BRAIN
ORIGINAL ARTICLE

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 Schaake,[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 Klein[1] and Joanne Trinh[1]

[†]These authors contributed equally to this work.

Deep intronic *FGF14* repeat expansions have been identified as a frequent genetic cause of late-onset cerebellar ataxias, explaining ≤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 used 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 *FGF14* 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 (standard deviation = 60), with 20 individuals showing repeat interruptions, including complex motifs such as GAG, GAAGGA, GAAGAAAGAA, GAAAAGAAGAAGGAAGAAGGAA, GAAAAGAAGAAGGAA and GCAGAAGAAGAAGAA. 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 >200 or <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 were correlated with pure GAA length ($P = 0.022$, $R^2 = 0.334$), with a higher mosaic frequency of interruptions in unaffected individuals compared with 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; and (iv) we provide evidence that mosaicism stabilizes pure GAA repeats in interrupted *FGF14* repeat expansions.

Received November 13, 2024. Revised March 24, 2025. Accepted April 26, 2025. Advance access publication May 17, 2025

1 Institute of Neurogenetics, University of Lübeck and University Hospital Schleswig-Holstein, Lübeck 23538, Germany
2 Department of Neurology, University of Lübeck and University Hospital Schleswig-Holstein, Lübeck 23538, 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, Lübeck 23538, Germany
9 Department of Molecular Genetics, The Hospital for Sick Children, Genetics & Genome Biology, University of Toronto, Toronto, ON M5G 0A4, Canada

Correspondence to: Prof. Joanne Trinh
Institute of Neurogenetics, University of Lübeck and University
Hospital Schleswig-Holstein, Campus Lübeck, BMF, Building 67
Room 12, Ratzeburger Allee 160, Lübeck 23538, Schleswig-Holstein, Germany
E-mail: joanne.trinh@uni-luebeck.de

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 to be 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 GAA<u>GGA</u> interruptions in patients and controls, where they concluded that the interruption was non-pathogenic.[6] However, interruptions have not been characterized further 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] Although 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 30–55 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 contribute collectively to repeat instability. These variations may remain undetected without assessment of single molecules. Data on the correlation between repeat length and AAO varies 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.[15-17] Thus, in-depth genetic methods might shed new light on this relationship, which is of importance for clinical care and patient counselling.

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 AAO, dependent on the interruption length and position, with high accuracy; and (iv) we provide evidence that mosaicism stabilizes pure GAA repeats in interrupted *FGF14* repeat expansions.

Materials and methods

Participant recruitment

The individuals analysed 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 centres 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

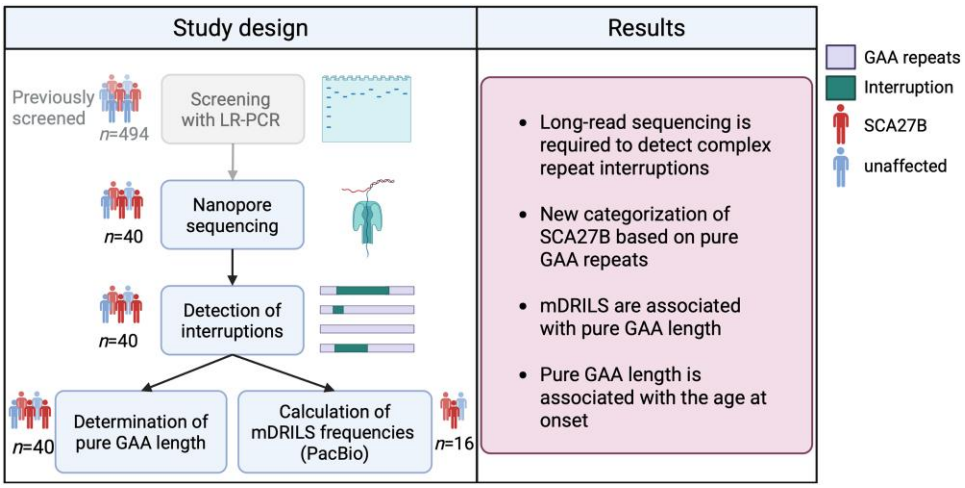

Figure 1 Overview of the study design. Initial screening of *FGF14* repeat expansions (≥250 repeats) was previously performed with long-range PCR.[1,17-19] In the present study, individuals with detected expansions were analysed further using long-read sequencing to detect interruptions, pure GAA tract length and somatic mosaicism. LR-PCR = long-range PCR; mDRILS = mosaic divergent repeat interruption affecting length and sequence; PacBio = PacBio sequencing. Created in BioRender. Laß, J. (2025) https://BioRender.com/a67g089.

