Identification of new repeat expansion diseases

Hiroyuki Ishiura[1,2]

[1] Department of Neurology, Graduate School of Medicine, The University of Tokyo
[2] Department of Neurology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences

Corresponding author
Hiroyuki Ishiura
Department of Neurology, Okayama University Graduate School of Medicine, Dentistry and Pharmaceutical Sciences
2-5-1 Shikata-cho, Kita-ku, Okayama 700-8558, JAPAN
ishiura@s.okayama-u.ac.jp
Tel 81-86-235-7365
Fax 81-86-235-7368

Short title: New repeat expansion diseases

Abstract (250 words)

Through a genetic study of benign adult familial myoclonus epilepsy (BAFME) type 1, TTTCA and TTTTA repeat expansions have been identified in intron 4 of *SAMD12*. Lengths of expanded repeats inversely correlated with age at onset of epilepsy. Gain-of-toxic function mechanisms are suggested by the presence of UUUCA-repeat-containing RNA foci. From families with BAFME who did not have repeat expansions in *SAMD12*, we identified expanded TTTCA and TTTTA repeats in *TNRC6A* and *RAPGEF2*. These findings indicated a strong correlation between the repeat motif and the phenotype, leading to the identification of other types of BAFME. We then conducted genetic analysis of neuronal intranuclear inclusion disease (NIID), oculopharyngeal myopathy with leukoencephalopathy (OPML), and oculopharyngodistal myopathy (OPDM). From the observation that NIID, OPML, and OPDM, in addition to fragile X-associated tremor/ataxia syndrome, have shared clinical features, a direct search for CGG repeat expansions successfully led to the identification of the causative genes. Here, I review recent studies on repeat expansions.

1. Introduction

Through linkage analysis, many hereditary conditions in large families have been uncovered since the 1990s. After the development of next-generation sequencing technology, many more causative genes have been found especially in families with *de novo* mutations and in small nuclear families.[1–3] However, finding the causative genes for some diseases is difficult. Disease-causing variants might be located in introns or regulatory regions where it is difficult for us to determine the pathogenicity of variants. Disease-causing variants might not be simple: large duplications, large deletions, insertions of unknown sequences, inversions, and repeat expansions are hardly visible with standard variant calling methods after short-read sequence analysis. There might be a reduced penetrance (some mutation carriers do not show a phenotype) or phenocopies (some mutation noncarriers show a similar phenotype to those of carriers), which substantially reduces the power of identifying disease-causing variants.

As an awardee of the Japanese Society of Neurology, I will provide an overview of the paper that contributed to my award, review recent studies by ours and others on hereditary neurological and muscular diseases, and show how we overcame the difficulties described above. The studies highlight the importance of noncoding repeat expansions. On the basis of our findings, we propose a new concept of repeat motif–phenotype correlation and suggest the role of repeat expansions in the pathogenesis of diseases.

2. Benign adult familial myoclonic epilepsy

2.1 Overview

Benign adult familial myoclonus epilepsy (BAFME) is an autosomal dominant disease characterized by infrequent epilepsy and cortical tremor. Many cases have been identified in Japan, and the disease concept was clinically, electrophysiologically, and pathologically established in Japan around 1990;[4–6] the first case report of the disease can be traced back to the 1920s.[4] Seizures can be easily controlled with antiepileptics such as sodium valproate, clonazepam, and primidone. Recently, the effects of low-dose perampanel have also been reported.[7] Patients show photosensitivity. Tremulous movement of the hands is also characteristic of the disease. This involuntary movement is always seen and not directly related to epilepsy. It is aggravated by mental stress, fatigue, or sleep deprivation, and it responds to alcohol intake in some patients. Electrophysiologically, it is considered to be myoclonus of cortical origin, as shown by giant somatosensory evoked potentials, C reflex, and a myoclonus-related cortical activity revealed by the jerk-locked back averaging method. It is virtually nonprogressive

or minimally progressive. Its diagnostic criteria have been proposed, in which the clinical features are well summarized.[8]

The disease is also known as familial adult myoclonic epilepsy (FAME), familial cortical myoclonic tremor with epilepsy (FCMTE), or autosomal dominant cortical tremor, myoclonus, and epilepsy (ADCME). Of note, Yasuda[6] originally named this disease "benign adult familial myoclonic epilepsy (BAFME)", which is the most widely used name in Japan, but the term "benign adult familial myoclonus epilepsy" is also used because cortical myoclonus in this disease seems to occur independently from seizures. Terms such as FAME and FCMTE are used in Europe and other countries; however, I prefer to use BAFME in this paper.

