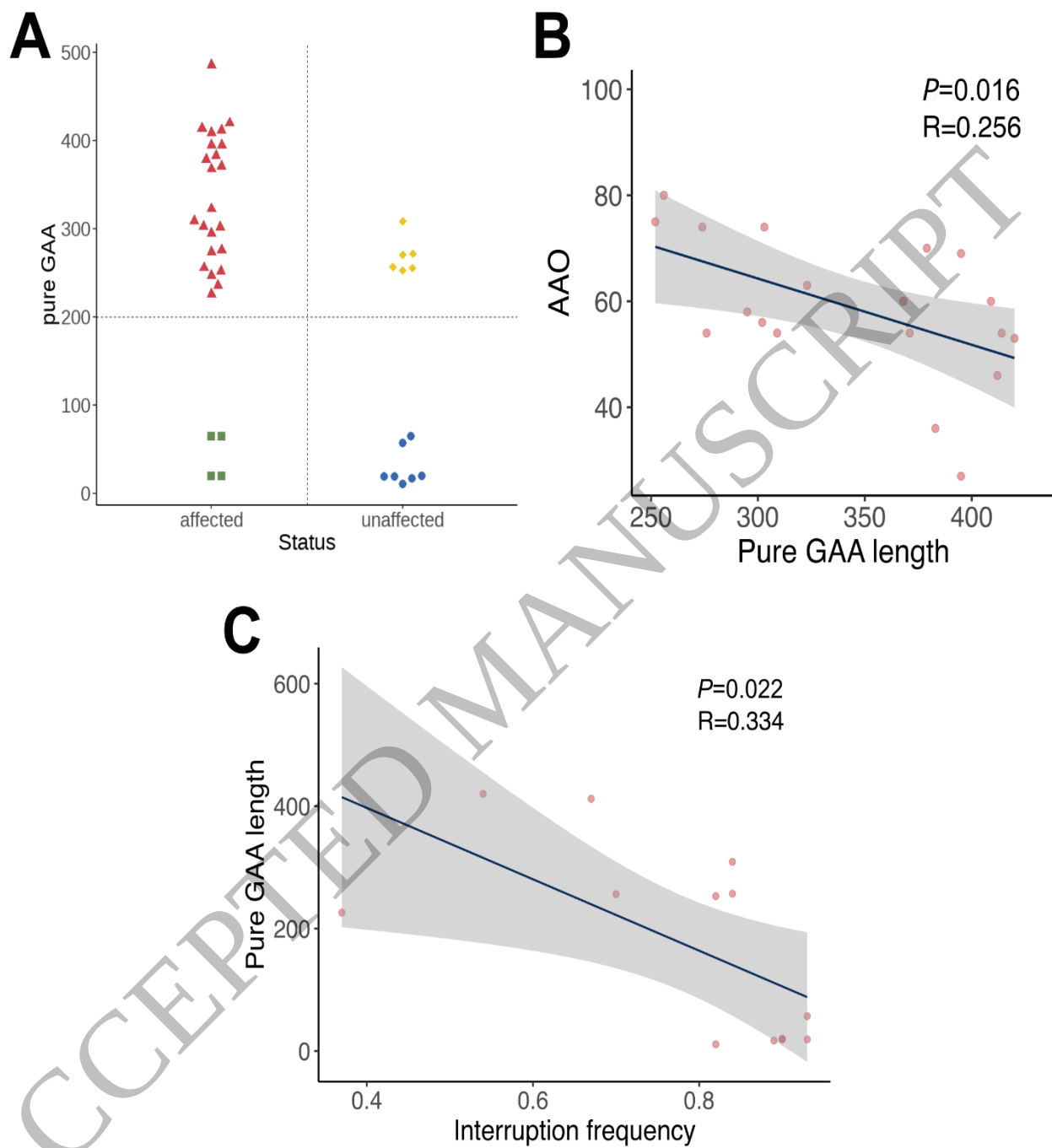


Figure 4
220x204 mm (x DPI)

