nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-018-0067-2

In the format provided by the authors and unedited.

Expansions of intronic TTTCA and TTTTA repeats in benign adult familial myoclonic epilepsy

**Hiroyuki Ishiura[1], Koichiro Doi[2], Jun Mitsui[1], Jun Yoshimura[2], Miho Kawabe Matsukawa[1], Asao Fujiyama[3], Yasuko Toyoshima[4], Akiyoshi Kakita[4], Hitoshi Takahashi[4], Yutaka Suzuki[5], Sumio Sugano[6], Wei Qu[2], Kazuki Ichikawa[2], Hideaki Yurino[7], Koichiro Higasa[8], Shota Shibata[1], Aki Mitsue[1], Masaki Tanaka[1], Yaeko Ichikawa[9], Yuji Takahashi[10], Hidetoshi Date[1], Takashi Matsukawa[1], Junko Kanda[1], Fumiko Kusunoki Nakamoto[1], Mana Higashihara[11], Koji Abe[12], Ryoko Koike[13], Mutsuo Sasagawa[14], Yasuko Kuroha[13], Naoya Hasegawa[15], Norio Kanesawa[16], Takayuki Kondo[17], Takefumi Hitomi[17,18], Masayoshi Tada[19], Hiroki Takano[20], Yutaka Saito[21], Kazuhiro Sanpei[22], Osamu Onodera[19], Masatoyo Nishizawa[23], Masayuki Nakamura[24], Takeshi Yasuda[25], Yoshio Sakiyama[26], Mieko Otsuka[27], Akira Ueki[43], Ken-ichi Kaida[28], Jun Shimizu[1], Ritsuko Hanajima[29], Toshihiro Hayashi[1], Yasuo Terao[30], Satomi Inomata-Terada[1], Masashi Hamada[1], Yuichiro Shirota[1], Akatsuki Kubota[1], Yoshikazu Ugawa[31], Kishin Koh[32], Yoshihisa Takiyama[32], Natsumi Ohsawa-Yoshida[33], Shoichi Ishiura[33,34], Ryo Yamasaki[35], Akira Tamaoka[36], Hiroshi Akiyama[37], Taisuke Otsuki[38], Akira Sano[24], Akio Ikeda[39], Jun Goto[40], Shinichi Morishita[2] and Shoji Tsuji[1,41,42]*

[1]Department of Neurology, The University of Tokyo Hospital, Tokyo, Japan. [2]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan. [3]Advanced Genomics Center, National Institute of Genetics, Shizuoka, Japan. [4]Department of Pathology, Brain Research Institute, Niigata University, Niigata, Japan. [5]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan. [6]Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan. [7]Department of Laboratory Medicine, Graduate School of Medical Sciences, Kanazawa University, Ishikawa, Japan. [8]Human Disease Genomics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [9]Department of Neurology, Kyorin University, Tokyo, Japan. [10]Department of Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan. [11]Department of Neurology, Tokyo Metropolitan Geriatric Hospital, Tokyo, Japan. [12]Department of Neurology, Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama University, Okayama, Japan. [13]Department of Neurology, Nishi-Niigata Chuo National Hospital, Niigata, Japan. [14]Department of Psychiatry, Toyosato Hospital, Ibaraki, Japan. [15]Department of Epilepsy, Nishi-Niigata Chuo National Hospital, Niigata, Japan. [16]Division of Cardiology, National Hospital Organization Takasaki General Medical Center, Gunma, Japan. [17]Department of Neurology, Kyoto University Graduate School of Medicine, Kyoto, Japan. [18]Department of Clinical Laboratory Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan. [19]Department of Neurology, Brain Research Institute, Niigata University, Niigata, Japan. [20]Department of Neurology, Tachikawa General Hospital, Niigata, Japan. [21]Department of Neurology, Sannocho Hospital, Niigata, Japan. [22]Department of Neurology, Sado General Hospital, Niigata, Japan. [23]Department of Neurology, Niigata Medical Center, Niigata, Japan. [24]Department of Psychiatry, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan. [25]Department of Neurology, Kurashikikinen Hospital, Okayama, Japan. [26]Department of Neurology, Jichi Medical University, Saitama Medical Center, Saitama, Japan. [27]Department of Neurology, International University of Health and Welfare, Tochigi, Japan. [28]Third Department of Internal Medicine, National Defense Medical College, Saitama, Japan. [29]Department of Neurology, School of Medicine, Kitasato University, Kanagawa, Japan. [30]Department of Cell Physiology, School of Medicine, Kyorin University, Tokyo, Japan. [31]Department of Neurology, Fukushima Medical University, Fukushima, Japan. [32]Department of Neurology, University of Yamanashi, Yamanashi, Japan. [33]Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan. [34]Faculty of Life and Medical Sciences, Doshisha University, Kyoto, Japan. [35]Department of Neurology, Neurological Institute, Faculty of Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan. [36]Department of Neurology, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan. [37]Hiratsuka Hospital, Kanagawa, Japan. [38]Epilepsy Hospital Bethel Japan, Miyagi, Japan. [39]Department of Epilepsy, Movement Disorders and Physiology, Kyoto University Graduate School of Medicine, Kyoto, Japan. [40]Department of Neurology, International University of Health and Welfare Mita Hospital, Tokyo, Japan. [41]Medical Genome Center, The University of Tokyo Hospital, Tokyo, Japan. [42]International University of Health and Welfare, Chiba, Japan. [43]Deceased: Akira Ueki. *e-mail: tsuji@m.u-tokyo.ac.jp

Supplementary Note

Intronic TTTCA and TTTTA repeat expansions in benign adult familial myoclonic epilepsy

Hiroyuki Ishiura[1], Koichiro Doi[2], Jun Mitsui[1], Jun Yoshimura[2], Miho Kawabe Matsukawa[1], Asao Fujiyama[3], Yasuko Toyoshima[4], Akiyoshi Kakita[4], Hitoshi Takahashi[4], Yutaka Suzuki[5], Sumio Sugano[6], Wei Qu[2], Kazuki Ichikawa[2], Hideaki Yurino[7], Koichiro Higasa[8], Shota Shibata[1], Aki Mitsue[1], Masaki Tanaka[1], Yaeko Ichikawa[9], Yuji Takahashi[10], Hidetoshi Date[1], Takashi Matsukawa[1], Junko Kanda[1], Fumiko Kusunoki Nakamoto[1], Mana Higashihara[11], Koji Abe[12], Ryoko Koike[13], Mutsuo Sasagawa[14], Yasuko Kuroha[13], Naoya Hasegawa[15], Norio Kanesawa[16], Takayuki Kondo[17], Takefumi Hitomi[17,18], Masayoshi Tada[19], Hiroki Takano[20], Yutaka Saito[21], Kazuhiro Sanpei[22], Osamu Onodera[19], Masatoyo Nishizawa[23], Masayuki Nakamura[24], Takeshi Yasuda[25], Yoshio Sakiyama[26], Mieko Otsuka[27], Akira Ueki[43], Ken-ichi Kaida[28], Jun Shimizu[1], Ritsuko Hanajima[29], Toshihiro Hayashi[1], Yasuo Terao[30], Satomi Inomata-Terada[1], Masashi Hamada[1], Yuichiro Shirota[1], Akatsuki Kubota[1], Yoshikazu Ugawa[31], Kishin Koh[32], Yoshihisa Takiyama[32], Natsumi Ohsawa-Yoshida[33], Shoichi Ishiura[33,34], Ryo Yamasaki[35], Akira Tamaoka[36], Hiroshi Akiyama[37], Taisuke Otsuki[38], Akira Sano[24], Akio Ikeda[39], Jun Goto[40], Shinichi Morishita[2], and Shoji Tsuji[1,41,42*]

[1] Department of Neurology, The University of Tokyo Hospital, Tokyo, Japan
[2] Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan
[3] Advanced Genomics Center, National Institute of Genetics, Shizuoka, Japan
[4] Department of Pathology, Brain Research Institute, Niigata University, Niigata, Japan
[5] Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan
[6] Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Tokyo, Japan
[7] Department of Laboratory Medicine, Graduate School of Medical Sciences, Kanazawa University, Ishikawa, Japan
[8] Human Disease Genomics, Center for Genomic Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
[9] Department of Neurology, Kyorin University, Tokyo, Japan
[10] Department of Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan
[11] Department of Neurology, Tokyo Metropolitan Geriatric Hospital, Tokyo, Japan
[12] Department of Neurology, Graduate School of Medicine, Dentistry, and

Pharmaceutical Sciences, Okayama University, Okayama, Japan
[13] Department of Neurology, Nishi-Niigata Chuo National Hospital, Niigata, Japan
[14] Department of Psychiatry, Toyosato Hospital, Ibaraki, Japan
[15] Department of Epilepsy, Nishi-Niigata Chuo National Hospital, Niigata, Japan
[16] Division of Cardiology, National Hospital Organization Takasaki General Medical Center, Gunma, Japan
[17] Department of Neurology, Kyoto University Graduate School of Medicine, Kyoto, Japan
[18] Department of Clinical Laboratory Medicine, Kyoto University Graduate School of Medicine, Kyoto, Japan
[19] Department of Neurology, Brain Research Institute, Niigata University, Niigata, Japan
[20] Department of Neurology, Tachikawa General Hospital, Niigata, Japan
[21] Department of Neurology, Sannocho Hospital, Niigata, Japan
[22] Department of Neurology, Sado General Hospital, Niigata, Japan
[23] Department of Neurology, Niigata Medical Center, Niigata, Japan
[24] Department of Psychiatry, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan
[25] Department of Neurology, Kurashikikinen Hospital, Okayama, Japan
[26] Department of Neurology, Jichi Medical University, Saitama Medical Center, Saitama, Japan
[27] Department of Neurology, International University of Health and Welfare, Tochigi, Japan
[28] Third Department of Internal Medicine, National Defense Medical College, Saitama, Japan
[29] Department of Neurology, School of Medicine, Kitasato University, Kanagawa, Japan
[30] Department of Cell Physiology, School of Medicine, Kyorin University, Tokyo, Japan
[31] Department of Neurology, Fukushima Medical University, Fukushima, Japan
[32] Department of Neurology, University of Yamanashi, Yamanashi, Japan
[33] Department of Life Sciences, Graduate School of Arts and Sciences, The University of Tokyo, Tokyo, Japan
[34] Faculty of Life and Medical Sciences, Doshisha University, Kyoto, Japan
[35] Department of Neurology, Neurological Institute, Faculty of Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan
[36] Department of Neurology, Faculty of Medicine, University of Tsukuba, Ibaraki, Japan
[37] Hiratsuka Hospital, Kanagawa, Japan
[38] Epilepsy Hospital Bethel Japan, Miyagi, Japan
[39] Department of Epilepsy, Movement Disorders and Physiology, Kyoto University Graduate School of Medicine, Kyoto, Japan
[40] Department of Neurology, International University of Health and Welfare Mita Hospital, Tokyo, Japan
[41] Medical Genome Center, The University of Tokyo Hospital, Tokyo, Japan
[42] International University of Health and Welfare, Chiba, Japan
[43] Deceased

RESULTS

Genome-wide linkage analysis and haplotype analysis

To identify the causative gene for BAFME, we first performed a genome-wide linkage study of four families including 16 patients with the disease and two individuals without the disease using high-density SNP arrays (Genome-wide Human SNP array 6.0, Affymetrix), the pipeline software SNP-HiTLink (Fukuda, Y. *et al.* SNP-HiTLink: a high-throughput linkage analysis system employing dense SNP data. *BMC Bioinformatics* **10**, 121 (2009)), and allegro version 2 (Gudbjartsson, D.F., Thorvaldsson, T., Kong, A., Gunnarsson, G., & Ingolfsdottir, A. Allegro version 2. *Nat. Genet.* **37**, 1015-1016 (2005)).

In addition to the locus with a cumulative multipoint LOD score of 3.1 at 8q22.1-8q24.13, there was a small peak with a LOD score of 1.5 on chromosome 1. This locus did not remain consistent when inter-marker distances used for the linkage analysis were narrowed, indicating that there were markers inconsistent with the linkage within the region that were not used in the original linkage analysis.

To further narrow down the candidate region, we conducted a detailed haplotype analysis using microsatellite (**Supplementary Table 1**) and high-density SNP markers. In addition to the four families used for the linkage study, two additional families were included in the haplotype analysis (**Fig. 1a**).

Determination of precise structures of expanded repeats by MinION sequencers

To determine the precise structures of the expanded repeats, we further sequenced two genomic DNA samples extracted from the peripheral blood leukocytes of patient II-6 in family F6906 and patient II-1 in family F6115 using MinION sequencers (Oxford Nanopore Technologies) with R9.4 and R9.5 flowcells. The respective two datasets had 9,680,911 and 8,359,922 reads, their average lengths were 11,585 and 10,279 bases, and the total bases were 112.16 and 85.94 billion bases. We aligned all the reads with the hg19 human reference genome using NGM-LR (https://github.com/philres/ngmlr) and then collected reads that mapped to the focal short tandem repeat (STR) site at chr8:119,379,052-119,379,172 in hg19. Precisely, since the reference repeat in hg19 is short and is of length approximately 120 bp, we used reads that were aligned within 100 kb surrounding the mutation site. Of these, some reads did not have repeat expansions at the site, confirming the repeat expansion was heterozygous. Thus, for each of the two samples, we used only reads with repeat expansions and assembled them into a contig using the miniasm assembler (Li, H.

Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103-2110 (2016)).

Subsequently, we reconfirmed the presence of the repeat expansion by aligning the raw MinION reads to the assembled contig. Specifically, **Supplementary Fig. 5** shows a dot plot between the assembled contig with the STR expansion and its corresponding hg19 region around the mutation site, for each of the two samples.

Finally, we attempted to examine the precise structure of the repeat expansions in II-6 in F6906 and II-1 in F6115. We investigated raw reads with a plenty of TTTCA/TTTTA occurrences that mapped to the mutation site; however, we found substantial variation in the repeat lengths possibly reflecting error reads in nanopore sequencing and/or somatic instability of the expanded repeats, making it difficult to determine the numbers of TTTCA/TTTTA occurrences in those raw reads.

We remark one possible estimation of TTTCA/TTTTA occurrences. We found one pair of two Nanopore reads from the plus and minus strands of an identical DNA fragment using the $1D^2$ sequencing kit (Oxford Nanopore Technologies) for each of II-6 in F6906 and II-1 in F6115. We could partly correct errors in one strand by using sequence information in the opposite strand, yielding the sequences of the mutant allele for the two samples, which can be found at **Supplementary Fig. 6a,d**. We then manually partitioned the repeat expansion into two motifs for II-6 in F6906 and count the number of TTTCA and TTTTA motifs using Tandem Repeat Finder (Version4.09, Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999)), indicating $(TTTTA)_{598}(TTTCA)_{458}$ (**Supplementary Fig. 6b**). Similarly, we divided the repeat expansion in II-1 in F6115 into $(TTTTA)_{221}(TTTCA)_{225}(TTTTA)_{81}$ (**Supplementary Fig. 6e**).

Nevertheless, care has to be taken to interpret these numbers of occurrences because they may not be highly accurate. Specifically, the match ratios between the observed and estimated pure repeat expansions were around 90% (**Supplementary Fig.6c,f**).