centre 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 behaviour 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 sampling, which was set as the 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′-CAGTTCCTGCCCAC ATAGAGC-3′) and FGF14-R (5′-AGCAATCGTCAGTCAGTGTAA GC-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 was set up using the Platinum SuperFi II Master Mix (Thermo Fisher Scientific), 5% dimethyl sulphoxide, 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 (CACGACGTTGT AAAACGAC) was attached to the forward primer, and a third FAM-labelled primer (FAM-M13F) was added to analyse 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 of >250, corresponding to products with a size greater than ~900 bp, repeat-primed PCR (RP-PCR), Sanger sequencing, Nanopore sequencing and PacBio sequencing were conducted.

Repeat-primed PCR

RP-PCR was used to analyse 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-CTTCTTCTT CTTCTTCTTCTT-3′; FGF14-RP-GAAGGA: 5′-M13F-TCCTTCTCCTTC TCCTTC-3′), and an FAM-labelled M13-primer (FAM-M13F). Cycling conditions are displayed in Supplementary Table 2. The products were analysed 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 analyse the repeat expansion in *FGF14*. The approach involved PCR amplification of the repeat tract, using 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

v.3.1 Cycle Sequencing Kit (Applied Biosystems) with primers FGF14-F and FGF14-R. The sequencing conditions included 25 cycles of 96°C for 10 s, 55°C for 5 s and 60°C for 3 min. Sequencing products were purified using sodium acetate precipitation and analysed 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 165 kb 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 10 min was added before the 4°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 h.

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 analysed with the software Nanostat (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 software Samtools (version 1.15).[25] The next step was the sorting and indexing of the reads with Samtools.

The trinucleotide repeats were analysed 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 was 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 [interquartile range (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 3 of 40 individuals were outside the limits of agreement (Fig. 2B).

PacBio sequencing was performed on 16 of 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 (1 of 16 individuals was outside the limits of agreement) (Fig. 2C–F).

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 11–486 repeat units across these individuals. For individuals with interruptions, the longest GAA tract was used for further analyses. The comparison between patients and unaffected individuals resulted in the identification of four distinct clusters (Figs 3A and 4A). Recent literature has identified SCA27B patients with a lower repeat number than 250.[6,27] 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, Figs 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, Figs 3A and 4A, indicated in yellow). Regarding AAO, the median AAO of SCA27B patients is reported to be 60 years (range 21–87 years),[28] which we used as a supposed threshold for the categorization of affected and pre-manifesting individuals in Fig. 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, Figs 3A and 4A, indicated in blue). The last group consisted of patients with a pure GAA tract <200 repeat units (other ataxia, Figs 3A and 4A, indicated in green).

