nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-019-0458-z

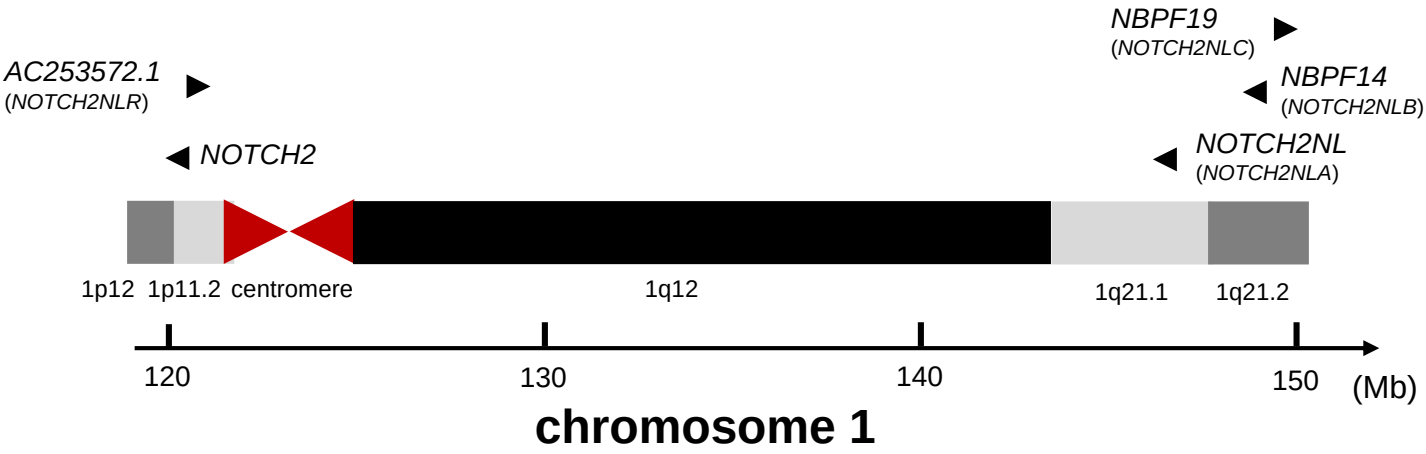
In the format provided by the authors and unedited.

Noncoding CGG repeat expansions in neuronal intranuclear inclusion disease, oculopharyngodistal myopathy and an overlapping disease

**Hiroyuki Ishiura[1], Shota Shibata[1], Jun Yoshimura[2], Yuta Suzuki[2], Wei Qu[2], Koichiro Doi[2,20], M. Asem Almansour[1], Junko Kanda Kikuchi[1], Makiko Taira[1,21], Jun Mitsui[1,3], Yuji Takahashi[1,4], Yaeko Ichikawa[1,5], Tatsuo Mano[1], Atsushi Iwata[1], Yasuo Harigaya[6], Miho Kawabe Matsukawa[1], Takashi Matsukawa[1,3], Masaki Tanaka[1,7], Yuichiro Shirota[1], Ryo Ohtomo[1], Hisatomo Kowa[1,22], Hidetoshi Date[1,4], Aki Mitsue[1], Hiroyuki Hatsuta[8,9], Satoru Morimoto[8], Shigeo Murayama[8], Yasushi Shiio[10], Yuko Saito[11], Akihiko Mitsutake[1], Mizuho Kawai[1], Takuya Sasaki[1], Yusuke Sugiyama[1], Masashi Hamada[1], Gaku Ohtomo[1], Yasuo Terao[1,23], Yoshihiko Nakazato[12], Akitoshi Takeda[9], Yoshio Sakiyama[13], Yumi Umeda-Kameyama[14], Jun Shinmi[1], Katsuhisa Ogata[15], Yutaka Kohno[16], Shen-Yang Lim[17], Ai Huey Tan[17], Jun Shimizu[1], Jun Goto[1,18], Ichizo Nishino[19], Tatsushi Toda[1], Shinichi Morishita[2] and Shoji Tsuji[1,3,7]*

[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]Department of Molecular Neurology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [4]Department of Neurology, National Center of Neurology and Psychiatry, Tokyo, Japan. [5]Department of Neurology, Kyorin University, Tokyo, Japan. [6]Department of Neurology, Maebashi Red Cross Hospital, Maebashi, Japan. [7]Institute of Medical Genomics, International University of Health and Welfare, Chiba, Japan. [8]Department of Neurology and Neuropathology (the Brain Bank for Aging Research), Tokyo Metropolitan Geriatric Hospital and Institute of Gerontology, Tokyo, Japan. [9]Department of Neurology, Osaka City University Graduate School of Medicine, Osaka, Japan. [10]Department of Neurology, Tokyo Teishin Hospital, Tokyo, Japan. [11]Department of Pathology and Laboratory Medicine, National Center Hospital, National Center of Neurology and Psychiatry, Tokyo, Japan. [12]Department of Neurology, Saitama Medical University, Saitama, Japan. [13]Department of Neurology, Jichi Medical University, Saitama Medical Center, Saitama, Japan. [14]Department of Geriatric Medicine, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan. [15]Department of Neurology, National Hospital Organization Higashisaitama Hospital, Saitama, Japan. [16]Department of Neurology, Ibaraki Prefectural University of Health Sciences, Ami, Japan. [17]Division of Neurology, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia. [18]Department of Neurology, International University of Health and Welfare Mita Hospital, Tokyo, Japan. [19]Department of Neuromuscular Research, National Institute of Neuroscience, National Center of Neurology and Psychiatry, Tokyo, Japan. [20]Present address: School of Bioscience and Biotechnology, Tokyo University of Technology, Tokyo, Japan. [21]Present address: Department of Integrative Genomics, Tohoku Medical Megabank Organization, Tohoku University, Sendai, Japan. [22]Present address: Department of Rehabilitation Science, Kobe University Graduate School of Health Sciences, Kobe, Japan. [23]Present address: Department of Cell Physiology, Kyorin University School of Medicine, Tokyo, Japan. *e-mail: tsuji@m.u-tokyo.ac.jp

a



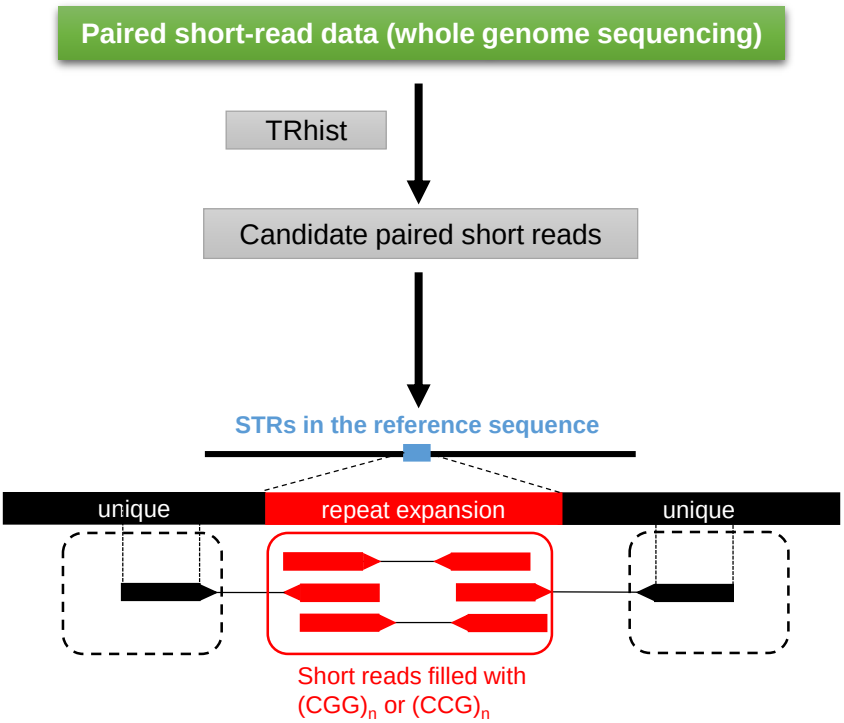
b

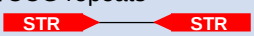
Gene/region	Chromo-somal band	Strand	Start	End	Qscore	Identity
NBPF19 (*NOTCH2NLC*)	1q21.2	+	149,370,802	149,410,843	40042	100.0%
NBPF14 (*NOTCH2NLB*)	1q21.2	-	148,659,853	148,700,939	39249	99.5%
NOTCH2NL (*NOTCH2NLA*)	1q21.1	-	146,209,112	146,250,193	39246	99.5%
NOTCH2	1p11.2	-	120,049,738	120,089,484	39018	99.2%
AC253572.1 (*NOTCH2NLR*)	1p11.2	+	120,705,588	120,743,847	37527	99.2%

Supplementary Fig. 1 Homologous regions around the CGG repeats in *NBPF19*.

(a) Schematic representation of the four highly homologous genes (*AC237572.1*, *NOTCH2*, *NOTCH2NL*, and *NBPF14*) and *NBPF19* are shown. Physical positions in hg38 are indicated. The five genes are located in the pericentric region of chromosome 1. The centromere and a long heterochromatin (1q12) exist between them. Parts of *NBPF19*, *NBPF14*, *NOTCH2NL*, and *AC253572.1* have also been recently annotated as *NOTCH2NLC*, *NOTCH2NLB*, *NOTCH2NLA*, and *NOTCH2NLR*, respectively [Fiddes, I.T. et al. *Cell* **173**, 1356–1369.e22 (2018) and Suzuki, I.K. et al. *Cell* **173**, 1370–1384 (2018)].

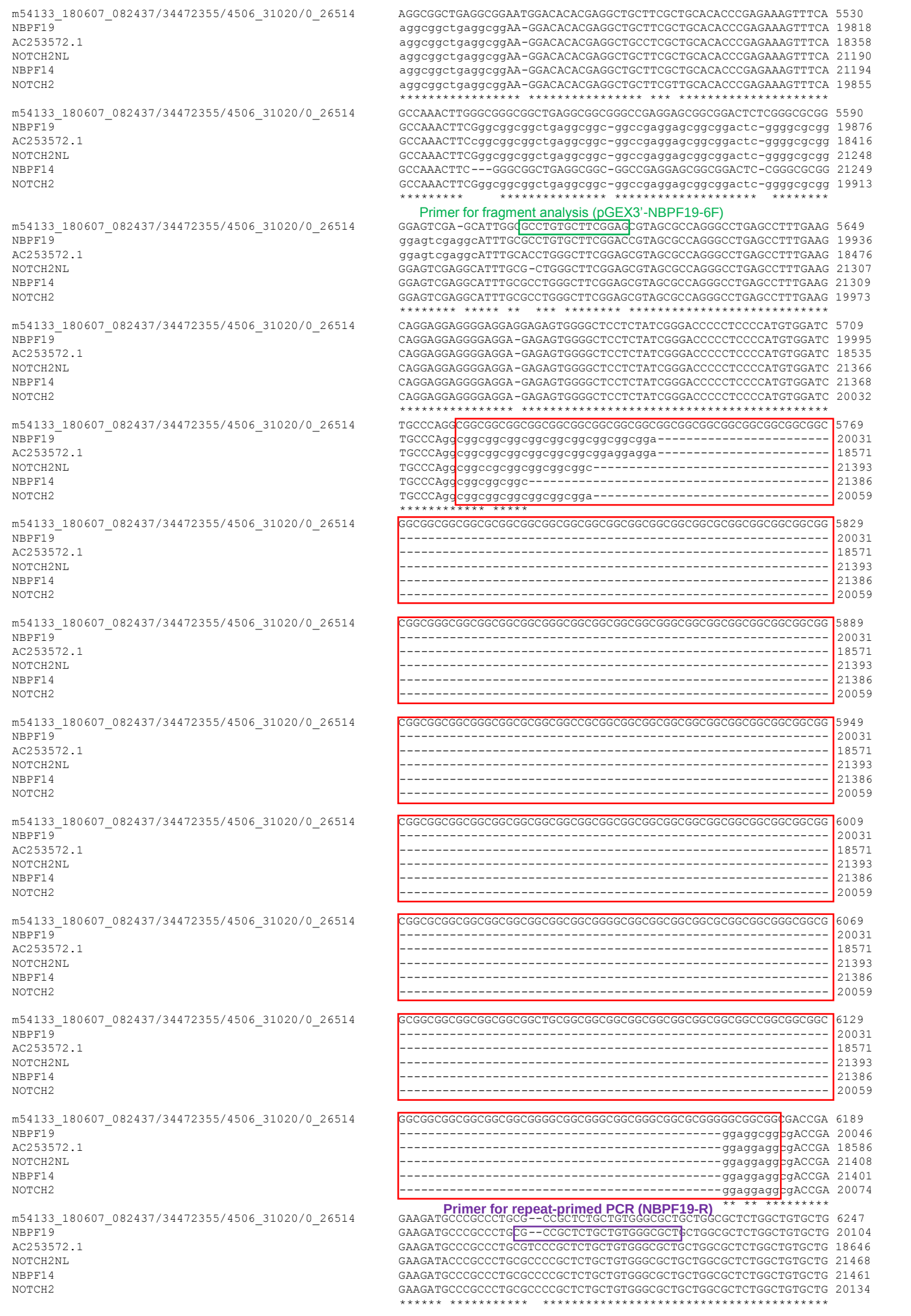
(b) To see sequences with high similarity in these regions, qscore and identity are calculated using BLAT [Kent, W.J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**:646–664 (2002)]. A portion of the *NBPF19* sequence (chr1:149,370,802-149,410,843 in hg38 that corresponds to 20 kb upstream and 20kb downstream of the CGG repeats in 5' UTR of *NBPF19*) is used as a query. Identities of 99.2%–99.5% are indicated.



Patient	F9193 proband (II-6)	F8504 proband	F9468 proband	F9785 proband
Number of paired reads in which both reads were filled with CGG/CCG repeats 	8	0	7	0
Number of paired reads in which one of the paired reads did not contain tandem repeats (non-repeat reads) 	16	7	14	7
Number of short reads (nonrepeat reads) that strongly supported alignment to *NBPF19*	0	3	2	1
Number of short reads (nonrepeat reads) that were aligned not only to *NBPF19* but also to other paralogous genes	6	4	11	6
Number of short reads (nonrepeat reads) that were not aligned to *NBPF19* or the paralogous genes (*AC253572.1*, *NOTCH2*, *NOTCH2NL*, and *NBPF14*)	10 (9 were in *AFF3*)	0	1	0

Supplementary Fig. 2 Identification of location of CGG/CCG repeats in families with NIID.

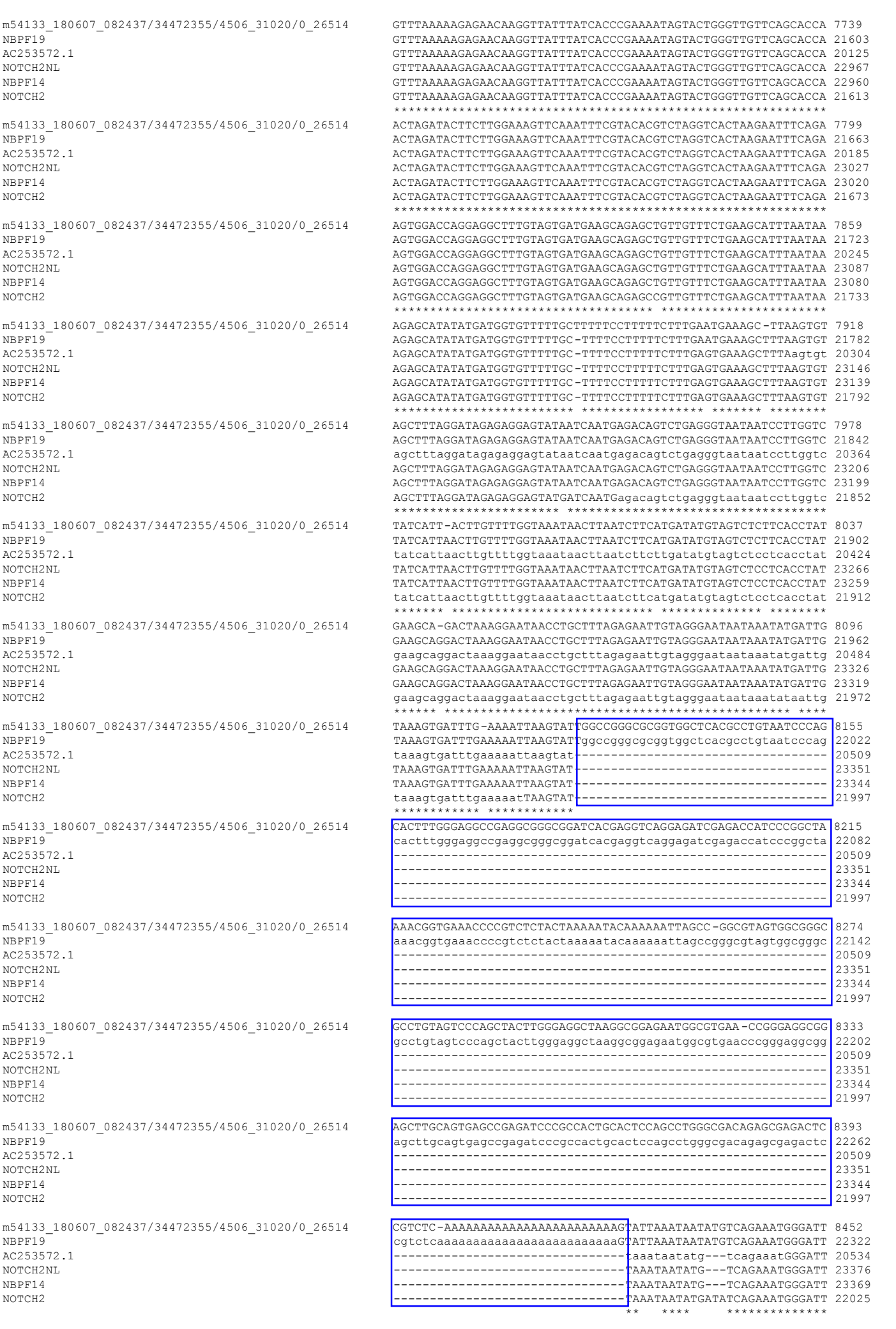
After short reads filled with CGG/CCG repeats were identified in four patients with NIID, reads paired with reads filled with CGG/CCG repeats were investigated. After trimming using quality score using sickle (version 1.33, https://github.com/najoshi/sickle), reads were visually investigated and mapped to hg38 using BLAT. In patients in F9193, F8504, F9468, and F9785, 6, 7, 13, and 7 reads were mapped to chromosome 1 (boxed with a blue line). In three patients, 3, 2, and 1 nonrepeat reads strongly supported the location of CGG/CCG repeats in *NBPF19* (boxed with a red line). In patient II-6 in F9193, another CGG/CCG repeat was suggested in *AFF3* at the fragile site FRA2A located outside the candidate region determined by linkage analysis (data not shown). STR, short tandem repeat.