Family history of homozygous patients with repeat expansion mutation in *SAMD12*

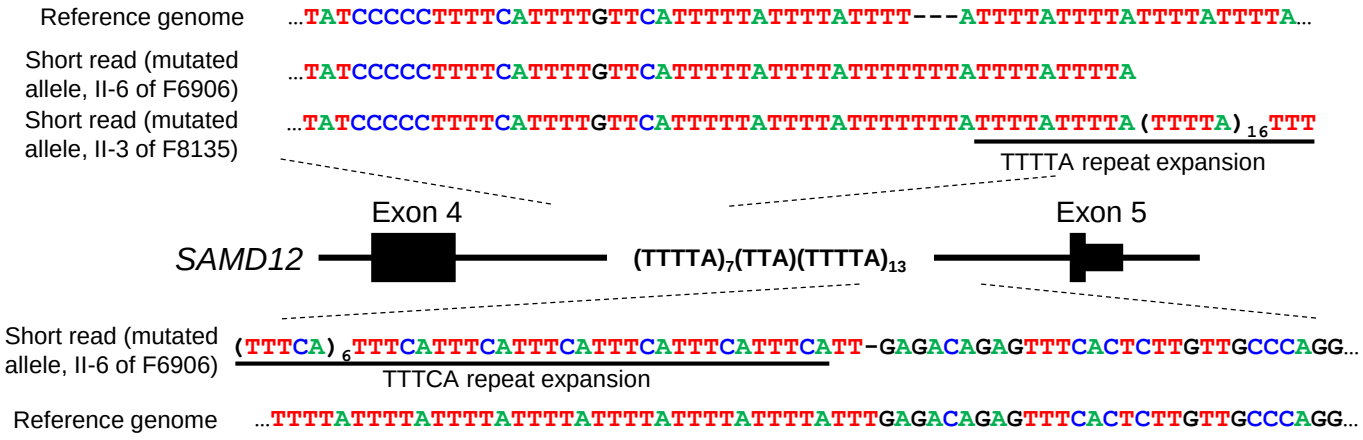
There was a parental consanguineous marriage in family F8140, but not in families F8115 or F8308. Although molecular diagnoses were not confirmed, the parents and the aunt of the homozygous patient in family F8115 had symptoms consistent with BAFME, supporting the assortative mating. In family F8398, information on the parents were unavailable.

Identification of TTTCA and TTTTA pentanucleotide repeat expansions in *TNRC6A* and *RAPGEF2* in other families with similar clinical presentations

In the F9283 family (BAFME6), we were able to identify short reads containing TTTCA repeats whose paired reads were aligned to the unique sequences near *TNRC6A* (which encodes trinucleotide repeat-containing 6A) on chromosome 16p12.1. The repeat expansion was identified in the upstream region of exon 1 of *TNRC6A* registered in the RefSeq database. RNA-seq data from control brains indicated that there are spliced transcripts flanking the repeat. The transcript is further confirmed to be expressed in the brain, placenta, kidney, and pancreas by RT-PCR analysis (**Supplementary Figs. 13** and **14**). These findings indicate that the expanded repeats are located in an intron (designated here as intron 1a). There is $(TTTTA)_{18}$ in the reference genome sequence in intron 1a (chr16:24,624,761-24,624,850). Parametric linkage analysis excluded the loci for the other known autosomal dominant BAFME loci on 2p11.2–q11.2, 3q26.32–q28, and 5p15.31–p15.1 (**Supplementary Fig. 15**). The ages at onset of myoclonic tremor ranged from early 20s to 70s in the family. In one patient, a giant somatosensory evoked potential (SEP), which is characteristic of cortical myoclonus, was observed. Of note, none of the patients in the family experienced seizures, although one patient showed borderline findings in electroencephalogram (sharp waves induced by photic stimulation).

In the F8241 family, TTTCA repeat expansions were similarly identified in intron 14 of *RAPGEF2* (which encodes Rap guanine nucleotide-exchange factor 2) on chromosome 4q32.1 using paired reads anchored to unique sequences. There is $(TTTTA)_5(TATTA)(TTTTA)_{12}$ in the reference genome sequence of intron 14 of *RAPGEF2* (chr4:160,263,679-160,263,768). Based on the manual realignment of short reads, we assumed that the expanded TTTCA repeats were located between the expanded TTTTA repeats and the short TTTTA repeats (**Supplementary Fig. 16** and **Fig. 6d**). The number of repeat units of the latter short TTTTA repeat was 3, but it was difficult to unequivocally determine the number of repeat units because it varied among the short reads presumably owing to the instability of the TTTTA repeats. Patient III-2 in family F8241 started to show myoclonic tremor from the age of 18. At the age of 27, he experienced a seizure, when anticonvulsants and β-blocker were started. Neither brain MRI abnormalities nor giant SEPs were observed.

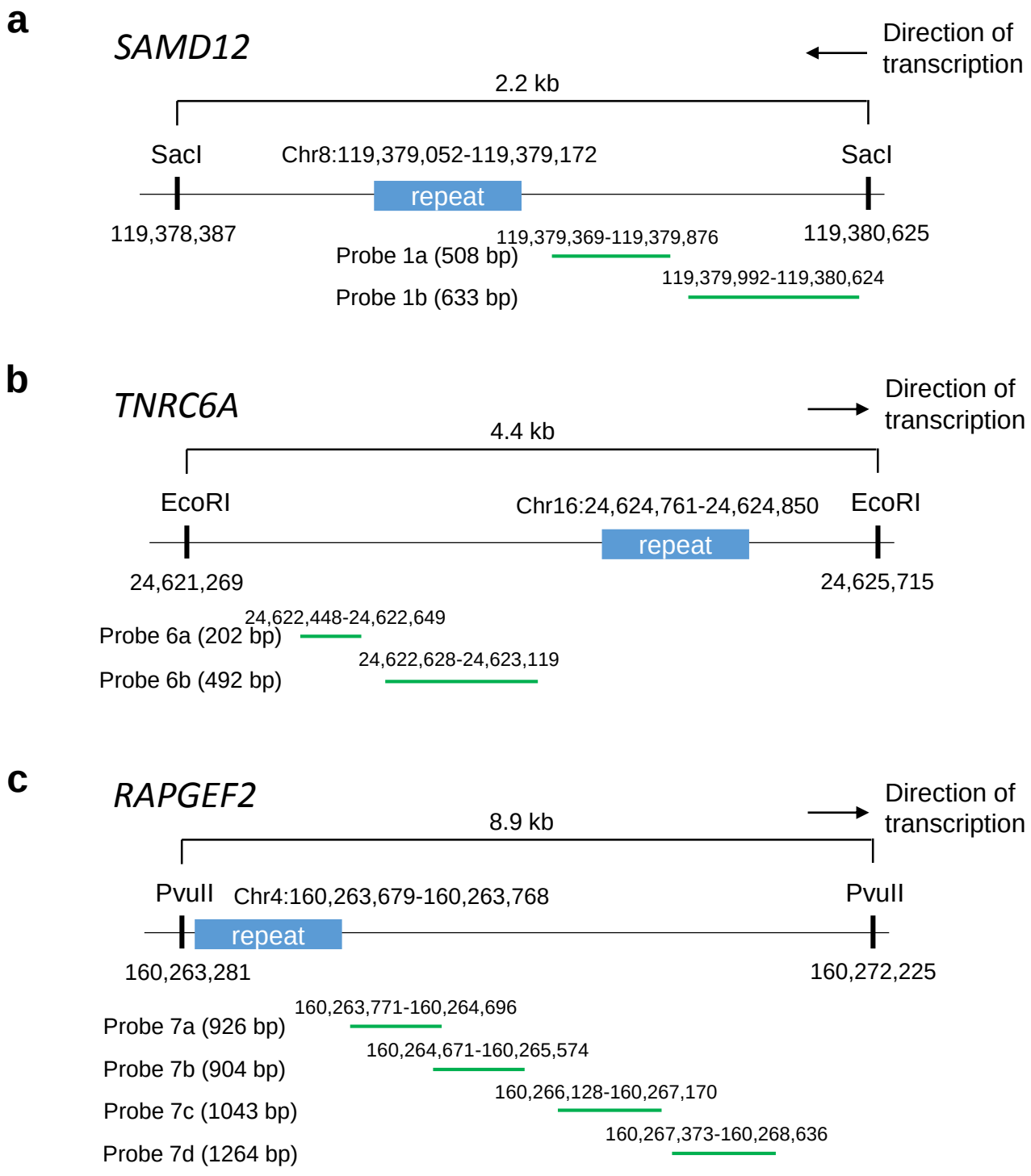
Supplementary Figure 1



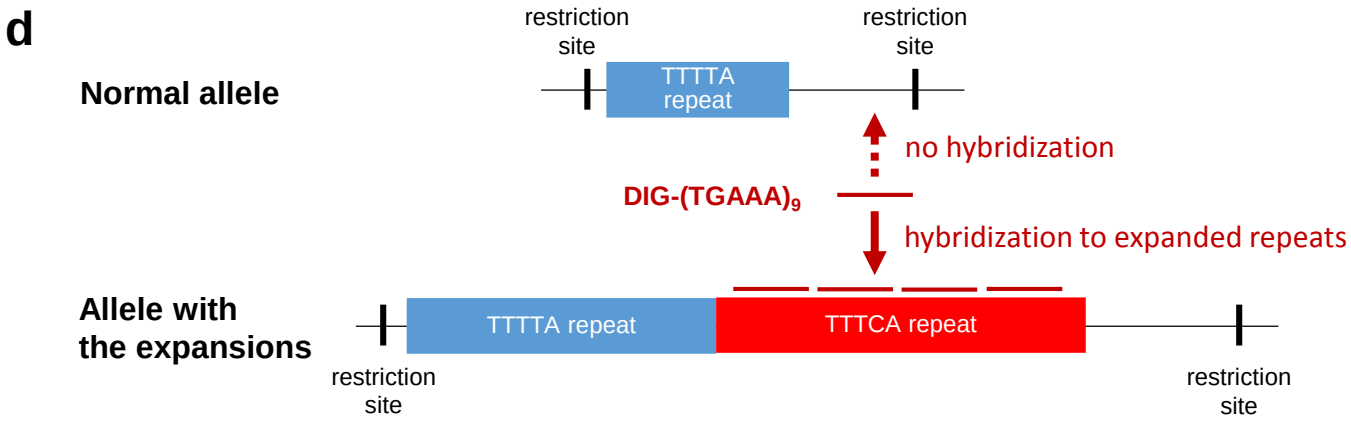
Short read aligned to upstream region of repeat in SAMD12 showing TTTTA repeat expansion

Top: Results of whole-genome sequence analysis of two patients suggest mutant alleles containing TTTTA repeat expansion in the upstream region.
Bottom: Results of whole-genome sequence analysis of a patient suggest TTTCA repeat expansion in the downstream region. The bottom diagram is the same as that in **Fig. 2b**.

Supplementary Figure 2



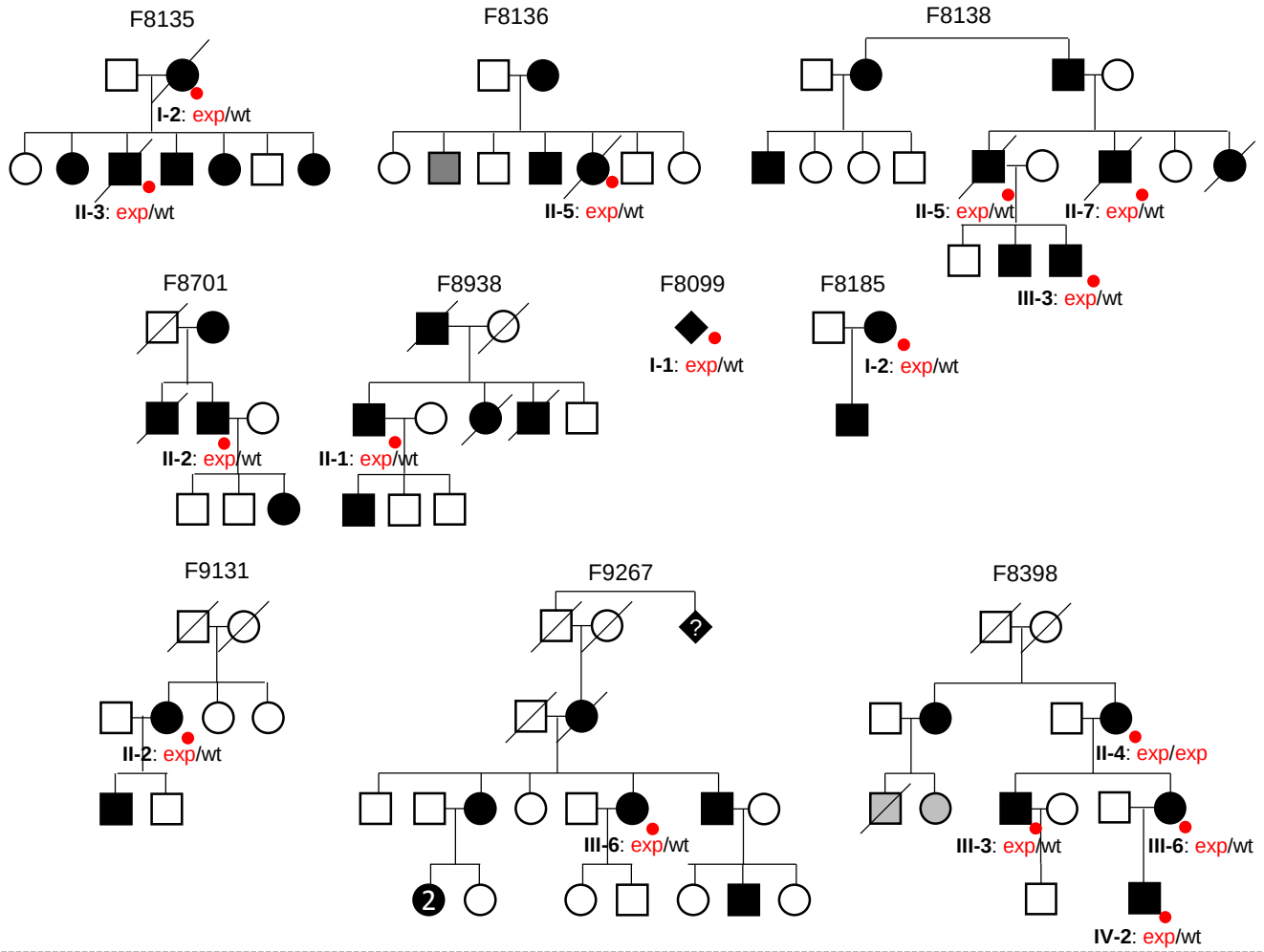
Supplementary Figure 2 (cont.)