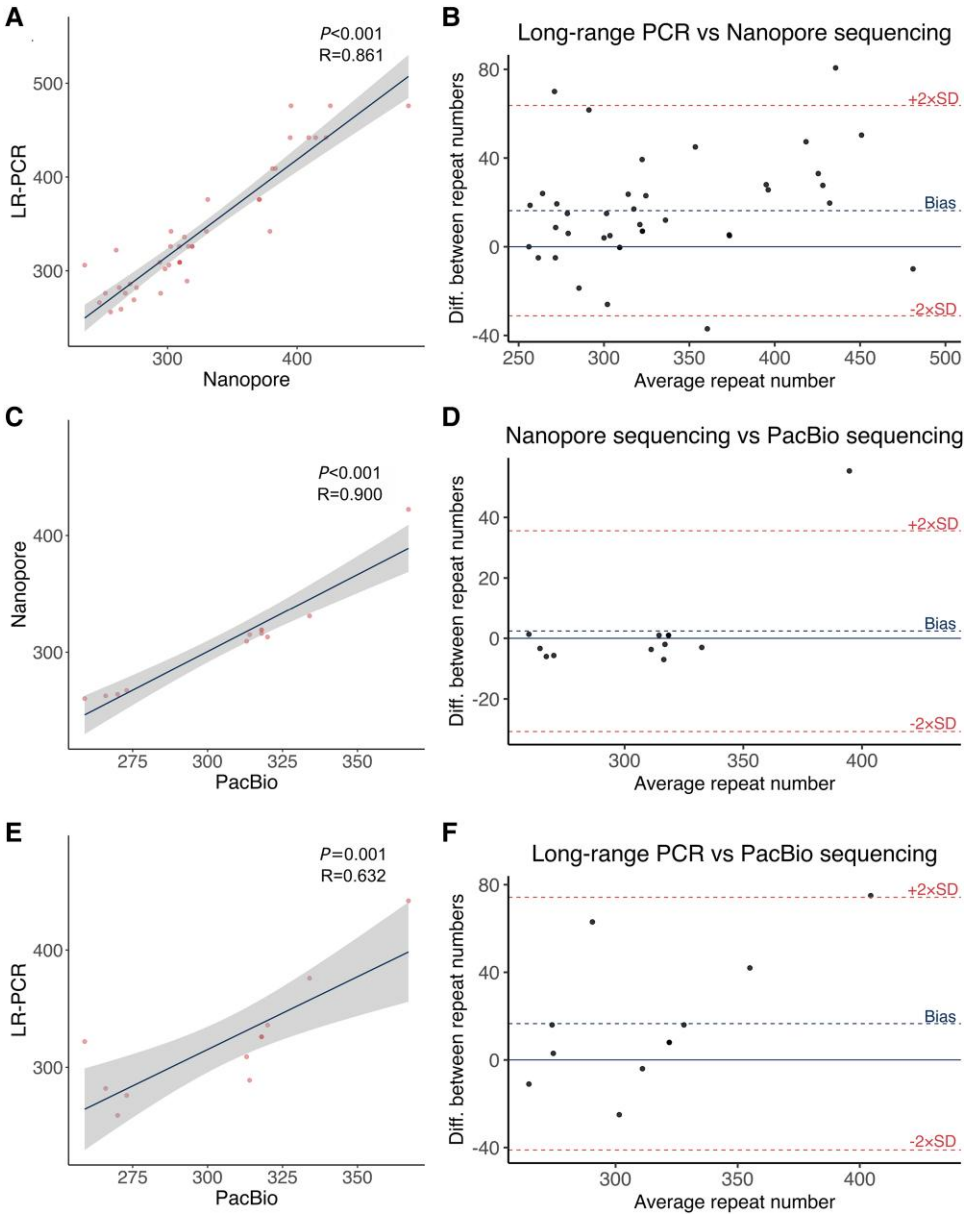

Figure 2 Comparison of the repeat length between different methods. (**A**) Correlation between LR-PCR and Nanopore. (**B**) Bland–Altman plots between LR-PCR and Nanopore. (**C**) Correlation between Nanopore and PacBio. (**D**) Bland–Altman plots between Nanopore and PacBio. (**E**) Correlation between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

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.

Novel repeat interruptions were found through long-read sequencing

Seven different repeat interruptions were detected (Tables 1 and 2). Among these, the most frequently observed interruption motif was $(GAAGGA)_n$, characterized by a single adenosine-to-guanine conversion (A-to-G). This interruption motif was found in 12 individuals (6 affected and 6 unaffected). The $(GAAGGA)_n$ interruption motif was repeated 2–158 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 $(GAA)_3GAAAAGAA$ $GAAGGAAGAAGGAA(GAA)_n$. 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 individual. 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 $(GAA)_{22}(GCAGAAGAAGAA)_3$ $(GCAGAA)(GCAGAAGAAGAAGAA)_{24}(GCAGAA)_{12}(GCAGCAGAA)_3(GCA$ $GAA)_{13}(GCAGCAGAAGCAGAA)_4(GCAGAAGAAGAAGAA)_6(GAA)_{22}GAG-$ $(GAA)_n$. This motif was observed in two affected members and one unaffected member of one family. However, the affected members