2.2 History of genetic analysis

Linkage analyses were performed on Japanese families[9–12] and a Chinese family,[13] and the results showed a linkage to the long arm of chromosome 8 (called BAFME type 1). Exons and splicing regions were thoroughly examined in a Japanese family, which did not suggest a pathogenic variant.[12]

Of note, similar diseases have been identified in Europe, Australia, New Zealand, and Thailand. Linkage studies showed linkages to chromosomes 2,[14] 5,[15] and 3,[16] which are called BAFME types 2, 3, and 4, respectively. The studies indicated genetic heterogeneity but also suggested a founder effect in the disease.[17]

2.3 Identification of repeat expansion mutations in BAFME

We enrolled 51 Japanese BAFME families in a study. By linkage analysis, we confirmed a linkage to chromosome 8, but failed to further narrow down the candidate region on 8q24. To further narrow down the candidate region, we performed haplotype analysis hypothesizing that there is a founder effect in the disease. This is partly because we learned from previous studies that a founder effect can exist if a frequency of disease is particularly high in a region.[18–21] As a result, we finally found a founder chromosome, resulting in the narrowing down of a candidate region from 4.2 Mb to 134 kb, which only included exon and surrounding introns of *SAMD12*.

After excluding exonic variants as a cause of the disease, we performed Sanger sequencing for intronic variants using polymerase chain reaction (PCR) products. The analysis unexpectedly revealed a repeat sequence, which appeared to be discordant between affected parent–offspring pairs. The observation prompted us to consider the presence of repeat expansions. Visual inspection of whole-genome sequencing data actually revealed the presence of repeat expansions in intron 4 of *SAMD12*, which

consisted of expanded TTTTA and TTTCA repeats, where the reference sequence mainly consists of TTTTA repeats. While the TTTTA repeats were expanded in a limited number of controls, no TTTCA repeats were found in control subjects, indicating an important role of TTTCA repeats. In line with this idea, RNA foci containing UUUCA repeats have been identified in neurons of the autopsied brains from patients with BAFME1.[22] Lengths of expanded repeats inversely correlated with age at onset of epilepsy. We considered that these findings suggested an RNA-mediated gain-of-function mechanism caused by TTTCA repeats.

At that time, we identified two BAFME families without repeat expansions in *SAMD12*. It was expected that expanded TTTCA repeats in other genes can also cause BAFME if the RNA-mediated gain-of-function mechanism in BAFME was true. Through the analysis, we noticed that repeat expansions can be identified by evaluating short reads. That is, if expanded TTTCA repeats are present, we should find many short reads with TTTCA repeats when we perform whole genome sequencing. Using TRhist,[23] which efficiently analyzes short reads filled with repeat sequences, we further identified TTTCA and TTTTA repeat expansions in *TNRC6A* and *RAPGEF2* from two families without repeat expansions in *SAMD12* (Fig. 1a,b). Again, whereas the reference sequence mainly consists of TTTTA repeats and the TTTTA repeats were sometimes expanded in controls, TTTCA is exclusively found in BAFME patients, indicating the importance of TTTCA repeats. The diseases are designated as BAFME types 6 and 7 for these expansions in *TNRC6A* and *RAPGEF2*, respectively.[22]

2.4 Repeat motif–phenotype correlation

After identification of expanded TTTCA and TTTTA repeats in Japanese BAFME families, we predicted the presence of TTTCA repeats in other forms of BAFME. Indeed, expanded TTTCA and TTTTA repeats have been identified in *STARD7*,[24] *MARCHF6*,[25] *YEATS2*[26], and *RAI1*[27] in BAFME types 2, 3, 4, and 8 respectively. TTTCA repeats were exclusively found in patients with BAFME, whereas a limited number of control subjects showed expanded TTTTA repeats. The findings strongly indicate the repeat motif–phenotype correlation in the disease.[28] The findings also suggest that expanded TTTCA repeats, rather than dysfunctions of causative genes, commonly play an important role in the disease (Fig. 2a).