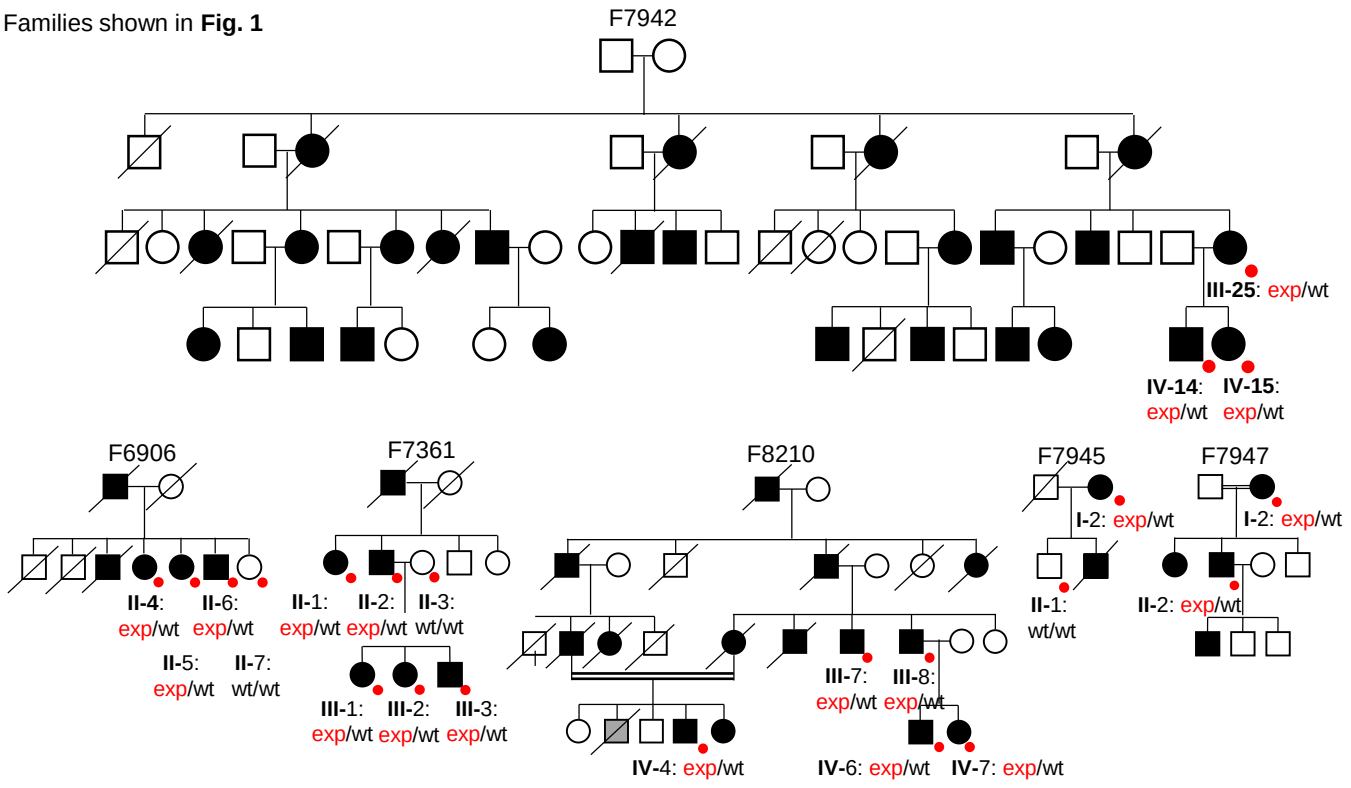
Schematic representations of the positions of the repeat, the probes used for Southern blot analysis, and the restriction sites

(a–c). Schematic representations of probes used in the study. Physical positions are based on hg19. Digoxigenin (DIG)-labeled probe synthesis are performed using PCR, with subcloned genomic fragments prepared for templates. Primers used for probe synthesis are listed in **Supplementary Table 5**. When indicated, multiple probes were mixed in hybridization buffer (Roche DIG Easy-Hyb) to increase hybridization signals.
(d) A schematic representation of unique and $(TGAAA)_9$ probes. DIG-labeled probes made from genomic fragments are hybridized to both normal and mutant alleles with the expansions, whereas DIG-$(TGAAA)_9$ probe is hybridized exclusively to mutant alleles containing TTTCA repeat expansions.

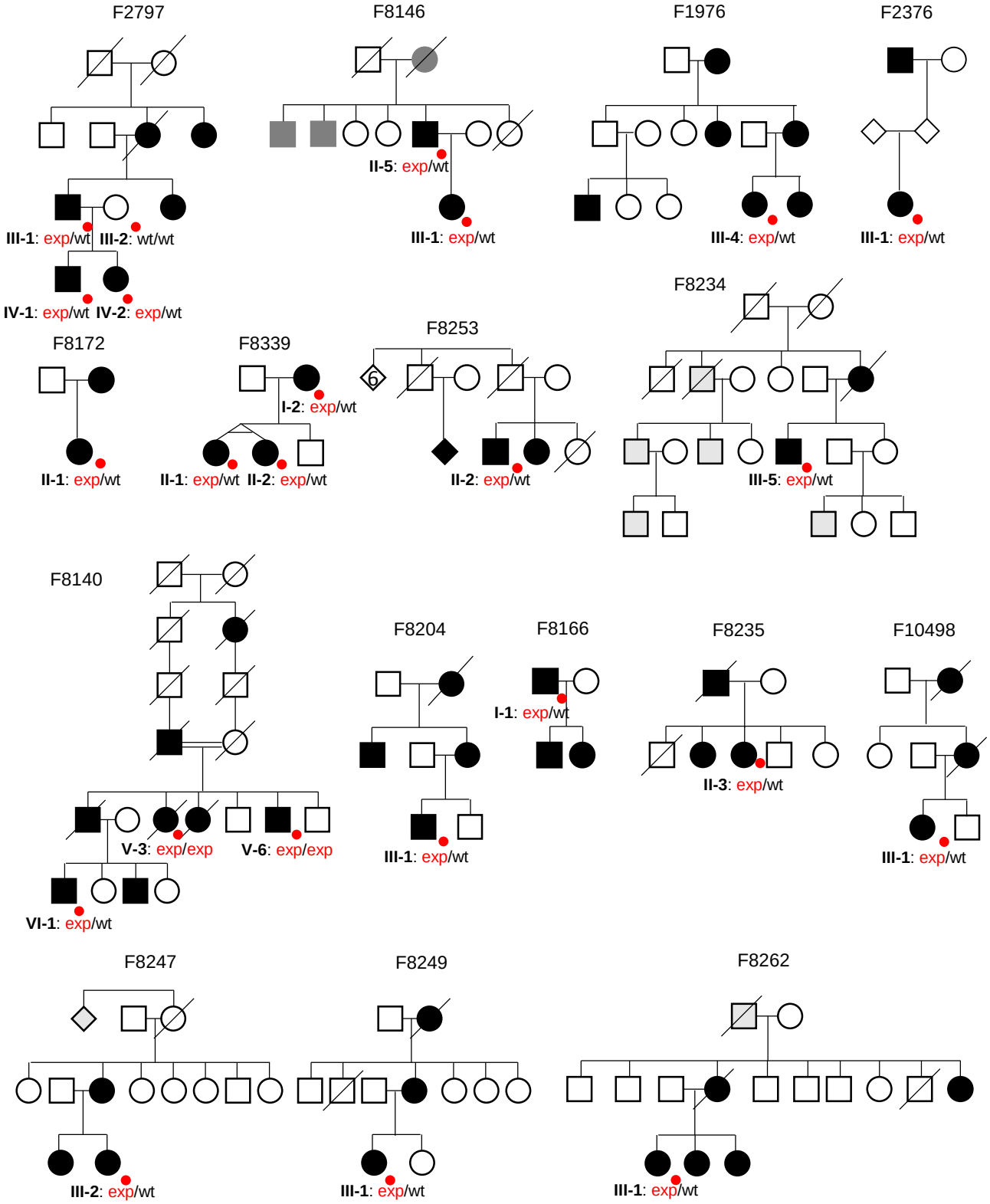
Supplementary Figure 3



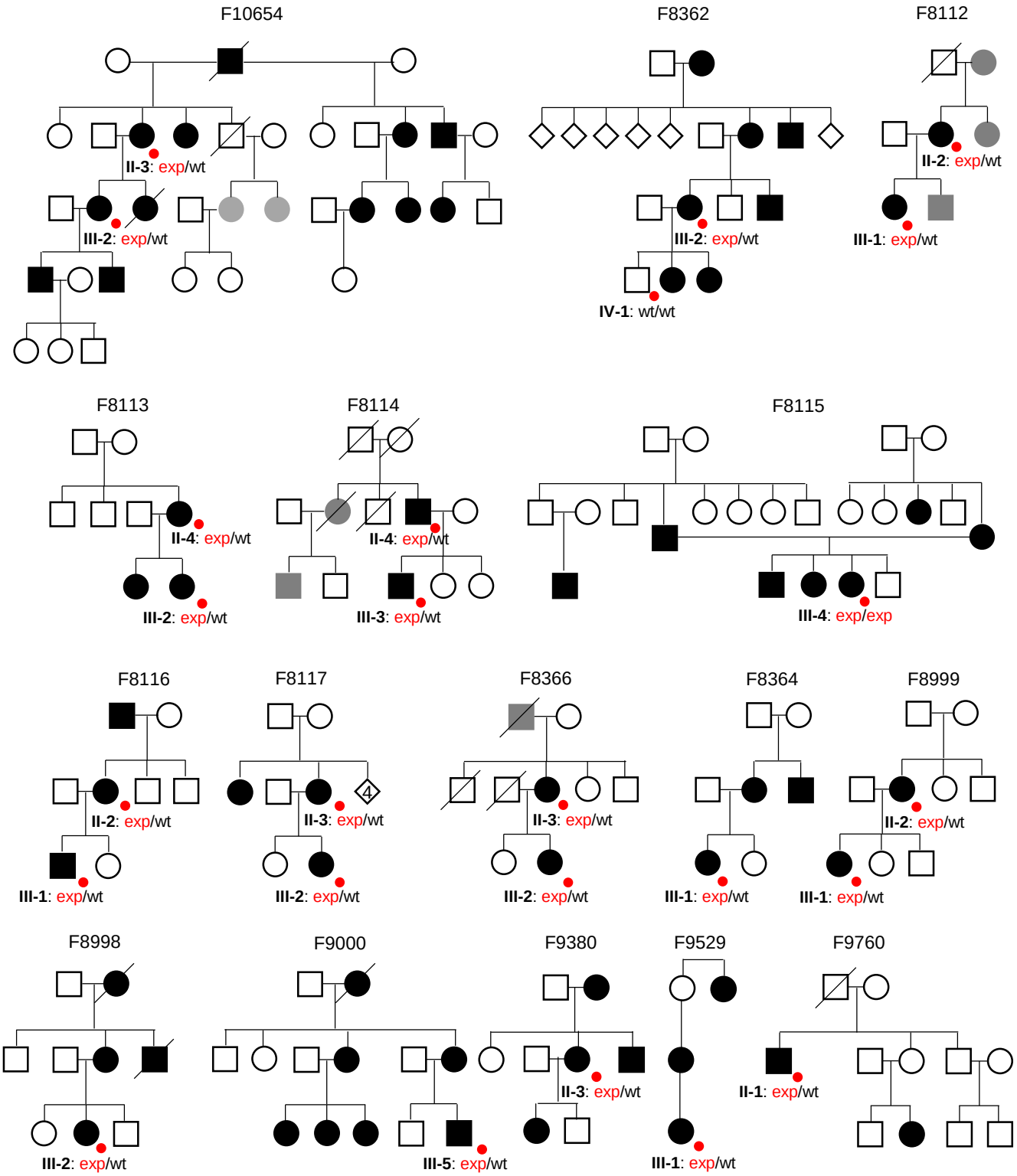
Families shown in **Fig. 1**



Supplementary Figure 3 (continued)



Supplementary Figure 3 (continued)



Families with repeat expansion mutations in *SAMD12* (repeat configuration 1)

Patients with BAFME are indicated by filled symbols. A diagonal line through a symbol indicates a deceased individual. Symbols with red dots represents individuals whose genomic DNA samples were available.

Supplementary Figure 4

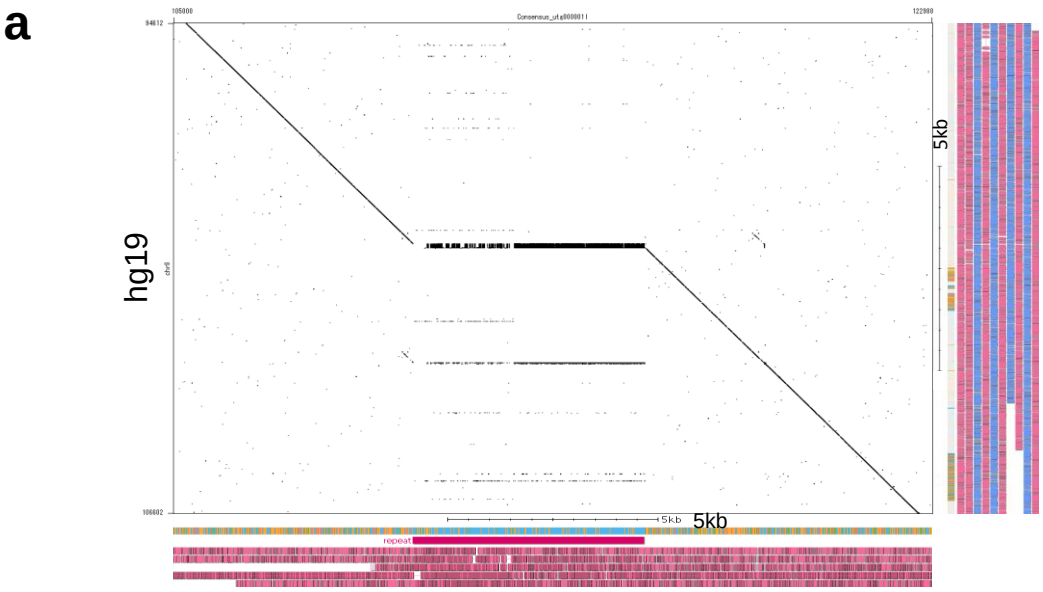
Marker	Physical position (hg19)	Haplotype of family F6906 (repeat configuration 1)	Haplotype of family F6115 (repeat configuration 2)	Minor allele frequency
rs4637872	119,293,911	G	A	A: 0.60, G: 0.40
rs3900767	119,307,293	G	A	G: 0.82, A: 0.18
rs7464659	119,332,157	A	A	A: 0.83, G: 0.17
rs59797347	119,333,061	T	T	not registered in 1kJPN
rs2450222	119,343,669	A	A	G: 0.21, A: 0.79
rs6994270	119,348,495	G	G	C: 0.88, G: 0.12
rs2514941	119,349,285	G	G	A: 0.32, G: 0.68
rs34200541	119,349,531-119,349,532	TG	TG	not registered in 1kJPN
rs2450315	119,350,656	C	C	T: 0.30, C: 0.70
rs2450203	119,352,278	A	A	C: 0.31, A: 0.69
rs111553876	119,354,490-119,354,491	delCA	delCA	not registered in 1kJPN
rs764446520	119,354,532	insAAA	insAAA	not registered in 1kJPN
-	119,355,641	insG	insG	not registered in dbSNP
rs2515029	119,356,466	C	C	T: 0.14, C: 0.86
rs2515028	119,357,221	C	C	G: 0.14, C: 0.86
rs17828843	119,368,718	G	G	G: 0.85, A: 0.15
rs9643124	119,376,989	G	G	G: 0.84, T: 0.16
3' end of the repeat	119,379,054	-	A	
Mutation	119,379,055-119,379,157	- $(TGAAA)_{exp}$ $(TAAAA)_{exp}$	$(TAAAA)_{exp}$ $(TGAAA)_{exp}$ $(TAAAA)_{exp}$	
rs10086119	119,389,854	G	G	A: 0.14, G: 0.86
rs12678322	119,394,878	A	A	A: 0.86, G: 0.14
rs5894426	119,399,723	insTAA	insTAA	not registered in 1kJPN
rs4876820	119,403,640	T	T	T: 0.71, C: 0.29
rs9297585	119,405,199	A	A	A: 0.71, G: 0.29
rs12546245	119,407,926	G	G	not registered in 1kJPN
rs62533397	119,418,994	C	T	C: 0.86, T: 0.14
rs7827332	119,424,808	A	G	A: 0.83, G: 0.17

111.7 kb

Disease haplotypes of families with repeat configurations 1 and 2 in *SAMD12*

Haplotypes were determined by SNPtyping, sequencing of BAC clones using PacBio RSII, and sequencing of genomic DNAs using 10X GemCode technology. Between rs3900767 and rs62533397, all the sequence variants matched except the repeat expansion mutations. Thus, it is likely that a common founder was present and one configuration derived from another configuration.