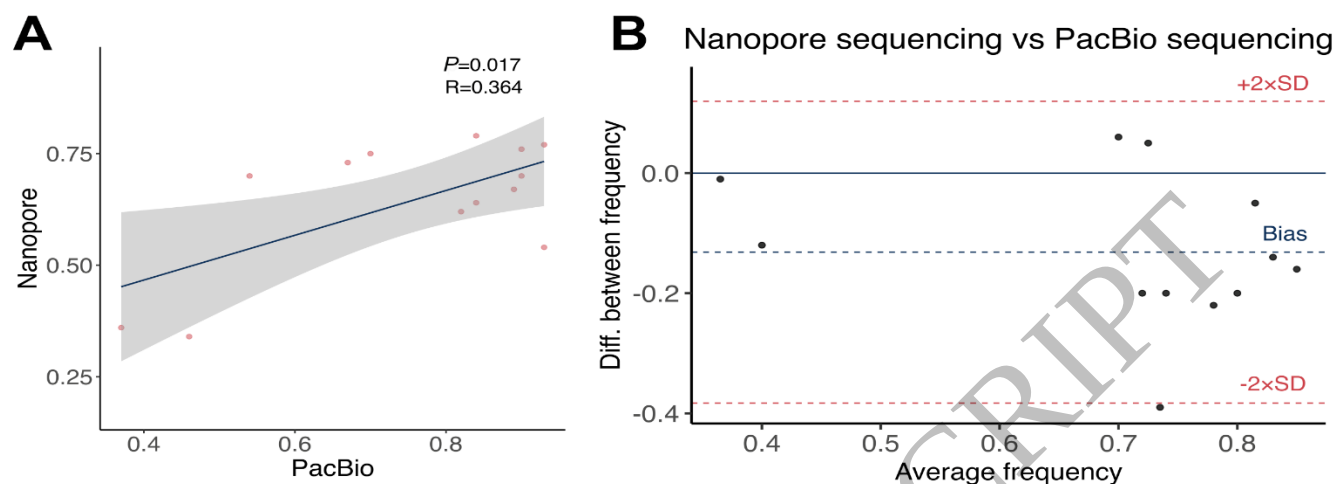


Figure 5
209x87 mm (x DPI)

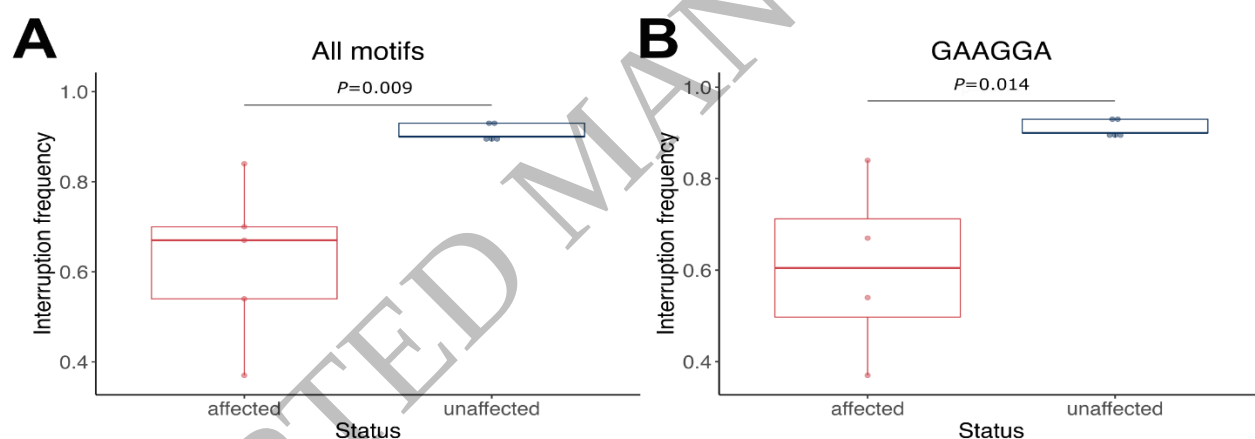


Figure 6
210x82 mm (x DPI)

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

ID	Status	Previous y Publishe d (PMID)	RN (ONT)	RN (LR- PCR)	RN (PacBio)	Interruption frequency (PacBio)	Pur e GA A	Interruption in the long allele (ONT)	Methods confirming interruptio n
L-3013	Unaffected	36493768	316	326	318	0.90	20	GAAGGA	PacBio, Sanger
L-3329	Unaffected	36493768	271	286	276	-	271	-	Sanger
L-3344	Unaffected	36493768	319	326	318	0.90	19	GAAGGA	PacBio, Sanger
L-3479	Unaffected	36493768	260	322	259	0.93	57	GAAGGA	PacBio, Sanger
L-3501	Unaffected	36493768	319	326	318	0.93	19	GAAGGA	PacBio, Sanger
L-3656	Unaffected	36493768	294	309	-	-	11	GAAGGA	-

