



HOKKAIDO UNIVERSITY

Title	Complete nanopore repeat sequencing of SCA27B (GAA-FGF14 ataxia) in Japanese
Author(s)	Miyatake, Satoko; Doi, Hiroshi; Yaguchi, Hiroaki et al.
Citation	Journal of neurology, neurosurgery and psychiatry, 95(12), 1187-1195 https://doi.org/10.1136/jnnp-2024-333541
Issue Date	2024-11-18
Doc URL	https://hdl.handle.net/2115/94097
Rights	This article has been accepted for publication in Journal of neurology, neurosurgery and psychiatry, 2024, following peer review, and the Version of Record can be accessed online at https://doi.org/10.1136/jnnp-2024-333541 . For the avoidance of doubt, this manuscript version is protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.
Type	journal article
File Information	FGF14_miyatake_20240201_main_JNNP.pdf



Original Research

Complete nanopore repeat sequencing of SCA27B (*GAA-FGF14* ataxia) in Japanese

Satoko Miyatake,^{1,2,†} Hiroshi Doi,^{3,†} Hiroaki Yaguchi,^{4,†} Eriko Koshimizu,^{1,†} Naoki Kihara,⁵ Tomoyasu Matsubara,⁵ Yasuko Mori,⁶ Kenjiro Kunieda,⁶ Yusaku Shimizu,⁷ Tomoko Toyota,⁸ Shinichi Shirai,⁴ Masaaki Matsushima,⁴ Masaki Okubo,³ Taishi Wada,³ Misako Kunii,³ Ken Johkura,⁹ Ryosuke Miyamoto,⁵ Yusuke Osaki,⁵ Takabumi Miyama,¹ Mai Satoh,¹ Atsushi Fujita,¹ Yuri Uchiyama,^{1,10} Naomi Tsuchida,^{1,10} Kazuharu Misawa,^{1,11} Kohei Hamanaka,^{1,12} Haruka Hamanoue,² Takeshi Mizuguchi,¹ Hiroyuki Morino,¹³ Yuishin Izumi,⁵ Takayoshi Shimohata,⁶ Kunihiro Yoshida,^{14,15} Hiroaki Adachi,⁸ Fumiaki Tanaka,³ Ichiro Yabe,⁴ and Naomichi Matsumoto^{1, 2, 10}

[†]These authors contributed equally to this work.

Author affiliations:

1. Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan
2. Department of Clinical Genetics, Yokohama City University Hospital, Yokohama, Japan
3. Department of Neurology and Stroke Medicine, Yokohama City University Graduate School of Medicine, Yokohama, Japan
4. Department of Neurology, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo, Japan
5. Department of Neurology, Tokushima University Graduate School of Biomedical Sciences, Tokushima, Japan
6. Department of Neurology, Gifu University Graduate School of Medicine, Gifu, Japan
7. Department of Neurology, Ina Central Hospital, Ina, Japan
8. Department of Neurology, University of Occupational and Environmental Health School of Medicine, Kitakyushu, Japan
9. Department of Neurology, Yokohama Brain and Spine Center, Yokohama, Japan
10. Department of Rare Disease Genomics, Yokohama City University Hospital, Yokohama, Japan
11. Riken Center for Advanced Intelligence Project, Tokyo, Japan

12. Institute for the Advanced Study of Human Biology, Kyoto University, Kyoto, Japan
13. Department of Medical Genetics, Tokushima University Graduate School of Biomedical Sciences, Tokushima, Japan
14. Department of Neurology, JA Nagano Koseiren, Kekeyu-Misayama Rehabilitation Center Kekeyu Hospital, Ueda, Japan
15. Department of Brain Disease Research, Shinshu University School of Medicine, Matsumoto, Japan

Corresponding authors: Prof. Ichiro Yabe and Prof. Naomichi Matsumoto

Prof. Ichiro Yabe

Corresponding author's address: Department of Neurology, Faculty of Medicine and Graduate School of Medicine, Hokkaido University, Sapporo 060-8638, Japan

Corresponding author's phone and fax: +81- 11-706-6028/+81- 11-700-5356

Corresponding author's e-mail address: yabe@med.hokudai.ac.jp

Prof. Naomichi Matsumoto

Corresponding author's address: Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa 236-0004, Japan

Corresponding author's phone and fax: +81-45-787-2606/+81-45-786-5219

Corresponding author's e-mail address: naomat@yokohama-cu.ac.jp

ABSTRACT

Background Although pure GAA expansion is considered pathogenic in SCA27B, non-GAA repeat motif is mostly mixed into longer repeat sequences. This study aimed to unravel the complete sequencing of *FGF14* repeat expansion to elucidate its repeat motifs and pathogenicity.

Methods We screened *FGF14* repeat expansion in a Japanese cohort of 460 molecularly undiagnosed adult-onset cerebellar ataxia patients and 1022 controls, together with 92 non-Japanese controls, and performed nanopore sequencing of *FGF14* repeat expansion.

