Unveiling pathogenic and non-pathogenic FGF14 repeat expansions: identification, sequence, and secondary structure

Christel Depienne

christel.depienne@uk-essen.de

University of Duisburg-Essen https://orcid.org/0000-0002-7212-9554

Lars Mohren

University Hospital Essen, University Duisburg-Essen

Friedrich Erdlenbruch

Department of Neurology and Center for Translational Neuro- and Behavioral Sciences (C-TNBS), Essen University Hospital, University of Duisburg-Essen, 45147 Essen, Germany. https://orcid.org/0009-0006-0654-836X

Elsa Leitão

University of Duisburg-Essen, University Hospital Essen https://orcid.org/0000-0001-5051-9714

Fabian Kilpert

University Hospital Essen and University of Duisburg-Essen

G. Sebastian Hönes

University Hospital Essen, University Duisburg-Essen

Andreas Thieme

University Hospital Essen, University Duisburg-Essen

Marc Sturm

Institute of Medical Genetics and Applied Genomics, University of Tübingen https://orcid.org/0000-0002-6552-8362

Joohyun Park

University of Tübingen

Agatha Schlüter

Institut d'Investigació Biomédica de Bellvitge

Montserrat Ruiz

Institut d'Investigació Biomédica de Bellvitge https://orcid.org/0000-0003-0466-2653

Moisés Morales de la Prida

Institut d'Investigació Biomèdica de Bellvitge (IDIBELL) https://orcid.org/0000-0002-9184-529X

Carlos Casasnovas

Hospital Universitari de Bellvitge https://orcid.org/0000-0003-1170-2676

Kerstin Becker
Cologne Center for Genomics, University Hospital Cologne https://orcid.org/0009-0009-7897-7181
Ulla Roggenbuck
Biometry and Epidemiology, University Hospital Essen, University of Duisburg-Essen
Sonali Pechlivanis
Biometry and Epidemiology, University Hospital Essen, University of Duisburg-Essen
Frank Kaiser
University of Duisburg-Essen
Matthis Synofzik
https://orcid.org/0000-0002-2280-7273
Thomas Wirth
Hôpitaux Universitaires de Strasbourg
Mathieu Anheim
University Hospital of Strasbourg
Tobias Haack
University of Tübingen
Paul Lockhart
Murdoch Childrens Research Institute https://orcid.org/0000-0003-2531-8413
Karl-Heinz Jöckel
University Hospital Essen, University of Duisburg-Essen
Aurora Pujol
ICREA/ IDIBELL https://orcid.org/0000-0002-9606-0600
Stephan Klebe
University Hospital Würzburg
Dagmar Timmann
University Hospital Essen, University of Duisburg-Essen, Germany
Sabine Kaya
University Hospital Essen, University Duisburg-Essen
Christopher Schroeder

Article

Keywords:

Posted Date: February 28th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3940197/v1

Additional Declarations: There is **NO** Competing Interest.

Abstract

Repeat expansions in *FGF14*cause autosomal dominant late-onset cerebellar ataxia (SCA27B), with pathogenic thresholds estimated between 250 and 300 AAG repeats. However, the complete sequences of pathogenic and non-pathogenic alleles remain largely unknown. Here, we identify *FGF14* repeat expansions as the most commonly missed cause of cerebellar ataxia using a combination of short-read genome sequencing, long-range PCR and nanopore sequencing. We compare *FGF14* alleles in 154 patients with ataxia and 802 controls and report an enrichment of pure *FGF14* expansions above 180-220 repeats in patients. Conversely, AAGGAG and interrupted alleles are more frequent in controls. Distinct 5'-flanking regions correlate with repeat stability. SCA27B patients have a characteristic phenotype, including frequent episodic ataxia and downbeat nystagmus, similar to those with a novel nonsense variant (SCA27A). Interestingly, pathogenic and non-pathogenic repeats form different secondary structures. Our results emphasize the importance of sequencing both the repeat expansion and its flanking region for accurate clinical interpretation.

Introduction

Spinocerebellar ataxias (SCAs) are a group of autosomal dominant, slowly progressive disorders characterized by impaired coordination and imbalance resulting in ataxia of stance and gait, limb ataxia, dysarthria and oculomotor signs[1]. Cognitive impairment, tremor, rigidity, bradykinesia, dystonia, spasticity, and polyneuropathy are frequently associated features. So far, at least 40 different SCA subtypes, classified according to their underlying locus/genetic cause, have been reported[2]. This list includes repeat expansions of CAG trinucleotides encoding polyglutamine (PolyQ) repeats (SCA1, 2, 3, 6, 7, 12, 17) as well as noncoding repeat expansions of penta- or hexanucleotides (SCA10, 12, 31, 36, 37). Furthermore, point variants have been described in at least 28 distinct genes[1,2]. Among those, rare point variants or microdeletions leading to the heterozygous loss-of-function of *FGF14* had previously been reported as SCA27 (renamed SCA27A)[3–5].

More recently, heterozygous noncoding GAA/CTT repeat expansions in *FGF14* have been identified as a frequent cause of late-onset cerebellar ataxia (SCA27B) in Canada, Australia, and Europe[6–8]. *FGF14* encodes the fibroblast growth factor 14, a gene expressing at least eight different isoforms, according to the Ensembl database. The Matched Annotation from the NCBI and EMBL-EBI (MANE) isoform (NM_004115.4), also called transcript 1, encodes a 247-amino acid (27 kDa) protein that is considered as the canonical sequence in Uniprot. However, this transcript is poorly expressed in all tissues according to the Genotype-Tissue Expression (GTEx) database. The most abundant transcript (transcript 2; NM_175929.3) is brain-specific, with the highest expression in the cerebellum, and encodes a 252-amino acid (28 kDa) protein. The two protein isoforms differ in their N-terminus as a result of distinct transcription start sites, leading to the inclusion of different first exons[9]. The N-terminus of isoform 1 contains a nuclear localization signal and is predicted to localize in the nucleus. In contrast, isoform 2 localizes at the axon initial segment of cerebellar Purkinje neurons[10], where it regulates the activity of

voltage-gated sodium[9] and potassium[11] channels. The repeat expansion responsible for SCA27B is located in intron 1 of isoform 2. The expansion is usually described as a GAA (or AAG) repeat expansion according to its genomic context, but as the gene is on the reverse strand, the expansion consists of CTT repeats in the gene context. The expansion occurs at a short tandem repeat highly polymorphic in humans and is associated with a probable founder effect in Canada, with an incidence up to 60% of families in this population[6]. The comparison of expansion sizes in patient and control populations has suggested that expansions above 300 AAG repeats are pathogenic with full penetrance while expansions between 250 and 299 repeats are associated with incomplete penetrance[6,7]. The consequence of the expansion is a reduction of *FGF14* expression from the expanded allele, likely leading to the heterozygous loss-of-function (haploinsufficiency) of isoform 2 in the cerebellum[6].

In this study, we identify *FGF14* repeat expansions as the most frequently missed genetic cause of cerebellar ataxia in two cohorts of patients, one with available genome data, and the other one with inconclusive diagnostic analyses. This finding led us to analyze *FGF14* alleles in a total of 988 individuals, including 154 patients from 134 independent families with cerebellar ataxia of unknown cause, 32 patients with other neurological disorders, and 802 control subjects, using a combination of long-range PCR, fluorescent gene fragment analysis and targeted nanopore sequencing. Our results provide molecular, phenotypic and mechanistic insights into how pure AAG repeat expansions lead to adult-onset cerebellar ataxia while expansions of other motifs or interrupted expansions are non-pathogenic.

Results

Identification of FGF14 expansions from genome data

We used STRling to identify STR expansions in short-read genome data of 80 patients with neurological disorders (76 from Germany and 4 from Spain), including 48 patients with cerebellar ataxia from 39 independent families (Fig. 1). This analysis detected outlier values associated with significant q values indicating possible *FGF14* AAG repeat expansions (chr13(hg38): 102,161,567 – 102,161,726) in 22 of the 48 patients with ataxia (19/39 families; 49%). Significant values were detected for both individuals of four families for which genome data were available. In contrast, only three of the 32 individuals with other neurological disorders (3/31 families; 10%) showed significant outlier values. In addition, STRling detected an expansion of another motif (AAGGAG) at the same locus in one family from Spain.

We set up LR-PCR and RP-PCR assays to validate these results. The existence of at least one large *FGF14* AAG allele (PCR product ≥ 700 bp; triplet repeat number ≥ 180) was confirmed for 18 of the 26 individuals, all with cerebellar ataxia. The eight remaining individuals, including the three with another neurological disorder, had a larger allele below 650 bp (i.e. ≤160 repeats) and were therefore considered negative for *FGF14* expansion/SCA27B. Of note, we also confirmed that all patients that had no outlier values detected by STRling only had small *FGF14* alleles (Supplementary Fig. 1; Supplementary Data 1).

Screening of FGF14 expansions in a second cohort

We then used the LR-PCR and RP-PCR assays to analyze 106 additional patients (95 new index cases and 11 family members) with cerebellar ataxia (Fig. 1, Fig. 2, Fig. 3A and Supplementary Fig. 2). Twenty-seven of the 95 index cases (28%) and five family members had at least one AAG allele ≥ 700 bp. Taken together with the expansions detected from genome data, 49 individuals from 40 families out of 134 independent families analyzed (30%) had an expanded AAG allele (≥ 180 repeats) estimated from fragment size or gel analysis. The AAG expansions segregated with the disorder in all families including at least two affected individuals available for genetic analysis (Fig. 2; Supplementary Fig. 3A). Four patients from three families had an expansion of the AAGGAG motif (Supplementary Fig. 2B). The AAGGAG expansion segregated in two affected individuals from a Spanish family but was present in only the index case but not the affected mother of a German family (Supplementary Fig. 3B) and was further considered as non-pathogenic, as already reported[6,7,17,18].

Targeted Nanopore Sequencing of FGF14 expanded alleles

Since LR-PCR and RP-PCR failed to provide the precise count of AAG repeats at the *FGF14* locus, we developed a nanopore sequencing assay of LR-PCR amplicons. In parallel, we analyzed and compared the distribution of *FGF14* alleles in a control cohort composed of 802 subjects. In total, we sequenced *FGF14* alleles in 59 patients and 64 control individuals (Fig. 1; Fig. 3B-C). We observed a strong correlation between allele sizes determined by fragment analysis or gel electrophoresis and the median number of repeats calculated from nanopore data (Fig. 3D). Combining fragment analysis and nanopore sequencing of LR-PCR amplicons provided a comprehensive view of allele sizes in both patients and controls.

Overall, thirty-eight patients from 31 families (31/134; 23%) had at least one allele composed of pure AAG repeats exceeding the current established threshold for pathogenicity (250 repeats) and were considered as having SCA27B (Fig. 2). The median number of repeats calculated from nanopore reads ranged from 254 to 937 AAG repeats (Supplementary Data 2). Twenty-six patients had a repeat number above 300 repeats (Fig. 2A) whereas 12 patients had a repeat number between 250 and 299 (Fig. 2B). Nevertheless, we noted that the thresholds of 250 and 300 repeats appeared somewhat arbitrary. For instance, in one family, two sisters with similar symptoms and age at onset (AAO) had median repeat numbers of 401 and 281 repeats, respectively (Fig. 2A). Similarly, in another family, the index case had 258 repeats and his affected mother had 224 repeats (Fig. 2B).

Five patients from four independent families showed biallelic repeat expansions (Fig. 3A-B). In three families, the largest allele was below 300 repeats and the lowest allele below 250 repeats: 222/278, 196/284, 165/272. The fourth family included three affected siblings; two siblings had 204/311 and 196/319 repeats whereas their affected sister only had one large pathogenic allele (325 repeats). One patient repeatedly exhibited two large alleles, one with ≥ 300 repeats and another between 250 and 300 repeats, in addition to a small allele, suggesting somatic variability of the expanded allele (Fig. 3A-B).

More generally, a high degree of somatic mosaicism around the mean value was detected for all individuals with repeat expansions, as highlighted by the positive correlation between the standard deviation and the allele size (Fig. 3E).

Distribution of FGF14 alleles in patients and controls

The 802 control subjects showed an overall different distribution of *FGF14* alleles compared to patients with ataxia (Fig. 4A-B), especially when considering only the larger allele (Supplementary Fig. 5B; Mann-Whitney, p = 0.0002). Large alleles composed of pure AAG ≥ 180 repeats were enriched in patients with ataxia with alleles ≥ 250 repeats showing a more significant enrichment than intermediate alleles (Fig. 4C-D). Conversely, AAGGAG expansions were more frequent in controls (Fisher's test: $p = 0.02$, OR 3.8; Fig. 4A). We observed true AAG interruptions (disrupting repeats in the middle) in smaller alleles only, while interruptions limited to 3' or 5' sides of the repeats were equally frequent in both alleles (Fisher's test: $p = 0.47$) but more frequent in control subjects than patients with ataxia (Fisher's test: $p = 0.01$, OR 4.2; Fig. 4; Supplementary Fig. 4). Out of 21 control subjects who had at least one allele above 250 repeats, 13 were composed of the non-pathogenic AAGGAG repeat motif (Supplementary Fig. 4; Supplementary Fig. 5). Eight control individuals had repeat expansions ≥ 250 AAG repeats and only two, aged 46 years old and 70 years old at the time of sampling, had a pure AAG repeat expansion above 300 repeats (313 and 319 repeats respectively; Supplementary Data 2).