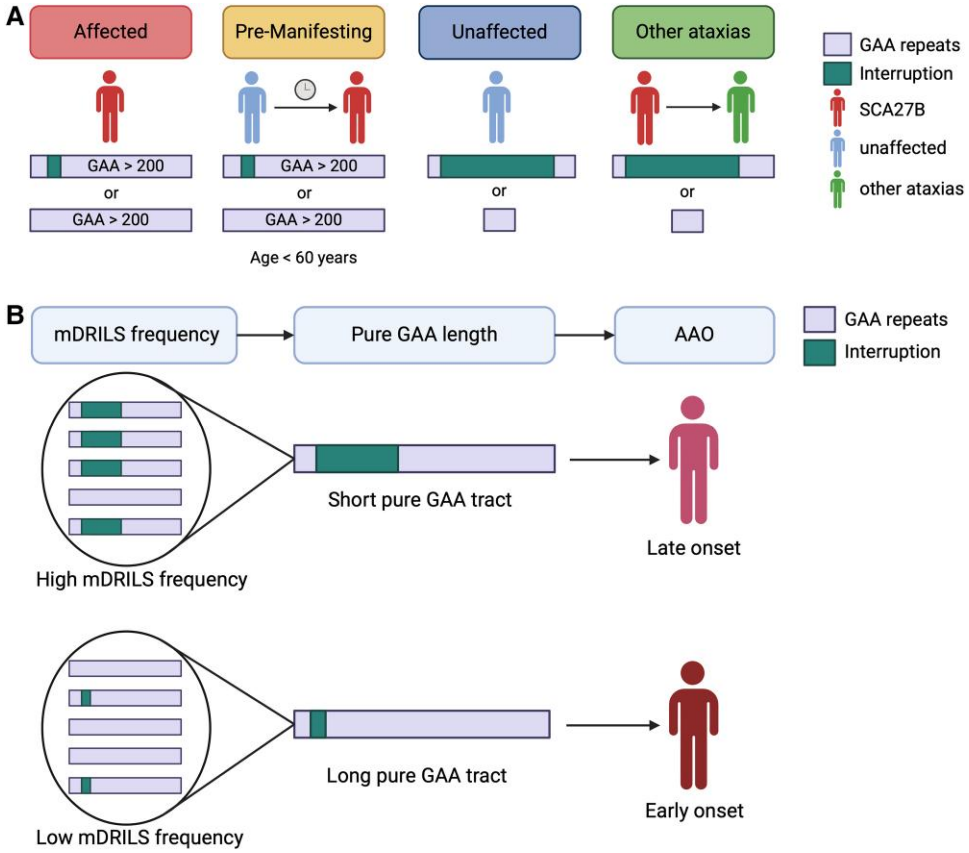

Figure 3 Schematic illustrations. (**A**) New categorization of individuals based on pure GAA length. Affected individuals with SCA27B have either >200 pure GAA repeats (i.e. without interruption) or >200 pure GAA repeats despite an interruption. Pre-manifesting carriers are individuals with an earlier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

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 and B). We used PacBio sequencing data for mosaic frequency calculations owing 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 included only 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) in comparison to patients (0.67, IQR: 0.54–0.70) ($P = 0.009$, $Z = -2.627$) (Fig. 6A). The most common interruption motif, $(GAA\underline{GGA})_n$, had a mosaic frequency ranging from 0.37 to 0.93. Patients with $(GAA\underline{GGA})_n$ exhibited a lower median frequency of 0.61 (IQR: 0.50–0.71) in comparison 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<u>A</u>GAAGAA<u>G</u>GAAGAA<u>G</u>GAA motif and the GAAA<u>A</u>GAAGAA<u>G</u>GAA motif were similar, at 0.82 and 0.84, respectively. The shortest interruption motif, <u>GAG</u>, had a slightly lower frequency of 0.70 in comparison to the other motifs. In contrast to the other motifs, the most complex motif had the overall lowest frequency, at 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.

Some interrupting sequences compromise the stability or toxicity of GAA repeats, whereas others may stabilize the repeat or exert more severe damaging effects. Thus, we investigated the 5′ flanking region of 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 10 of the 20 individuals with an interruption, the motif CTTTCTGTGTAGTCATAGTACCCC was detected.