Results In the Japanese population, the GCA motif was predominantly observed as the non-GAA motif, whereas the GGA motif was frequently detected in non-Japanese controls. The 5'-common flanking variant was observed in all Japanese GAA repeat alleles within normal length, demonstrating its meiotic stability against repeat expansion. In both patients and controls, pure GAA repeat was up to 400 units of length, whereas non-pathogenic GAA-GCA repeat was larger, up to 900 units, but they evolved from different haplotypes, as rs534066520, located just upstream of the repeat sequence, completely discriminated them. Both $(\text{GAA})_{\geq 250}$ and $(\text{GAA})_{\geq 200}$ were enriched in patients, whereas $(\text{GAA-GCA})_{\geq 200}$ was similarly observed in patients and controls, suggesting the pathogenic threshold of $(\text{GAA})_{\geq 200}$ for cerebellar ataxia. We identified 14 patients with SCA27B (3.0%), but their SNP genotype indicated different founder alleles between Japanese and Caucasians. The low prevalence of SCA27B in Japanese may be due to the lower allele frequency of $(\text{GAA})_{\geq 250}$ in the Japanese population than in Caucasians (0.15% vs. 0.32%-1.26%).

Conclusions *FGF14* repeat expansion has unique features of pathogenicity and allelic origin, as revealed by a single ethnic study.

What is already known on this topic: Intronic pure GAA repeat expansion in *FGF14* is a common cause of adult-onset cerebellar ataxia, SCA27B. Non-GAA repeat motif is often mixed into the repeat sequence as the repeat length increases, and their pathogenic significance has remained unclear. Some patients with cerebellar ataxia had $(\text{GAA})_{200-249}$ repeat units, which is below the pathogenic threshold of $(\text{GAA})_{\geq 250}$ repeat units.

What this study adds: GCA motif was predominantly observed as the non-GAA motif in Japanese, whereas GGA motif was frequently observed in non-Japanese. The length of the pure GAA stretch seemed to be limited, while benign GAA-GCA repeat were larger. rs534066520, located just upstream of the repeat, discriminated GAA expansion alleles from GAA-GCA expansion alleles. Both $(\text{GAA})_{\geq 250}$ and $(\text{GAA})_{\geq 200}$ were enriched in a Japanese patient cohort, while $(\text{GAA-GCA})_{\geq 200}$ was similarly observed in patients and controls.

How this study might affect research, practice or policy: Clinically, $(\text{GAA})_{\geq 200}$ can be considered as a possibly pathogenic threshold for cerebellar ataxia. Genetically, *GAA-FGF14* locus is a conceptually new locus: pathogenic pure GAA repeat motif with limited repeat length (up to 400 units), co-existing with differently evolved, benign GAA-GCA repeat motif with much larger repeat length (over 900 units).

INTRODUCTION

Dominantly inherited, intronic GAA repeat expansions in the fibroblast growth factor 14 (*FGF14*) gene have recently been identified as a common cause of adult-onset cerebellar ataxia [*GAA-FGF14* ataxia, spinocerebellar ataxia 27B (SCA27B), MIM 620174].^{1,2} This locus is defined as having 50 GAA repeats in the general population, whereas patients with SCA27B have been reported to have more than 250 GAA repeats. Meanwhile, non-GAA repeat motifs are mixed within the large repeat sequence (explained as “decreased motif purity”¹), in which they are considered not to be pathogenic. In fact, repeat-primed PCR (RP-PCR) for detecting GAA repeat stretches occasionally shows an interrupted waveform instead of a clean sawtooth pattern, suggesting the presence of non-GAA sequences. However, what kinds of sequence motifs are present has not been completely clarified, and nor has how they affect motif purity, to what extent they are elongated, and whether they are truly non-pathogenic. Therefore, we performed a complete sequence analysis of *FGF14* repeat expansion by nanopore long-read sequencing in a Japanese cohort of molecularly undiagnosed adult-onset cerebellar ataxia patients and control individuals. The goal here was to clarify the molecular characteristics of the *GAA-FGF14* locus, as well as the association between sequence expansion and disease.

METHODS

Subjects

A total of 460 Japanese patients with molecularly undiagnosed adult-onset ataxia, 1022 Japanese control individuals, and 92 non-Japanese controls were included in this study. The 1022 Japanese controls consisted of 837 unaffected healthy controls and 185 other disease controls with congenital, pediatric, or multiorgan disorders without neurodegeneration. The 837 unaffected controls included 200 elderly controls and 637 middle-aged controls (table 1). Clinical information was collected from medical records via the medical doctors who saw the patients.

Table 1 Study participants

	PromethION screening	PCR-based screening	Total number of tested individuals
Patients with adult-onset ataxia	61	399	460
Japanese controls	230	792	1022
Non-Japanese controls	92	0	92

Genetic analyses

All genetic analyses were performed with genomic DNA extracted using standard methods, from peripheral blood leukocytes, lymphoblastoid cells established from the patient's lymphocytes, or frozen autopsied brain tissue. For 61 patients, 230 Japanese controls, and 92 non-Japanese controls, we performed long-read whole-genome sequencing (LR-WGS) on PromethION (Oxford Nanopore Technologies, Oxford, UK). For 399 patients and 792 Japanese controls, we first screened *FGF14* repeat expansion using PCR-based assay as described below, and then performed either targeted long-read sequencing (T-LRS) using GridION (Oxford Nanopore Technologies) or fragment analysis of screening-positive samples (table 1).