3. CGG or CCG repeat expansion disorders

3.1 Fragile X syndrome and fragile X-associated tremor/ataxia syndrome

When I began my research, the well-known CGG repeat disorders were fragile

X syndrome (FXS) and fragile X-associated tremor/ataxia syndrome (FXTAS). FXS is an X-linked disease characterized by intellectual disability. In most cases, the cause of the disease is CGG repeat expansions (>200 repeat units) located in the 5' untranslated region of *FMR1*. This was one of the first pathogenic repeat expansions discovered in the history of this field.[29] Very interestingly, relatively short CGG repeats in *FMR1* (55–200 repeat units called premutation) have been shown to be associated with a late-onset disease with leukoencephalopathy, ataxia, and tremor called FXTAS.[30] Whereas a full mutation (>200 repeat units) causes hypermethylation and transcriptional silencing of FMR1 leading to FXS, the premutation found in FXTAS is shown to escape transcriptional silencing and rather causes gain of toxic function.

3.2 History of genetic analysis of CGG repeat expansion disorders

Neuronal intranuclear inclusion disease (NIID) is a heterogeneous disorder characterized pathologically by the presence of intranuclear inclusions.[31] Originally, patients from Europe, mainly with juvenile onset, have been described.[32–34] On the other hand, adult-onset cases were reported from Japan. In particular, after it was shown that skin biopsy is useful for the antemortem diagnosis of NIID,[35] many cases have been identified. Most of the patients mainly show cognitive dysfunction. Brain MRI revealed leukoencephalopathy and characteristic diffusion hyperintensities in the corticomedullary junctions. Some patients mainly show peripheral neuropathy. Other features include autonomic dysfunction, ataxia, parkinsonism, retinopathy, and encephalitic episodes.[36]

Oculopharyngodistal myopathy (OPDM) was first reported in Japan,[37] which is characterized by ocular, facial, pharyngeal, and distal muscle weakness. OPDM has been shown to be distinct from oculopharyngeal muscular dystrophy (OPMD) caused by GCN repeats encoding polyalanine stretches in *PABPN1*.[38,39]

3.3 Identification of CGG repeat expansions in NIID, oculopharyngeal myopathy with leukoencephalopathy (OPML), and OPDM

The initial idea of directly searching repeat sequences was based on our observation that the clinical and pathological features of NIID are similar to those of FXTAS. Another clue came from a family we encountered, whose members were affected by both oculopharyngeal myopathy and leukoencephalopathy with characteristic diffusion hyperintensities in the corticomedullary junctions. We named the disease "oculopharyngeal myopathy with leukoencephalopathy (OPML)". From these clinical observations, in addition to our experiences in BAFME research, we hypothesized that

the expanded CGG repeats are the causes of NIID, OPML, and OPDM.

Whole-genome sequencing of affected families in addition to repeat-primed PCR and long-read sequence analysis actually revealed expanded CGG repeats in *NOTCH2NLC* (formerly known as *NBPF19*), *LOC642361*/*NUTM2B-AS1*, and *LRP12* in NIID, OPML, and OPDM, respectively.[40] Due to the presence of homologous sequences with high local similarity (>99% identity) to *NOTCH2NLC*, namely, *NOTCH2*, *NOTCH2NLR*, *NOTCH2NLA*, and *NOTCH2NLB*, precise localization of the expanded CGG repeat was challenging and required confirmation through long-read sequencing (Fig. 1c).