Variability of the flanking region

Interestingly, the 5' region flanking the repeats (3' region in the context of the gene) was highly variable and drastically differed in expanded versus non-expanded alleles (Fig. 5A). We observed seven different sequences following a constant CTTTCT motif (chr13:102,161,558 – 102,161,563) upstream of the repeats. These 5'-flanking sequences were either directly followed by AAG repeats, or preceded by short AG-rich sequences (e.g., AAGAAAGAG or AAGAG) that we considered as 'pre-repeat' (Fig. 5B; Supplementary Fig. 6A-B). Pre-repeats were more frequently observed in larger (a2) alleles (Fisher's test: $p = 3.227\times10^{-8}$, OR 7.5; Fig. 5C; Supplementary Fig. 6C-D). The variable GTG sequence (present in the hg38 reference genome) was the most frequently associated with *FGF14* expansions although the range of repeats observed was highly variable (range: 36–512). GG and GGG sequences were only detected in expanded alleles (326–937 repeats). Conversely, the GTTAGTCATAGTACCCC was strikingly associated with small alleles (9–21 triplet repeats) only. Interestingly, one nearly identical sequence, differing only in the final two nucleotides (GTTAGTCATAGTACCAG), was associated with 203/207 repeats in both affected individuals of family SPG-79. This suggests that the four consecutive cytosines at the end of the sequence play a key role in preserving the stability of the adjacent repeats.

Intermediate FGF14 alleles

Eleven patients exhibited *FGF14* AAG repeat expansions ranging from 180 to 249 repeats (Supplementary Fig. 3A). According to the current threshold of 250 repeats for SCA27B diagnosis, these patients are

classified as negative. For instance, in family E19-0805, the affected mother (M79607) had a median repeat value of 224, prompting further investigation into alternative genetic causes for her ataxia. Despite analyzing genome data from both family members, no other ataxia-associated variants were identified, suggesting the *FGF14* expansion as the primary culprit. Our exploration revealed that alleles below 250 repeats could contribute to disease, as evidenced by segregation in a Spanish family with two affected individuals (SPG-79; Supplementary Fig. 3A). In another case (M91143), a female patient harbored a variant of unknown significance in *PUM1* (NM_001020658.2: c.2180T > C; p.(Ile727Thr)) alongside 236 AAG repeats in *FGF14*. The remaining families had no identified variants in ataxia-associated genes. Among patients without genome data, one male patient (M87909) had a pathogenic SCA6 expansion (21 repeats) alongside 209 repeats in *FGF14*. This suggests the potential involvement of additional genetic or non-genetic factors in the disease manifestation, with *FGF14* intermediate alleles likely acting as susceptibility factors.

Clinical comparisons

We divided patients with cerebellar ataxia into four distinct groups for clinical comparison: 1) patients with a median number of AAG repeats ≥ 300 ($n = 26$); 2) patients with 250–299 median repeats ($n = 12$); 3) patients with an intermediate allele (180–249; $n = 10$); and 4) patients negative for SCA27B ($n = 90$; Supplementary Data 3 and 4). Furthermore, we included the clinical data of a German family (affected father-daughter pair) with a novel pathogenic nonsense variant in *FGF14* (SCA27A;NM_175929.3:c.239T > G; p.(Leu80*);NM_004115.4(MANE):c.224T > G; p.(Leu75*)) identified by routine exome sequencing (Fig. 2C; Fig. 8A-B).

Overall, most patients with *FGF14* expansion ≥ 250 repeats (29/38; 76%) had a highly recognizable phenotype, characterized by the association of slowly progressive cerebellar signs accompanied by episodic symptoms of ataxia and/or downbeat nystagmus (DBN) that often present as first symptoms (Supplementary Data 3). Nine patients exhibited cerebellar symptoms without episodic features or downbeat nystagmus (DBN). Six patients displayed a phenotype with additional signs or a distinct disease course. Detailed case reports of these individuals (clinical outliers; #1–6 on Fig. 7F-G) are available in Supplementary data.

Patients with more than 300 repeats and patients with 250–299 repeats exhibited comparable clinical characteristics that slightly differed from *FGF14*-negative patients with ataxia. For example, we observed a significantly higher occurrence of early cerebellar oculomotors signs (95%; 92% and 96%, respectively, compared to 63% in *FGF14*-negative patients). In particular, DBN was present in 47% (50% and 46%) at the first examination, compared to only 3% in patients negative for *FGF14* (Fig. 6A). Patients with SCA27B had a lower occurrence of dysarthria on the first examination (24%; 25% and 23%, respectively) compared to 62% in the *FGF14* negative group; Fig. 6A). There was less cognitive impairment in patients with *FGF14* expansions compared to the *FGF14* negative group either at the first or last examinations (71% of negative patients had cognitive impairment versus 38% in the *FGF14*-positive group; Fig. 6B). These statistical differences remained even when removing the six patients behaving as clinical outliers.

The progression of SCA27B ataxia, assessed through SARA and ICARS scores, was generally slow, with a mean increase in SARA scores of 8.7 points over 30 years (Fig. 7F-G). Excluding the six patients considered as 'clinical outliers' resulted in even slower progression (4.5 points over 30 years; Supplementary Fig. 7D-E). However, variability was high in both SCA27B groups (250 ≤ repeats < 300 and ≥ 300 repeats). The linear regression model suggests that disease duration accounts for only 7.9% of observed variance, indicating that other factors exert a more substantial impact. Interestingly, a substantial proportion of SCA27B patients reported worsening symptoms in the mornings (67%; 71% and 64% in *FGF14*-positive groups compared to 25% in the *FGF14*-negative group; Fig. 6D). More than half of the patients who received 4-aminopyridine/fampridine reported symptom improvement (57%; 62% and 53%, respectively; Supplementary Fig. 7F-G). Treatment response to acetazolamide was also reported (29%; 0% and 50%, respectively).

The patient harboring the pathogenic nonsense variant in *FGF14* (p.Leu80*; M96962) exhibited symptoms very similar to the groups of patients with ≥ 250 repeats. He had a slowly progressive cerebellar syndrome, episodic worsening of cerebellar symptoms and DBN. In addition, he also experienced episodic dystonia of the left hand. While he developed his first symptoms at the age of 36, his daughter had tremors at the age of 5 years and developed gait instability around 30 years.

Correlations between repeat number and age at onset (AAO)

The mean AAO in our combined cohorts was 55.2 years (range 33–76; Table 1; Fig. 7A) in patients with 250–299 repeats and 51.0 years (21–75 years) in patients with ≥ 300 repeats, whereas patients of the negative group had a mean AAO of 46.0 years (1–82 years). This difference was, however, not significant because variation in all groups was high. Seven patients presented with an earlier form of the disease, with an onset before 40 years old (21–38); all but two had ≥ 300 repeats. Accordingly, we observed a significant inverse correlation between the number of AAG repeats and the AAO (Fig. 7D; Supplementary Fig. 7A). Nevertheless, 81% of the variation is independent of the number of AAG repeats (R^2 = 0.19). The variability of the AAO was illustrated by patient M90982 (258 repeats) who started to show symptoms at 34 years old whereas individual M91231 (259 repeats) only experienced the first symptoms at the age of 72 years.

Table 1
Mean age at disease onset stratified by *FGF14* genetic status.

Patient group	This study			Meta-analysis		
	n	Mean AAO	AAO Range	*n*	Mean AAO	AAO Range
AAG expansion negative	90	46	1–82	-	-	-
AAG expansion 180 ≤ exp < 250	10	53	9–72	-	-	-
AAG expansion 250 ≤ exp < 300	12	55.2	33–76	35	62.5	33–79
AAG expansion exp ≥ 300	26	51	21–75	104	59.1	21–78
AAG expansion exp ≥ 250	38	52.3	21–76	139	60.0	21–79
Truncating	2	20.5	5–36	7	19.4	2–47
Missense (p.Phe150Ser)	-	-	-	24	20.5	6–40

n: number of patients; AAO: age at onset

We conducted a meta-analysis to assess the correlation between AAO and expansion size, pooling data from our study and four prior studies.[7,19–21]. We also included data of patients with truncating variants[4,5,22,23] or the F150S (F145S in MANE isoform 1) missense variant in *FGF14*[3,24] in the comparison (Table 1; Supplementary Data 5). Overall, we confirmed a significant inverse correlation between AAO and expansion size (Fig. 7E; supplementary Fig. 7B). The aggregated data also showed a tendency for patients with *FGF14* expansions between 250 and 299 repeats to be later affected on average (62.5 years; $n = 37$) than patients with ≥ 300 repeats (59.1 years; $n = 110$), but despite increased statistical power, the difference remains not significant due to high variation in both groups (Fig. 7C). Nonetheless, patients with expansions from both groups exhibited significantly later AAO compared to those with truncating (19.4 years, range: 2–47) or F150S (20.5 years, range: 6–40) variants (Fig. 7C).

Frequency of SCA27B in the German cohort

To assess the prevalence of SCA27B relative to other dominantly inherited ataxia subtypes, we compared the number of patients diagnosed in the ataxia outpatient clinic (Department of Neurology, University Hospital Essen). Among the diagnoses, SCA6 (*CACNA1A*) accounted for 40 patients, SCA27B for 36 patients, SCA3 (*ATXN3*) for 35 patients, SCA1 (*ATXN1*) for 17 patients, Episodic Ataxia 2 (point variant in *CACNA1A*) for 12 patients, SCA14 (*PRKCG*) for 7 patients, SCA2 (*ATXN2*) for 6 patients, and SCA8 (*ATXN8OS*) for 4 patients. Additionally, rarer forms such as SCA28 and SCA49 (two each), SCA7, SCA13, SCA15, and SCA27A (one each) were observed. This indicates that SCA27B ranks among the most prevalent SCA subtypes, representing approximately one-fifth of all diagnoses in patients with cerebellar ataxia. This finding independently supports the high frequency of SCA27B diagnoses observed in another German cohort.[25]

Secondary structures associated with FGF14 expansions

We used CD spectroscopy to assess the potential of secondary structure formation of the different *FGF14* antisense repeat expansions AAG and AAGGAG as well as the complementary sense sequences CTT and CTCCTT on DNA and RNA. The AAG-DNA 25-mer formed an antiparallel homoduplex while CD spectroscopy of the AAGGAG-DNA oligo revealed formation of a parallel homoduplex[26–29](Fig. 8C). The RNA counterparts CUU and CUCCUU did not obtain any secondary structure under the tested conditions, confirmed by a single positive band at 270 nm[30] (Fig. 8C). Interestingly, the non-pathogenic AAGGAG-RNA oligo folded into a parallel guanine-quadruplex (G4) with a positive band around 260 nm, a negative band around 240 nm and positive values around 210 nm. Presence of a G4 was further confirmed by a G4-specific decrease in the stability detected, shifting from a parallel G4 structure (with 100 mM K) to a hairpin structure (with 100 mM Li)[31,32] (Fig. 8C). Of note, for the pathogenic AAG-RNA repeat we detected a CD spectrum related to an A-form RNA structure, with a negative peak around 210 nm that reflects intra-strand interaction of RNA duplexes[33].

Discussion

In this study, we identify intronic *FGF14* AAG expansions (SCA27B) as a major genetic contributor to cerebellar ataxia that had previously been overlooked by short-read technologies. Bioinformatics tools such as STRling[13] can be used to detect this expansion from short-read genome data with a good predictive value (70%) and excellent sensitivity (100%; Supplementary Fig. 2). Although STRling largely underestimates the repeat count, the identification of a significant outlier value is an indicator of a possible expansion, requiring validation through an alternative method such as LR-PCR, RP-PCR and/or targeted long-read sequencing. A tool that can easily be implemented in routine diagnosis is ExpansionHunter[14] using the STRipy interface[15]. When used with off-target/extended mode, this tool provides a closer estimation of repeat numbers but the values cannot be fully trusted and need validation by a more reliable method.