Primer for fragment analysis (pGEX3'-NBPF19-6F)
Primer for repeat-primed PCR (NBPF19-R)

Primer for fragment analysis (pGEX3'-NBPF19-5R2)

```
m54133_180607_082437/34472355/4506_31020/0_26514      CGCGACCCCCGCGCATGGTGAGTATCGGGCTGAGGGGCGCTGTCCGCGGCGCCCCGGGGCT 6307
NBPF19                                                CGCGACCCCCGCGCATGGTGAGTATCGGGCTGAGGGGCGCTGTCCGCGGCGCCCCGGGGCT 20164
AC253572.1                                            GGCGGCCCCCGCGCATGGTGAGTATCGGGCTGAGGGGCGCTGTCCGCGGCGCCCCGGGGCT 18706
NOTCH2NL                                              CGCGACCCCCGCGCATGGTGAGTATCGGGCTGAGGGGCGCTGTCCGCGGCGCCCCGGGGCT 21528
NBPF14                                                CGCGACCCCCGCGCATGGTGAGTATCGGGCTGAGGGGCGCTGTCCGCGGCGCCCCGGGGCT 21521
NOTCH2                                                CGCGGCCCCCGCGCATGGTGAGTATCGGGCTGAGGGGCGCTGTCCGCGGCGCCCCGGGGCT 20194
                                                       *** ********************************************************

m54133_180607_082437/34472355/4506_31020/0_26514      GCCACCTGGGGCGACCCTTCTCCCCCTCGGTCCTTCTCTGTGTGGGAA-GCCAGGCTCGG 6366
NBPF19                                                GCCACCTGGGGCGACCCTTCTCCCCCTCGGTCCTTCTCTGTGTGGGAAGGCCAGGCTCGG 20224
AC253572.1                                            GCCACCTGGGGCGACCCTTCTCCCCCTCAGTCCTTCTCTGTGTGGGAAGGCCAGGCTCGG 18766
NOTCH2NL                                              GCCACCTGGGGCGACCCTTCTCCCCCTCGGTCCTTCTCTGTGTGGGAAGGCCAGGCTCGG 21588
NBPF14                                                GCCACCTGGGGCGACCCTTCTCCCCCTCGGTCCTTCTCTGTGTGGGAAGGCCAGGCTCGG 21581
NOTCH2                                                GCCACCTGGGGCGACCCTTCTCCCCCTCGGTCCTTCTCTGTGTGGGAAGGCCAGGCTCGG 20254
                                                      ************************* ******************** ***********

m54133_180607_082437/34472355/4506_31020/0_26514      CCGCCGGCGCGGAGCGAGGCCACTCGCTGGGTTCCCGAGAGTTTGGACATCGCCGGGGG- 6425
NBPF19                                                CCGCCGGCGCGGAGCGAGGCCACTCGCTGGGTTCCCGAGAGTTTGGACATCGCCGGGGGC 20284
AC253572.1                                            CCGCCGGCGCGGAGTGAGGCCACTCGCTGGGTTCCCAAGAGTTTGGACATCGCCGGGGGC 18826
NOTCH2NL                                              CCGCCGGCGCGGAGCGAGGCCACTCGCTGGGTTCCCGAGAGTTTGGACATCGCCGGGGGC 21648
NBPF14                                                CCGCCGGCGCGGAGCGAGGCCACTCGCTGGGTTCCCGAGAGTTTGGACATCGCCGGGGGC 21641
NOTCH2                                                CCGCCGGCGCGGAGCGAGGCCACTCGCTGGGTTCCCAAGAGTTTGGACATCGCCGGGGGC 20314
                                                      ************** ********************** ***********************

m54133_180607_082437/34472355/4506_31020/0_26514      CCCTCCCGTGGTGCCCCGCCAACCGCCGGGGTTCCCCGCCAACCGCCGGGGTTCCCCGCC 6485
NBPF19                                                CCCTCCCGTGGTGCCCCGCCAACCGCCGGGGTTCCCCGCCAACCGCCGGGGTTCCCCGCC 20344
AC253572.1                                            CCCTCCCGTGGTG--------------------CCCCGCCAACCGCTGGGGTTCCCCGCC 18866
NOTCH2NL                                              CCCTCCCGTGGTGCCCCGCCAACCGCCGGGGTTCCCCGCCAACCGCCGGGGTTCCCCGCC 21708
NBPF14                                                CCCTCCCGTGGTGCCCCGCCAACCGCCGGGGTTCCCCGCCAACCGCCGGGGTTCCCCGCC 21701
NOTCH2                                                CCCTCCCGTGGTG--------------------CCCCGCCAACCGCCGGGGTTCCCCGCC 20354
                                                      *************                     ************* *************

m54133_180607_082437/34472355/4506_31020/0_26514      GCCTCTGCTCCCCGCGGCCCGGGACCCCTCACACGCCTCCTCGGCAGGAGGGAGGCCGGC 6545
NBPF19                                                GCCTCTGCTCCCCGCGGCCCGGGACCCCTCACACGCCTCCTCGGCAGGAGGGAGGCCGGC 20404
AC253572.1                                            GCCTCTGCTCCCCGCGGCCCGGGACCCCTCACACGCCTCCTCGGCAGGAGGGAGGCCGGC 18926
NOTCH2NL                                              GCCTCTGCTCCCCGCGGCCCGGGACCCCTCACACGCCTCCTCGGCAGGAGGGAGGCCGGC 21768
NBPF14                                                GCCTCTGCTCCCCGCGGCCCGGGACCCCTCACACGCCTCCTCGGCAGGAGGGAGGCCGGC 21761
NOTCH2                                                GCCTCTGCTCCCCGCGGCCCGGGACCCCTTACACGCCTCCTCGGCAGGAGGGAGGCCGGC 20414
                                                      **************************** ****************************

m54133_180607_082437/34472355/4506_31020/0_26514      AGCAAGTCTCAGAAACTCCTTTTTCGTAGTGCCAGGGTGCAGGGAGGTGGGCAGTTTTGC 6605
NBPF19                                                AGCAAGTCTCAGAAACTCCTTTTTCGTAGTGCCAGGGTGCAGGGAGGTGGGCAGTTTTGC 20464
AC253572.1                                            AGCAAGTCTCAGAAACTCCTTTTTCGTAGTGCCAGGGTGCAGGGAGGTGGGCAGTTTTGC 18986
NOTCH2NL                                              AGCAAGTCTCAGAAACTCCTTTTTCGTAGTGCCAGGGTGCAGGGAGGTGGGCAGTTTTGC 21828
NBPF14                                                AGCAAGTCTCAGAAACTCCTTTTTCGTAGTGCCAGGGTGCAGGGAGGTGGGCAGTTTTGC 21821
NOTCH2                                                AGCAAGTCTCAGAAACTCCTTTTTCGTAGTGCCAGGGTGCAGGGAGGTGGGCAGTTTTGC 20474
                                                      ************************************************************

m54133_180607_082437/34472355/4506_31020/0_26514      CCTTCAGGTTCCGCGTTTCTTGGGGTCGAGCGAGAGCCGACGGCGGGCCTCGGAGGGGCT 6665
NBPF19                                                CCTTCAGGTTCCGCGTTTCTTGGGGTCGAGCGAGAGCCGACGGCGGGCCTCGGAGGGGCT 20524
AC253572.1                                            CCTTCAGGTTCCGCGTTTCTTGGGGTCGAGCGAGAGCCGACGGCGGGCCTCGGAGGGGCT 19046
NOTCH2NL                                              CCTTCAGGTTCCGCGTTTCCTGGGGTCGAGCGAGAGCCGACGGCGGGCCTCGGAGGGGCT 21888
NBPF14                                                CCTTCAGGTTCCGCGTTTCTTGGGGTCGAGCGAGAGCCGACGGCGGGCCTCGGAGGGGCT 21881
NOTCH2                                                CCTTCAGGTTCCGCGTTTCTTGGGGTCGAGCGAGAGCCGACGGCGGGCCTCGGAGGGGCT 20534
                                                      ******************** ***************************************

m54133_180607_082437/34472355/4506_31020/0_26514      GAGCGAAGGCGTGCCAGATTCTGGCGTGGAGAGCGGGGGCAGGGCCGCCAAGCCAAACGG 6725
NBPF19                                                GAGCGAAGGCGTGCCAGATTCTGGCGTGGAGAGCGGGGGCAGGGCCGCCAAGCCAAACGG 20584
AC253572.1                                            GAGCGAAGGAATGCCAGATTCTGGCGTGGAGAGCGGGGGCAGGGCCGCCAAGCCAAACGG 19106
NOTCH2NL                                              GAGCGAAGGAATGCCAGATTCTGGCGTGGAGAGCGGGGGCAGGGCCGCCAAGCCAAACGG 21948
NBPF14                                                GAGCGAAGGCGTGCCAGATTCTGGCGTGGAGAGCGGGGGCAGGGCCGCCAAGCCAAACGG 21941
NOTCH2                                                GAGCGAAGGAATGCCAGATTCTGGCGTGGAGAGCGGGGGCAGGGCCGCCAAGCCAAACGG 20594
                                                      ********* ************************************************

m54133_180607_082437/34472355/4506_31020/0_26514      CCTGCACCTTCGCAGCC-GCCTCGCCTTTGCCAGGGGGCGGCACATGGGCCGGGTGTGTG 6784
NBPF19                                                CCTGCACCTTCGCAGCCAGCCTCGCCTTTGCCAGGGGGCGGCACATGGGCCGGGTGTGTG 20644
AC253572.1                                            CCTGCAGCTTCGCAGCCAGCCTCGCCTTTGCCAGGGGGCGGCACATGGGCCGGGTGTGTG 19166
NOTCH2NL                                              CCTGCACCTTCGCAGCCAGCCTCGCCTTTGCCAGGGGGCGGCACATGGGCCGGGTGTGTG 22008
NBPF14                                                CCTGCACCTTCGCAGCCAGCCTCGCCTTTGCCAGGGGGCGGCACATGGGCCGGGTGTGTG 22001
NOTCH2                                                CCTGCACCGTCGCAGCCAGCCTCGCCTTTGCCAGGGGGCGGCACATGGGCCGGGTGTGTG 20654
                                                      ****** * ******** ******************************************

m54133_180607_082437/34472355/4506_31020/0_26514      GGCTTGGTTTGGATGGGGACGGGGTTTTGCGGCGCGCCTGAGTTTTGACACTCCAACCCC 6844
NBPF19                                                GGCTTGGTTTGGATGGGGACGGGGTTTTGCGGCGCGCCTGAGTTTTGACACTCCAACCCC 20704
AC253572.1                                            GGCTTGGTTTGGATGGGGACGGGGTTTTGCGGCGCGCCTGAGTTTTGACACTCCAACCCC 19226
NOTCH2NL                                              GGCTTGGTTTGGATGGGGACGGGGTTTTGCGGCGCGCCTGAGTTTTGACACTCCAACCCC 22068
NBPF14                                                GGCTTGGTTTGGATGGGGACGGGGTTTTGCGGCGCGCCTGAGTTTTGACACTCCAACCCC 22061
NOTCH2                                                GGCTTGGTTTGGATGGGGACGGGGTTTTGCGGCGCGCCTGAGTTTTGACACTCCAACCCC 20714
                                                      ************************************************************

m54133_180607_082437/34472355/4506_31020/0_26514      ACCGAAAGTCCGGAGGAGCCGGGTGTGCTGCTCGCGTCTTTGAAAGGTGGAGGCAGGAGA 6904
NBPF19                                                ACCGAAAGTCCGGAGGAGCCGGGTGTGCTGCTCGCGTCTTTGAAAGGTGGAGGCAGGAGA 20764
AC253572.1                                            ACCGAAAGTCCGGGGGAGCCGTGTGTGCTGCTCGCGTCTTTGAAAGGTGGAGGCAGGAGA 19286
NOTCH2NL                                              ACCGAAAGTCCGGAGGAGCCGGGTGTGCTGCTCGCGTCTTTGAAAGGTGGAGGCAGGAGA 22128
NBPF14                                                ACCGAAAGTCCGGAGGAGCCGGGTGTGCTGCTCGCGTCTTTGAAAGGTGGAGGCAGGAGA 22121
NOTCH2                                                ACCGAAAGTCCGGGGGAGCCGTGTGTGCTGCTCGCGTCTTTGAAAGGTGGAGGCAGGAGA 20774
                                                      ************* ******* ************************************

m54133_180607_082437/34472355/4506_31020/0_26514      AGTAGGGCAACTGGTGTGGCTGCATGCTGAGGCACATGATTT-AAAATCTCAACTGCTGT 6963
NBPF19                                                AGTAGGGCAACTGGTGTGGCTGCATGCTGAGGCACATGATTTTAAAAATCTCAACTGCTGT 20824
AC253572.1                                            AGTAGGGCAACTGGTGTGGCTGCATGCTGAGGCACATGATTTAAAAATCTCAACTGCTGT 19346
NOTCH2NL                                              AGTAGGGCAACTGGTGTGGCTGCATGCTGAGGCACATGATTTAAAAATCTCAACTGCTGT 22188
NBPF14                                                AGTAGGGCAACTGGTGTGGCTGCATGCTGAGGCACATGATTTAAAAATCTCAACTGCTGT 22181
NOTCH2                                                AGTAGGGCAACTGGTGTGGCTGCATGCTGAGGCACATGATTTAAAAATCTCAACTGCTGT 20834
                                                      ****************************************** *****************
```