Very interestingly, CGG repeat expansions in *NOTCH2NLC* were found not only in patients with familial NIID, but also in those with sporadic NIID. Some patients with OPDM were negative for *LRP12*, indicating further genetic heterogeneity. In summary, our research suggests that NIID, which causes leukoencephalopathy and neuropathy, and OPDM, a peculiar muscle disorder, represent a disease spectrum with a common basis of CGG repeat expansions.

In this paper, we refer to the repeat sequence in *NOTCH2NLC* as a CGG repeat in order to focus on the disease spectrum shared with other CGG repeat disorders. However, it should be noted that this repeat is also sometimes described as a GGC repeat.

3.4 Repeat motif–phenotype correlation

After our publication, patients with expanded CGG repeats in *NOTCH2NLC* were also reported to have the oculopharyngodistal myopathy phenotype (termed OPDM type 3), although more than half of these cases also present cognitive decline, leukoencephalopathy, and neuropathy, signs frequently found in NIID.[41,42] Moreover, expanded CGG repeats in *GIPC1* and expanded CGG repeats *RILPL1* have been identified as causes of OPDM types 2[43,44] and 4[45], respectively. More recently, expanded CCG repeats in *ABCD3* have been identified in European patients with OPDM.[46] Regarding the clinical phenotype, the repeat expansion in *LOC642361*/*NUTM2B-AS1*, initially identified in OPML families, has also been found in OPDM cases.[47–50]

Taken together, expanded CGG or CCG repeats seem to expand a new disease spectrum from leukoencephalopathies to oculopharyngeal-type myopathies. We propose the disease spectrum as FNOP-spectrum disorder after the names of **F**XTAS, **N**IID, **OP**ML, and **OP**DM.[51] The findings that these diseases share some clinical features strongly indicates that there is a common disease pathomechanism caused by expanded CGG or CCG repeats (Fig. 2b).[23]

There seems to be accumulating evidence showing that very long CGG repeats

in these genes cause methylation and transcriptional silencing. Very interestingly, for example, individuals with very long CGG repeats in *NOTCH2NLC* are asymptomatic.[52] The findings support the original idea that expanded CGG repeats cause the spectrum of disease probably through gain of toxic function rather than haploinsufficiency.

4. Conclusion

In this review, I summarized recent studies on repeat expansion diseases by ours and others. The identification of the repeat motif–phenotype correlation, namely, the same expanded repeat motifs in similar diseases, strongly indicates the gain-of-function mechanism in these diseases as the primary pathomechanism rather than dysfunctions of causative genes. Further studies on molecular mechanisms of these diseases and novel therapeutic trials are awaited in this field.

ACKNOWLEDGMENTS

The author thanks Professors Shoji Tsuji and Tatsushi Toda for their great support. The author also thanks the patients and their family members as well as collaborators for participating in the study. This work was supported by JSPS KAKENHI Grant Number JP20H03588 and JP23K27514, and by AMED Grant Number JP23ek0109673.

CONFLICT OF INTEREST

The author declares no conflict of interest except for pending patents on methods of genetic analysis of repeat expansions.

REFERENCES

1. Ng SB, Buckingham KJ, Lee C, et al. Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 2010; 42: 30–5.

2. Ng SB, Bigham AW, Buckingham KJ, et al. Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat Genet 2010; 42: 790–3.

3. Vissers LELM, de Light J, Gilissen C, et al. A de novo paradigm for mental retardation. Nat Genet 2010; 42: 1109–12.

4. Inazuki G, Naito H, Ohama E, et al. A clinical study and neuropathological findings of a familial disease with myoclonus and epilepsy–the nosological place of familial essential myoclonus and epilepsy (FEME). Psych Neurol Jpn 1990; 92: 1–21.