The frequency of pathogenic expansions found in our combined ataxia cohorts, considering the previously established threshold of 250 repeats, was remarkably high (23%), equivalent to or even higher than previous reports[6,7,19–21,25,34–36]. The difference in incidence is possibly linked to the population studied but it could also result from the sensitivity of the genetic test used. We improved the LR-PCR conditions to minimize the amplification bias towards the smaller allele usually observed under standard conditions. This method could detect the largest expansion (937 repeats) reported so far. Current diagnostic strategies for SCA27B rely on sequential genetic tests based on the detection of normal and intermediate alleles (1000 bp being the usual limit of detection using fragment size analysis) combined with RP-PCR[37]. However, due to their inherent limitations, both tests may miss expansions that are unusually large and/or composed of other repeat motifs. A complete sequencing of *FGF14* alleles should thus be part of routine diagnostic procedures. Targeted nanopore sequencing is a cost-effective method that can easily be implemented to sequence *FGF14* expansions. Its known drawback is the higher error

rate compared to other sequencing methods. However, errors can usually be easily distinguished from true sequence variations if the sequencing depth is sufficient, based on their random occurrence in single reads (errors) or presence in multiple reads (true variants). For example, nanopore sequencing leads to systematic errors (AAC instead of AAG) at the end of long reads. Moreover, smaller DNA fragments are preferentially sequenced by this technology. Since patients with *FGF14* repeat expansions show a high somatic variability that is correlated with expansion size, preferential sequencing of small fragments leads to underestimate repeat number in patients with large repeat expansions. To circumvent this problem, we used the median number of repeats instead of the mean, which is also closer to the size observed on gel.

The repeat expansion was mainly identified in patients with adult-onset cerebellar ataxia and was associated in 76% with a very recognizable phenotype that consisted of slowly progressive cerebellar ataxia with accompanying episodic symptoms of ataxia, often the initial manifestation, and/or downbeat nystagmus[38]. We found that episodic symptoms are not restricted to ataxia, but can also include episodic choreatiform movements, as described in *Fgf14* knock-out mice[39]. Mild pyramidal tract signs were observed in one patient. In three cases, cerebellar ataxia was more severe. In one patient with biallelic expansion, an anxiety disorder likely contributed. Additionally, marked cerebellar atrophy, uncommon in SCA27B, was observed in two cases, indicating potential secondary cerebellar disease of unknown etiology. Two patients developed Parkinsonism and dementia in later stages, possibly due to idiopathic Parkinson's disease and vascular encephalopathy, respectively. Of note, the patient with later onset of Parkinsonism and dementia previously had tentative clinical diagnosis of multiple system atrophy, cerebellar subtype (MSA-C). This would be the third patient with this diagnosis in whom a *FGF14* pathogenic expansion is identified[19,21]. However, in the context of our patient, a careful retrospective examination ruled out the MSA-C diagnosis and revealed that this patient actually had a typical SCA27B phenotype with later onset of second diagnoses that are frequent in elderly people. Secondary disease as a possible confounder for SCA27B with fast progression has already been discussed[19].

The mean AAO in our cohort was around 52.3 years, but it strikingly varied from 21 to 76 years. Although most patients have a late-onset of first symptoms, seven patients presented with an earlier form of the disease, with an onset before the age of 40. We observed a positive correlation between the number of AAG repeats and AAO, but the contribution of repeat number to AAO is rather limited. Accordingly, we observed an important individual variation independent of the number of AAG repeats, as illustrated by the two patients with 258 and 259 repeats, but started showing symptoms at 34 and 72 years of age.

The family history of patients with SCA27B implied incomplete penetrance, both in patients with 250–299 repeats and patients with ≥ 300 repeats. Incomplete penetrance above 300 repeats is further supported by the identification of pure repeat expansions above this threshold in two individuals of the HNR control population. Overall, the frequency of pathogenic *FGF14* expansions in the control population we studied, which is representative of the population from North Rhine-Westphalia, using 250 repeats as a threshold, is 7 out of 802 (0.87%). In particular, we observed two control individuals (46 and 70 years old) with a pure AAG repeat expansion ≥ 300 (313 and 319 repeats, respectively). We did not have details

about their neurological status and these individuals may develop symptoms of SCA27B later in life. But it also suggests that incomplete penetrance exists within this range of repeats and that the threshold for complete penetrance, if it exists, may thus be higher i.e. 320–335 repeats, as suggested by Rafehi *et al*.[7] Pellerin *et al*. demonstrated that *FGF14* repeat expansions tend to contract when inherited from the father, whereas they typically elongate when transmitted from the mother[6]. This aspect may at least partly explain the incomplete penetrance observed in some families and should be taken into account for genetic counseling.

We observed one family with a typical SCA27B phenotype in which the proband had 258 repeats and his affected mother 224 repeats. Genome sequencing did not reveal any other possible cause for the ataxia in this family. Furthermore, we also observed a significant enrichment of intermediate alleles (180–249 repeats) in patients with ataxia compared to the control population. This strongly suggests that expansions below 250 may predispose to SCA27B, as also already suggested by Pellerin *et al*.[40] and Hengel *et al*.[25] Among the eight families with intermediate alleles, two patients exhibited episodic symptoms and a possible SCA27B phenotype. However, the phenotypes appeared more heterogeneous, suggesting that some patients either had an additional or a different genetic cause. Additional causes include a pathogenic SCA6 expansion. Intermediate alleles could thus be contributing to the disease only in association with other factors. However, a specific analysis including a large number of such patients is needed to confirm or rule out the pathogenicity of intermediate alleles. Altogether, our observations currently support a reassessment of pathogenic thresholds to 220 for pathogenic alleles with incomplete penetrance and at least 320 for full penetrance. Genetic modifiers influencing the AAO and penetrance are likely multiple and possibly involve genetic variations at genes encoding DNA repair genes, like in other repeat expansion disorders[41]. Nevertheless, the observation of several cases with biallelic expansions also suggests that the size of the smaller allele and/or other eQTLs at the locus play an additive effect and partially determine the penetrance of alleles < 320 repeats.

The pathophysiological mechanism associated with *FGF14* expansions is likely a loss-of-function of isoform 2. This is supported by similar clinical features in patients with point mutations and repeat expansions. The p.Phe150Ser (F150S; F145S in isoform 1) variant, which was the first variant identified in *FGF14*,[3] also leads to the loss of interaction with voltage-gated channels and a loss of FGF14 capacity to control their activity[42]. However, patients with nonsense or the F150S variant are on average earlier affected, suggesting that the loss-of-function of isoform 2 may be incomplete as a result of the expansion (i.e. expanded alleles would be hypomorphic compared to point mutations). Another possibility is that loss-of-function of isoform 2 would occur only when the repeats extend beyond an even higher threshold in neurons, as a result of aging. This hypothesis could explain the clinical variability observed in patients and also that intermediate alleles have a probability to expand beyond the pathogenic threshold only in patients who have an overall lower control of microsatellite stability. It is also possible that the number of repeats detected in peripheral tissues such as blood might not always reflect the number of repeats present in the brain or cerebellum.

The somatic variability of pure AAG repeat expansions is remarkable. Although positively correlated with expansion size, we also see an individual variability of this phenomenon. In particular, somatic variability is influenced by the presence of interruptions and by the sequence of the 5' flanking region (Supplementary Fig. 5D; Fig. 5B). The association of specific flanking regions with repeat stability or instability has already been reported by Pellerin *et al*.[43] In this study, we extend this observation by showing that the four cytosines present in the flanking region associated with small alleles are crucial in this process. We hypothesize that these cytosines control the overall ability of the repeats to form secondary structures favoring repeat number amplification. This observation is to relate to secondary structures possibly formed by the repeats. We detected differences in the ability of pure pathogenic AAG and non-pathogenic AAGGAG repeats to form secondary structures at the DNA level but also at the RNA level. Noteworthy, the non-pathogenic AAGGAG RNA sequences can form a parallel G4 while pathogenic AAG repeats form an A-form RNA. G4 structures formed by other repeat expansions, including those found in *RFC1*/CANVAS, have been previously linked to a decrease in gene expression[44]. Furthermore, intronic AAG repeat expansions leading to Friedreich ataxia were shown to form R-loops that impair the transcription of *FXN*[45–47]. However, AAG is the repeat motif present in *FXN* whereas in *FGF14*, the repeats present at the RNA level are CUU repeats (Fig. 8B). Hence, in the context of *FGF14*/SCA27B, there is a potential scenario where the capacity of AAGGAG repeats to generate G4 structures serves as a protective mechanism against the formation of other structures, such as R-loops on the complementary strand. A more comprehensive investigation of how these secondary structures affect gene expression is essential for *FGF14*, but requires studying this effect in the appropriate tissue or cell type, given the predominant expression of this gene in the brain and cerebellum.

In conclusion, this study reveals the complete sequence of pathogenic and non-pathogenic *FGF14* expansions. We suggest that pure AAG expansions are pathogenic from a lower threshold (comprised between 180 and 220 repeats) and account for at least 23% of patients with adult-onset cerebellar ataxia in European populations while interrupted expansions or expansions composed of other hexanucleotide repeats motifs are non-pathogenic. For this reason, diagnostic tests should incorporate not only the assessment of repeat numbers but also include a comprehensive sequencing of the expansion.

Methods

Patients & subjects. Patient inclusion was part of the project "Identification of tandem repeat EXPAnsions in unsolved Neurological Disorders" (EXPAND). The study has received the approval of the ethics committee of University Hospital Essen (21-10155-BO). All patient and subjects' consent was obtained according to the Declaration of Helsinki. At the time of the study, 76 patients with neurological disorders remaining without any identified genetic cause after an exome or a genome analysis were recruited from the Department of Neurology, University Hospital Essen, and had their genome sequenced. Among these, 44 patients from 35 independent families had spinocerebellar ataxia as a main clinical feature. Additionally, data of four families with cerebellar ataxia from Spain were included in the analysis. The *FGF14* repeat expansion was screened for in an independent cohort of 95 index cases with cerebellar

ataxia without available genome data and 11 additional affected family members. One patient had a point variant in *FGF14* (NM_175929.3: c.239T > G; p.Leu80*) identified by routine exome sequencing. This study also included 802 control subjects: 30 anonymous blood donors and 772 participants from the Heinz-Nixdorf Recall (HNR) Study[12].

Genome sequencing & expansion calling. Short-read genome sequencing was performed at the Cologne Genome Center (Cologne, Germany) for 51 patients and as part of routine diagnosis at the Institute for Medical Genetics and Applied Genomics (University of Tübingen, Germany) for 25 patients. Libraries were prepared with the DNA tagmentation based library preparation kit (Illumina) without PCR, with 300–500 ng genomic DNA input. Library preparation was followed by clean up and/or size selection using SPRI beads (Beckman Coulter Genomics). After library quantification (Qubit, Life Technologies) equimolar amounts of library were pooled. The library pools were quantified using the Peqlab KAPA Library Quantification Kit and the Applied Biosystems 7900HT Sequence Detection System and then sequenced on an Illumina NovaSeq6000 sequencing instrument with a paired-end 2x150bp protocol. Fastq data were mapped to the hg38 reference genome using an in-house pipeline: raw sequencing data underwent preprocessing using cutadapt[48] to remove adapter sequences. Read mapping used bwa-mem[49], bwa-mem 2[50], and bwa-meme[51]. Duplicate reads were removed with samblaster[52]. Sorted and indexed CRAM files were generated by samtools[53]. In an initial phase, we compared the ability of ExpansionHunter DeNovo[54] and STRling[13] to detect known repeat expansions from short-read genome data including TTTTA/TTTCA repeat expansions in *MARCHF6*[55] and *STARD7*[56], and a full (> 200) CGG repeat expansion in *FMR1* (Fragile X). Both tools performed similarly but we chose STRling based on its ability to detect a more accurate number of repeats compared to ExpansionHunter DeNovo. We then used STRling (version 0.5.2) to call short tandem repeats on the processed and mapped sequencing data at the genome-wide level. Known repeat expansions were additionally called by ExpansionHunter[14] v5.0 and STRipy[15] using the extended mode (i.e. with off-target regions). AAG is the motif detected by the bioinformatics tools (equivalent to GAA, avoids redundancy of repeated motifs). In this study, we decided to keep this nomenclature, especially as the repeat expansion generally starts with an AAG motif after the pre-repeat (see Fig. 5A).

Long-range PCR amplification. Repeat expansions at the *FGF14* locus were amplified by Long-Range PCR (LR-PCR) from genomic DNA extracted from blood using a protocol adapted from Rafehi *et al*[7]. Notably, we reduced the number of cycles of the LR-PCR to limit an enhanced amplification of smaller alleles. The amplification of *FGF14* repeat alleles was performed from 50 ng genomic DNA in 25 µl using the HotStarTaq DNA Polymerase (Qiagen, Hilden, Germany) and 0.20 µM of each of the following primers: FGF14_RPP_ F1: AGCAATCGTCAGTCAGTGTAAGC; FGF14_LRP_ R1: CAGTTCCTGCCCACATAGAGC. The PCR program comprised an initial step at 95°C of 15 min; followed by 28 cycles, each consisting of 30 seconds at 95°C, 30 seconds at 60°C and 2 minutes at 72°C; and a final step at 72°C of 10 min. LR-PCR amplicons were analyzed on a 1.3% agarose gel. The PCR was performed with a FAM-marked forward primer for gene fragment analysis on ABI 3130xl DNA Analyzer (Applied Biosystems, Waltham, MA). The

size of FGF14 alleles below 700–1200 bp were quantified using the GeneMarker software (SoftGenetics LLC, PA).