PCR-based *FGF14* repeat expansion screening

We performed screening of *FGF14* repeat expansion via the following steps.

Step 1: Flanking PCR

Flanking PCR was performed as previously described,¹ except that we used KOD Fx Neo (TOYOBO, Osaka, Japan) as a DNA polymerase (supplemental table 1). Samples with a PCR amplicon of 700 bp or larger, corresponding to an *FGF14* repeat count of approximately 150 units or more on agarose gel electrophoresis, were then subjected to two kinds of repeat-primed PCRs (RP-PCRs).

Step 2: RP-PCR

To detect pure GAA repeat stretches, we performed RP-PCR as described previously.¹ To detect GAA-GCA repeat stretches, we performed RP-PCR following the same PCR conditions as for the pure GAA repeat stretch, except for a reverse primer targeting the GAA-GCA repeat sequence. Primer information and PCR settings are presented in supplemental table 1, and the representative RP-PCR patterns for pure GAA and non-pure GAA repeat stretch are shown in supplemental figure 1. Samples that showed a complete sawtooth pattern only with RP-PCR for the pure GAA repeat stretch were considered positive for pure GAA *FGF14* repeat expansion, and further subjected to T-LRS if sufficient DNA was available. Otherwise, we performed fragment analysis to confirm the repeat length.

Step 3: Fragment analysis

For eight unaffected controls who were positive for pure GAA repeat expansion but without sufficient DNA for T-LRS, fragment analysis was performed as described in supplemental table 1. Five samples whose repeat length had been confirmed by T-LRS also underwent fragment analysis at the same time to draw a standard curve. For eight samples, the repeat length equivalent to the repeat length in T-LRS was calculated using the standard curve method.

Next-generation sequencing

LR-WGS on the PromethION sequencer (Oxford Nanopore Technologies) or T-LRS using the adaptive sampling option implemented on GridION (Oxford Nanopore Technologies) was performed in accordance with the manufacturer's instructions at Yokohama City University, as described previously.^{3,4} For T-LRS, we assigned the entire genomic *FGF14* (NM_175929.3) region and the regions 100 kb upstream/downstream of it using a bed file (Chr13:101610804–102502457 on GRCh38), and used a time-lag sampling method (two samples on one flowcell).⁴ For data analysis, sequences were base-called using Guppy v6.4.2 in either high-accuracy mode or super-accuracy mode, and aligned to GRCh38 using either minimap2 v2.14 (<https://github.com/lh3/minimap2>) or LAST 1132 (<https://gitlab.com/mcfrith/last>). Tandem-genotypes v1.9.0 (<https://github.com/mcfrith/tandem-genotypes>) was used to find changes in the length of tandem repeats of selected loci with the crude allele prediction option, and consensus sequence generation was performed with the merge option. Detailed repeat analyses were performed using RepeatAnalysisTools (<https://github.com/PacificBiosciences/apps-scripts/tree/master/RepeatAnalysisTools>). The rs72665334 genotype was confirmed by manual inspection of GAA-*FGF14* ataxia patients' LR data on IGV with a read depth of more than 20×. For phasing, we analyzed T-

LRS data from 10 individuals with pure GAA repeat expansion ≥ 200 and 15 individuals with GAA-GCA repeat expansion ≥ 200 using PEPPER-Margin-DeepVariant (<https://github.com/kishwarshafin/pepper>).⁵ We constructed haplotypes nearby the *FGF14* repeat locus after confirming the detected SNVs by manual inspection of phased bam files (of long read sequencing data) using Integrative Genomics Viewer.⁶

Statistical analysis

To assess the significance of differences between the two groups in terms of continuous and categorical variables, the Mann–Whitney U test and Fisher’s exact test were employed. Correlation between the number of repeat units and age at onset was assessed using Pearson’s correlation test. All statistical analyses were performed using GraphPad Prism v8.0.2 (GraphPad Software, San Diego, CA USA) software. A two-tailed *P* value < 0.05 was considered to reflect statistical significance.

RESULTS

GAA-GCA repeat has a tendency to be larger than pure GAA repeat in Japanese

For 61 patients with molecularly undiagnosed ataxia and 230 individuals with either other non-neurodegenerative disorders or parents unaffected by such diseases, LR-WGS had been performed for research purposes using the nanopore PromethION sequencer (Oxford Nanopore Technologies) in our laboratory. Thus, we surveyed the data and confirmed the repeat unit count and complete repeat sequence of the *FGF14* repeat locus among patients and controls. In Japanese, GCA is the dominantly observed motif with the configuration variation of “(GAA)_{1–4}(GCA)₁” as the non-GAA motif, while GGA motifs were also observed in 92 non-Japanese individuals. This is compatible with a previous report describing that the GGA motif was frequently observed in a predominantly Caucasian cohort.¹ While the length of the pure GAA repeat stretch seemed to be within a limited extent (up to approximately 400 repeat units), repeat length with non-GAA motif can be much more to the maximum of 2760 bp, corresponding to approximately 900 repeat units, in our cohort. Among 33 individuals with GCA repeat expansion, one individual had a repeat length of 501 bp corresponding to 167 repeat units while the rest of 32 individuals had repeat length of approximately 850 bp (corresponding to approximately 280 repeat units) or more. When the repeat length exceeded 1500 bp (corresponding to approximately 500 repeat units), the repeat sequence always had a non-GAA motif.