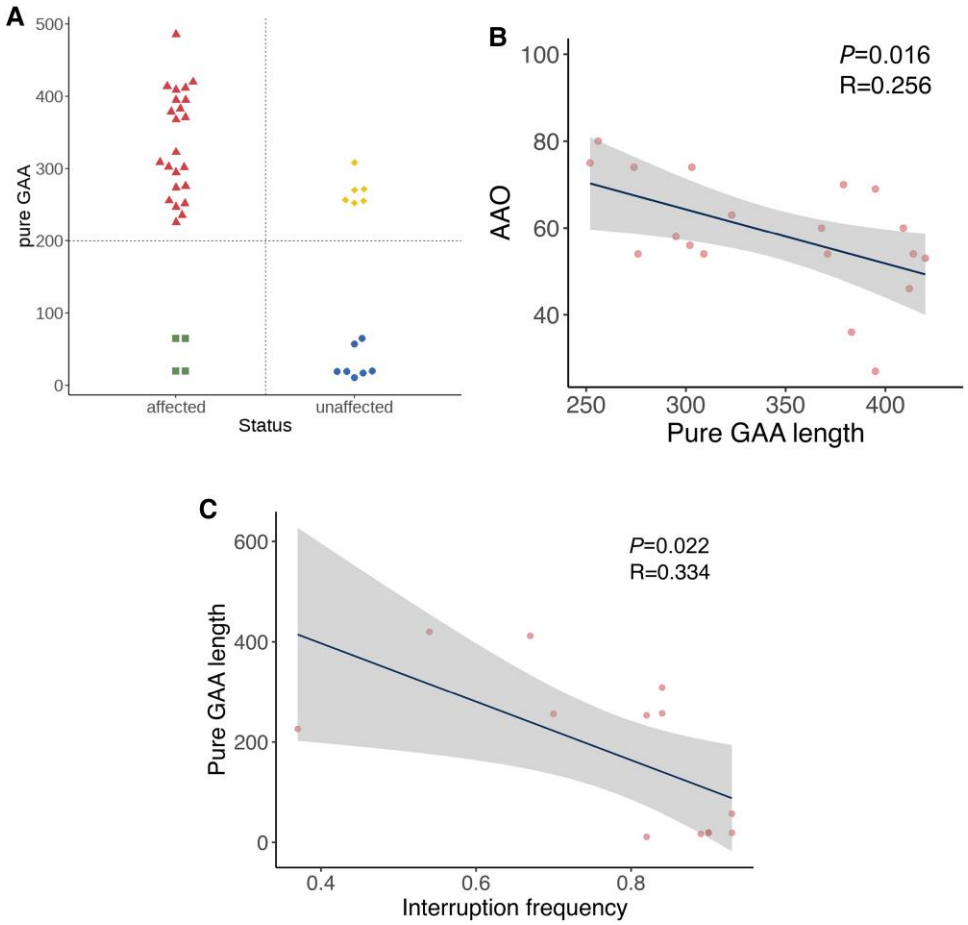

Figure 4 Analysis of the pure GAA length, determined by nanopore sequencing. (**A**) Differences between affected and unaffected repeat expansion carriers in pure GAA length. Shape coding: triangle = affected; diamond = pre-manifesting; circle = unaffected; square = other ataxias. (**B**) Correlation 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 $\alpha = 0.013$.

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 was not correlated with complex motifs or 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 AAO in a much more common condition, SCA27B. In XDP, the *TAF1* insertion is present only 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, because 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 might 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 pre-manifesting 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 to have SCA27B and carrying a large repeat interruption with <200 uninterrupted GAA repeats, the diagnosis needs to be revisited and probably 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 included only individuals

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	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	GAAAAGAAGAAGGAA	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	GAAAAGAAGAAGGAAGAAGGAA	PacBio, Sanger
L-10410	Unaffected	36493768	281	292	–	–	243	GAAAAGAAGAAGGAAGAAGGAA	Sanger
L-14996	Unaffected	37246629	309	309	–	–	66	GCAGAAGAAGAAGAA	–

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

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	GCAGAAGAAGAAGAA	PacBio, Sanger
L-14997	Patient	37246629	309	309	–	–	66	GCAGAAGAAGAAGAA	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	GAG	PacBio, Sanger

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

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, because this group might also harbour uninterrupted GAA repeat stretches of ~200 repeats in length. This might significantly broaden the group of *FGF14* GAA repeat expansion carriers who will be likely to manifest disease but are currently not considered to be at risk, according to present published recommendations.