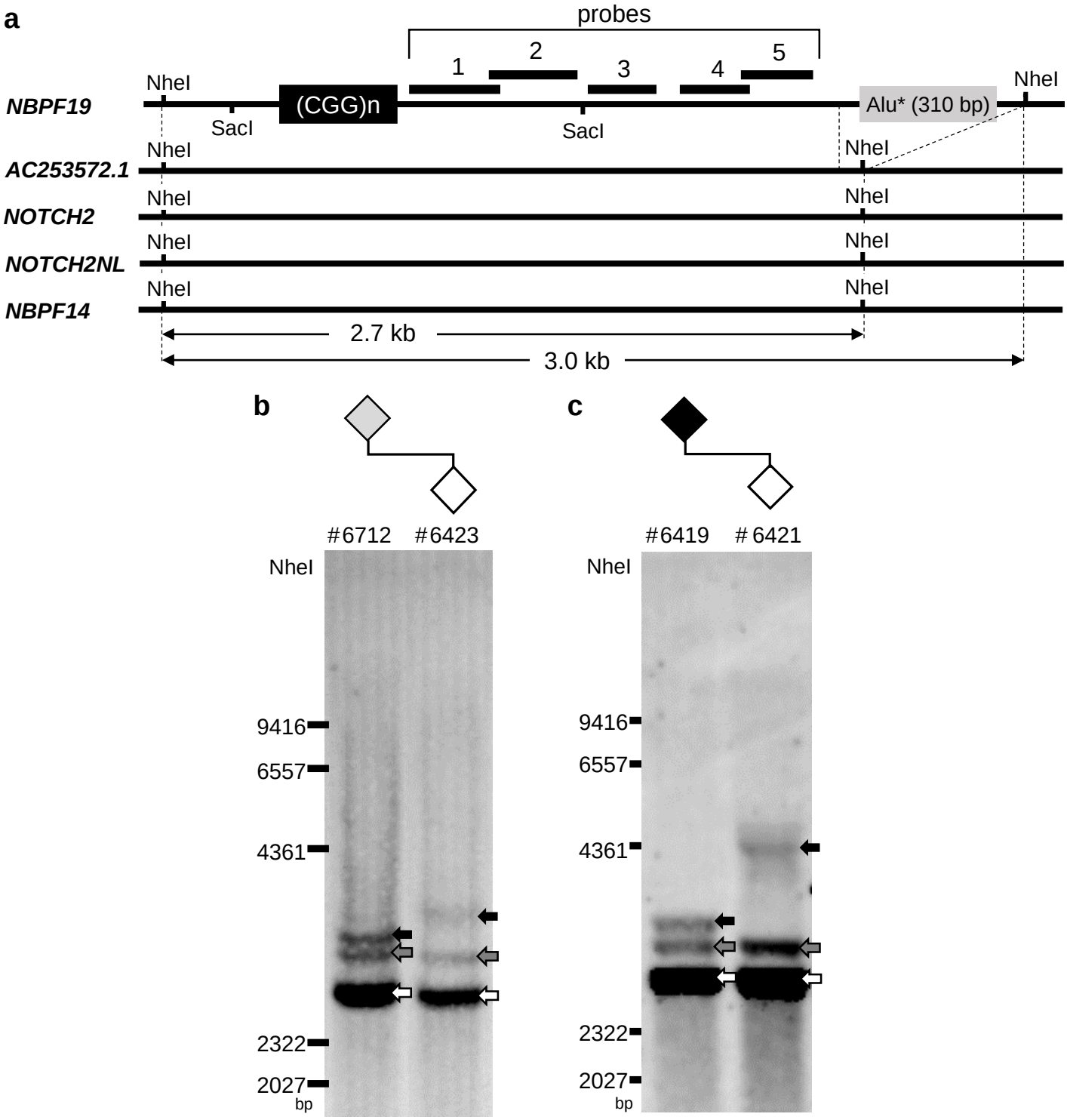
```
m54133_180607_082437/34472355/4506_31020/0_26514      TATTCTTTCCGAGGCGC-GAGCTCTGCTGCTGTTTCAGGCTGTGTCCAGACCCAGGAATG 7022
NBPF19                                                TATTCTTTCCGAGGCGCGGAGCTCTGCTGCTGTTTCAGGCTGTGTCCAGACCCAGGAATG 20884
AC253572.1                                            TATTCTTTCCGAGGCGCGGAGCTCTGCTGCTGTTTCAGGCTGTGTCCAGACCCAGGAATG 19406
NOTCH2NL                                              TATTCTTTCTGAGGCGCGGAGCTCTGCTGCTGTTTCAGGCTGTGTCCAGACCCAGGAATG 22248
NBPF14                                                TATTCTTTCTGAGGCGCGGAGCTCTGCTGCTGTTTCAGGCTGTGTCCAGACCCAGGAATG 22241
NOTCH2                                                TATTCTTTCCGAGGCGCGGAGCTCTGCTGCTGTTTCAGGCTGTGTCCAGACCCAGGAATG 20894
                                                      ********* ******* ****************************************

m54133_180607_082437/34472355/4506_31020/0_26514      TGGTGTGACGATCACCAATTCCTCCAACCTGGCAGCAGCATTTGCTGCTCCTTTTGGCAT 7082
NBPF19                                                TGGTGTGACGATCACCAATTCCTCCAACCTGGCAGCAGCATTTGCTGCTCCTTTGGCATG 20944
AC253572.1                                            TGGTGTGACGATCACCAATTCCTCCAACCTGGCAGCAGCATTTGCTGCTCCTTTGGCATG 19466
NOTCH2NL                                              TGGTGTGACGATCACCAATTCCTCCAACCTGGCAGCAGCATTTGCTGCTCCTTTGGCATG 22308
NBPF14                                                TGGTGTGACGATCACCAATTCCTCCAACCTGGCAGCAGCATTTGCTGCTCCTTTGGCATG 22301
NOTCH2                                                TGGTGTGACGATCACCAATTCCTCCAACCTGGCAGCAGCATTTGCTGCTCCTTTGGCATG 20954
                                                      ********************************************************* *

m54133_180607_082437/34472355/4506_31020/0_26514      GGCTGGGGGTGGGCACGGGCGGGTGAGAAAAAGTGGATACGTCAATTCAAAGGGCTTCCT 7142
NBPF19                                                GCTGGGGGTGGGGCACGGGCGGGTGAGAAAAAGTGGATACGTCAATTCAAAGGGCTTCCT 21004
AC253572.1                                            GCTGGGGGTGGGGCACGGGCGGGTGAGAAAAAGTGGATACGTTAATTCAAAGGGCTTCCT 19526
NOTCH2NL                                              GCTGGGGGTGGGGCACGGGCGGGTGAGAAAAAGTGGATACGTCAATTCAAAGGGCTTCCT 22368
NBPF14                                                GCTGGGGGTGGGGCACGGGCGGGTGAGAAAAAGTGGATACGTCAATTCAAAGGGCTTCCT 22361
NOTCH2                                                GCTGGGGGTGGGGCACGGGCGGGTGAGAAAAAGTGGATACGTTAATTCAAAGGGCTTCCT 21014
                                                      *   ****  ********************************** ****************

m54133_180607_082437/34472355/4506_31020/0_26514      TAGAAAGCTTCTTTATGGTTGGATGTTTCTAACACGGTTGGACCAAGGAAAGGGAATCAA 7202
NBPF19                                                TAGAAAGCTTCTTTATGGTTGGATGTTTCTAACACGGTTGGACCAAGGAAAGGGAATCAA 21064
AC253572.1                                            TAGAAAGCTTCTTTATGGTTGGATGTTTCTAACACGGTTGGACCAAGGAAAGGGAATCAA 19586
NOTCH2NL                                              TAGAAAGCTTCTTTGTGGTTGGATGTTTCTAACACGGTTGGACCAAGGAAAGGGAATCAA 22428
NBPF14                                                TAGAAAGCTTCTTTGTGGTTGGATGTTTCTAACACGGTTGGACCAAGGAAAGGGAATCAA 22421
NOTCH2                                                TAGAAAGCTTCTTTATGGTTGGATGTTTCTAACACGGTTGGACCAAGGAAAGGGAATCAA 21074
                                                      ************** ********************************************

m54133_180607_082437/34472355/4506_31020/0_26514      ATCATATCTTCCCCATCCCACCCACACTTTTAGATTCTACATTTCTTCAGACCCTTTAGT 7262
NBPF19                                                ATCATATCTTCCCCATCCCACCCACACTTTTAGATTCTACATTTCTTCAGACCCTTTagt 21124
AC253572.1                                            ATCATATCTTCCCCATCCCACCCACACTTTTAGATTCTACATTTCTTCAGACCCTTTagt 19646
NOTCH2NL                                              ATCATATCTTCCCCATCCCACCCACACTTTTAGATTCTACATTTCTTCAGACCCTTTagt 22488
NBPF14                                                ATCATATCTTCCCCATCCCACCCACACTTTTAGATTCTACATTTCTTCAGACCCTTTagt 22481
NOTCH2                                                ATCATATCTTCCCCATCCCACCCACACTTTTAGATTCTACATTTCTTCAGACCCTTTagt 21134
                                                      ************************************************************

m54133_180607_082437/34472355/4506_31020/0_26514      GGAAATAACTTGGGCTTTGGAGCCCTGCTCTGTAATCCTGGATTC-AATCCAGCTCTGCT 7321
NBPF19                                                ggaaataacttgggctttggagccctgctctgtaatcctggattcaaatccagctctgct 21184
AC253572.1                                            ggaaataacttgggctttggagccctgctctgtaatcctgtattcaaatccagctctgct 19706
NOTCH2NL                                              ggaaataacttgggctttggagccctgctctgtaatcctggattcaaatccagctctgct 22548
NBPF14                                                ggaaataacttgggctttggagccctgctctgtaatcctggattcaaatccagctctgct 22541
NOTCH2                                                ggaaataacttgggctttggagccctgctctgtaatcctggattcaaatccagctctgct 21194
                                                      ************************************** **** **************

m54133_180607_082437/34472355/4506_31020/0_26514      ATATGAACTCACCAATAAAATGGGAATGAAATTATCCAGCTAATAGGATTAATTAAAAT- 7380
NBPF19                                                aTATGAACTCACCAATAAAATGGGAATGAAATTATCCAGCTAATAGGATTAATTAAAATA 21244
AC253572.1                                            aTATGAACTCACCAATAAAATGGGAAAGAAATTATCCAGCTAATAGGATTAATTAAAATA 19766
NOTCH2NL                                              aTATGAACTCACCAATAAAATGGGAATGAAATTATCCAGCTAATAGGATTAATTAAAATA 22608
NBPF14                                                aTATGAACTCACCAATAAAATGGGAATGAAATTATCCAGCTAATAGGATTAATTAAAATA 22601
NOTCH2                                                aTATGAACTCACCAATAAAATGGGAAAGAAATTATCCAGCTAATAGGATTAATTAAAATA 21254
                                                      ************************** ******************************

m54133_180607_082437/34472355/4506_31020/0_26514      ATGTATATTAAACGGATGACACTCCAAAGGTG-TCAGAAAGTCCTTTGAGGGCCGAGGCT 7439
NBPF19                                                ATGTATATTAAACGGATGACACTCCAAAGGTGTTCAGAAAGTCCTTTGAGGGCCGAGGCT 21304
AC253572.1                                            ATGTATATTAAACGGATGACACTCCAAAGGTGTTCAGAAAGTCCTTTCAGGGCCGAGGCT 19826
NOTCH2NL                                              ATGTATATTAAACGGATGACACTCCAAAGGTGTTCAGAAAGTCCTTTGAGGGCCGAGGCT 22668
NBPF14                                                ATGTATATTAAACGGATGACACTCCAAAGGTGTTCAGAAAGTCCTTTGAGGGCCGAGGCT 22661
NOTCH2                                                ATGTATATTAAACGGATGACACTCCAAAGGTGTTCAGAAAGTCCTTTCAGGGCCGAGGCT 21314
                                                      ******************************** ************** ************

m54133_180607_082437/34472355/4506_31020/0_26514      GTGTTTTACTGCTTCTGAGATTGATGCTCCCAGTGGGATGGTGTCTTGTGCATATTTGTT 7499
NBPF19                                                GTGTTTTACTGCTTCTGAGATTGATGCTCCCAGTGGGATGGTGTCTTGTGCATATTTGTT 21364
AC253572.1                                            GTGTTTTACTGCTTCTGAGATTGATGCTCCCAGTGGGATGCTGTCTTGTGCATATTTGTT 19886
NOTCH2NL                                              GTGTTTTACTGCTTCTGAGATTGATGCTCCCAGTGGGATGGTGTCTTGTGCATATTTGTT 22728
NBPF14                                                GTGTTTTACTGCTTCTGAGATTGATGCTCCCAGTGGGATGGTGTCTTGTGCATATTTGTT 22721
NOTCH2                                                GTGTTTTACTGCTTCTGAGATTGATGCTCCCAGTGGGATGCTGTCTTGTGCATATTTGTT 21374
                                                      ****************************************** ******************

m54133_180607_082437/34472355/4506_31020/0_26514      GAACGGAACGCAGCCCCTTCTTTTTCTGAGCCTAATCTGAATAAAGATTTTTACCTTCAC 7559
NBPF19                                                GAACGGAAAGCAGCCCCTTCTTTTTCTGAGCCTAATCTGAATAAAGATTTTTACCTTCAC 21424
AC253572.1                                            GAACGGAAAGCAGCCCCTTCTTTTTCTGAGCCTAATCTGAATAAAGATTTTTACCTTCAC 19946
NOTCH2NL                                              GAACGGAAAGCAGCCCCTTCTTTTTCTGAGCCTAATCTGAATAAAGATTTTTACCTTCAC 22788
NBPF14                                                GAACGGAAAGCAGCCCCTTCTTTTTCTGAGCCTAATCTGAATAAAGATTTTTACCTTCAC 22781
NOTCH2                                                GAACGGAAAGCAGCCCCTTCTTTTTCTGAGCCTAATCTGAATAAAGATTTTTACCTTCAC 21434
                                                      ******** ************************************************

m54133_180607_082437/34472355/4506_31020/0_26514      CAAAAAAACTAGTACATGAACTACATCTGTTTTCAAAGAAAAGGTTAGGAAGATGATGAG 7619
NBPF19                                                C-AAAAAACTAGTACATGAACTACATCTGTTTTCAAAGAAAAGGTTAGGAAGATGATGAG 21483
AC253572.1                                            C-AAAAAACTAGTACATGAACTACATCTGTTTTCAAAGAAAAGGTTAGGAAGATGATGAG 20005
NOTCH2NL                                              C-AAAAAACTAGTACATGAACTACATCTGTTTTCAAAGAAAAGGTTAGGAAGATGATGAG 22847
NBPF14                                                C-AAAAAACTAGTACATGAACTACATCTGTTTTCAAAGAAAAGGTTAGGAAGATGATGAG 22840
NOTCH2                                                C-AAAAAACTAGTACATGAACTACATCTGTTTTCAAAGAAAAGGTTAGGAAGATGATGAG 21493
                                                      * ************************************************************

m54133_180607_082437/34472355/4506_31020/0_26514      CTTCAGAAAAATATGGTCTTTGTTCATTGTTAACAGTCAGTCGACATGTCAATCGCAGAT 7679
NBPF19                                                CTTCAGAAAAATATGGTCTTTGTTCATTGTTAACAGTCAGTCGACATGTCAATCGCAGAT 21543
AC253572.1                                            CTTCAGAAAAATATGGTCTTTGTTCATTGTTAACAGTCAGTCGACATGTCAATCGCAGAT 20065
NOTCH2NL                                              CTTCAGAAAAATATGGTCTTTGTTCATTGTTAACAGTCAGTCGACATGTCAATCGCAGAT 22907
NBPF14                                                CTTCAGAAAAATATGGTCTTTGTTCATTGTTAACAGTCAGTCGACATGTCAATCGCAGAT 22900
NOTCH2                                                CTTCAGAAAAATATGGTCTTTGTTCATTGTTAACAGTCAGTCGACATGTCAATCGCAGAT 21553
                                                      ************************************************************
```