Targeted sequencing with Oxford Nanopore Technologies (ONT). LR-PCR products were sequenced for 123 individuals: 100 subjects (53 patients with cerebellar ataxia and 47 control subjects) with an allele ≥ 700 bp, and 23 samples (17 controls and 6 patients) with an allele between 487 and 685 bp (due to discrepancies between values estimated on agarose gel and fragment analysis). LR-PCR amplification was performed without FAM-labelled primer in a total volume of 75 µl. LR-PCR amplicons were purified using the DNA Clean & Concentrator (Zymo Research). We use the SQK-LSK109 ligation-based sequencing kit (Oxford Nanopore) and the native barcoding protocol (EXP-NBD196) to prepare the libraries and multiplex samples, respectively. In brief, this procedure involves the following stages: 1) End-prep, where 200 fmol of each purified amplicon undergoes incubation with NEBNext Ultra II End Repair/dA-tailing Module Reagents at 20°C for 5 min and 65°C for 5 min in a 96-well plate; 2) Native barcoding ligation, comprising incubation with native barcodes and NEB Blunt/TA Ligase Master Mix at 20°C for 20 min and 65°C for 10 min; 3) Pooling of barcoded amplicons and purification using AMPure XP beads (Beckman Coulter); 4) Adapter ligation, involving a 10-minute incubation with Adapter Mix II Expansion/NEBNext Quick Ligation Reaction Module followed by clean-up with AMPure XP beads. All steps were executed in accordance with the manufacturer's recommendations. Approximately 15 ng of the final prepared library was loaded onto a MinION Mk1B R9.4.1 FLO-MIN106 flow cell, and nanopore sequencing was conducted for up to 24 hours, and monitored using the MinKNOW software.

Basecalling and analysis of nanopore data were performed using command line Snakemake[16] workflows available at GitHub (https://github.com/kilpert/FGF14_basecalling.git; https://github.com/kilpert/FGF14_analyses.git). This workflow takes fast5 files as input and utilizes several tools to generate sequence_summary.txt, final_summary.txt, and fastq.gz files. Guppy[57] (version 6.4.6) was used for basecalling, pycoQC[58] (version 2.5.2) and NanoPlot[59] (version 1.41.6) for quality control. The command line version of Guppy, guppy_basecaller, used the parameters: ""--recursive --compress_fastq --do_read_splitting --calib_detect --records_per_fastq 0 --enable_trim_barcodes". After comparing the basecalling hac (dna_r9.4.1_450bps_hac.cfg) and sup (dna_r9.4.1_450bps_sup.cfg) models, the "sup" model was selected for further analysis. The "pass"-reads of samples sequenced on multiple runs were pooled. The initial fastq files were quality trimmed and filtered with BBMap[60] bbduk.sh using the parameters: "-Xmx2g qin = 33 minlen = 200 qtrim = lr trimq = 10 maq = 10 maxlen = 100000". Two 25 bp flanking sequences upstream chr13(hg38):102161532–102161557, ATATCAATATTCTCTATGCAACCAA) and downstream (chr13:102161726–102161751, TAGAAATGTGTTTAAGAATTCCTCA) of the repeat expansion were used to filter for all reads that had both flanking sequences, allowing 2 mismatches per flanking sequence using bbduk.sh with --literal and --edist parameters. Reads where both flanking sequences showed a -strand orientation were subsequently converted to their reverse-complement (+ strand) sequence. All other reads were discarded. In the next step, flanking sequences were trimmed off from both sides using Cutadapt[48] to leave only the repeat containing region, as well as the invariable and variable region (5' flanking region) and pre-repeat. We considered only reads containing 2x AAG (i.e.

AAGAAG). The workflow calculates statistics and generates plots (custom Python and R scripts) to characterize the nature of the repeat expansion in length and motif composition. Specific alleles (length and motif) can be defined manually for enhanced visualization.

For each sample, reads were visually inspected in Geneious Prime® 2019 (Biomatters Ltd.) for identification of the sequences corresponding to the invariable region, 5' flanking region, pre-repeat, main motif, additional repeat motif, and interruptions. Differences between alleles (inclusion or exclusion of sequences, minimum or maximum length) were used to separate reads from different alleles into a1 (smaller), a2 (larger) and a3 (intermediate, mosaic cases). Plots were generated for the separated alleles using up to 300 random reads per allele. For each allele and sample, we calculated the median size of the repeat region (after subtracting from the length of each sequence, the sizes of the invariable region, 5' flanking region and pre-repeat), the median number of triplets and its standard deviation. For comparisons with the sizes obtained by fragment analysis, we increased the median size by 146 bp, which were previously removed by trimming.

Repeat-primed PCR. *FGF14* AAG repeat expansions were amplified by repeat-primed PCR (RP-PCR) with 6-FAM labeled FGF14_RPP_F1 (AGCAATCGTCAGTCAGTGTAAGC), and non-labeled RPP_M13R (CAGGAAACAGCTATGACC) and FGF14_RPP_AAG_RE_R1 (CAGGAAACAGCTATGACCCTTCTTCTT-CTTCTTCTTCTT). AAGGAG expanded alleles were amplified using FAM-FGF14_RPP_F1, P3 (TACGCATCCCAGTTTGAGACG), and P3_AAGGAG (TACGCATCCCAGTTTGAGACGAAGGAGAAGGA-GAAGGAGAAG). PCR was performed from 100 ng genomic DNA, with 0.8 µM primer FGF14_RPP_ F1, 0.8 µM primer RPP_M13R or P3, and 0.26 µM primer FGF14_RPP_AAG_RE_R1 or P3_AAGGAG using the HotStarTaq Master Mix (QIAGEN). The PCR program consisted in 95°C for 15 min, followed by 35 cycles (94°C for 30 s, 61°C for 30 s, and 72°C from for 2 min) and a final extension step at 72°C for 10 min. RP-PCR products were detected on an ABI 3130xl DNA Analyzer and analyzed using GeneMarker software (SoftGenetics).

Statistics. Fisher's tests (two-sided) were performed to determine associations between: (i) a sample belonging to a class of triplet numbers (either in all alleles or just in the larger allele of an individual) and being an ataxia patient; (ii) a patient belonging to a class of *FGF14* pathogenic expansion size or point mutations and presenting certain symptoms; (iii) a patient belonging to a class of *FGF14* pathogenic expansion size and having a certain family history presentation; and (iv) a patient belonging to a class of *FGF14* pathogenic expansion size and having a certain response to treatment. Odds ratios were $\log_2$ transformed and indicate enrichment or depletion, for positive or negative values, respectively. *P*-values were adjusted for multiple comparisons using Bonferroni correction. Mann-Whitney U test (two-sided) followed by Bonferroni correction for multiple testing (when applicable) was used to assess (i) the median number of repeats difference between ataxia patients and controls; (ii) the age at onset / age at last examination / disease duration difference between various classes of *FGF14* pathogenic expansion size and point mutations; (iii) the difference in the number of triplets quantified by ExpansionHunter or Nanopore/STR fragment size for samples for which STRling detected or not FGF14 expansions; and (iv)

the standard deviation of the number of triplets difference between various categories of main repeats and interruptions. The details and results of all statistical tests performed appear in the Source Data file.

Circular dichroism spectroscopy. The AAG and AAGGAG repeats and the complementary counterparts were purchased as 25-mer DNA and RNA oligos from Microsynth (Switzerland) and dissolved in nuclease-free water at 100 µM. The sequences of the oligos are displayed in Fig. 8C and are identical for DNA and RNA except that thymidines were replaced by uracils in RNA. The final concentration of oligos was 50 µM in 10 mM cacodylate buffer (pH 7.2) either supplemented with 100 mM K^+ or Li^+ as indicated. Secondary structure formation was carried out by heating the oligos to 95°C for 5 min and slowly decreasing the temperature with a ramp rate of 0.01°C/s to 20°C using the LightCycler® LC480II (Roche). The folded oligos were kept at 4°C overnight and measured the next day. Circular dichroism (CD) spectra were measured over a spectral range of 200–340 nm on a Jasco J-710 CD spectropolarimeter coupled to a Jasco PFD-3505 Peltier temperature controller. All measurements were carried out at 20°C in a quartz cuvette with a 1 mm path length using a scanning speed of 200 nm/min, a response time of 2 s, a bandwidth of 1 nm, and an accumulation of four spectra.

Declarations

Resources

Ensembl: https://www.ensembl.org

Uniprot: https://www.uniprot.org/

GTEx: https://www.gtexportal.org/

STRipy: https://stripy.org/

Acknowledgements

We thank all patients, family members, and control subjects who participated in this study. We also thank Nele Bohne and Dr. Theresa Kühnel for their contributions to establishing the LR-PCR analysis and Dr. Christine Beuck for her expert technical guidance with the CD spectroscopy measurements.

Author contributions

L.M., F. E., E. L., G.S.H., and S. Ka. contributed to data acquisition and data analysis. L.M., E. L., and S.Ka performed the experiments. E.L., F. Ki. and C.S. performed computational analyses. C.D. conceived and supervised the study. A.T., M.St., J.P., A.S., M.R., M.M.dlP., C.C., K.B., U.R., S.P., F.J.K., M.Sy., T.W., M.A., T.B.H., P.J.L., K-H.J., A.P., S.Kl., D.T. contributed genetic data, patient data, analysis and/or critically revised the manuscript for intellectual content. C.D, F.E. and D.T. drafted the manuscript. All authors reviewed and approved the manuscript.

Funding

This work was supported by the Tom Wahlig foundation, the University Hospital Essen, and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Research Infrastructure West German Genome Center (project 407493903) as part of the Next Generation Sequencing Competence Network (project 423957469). Short-read genome sequencing was carried out at the production site Cologne (Cologne Center for Genomics; CCG). C.D. received the DFG 458099954 as part of the DFG Sequencing call #3. S.Kl. was supported by the German Federal Ministry of Education and Research (BMBF) through the TreatHSP consortium (01GM1905C). M.S. and -as an associated partner- D.T. and S. Kl. were also supported by the DFG (grant number 441409627), as part of the PROSPAX consortium under the frame of European Joint Programme on Rare Diseases (EJP-RD), under the EJP RD COFUND-EJP N° 825575. J.P. was supported by the Clinician Scientist program “PRECISE.net” funded by the Else Kröner-Fresenius-Stiftung. F.J.K. received funding from the DFG Research Unit FOR2488. Genome sequencing in Spanish patients was supported by the Undiagnosed Rare Diseases Program of Catalonia (URDCat; PERIS SLT002/16/00174) from the Autonomous Government of Catalonia; the Biomedical Research Networking Center on Rare Diseases (CIBERER, ACCI19-759); the Hesperia Foundation (Royal House of Spain), and “La Marató de TV3” Foundation with project 202006–30 to C.C. and A.P. and the Instituto de Salud Carlos III co-funded by the “Fondo Europeo de Desarrollo Regional (FEDER), Unión Europea, una manera de hacer Europa” (FIS PI20/00758) to C.C., and IMPACT-Genomica (IMP/00009) to A.P.. M.R. was funded by the Center for Biomedical Research on Rare Diseases, an initiative of the Instituto de Salud Carlos III. We thank the CERCA Program/Generalitat de Catalunya for institutional support.

Competing interests

The authors report no competing interests

Inclusion & Ethics

Informed consents have been obtained for each patient/participant included in this study. Patient/participant/samples were pseudonymized for the genetic study at each participating center. Participants receive no compensation for inclusion in the genetic study. Genetic and clinical data shared in the context in this study cannot be used to identify individuals. Participating centers and data collection sites had study protocols approved by the local Institutional Review Boards (IRB) the Ethik Kommission der Medizinischen Fakultät der Universität Duisburg-Essen. Researchers and clinicians from participating centers contributing either data or intellectual input were involved at all stages of the study from design, implementation, drafting, and revising the manuscript, and are coauthors of the article.

Data availability

All data and scripts included in this study, with the exception of genome data, are available from the Supplementary tables or in the Github repository. The consent forms of the EXPAND study do not allow

depositing genome data in a public repository.

Code availability

Custom scripts used in this study are available on GitHub using the following links: https://github.com/kilpert/FGF14_basecalling.git; https://github.com/kilpert/FGF14 _analyses.git.