After detecting individuals with GAA-*FGF14* SCA27B using PCR-based assay, we confirmed their accurate repeat sequencing and repeat length by T-LRS on a GridION sequencer. After complete sequence data from PromethION and GridION sequencers were visualized together in a waterfall-style plot, a characteristic tendency for *FGF14* repeat expansion in the Japanese cohort was consistently noted (figure 1).

A recent study has reported that there is a 17-bp deletion-insertion common 5'-flanking variant (5'-CFV) that is associated with enhanced mitotic stability of the *FGF14* repeat locus.⁷ Regarding 91 individuals shown in figure 1, all alleles with an *FGF14* repeat length of normal range (less than approximately 120 bp, corresponding to 40 repeat units) have 5'-CFV, except for one individual with minimally altered 5'-CFV (TAGTCATAGTACCCCAA to TAGTCATAGTACCTTAA), while none of the alleles with either pure GAA or non-GAA repeat expansion of more than 50 repeat units has 5'-CFV. We also assessed alleles from Japanese controls with biallelic *FGF14* repeat within normal length (up to approximately 40 repeat units) and found that 94.8% of the evaluated alleles (398/420 alleles) had 5'-CFV and remaining alleles (5.2%, 22/420 alleles) had 5'-CFV with minimally but diversely altered sequence (supplemental figure 2). Taken together, these results are compatible with the previous study. Next, we investigated whether there is a single nucleotide polymorphism (SNP) in the flanking region of the *FGF14* repeat locus, which could distinguish alleles with either pure GAA or GAA-GCA repeat expansion. rs534066520 (GRCh38, chr13:102161565 T>A), which is just upstream of the 5'-CFV insertion site, had the alternate genotype A in all GAA-GCA repeat expansion alleles, whereas the wild-type genotype (i.e., T) was found in pure GAA repeat expansion alleles. The Japanese alleles with *FGF14* repeat within normal length never had GAA-GCA repeat motif, but had the wild-type rs534066520 genotype (supplemental figure 2). Although this is a limited observation, this SNP might be a marker to differentiate the possible pure GAA repeat expansion allele from all alleles positive on *FGF14* flanking PCR (which has no 5'-CFV). Phasing of 25 T-LRS data around *FGF14* repeat locus obtained from the individuals with either pure GAA repeat expansion ≥ 200 or GAA-GCA repeat expansion ≥ 200 further suggested that these two alleles may have different haplotypes which could be derived from different ancestors (supplemental figure 3).

rs72665334 (chr13:102150076 C>T on GRCh38) was reported to be associated with downbeat nystagmus syndrome by GWAS,⁸ as part of a disease haplotype shared by three Australian patients with GAA-*FGF14* ataxia,² and to be associated with GAA-*FGF14* with repeat count ≥ 250 .⁹ The minor allele frequency of the SNP is 0.000378 in Japanese Multi-Omics Reference Panel (jMorp) 54KJPN (sample size: approximately 54,000 Japanese; <https://jmorp.megabank.tohoku.ac.jp/>) and 0.00116 in the genome Aggregation Database (gnomAD) v4.1.0 for East Asians (EAS), while the overall frequency in gnomAD v4.1.0 is 0.0481

(https://gnomad.broadinstitute.org/variant/13-102150076-C-T?dataset=gnomad_r4). Its frequency is also up to 7% in the European population in the 1000 Genomes Project.¹⁰ These data suggest that this SNP is particularly rare in Japan compared with its rate in the Caucasian population. Among 14 patients with *GAA-FGF14* in our cohort, none had rs72665334, implying that *GAA-FGF14* repeat expansion in the Japanese may have arisen from a different haplotype background than that in Caucasian patients.

Pure GAA ≥ 200 repeat units is significantly enriched in Japanese patients with adult-onset ataxia

In addition to the LR-WGS analyses of 61 patients and 230 controls, we extended our PCR-based *GAA-FGF14* screening to 399 patients with molecularly undiagnosed adult-onset ataxia, 592 unaffected middle-aged parents of a child with congenital/pediatric diseases, and 200 unaffected elderly individuals; thus, our cohort consisted of 460 patients and 1022 controls in total (table 1). We performed T-LRS on the individuals for whom PCR-based screening suggested the existence of pure GAA expansion.