m54133_180607_082437/34472355/4506_31020/0_26514 GTTTAAAAAGAGAACAAGGTTATTTATCACCCGAAAATAGTACTGGGTTGTTCAGCACCA 7739
NBPF19 GTTTAAAAAGAGAACAAGGTTATTTATCACCCGAAAATAGTACTGGGTTGTTCAGCACCA 21603
AC253572.1 GTTTAAAAAGAGAACAAGGTTATTTATCACCCGAAAATAGTACTGGGTTGTTCAGCACCA 20125
NOTCH2NL GTTTAAAAAGAGAACAAGGTTATTTATCACCCGAAAATAGTACTGGGTTGTTCAGCACCA 22967
NBPF14 GTTTAAAAAGAGAACAAGGTTATTTATCACCCGAAAATAGTACTGGGTTGTTCAGCACCA 22960
NOTCH2 GTTTAAAAAGAGAACAAGGTTATTTATCACCCGAAAATAGTACTGGGTTGTTCAGCACCA 21613
m54133_180607_082437/34472355/4506_31020/0_26514 ACTAGATACTTCTTGGAAAGTTCAAATTTCGTACACGTCTAGGTCACTAAGAATTTCAGA 7799
NBPF19 ACTAGATACTTCTTGGAAAGTTCAAATTTCGTACACGTCTAGGTCACTAAGAATTTCAGA 21663
AC253572.1 ACTAGATACTTCTTGGAAAGTTCAAATTTCGTACACGTCTAGGTCACTAAGAATTTCAGA 20185
NOTCH2NL ACTAGATACTTCTTGGAAAGTTCAAATTTCGTACACGTCTAGGTCACTAAGAATTTCAGA 23027
NBPF14 ACTAGATACTTCTTGGAAAGTTCAAATTTCGTACACGTCTAGGTCACTAAGAATTTCAGA 23020
NOTCH2 ACTAGATACTTCTTGGAAAGTTCAAATTTCGTACACGTCTAGGTCACTAAGAATTTCAGA 21673
m54133_180607_082437/34472355/4506_31020/0_26514 AGTGGACCAGGAGGCTTTGTAGTGATGAAGCAGAGCTGTTGTTTCTGAAGCATTTAATAA 7859
NBPF19 AGTGGACCAGGAGGCTTTGTAGTGATGAAGCAGAGCTGTTGTTTCTGAAGCATTTAATAA 21723
AC253572.1 AGTGGACCAGGAGGCTTTGTAGTGATGAAGCAGAGCTGTTGTTTCTGAAGCATTTAATAA 20245
NOTCH2NL AGTGGACCAGGAGGCTTTGTAGTGATGAAGCAGAGCTGTTGTTTCTGAAGCATTTAATAA 23087
NBPF14 AGTGGACCAGGAGGCTTTGTAGTGATGAAGCAGAGCTGTTGTTTCTGAAGCATTTAATAA 23080
NOTCH2 AGTGGACCAGGAGGCTTTGTAGTGATGAAGCAGAGCCGTTGTTTCTGAAGCATTTAATAA 21733
m54133_180607_082437/34472355/4506_31020/0_26514 AGAGCATATATGATGGTGTTTTTGCTTTTTCCTTTTTCTTTGAATGAAAGC-TTAAGTGT 7918
NBPF19 AGAGCATATATGATGGTGTTTTTGC-TTTTCCTTTTTCTTTGAATGAAAGCTTTAAGTGT 21782
AC253572.1 AGAGCATATATGATGGTGTTTTTGC-TTTTCCTTTTTCTTTGAGTGAAAGCTTTAagtgt 20304
NOTCH2NL AGAGCATATATGATGGTGTTTTTGC-TTTTCCTTTTTCTTTGAGTGAAAGCTTTAAGTGT 23146
NBPF14 AGAGCATATATGATGGTGTTTTTGC-TTTTCCTTTTTCTTTGAGTGAAAGCTTTAAGTGT 23139
NOTCH2 AGAGCATATATGATGGTGTTTTTGC-TTTTCCTTTTTCTTTGAGTGAAAGCTTTAAGTGT 21792
m54133_180607_082437/34472355/4506_31020/0_26514 AGCTTTAGGATAGAGAGGAGTATAATCAATGAGACAGTCTGAGGGTAATAATCCTTGGTC 7978
NBPF19 AGCTTTAGGATAGAGAGGAGTATAATCAATGAGACAGTCTGAGGGTAATAATCCTTGGTC 21842
AC253572.1 agctttaggatagagaggagtataatcaatgagacagtctgagggtaataatccttggtc 20364
NOTCH2NL AGCTTTAGGATAGAGAGGAGTATAATCAATGAGACAGTCTGAGGGTAATAATCCTTGGTC 23206
NBPF14 AGCTTTAGGATAGAGAGGAGTATAATCAATGAGACAGTCTGAGGGTAATAATCCTTGGTC 23199
NOTCH2 AGCTTTAGGATAGAGAGGAGTATGATCAATGagacagtctgagggtaataatccttggtc 21852
m54133_180607_082437/34472355/4506_31020/0_26514 TATCATT-ACTTGTTTTGGTAAATAACTTAATCTTCATGATATGTAGTCTCTTCACCTAT 8037
NBPF19 TATCATTAACTTGTTTTGGTAAATAACTTAATCTTCATGATATGTAGTCTCTTCACCTAT 21902
AC253572.1 tatcattaacttgttttggtaaataacttaatcttcttgatatgtagtctcctcacctat 20424
NOTCH2NL TATCATTAACTTGTTTTGGTAAATAACTTAATCTTCATGATATGTAGTCTCCTCACCTAT 23266
NBPF14 TATCATTAACTTGTTTTGGTAAATAACTTAATCTTCATGATATGTAGTCTCCTCACCTAT 23259
NOTCH2 tatcattaacttgttttggtaaataacttaatcttcatgatatgtagtctcctcacctat 21912
m54133_180607_082437/34472355/4506_31020/0_26514 GAAGCA-GACTAAAGGAATAACCTGCTTTAGAGAATTGTAGGGAATAATAAATATGATTG 8096
NBPF19 GAAGCAGGACTAAAGGAATAACCTGCTTTAGAGAATTGTAGGGAATAATAAATATGATTG 21962
AC253572.1 gaagcaggactaaaggaataacctgctttagagaattgtagggaataataaatatgattg 20484
NOTCH2NL GAAGCAGGACTAAAGGAATAACCTGCTTTAGAGAATTGTAGGGAATAATAAATATGATTG 23326
NBPF14 GAAGCAGGACTAAAGGAATAACCTGCTTTAGAGAATTGTAGGGAATAATAAATATGATTG 23319
NOTCH2 gaagcaggactaaaggaataacctgctttagagaattgtagggaataataaatataattg 21972
m54133_180607_082437/34472355/4506_31020/0_26514 TAAAGTGATTTG-AAAATTAAGTATTGGCCGGGCGCGGTGGCTCACGCCTGTAATCCCAG 8155
NBPF19 TAAAGTGATTTGAAAAATTAAGTATTggccgggcgcggtggctcacgcctgtaatcccag 22022
AC253572.1 taaagtgatttgaaaaattaagtat---------- 20509
NOTCH2NL TAAAGTGATTTGAAAAATTAAGTAT---------- 23351
NBPF14 TAAAGTGATTTGAAAAATTAAGTAT---------- 23344
NOTCH2 taaagtgatttgaaaaatTAAGTAT---------- 21997
m54133_180607_082437/34472355/4506_31020/0_26514 CACTTTGGGAGGCCGAGGCGGGCGGATCACGAGGTCAGGAGATCGAGACCATCCCGGCTA 8215
NBPF19 cactttgggaggccgaggcgggcggatcacgaggtcaggagatcgagaccatcccggcta 22082
AC253572.1 ---------- 20509
NOTCH2NL ---------- 23351
NBPF14 ---------- 23344
NOTCH2 ---------- 21997
m54133_180607_082437/34472355/4506_31020/0_26514 AAACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCC-GGCGTAGTGGCGGGC 8274
NBPF19 aaacggtgaaaccccgtctctactaaaaatacaaaaaattagccgggcgtagtggcgggc 22142
AC253572.1 ---------- 20509
NOTCH2NL ---------- 23351
NBPF14 ---------- 23344
NOTCH2 ---------- 21997
m54133_180607_082437/34472355/4506_31020/0_26514 GCCTGTAGTCCCAGCTACTTGGGAGGCTAAGGCGGAGAATGGCGTGAA-CCGGGAGGCGG 8333
NBPF19 gcctgtagtcccagctacttgggaggctaaggcggagaatggcgtgaacccgggaggcgg 22202
AC253572.1 ---------- 20509
NOTCH2NL ---------- 23351
NBPF14 ---------- 23344
NOTCH2 ---------- 21997
m54133_180607_082437/34472355/4506_31020/0_26514 AGCTTGCAGTGAGCCGAGATCCCGCCACTGCACTCCAGCCTGGGCGACAGAGCGAGACTC 8393
NBPF19 agcttgcagtgagccgagatcccgccactgcactccagcctgggcgacagagcgagactc 22262
AC253572.1 ---------- 20509
NOTCH2NL ---------- 23351
NBPF14 ---------- 23344
NOTCH2 ---------- 21997
m54133_180607_082437/34472355/4506_31020/0_26514 CGTCTC-AAAAAAAAAAAAAAAAAAAAAAGTATTAAATAATATGTCAGAAATGGGATT 8452
NBPF19 cgtctcaaaaaaaaaaaaaaaaaaaaaaaGTATTAAATAATATGTCAGAAATGGGATT 22322
AC253572.1 ----------taaataatatg---tcagaaatGGGATT 20534
NOTCH2NL ----------TAAATAATATG---TCAGAAATGGGATT 23376
NBPF14 ----------TAAATAATATG---TCAGAAATGGGATT 23369
NOTCH2 ----------TAAATAATATGATATCAGAAATGGGATT 22025

```
m54133_180607_082437/34472355/4506_31020/0_26514      ACTCTGACACATAGGGCTTCTTTTTATATAGGATCTC-GTGCATAAAACTATGCCCTCAC 8511
NBPF19                                                ACTCTGACACATAGGGCTTCTTTTATTATAGGATCTCAGTGCATAAAACTATGCCCTCCA 22382
AC253572.1                                            ACTCTGACACATAGGGCTTCTTTTATTATAGGATCTCAGTGCATAAAACTATGCCCTCCA 20594
NOTCH2NL                                              ACTCTGACACATAGGGCTTCTTTTATTATAGGATCTCAGTGCGTAAAACTATGCCCTCCA 23436
NBPF14                                                ACTCTGACACATAGGGCTTCTTTTATTATAGGATCTCAGTGCGTAAAACTATGCCCTCCA 23429
NOTCH2                                                ACTCTGACACATAGGGCTTCTTTTATTATAGGATCTCAGTGCATAAAACTATGCCCTCCA 22085
                                                      ********************  *********** **** ************** 

m54133_180607_082437/34472355/4506_31020/0_26514      CTCTGCTCTGCTCTGCTGAGGTGCTGCGATTATTAGCTAGCTTTTAGGTAGTCAAGTTGA 8571
NBPF19                                                C----CTCTGCTCTGCTGAGGTGCTGCGATTATTAGCTAGCTTTTAGGTAGTCAAGTTGA 22438
AC253572.1                                            C----CTCTGCTCTGCTGAGGTGCTGCGATTATTAGCTAGCTTTTAGGTAGTCAAGTTGA 20650
NOTCH2NL                                              C----CTCTGCTCTGCTGAGGTGCTGCGATTATTAGCTAGCTTTTAGGTAGTCAAGTTGA 23492
NBPF14                                                C----CTCTGCTCTGCTGAGGTGCTGCGATTATTAGCTAGCTTTTAGGTAGTCAAGTTGA 23485
NOTCH2                                                C----CTCTGCTCTGCTGAGGTGCTGCGATTATTAGCTAGCTTTTAGGTAGTCAAGTTGA 22141
                                                      *    ************************************************** 

m54133_180607_082437/34472355/4506_31020/0_26514      GTGTTAGTCAAGTTGACTCAAAGGATAGGATTTAGATATTTTGATTTTTAGACACGGGTT 8631
NBPF19                                                GTGTTAGTCAAGTTGACTCAAAGGATAGGATTTAGATATTTTGATTTTTAGACAGGTGTT 22498
AC253572.1                                            GTGTTAGTCAAGTTGACTCAAAGGATAGGATTTAGATATTTTGATTTTTAGACAGGTGTT 20710
NOTCH2NL                                              GTGTTAGTCAAGTTGACTCAAAGGATAGGATTTAGATATTTTGATTTTTAGACAGGTGTT 23552
NBPF14                                                GTGTTAGTCAAGTTGACTCAAAGGATAGGATTTAGATATTTTGATTTTTAGACAGGTGTT 23545
NOTCH2                                                GTGTTAGTCAAGTTGACTCAAAGGATAGGATTTAGATATTTTGATTTTTAGACAGGTGTT 22201
                                                      ********************************************************* * *** 

m54133_180607_082437/34472355/4506_31020/0_26514      C-TTAGATGCTAAACTAATGCAGTAGAAGAGATTTAGAAAGCCTGTTTCTTGGTACAAAT 8690
NBPF19                                                CTTTAGATGCTAAACTAATGCAGTAGAAGAGATTTAG-AAGCCTGTTTCTTGGTACAAAT 22557
AC253572.1                                            CTTTAGATGCTAAACTAATGCAGTAGAAGAGATTTAG-AAGCC-GTTTCTTGGTACAAAT 20768
NOTCH2NL                                              CTTTAGATGCTAAACTAATGCAGTAGAAGAGATTTAG-AAGCCTGTTTCTTGGTACAAAT 23611
NBPF14                                                CTTTAGATGCTAAACTAATGCAGTAGAAGAGATTTAG-AAGCCTGTTTCTTGGTACAAAT 23604
NOTCH2                                                CTTTAGATGCTAAACTAATGCAGTAGAAGAGATTTAG-AAGCC-GTTTCTCGGTACAAAT 22259
                                                      * ********************************** ***** ****** ********* 

m54133_180607_082437/34472355/4506_31020/0_26514      CTTAAATCTTCTTAGAGTCTACAGTGAAGGCTTCTTGAATCCTCTTAATTTATTA-TGTG 8749
NBPF19                                                CTTAAATCTTCTTAGAGTCTACAGTGAAGGCTTCTTGAATCCTCTTAATTTATTATTGTG 22617
AC253572.1                                            CTTAAATCTTCTTAGAGTCTACAGTGAAGGCTTCTTGAATCCTCCTAATTTATTATTGTG 20828
NOTCH2NL                                              CTTAAATCTTCTTAGAGTCTACAGTGAAGGCTTCTTGAATCCTCCTAATTTATTATTGTG 23671
NBPF14                                                CTTAAATCTTCTTAGAGTCTACAGTGAAGGCTTCTTGAATCCTCCTAATTTATTATTGTG 23664
NOTCH2                                                CTTAAATCTTCTTAGAGTCTACAGTGAAGGCTTCTTGAATCCTCTTAATTTATTATTGTG 22319
                                                      ****************************************** ********** **** 

m54133_180607_082437/34472355/4506_31020/0_26514      TTTTGCTAAGGTACCAATTACAGCACACCCTTCTAAACTTTTGAGTGCTCCTGTGTTTGG 8809
NBPF19                                                TTTTGCTAAGGTACCAATTACAGCACACCCTTCTAAACTTTTGAGTGCTCCTGTGTTTGG 22677
AC253572.1                                            TTTTGCTAAGGTACCAATTACAGCACACCCTTCTAAACTTTTGAGTGTTCCTGTGTTTGG 20888
NOTCH2NL                                              TTTTGCTAAGGTACCAATTACAGCACACCCTTCTAAACTTTTGAGTGTTCCTGTGTTTGG 23731
NBPF14                                                TTTTGCTAAGGTACCAATTACAGCACACCCTTCTAAACTTTTGAGTGTTCCTGTGTTTGG 23724
NOTCH2                                                TTTTGCTAAGGTACCAGTTGCAGCACACCCTTCTAAACTTTTGAGTGTTCCTGTGTTTGG 22379
                                                      **************** ** ************************** ************ 
```

Supplementary Fig. 3 Multiple sequence alignment of a long read, *NBPF19*, *AC253572.1*, *NOTCH2NL*, *NBPF14*, and *NOTCH2*.

Multiple sequence alignment of a long-read sequence obtained by single-molecule, real-time sequencing, as well as the corresponding regions in *NBPF19*, *AC253572.1*, *NOTCH2NL*, *NBPF14*, and *NOTCH2* using Clustal W2 [Larkin, M. A., et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007)]. The five long reads spanning the CGG repeats in *NBPF19* were subjected to error-correction using Canu (version 1.7) [Koren, S., et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**, 722–736 (2017)] and then assembled using racon (version 1.3.1) [Vaser, R., et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746 (2017)]. CGG repeat expansions wereshown by red boxes. An *NBPF19*-specific insertion of *Alu* sequence was shown by blue boxes, which confirmed that the expanded CGG repeats were located in *NBPF19*. One of the primer sequence (NBPF19-R, **Supplementary Table 4**) for repeat-primed PCR analysis (shown by a purple box) and a primer pair (pGEX3'-NBPF19-6F and NBPF19-5R2, **Supplementary Table 7**) for fragment analysis (shown by green boxes) were designed to avoid nonspecific amplification.

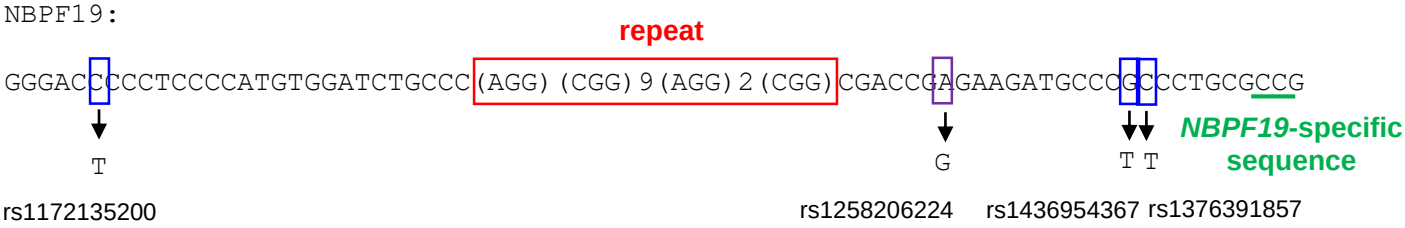


Supplementary Fig. 4 Intergenerational instability of the CGG repeats in *NBPF19*.

(a) SacI/NheI digestion sites around the CGG repeats in the 5' UTR of *NBPF19* are shown. An *Alu* sequence (starred) on the downstream of the CGG repeats is absent in the other 4 highly homologous genes (*AC253572.1*, *NOTCH2*, *NOTCH2NL*, and *NBPF14*). This enabled us to distinguish the *NBPF19* alleles from other highly homologous genes in Southern blot analysis using NheI-digested genomic DNA (gDNA). Restriction fragments generated from *NOTCH2*, *AC253572.1*, *NBPF14*, and *NOTCH2NL* are estimated to be 2,696 bp, 2,691 bp, 2,696 bp, and 2,707 bp, respectively, whereas that from *NBPF19* is estimated to be 3,009 bp based on hg38.

(b, c) Southern blot analysis of parent-offspring pairs in the branches of F6321 using NheI-digested gDNA, where we use probes 1–5 to enhance the signal intensity of target bands. White arrows indicate fragments derived from the 4 genes (*NOTCH2*, *AC253572.1*, *NBPF14*, and *NOTCH2NL*) that do not carry the *Alu* sequence designated by a star in **(a)** and gray arrows indicate wild type *NBPF19* alleles that carry the *Alu* sequence. Black arrows indicate *NBPF19* alleles with expanded CGG repeats. The results showed that the sizes of the CGG repeats in *NBPF19* become larger in the successive generations. The parent indicated by a gray symbol in **(b)** only showed abnormalities in the nerve conduction study. Each experiment was conducted once. The uncropped images are in **Supplementary Fig. 17**.

a



b

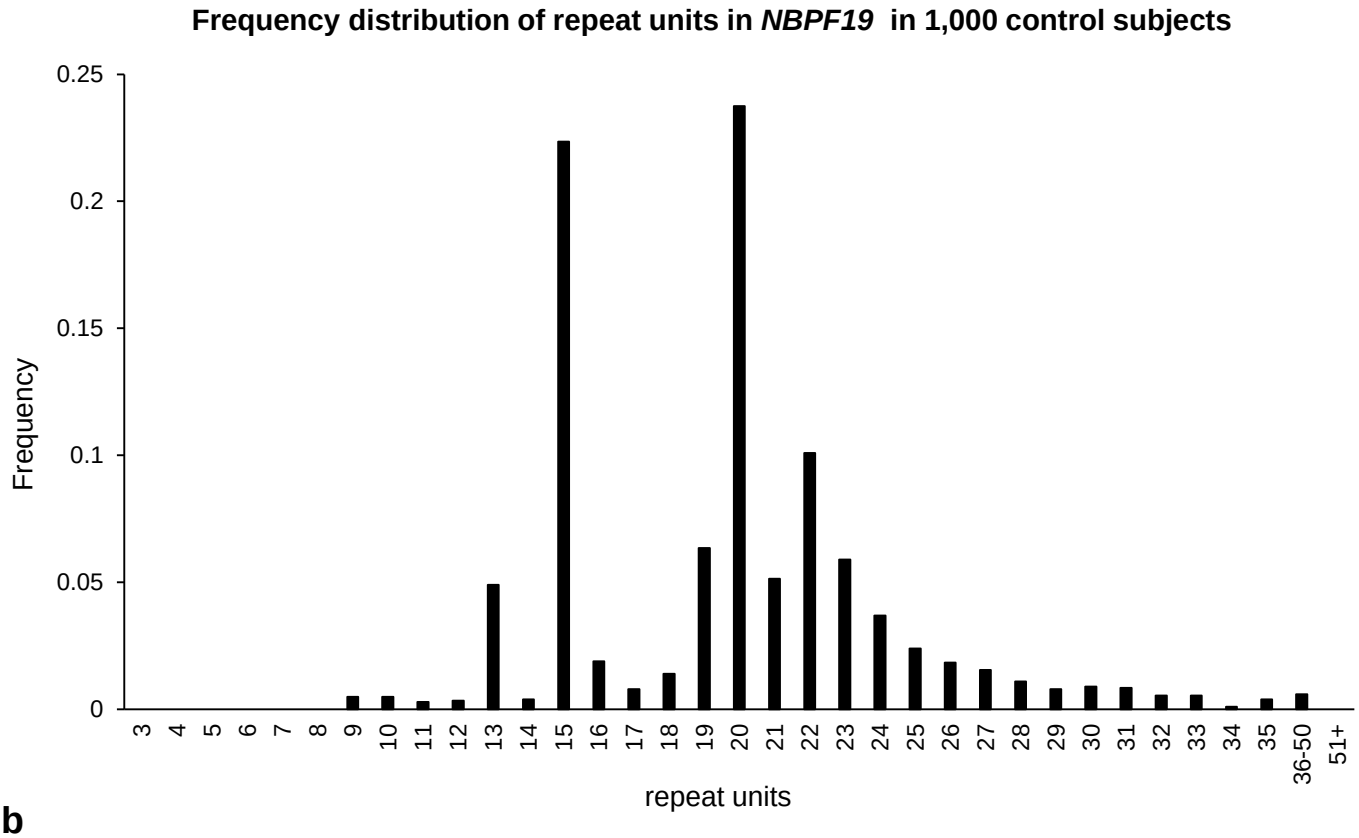
Repeat configuration	Presence of rs1172135200, rs1436954367, and rs1376391857 (3 SNVs)	Presence of rs1258206224	# of observed alleles
$(AGG)(CGG)_n(AGG)_2(CGG)$	No	No	239
$(AGG)(CGG)_n(AGG)_3$	No	No	85
$(AGG)(CGG)_n(AGG)(CGG)$	No	No	9
$(AGG)(CGG)_n$	No	No	6
$(AGG)(CGG)_n(AGG)_3(CGG)$	No	No	4
$(AGG)(CGG)_n(AGG)(CGG)_2$	No	No	1
$(AGG)(CGG)_n(AGG)_2$	No	No	1
$(AGG)(CGG)_n(AGG)_2(CGG)_3(AGG)_2(CGG)$	No	No	1
$(AGG)(CGG)_n(AGG)_2(CGG)_2(AGG)_2(CGG)$	No	No	1
$(AGG)(CGG)_9(AGG)_3$	Yes	No	14
$(AGG)(CGG)_n(AGG)_2(CGG)$	No	Yes	3

Supplementary Fig. 5 Repeat configurations of CGG and flanking repeats in *NBPF19* in control subjects as revealed by CCS analysis.

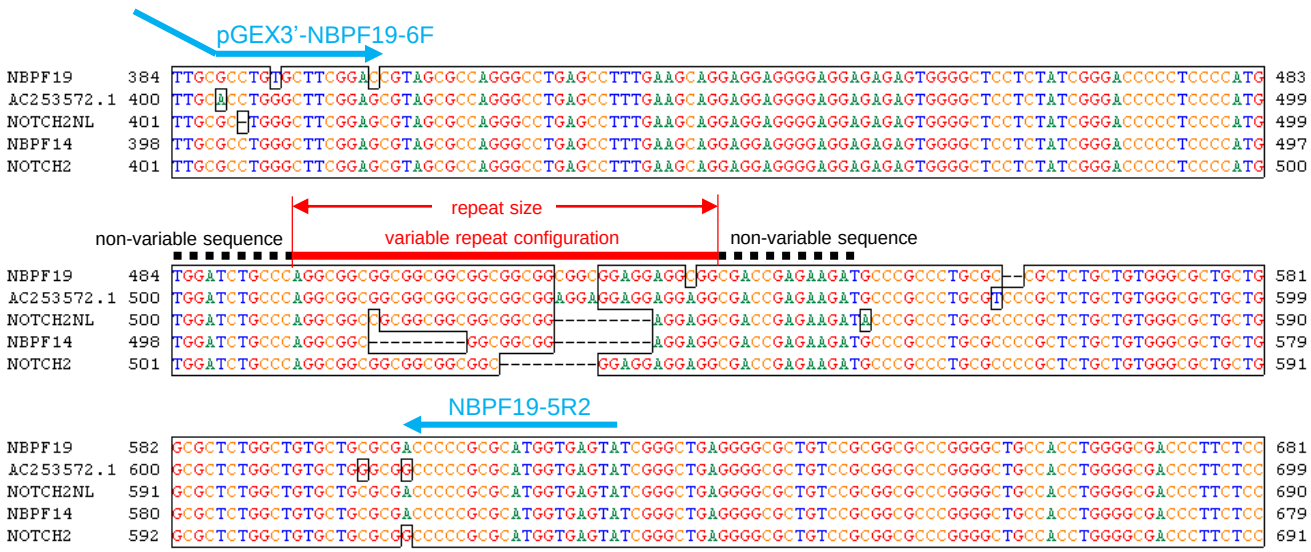
(a) The CGG and flanking repeats in the 5' UTR of *NBPF19* is $(AGG)(CGG)_9(AGG)_2(CGG)$ in the reference sequence (hg38). To determine the number of repeat units, repeat configurations and single nucleotide variants in the flanking sequences, circular consensus sequencing (CCS) analysis was performed for pooled barcoded PCR products from 182 control subjects. CCS reads were confirmed to have *NBPF19*-specific sequence shown by a green underline.