Supplementary Figure 5

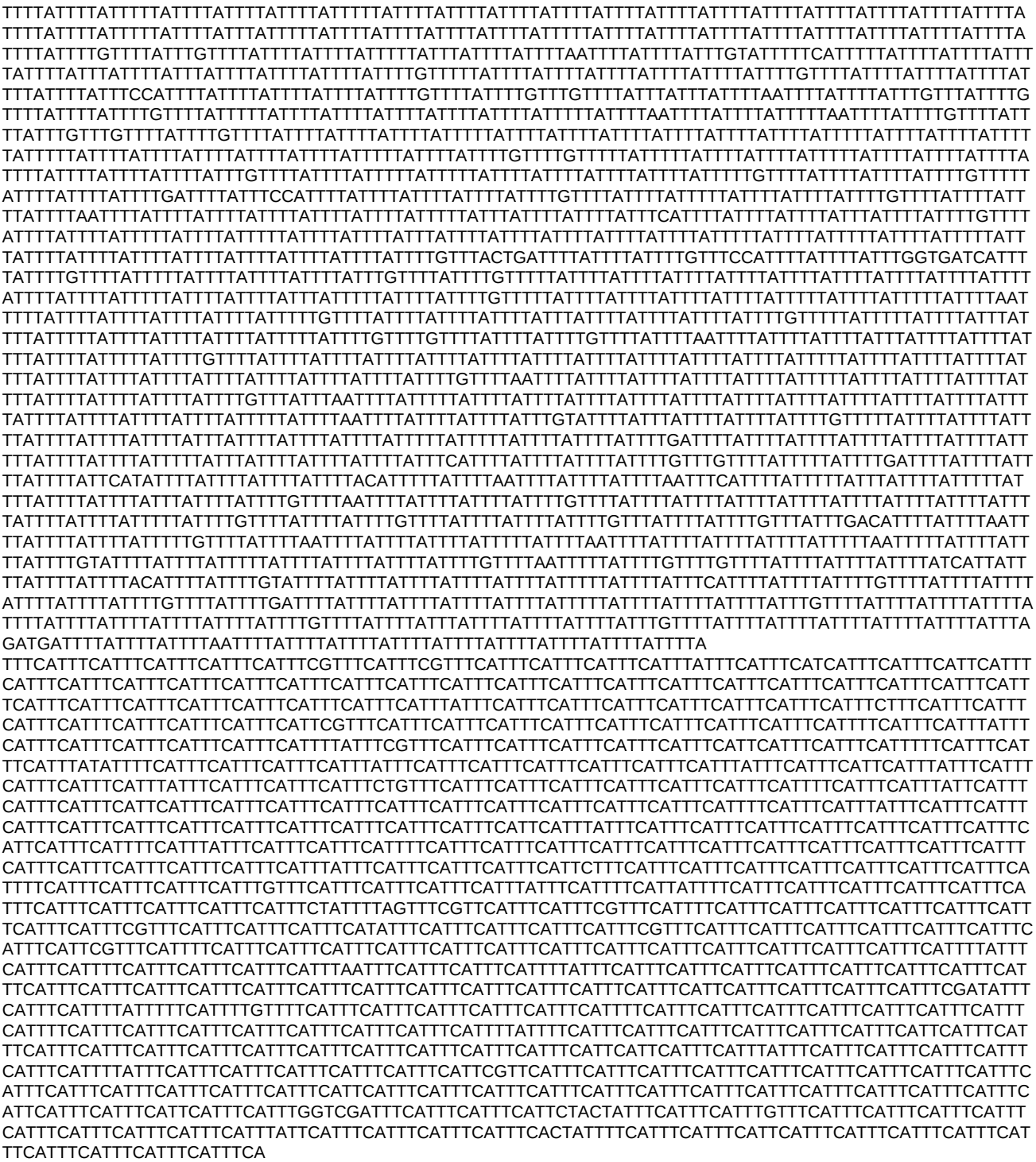


Dot plot of assembled Nanopore reads indicating the repeat expansion

Dot plots between the assembled contig with the STR expansion for each of the two samples (X-axis) and its corresponding hg19 region surrounding the STR site (chr8:119,379,052-119,379,172) without an expansion (Y-axis). Below the X-axis, the interleaved patterns of four colors show genomic sequence compositions, the red bars represent STR expansions in the two samples, and the respective red and blue-colored bars show nanopore read alignments in the plus and minus strands. Raw nanopore reads with STR expansions were aligned to the assembled contigs in the x-axis, supporting the presence of the STR expansions. In contrast, reads without STR expansions were aligned with the hg19 reference genome in the y-axis. These alignments are shown to the right of the y-axis, indicating the heterozygosity of the STR expansions. **(a)** data from patient II-6 in family F6906 (repeat configuration 1 in *SAMD12*). **(b)** data from patient II-1 in family F6115 (repeat configuration 2 in *SAMD12*).

Supplementary Figure 6

a Patient II-6 in family F6906 (repeat configuration 1)



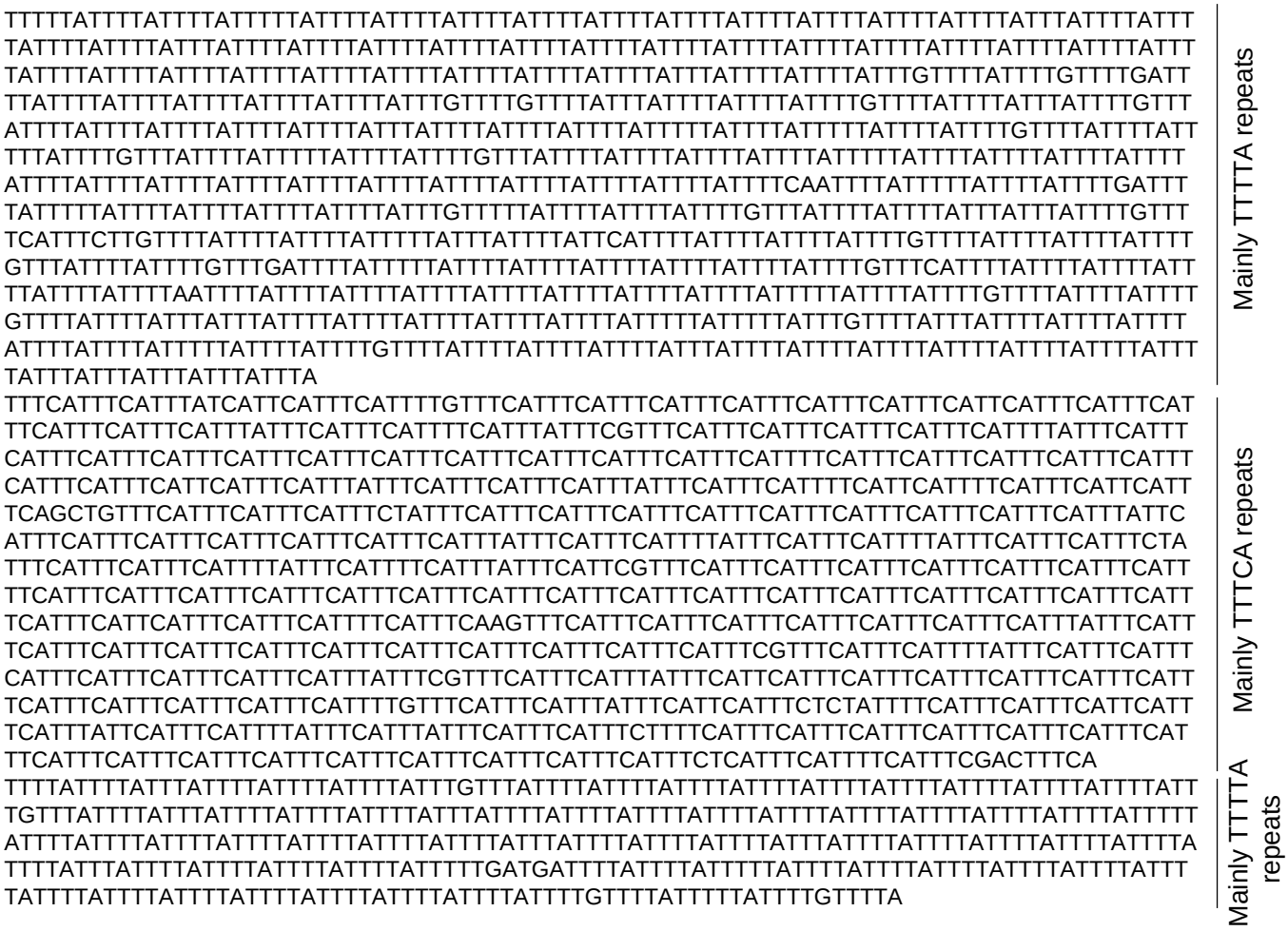
b $(TTTTA)_{598}\ (TTTCA)_{458}$

c

	Match ratio
Observed read (part 1) vs pure TTTTA repeat	87%
Observed read (part 2) vs pure TTTCA repeat	91%

Supplementary Figure 6 (continued)

d Patient II-1 in family F6115 (repeat configuration 2)



e $(TTTTA)_{221}$ $(TTTCA)_{225}$ $(TTTTA)_{81}$

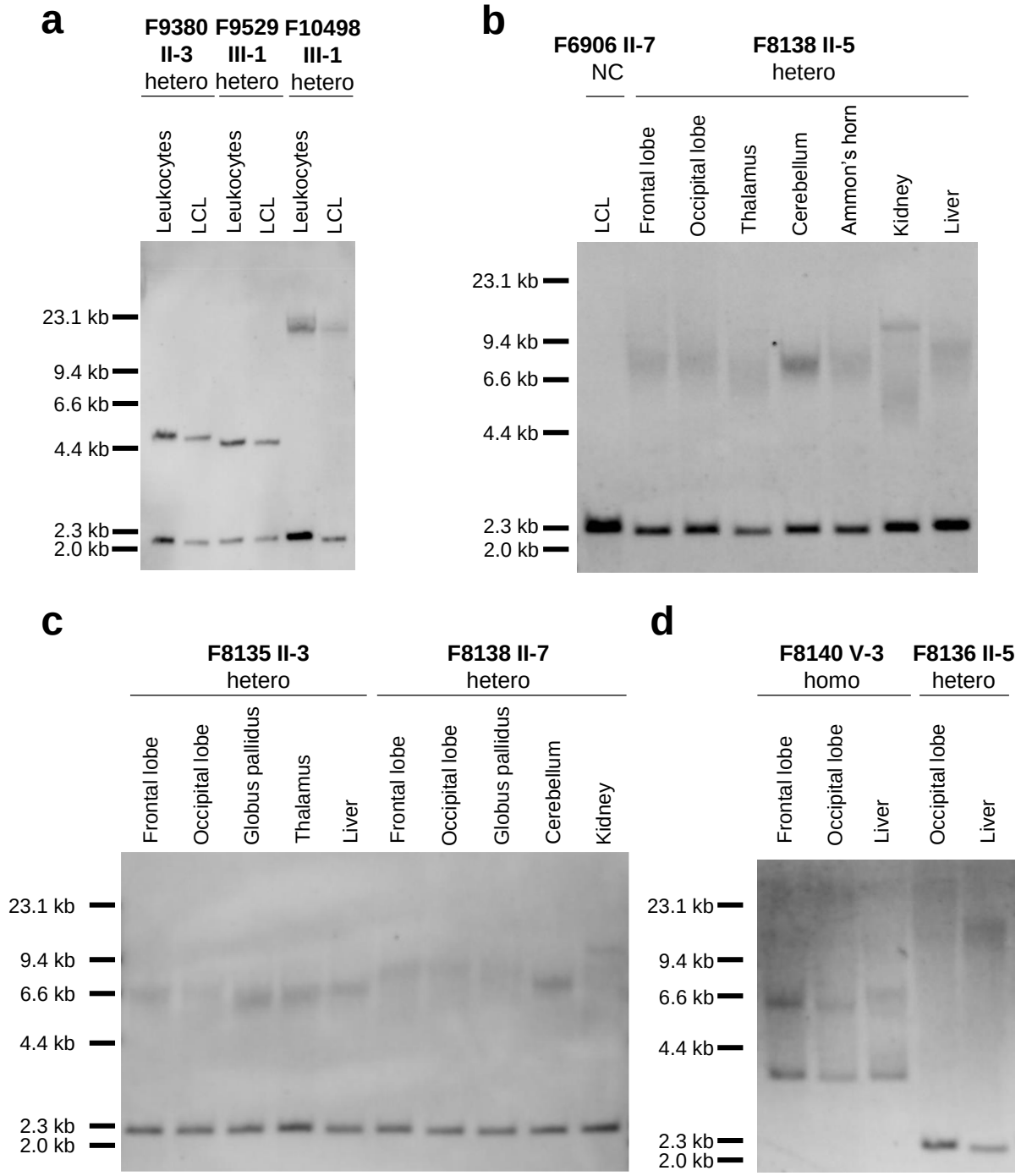
f

	Match ratio
Observed read (part 3) vs pure TTTTA repeat	88%
Observed read (part 4) vs pure TTTCA repeat	89%
Observed read (part 5) vs pure TTTTA repeat	90%

Nanopore reads indicating the repeat expansion in *SAMD12* using 1D^2 technology

Raw Nanopore 1D^2 reads of repeat configuration 1 (patient II-6 in family F6906, **a**) and repeat configuration 2 (patient II-1 in family F6115, **d**) of *SAMD12* are shown. The number of TTTCA and TTTTA motifs counted using Tandem Repeat Finder (Version4.09) are shown in **(b)** and **(e)**, respectively. Care has to be taken to interpret these numbers of motifs because they may still include errors. The match ratios between observed read and estimated pure repeats are around 90% (**c** and **f**, respectively).