In the first reports on *FGF14*-GAA ataxia,^{1,2} the pathogenic threshold had been proposed with a cut-off of 250 repeat units (≥ 300 –335 repeats as full mutation, ≥ 250 as likely pathogenic with reduced penetrance), but the possibility of pathogenic potential upon a repeat count of 200–250 was later raised in a few papers.^{9, 11, 12} As for the repeat motif, non-GAA motifs such as GCA and GGA other than the pure GAA motif were suggested to be non-pathogenic.^{1, 11} To evaluate which repeat motif and length are associated with ataxia, we set the threshold as ≥ 200 repeats and compared the frequency of individuals with respective repeat motif and length in both patients and controls: (1) pure GAA repeat with ≥ 250 repeat units, (2) pure GAA repeat with ≥ 200 repeat units, (3) GAA-GCA repeat with ≥ 200 repeat units, and (4) any triplet repeat units with ≥ 200 repeats. The alleles of (1) pure GAA repeat with ≥ 250 repeat units were enriched in patients (8/460, 1.7%) compared with the level in controls [3/1022, 0.29%; OR 6.012 (95% CI 1.802–21.03), $P=0.005$]. The alleles of (2) pure GAA repeat with ≥ 200 repeat units were similarly enriched in patients (14/460, 3%) compared with the level in controls (5/1022, 0.49%, OR 6.385, 95% CI 2.246–16.18, $P=0.0001$). Meanwhile, the frequencies of the alleles of (3) GAA-GCA repeat with ≥ 200 repeat units and (4) any repeat with ≥ 200 repeat units were not significantly different between patients [(3) 16/460, 3.5% and (4) 30/460, 6.5%] and controls [(3) 44/1022, 4.3%, $P=0.6$ and (4) 49/1022, 4.8%, $P=0.2$], respectively (figure 2). Our results support the scenario that pure GAA repeat with ≥ 200 repeats may be pathogenic with reduced penetrance, and GAA-GCA repeat expansion is not pathogenic despite the tendency to be longer than pure GAA repeat expansion (figure 2 and supplemental table 2).

Identification of GAA-*FGF14* SCA27B ataxia in Japanese ataxia cohort and its clinical analysis

On the basis of the above observation, we identified 14 patients with GAA-*FGF14* ataxia, eight of whom have a repeat count of ≥ 250 (tables 2, 3, supplemental tables 3, 4, and supplemental figure 4). The detection rate in our limited Japanese cohort was 3%. Seven of the patients were male, and seven of them had a family history compatible with autosomal dominant inheritance. Mean age at onset was 69.1 years, with mean disease duration of 9 years. Although there was only a single case observation of Patient 2, the Scale for the Assessment and Rating of Ataxia (SARA), scored from 2007 to 2016, implied a slowly progressive course of this disease (supplemental figure 5).

Table 2 Clinical features of patients with SCA27B, (GAA)_{>250}

Patient	1	2^a	3	4	5^b	6	7	8
Age at onset	55	64	73	74	78	56	66	68
Family history	+	-	+	+	-	+	-	-
Long-read sequencing platform	GridION	GridION	GridION	GridION	PromethION	PromethION	GridION	GridION
GAA repeat unit motif count	17/402	14/363	8/300	10/293	9/282	17/269	9/259	26/254
Pathogenicity	Full mutation	Full mutation	Full mutation	Pathogenic	Pathogenic	Pathogenic	Pathogenic	Pathogenic
rs72665334 genotype	C/C	C/C	C/C	C/C	C/C	C/C	C/C	C/C
Clinical features								
Initial symptom	Unstable gait	Unstable gait, dysarthria	Unstable gait, dysarthria	Dysesthesia of lower extremities	Unstable gait	Unstable gait	Lightheadedness	NA
Episodic symptom	-	NA	+	-	-	-	-	NA
Nystagmus (any/downbeat)	+/+	NA	-/-	+/+	+/-	-/-	+/+	+/+
Ataxia (gait/appendicular)	+/+	+/+	+/-	+/+	+/+	+/+	-/+	+/NA
Vertigo/dizziness	-	NA	+	+	+	-	+	NA
Upper motor neuron sign	-	NA	-	-	+	-	-	NA
Lower motor neuron sign	-	NA	-	-	-	-	-	NA
Sensory impairment	-	NA	-	+	-	+	-	NA
Vestibular dysfunction	-	NA	-	-	NA	-	-	NA
Cerebellar atrophy	+	+	+	-	+	±	-	NA
Abnormality in NCS	-	NA	NA	+	NA	+	-	NA

^a: lost to follow-up; ^b: deceased; rs72665334 genotype: C=wild type, T=alternate; NCS: nerve conduction study; sensory impairment includes superficial and deep sensation.

Table 3 Clinical features of patients with SCA27B, (GAA)₂₀₀₋₂₅₀

Patient	9	10^b	11	12	13	14^b
Age at onset	71	73	72	77	75	65
Family history	-	+	+	-	-	+
Long-read sequencing platform	GridION	GridION	GridION	PromethION	GridION	GridION
GAA repeat unit motif count	9/242	10/240	9/230	9/229	152/227	10/213
Pathogenicity	Possibly pathogenic	Possibly pathogenic	Possibly pathogenic	Possibly pathogenic	Possibly pathogenic	Possibly pathogenic
rs72665334 genotype	C/C	C/C	C/C	C/C	C/C	C/C
Clinical features						
Initial symptom	Oscillopsia	Dizziness, vomiting	Unstable gait, dysarthria + (fluctuating)	Dysarthria, dysphagia, unstable gait +	Unstable gait	Unstable gait
Episodic symptom	-	+			-	-
Nystagmus (any/downbeat)	+/+	+/-	+/-	-/-	+/-	-/-
Ataxia (gait/appendicular)	-/+	+/+	+/+	+/+	+/+	-/+
Vertigo/dizziness	-	+	-	+	+	NA
Upper motor neuron sign	+	-	-	+	-	-
Lower motor neuron sign	-	-	-	+	-	-
Sensory impairment	-	-	-	+	+	-
Vestibular dysfunction	-	NA	-	+	-	-
Cerebellar atrophy	-	+ (mild-slight)	-	+	+	+
Abnormality in NCS	-	NA	-	+	NA	NA

^b: deceased; rs72665334 genotype: C=wild type, T=alternate; NCS: nerve conduction study; sensory impairment includes superficial and deep sensation.