References

1. Coarelli G, Coutelier M, Durr A. Autosomal dominant cerebellar ataxias: new genes and progress towards treatments. *Lancet Neurol*. Aug 2023;22(8):735-749. doi:10.1016/S1474-4422(23)00068-6
2. Ashizawa T, Oz G, Paulson HL. Spinocerebellar ataxias: prospects and challenges for therapy development. *Nat Rev Neurol*. Oct 2018;14(10):590-605. doi:10.1038/s41582-018-0051-6
3. van Swieten JC, Brusse E, de Graaf BM*, et al*. A mutation in the fibroblast growth factor 14 gene is associated with autosomal dominant cerebellar ataxia [corrected]. *Am J Hum Genet*. Jan 2003;72(1):191-9. doi:10.1086/345488
4. Choquet K, La Piana R, Brais B. A novel frameshift mutation in FGF14 causes an autosomal dominant episodic ataxia. *Neurogenetics*. Jul 2015;16(3):233-6. doi:10.1007/s10048-014-0436-7
5. Dalski A, Atici J, Kreuz FR, Hellenbroich Y, Schwinger E, Zuhlke C. Mutation analysis in the fibroblast growth factor 14 gene: frameshift mutation and polymorphisms in patients with inherited ataxias. *Eur J Hum Genet*. Jan 2005;13(1):118-20. doi:10.1038/sj.ejhg.5201286
6. Pellerin D, Danzi MC, Wilke C*, et al*. Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia. *N Engl J Med*. Jan 12 2023;388(2):128-141. doi:10.1056/NEJMoa2207406
7. 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*. Jan 5 2023;110(1):105-119. doi:10.1016/j.ajhg.2022.11.015
8. Rafehi H, Read J, Szmulewicz DJ*, et al*. An intronic GAA repeat expansion in FGF14 causes the autosomal-dominant adult-onset ataxia SCA27B/ATX-FGF14. *Am J Hum Genet*. Jun 1 2023;110(6):1018. doi:10.1016/j.ajhg.2023.05.005
9. Laezza F, Lampert A, Kozel MA*, et al*. FGF14 N-terminal splice variants differentially modulate Nav1.2 and Nav1.6-encoded sodium channels. *Mol Cell Neurosci*. Oct 2009;42(2):90-101. doi:10.1016/j.mcn.2009.05.007
10. Xiao M, Bosch MK, Nerbonne JM, Ornitz DM. FGF14 localization and organization of the axon initial segment. *Mol Cell Neurosci*. Sep 2013;56:393-403. doi:10.1016/j.mcn.2013.07.008
11. Pablo JL, Pitt GS. FGF14 is a regulator of KCNQ2/3 channels. *Proc Natl Acad Sci U S A*. Jan 3 2017;114(1):154-159. doi:10.1073/pnas.1610158114
12. Schmermund A, Möhlenkamp S, Stang A*, et al*. Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors,

Evaluation of Coronary Calcium and Lifestyle. *Am Heart J*. Aug 2002;144(2):212-8. doi:10.1067/mhj.2002.123579

13. Dashnow H, Pedersen BS, Hiatt L*, et al*. STRling: a k-mer counting approach that detects short tandem repeat expansions at known and novel loci. *Genome Biol*. Dec 14 2022;23(1):257. doi:10.1186/s13059-022-02826-4
14. Dolzhenko E, van Vugt J, Shaw RJ*, et al*. Detection of long repeat expansions from PCR-free whole-genome sequence data. *Genome Res*. Nov 2017;27(11):1895-1903. doi:10.1101/gr.225672.117
15. Halman A, Dolzhenko E, Oshlack A. STRipy: A graphical application for enhanced genotyping of pathogenic short tandem repeats in sequencing data. *Hum Mutat*. Jul 2022;43(7):859-868. doi:10.1002/humu.24382
16. Molder F, Jablonski KP, Letcher B*, et al*. Sustainable data analysis with Snakemake. *F1000Res*. 2021;10:33. doi:10.12688/f1000research.29032.2
17. 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*. Aug 2023;38(8):1575-1577. doi:10.1002/mds.29552
18. 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*. Jun 2023;38(6):1107-1109. doi:10.1002/mds.29390
19. Wilke C, Pellerin D, Mengel D*, et al*. GAA-FGF14 ataxia (SCA27B): phenotypic profile, natural history progression and 4-aminopyridine treatment response. *Brain*. Oct 3 2023;146(10):4144-4157. doi:10.1093/brain/awad157
20. Wirth T, Clement G, Delvallee C*, et al*. Natural History and Phenotypic Spectrum of GAA-FGF14 Sporadic Late-Onset Cerebellar Ataxia (SCA27B). *Mov Disord*. Jul 20 2023;doi:10.1002/mds.29560
21. Ando M, Higuchi Y, Yuan J*, et al*. Clinical variability associated with intronic FGF14 GAA repeat expansion in Japan. *Ann Clin Transl Neurol*. Nov 2 2023;doi:10.1002/acn3.51936
22. Miura S, Kosaka K, Fujioka R*, et al*. Spinocerebellar ataxia 27 with a novel nonsense variant (Lys177X) in FGF14. *Eur J Med Genet*. Mar 2019;62(3):172-176. doi:10.1016/j.ejmg.2018.07.005
23. Humbertclaude V, Krams B, Nogue E*, et al*. Benign paroxysmal torticollis, benign paroxysmal vertigo, and benign tonic upward gaze are not benign disorders. *Dev Med Child Neurol*. Dec 2018;60(12):1256-1263. doi:10.1111/dmcn.13935
24. Brusse E, de Koning I, Maat-Kievit A, Oostra BA, Heutink P, van Swieten JC. Spinocerebellar ataxia associated with a mutation in the fibroblast growth factor 14 gene (SCA27): A new phenotype. *Mov Disord*. Mar 2006;21(3):396-401. doi:10.1002/mds.20708
25. 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*. Aug 2023;38(8):1557-1558. doi:10.1002/mds.29559

26. Kypr J, Kejnovska I, Renciuk D, Vorlickova M. Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res*. Apr 2009;37(6):1713-25. doi:10.1093/nar/gkp026
27. Vorlickova M, Kejnovska I, Bednarova K, Renciuk D, Kypr J. Circular dichroism spectroscopy of DNA: from duplexes to quadruplexes. *Chirality*. Sep 2012;24(9):691-8. doi:10.1002/chir.22064
28. Vorlickova M, Kejnovska I, Sagi J*, et al*. Circular dichroism and guanine quadruplexes. *Methods*. May 2012;57(1):64-75. doi:10.1016/j.ymeth.2012.03.011
29. Del Villar-Guerra R, Trent JO, Chaires JB. G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew Chem Int Ed Engl*. Jun 11 2018;57(24):7171-7175. doi:10.1002/anie.201709184
30. Chauca-Diaz AM, Choi YJ, Resendiz MJ. Biophysical properties and thermal stability of oligonucleotides of RNA containing 7,8-dihydro-8-hydroxyadenosine. *Biopolymers*. Mar 2015;103(3):167-74. doi:10.1002/bip.22579
31. Bhattacharyya D, Mirihana Arachchilage G, Basu S. Metal Cations in G-Quadruplex Folding and Stability. *Front Chem*. 2016;4:38. doi:10.3389/fchem.2016.00038
32. Zaccaria F, Fonseca Guerra C. RNA versus DNA G-Quadruplex: The Origin of Increased Stability. *Chemistry*. Nov 2 2018;24(61):16315-16322. doi:10.1002/chem.201803530
33. Garg A, Heinemann U. A novel form of RNA double helix based on G.U and C.A(+) wobble base pairing. *RNA*. Feb 2018;24(2):209-218. doi:10.1261/rna.064048.117
34. Pellerin D, Wilke C, Traschutz A*, et al*. Intronic FGF14 GAA repeat expansions are a common cause of ataxia syndromes with neuropathy and bilateral vestibulopathy. *J Neurol Neurosurg Psychiatry*. Jun 30 2023;doi:10.1136/jnnp-2023-331490
35. Novis LE, Frezatti RS, Pellerin D*, et al*. Frequency of GAA-FGF14 Ataxia in a Large Cohort of Brazilian Patients With Unsolved Adult-Onset Cerebellar Ataxia. *Neurol Genet*. Oct 2023;9(5):e200094. doi:10.1212/NXG.0000000000200094
36. 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*. Dec 2023;30(12):3828-3833. doi:10.1111/ene.16039
37. Bonnet C, Pellerin D, Roth V*, et al*. Optimized testing strategy for the diagnosis of GAA-FGF14 ataxia/spinocerebellar ataxia 27B. *Sci Rep*. Jun 15 2023;13(1):9737. doi:10.1038/s41598-023-36654-8
38. Ashton C, Indelicato E, Pellerin D*, et al*. Spinocerebellar ataxia 27B: episodic symptoms and acetazolamide response in 34 patients. *Brain Commun*. 2023;5(5):fcad239. doi:10.1093/braincomms/fcad239
39. Wang Q, Bardgett ME, Wong M*, et al*. Ataxia and paroxysmal dyskinesia in mice lacking axonally transported FGF14. *Neuron*. Jul 3 2002;35(1):25-38. doi:10.1016/s0896-6273(02)00744-4
40. Pellerin D, Heindl F, Wilke C*, et al*. Intronic FGF14 GAA repeat expansions are a common cause of downbeat nystagmus syndromes: frequency, phenotypic profile, and 4-aminopyridine treatment response. *medRxiv*. Aug 5 2023;doi:10.1101/2023.07.30.23293380

41. Rajagopal S, Donaldson J, Flower M, Hensman Moss DJ, Tabrizi SJ. Genetic modifiers of repeat expansion disorders. *Emerg Top Life Sci*. Oct 20 2023;doi:10.1042/ETLS20230015
42. Laezza F, Gerber BR, Lou JY*, et al*. The FGF14(F145S) mutation disrupts the interaction of FGF14 with voltage-gated Na+ channels and impairs neuronal excitability. *J Neurosci*. Oct 31 2007;27(44):12033-44. doi:10.1523/JNEUROSCI.2282-07.2007
43. Pellerin D, Gobbo GD, Couse M*, et al*. A common flanking variant is associated with enhanced meiotic stability of the FGF14 -SCA27B locus. *bioRxiv*. Jun 30 2023;doi:10.1101/2023.05.11.540430
44. 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*. Oct 2023;299(10):105202. doi:10.1016/j.jbc.2023.105202
45. Groh M, Lufino MM, Wade-Martins R, Gromak N. R-loops associated with triplet repeat expansions promote gene silencing in Friedreich ataxia and fragile X syndrome. *PLoS Genet*. May 2014;10(5):e1004318. doi:10.1371/journal.pgen.1004318
46. Butler JS, Napierala M. Friedreich's ataxia–a case of aberrant transcription termination? *Transcription*. 2015;6(2):33-6. doi:10.1080/21541264.2015.1026538
47. Zhang J, Fakharzadeh A, Pan F, Roland C, Sagui C. Atypical structures of GAA/TTC trinucleotide repeats underlying Friedreich's ataxia: DNA triplexes and RNA/DNA hybrids. *Nucleic Acids Res*. Sep 25 2020;48(17):9899-9917. doi:10.1093/nar/gkaa665
48. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal*. 2011;1(1)(EMBnet Stichting):10. doi:https://doi.org/10.14806/ej.17.1.200
49. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013;doi:https://doi.org/10.48550/ARXIV.1303.3997
50. Vasimuddin M, Misra, S., Li, H., & Aluru, S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. *In 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS) 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS) IEEE* 2019;doi:https://doi.org/10.1109/ipdps.2019.00041
51. Jung Y, Han D. BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics*. Apr 28 2022;38(9):2404-2413. doi:10.1093/bioinformatics/btac137
52. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics*. Sep 1 2014;30(17):2503-5. doi:10.1093/bioinformatics/btu314
53. Danecek P, Bonfield JK, Liddle J*, et al*. Twelve years of SAMtools and BCFtools. *Gigascience*. Feb 16 2021;10(2)doi:10.1093/gigascience/giab008
54. Dolzhenko E, Bennett MF, Richmond PA*, et al*. ExpansionHunter Denovo: a computational method for locating known and novel repeat expansions in short-read sequencing data. *Genome Biol*. Apr 28 2020;21(1):102. doi:10.1186/s13059-020-02017-z
55. Florian RT, Kraft F, Leitao E*, et al*. Unstable TTTTA/TTTCA expansions in MARCH6 are associated with Familial Adult Myoclonic Epilepsy type 3. *Nat Commun*. Oct 29 2019;10(1):4919. doi:10.1038/s41467-019-12763-9

56. Corbett MA, Kroes T, Veneziano L*, et al*. Intronic ATTTC repeat expansions in STARD7 in familial adult myoclonic epilepsy linked to chromosome 2. *Nat Commun*. Oct 29 2019;10(1):4920. doi:10.1038/s41467-019-12671-y
57. Community N. Downloads - Release notes. Oxford Nanopore Technologies. . doi:https://community.nanoporetech.com/downloads/guppy/release_notes.
58. Leger AL, T. . pycoQC, interactive quality control for Oxford Nanopore Sequencing. *J Open Source Softw* 2019;4:1236.
59. De Coster W, Rademakers R. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics*. May 4 2023;39(5)doi:10.1093/bioinformatics/btad311
60. Offical) BBN. BBMap short read aligner, and other bioinformatic tools. doi:https://github.com/BioInfoTools/BBMap/tree/master.