(b) We observed 11 repeat configurations and single nucleotide variants (SNVs) in the flanking sequences in *NBPF19*. One allele carrying three SNVs (rs1172135200, rs1436954367, and rs1376391857) in the flanking sequences, all of which carried a configuration $(AGG)(CGG)_9(AGG)_3$, and another allele carrying rs1258206224 with a configuration of $(AGG)(CGG)_n(AGG)_2(CGG)$ were observed in 14 and 3 controls, respectively. On the basis of these observations, distribution of number of the CGG repeat unit (shown by "n") was determined (**Fig. 4f**).

a



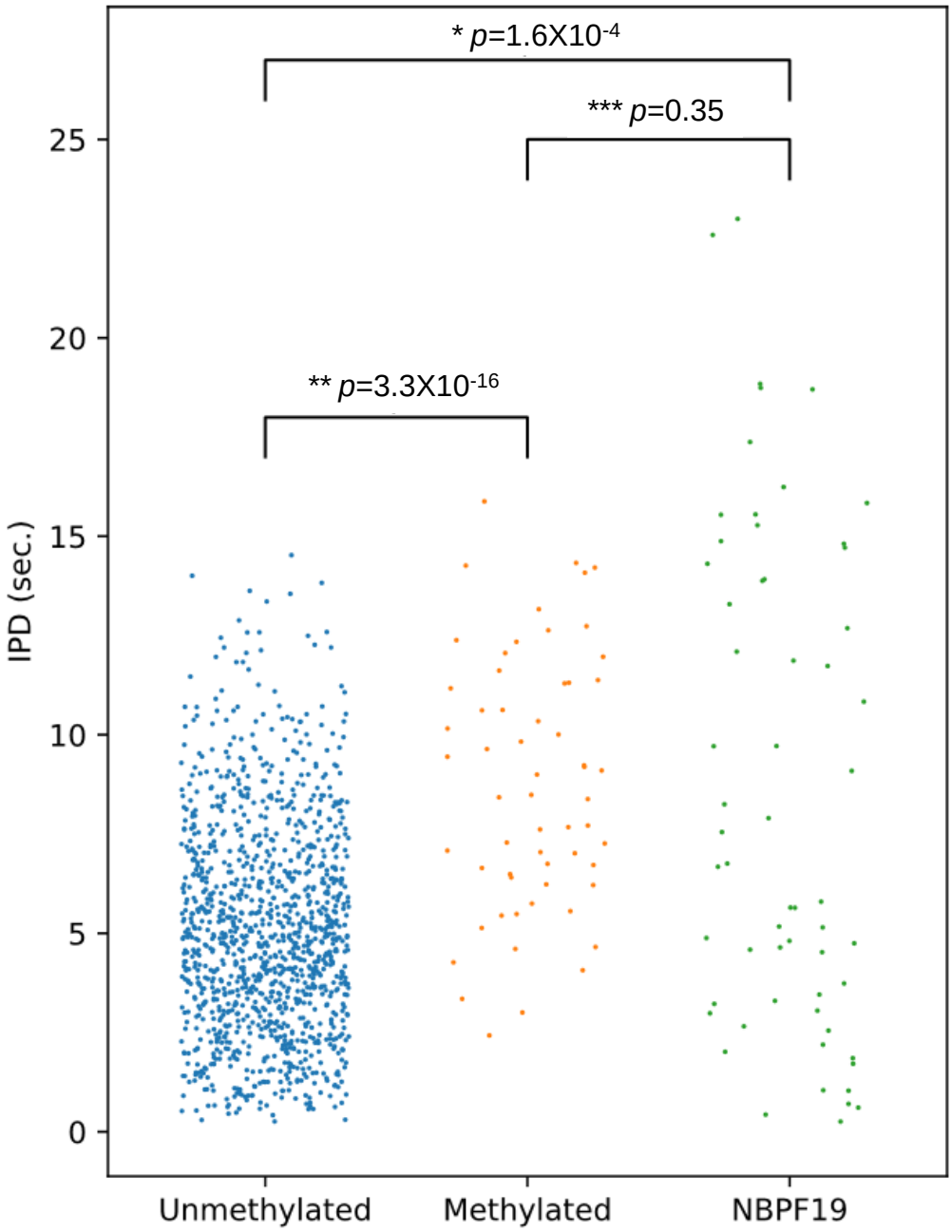
b



Supplementary Fig. 6 Frequency distribution of repeat sizes in *NBPF19* in 1,000 control subjects as revealed by fragment analysis

(a) Frequency distribution of repeat sizes of the CGG repeats and the flanking variable repeat sequences in *NBPF19* of 1,000 control subjects was determined by fragment analysis of PCR products obtained using *NBPF19*-specific primer pair (pGEX3'-NBPF19-6F and NBPF19-5R2). In the reference sequence (hg38), the repeat size is 13 repeat units, namely, $(AGG)(CGG)_9(AGG)_2(CGG)$.

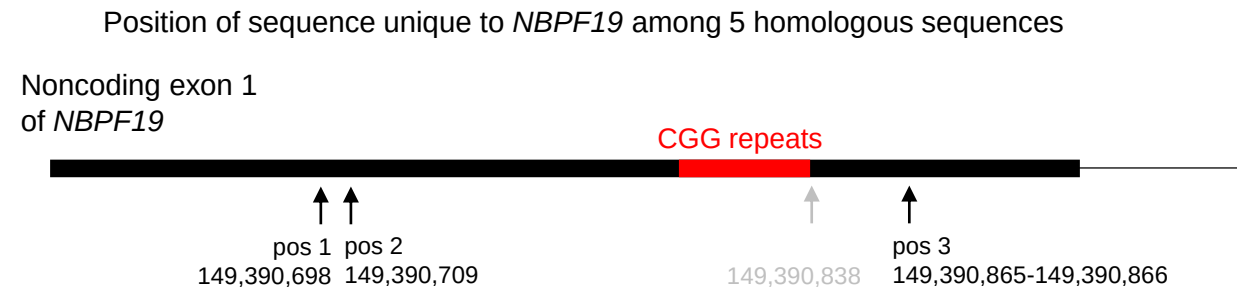
(b) Multiple sequence alignment of the five homologous sequences (*NBPF19*, *AC253572.1*, *NOTCH2NL*, *NBPF14*, and *NOTCH2*) using Clustal W2 is shown. Variable repeat sequences including CGG repeats are shown below the red line. In the fragment analysis, repeat sizes were determined as the lengths in repeat units between the flanking non-variable sequences (shown below the black dotted lines). Primers used in the analysis are shown by the blue arrows (pGEX3'-NBPF19-6F and NBPF19-5R2). Numbers shown in the figures indicate relative distances from 149,390,308 (*NBPF19*), 120,723,618 (*AC253572.1*), 146,229,332 (*NOTCH2NL*), 148,680,074 (*NBPF14*), and 120,069,958 (*NOTCH2*).



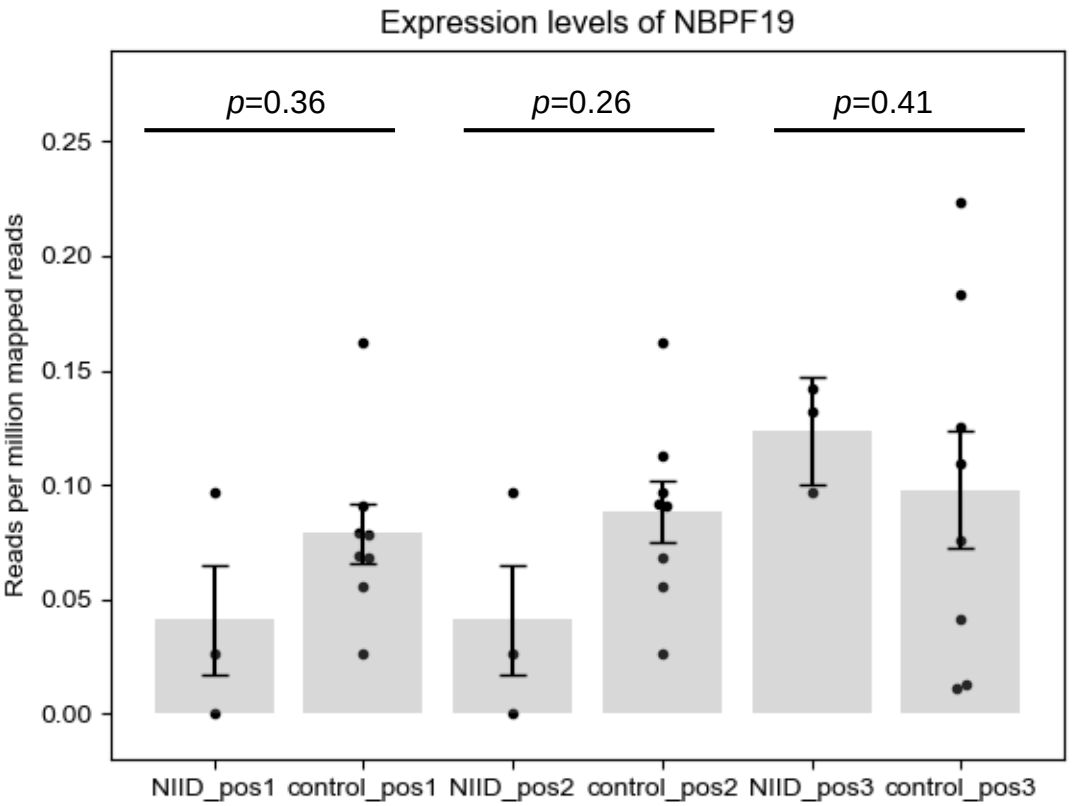
Supplementary Fig. 7 Inter-pulse durations (IPDs) in CGG sites examined by SMRT sequencing.

We first created a reference IPD set for the hypomethylated CGGs and hypermethylated CGGs using whole-genome bisulfite sequencing data and PacBio Sequel sequencing data (both obtained from the same individual). The reference benchmark set had 303 hypomethylated CGG repeat regions with 1,220 CpGs and 14 hypermethylated regions with 59 CpGs. We observed a significant difference in IPD statistics (on cytosine sites of CGG) between the methylated (n=59) and unmethylated (n = 1,220) CpG sites (* p=3.3X10^{-16}, one-sided) using Mann–Whitney U test, demonstrating that IPD is informative in inferring CpG methylation status of CGG repeats. We next examined whether the expanded CGG repeat in the 5' UTR of *NBPF19* was similar to hypomethylated CGG repeats or hypermethylated CGG repeats in terms of IPD statistics of CpG sites, and we checked the null hypothesis of independence of IPD statistics using Mann-Whitney U test. We found that the IPD distribution on cytosine sites of the expanded CGG repeat in the 5' UTR of *NBPF19* (n=60) was similar to that of hypermethylated CGG repeats (n = 59) (*** p=0.35, two-sided test) but was significantly dissimilar to that of hypomethylated CGG repeats (n = 1,220) (** p=1.6X10^{-4}, one-sided test), showing that the expanded CGG repeat in the 5' UTR of *NBPF19* was regionally hypermethylated as a whole.

a



b



Supplementary Fig. 8 Expression levels of *NBPF19* in brains examined by RNA-seq.

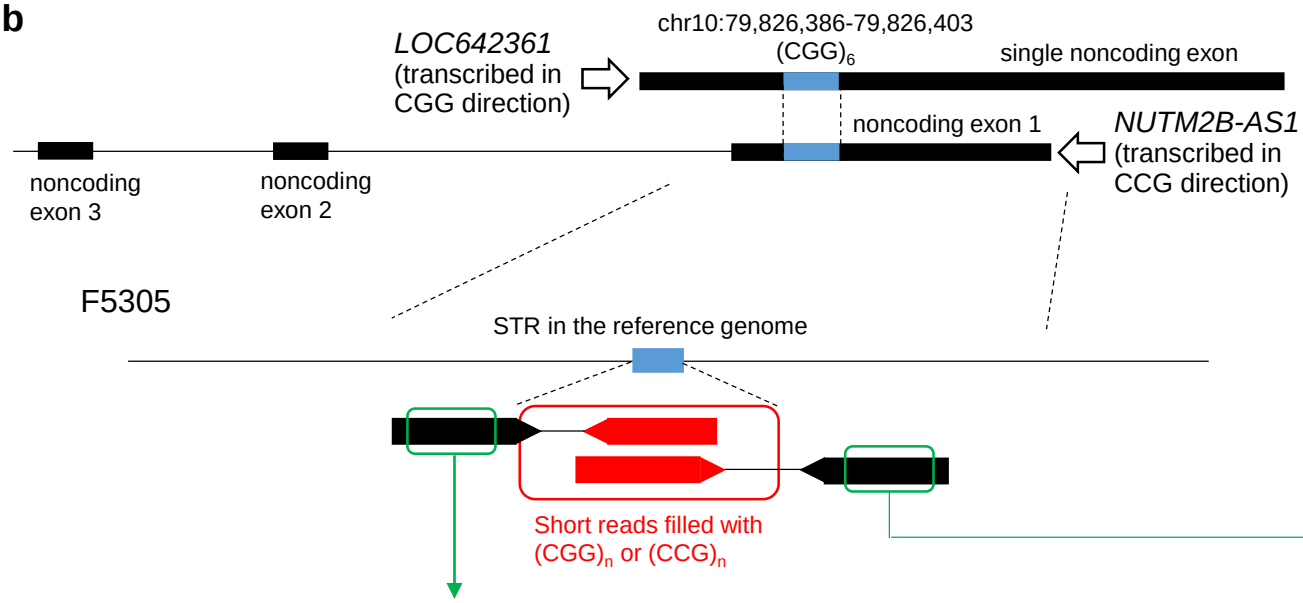
(a) There are 4 positions in noncoding exon 1 of *NBPF19* whose sequences are unique to *NBPF19* among the five homologous sequences in *AC253572.1*, *NOTCH2*, *NOTCH2NL*, *NBPF19*, and *NBPF14*. Physical positions in hg38 are shown. From RNA-seq data from 3 patients with NIID and 8 control subjects (occipital lobe), read per million mapped reads of the positions were calculated. Because one of the position is just downstream of the CGG repeats (chr1:149,390,838 in hg38), which made precise alignment difficult, we did not calculate coverages of the position.

(b) Expression levels of *NBPF19* were assessed using read per million mapped reads in the three positions as described above. We did not see any statistical significant differences between NIID (n=3) and control subjects (n=8, Wilcoxson rank sum tests, two-sided). The data are shown as means and standard errors of means.

a

Patient	F5305 III-5
Number of paired reads in which both reads were filled with CGG/CCG repeats 	4
Number of paired reads in which one of the paired reads did not contain tandem repeats (nonrepeat reads) 	9
Number of short reads (nonrepeat reads) that supported alignment to *LOC642361/NUTM2B-AS1*	7
Number of short reads (nonrepeat reads) that were aligned not only to *LOC642361/NUTM2B-AS1* but also to *LINC00863/NUTM2A-AS1 and FLJ22063/AMMECR1L*	2