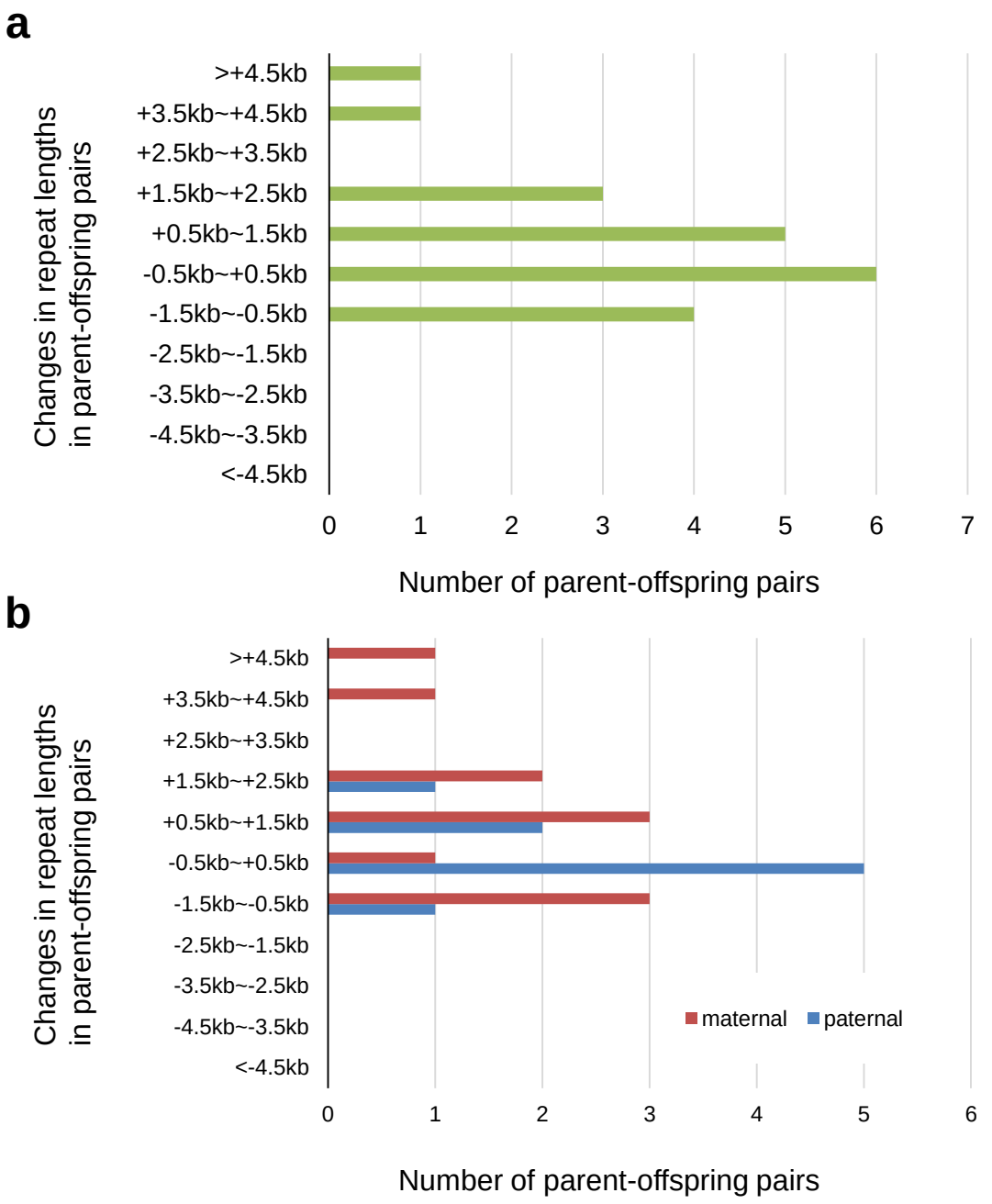
Supplementary Figure 7



Somatic instability of the repeat expansion in *SAMD12* in various tissues/cells

All the experiments were performed after SacI digestion of genomic DNA and hybridization was performed using probes 1a and 1b. **(a)** Lengths of the repeat expansion of leukocytes and those of lymphoblastoid cell lines, which were comparable. **(b–d)** Length of the repeat expansions in various part of the brains and other tissues. Genomic DNA from patient II-5 of family F8138, patient II-3 of family F8135, patient II-7 of family F8138, and patient II-5 of family F8136 show smear patterns, whereas a homozygous patient (V-3 of family F8140) shows rather discrete bands. Genomic DNAs from cerebellum tend to show less somatic instability. The experiment was performed once. Uncropped image is shown in **Supplementary Data**. Hetero, heterozygote; homo, homozygote; LCL, lymphoblastoid cell line; NC, non-carrier.

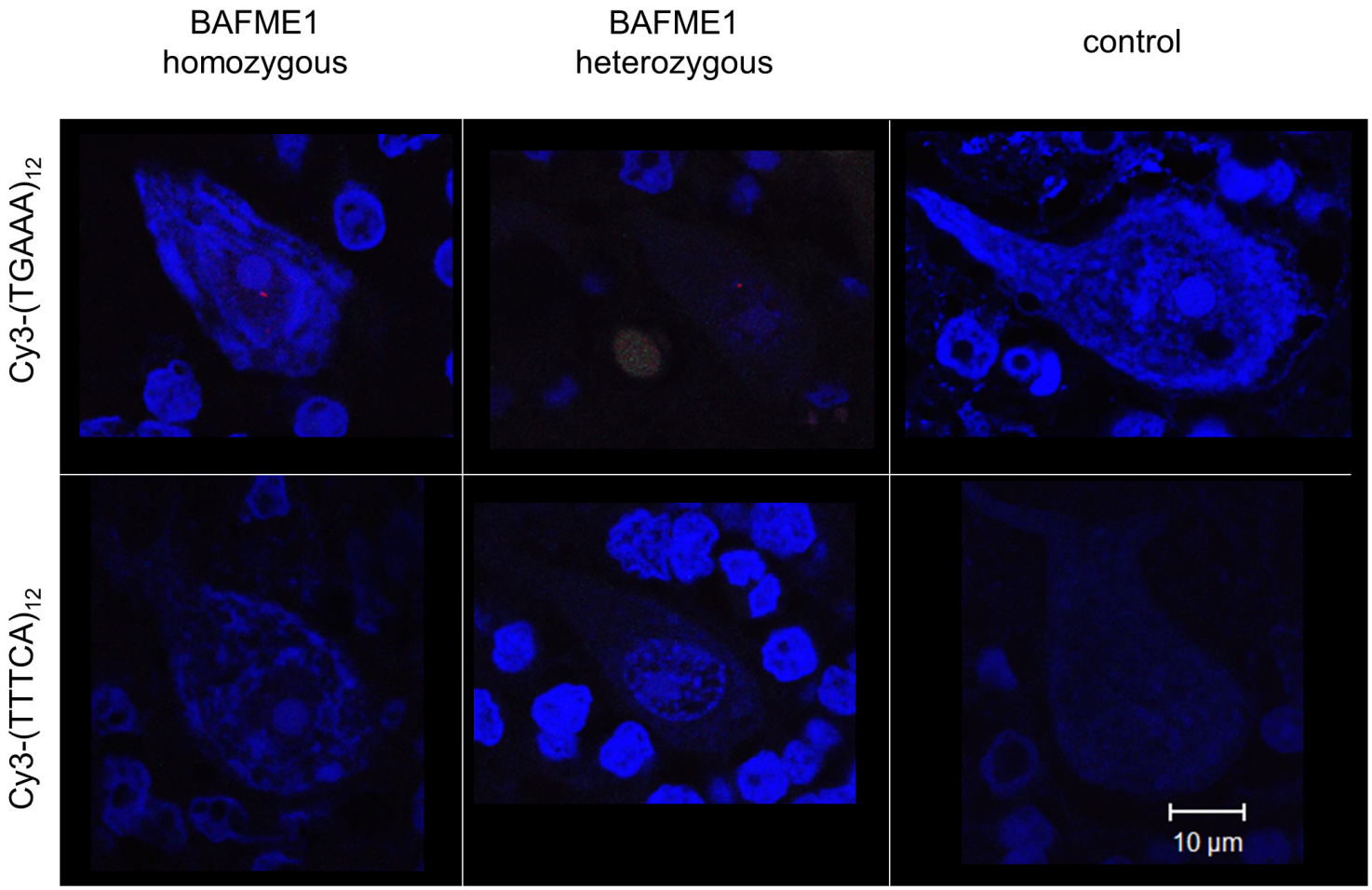
Supplementary Figure 8



Intergenerational instability of repeat lengths in *SAMD12*

(a) Intergenerational instability in parent-offspring pairs. The average increase in the lengths of expanded repeats observed in parent-offspring pairs (n = 20) was +0.89 kb (95% confidence interval, +0.12 to +1.48 kb).
(b) Intergenerational instability of repeat lengths in father-offspring pairs and mother-offspring pairs. Expansions tended to be larger in maternal transmissions (n=11) (+1.2 kb, 95% confidence interval, -0.03 to +2.5 kb) than in paternal transmission (n=9) (+0.28 kb, 95% confidence interval, -0.38 to +0.94 kb), although the difference was not statistically significant (*P*=0.29, two-tailed Wilcoxon rank sum test).

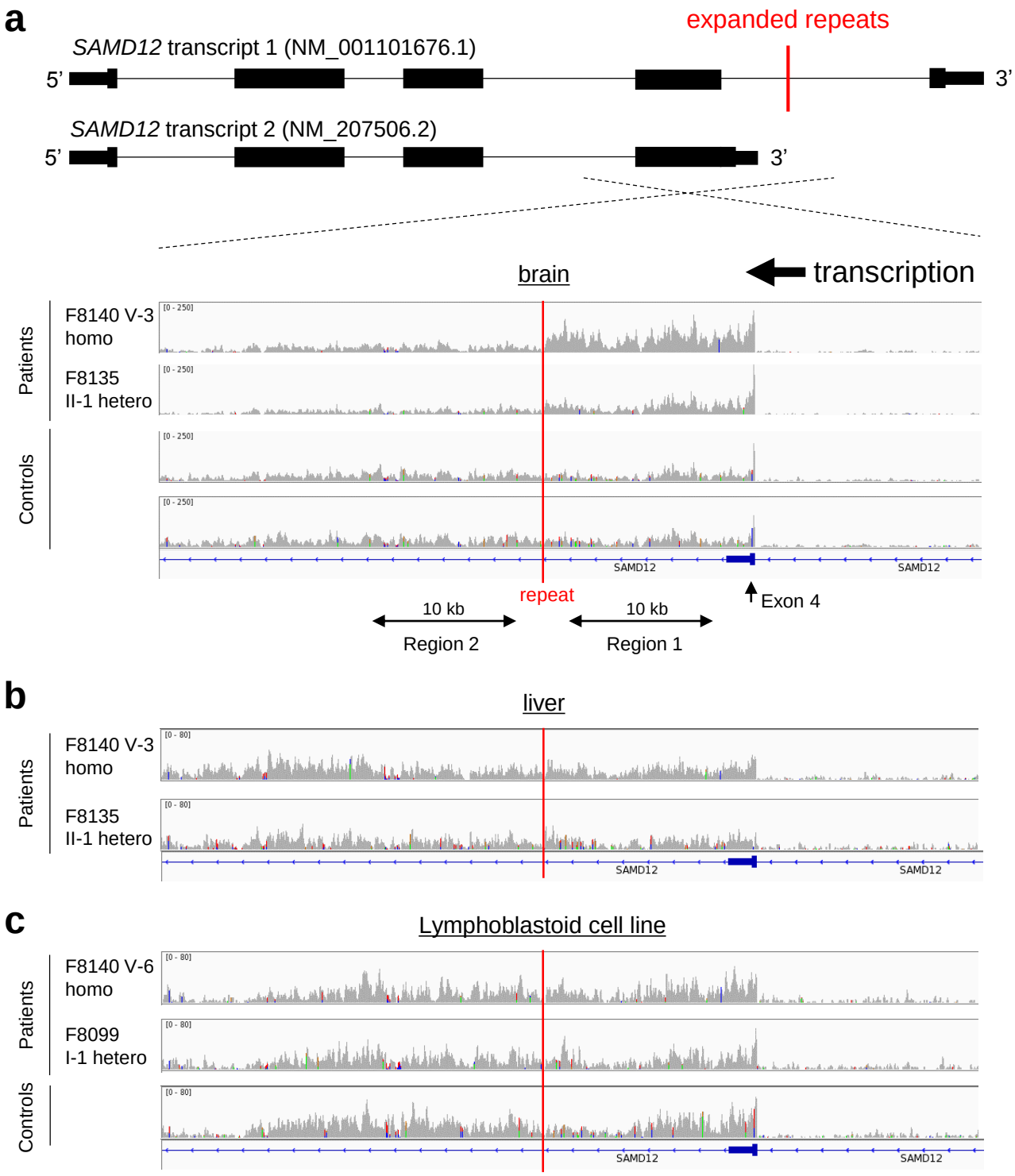
Supplementary Figure 9



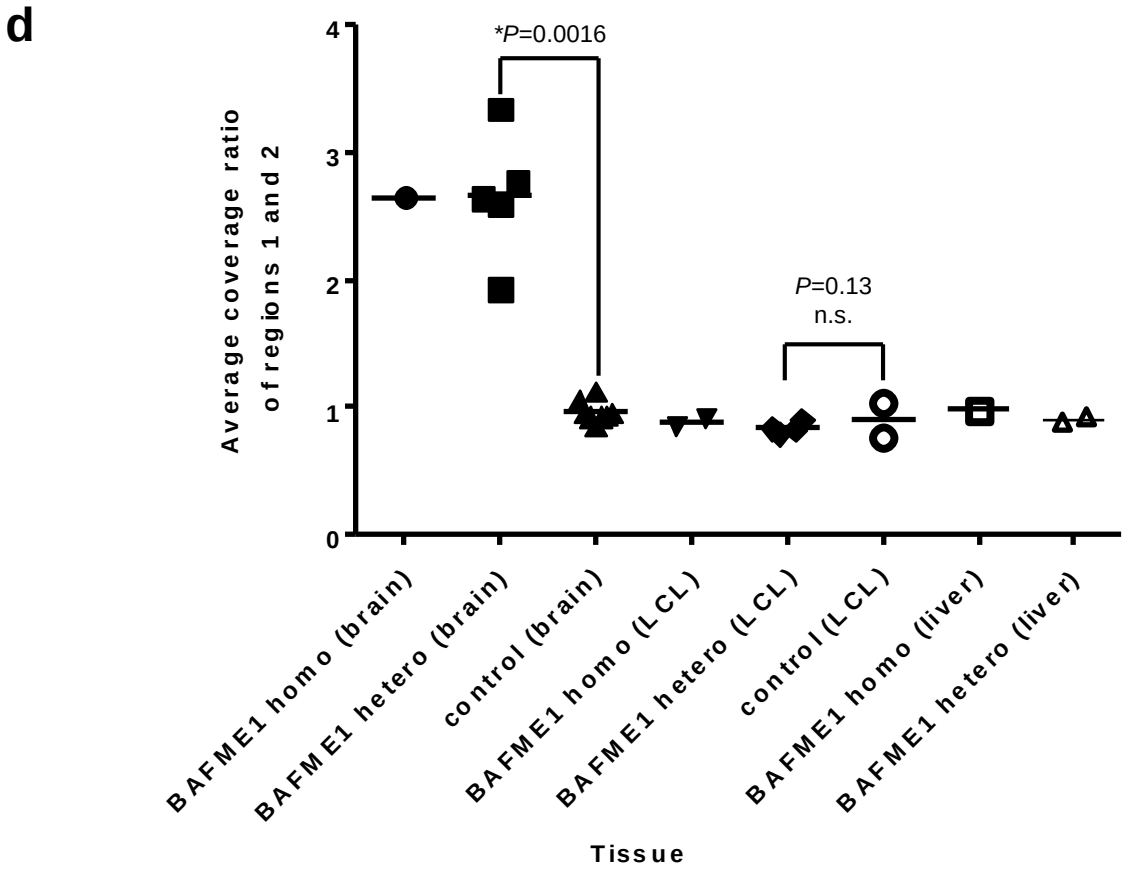
RNA foci observed in cerebella of patients carrying expanded repeats in homozygous or heterozygous state

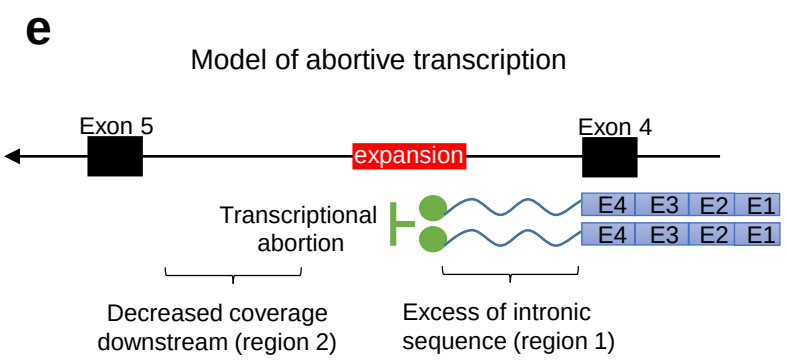
Fluorescent *in situ* hybridization (FISH) analysis using oligonucleotides (Cy3-$(TGAAA)_{12}$ and Cy3-$(TTTCA)_{12}$) revealed RNA foci in cerebella of the patients carrying expanded repeats in the homozygous or heterozygous state, whereas no RNA foci were observed in the control cerebellum. The experiments were performed twice with similar results.