Figures

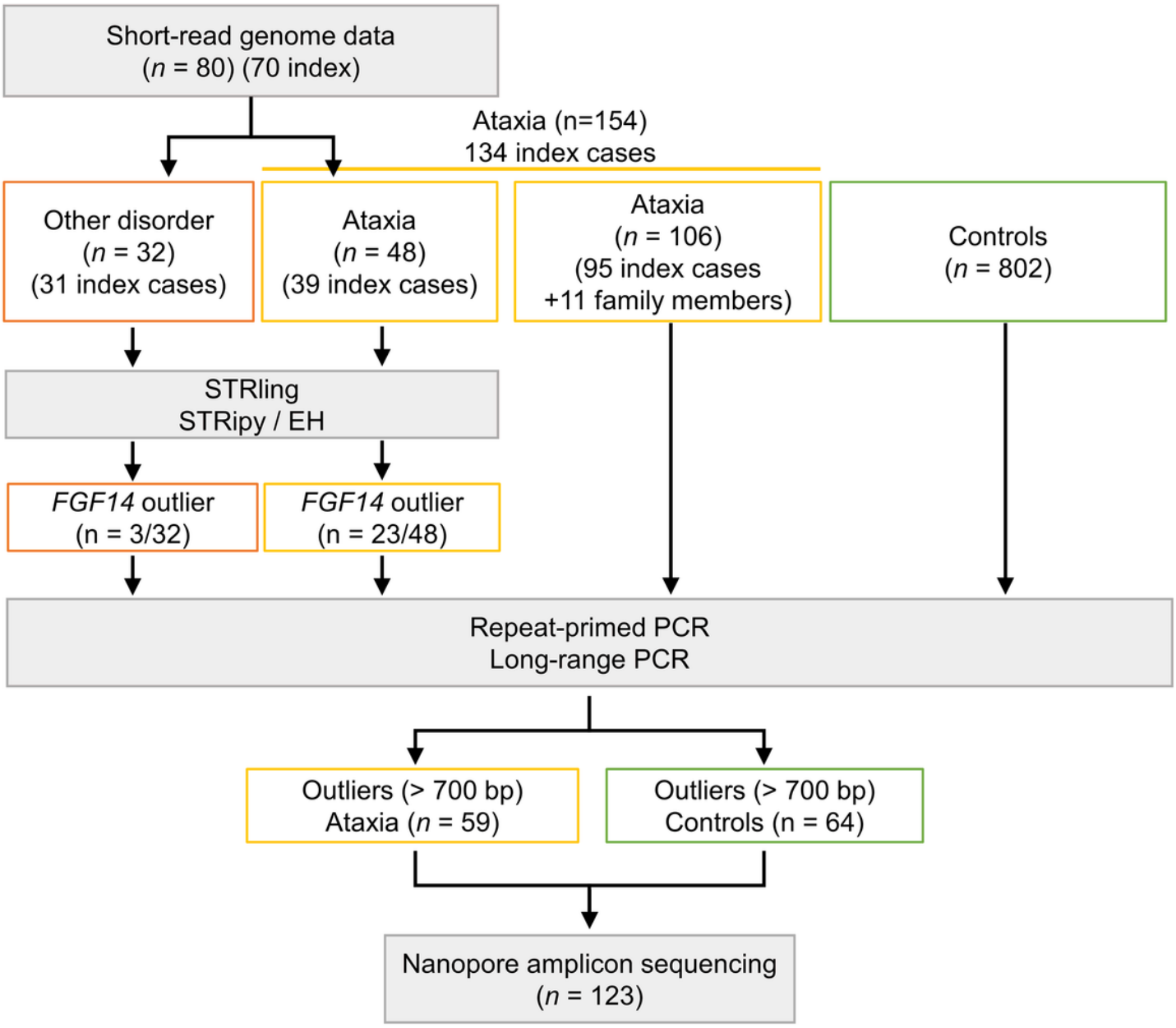

Figure 1

Flowchart illustrating the design of the study. This figure visually represents the sequential steps, methods used, and distribution of participants at each stage.

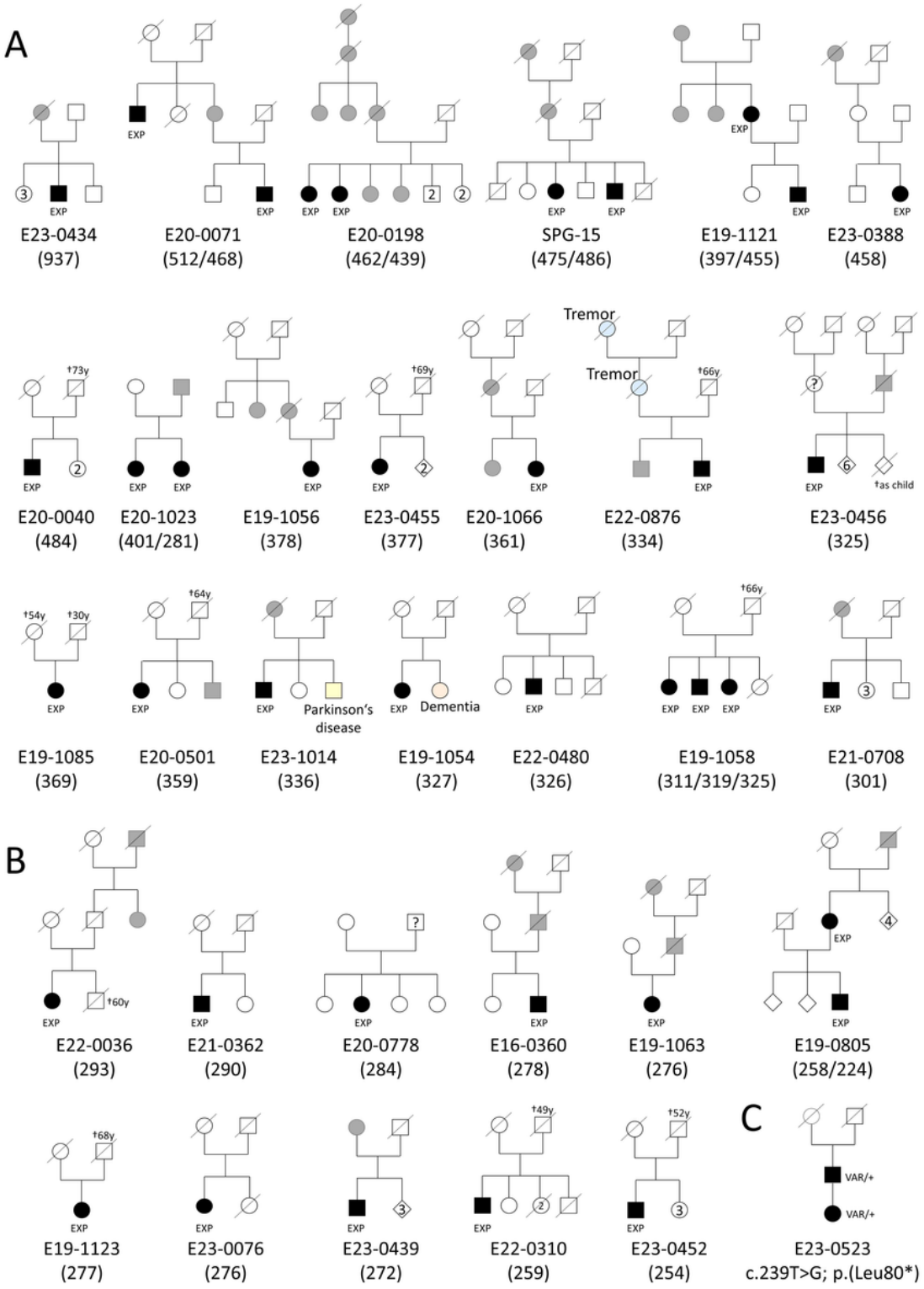

Figure 2

Pedigrees of families with *FGF14* pathogenic repeat expansions. **A)**Pedigrees of families in which at least one affected subject had a number of AAG repeats ≥300. **B)** Pedigrees of families in which at least one affected subject had a repeat number comprised between 250 and 299. Black symbols indicates subjects examined and sampled in the study. Grey symbols indicate subjects reported to be affected on history but that could not be examined. The number in brackets indicates the median number of repeats for the

affected individuals of this family. EXP indicates individuals with a *FGF14* expansion. **C)** Pedigree of the family with NM_175929.3: c.239T>G; p.(Leu80*).

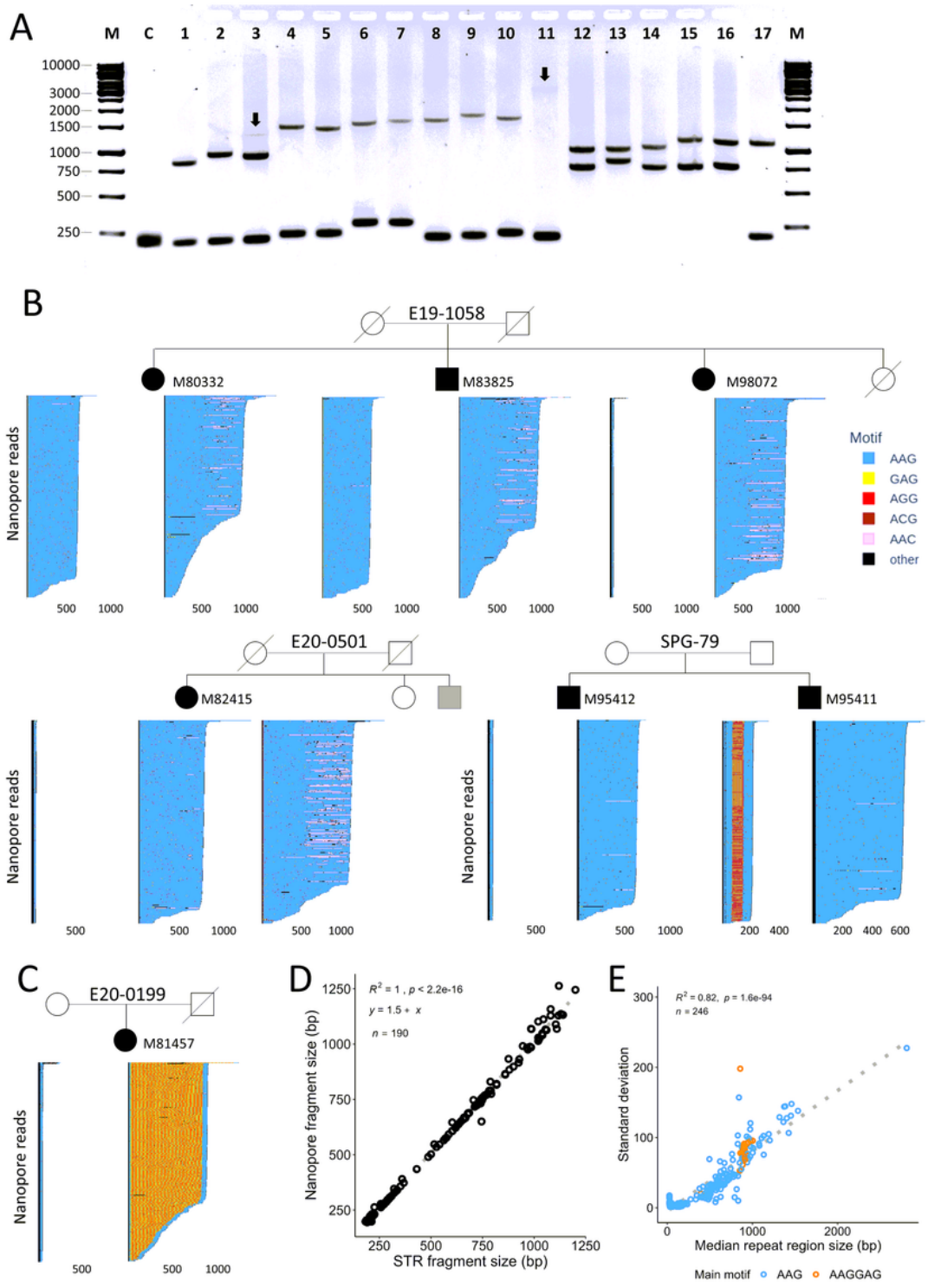

Figure 3

Analysis of *FGF14* repeat expansions in affected subjects using LR-PCR and nanopore sequencing. A) Gel electrophoresis of selected LR-PCR products spanning the *FGF14*(AAG) STR locus. From left to right,

C: negative control; lanes 1-2: examples of *FGF14* expansions between 220 and 299 repeats (1-M79607, 2-M90982); lane 3: individual showing one small and two large alleles (somatic mosaicism of the largest allele; M82415); lanes 4-11: examples of *FGF14* expansions above 300 repeats (4-M87668, 5-M84267, 6-M93354, 7-M81456, 8-M95289, 9-M80996, 10-M80920, 11-M96642 (937 repeats)); lanes 12-16: individuals with biallelic expansions (12-M96652, 13-M97638, 14-M83382, 15-M83825, 16-M80332); lane 17 corresponds to the sibling of individuals 15 and 16 who both have biallelic expansions (family E19-1058, M98072); she has a heterozygous expansion of 325 repeats. **B)** Schematic representation of nanopore reads sequenced for selected family members. *FGF14* alleles have been separated based on their flanking regions (see methods). 300 randomly chosen reads are displayed in each graph. AAG repeats appear in blue. GAG, AGG, and AAC repeats appear in yellow, red and light pink respectively. Other sequences appear in black. The panel above shows the segregation of the *FGF14* alleles in family E19-1058 comprising three affected siblings, two with biallelic expansions and one sister with a single expanded allele. Lower-left panel: individual M82415 (E20-0501) showed reads that could be partitioned into three different alleles: one small allele and two large alleles (somatic mosaicism). Lower-right panels: family SPG-79 including two affected family members with intermediate *FGF14* alleles (203 and 207 repeats). Note interruptions in the middle of the same allele of individual M95411. **C)** Nanopore reads from an individual with a small and an expanded AAGGAG allele. **D)** Correlation between the median number of repeats detected by nanopore sequencing and expansion size estimated from fragment size analysis. Due to the observed very high correlation, we could confidently use the values calculated by nanopore sequencing to fill in missing allele values > 700-1200 bp (i.e. alleles too large to be analyzed by fragment size analysis). **E)** Standard deviation in allele size calculated from nanopore reads showing that somatic instability is positively correlated with expansion size.