b



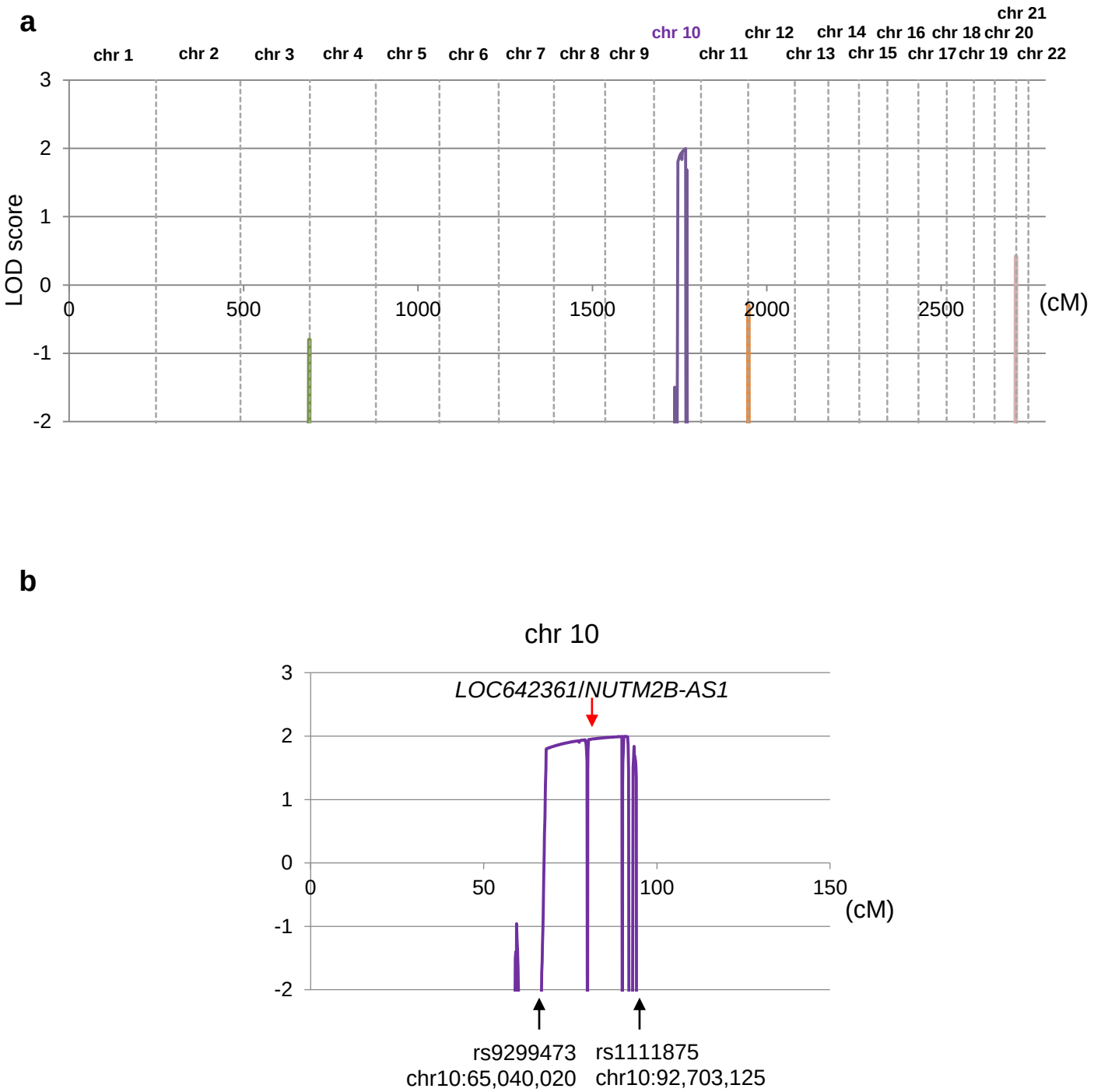

```
Short read            5′…TGGGGGCTCCCGAGAGAAGCTGGCCTGCGGGCGGGCCGGACGCGCTGCGCGGACGGGACGG…
LOC642361/NUTM2B-AS1  5′…TGGGGGCTCCCGAGAGAAGCTGGCCTGCGGGCGGGCCGGACGCGCTGCGCGGACGGGACGG…
LINC00863/NUTM2A-AS1  5′…TGGAGGCTCCCGAGAGAAGCTGGCCTGCGGGCGGGCCGGACGAGCTGCGCGGACGGGGCGG…
FLJ22063/AMMECR1L     5′…TGGGGGCTCCCGAGAGAAGCTGGCCTGCGGGCGGGCCGGACGCGCTGCGCGGACGGGGCGG…
```

```
Short read            5′…GTGAGGTAATGGAGGAAGAATGTAGGGAGTCCTCCGGTGGACCACACTCCTTCTAGCG…
LOC642361/NUTM2B-AS1  5′…GTGAGGTAATGGAGGAAGAATGTAGGGAGTCCTCCGGTGGACCACACTCCTTCTAGCG…
LINC00863/NUTM2A-AS1  5′…GTGAGATAATGGAGGAAGAATGTAGGGAGACCTCCAGTGGACCGCACTCCTTCTAGCG…
FLJ22063/AMMECR1L     5′…GTGAGGTAATGGAGGAAGAATGTAGGGAGTCCTCCAGTGGACCACACTCCTTCTAGCG…
```

Supplementary Fig. 9 Short reads indicating CGG repeat expansion in *LOC642361/NUTM2B-AS1*.

(a) Nine nonrepeat reads paired with reads filled with CGG/CCG repeats were identified in patient III-5 in F5305. Seven of the nine reads were mapped to the *LOC642361/NUTM2B-AS1* region best by BLAT. STR, short tandem repeat.

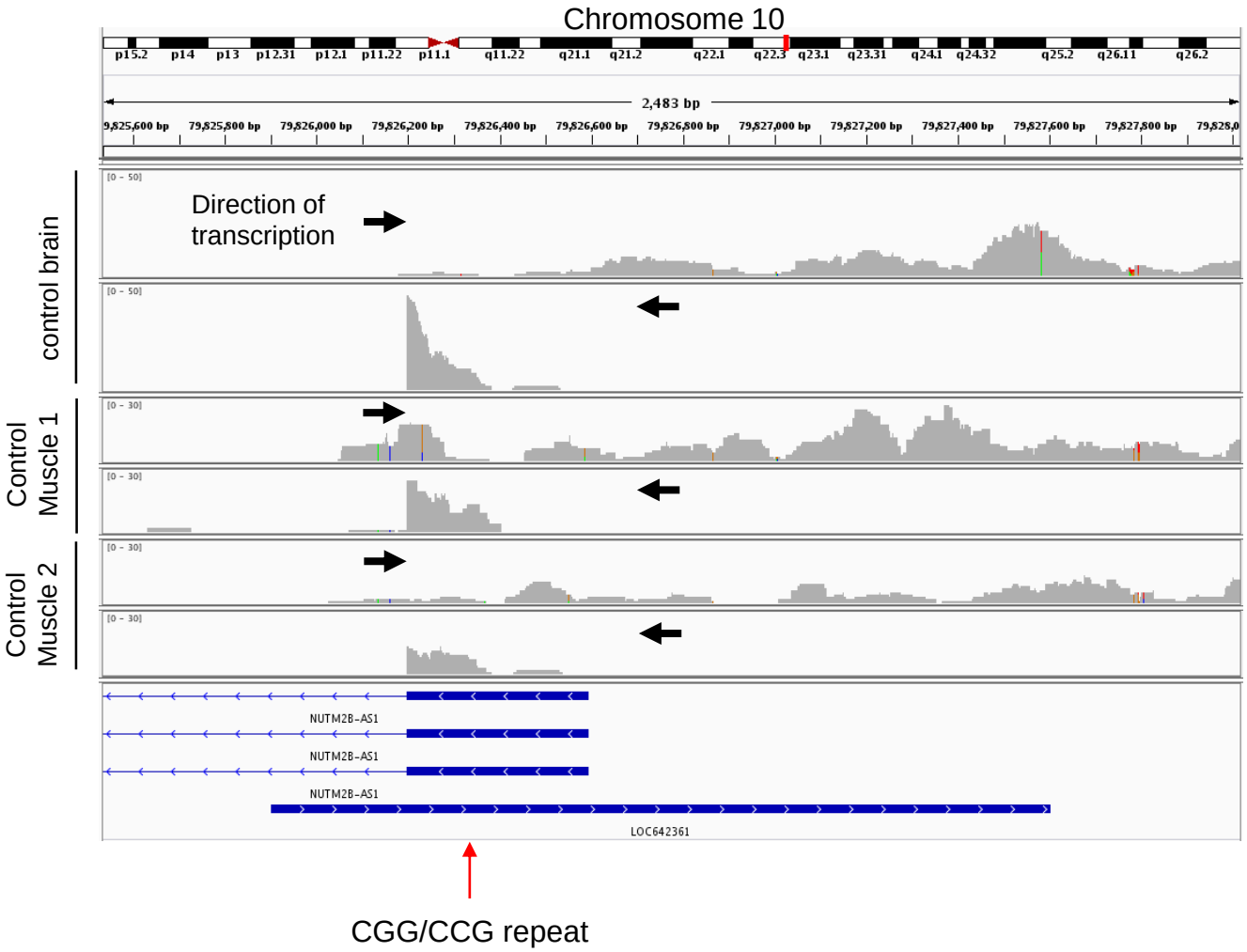
(b) Alignment of nonrepeat reads paired with reads filled with CGG/CCG repeats indicates that CGG repeat expansion is located in *LOC642361/NUTM2B-AS1*. Reads are shown in the same strand as the direction of transcription of *LOC642361*. Homologous sequences of *LOC642361/NUTM2B-AS1* and mismatches among them are shown in red squares.



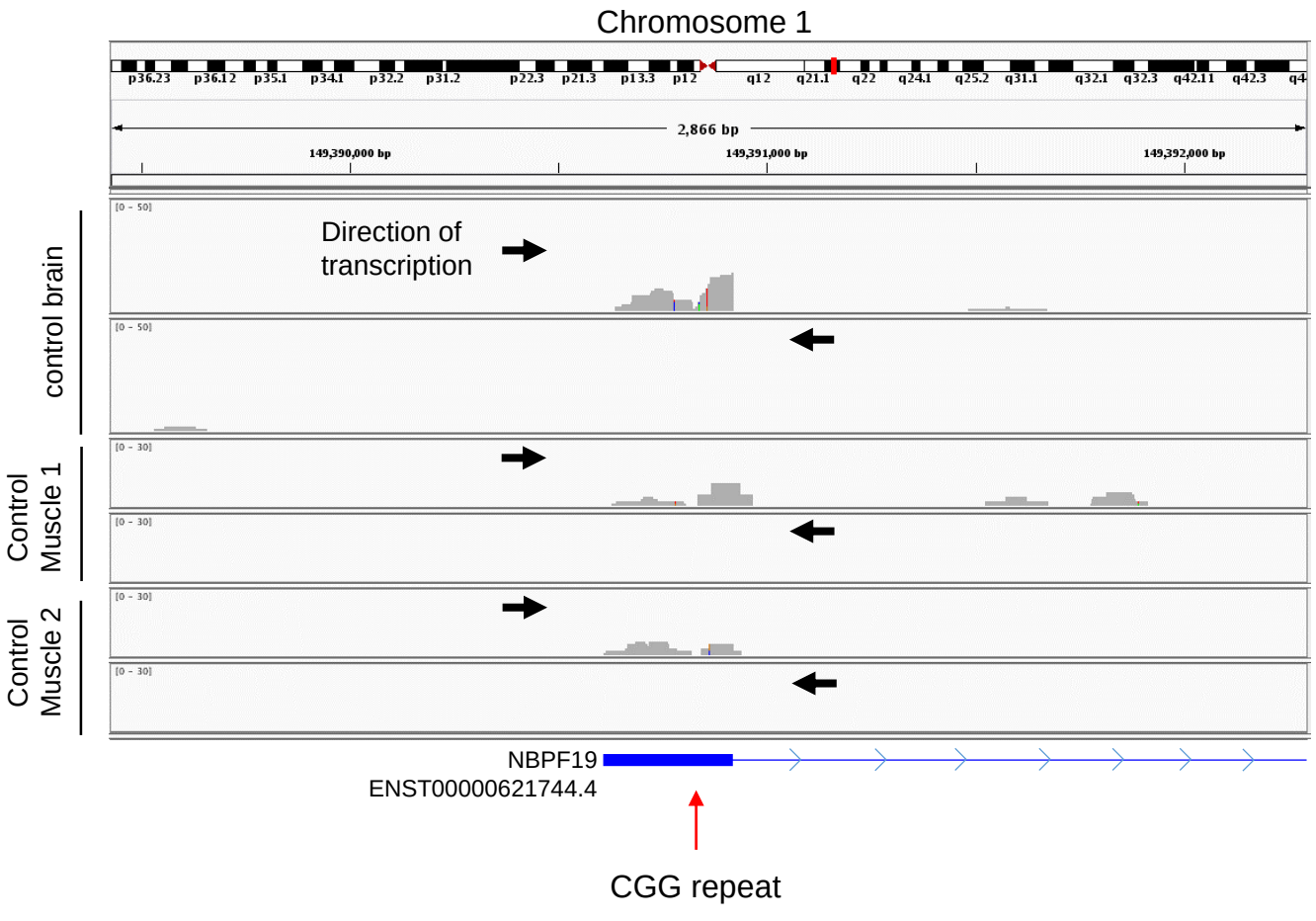
Supplementary Fig. 10 Linkage analysis of family (F5305) with OPML.

Parametric linkage analysis results of family with OPML (F5305, **Fig. 5c**) for all chromosomes **(a)** and candidate regions **(b)** are shown. Chromosome 10 is the only chromosome that shows LOD score of above 1. Boundary markers with physical positions in hg38 are indicated below. The locus of *LOC642361/NUTM2B-AS1* is indicated by the red arrow.

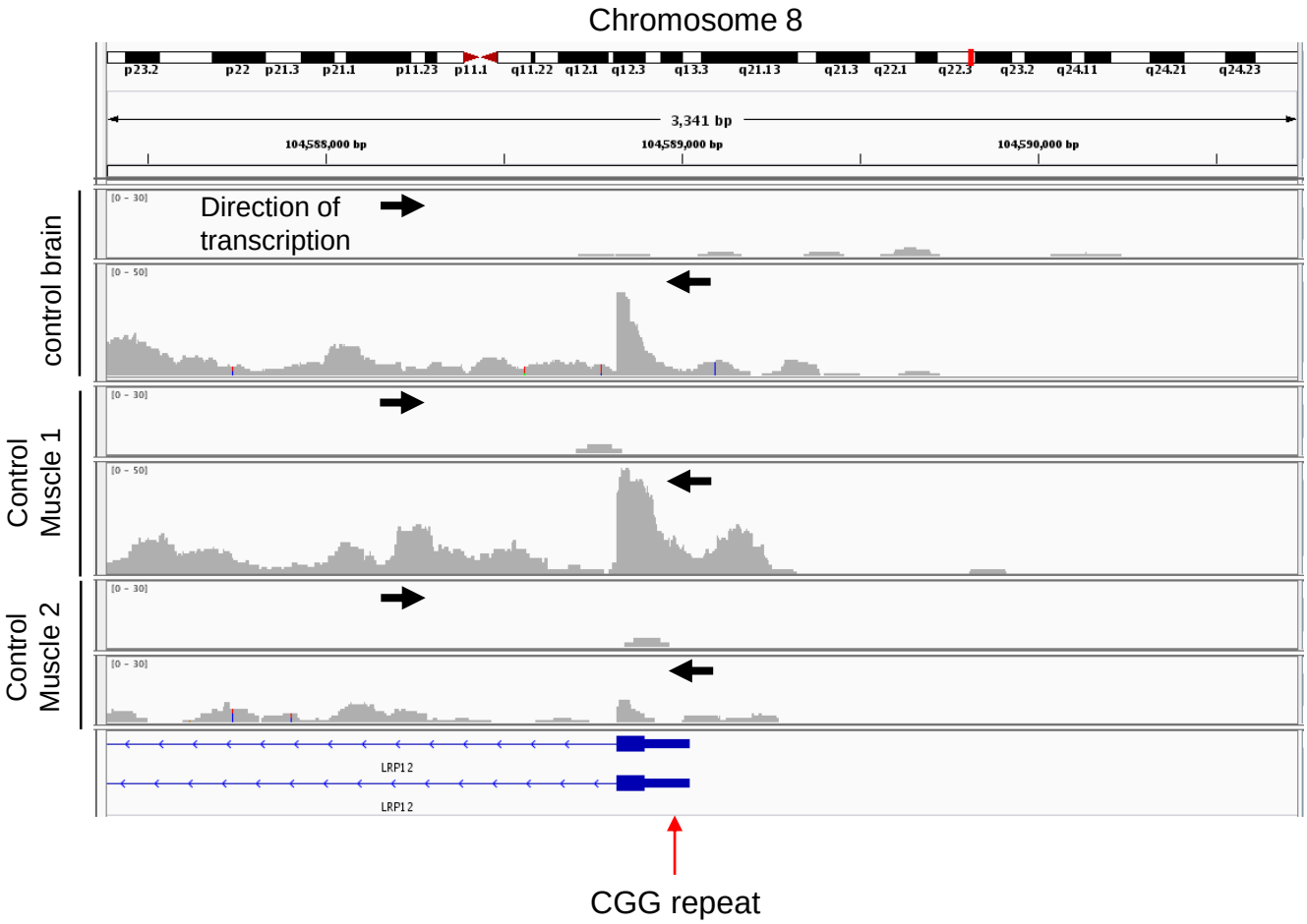
a *LOC642361/NUTM-2B*



b *NBPF19*



c *LRP12*



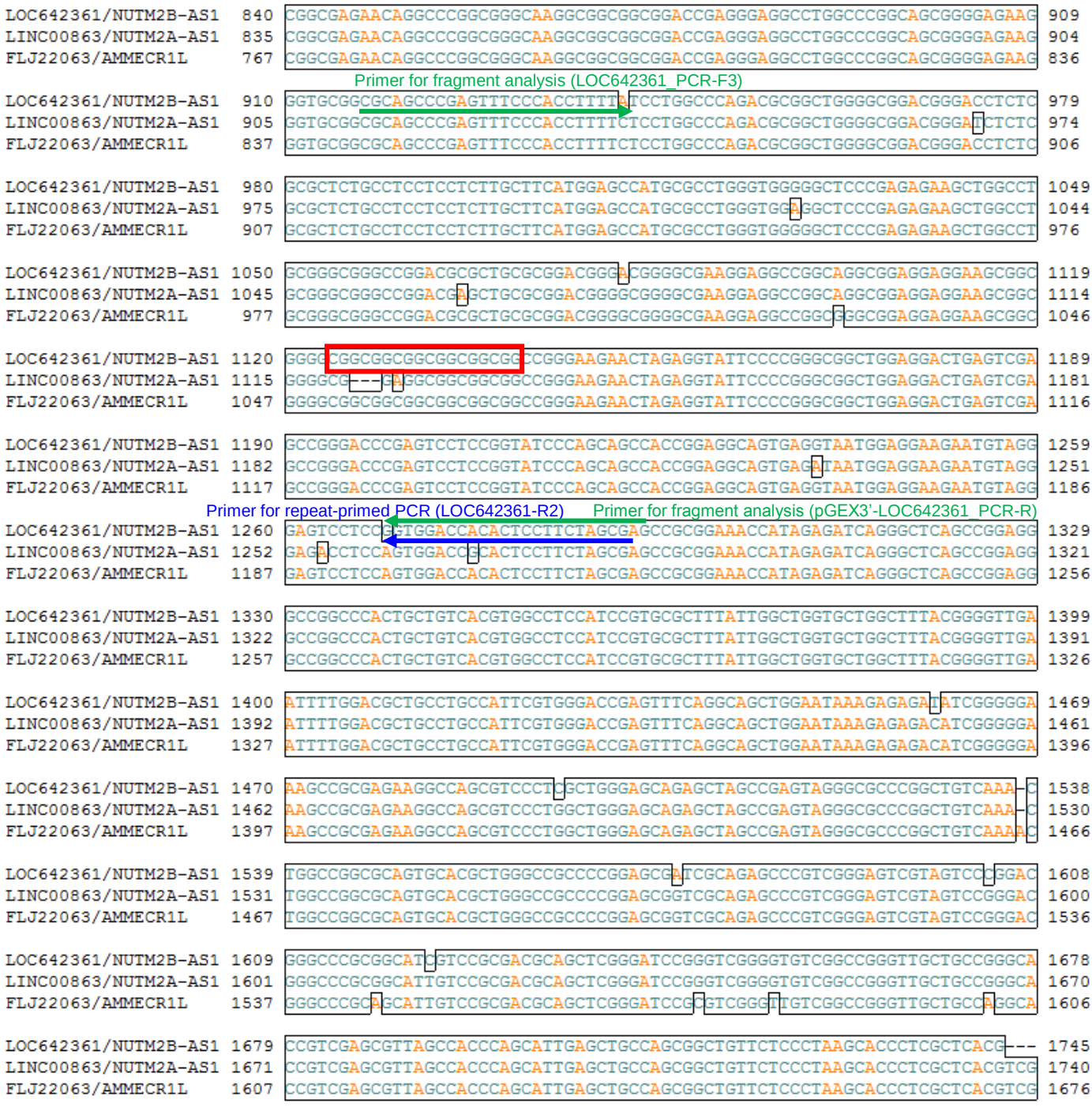
Supplementary Fig. 11 Bidirectional transcription of CGG/CCG repeats in *LOC642361/NUTM2B-AS1*.