5. Ikeda A, Kakigi R, Funai N, et al. Cortical tremor: a variant of cortical reflex

myoclonus. Neurology 1990; 40: 1561–5.
6. Yasuda, T. Benign adult familial myoclonic epilepsy (BAFME). Kawasaki Med J 1991; 17: 1–13.
7. Oi K, Neshige S, Hitomi T, et al. Low-dose perampanel improves refractory cortical myoclonus by the dispersed and suppressed paroxysmal depolarization shifts in the sensorimotor cortex. Clin Neurophysiol 2019; 130: 1804–12.
8. Kobayashi K, Hitomi T, Matsumoto R, et al. Nationwide survey in Japan endorsed diagnostic criteria of benign adult familial myoclonus epilepsy. Seizure 2018; 61:14–22.
9. Plaster NM, Uyama E, Uchino M, et al. Genetic localization of the familial adult myoclonic epilepsy (FAME) gene to chromosome 8q24. Neurology 1999; 53: 1180–3.
10. Mikami M, Yasuda T, Terao A, et al. Localization of a gene for benign adult familial myoclonic epilepsy to chromosome 8q23.3-q24.1. Am J Hum Genet 1999; 65: 745–51.
11. Suzuki T. Clinical genetics and linkage analysis of familial essential myoclonus and epilepsy (FEME). Niigata Igakkai Zasshi (in Japanese) 2002; 116: 535–45.
12. Mori S, Nakamura M, Yasuda T, et al. Remapping and mutation analysis of benign adult familial myoclonic epilepsy in a Japanese pedigree. J Hum Genet 2011; 56: 742–7.
13. Cen ZD, Xie F, Lou DN, et al. Fine mapping and whole-exome sequencing of a familial cortical myoclonic tremor with epilepsy family. Am J Med Genet Part B 2015; 168: 595–9.
14. Guerrini R, Bonanni P, Patrignani A, et al. Autosomal dominant cortical myoclonus and epilepsy (ADCME) with complex partial and generalized seizures: A newly recognized epilepsy syndrome with linkage to chromosome 2p11.1-q12.2. Brain 2001; 124: 2459–75.
15. Depienne C, Magnin E, Bouteiller D, et al. Familial cortical myoclonic tremor with epilepsy: the third locus (FCMTE3) maps to 5p. Neurology 2010; 74: 2000–3.
16. Yeetong P, Ausavarat S, Bhidayasiri R, et al. A newly identified locus for benign adult familial myoclonic epilepsy on chromosome 3q26.32-3q28. Eur J Hum Genet 2013; 21: 225–8.
17. Henden L, Freytag S, Afwi Z, et a. Identity by descent fine mapping of familial adult myoclonus epilepsy (FAME) to 2p11.2-2q11.2. Hum Genet 2016; 135: 1117–25.
18. Ishiura H, Sako W, Yoshida M, et al. The TRK-fused gene is mutated in hereditary motor and sensory neuropathy with proximal dominant involvement. Am J Hum Genet 2012; 91: 320–9.
19. Ishiura H, Takahashi Y, Mitsui J, et al. C9ORF72 repeat expansion in amyotrophic lateral sclerosis in the Kii peninsula of Japan. Arch Neurol 2012; 69: 1154–8.
20. Taira M, Ishiura H, Mitsui J, et al. Clinical features and haplotype analysis of newly