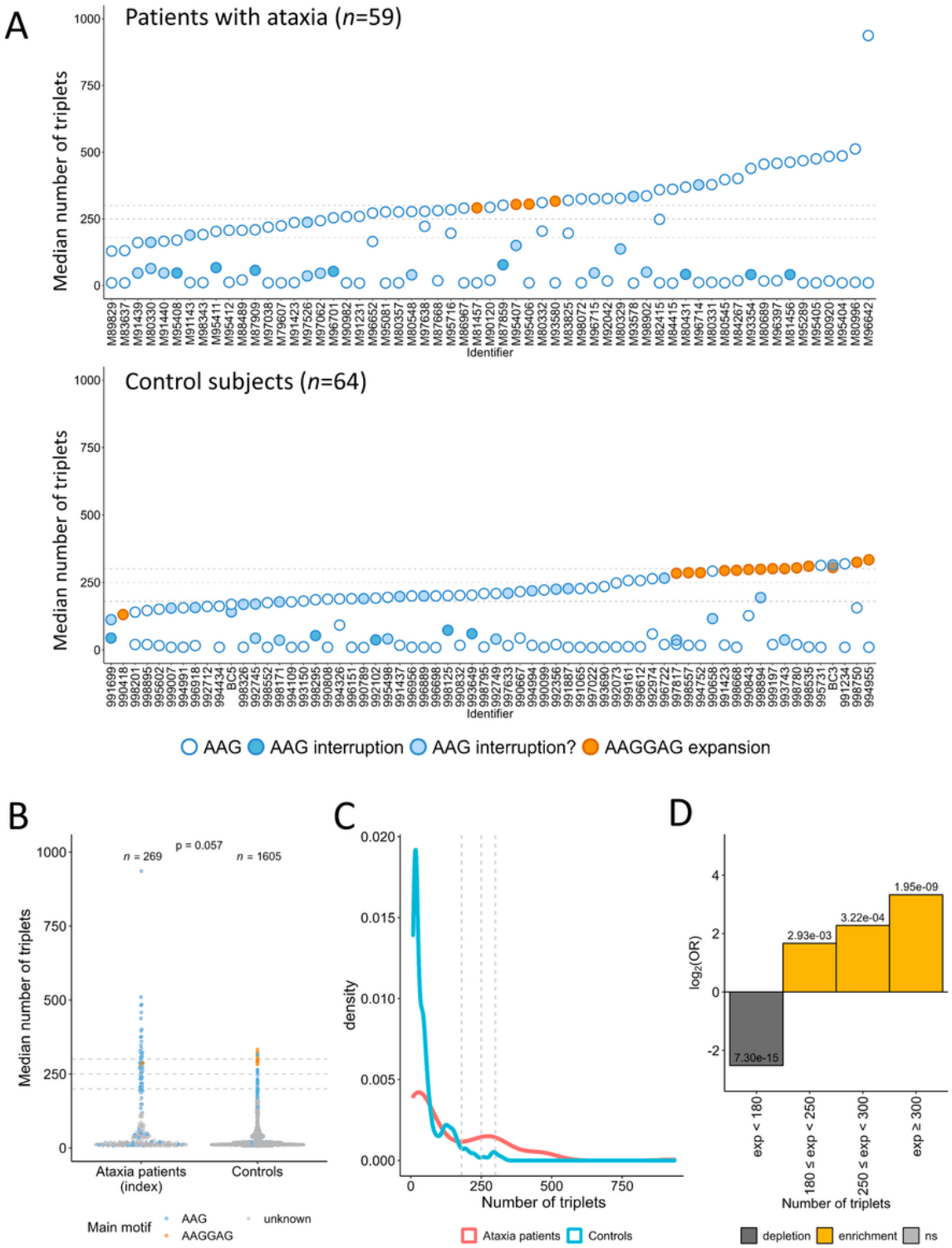

Figure 4

Distribution of *FGF14* alleles in patients with cerebellar ataxia and control subjects. A) Median number of triplets of both alleles for the 59 patients with ataxia and 64 control individuals sequenced by nanopore sequencing. Pure AAG alleles are depicted as blue dots with a white fill. AAGGAG alleles appear in orange. Alleles with interruptions are depicted as blue dots with a dark blue fill. Alleles with interruptions limited to the 5' or 3' of the expansion are depicted as blue dots with a light blue fill. **B)**Comparison of median allele

sizes (including all alleles) in index patients with cerebellar ataxia (*n*=134) and control subjects (*n*=802). **C)** Density plot showing the different distributions of the number of triplets in the larger allele for index patients with cerebellar ataxia (*n*=134) and control subjects (*n*=802). **D)** Log odds ratio according to repeat numbers in all alleles (134 index patients with cerebellar ataxia and 802 control subjects) showing a significant enrichment of alleles > 180 repeats in patients with cerebellar ataxia. Figures similar to B) and D) but considering only the large alleles appear in Supplementary Figure 5.

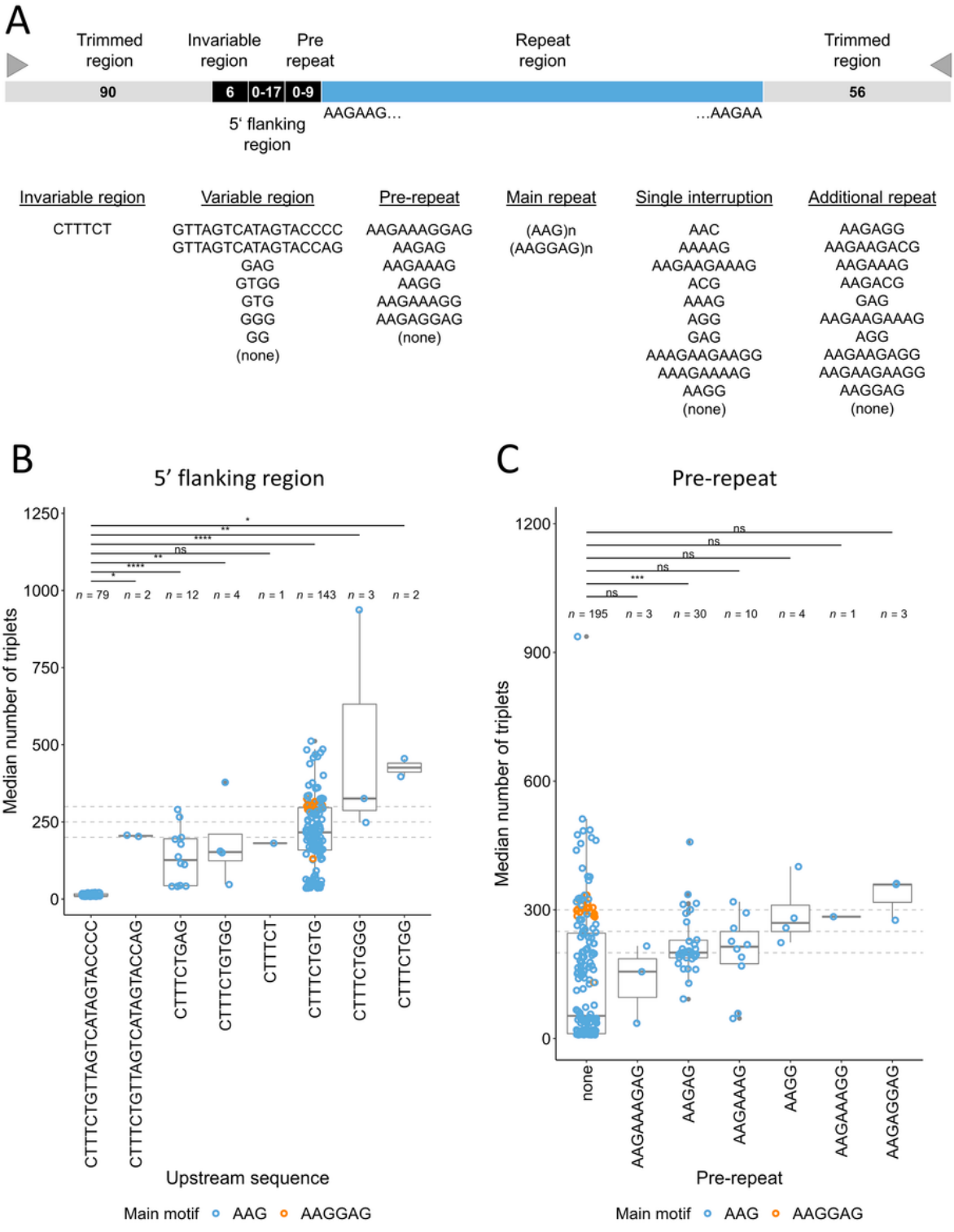

Figure 5

Effect of 5’ flanking regions on repeat instability. **A)** Schematic representation of the different parts composing *FGF14* repeat expansions. An invariable CTTTCT motif is usually followed by a variable 5’ region. A pre-repeat can be present in some individuals before the repeats. Some alleles are interrupted by one or several other motifs called interruptions. **B)** Median number of triplets for each allele depending on the flanking region sequence. GTTAGTCATAGTACCCC is present in small alleles (≤21 repeats) only. Other sequences show higher number of repeats, suggesting higher instability of these associations. **C)**Median number of triplets for each allele depending on the pre-repeat motif. Graphs displayed in panels B) and C) include both patients with ataxia and controls. Graphs presenting data for patients with ataxia and controls separately appear in Supplementary Figure 6.

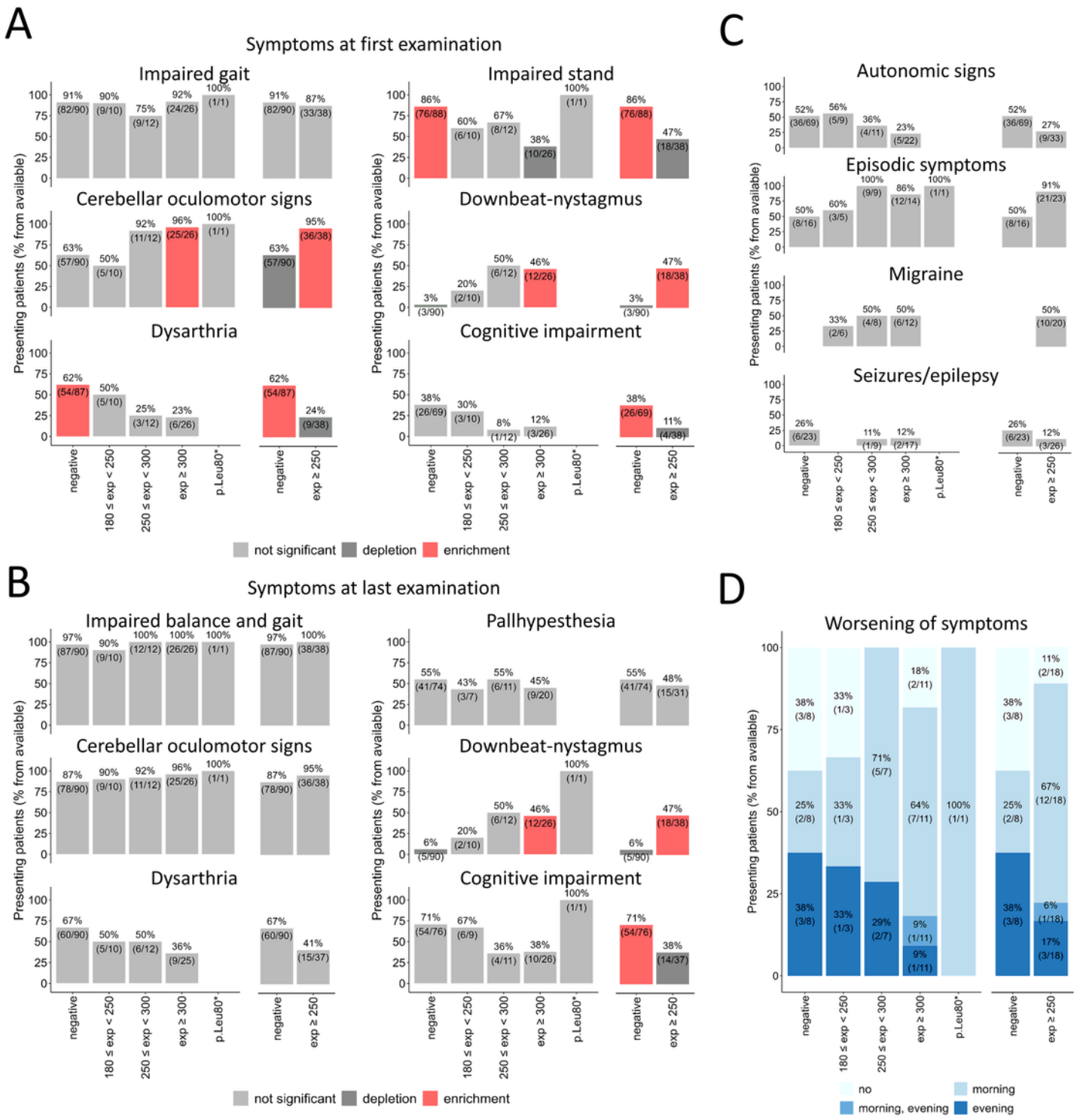

Figure 6

Clinical characteristics of individuals with *FGF14* pathogenic expansions and point mutations. **A)** Bar graphs showing the percentage of patients presenting impaired gait, impaired stand, cerebellar oculomotor signs, downbeat-nystagmus, dysarthria and cognitive impairment at first examination. **B)** Bar graphs showing the percentage of patients presenting impaired gait and stand, cerebellar oculomotor signs, downbeat-nystagmus, dysarthria, pallhypersthesia and cognitive impairment at last examination.