Otherwise, our data imply that not all interruptions can be considered non-pathogenic, because, for instance, five ataxia patients with interruptions still had a remaining pure GAA expansion of >200. Given that many studies screened unspecified ataxia cohorts

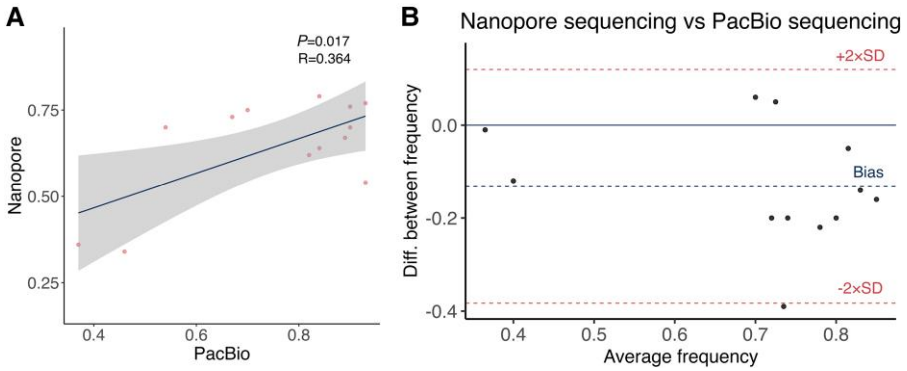

Figure 5 Comparison of the interruption frequency of all motifs between nanopore and PacBio sequencing. (**A**) Correlation between Nanopore and PacBio sequencing. (**B**) Bland–Altman plot between Nanopore and PacBio sequencing. A linear regression model was used for statistical analysis.

for repeat expansions in *FGF14* to debunk the clinical phenotype of SCA27B without considering interruptions, our data demonstrate that 14% (4 of 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 with or without 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 relationships, not only regarding the AAO but also in the context of disease severity and progression. We suggest that it is feasible to use long-read sequencing in *FGF14* repeat expansion disorders in future clinical practice.

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 also assume a similar AAO-modifying effect by mDRILS in *FGF14*, with a higher mDRILS frequency resulting in shorter uninterrupted GAA tracts, which, in turn, are associated with a later AAO (Fig. 3B). Given that a relationship between repeat length and AAO 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 to better detect phenotypic correlations and impact on repeat stability.[15] The mDRILS in both 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).[29] As with other disease-associated repeat expansions, including the GAA expansions causing Friedreich's ataxia,[10,30] SCA27B-associated *FGF14* repeat expansions arise on specific haplotypes and can have interruptions in the repeats, suggesting the contribution of *cis*-elements both flanking and within the unstable repeat, respectively.[1,11,31,32] 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 might predispose to parent-to-offspring (germline) repeat expansions.[1,11,31,32] Possible mechanistic paths, by which such flanking sequence changes might act, might 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 postnatal tissues. We did not observe an association of the motif within the 5′-flanking region of the short allele and interruption motifs. However, our sample size is small, and a larger number of 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 might act to modulate repeat instability, germline or somatic instability, can be 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 that differences in chromatin compaction might 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 might alter formation of RNA foci 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 might be targeted therapeutically; for instance, these pathogenic structures can be bound by the ligands TMPyP4 and BRACO-19.[40,43] 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.[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 in families affected by XDP, providing evidence that a higher

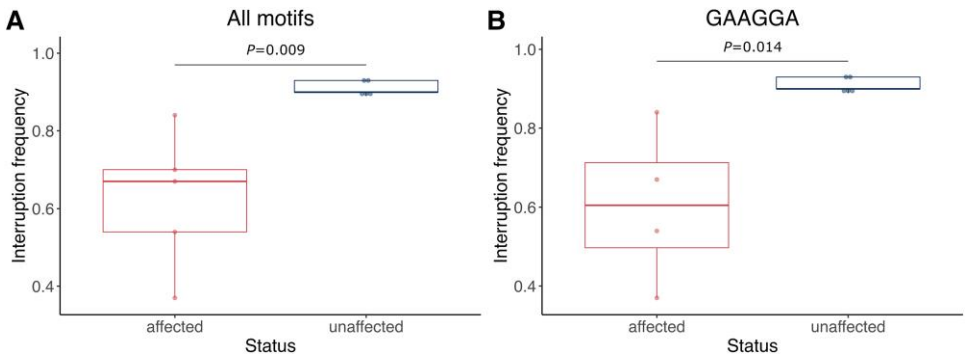

Figure 6 Analysis of the interruption frequency. Differences between affected and unaffected individuals in interruption frequency of all motifs (**A**) and the $(GAAGGA)_n$ motif (**B**). Mann–Whitney U-tests were performed for statistical analysis. The adjusted significance level is $\alpha = 0.013$.

frequency might stabilize the repeats across generations.[51] Thus, future studies should aim for an analogous exploration of *FGF14* mDRILS across time within individual patients and across generations.