identified Japanese patients with gelsolin-related familial amyloidosis of Finnish type. Neurogenetics 2012; 13: 237–43.
21. Mano KK, Matsukawa T, Mitsui J, et al. Atypical parkinsonism caused by Pro105Leu mutation of prion protein: A broad clinical spectrum. Neurol Genet 2016; 2: e48.
22. 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.
23. Doi K, Monjo T, Hoang PH, et al. Rapid detection of expanded short tandem repeats in personal genomics using hybrid sequencing. Bioinformatics 2014; 30: 815–22.
24. 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 2019; 10: 4920.
25. Florian RT, Kraft F, Leitão E, et al. Unstable TTTTA/TTTCA expansions in MARCH6 are associated with familial adult myoclonic epilepsy type 3. Nat Commun 2019; 10: 4919.
26. Yeetong P, Pongpanich M, Srichomthong C, et al. TTTCA repeat insertions in an intron of YEATS2 in benign adult familial myoclonic epilepsy type 4. Brain 2019; 142: 3360–6.
27. Yeetong P, Dembélé ME, Pongpanich M, et al. Pentanucleotide Repeat Insertions in RAI1 Cause Benign Adult Familial Myoclonic Epilepsy Type 8. Mov Disord 2024; 39: 164–172.
28. Ishiura H, Tsuji S. Advances in repeat expansion diseases and a new concept of repeat motif-phenotype correlation. Curr Opin Genet Dev 2020; 65: 176–85.
29. Depienne C, Mandel J-L. 30 years of repeat expansion disorders: What have we learned and what are the remaining challenges? Am J Hum Genet 2021; 108: 764–85.
30. Hagerman RJ, Leehey M, Heinrichs W, et al. Intention tremor, parkinsonism, and generalized brain atrophy in male carriers of fragile X. Neurology 2001; 57: 127–30.
31. Takahashi-Fujigasaki J. Neuronal intranuclear hyaline inclusion disease. Neuropathology 2003; 23: 351–9.
32. Lindenberg R, Rubinstein LJ, Herman MM, et al. A light and electron microscopy study of an unusual widespread nuclear inclusion body disease. A possible residuum of an old herpesvirus infection. Acta Neuropathol (Berl) 1968; 10: 54-73.
33. Haltia M, Somer H, Palo J, Johnson WG. Neuronal intranuclear inclusion disease in identical twins. Ann. Neurol. 1984; 15: 316–321.
34. Kimber TE, Blumbergs PC, Rice JP, et al. Familial neuronal intranuclear inclusion disease with ubiquitin positive inclusions. J Neurol Sci 1998; 160: 33–40.
35. Sone J, Tanaka F, Koike H, et al. Skin biopsy is useful for the antemortem diagnosis of neuronal intranuclear inclusion disease. Neurology 2011; 76: 1372–6.

36. Sone J, Mori K, Inagaki T, et al. Clinicopathological features of adult-onset neuronal intranuclear inclusion disease. Brain 2016; 139: 3170–86.
37. Satoyoshi E, Kinoshita M. Oculopharyngodistal myopathy. Arch Neurol. 1977; 34: 89–92.
38. Uyama E, Uchino M, Chateau D, Tomé FM. Autosomal recessive oculopharyngodistal myopathy in light of distal myopathy with rimmed vacuoles and oculopharyngeal muscular dystrophy. Neuromuscul Disord. 1998; 8: 119–125.
39. Minami N, Ikezoe K, Kuroda H, et al. Oculopharyngodistal myopathy is genetically heterogeneous and most cases are distinct from oculopharyngeal muscular dystrophy. Neuromuscul Disord 2001; 11: 699–702.
40. Ishiura H, Shibata S, Yoshimura J, et al. Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease. Nat Genet 2019; 51: 1222–32.
41. Ogasawara M, Iida A, Kumutpongpanich T, et al. CGG expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy with neurological manifestations. Acta Neuropathol Commun. 2020; 8: 204.
42. Yu J, Deng J, Guo X, et al. The GGC repeat expansion in NOTCH2NLC is associated with oculopharyngodistal myopathy type 3. Brain 2021; 144: 1819–32.
43. Deng J, Yu J, Li P, et al. Expansion of GGC repeat in GIPC1 is associated with oculopharyngodistal myopathy. Am J Hum Genet 2020; 106: 793–804.
44. Xi J, Wang X, Yue D, et al. 5' UTR CGG repeat expansion in GIPC1 is associated with oculopharyngodistal myopathy. Brain 2021; 144: 601–14.
45. Yu J, Shan J, Yu M, et al. The CGG repeat expansion in RILPL1 is associated with oculopharyngodistal myopathy type 4. Am J Hum Genet 2022; 109: 533–541.
46. Cortese A, Beecroft SJ, Facchini S, et al. A CCG expansion in ABCD3 causes oculopharyngodistal myopathy in individuals of European ancestry. Nat Commun 2024; 15: 6327.
47. Gu X, Yu J, Jiao K, et al. Non-coding CGG repeat expansion in LOC642361/NUTM2B-AS1 is associated with a phenotype of oculopharyngodistal myopathy. J Med Genet 2024; 61: 340–346.
48. Shi Y, Cao C, Zeng Y, et al. CGG repeat expansion in LOC642361/NUTM2B-AS1 typically presents as oculopharyngodistal myopathy. J Genet Genomics 2024; 51: 184–186.
49. Pongpakdee S, Apiwattanakul M, Termglinchan T, et al. CGG/CCG Repeat Expansions in LOC642361/NUTM2B-AS1 in Thai Patients With Oculopharyngodistal Myopathy. Neurol Genet 2024; 10: e200170.