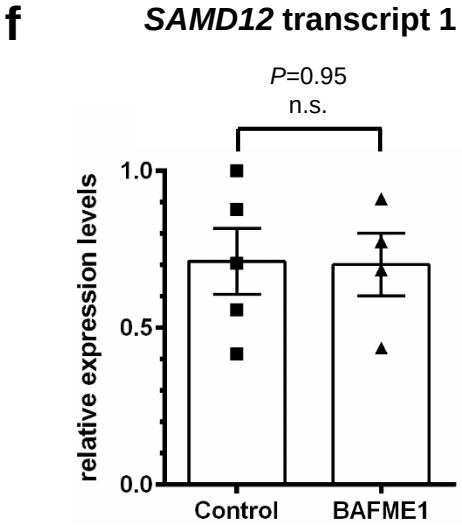
Supplementary Figure 10



Supplementary Figure 10 (continued)



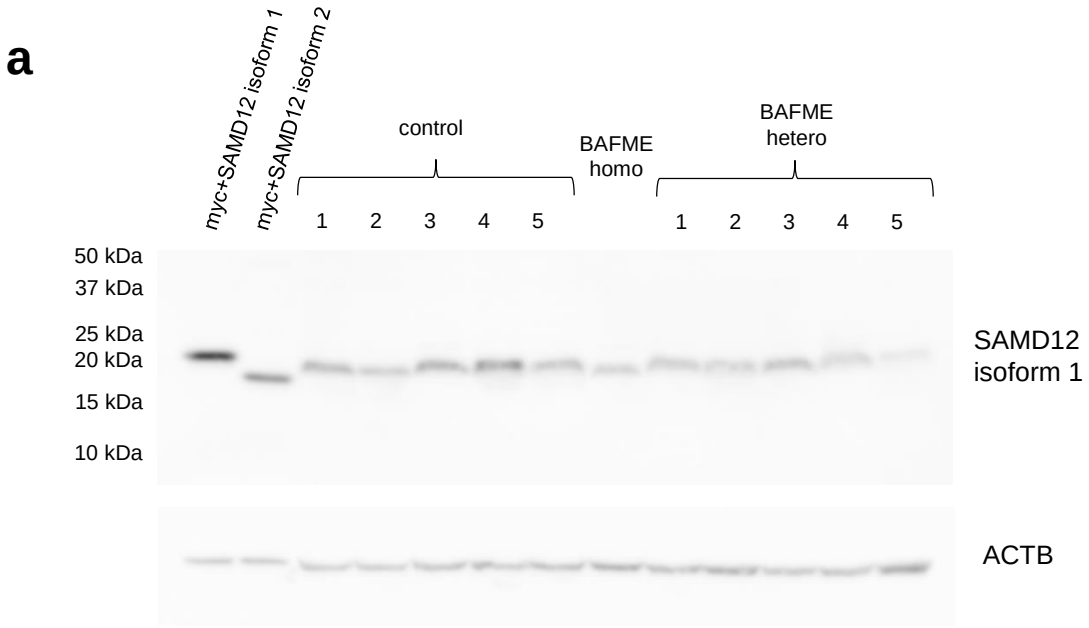


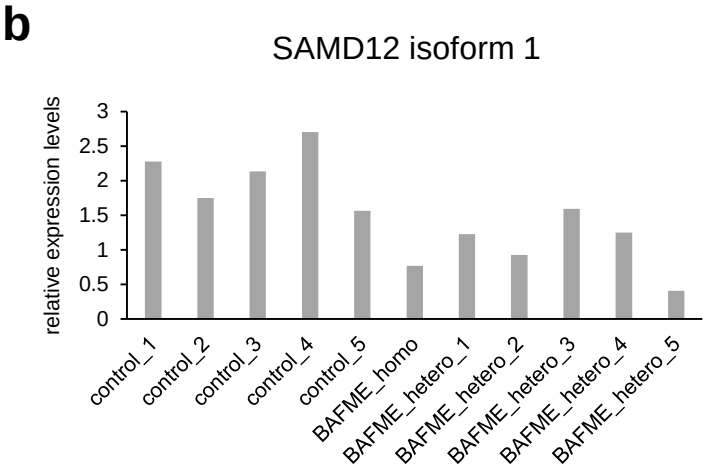


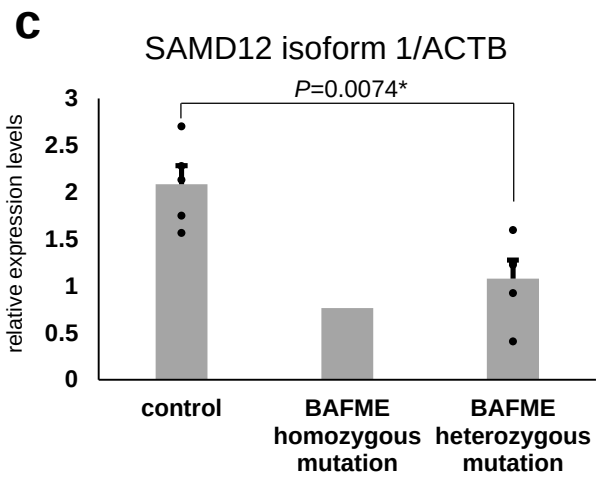
Abortive transcription in the brain was suggested by RNA-seq analysis

(a–c) Aligned short reads visualized using Integrative Genomic Viewer. Schematic representations of *SAMD12* transcripts are shown above. In homozygous (one patient) and heterozygous patients (two representative patients), read coverages increased in the region upstream of the expanded repeats (region 1) compared with those of the region downstream of the expanded repeats (region 2). This increase is presumably due to the abortive transcription at the expanded repeats. No such increase occurred in the liver (**b**) and lymphoblastoid cell lines (**c**) of the patients. Similar results were independently obtained in six brains from patients with BAFME1, eight control brains, six lymphoblastoid cell lines from patients with BAFME1, two control lymphoblastoid cell lines, and three livers from patients with BAFME1.
(d) Quantification of average read coverages of regions 1 and 2. The read coverage ratios of brains of a homozygous BAFME1 patient (n=1), heterozygous BAFME1 patients (n=5), and controls (n=8), those of lymphoblastoid cell lines of homozygous BAFME1 patients (n=2), heterozygous BAFME1 patients (n=4), and family members without disease (n=2), and those of livers of a homozygous patient (n=1) and heterozygous BAFME1 patients (n=2) are shown. All the patients have repeat expansion mutations in *SAMD12* (repeat configuration 1). The lines in the figure indicate the mean of the ratios. Wilcoxon's rank sum test was used for statistical analysis. "*" indicates that there is a significant difference ($p<0.05$).
(e) A model of abortive transcription. An increase in the read coverage ratio of region 1 compared with that of region 2, observed only in the brain from the patients, was supposed to be caused by abortive transcription due to abnormal repeat expansions (Ameur *et al.* Nat. Struct. Mol. Biol. **18**, 1435–1440 (2011)). E1–4 indicates exons 1–4.
(f) Expression levels of *SAMD12* transcript 1 in the brain (heterozygous BAFME1 patients (n=4) and controls (n=5)) determined by qRT-PCR analysis, using primers on exon 4/5 junctions and exon 5. The expression level of *RPL13A* was used as the endogeneous control. The data are shown as means and standard errors of the mean. The expression level of *SAMD12* transcript 1 was not altered in the brain from the patients (two-tailed Student's t-test) even when abortive transcription products are present. Note that the exact quantification of *SAMD12* transcript 2 is difficult in the presence of abortive transcription products.
Homo, homozygous; hetero, heterozygous; LCL, lymphoblastoid cell line; n.s., not significant.

Supplementary Figure 11



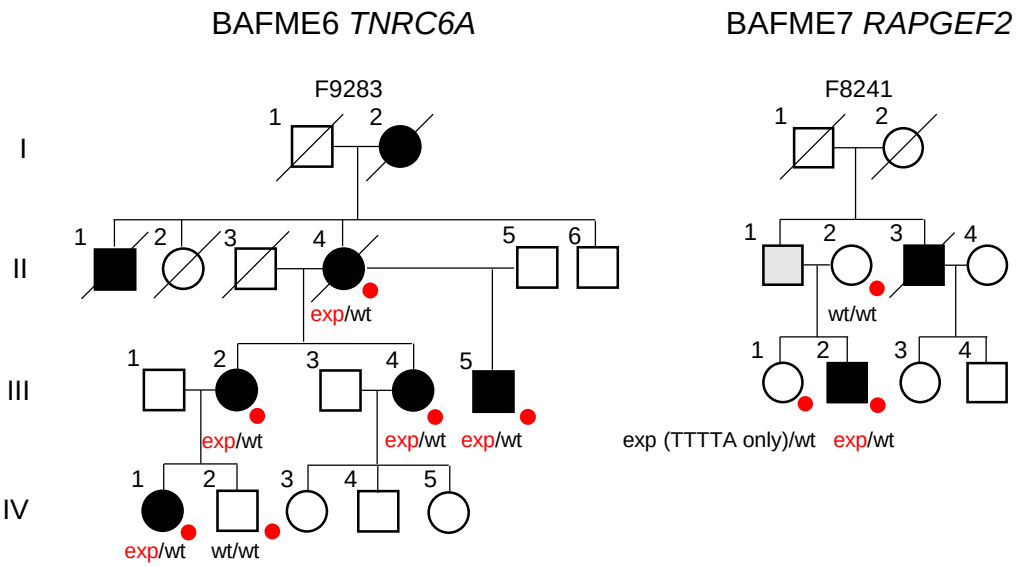




Expression levels of SAMD12 protein in brains

(a) The upper panel is the image of western blot of SAMD12 protein (isoform 1) of brains from patients with BAFME1 (occipital lobe, a homozygous patient (n=1) and heterozygous patients (n= 5)) and controls (occipital lobe, n=5). The lower panel is the image of western blot of β-actin.
(b) The graphic representation of the relative expression of SAMD12 protein (isoform 1) with respect to β-actin expression in each brain.
(c) The bar graph of relative expression levels of SAMD12 isoform 1 (a homozygous patient (n=1), heterozygous patients (n= 5), and controls (n=5)). The expression levels of SAMD12 isoform 1 in brains from patients with heterozygous repeat expansions were lower than those of the control brains (two-tailed t-test, P=0.0074, 95% confidence interval, 0.59 to 1.57 vs. 1.59 to 2.58). The data are shown as means and standard errors of the mean. "*" indicates that there is a significant difference (P<0.05).
The experiment was independently repeated twice with similar results. Uncropped image is shown in **Supplementary Data**.

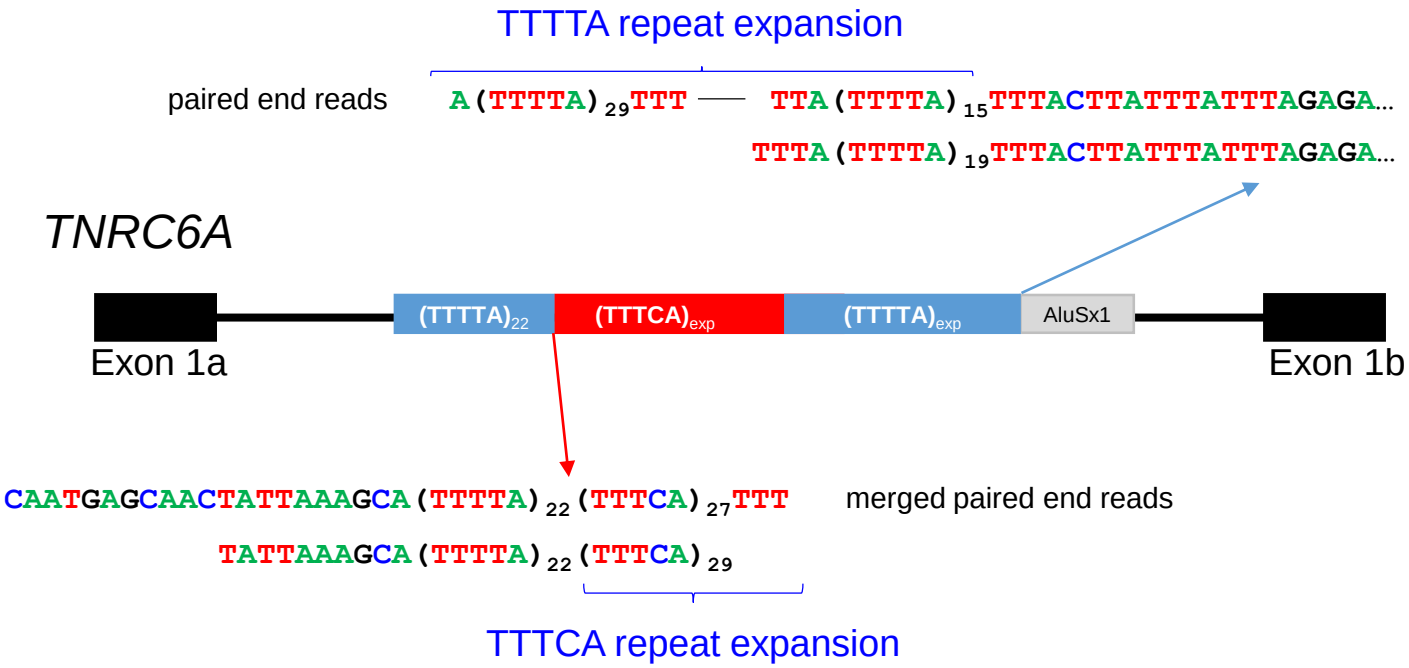
Supplementary Figure 12



TTTCA repeat expansions are identified in *TNRC6A* and *RAPGEF2* in families F9283 and F8241, respectively

The family F9283 had a TTTCA repeat expansion mutation in *TNRC6A*, whereas another family F8241 had a TTTCA repeat expansion mutation in *RAPGEF2*.