Stranded RNA-seq data of a control brain and two control muscles using random primers in reverse transcription reactions are shown. Short reads are aligned to the reference sequence (hg38) using STAR [Dobin, A., et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013)]. Reads are divided into two files according to the direction of transcription. Only reads with mapping quality equal or more than 5 are shown using the Integrative Genomic Viewer [Robinson, J.T., et al. Integrative Genomic Viewer. *Nat. Biotechnol.* **29**, 24–26 (2011)].

(a) The CGG/CCG repeats in *LOC642361/NUTM2B-AS1* were bidirectionally transcribed, although coverages at the CGG/CCG repeat were underrepresented presumably owing to its high GC content.

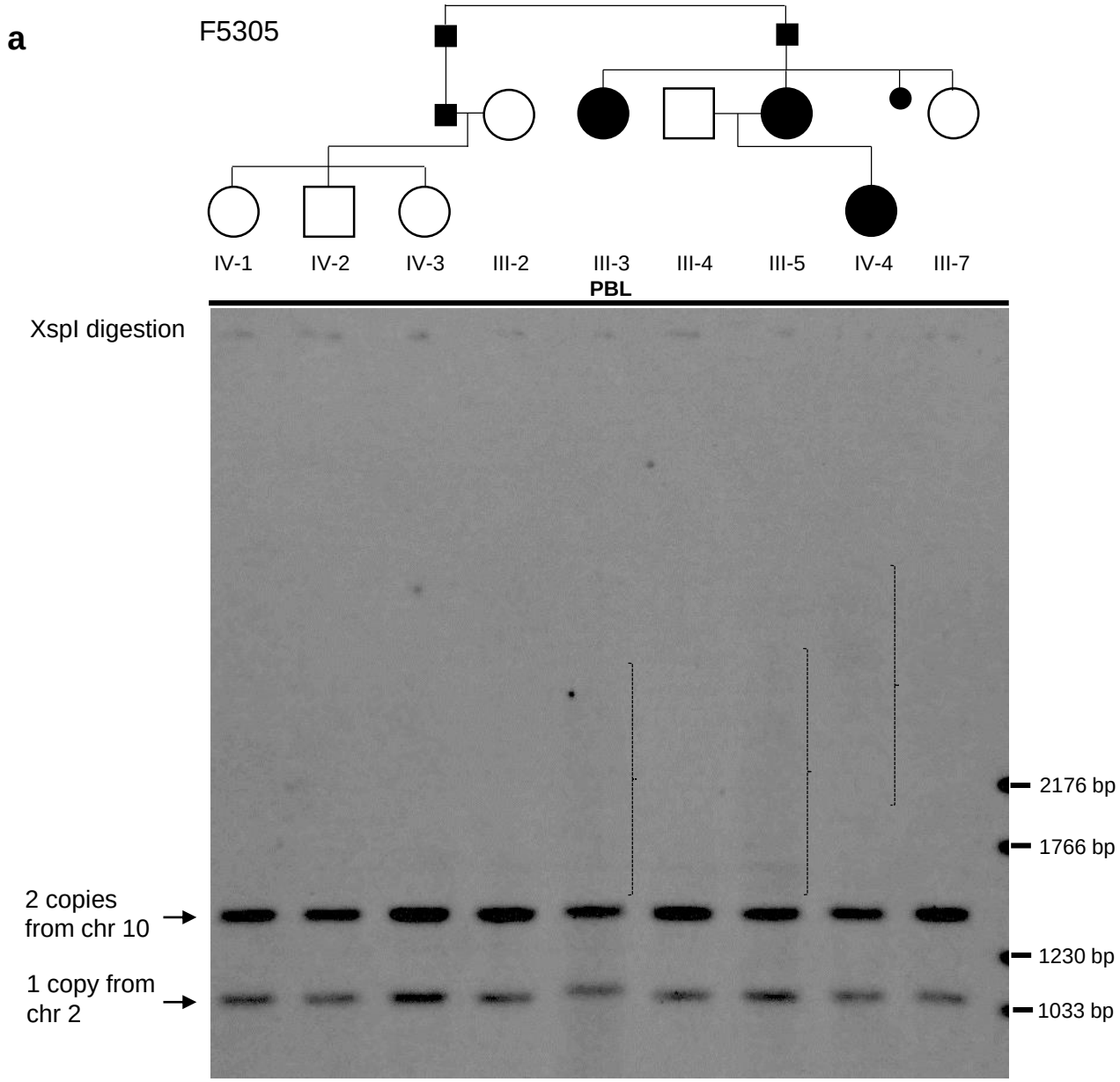
(b) No signals suggesting bidirectional transcription were observed in the CGG repeats in 5' UTR of *NBPF19*, although a mapping problem remains in the locus considering other highly homologous sequences.

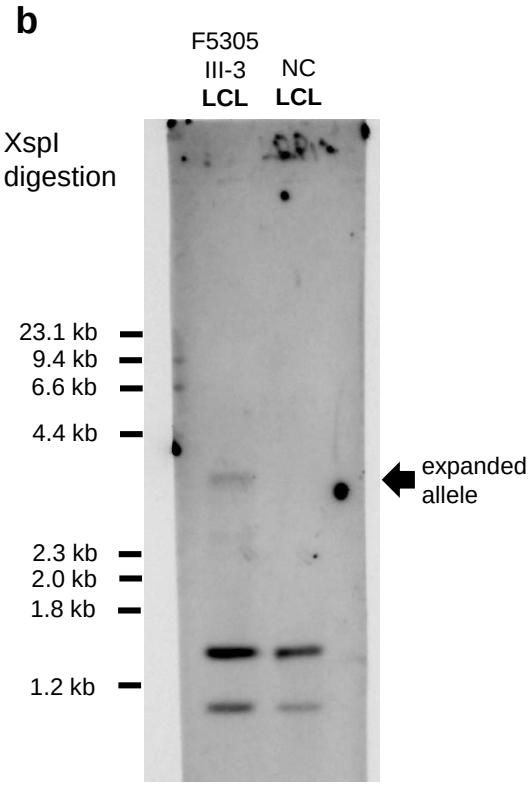
(c) Most of the reads in exon 1 of *LRP12* were sense reads, whereas only trivial antisense reads were observed.



Supplementary Fig. 12 Multiple sequence alignments of homologous genes of *LOC642361/NUTM2B-AS1*.

Multiple sequence alignment of sequence around the CGG/CCG repeats in *LOC642361/NUTM2B-AS1* with homologous sequences of *LINC00863/NUTM2A-AS1* (chromosome 10) and *FLJ22063/AMMECR1L* (chromosome 2) using Clustal W2. Sequences are derived from hg38. The position of CGG repeat expansion mutations is shown in red box. The primer sequence (LOC642361-R2, **Supplementary Table 4**) for repeat-primed PCR analysis (shown by blue arrow) and a primer pair (LOC642361_PCR-F3 and pGEX3'-LOC642361_PCR-R, **Supplementary Table 7**) for fragment analysis (shown by green arrows) were designed to avoid nonspecific amplification.





Supplementary Fig. 13 Southern blot analysis of *LOC642361*/*NUTM2B-AS1*.

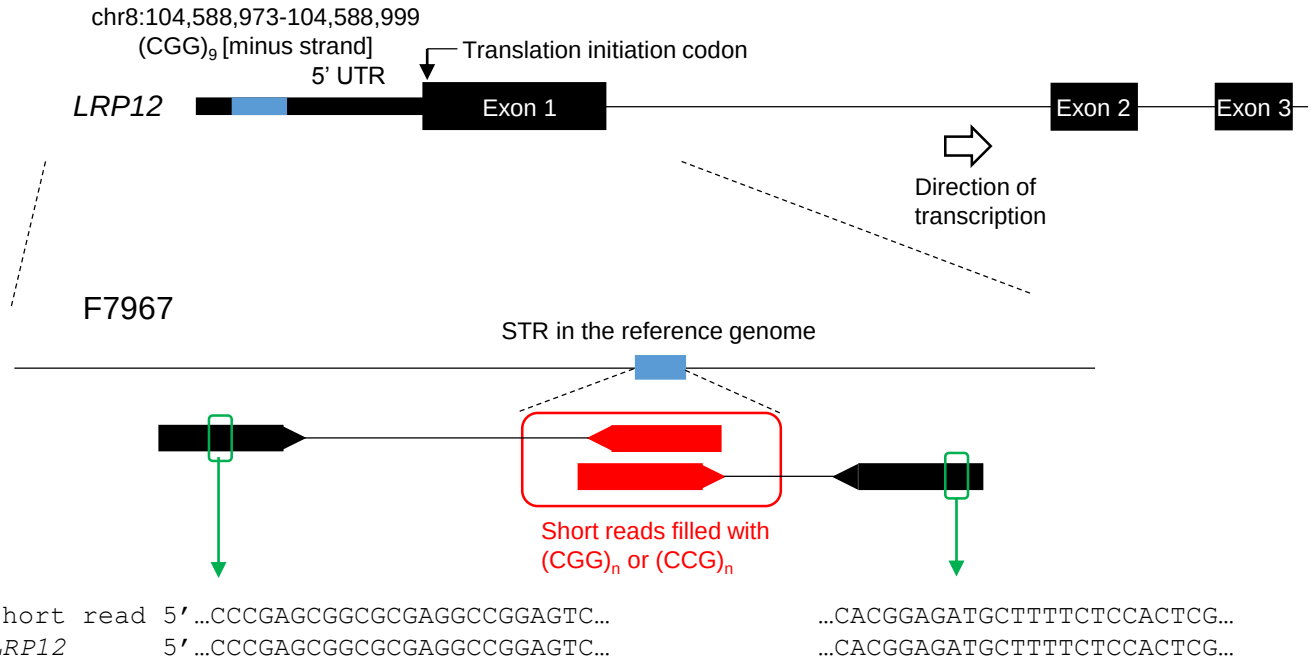
(a) Southern blot analysis was performed using probes targeting flanking regions of the CGG repeats in *LOC642361*/*NUTM2B-AS1* in chromosome 10. The probes were also predicted to hybridize to the other two similar sequences (*LINC00863*/*NUTM2A-AS1* in chromosome 10 and *FLJ22063*/*AMMECR1L* in chromosome 2). Predicted fragment sizes based on hg38 are 1.4 kb (*LOC642361*/*NUTM2B-AS1*), 1.4 kb (*LINC0863*/*NUTM2A-AS1*), and 1.1 kb (*FLJ22063*/*AMMECR1L*). Strong somatic instability of the CGG repeats was observed in genomic DNAs from peripheral blood leukocytes (PBL). The experiment was conducted once. The uncropped image is in **Supplementary Fig. 17**.

(b) An expanded allele of 2.1 kb (corresponding to 700 repeat units) was observed in genomic DNA from lymphoblastoid cell line of patient III-3 of family F5305. NC: normal control. The experiments were conducted twice with similar results. The uncropped image is in **Supplementary Fig. 17**.

a

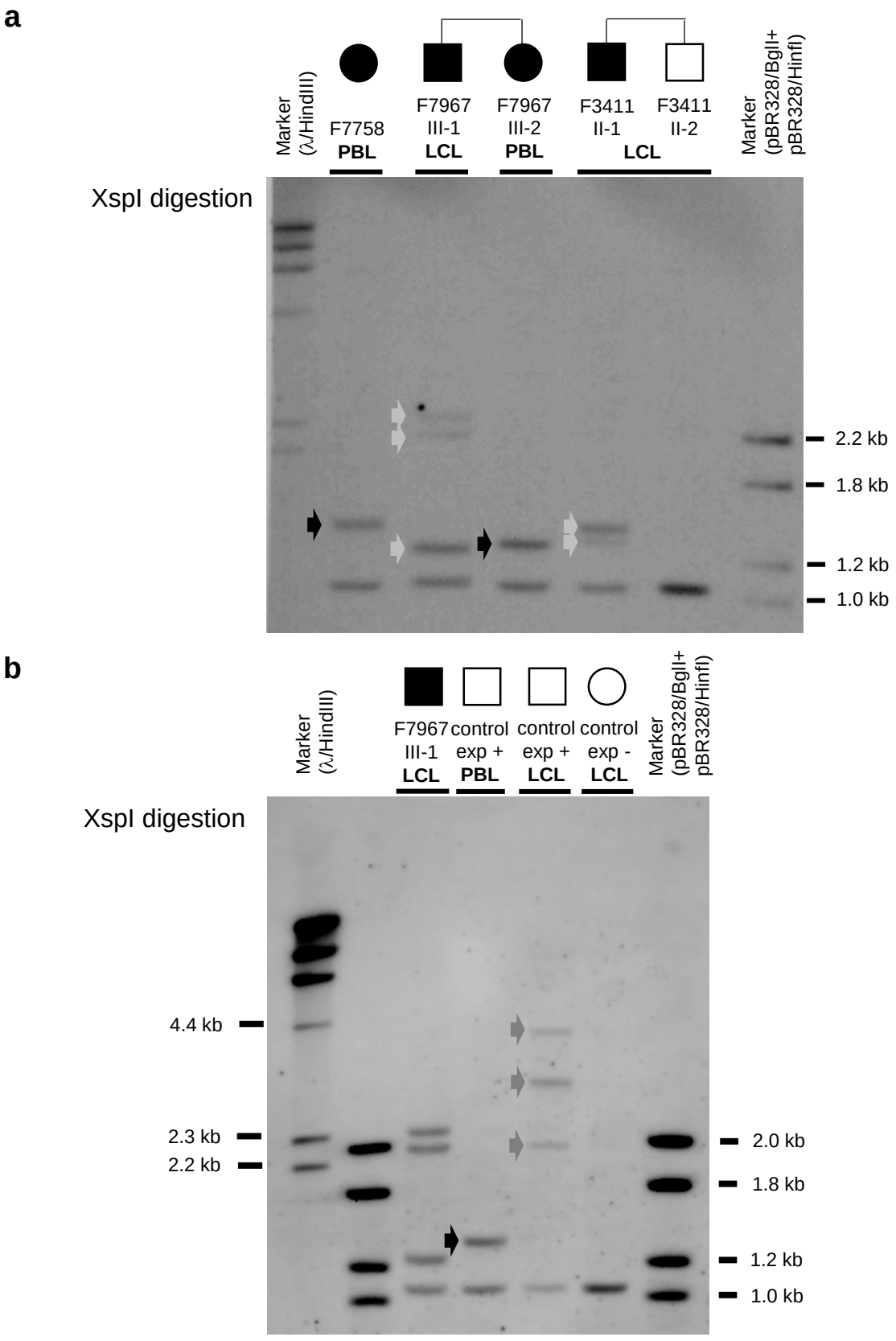
Patient	F7967 III-1
Number of paired reads in which both reads were filled with CGG/CCG repeats 	0
Number of paired reads in which one of the paired reads did not contain tandem repeats (nonrepeat reads) 	3
Number of short reads (nonrepeat reads) that supported alignment to *LRP12*	3

b



Supplementary Fig. 14 Short reads indicating CGG repeat expansion in *LRP12*.

(a) Three nonrepeat reads paired with reads filled with CGG/CCG repeats were identified in patient III-1 in F7967. All the three reads were mapped to the *LRP12* region by BLAT. STR, short tandem repeat.
(b) Alignment of nonrepeat reads paired with reads filled with CGG/CCG repeats indicates that CGG repeat expansion is located in 5' UTR of *LRP12*. Reads are shown in the same strand as the direction of transcription of *LRP12*.



Supplementary Fig. 15 Southern blot analysis of patients with oculopharyngodistal myopathy and controls.

(a) Southern blot analysis of patients with OPDM1. In genomic DNAs from lymphoblastoid cell lines (LCLs), multiple bands presumably derived from somatic instabilities (gray arrows) were observed, whereas single expanded bands (280 and 380 bp, black arrows) were observed in genomic DNAs from peripheral blood leukocytes (PBL). This experiment was conducted once. The uncropped image is in **Supplementary Fig. 17**.

(b) In the two controls who had the longest repeats as suggested by repeat-primed PCR analysis, whose ages at blood sampling were 63 years and 25 years, the expanded CGG repeat sizes exceeded 300 bp (black and gray arrows) and multiple bands were observed in genomic DNA from LCL (gray arrows). This experiment was conducted once. The uncropped image is in **Supplementary Fig. 17**.

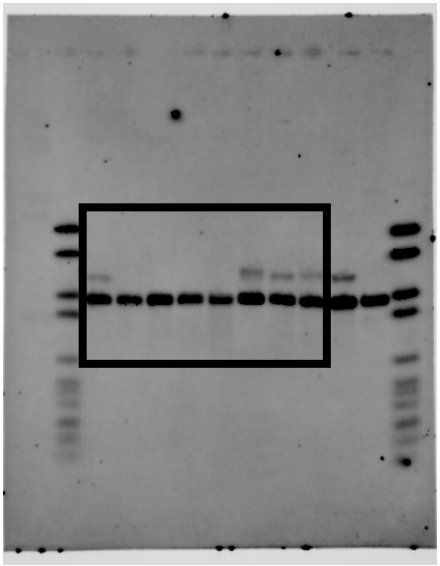
Exp +, carrier of expansion; exp -, noncarrier of expansions.

marker	Physical position (hg19)	Physical Position (hg38)	F3411 II-1 (sporadic)	F7758 proband (familial)	F7967 III-1 (familial)	Allele frequency in dbSNP
rs17202607	104,121,517	103,109,289	G	A	G	G: 0.93
rs2246446	104,138,724	103,126,496	A	A	A	A: 0.77
rs3110474	104,760,352	103,748,124	T	T	T	T: 0.80
rs2441863	105,159,724	104,147,496	G	(A/G)	A	A: 0.72
rs2441882	105,183,912	104,171,684	G	G	G	G: 0.72
rs74743202	105,394,388	104,382,160	T	T	T	T: 0.05
rs7845549	105,544,860	104,532,632	T	T	T	T: 0.43
rs34145689	105,547,841	104,535,613	A	A	A	A: 0.06
rs7842204	105,508,868	104,496,640	T	T	T	T: 0.15
rs11783651	105,545,808	104,533,580	G	G	G	G: 0.06
CGG repeat	105,601,201-105,601,227	104,588,973–104,588,999	expansion	expansion	expansion	
rs143338216	105,601,312	104,589,084	G	G	G	G: 0.002
rs33999897	105,635,296	104,623,068	G	G	G	G: 0.06
rs11782865	105,657,322	104,645,094	A	A	A	A: 0.18
rs35258168	105,670,614	104,658,386	A	A	A	A: 0.06
rs17282623	106,355,197	105,342,969	C	C	C	C: 0.90
rs1348262	106,685,293	105,673,065	G	G	T	G: 0.45
rs7824454	107,890,328	106,878,100	G	G	A	G: 0.87
rs16875856	108,224,687	107,212,459	A	A	G	A: 0.73
rs13253953	108,709,307	107,697,079	C	C	C	C: 0.79
rs11785730	108,785,847	107,773,619	G	A	G	G: 0.58
rs4538845	108,788,472	107,776,244	T	G	T	T: 0.58

Supplementary Fig. 16 Haplotype analysis of three families with oculopharyngodistal myopathy type 1.