50. Ishiura H. Recent progress in oculopharyngodistal myopathy research from clinical and genetic viewpoints. J Neuromuscul Dis 2025 in press.
51. Ishiura H, Tsuji S, Toda T. Recent advances in CGG repeat diseases and a proposal of fragile X-associated tremor/ataxia syndrome, neuronal intranuclear inclusion disease, and oculophryngodistal myopathy (FNOP) spectrum disorder. J Hum Genet 2023; 68: 169–174.
52. Fukuda H, Yamaguchi D, Nyquist K, et al. Father-to-offspring transmission of extremely long NOTCH2NLC repeat expansions with contractions: genetic and epigenetic profiling with long-read sequencing. Clin Epigenetics 2021; 13: 204.

Figure legend

Figure 1 Method for identifying repeat expansions from short-read sequencing data
(a) Paired-end short reads consist of opposing read 1 and read 2, typically ranging from 100 to 150 base pairs in length. The total length of the DNA fragment is often between 300 and 500 base pairs. This panel shows the type of pair-end reads obtained when a short tandem repeat (STR) is present and has not undergone expansion. The blue represents surrounding (unique) sequences, while the red represents repeat sequences.
(b) When repeat expansion sequences are present, they can be classified into the following three scenarios: 1. Both paired-end reads are unique. 2. One read contains repeat sequences, and the remainder of the read is unique. 3. Both paired-end reads consist entirely of repeat sequences. When both repeat and unique sequences are present, the location of the repeat expansion sequence can be determined based on the unique sequences.
(c) *NOTCH2NLC* has homologous genes with highly similar sequences, namely, *NOTCH2*, *NOTCH2NLR*, *NOTCH2NLA*, and *NOTCH2NLB*, all of which are located on chromosome 1. Some unique sequences from paired-end reads may map equivocally to these homologous genes, making it difficult to determine the location of the repeat expansion sequence in such cases (left). On the other hand, several unique sequences from paired-end reads, although mismatched with the sequences of other homologous genes, perfectly align with the *NOTCH2NLC* sequence, strongly suggesting that the CGG repeat expansion is present in *NOTCH2NLC* (right). This was confirmed by long-read sequencing analysis. Similar sequences also exist on the genome for *LOC642361*/*NUTM2B-AS1*.

Figure 2 Pathomechanisms of repeat expansion diseases
(a) Proposed pathomechanism of benign adult familial myoclonus epilepsy (BAFME). Expansions of intronic TTTCA and TTTTA repeats have been found in six genes in BAFME. The findings strongly indicate that the expanded repeats, especially TTTCA repeats, rather than dysfunctions of each gene, play an important role in the pathomechanism of BAFME, probably through abnormal RNA.
(b) Proposed pathomechanism of CGG or CCG repeat expansion diseases including fragile X-associated tremor/ataxia syndrome, neuronal intranuclear inclusion disease, oculopharyngeal myopathy with leukoencephalopathy, and oculopharyngodistal myopathy. The findings that expanded CGG or CCG repeats in seven genes/loci cause diseases with shared clinical features strongly suggests a common mechanism caused by

CGG or CCG repeats, probably through an abnormal RNA with expanded repeats and/or an abnormal protein with an expanded amino acid translated by the abnormal RNA.