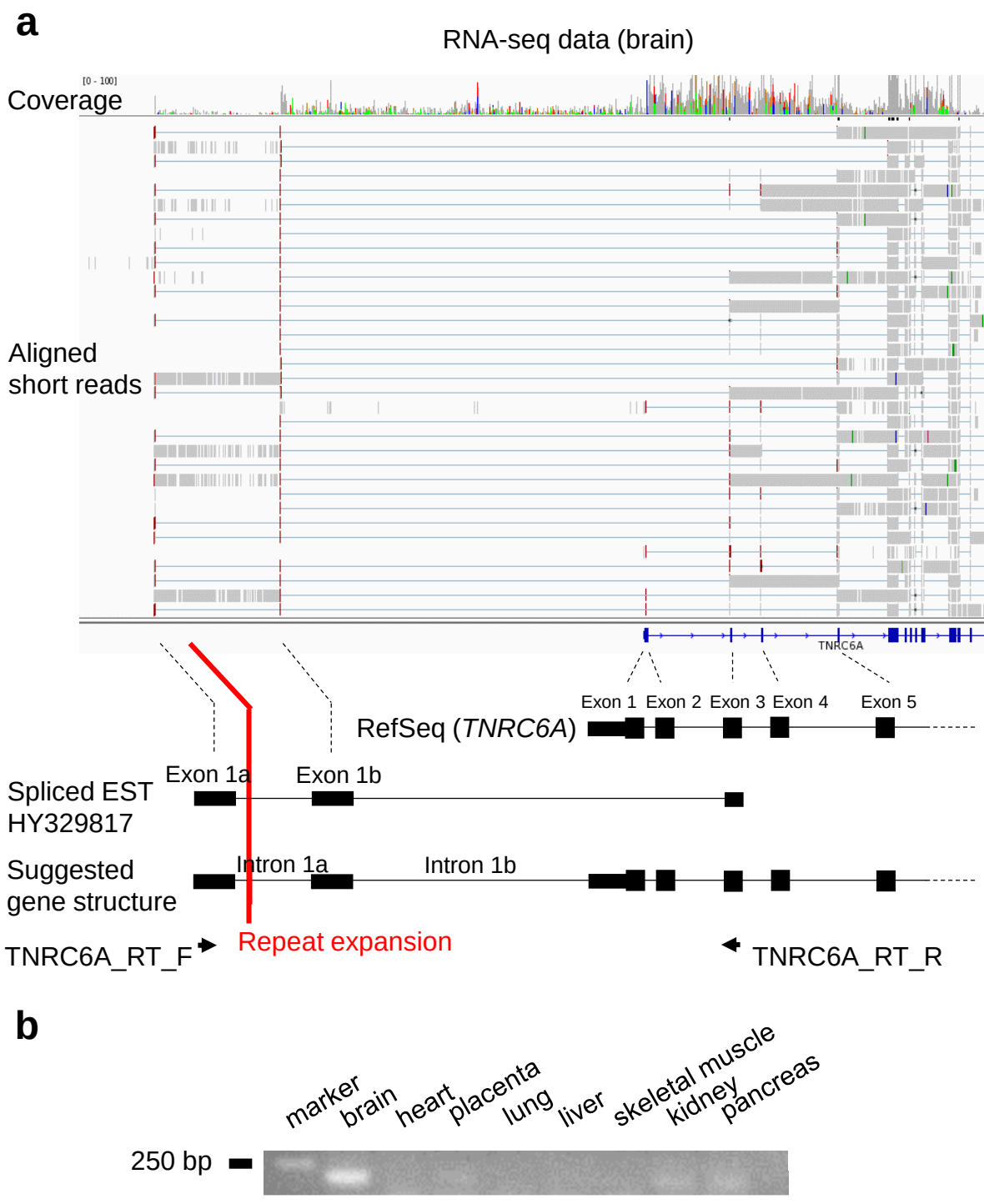
Supplementary Figure 13



TTTCA and TTTTA repeat expansion in *TNRC6A*

Manual alignment of short reads revealed that the mutant allele consists of expanded TTTCA repeats between the short (22) and expanded TTTTA repeats. There is $(TTTTA)_{18}$ in the reference genome sequence (chr16:24,624,761-24,624,850 in hg19).

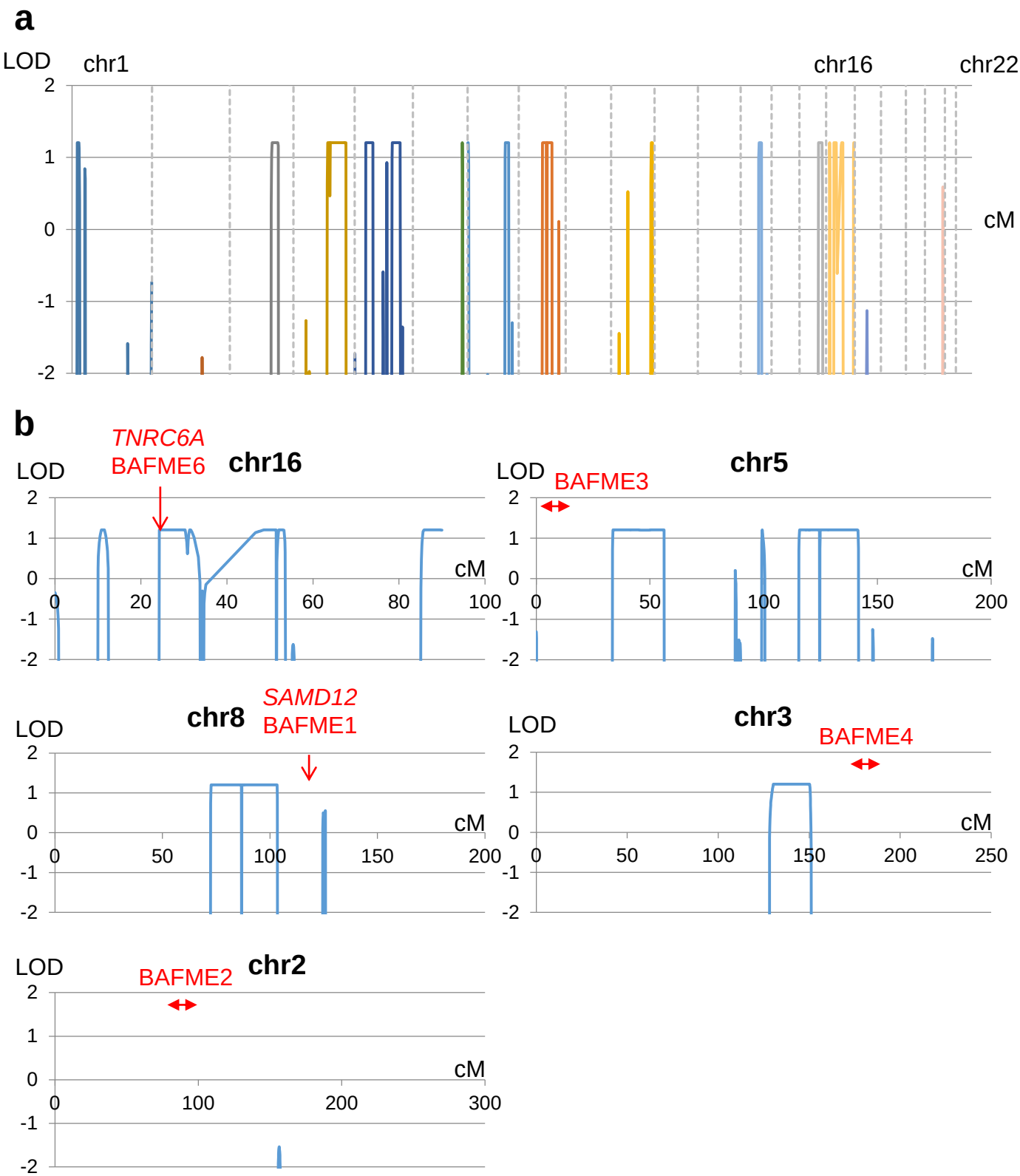
Supplementary Figure 14



Localization of expanded TTTCA repeats in *TNRC6A*

The repeat expansion was identified in the upstream region of exon 1 of *TNRC6A* registered in the RefSeq database. Brain RNA-seq data indicate that there are spliced transcripts flanking the repeat, which is also registered as spliced EST (HY329817) **(a)**. The transcript is further confirmed to be expressed in the brain, placenta, kidney, and pancreas by RT-PCR analysis **(b)**. These findings indicate that the expanded repeats are located in an intron (designated as intron 1a) of a transcript of *TNRC6A* expressed in the brain. The experiment was performed twice with similar results. Uncropped image is shown in **Supplementary Data**.

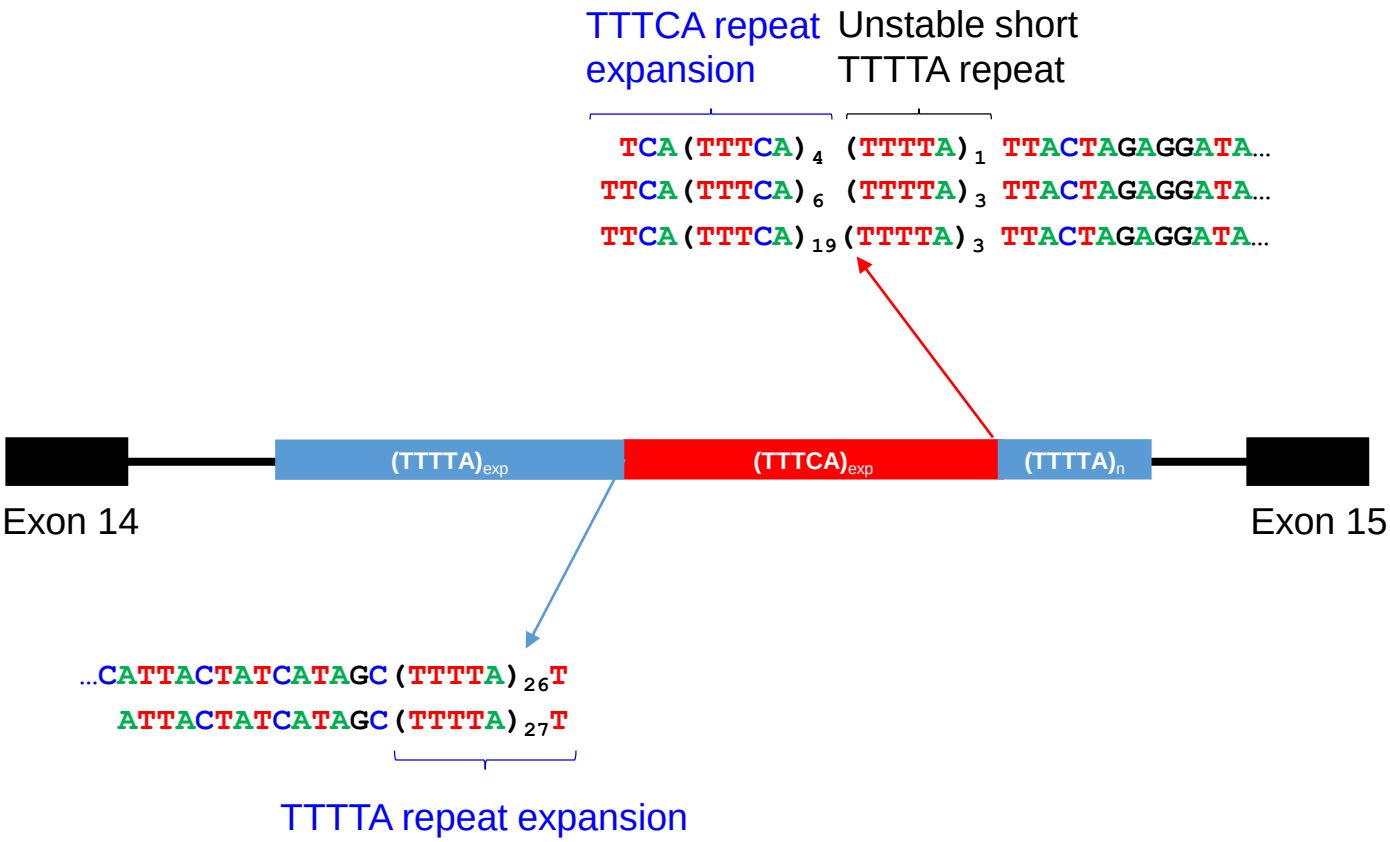
Supplementary Figure 15



Parametric linkage analysis of BAFME6 family

Parametric linkage analysis of all chromosomes **(a)** and known loci of BAFME **(b)**. The double-headed arrows indicate the known candidate regions. The analysis excludes other loci of BAFME1, BAFME2, BAFME3, and BAFME4.

Supplementary Figure 16



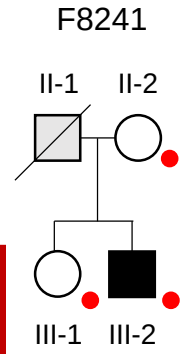
TTTCA and TTTTA repeat expansion in *RAPGEF2*

Manual alignment of short reads revealed that the mutant allele consists of expanded TTTCA repeats between the expanded and short (designated as “n”) TTTTA repeats. There is $(TTTTA)_5(TATTA)(TTTTA)_{12}$ in the reference genome sequence (chr4:160,263,679-160,263,768 in hg19)

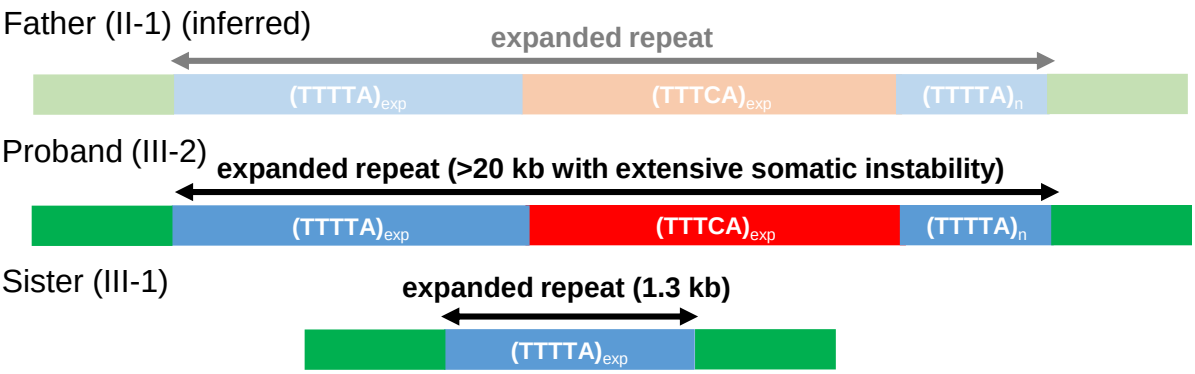
Supplementary Figure 17

a

	hg19 (Mb)	Hetero-zygosity								
D4S2976	155.80	0.81	(135)	?			131	141	135	131	135	141
D4S1629	158.33	0.68	(158)	?			166	166	158	166	158	166
D4S3354	159.55	0.80	(184)	?			180	176	184	180	184	176
D4S3328	159.69	0.62	(176)	?			176	179	176	176	176	179
RAPGEF2 repeat			?	?			$(TTTTA)_7TATTA(TTTTA)_{14}$	$(TTTTA)_5TATTA(TTTTA)_{11}$	$(TTTTA)_{exp}$	$(TTTTA)_7TATTA(TTTTA)_{14}$	(TTTCA)exp+ (TTTTA)exp	$(TTTTA)_5TATTA(TTTTA)_{11}$
D4S2918	160.33	0.38	(179)	?			181	181	179	181	179	181
D4S2980	160.61	0.76	(106)	?			106	106	106	106	106	106
D4S2982	161.09	0.84	(210)	?			194	210	210	194	210	210
D4S3033	161.73	0.76	(240)	?			238	240	240	238	240	240



b



Haplotype analysis of F8241

(a) Haplotype analysis suggests that patients III-1 and III-2 in family F8241 have the same paternal allele. This is also confirmed by dense SNP data analysis (data not shown). **(b)** The expanded repeat of the proband (III-2) shows a massive expansion (>20 kb) and somatic instability. The allele with the expansion in the sister without the disease consists only of TTTTA repeat, raising the possibility that the expanded TTTCA repeats could have been deleted when transmitted from the father, who also had tremulous movements in his hands.