C) Bar graphs showing the percentage of patients presenting autonomic signs, episodic symptoms, migraine, and seizures/epilepsy. For all graphs presented in panels A, B) and C), categories showing significant enrichment appear in red, significant depletion in dark gray while non-significant differences appear in light grey (Fisher's test). **D)** Bar graphs showing the percentage of patients presenting a worsening of symptoms on morning, evening, morning and evening or no worsening of symptoms depending on the day time.

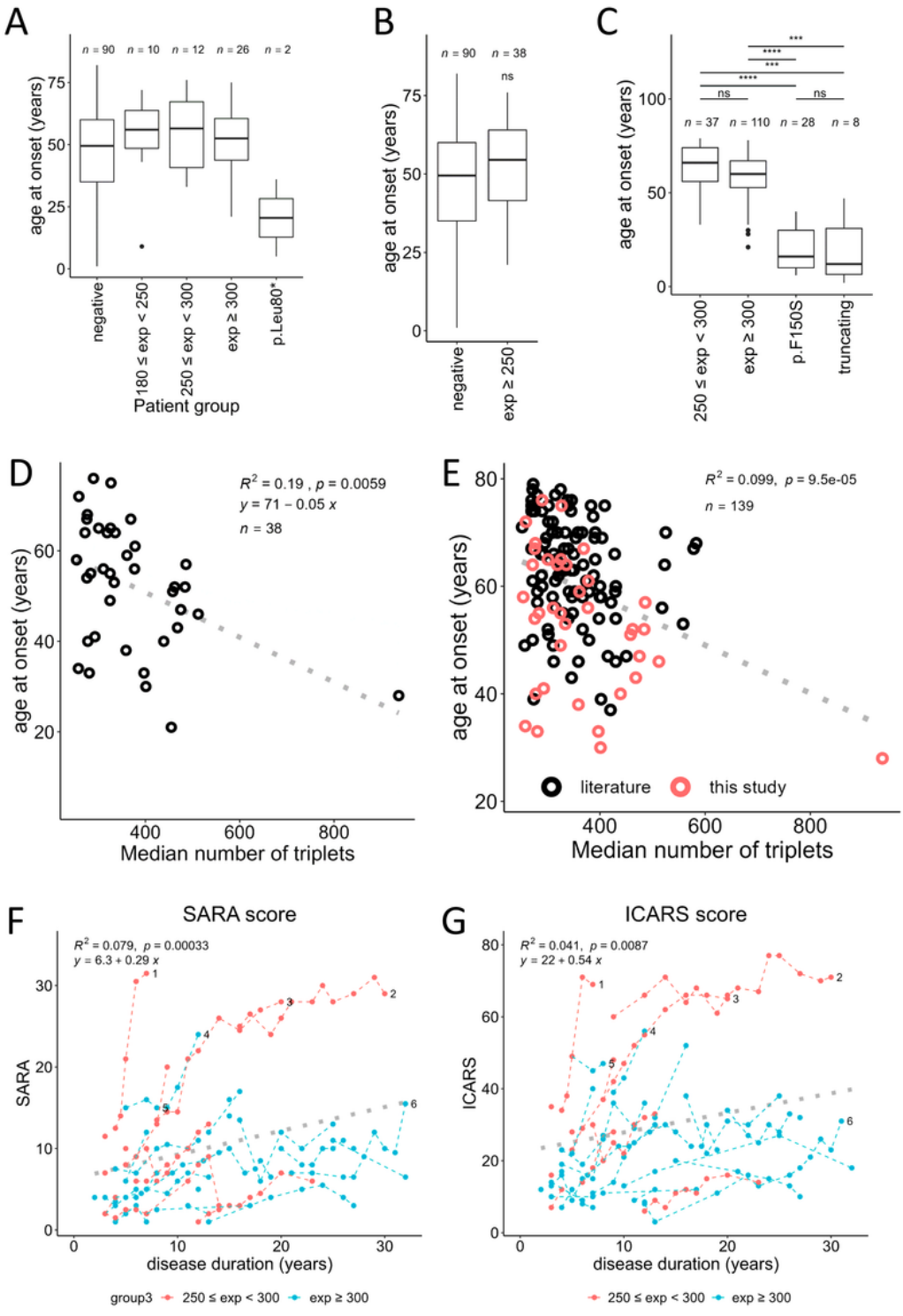

Figure 7

Analysis of the age at onset (AAO) and disease progression in patients with SCA27B/SCA27A. A) Comparison of the AAO in patients negative for SCA27A/SCA27B, patients with intermediate alleles (180-249), patients with *FGF14* expansions comprised between 250 and 299 repeats, patients with repeat size ≥300 repeats and the two patients with p.(Leu80*). **B)**Comparison of the age at onset in patients negative for SCA27A/SCA27B, and patients with SCA27B (repeat size ≥250 repeats). **C)** Meta-analysis comparing the age at onset in patients with *FGF14* expansions comprised between 250 and 299 repeats, patients with repeat size ≥300 repeats, patients with nonsense or frameshift variants in FGF14 or patients with p.Phe150Ser, showing that patients with pathogenic point variants (SCA27A) have an earlier age at onset than patients with repeat expansions (SCA27B). **D)**Correlation between the age at onset and *FGF14*AAG repeat number including only patients from this study. **E)** Correlation between the age at onset and *FGF14* AAG repeat number taking all patients from this study (red) and patients from previous studies (black) into account. **F)** SARA scores of 36 patients with *FGF14* repeat expansions (*n*=147 measurements). **G)** ICARS scores of 36 patients with *FGF14* repeat expansions (*n*=measurements). In both graphs shown in panels F) and G), patients with *FGF14*expansions comprised between 250 and 299 repeats appear in red while patients with ≥ 300 repeats appear in blue. Scores from the same patients at different time points are connected with dashed lines. Numbered last data points mark lines corresponding to atypical patients: M83382/M95716 (1), M90120 (2), M97638 (3), M80332 (4), M81456/M91395 (5) and M80920 (6).

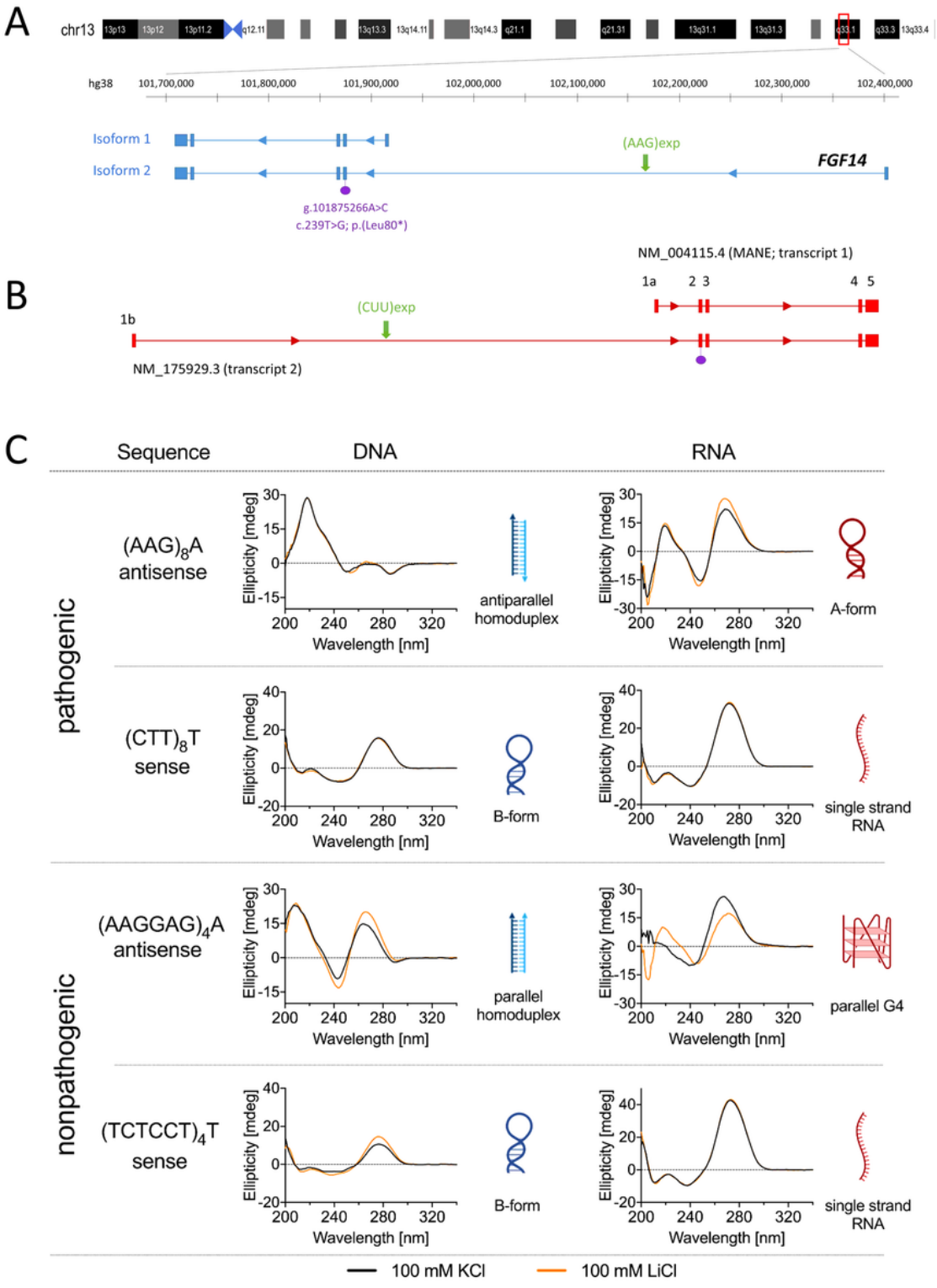

Figure 8

AAG and AAGGAG form different secondary structures at the DNA and RNA level. A) Schematic representation of the region on chromosome 13q33.1 containing the *FGF14*gene showing isoforms 1 (ENST00000376143.5; NM_004115.4) and 2 (ENST00000376131.9; NM_175929.3), which have alternative first exons. The gene is on the reverse strand. The green arrows show the location of the AAG expansion in intron 1 of isoform 2. The location of the novel nonsense variant (NM_175929.3: c.239T>G;

p.Leu80*) reported in this study is indicated in purple. **B)** Schematic representation of *FGF14* pre-mRNA isoforms 1 and 2. The expansion (green arrow) is composed of CUU repeats in RNA context. **C)** Secondary structures formed by AAG and AAGGAG repeats at the DNA and RNA level, assessed by circular dichroism spectroscopy. AAG repeats form an antiparallel homoduplex whereas AAGGAG repeats form a parallel homoduplex at the DNA level. At the RNA level, the AAGGAG repeats fold into a parallel guanine-quadruplex (G4) while AAG repeats adopt an A-form RNA structure. On the contrary, the CTT and TCTCCT repeats adopt a B-form and CUU and UCUCCU repeats did not form any particular secondary structure under the tested conditions.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryData070224.xlsx
- SourceData.xlsx
- Supplementaryfiguresandcasereports.pdf