Average repeat count among patients was 271.6 based on the motif count of consensus repeat sequence. A negative correlation between the age of onset and repeat count was observed (Pearson's correlation, $R^2=0.2590$, $P=0.0631$) (supplemental figure 6).

Among the patients for whom clinical information was available, ataxia is the most observed feature (appendicular ataxia 12/13, 92.3%; gait ataxia 11/14, 78.6%), followed by dysarthria (11/13, 84.6%), oculomotor symptoms (nystagmus of any kind 9/13, 69.2%; saccadic eye movement 8/12, 66.7%), vertigo/dizziness (7/11, 63.6%), and sensory impairment (4/12, 33.3%). Cerebellar atrophy was observed in 69.2% (9/13), but was mostly very slight to mild. Abnormal findings on nerve conduction study were observed in 3/7 patients (42.9%), while vestibular dysfunction was relatively uncommon (1/10, 10%). When comparing patients with ≥ 250 repeat units to those with 200–249, we noticed that these two groups of patients generally manifest similar clinical characteristics (figure 3), further supporting the assertion that a repeat count of 200–249 may also cause SCA27B. Among these patients, episodic symptom and diplopia/oscillopsia were more frequently observed in the latter group (episodic symptom: 1/6, 16.7% vs. 3/6, 50%; diplopia/oscillopsia 0/6, 0% vs. 2/6, 33.3%), while downbeat nystagmus and positional nystagmus were more common in the former group (downbeat nystagmus: 4/7, 57.1% vs. 1/6, 16.7%; positional nystagmus: 3/6, 50% vs. 1/6, 16.7%), but no statistical significance was reached for the differences in any of these features.

DISCUSSION

Nanopore sequencing of the complete repeat sequence of the *GAA-FGF14* locus from 58 Japanese individuals, including 32 patients with ataxia and 26 controls with *FGF14* repeat expansion of any triplet motif (> 50 repeat units), revealed its unexpected repeat count and motif distribution in Japanese individuals: Pathogenic pure GAA stretch had only limited length (up to 400 repeat units), whereas benign GAA-GCA repeat mostly existed with > 280 repeat units, being larger than the pure GAA stretch (up to 900 repeat units). We added 33 elongated repeat sequences of non-Japanese to the control group, and confirmed the same trend; we also found that the GGA motif is more commonly observed than in Japanese controls (16/33 individuals vs. 2/26 in Japanese controls). In repeat expansion diseases, it may be assumed that the pathogenic repeat expansion can evolve from one founder haplotype through replication slippage or non-homologous recombination,¹³ or by an interrupting mutation in a reference repeat sequence. This can potentially lead to an aberrant secondary/tertiary nucleotide structure that is more prone to DNA damage as well as the initiation of repeat expansion or genomic instability across meiotic and mitotic cell divisions. Across the generations, repetitive sequences acquire instability with an

intermediately expanded sequence, and then reach full expansion to manifest disease. Generally, a longer repeat length is likely to have a greater pathogenic impact. In a few recently identified non-coding repeat expansion diseases, such as spinocerebellar ataxia 31 (SCA31),¹⁴ familial adult myoclonic epilepsy (FAME),¹⁵ and cerebellar ataxia neuropathy vestibular areflexia syndrome (CANVAS),¹⁶ wild-type or benign repeat motifs may be expanded as polymorphisms. In such loci, pathogenic repeat expansion includes the aberrant, pathogenic repeat motif stretch as at least part or all of the repetitive sequence, and the length of pathogenic repeat expansion is generally longer than that of polymorphic expansion. Thus, it has also been considered that an aberrant repeat motif stretch with longer expansion may imply pathogenicity. However, our data suggest that the *GAA-FGF14* locus is a conceptually new locus with a pathogenic GAA (wild-type) repeat motif expansion with limited length, co-existing with much larger, but benign GAA-GCA (aberrant) repeat expansion.

One of the merits of applying the long-read sequencing approach to repeat expansion disease is that we can identify the complete repeat sequence along with the surrounding sequence, which enables us to phase the two alleles and confirm the repeat expansion-associated haplotype. Our cohort replicated the recently reported finding for 5'-CFV as a sequence that is possibly protective and can confer stability and avoid pathogenic expansion, while the marker SNP (rs534066520) possibly linked with GAA-GCA expanded allele was also detected. supplemental table 5 presents the association of the repeat motif, the neighboring SNP and sequence (5'-CFV), and their consequences. This SNP may imply that GAA-GCA repeat expansion did not evolve from a pure GAA expansion allele (figure 1). This observation could explain the seemingly peculiar distribution: non-pathogenic repeats (GAA-GCA) are larger than pathogenic repeats (GAA) of *FGF14*.