L-3657	Unaffected	36493768	309	326	-	-	309	-	Sanger
L-8761	Unaffected	36493768	264	259	270	0.84	257	<u>GAAA</u> GAAGAAGGAA	PacBio, Sanger
L-8782	Unaffected	36493768	256	256	-	-	256	-	-
L-8934	Unaffected	36493768	315	289	314	0.89	17	GAAGGA	PacBio, Sanger
L-10408	Unaffected	36493768	263	282	266	0.82	253	<u>GAAA</u> GAAGAAGGAAGAAAG <u>GAA</u>	PacBio, Sanger
L-10410	Unaffected	36493768	281	292	-	-	243	<u>GAAA</u> GAAGAAGGAAGAAAG <u>GAA</u>	Sanger
L-14996	Unaffected	37246629	309	309	-	-	66	<u>GC</u> AGAAGAAGAAGAA	-

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	GAAGGA	PacBio, Sanger
L-14630	Patient	36493768	313	336	320	0.84	309	GAAGGA	PacBio, Sanger
L-14846	Patient	38487929	371	376	-	-	371	-	Sanger
L-14853	Patient	38487929	298	302	-	-	20	GAAGGA	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	<u>GC</u> AGAAGAAGAAGAA	PacBio, Sanger
L-14997	Patient	37246629	309	309	-	-	66	<u>GC</u> AGAAGAAGAAGAA	Sanger
L-15166	Patient	36493768	331	376	334	0.37	226	GAAGGA	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	GAAGAG	Not confirmed
L-15891	Patient	36493768	301	306	-	-	20	GAAGGA	Sanger
L-17665	Patient	36493768	422	442	367	0.67	412	GAAGGA	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	GAAGAAAGAA	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	<u>GAG</u>	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.

Efficacy made Convenient



TYSABRI SC injection with the potential to administer **AT HOME** for eligible patients*

Efficacy and safety profile comparable between TYSABRI IV and SC^{†,2}

[†]Comparable PK, PD, efficacy, and safety profile of SC to IV except for injection site pain.^{1,2}

CLICK HERE TO DISCOVER MORE ABOUT
TYSABRI SC AND THE DIFFERENCE IT MAY
MAKE TO YOUR ELIGIBLE PATIENTS

Supported by



A Biogen developed and funded JCV antibody index PML risk stratification service, validated and available exclusively for patients on or considering TYSABRI.



*As of April 2024, TYSABRI SC can be administered outside a clinical setting (e.g. at home) by a HCP for patients who have tolerated at least 6 doses of TYSABRI well in a clinical setting. Please refer to section 4.2 of the SmPC.¹

TYSABRI is indicated as single DMT in adults with highly active RRMS for the following patient groups:^{1,2}

- Patients with highly active disease despite a full and adequate course of treatment with at least one DMT
- Patients with rapidly evolving severe RRMS defined by 2 or more disabling relapses in one year, and with 1 or more Gd+ lesions on brain MRI or a significant increase in T2 lesion load as compared to a previous recent MRI

Very common AEs include nasopharyngitis and urinary tract infection. Please refer to the SmPC for further safety information, including the risk of the uncommon but serious AE, PML.^{1,2}

Abbreviations: **AE:** Adverse Event; **DMT:** Disease-Modifying Therapy; **Gd+:** Gadolinium-Enhancing; **HCP:** Healthcare Professional; **IV:** Intravenous; **JCV:** John Cunningham Virus; **MRI:** Magnetic Resonance Imaging; **PD:** Pharmacodynamic; **PK:** Pharmacokinetic; **PML:** Progressive Multifocal Leukoencephalopathy; **RRMS:** Relapsing-Remitting Multiple Sclerosis; **SC:** Subcutaneous.

References: 1. TYSABRI SC (natalizumab) Summary of Product Characteristics. 2. TYSABRI IV (natalizumab) Summary of Product Characteristics.

Adverse events should be reported. For Ireland, reporting forms and information can be found at www.hpra.ie. For the UK, reporting forms and information can be found at <https://yellowcard.mhra.gov.uk/> or via the Yellow Card app available from the Apple App Store or Google Play Store. Adverse events should also be reported to Biogen Idc on MedInfoUKI@biogen.com 1800 812 719 in Ireland and 0800 008 7401 in the UK.