Supplementary Table 1 Primers used for analysis of microsatellites

Primers	Sequence
D8S0379i	
M13(-21)-D8S0379i-F	5’-TGTAAAACGACGGCCAGTATAAATGGCTGCATG CTAG-3’
D8S0379i-R	5’-CTCAGATAATGGAATGAAGACTAA-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D8S522	
M13(-21)-D8S522-F	5’-TGTAAAACGACGGCCAGTGAAAACAAAACAGG CTCC-3’
D8S522-R	5’-TGACCAAACAACTAGACACC-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D4S2976	
M13(-21)-D4S2976F	5’-TGTAAAACGACGGCCAGTTTCCACCCCCAAA AGA-3’
D4S2976R	5’-CAAAGGTTTTTAACATCCCA-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D4S1629	
M13(-21)-D4S1629F	5’-TGTAAAACGACGGCCAGTTGGTTCTGCTTTTT CTCTCC-3’
D4S1629R	5’-TTTAACAGACAAATGACAAATCTG-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D4S3354	
M13(-21)-D4S3354F	5’-TGTAAAACGACGGCCAGTGGTGACATAGGCAT CACACA-3’
D4S3354R	5’- TCTCCAATTCAGAGAGCTGC-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D4S3328	
M13(-21)-D4S3328F	5’- TGTAAAACGACGGCCAGTCCAAGATCACACTA CTGCCC-3’
D4S3328R	5’- AGCCTCCTCTGTCAAGGAGT-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D4S2918	
M13(-21)-D4S2918F	5’- TGTAAAACGACGGCCAGTGCAGAAGCCTGG TCAT-3’
D4S2918R	5’- TCAGCCCCTTAATCACA-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D4S2980	
M13(-21)-D4S2980F	5’-TGTAAAACGACGGCCAGTCCTCAAGGTATCA GGCTGC-3’
D4S2980R	5’-TTAAGTACATGGTCCGTCCC-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D4S2982	
M13(-21)-D4S2982F	5’-TGTAAAACGACGGCCAGTCCAGTCAACCCCA ACAC-3’
D4S2982R	5’-GAAATGAAACATGATGACACC-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
D4S3033	
M13(-21)-D4S3033F	5’-TGTAAAACGACGGCCAGTTGCCTTTAAGAATG CAATG-3’
D4S3033R	5’-TGGTGTCAACGCAGTAAAT-3’
Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’

Supplementary Table 2 Primer list used for repeat-primed PCR analysis

Primer	Sequence
(1) Repeat-primed PCR analysis of *SAMD12* targeting TTTTA	
P1, Fam-SAMD12_ R	5’-fam-CTTGAGCCCCAGACAAGAAT-3’
P2, M13R-AAAAT4	5’-CAGGAAACAGCTATGACC$(AAAAT)_4$-3’
P2anchor, M13R	5’-CAGGAAACAGCTATGACC-3’
(2) Repeat-primed PCR analysis of *SAMD12* targeting TTTCA	
P3, Fam-SAMD12_F	5’-fam-TGCTACTGTAAAAAGATAAACAAAATG-3’
P4, M13R-TTTCA5	5’-CAGGAAACAGCTATGACC$(TTTCA)_5$-3’
P4anchor, M13R	5’-CAGGAAACAGCTATGACC-3’
(3) Repeat-primed PCR analysis of *TNRC6A* targeting TTTCA	
P5, TNRC6A _F	5’-TGTGGAGCTCCAAAGCCCTG-3’
P6, M13(-21)-TGAAA5	5’-TGTAAAACGACGGCCAGT$(TGAAA)_5$-3’
P6anchor, Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
(4) Repeat-primed PCR analysis of *TNRC6A* targeting TTTTA	
P7, TNRC6A_R	5’-CCAAGTATGCTCCTTCCCTACTT-3’
P8, M13(-21)-TTTTA5	5’-TGTAAAACGACGGCCAGT$(TTTTA)_5$-3’
P8anchor, Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
(5) Repeat-primed PCR analysis of *RAPGEF2* targeting TTTTA	
P9, RAPGEF2_intron14_F	5’-TTAGTCATTTTCTCTGTCTGGGAG-3’
P10, M13(-21)-TAAAA5	5’-TGTAAAACGACGGCCAGT$(TAAAA)_5$-3’
P11anchor, Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’
(6) Repeat-primed PCR analysis of *RAPGEF2* targeting TTTCA	
P11, RAPGEF2_intron14_R	5’-GCAGTATGTACACATAGCAGTGC-3’
P12, M13(-21)-TTTCA5	5’-TGTAAAACGACGGCCAGT$(TTTCA)_5$-3’
P12anchor, Fam-M13(-21)	5’-fam-TGTAAAACGACGGCCAGT-3’

Supplementary Table 4 Number of reads filled with 5'-TTTCA/5'-TGAAA, 5'-CTTCA/5'-TGAAG, and 5'-GTTCA/5'-TGAAC motifs in RNA-seq and whole-genome sequence analyses.

a

RNA-seq

ID	Disease	Source	Number of reads	Read length	Number of reads filled with 5'-TTTCA or 5'TGAAA motif	Number of reads filled with 5'-CTTCA or 5'-TGAAG motif	Number of reads filled with 5'-GTTCA or 5'-TGAAC motif
F8136 I-1	BAFME1	liver	155,503,230	101PE	23	0	0
F8140 V-3	BAFME1 hom	liver	156,933,316	101PE	16	0	0
F8135 II-1	BAFME1	liver	137,428,560	101PE	12	0	0
F8140 V-6	BAFME1 hom	LCL	131,550,452	101PE	23	0	0
F8398 II-4	BAFME1 hom	LCL	139,338,450	101PE	38	0	0
F8398 III-3	BAFME1	LCL	159,131,286	101PE	67	0	0
F8146 III-1	BAFME1	LCL	155,302,968	101PE	3	0	0
F2376 III-1	BAFME1	LCL	141,962,356	101PE	40	0	0
F8099 I-1	BAFME1	LCL	140,966,042	101PE	12	0	0
F6115 I-1	unaffected	LCL	149,959,512	101PE	0	0	0
F6906 II-7	unaffected	LCL	150,014,140	101PE	0	0	0

b

DNA-seq

ID	Disease	Source	Number of reads	Read length	Number of reads filled with 5'-TTTCA or 5'TGAAA motif	Number of reads filled with 5'-CTTCA or 5'-TGAAG motif	Number of reads filled with 5'-GTTCA or 5'-TGAAC motif
F6906 II-6	BAFME1	blood	1,523,327,112	100PE	400	0	0
F8136 I-1	BAFME1	brain	1,144,146,288	101PE	1,112	10	1
F8135 II-1	BAFME1	brain	1,112,108,282	150PE	323	9	0
F8140 V-3	BAFME1 hom	brain	904,833,992	150PE	526	3	0
F9283 III-2	BAFME6	blood	1,146,555,396	150PE	385	0	0
F8241 III-2	BAFME7	blood	1,325,596,586	150PE	818	0	0

BAFME1; benign adult familial myoclonic epilepsy caused by TTTCA pentanucleotide repeat expansion in *SAMD12,*
BAFME6; benign adult familial myoclonic epilepsy caused by TTTCA pentanucleotide repeat expansion in *TNRC6A,*
BAFME7; benign adult familial myoclonic epilepsy caused by TTTCA pentanucleotide repeat expansion in *RAPGEF2,*
hom; homozygous, LCL; lymphoblastoid cell line, PE; paired end

Supplementary Table 5 The number of reads containing various repeat motifs shown by whole-genome sequence analysis using TRhist.

	BAFME1	BAFME1	BAFME6	BAFME7	control	control	control	control	control	control	control
Family	F8135	F8140	F9283	F8241	5763_1	5764_1	4303_2	5765_1	6071_1	5766_1	5767_1
Individual	II-1	V-3	III-2	III-2							
Average coverage	44.1X	32.8X	47.7X	55.0X	71.0X	83.1X	49.2X	53.1X	50.8X	36.9X	40.5X
Repeat motif											
C	14223	19902	32241	18794	58408	71353	36671	33797	15309	26468	20690
AACCCT	25236	32227	10564	6196	23133	13819	8681	7322	9092	8175	7298
A	2520	3922	5038	3176	6072	6343	5364	4161	1735	3425	2812
AAAGG	124	192	456	158	240	483	211	131	198	143	118
AAAG	160	187	425	330	331	512	203	199	410	218	251
AAATG	323	536	385	818	0	0	0	0	0	0	0
AAAAT	253	145	285	200	122	247	105	149	26	192	86
AATGG	67	80	142	16	33	115	7	102	189	24	15
AAG	108	116	70	142	72	103	220	49	70	71	61
AATAG	35	62	67	6	4	0	29	25	7	35	2
AT	47	65	47	4	171	358	54	60	21	36	208
AAAAG	70	14	35	116	124	38	52	11	146	72	78
AGC	0	2	8	0	0	4	10	0	9	0	9
AACAT	0	0	18	0	10	0	0	7	0	0	0
AAAGGATGAGGG	11	9	14	10	24	6	7	14	15	4	6
AG	5	4	11	2	5	3	2	1	7	5	6
AAGAG	9	4	9	6	13	12	6	4	3	0	1
ATCC	0	4	4	0	4	10	4	4	2	6	1
AAACC	14	10	2	4	15	4	9	4	2	0	5
AGCAT	0	7	0	98	11	0	0	0	2	0	57
ACC	7	12	2	4	1	1	0	0	1	5	1
AATATATATATG	1	0	1	10	0	0	0	3	0	0	0
AATAC	17	0	0	18	0	0	5	10	20	10	0
AAGGAG	30	2	0	0	0	2	1	0	0	2	0
AAGAGG	0	0	0	0	0	29	0	0	0	0	0
AAGAC	0	0	0	0	96	0	0	0	0	0	0
AAAGTTCATGGT	4	4	0	16	23	0	0	0	0	0	3
AAACT	0	0	0	0	0	0	0	0	10	0	0

In the probands of families F9283 and F8241, large numbers of reads containing 5'-AAATG (equivalent to 5'-TTTCA in the opposite strand) were similarly observed as in the two BAFME patients with TTTTCA repeat expansions in *SAMD12* (F8135 and F8140). Neither AAATG (=TTTCA), AACTG (=GTTCA), nor AAGTG (=CTTCA) repeat expansions were observed in the seven controls.

Supplementary Table 6 TTTCA/TGAAA or TTTTA/TAAAA repeats in the reference sequence (hg19)

a

	$(TTTCA)_{>9}/(TGAAA)_{>9}$ repeat	$(TTTTA)_{>9}/(TAAAA)_{>9}$ repeat
In the reference sequence (hg19)	6	711
In the candidate region of BAFME2 on chromosome 2p	0	3
Those located in intron	0	1
In the candidate region of BAFME3 on chromosome 5p	0	2
Those located in intron	0	2
In the candidate region of BAFME4 on chromosome 3q	0	0
Those located in intron	0	0

b

Repeat motif (in the transcriptional direction)	Gene	Physical position (hg19)	Repeat sequence in the reference sequence in the transcriptional direction
TTTCA	*MED12L*	chr3:151,086,478-151,086,539	$A(TTTCA)_{12}T$
	ZDHHC14	chr6:157,955,566-157,955,632	$A(TTTCA)_{12}T$
TGAAA	*GNAQ*	chr9:80,263,486-80,263,539	$AAA(TGAAA)_{10}T$
TTTTA	*MUSK*	chr9:113,463,975-113,464,207	$TTTA(TTTTA)_{46}TTT$
TAAAA	*KCNIP4*	chr4:21,716,410-21,716,715	$AAA(TAAAA)_{60}TAA$

c

Locus	Gene	Physical position (hg19)	Repeat sequence in the reference sequence in the transcriptional direction
BAFME2 candidate region	*STARD7*	chr2:96,862,805-96,862,862	$TTTA(TTTTA)_{10}TTTT$
BAFME3 candidate region	*MARCH6*	chr5:10,356,456-10,356,523	$TTTA(TTTTA)_{12}TTTT$
BAFME3 candidate region	*TRIO*	chr5:14,421,337-14,421,394	$TTTA(TTTTA)_{10}TTTT$

(a) The number of $(TTTCA)_{>9}/(TGAAA)_{>9}$ or $(TTTTA)_{>9}/(TAAAA)_{>9}$ repeat in the reference sequence (hg19) are shown. Of these, those located in the candidate regions of BAFME2, BAMFE3, and BAFME4 and those located in introns are shown.

(b) The longest intronic TTTTA/TAAAA or TTTCA/TGAAA repeats in the reference sequence.

(c) $(TTTCA)_{>9}/(TGAAA)_{>9}$ or $(TTTTA)_{>9}/(TAAAA)_{>9}$ repeat located in introns of the candidate regions of BAFME were shown.

Supplementary Table 7 Primers used for generation of digoxigenin-labeled probes used in Southern blot hybridization analysis.

Primers	Restriction enzyme	Lengths of the probe	Sequences
***SAMD12* (BAFME1)**	SacI		
SAMD12-probe-F2		508 bp	5'-TACAGGGGTTTTTGCCAGAC-3'
SAMD12-probe-R3			5'-CCAACCTCAAATGCATTGACCAC-3'
SAMD12-probe-F3		633 bp	5'-GGATCATTCCTGATAGAAAAAGTCC-3'
SAMD12-probe-extremeR			5'-AGCGGCTAATGGTAGATAAGG-3'
***TNRC6A* (BAMFE6)**	EcoRI		
TNRC6A-probe-F		202 bp	5'-ATTCACATACCTTTCTGTCTCCC-3'
TNRC6A-probe-R2			5'-GCATTCTACACTGTCTTAGTGA-3'
TNRC6A-probe-F2		492 bp	5'-TCACTAAGACAGTGTAGAATGC-3'
TNRC6A-probe-R			5'-TCACAATTCACATTTGGCTTCC-3'
***RAPGEF2* (BAFME7)**	PvuII		
RAPGEF2-long-F1		946 bp	5'-ACTAGAGGATATTTGCTCAGTCGG-3'
RAPGEF2-long-R1			5'-GGTGATTCTCAGCACATCCTACTATA-3'
RAPGEF2-long-F2		904 bp	5'-TATAGTAGGATGTGCTGAGAATCACC-3'
RAPGEF2-long-R2			5'-TTTAATTGCACAACTTTTCCCTTC-3'
RAPGEF2-long-F3		1043 bp	5'-AACTTTACTCACTGGAATGAGGC-3'
RAPGEF2-long-R3			5'-ACAGCATCCTGGGTAAATCTAATAATC-3'
RAPGEF2-long-F4		1264 bp	5'-TATTGTCCTAAATTGCTTTGCC-3'
RAPGEF2-long-R4			5'-GGAAAGATGAAGTGTAATTTAGCTCC-3'

Supplementary Table 8 Primer sequences used for BAC cloning

Primer	Direction	Sequence
centromeric primer	Forward	5'- CACATCTTGCTAACTAATAGCTGG-3'
	Reverse	5'- CCCAGCCCTCATTTTGTTTATC-3'
telomeric primer	Forward	5'- TGCTGGCTAGAGGAACCATT-3'
	Reverse	5'- TCCTTAAGATTTCCCCCGAC-3'

Supplementary Table 9 The ages at death and the causes of death of autopsied patients

ID	Age at death	Sex	Cause of death
F8135 I-2 (het)	77	F	pneumonia
F8136 II-5 (het)	46	F	ovarian cancer
F8135 II-3 (het)	66	F	gastric cancer
F8138 II-7 (het)	78	M	drowning in bathtub
F8138 II-5 (het)	83	M	myocardial infarction
F8140 V-3 (hom)	71	F	pneumonia

het, heterozygote; hom, homozygote

Supplementary Table 10 Primer sequences used for RT-PCR

Transcript	Direction	Location	Sequence
SAMD12 transcript 1	Forward	Exon 4/5 junction	5’- CACACAAGCATCATCTGAGGG-3’
	Reverse	Exon 5	5’- GCAGTGCCATGGGTTAGTCT-3’
RPL13A	Forward	Exon 1	5’- AACCTCCTCCTTTTCCAAGC -3’
	Reverse	Exon 2	5’- GCAGTACCTGTTTAGCCACGA -3’
TNRC6A	Forward	Exon 1a	5’-GGTGCCCCATCATTTTGCAGC-3’
	Reverse	Exon 3	5’-CAACTGTTCTTCTTCTTGCACTAAATCC-3’