Overall, all methods (RP-PCR, Sanger sequencing, Nanopore sequencing and PacBio sequencing) were concordant and detected the same interruptions. We identified confidently 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 both unaffected and ataxia patients but have not yet explored the position of interruptions and their impact on repeat expansion pathogenicity at the individual level.[6]

Limitations of our study include the relatively small sample size in comparison to other studies. Moreover, we cannot support our estimate of pathogenicity based on advanced genetic methods independently, because we do not investigate *FGF14* expression, and further biomarkers robustly distinguishing SCA27B from other ataxias are not available to date. Thus, the development of such, preferably accessible, biomarkers would be highly beneficial to the field. Future studies with larger sample size, longitudinal examinations and more in-depth clinical data are needed to investigate not only the influence of pure GAA repeat number on AAO but also a potential influence on disease severity and progression. Monitoring the mDRILS and their impact on disease progression and severity over a lifetime would be valuable. 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.

Conclusion

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. Although 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 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 Professor 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 created, in part, 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. J.T. is a consultant to Acurex. The remaining authors report no competing interests.

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; 110:105-119.
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;388:128-141.
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;273:14588-14595.
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: 1575-1577.
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.
6. Mohren L, Erdlenbruch F, Leitao E, et al. Identification and characterisation of pathogenic and non-pathogenic *FGF14* repeat expansions. *Nat Commun*. 2024;15:7665.
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;11:96-104.
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;32:1-27.
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;25:476-499.
10. Cleary JD, Pearson CE. The contribution of *cis*-elements to disease-associated repeat instability: Clinical and experimental evidence. *Cytogenet Genome Res*. 2003;100:25-55.
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;56:1366-1370.
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: 897-909.e21.
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;33:1108-1118.
14. Trinh J, Luth T, Schaake S, et al. Mosaic divergent repeat interruptions in XDP influence repeat stability and disease onset. *Brain*. 2023;146:1075-1082.
15. Paulson H. Repeat expansion diseases. *Handb Clin Neurol*. 2018; 147:105-123.
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;146:4144-4157.
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;30:3828-3833.
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;271:1023-1027.
19. Milovanovic A, Dragasevic-Miskovic N, Thomsen M, et al. *RFC1* and *FGF14* repeat expansions in Serbian patients with cerebellar ataxia. *Mov Disord Clin Pract*. 2024;11:626-633.
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;38:1107-1109.
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;42:128-128k.
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;11:885-897.
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;13:126.
24. Li H. Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*. 2018;34:3094-3100.
25. Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25:2078-2079.
26. Harris RS, Cechova M, Makova KD. Noise-cancelling repeat finder: Uncovering tandem repeats in error-prone long-read sequencing data. *Bioinformatics*. 2019;35:4809-4811.
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;38:1557-1558.
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(®)*. University of Washington, Seattle; 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.
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.
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;25:393-403.
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;95:1187-1195.
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:e1000257.
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;12:41-50.
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;280:4498-4503.
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.
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;434:29-34.
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: 488-505.
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: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; 299:105202.
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: 4361-4374.
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;52:2698-2710.
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: 28054.
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;300:107138.
45. Stochmanski SJ, Therrien M, Laganiere J, et al. Expanded *ATXN3* frameshifting events are toxic in *Drosophila* and mammalian neuron models. *Hum Mol Genet*. 2012;21:2211-2218.
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;9:1957-1966.
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;14:2649-2660.
48. Aviner R, Lee TT, Masto VB, Li KH, Andino R, Frydman J. Polyglutamine-mediated ribotoxicity disrupts proteostasis and stress responses in Huntington's disease. *Nat Cell Biol*. 2024;26:892-902.
49. Stein KC, Morales-Polanco F, van der Lienden J, Rainbolt TK, Frydman J. Ageing exacerbates ribosome pausing to disrupt cotranslational proteostasis. *Nature*. 2022;601:637-642.
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;273:26998-27008.
51. Laß J, Lüth T, Schlüter K, et al. Stability of mosaic divergent repeat interruptions in X-linked Dystonia-Parkinsonism. *Mov Disord*. 2024;39:1145-1153.