The allele frequency of rs534066520 has been registered in the gnomAD v4.1.0, as 0.001848 in global (181/97926 alleles), 0.003029 in non-Finnish European (127/41926 alleles), 0.001138 in Admixed American (11/9668 alleles), and 0.0003399 in East Asian (1/2942 alleles) (https://gnomad.broadinstitute.org/variant/13-102161565-T-A?dataset=gnomad_r4), while it is 0.0339 in Korean Reference Genome Database (sample size 2922; <https://www.ncbi.nlm.nih.gov/snp/rs534066520>). Although we did not find data on the allele frequency in Japanese, difference in allele frequency between Korean and Caucasian community may explain why GAA-GCA repeat allele is rare in non-Japanese, considering Korean and Japanese are both in Far East Asia.

Our observation of a single ethnic cohort (Japanese) warranted a finer resolution for the case-control study, and thus provides additional evidence to support the threshold for possible manifestation of SCA27B as pure (GAA)_{≥200} with reduced penetrance, as speculated in previous reports.^{9, 11, 12} Specifically, our work shows

that pure (GAA)_{≥200} repeat units were significantly enriched in Japanese patients with adult-onset ataxia, and patients with (GAA)_{200–249} and those with (GAA)_{≥250} had very similar clinical characteristics. Furthermore, we showed that wild-type rs534066520 was associated with both (GAA)_{≥250} and (GAA)_{200–249} repeat units.

Our study using only a Japanese population could advantageously identify rs534066520 (reference allele T) associated with pure GAA expansion from 50 to 400 repeat units but not with GAA-GCA expansions (supplemental table 5). This may support (GAA)_{200–249} may originated from the same founder allele of (GAA)_{≥250}. If multiple populations were involved in this study, it might be difficult to identify rs534066520 which could discriminate the founder allele of (GAA)_{200–249} from that of GAA-GCA expansions.

We identified 14 patients with GAA-*FGF14* SCA27B out of 460 individuals in the molecularly undiagnosed adult-onset Japanese ataxia cohort; thus, the detection rate for our cohort was 3%. If the criteria to consider a case as SCA27B were limited to pure GAA repeat with ≥ 250 repeats, the detection rate would be 1.7% (8/460). GAA-*FGF14* SCA27B is currently considered to be one of the most common forms of SCA regardless of country/ethnicity. The prevalence has been reported to be 61% in French-Canadians,¹ 14%–18% in Germans,¹ 2,¹¹ 14%–15% in Australians,^{1,2} 13% in French,¹⁷ 9% in Brazilians,¹⁸ and 10% in Indians,¹ under the pathological threshold setting of pure GAA repeat with ≥ 250 repeats. However, a recent report showed relatively low prevalence (1.2%, 11/940) in Japan under the pathological threshold setting of pure GAA repeat with a repeat count of ≥ 250 ,¹⁹ which our data further replicated. One possible reason why GAA-*FGF14* SCA27B is relatively rare in Japan is that the allele frequency with pure GAA repeat with ≥ 250 repeats in the Japanese population is much lower (0.15 %) than those in Caucasian controls, such as 1.26%¹ in Germans, 0.72% in French-Canadians,¹ and 0.32% in mixed samples of Australians and those only known to be of Caucasian ancestry.² This may be related to the fact that rs72665334, a GAA-*FGF14* repeat expansion-associated SNP, is very rare in the Japanese. (The minor allele frequency of the SNP is 0.000378 in jMorp 54KJPN versus up to 7% in the European population in the 1000 Genomes Project.)

Although our cohort has a limited number of patients with SCA27B, their clinical features are as previously reported: later onset, slowly progressive, pancerebellar feature with predominant unstable gait, oculomotor symptoms, very mild to slight cerebellar atrophy, and with episodic/fluctuating symptom or downbeat nystagmus as possible key features for clinicians to consider this disease. Some cases were accompanied by sensory symptom and neuropathy, and thus CANVAS can be one of the differential diagnoses, which is also compatible with previous reports.²⁰

Acknowledgement

We thank all participants for their cooperation in this research. We are grateful to Ms. N. Watanabe, Ms. K. Takabe, and Ms. S. Sugimoto from the Department of Human Genetics, Yokohama City University Graduate School of Medicine, for their technical assistance. We also thank Edanz (<https://jp.edanz.com/ac>) for editing a draft of this manuscript.

Contributors

S.M., H.D., H.Y., F.T., I.Y., and N.M. contributed to the conception and design of the study; E.K., N.K., T.Ma., Y.M., K.K., Y.S., T.T., S.S., M.M., M.O., T.W., M.K., K.J., R.M., Y.O., T.Miy., M.S., A.F., Y.U., N.T., K.M., K.H., H.H., T.Miz., H.M., Y.I., T.S., K.Y., H.A., F.T., I.Y., and N.M. contributed to the acquisition and analysis of data; S.M. and N.M. contributed to drafting the text. All authors approved the manuscript.

Funding

This work was supported by the Japan Agency for Medical Research and Development (AMED) under grant numbers JP23ek0109674, JP23ek0109549, JP23ek0109617 (N. Matsumoto), JP23ek0109648 (F. Tanaka, N. Matsumoto, H.Doi), and JP23bm1423020 (H.Doi, F. Tanaka); JSPS KAKENHI under grant numbers JP23H02829 (S. Miyatake) and JP21K07869 (E. Koshimizu); the Research Committee on the Medical Basis of Motor Ataxias, Health and Labor Sciences Research Grants of the Ministry of Health, Labor and Welfare, Japan under grant number 23FC1010 (I.Yabe); and the Takeda Science Foundation with no grant numbers (T. Mizuguchi, N. Matsumoto).

Competing interests

Nothing to report.

Patient consent for publication

Not applicable.

Ethics approval

The experimental protocol was approved by the Committee for Ethical Issues at Yokohama City University School of Medicine under reference number A170525011/B230600048(modified). Written informed consent was obtained from all individuals or from the guardians for the pediatric patients, in accordance with the Declaration of Helsinki.

Data availability statement

The anonymized data obtained in this study are available from the corresponding author, upon reasonable request.

Supplemental material

This content has been supplied by the author(s) .

REFERENCES

1. Pellerin D, Danzi MC, Wilke C, et al. Deep Intronic FGF14 GAA Repeat Expansion in Late-Onset Cerebellar Ataxia. *N Engl J Med* 2023;388:128-41.
2. 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* 2023;110:105-19.
3. Miyatake S, Yoshida K, Koshimizu E, et al. Repeat conformation heterogeneity in cerebellar ataxia, neuropathy, vestibular areflexia syndrome. *Brain* 2022;145:1139-50.
4. Miyatake S, Koshimizu E, Fujita A, et al. Rapid and comprehensive diagnostic method for repeat expansion diseases using nanopore sequencing. *NPJ Genom Med* 2022;7:62.
5. Shafin K, Pesout T, Chang PC, et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods* 2021;18:1322-32.
6. Robinson JT, Thorvaldsdottir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol* 2011;29:24-6.
7. Pellerin D, Gobbo GD, Couse M, et al. A common flanking variant is associated with enhanced meiotic stability of the FGF14 -SCA27B locus. *bioRxiv* 2023.
8. Strupp M, Maul S, Konte B, et al. A Variation in FGF14 Is Associated with Downbeat Nystagmus in a Genome-Wide Association Study. *Cerebellum* 2020;19:348-57.
9. Pellerin D, Heindl F, Wilke C, et al. GAA-FGF14 disease: defining its frequency, molecular basis, and 4-aminopyridine response in a large downbeat nystagmus cohort. *EBioMedicine* 2024;102:105076.
10. Genomes Project C, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature* 2015;526:68-74.
11. 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;38:1557-8.

12. Mizushima K, Shibata Y, Shirai S, et al. Prevalence of repeat expansions causing autosomal dominant spinocerebellar ataxias in Hokkaido, the northernmost island of Japan. *J Hum Genet* 2024;69:27-31.
13. Tautz D, Trick M, Dover GA. Cryptic simplicity in DNA is a major source of genetic variation. *Nature* 1986;322:652-6.
14. Sato N, Amino T, Kobayashi K, et al. Spinocerebellar ataxia type 31 is associated with "inserted" pentanucleotide repeats containing (TGGAA)_n. *Am J Hum Genet* 2009;85:544-57.
15. Ishiura H, Doi K, Mitsui J, et al. Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy. *Nat Genet* 2018;50:581-90.
16. Cortese A, Simone R, Sullivan R, et al. Biallelic expansion of an intronic repeat in RFC1 is a common cause of late-onset ataxia. *Nat Genet* 2019;51:649-58.
17. Wirth T, Clement G, Delvallee C, et al. Natural History and Phenotypic Spectrum of GAA-FGF14 Sporadic Late-Onset Cerebellar Ataxia (SCA27B). *Mov Disord* 2023;38:1950-6.
18. 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* 2023;9:e200094.
19. 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.
20. 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* 2024;95:175-9.

FIGURE LEGENDS

Figure 1. Complete *FGF14* repeat sequence in Japanese. Waterfall plot of (A) Japanese controls (n=26), (B) Japanese and non-Japanese controls (n=26 and 33, respectively), and (C) patients with SCA (n=32, including 14 patients with SCA27B). Left panels show sequences of the smaller allele with shorter repeat expansion from each individual stacked in order of repeat length, while right panels show those of the larger allele. Note that Japanese control alleles with repeat length of 1000 bp or more have GCA motif, while non-Japanese control alleles with such repeat length may have GGA motif.

Figure 2. Pure GAA with ≥ 200 repeat units was enriched in patients with SCA. (A) Odds ratios associated with manifesting SCA for the four different repeat lengths and motif characteristics. (B) The numbers of controls and patients in each category.

Figure 3. Comparison of clinical features between patients with pure GAA repeat units of ≥ 250 and pure GAA repeat units of 200–249. For each clinical sign shown on the vertical axis, the proportion of positive signs is shown in red for patients with pure GAA repeat units of ≥ 250 (left panel) and in blue for patients with pure GAA repeat units of 200–249 (right panel).