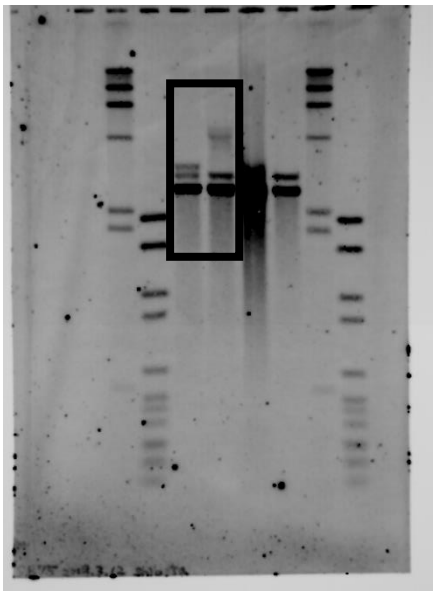
Haplotypes were reconstructed using single nucleotide variants genotyped using Affymetrix Genome Wide SNP array 6.0 in three families (F3411, F7758, and F7967). In Families F7758 and F7967, multiple affected individuals were observed, whereas in family F3411 only one affected individual (sporadic case) was observed. In this analysis, we used hg19 as the reference sequence. First, homozygosity haplotypes were reconstructed (Miyazawa et al. Homozygosity haplotype allows a genomewide search for the autosomal segments shared among patients. *Am J Hum Genet* **80**;1090–1102, (2007)) and shared regions among the three patients were visually confirmed (yellow). In addition to SNP array analysis, we also utilized10X GemCode Technology and compared each haploblock from three families from chr8:105,384,931 to chr8:105,657,322, avoiding genotypes within 10 kb of the boundaries of the haploblock indicated by longranger software. We selected single nucleotide variants with equal or more than 10 coverages from phased genotypes generated by 10X GemCode Technology. All the phased variants of the three families were matched as indicated by orange. These analyses suggested a common founder chromosome among these OPDM1 families.

Fig. 4e

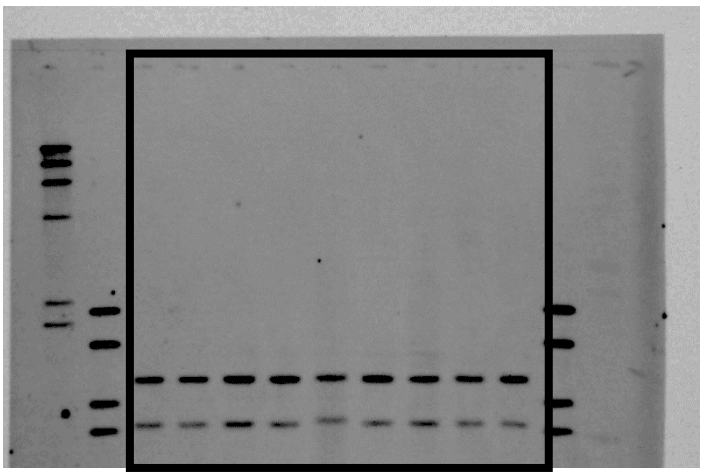


Supplementary Fig. 4b

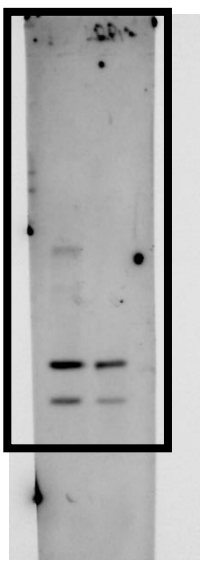
Supplementary Fig. 4c



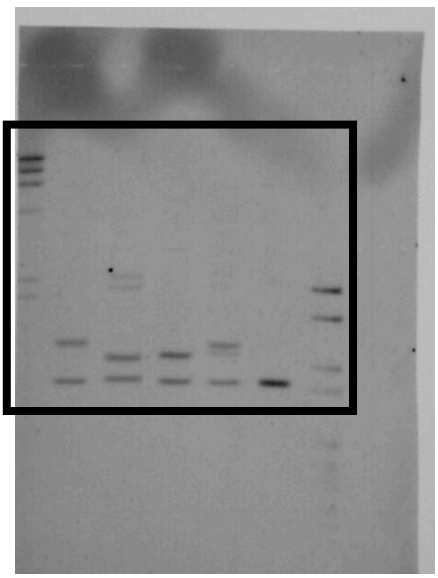
Supplementary Fig. 13a



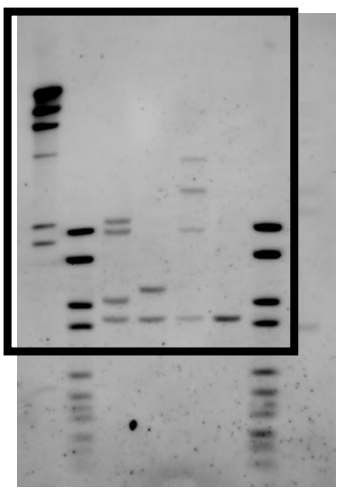
Supplementary Fig. 13b



Supplementary Fig. 15a



Supplementary Fig. 15b



Supplementary Fig. 17 Uncropped images of Southern blot analysis

Family	CGG repeat expansions in *NBPF19*	Family history	Diagnostics of proband	Number of patients examined	Reference/Notes
12 familial cases with NIID					
F6321	+	+	DWI, skin biopsy, autopsy	4	Yamaguchi, N. et al. *Intern. Med.* **57**, 3459–3462 (2018)
F8504	+	+	DWI, brain biopsy	1	
F9193	+	+	DWI, skin biopsy	4	
F9703	+	+	DWI, skin biopsy	1	Mitsutake, A. et al. *Neurol. Clin. Neurosci.* **7**, 136–138 (2019).
F10198	+	+	DWI, skin biopsy	1	
F10522	+	+	DWI, skin biopsy	1	
F10524	+	+	DWI, skin biopsy	1	Takumida, H. et al. *Geriatr. Geront. Int.* **17**, 2623–2625 (2017)
F10709	+	+	DWI, skin biopsy	1	
F10983	+	+	DWI, skin biopsy	1	
F11312	+	+	DWI, skin biopsy	1	
F11393	+	+	DWI, skin biopsy	3	
F11624	+	+	DWI, skin biopsy	1	
14 sporadic cases with NIID					
F7756	+	-	DWI, skin biopsy, rectal biopsy	1	
F8791	+	-	DWI, skin biopsy	1	Sasaki, T. et al. *Neurol. Clin. Neurosci*. **3**, 185–187 (2015)
F8821	+	-	DWI, skin biopsy	1	
F8832	+	-	DWI, skin biopsy	1	
F8982	+	-	DWI; skin, muscle, and peripheral nerve biopsy	1	Morimoto, S. et al. *J. Neurol. Sci.* **372**, 447–449 (2017)
F9224	+	-	DWI, skin biopsy	1	
F9468	+	-	DWI, skin biopsy	1	
F9681	+	-	DWI, skin biopsy	1	
F9714	+	-	DWI, skin biopsy	1	
F10587	+	-	DWI	1	
F10588	+	-	DWI, skin biopsy	1	
F11596	+	-	DWI, skin biopsy	1	
F8805	-	-	DWI, skin biopsy	1	
F10449	-	-	DWI	1	
Two cases with unavailable family history					
F8775	+	?	Autopsy	1	
F8776	+	?	Autopsy	1	

Supplementary Table 1 Japanese families with NIID enrolled in the study

DWI, diffusion-weighted imaging; ?: unavailable family history.

(a) 150 bp paired end sequencing analysis

Disease	**NIID1 F9193**	**OPML1 F5305**	**OPDM1 F7967**	**control**	**control**	**control**	**control**	**control**	**control**	**control**
Number of Total Reads (million reads)	1,143	574	658	1,823	2,202	1,183	1,265	1,277	931	1,006
Motif										
AACCCT	10,214	6,872	6,442	23,133	13,819	8,681	7,322	9,092	8,175	7,298
AAAG	279	95	207	331	512	203	199	410	218	251
AAAGG	257	38	93	240	483	211	131	198	143	118
AAAAG	123	60	28	124	38	52	11	146	72	78
AAG	98	48	152	72	103	220	49	70	71	61
AAAAT	84	108	28	122	247	105	149	26	192	86
AATGG	60	34	45	33	115	7	102	189	24	15
AATAG	50	51	25	4	0	29	25	7	35	2
AAGAC	0	0	0	96	0	0	0	0	0	0
AGATG	40	0	0	0	0	0	0	0	0	0
AAGGAG	0	34	0	0	2	1	0	0	2	0
CCG	36	16	3	0	0	0	0	0	0	0
AAACC	7	3	1	15	4	9	4	2	0	5
ACCCTG	6	15	0	3	2	0	0	0	1	0
AAGAG	4	1	3	13	12	6	4	3	0	1
AGCAT	4	0	0	11	0	0	0	2	0	57
AAGAGG	0	0	0	0	29	0	0	0	0	0
ATCC	3	5	5	4	10	4	4	2	6	1
AAACT	0	1	0	0	0	0	0	10	0	0
AAAGC	0	75	0	0	0	0	0	0	0	0
AACAT	0	0	0	10	0	0	7	0	0	0
AATAC	0	5	7	0	0	5	10	20	10	0
AGC	0	0	0	0	4	10	0	9	0	9

(b) 126 bp paired end sequencing analysis

Disease	**NIID1 F8504**	**NIID1 F9468**	**NIID1 F9785**	**control**
Number of Total Reads (million reads)	1,189	1,414	920	1,004
Motif				
AACCCT	11,042	17,336	11,024	11,804
AAG	398	356	280	337
AAAG	310	465	289	410
AAAAT	107	143	48	107
AAAAG	79	143	14	58
AAAGG	59	213	97	90
AAGGAG	43	10	109	12
AATGG	38	31	12	12
AGCAT	0	164	0	0
AGC	1	4	118	6
AATAG	6	24	12	55
CCG	8	24	7	0
AAGAG	7	22	11	8
AAACC	1	15	3	25
ACCTCT	0	13	0	0

Supplementary Table 2 Results from TRhist

Data from whole-genome sequence analysis of 150 bp **(a)** and 126 bp **(b)** paired-end reads. Only repeat motifs with 3–6 bases that any of the subjects showing more than 9 reads have been observed are shown. Reads filled with CCG(=CGG) repeats are observed in patients with NIID1 (red), OPML1 (blue), and OPDM1 (green). NIID1, neuronal intranuclear inclusion disease type 1; OPML1, oculopharyngeal myopathy with leukoencephalopathy type 1; OPDM1, oculopharyngodistal myopathy type 1.

Supplementary Table 3 Raw and corrected long reads

read name	error-correction	start	end	length of STRs	size of STR unit	Number of STR copies	% of matches	% of mismatches	% of indels	% of A	% of C	% of G	% of T	Consensus
@m54133_180606_220849/33489872/0_17215/0_17215	No	9556	9972	416	3	134.0	81	6	13	0	32	64	2	GCG
@m54133_180606_220849/33489872/0_17215/0_17215	Yes	9429	9859	430	3	141.3	88	2	10	0	32	67	0	GGC
@m54133_180607_082437/39846713/66679_93243/0_26564	No	575	1000	425	3	141.3	84	6	10	0	33	63	3	GGC
@m54133_180607_082437/39846713/66679_93243/0_26564	Yes	523	955	432	3	144.0	93	1	6	0	33	66	0	GCG
@m54133_180606_220849/12256111/0_22721/0_22721	No	10386	10813	427	3	139.7	82	7	11	2	32	63	2	CGG
@m54133_180606_220849/12256111/0_22721/0_22721	Yes	10347	10782	435	3	144.3	91	3	6	0	32	66	0	GGC
@m54133_180607_082437/34472355/31065_59831/0_28766	No	4965	5424	459	3	155.7	85	4	11	0	32	66	0	GGC
@m54133_180607_082437/34472355/31065_59831/0_28766	Yes	4924	5378	454	3	153.3	89	2	9	0	33	66	0	GGC
@m54133_180607_082437/34472355/4506_31020/0_26514	No	5603	6081	478	3	158.7	83	5	12	1	32	64	1	CGG
@m54133_180607_082437/34472355/4506_31020/0_26514	Yes	5628	6088	460	3	154.3	92	2	6	0	32	67	0	GGC

Rows with white background and those with yellow background show read names, properties of reads and nucleotide sequences before error correction and those after error correction by Canu, respectively.

Repeat sequence

GCGGCGGCGGCGGCGGCCTGCGGCGGCGGCGGCGGCGCGGTCGGCGGGCGGCGGGCGGCGGCGGCGGCGTCGGCGGCGGCGGCGGCTGCGGGCGGCGGCGGCGGT
GCGGCGCGGCCGGCCGCGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCCGGCGGGCGGCGGTCGGCCGCGGCGGCGGCGCGGGCGGCGGCGGCGGCGG
CTGGCGGGCGGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCGCGGGCGAAGCGGCGGCGGCGGCGGTCGGCGGCGCGGCGGCGGCCGGCGGCGGCGG
CGGCGGGCGGCGGCGGCGGCGGCTGCGAGCGGTGGCGGCGGGCGGCGGCGGCGGCCGGCGCGGCGGCGGCGGCGGCTGCGGCGGGGCGGCGGGGGGGGCGGCG
GGCGGCGGCGGCGGCGGCGGCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGGCGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG
GCGGCGCGGCGGCGGCGGCTGCGGCGGCTGCGGCGGCGGCGGCGGCGTCTGCGGCGGCGGCGGCGGCGGCGGTGGCGGCGGCGCGGCGGCTGCGGCGGCGCGGCGGC
GGCGGCG
GCGGCGGCGGCGCGGCGGCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCCGGCGGCGGCGCGGCGGCGGCGGCGGCGGCGCGGCGGCG
GCGGCGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGGCGGCGGCGCGCGGCGGCGCGGCGGCGGCGGCGGCG
GGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG
GCGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG
CGGCGGCG
CGGCGGCGGCGGCTGTGCGGAGGCGGCGGGCGGCGGCGGGGCGGCGGCGCGGCGGCGGCGGCGGCGGCGTCGGCGGCGGCGGCGGCGGCGCGCCGGCGGCGG
CGCGGCGGCGGCGGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGGCGGCGCGGCGGCGGCGGCGGCGGCAGGC
GGCGGCGGCGGAGGCGGCTTTGGCTTCGGCGGCATGGCGGCGGCGGCGGCGGCGGCGGATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCGGCGGAGG
CGGCGGCGGGGCGCGGCGGCGGCGGCTGCCGGCGGCGGCGGGCGGCGGGGCGGCGGCGGTCCGGCCGGCGGCGGCAGAGCGGCGGCAGGCGGCGGCCGGCGGCGG
CGGCGGCG
GGCGGCGGCGGCGGTGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGCGCCGGCGGC
GGCGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGC
GCGGCGGCGGCGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG
CGGCGGCGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG
GCGGCGGGGCGGCGGGCGGCGGCGCGGCGGCGGGCGGCGGCGGCG

flanking sequence (5' end)	flanking sequence (3' end)
GTGGTGACTCCTCTATCGGGACGCCCCTCCCATTGTATCTGGCCCAGGCT	CGGGCGCGGCGAACCGAGAATATGCCCGCCCTGCGCAGCTCTGACTGCTG
GGGAGAGTGGGGCTCCTCTATCGGGACCCCCTCCCCATTGGATCTGCCCA	AACCGAGAAGATGCCCGCCCTGCGCCGCTCTCTGCTGTGGGCGCTGCTGG
GGGCTCCTCTATCGGACCCCCTTCGCCCATTCGTGGATCTGCCCATCGCG	ACCGGAGATGGCCCCGCCCTGCGCCGTGCTCTGCTGTGGGGGCTGCTGGC
AGAGTGGGGCTCCTCTATCGGGACCCCCTCCCCATGTGGATCTGCCCATC	ACCGAGAAGATGCCCGCCCTGCGCCGCTCTGCTGTGGGCGCTGCTGGCGC
AGAGAGAGTGGGGATACTCTAATCGGGACCCCCTCCCCATGGGATCTGCC	ACCGAGAAGATGCCACGCCATGCGCCGCTCTGATGTGGGCGATGCTGGCG
AGAGAGTGGGGCTCCTCTATCGGGACCCCCTCCCCATGTGGATCTGCCCA	ACCGAGAAGATGCCCGCCCTGCGCCGCTCTGCTGTGGGCGCTGCTGGCGC
GAGAGGGGGCTCCTCTATCGTGACCCCCTCCCCATGTTGTCTGCCCCCCA	ACCGAGAAGAATGCCCGCCCTGGCCCGCTCTGCTGTGGGCGCTGCTTCTG
AGAGAGTGGGGCTCCTCTATCGGGACCCCCTCCCCATGTGGATCTGCCCA	ACCGAGAAGATGCCCGCCCTGCGCCGCTCTGCTGTGGGCGCTGCTGGCGC
AGAGTGGGGCTCCTATATCGGGACCCCCTCCCCATGTGATCTGCCCAGGT	ACCGAGAAGATTTGCCCGCCCTGCGACGCTCTGATGTGGGCTGGGCTGGC
AGAGAGTGGGGCTCCTCTATCGGGACCCCCTCCCCATGTGGATCTGCCCA	ACCGAGAAGATGCCCGCCCTGCGCCGCTCTGCTGTGGGCTGCTGCTGGCG

Primer	Buffer/ Final conc.	Sequence
NBPF19	GCII	
NBPF19-R	180 nM	5’-AGCGCCCACAGCAGAGCGGC-3’
pGEX3’-$(CGG)_5$	9 nM	5’-CCGGGAGCTGCATGTGTCAGAGGCGGCGGCGGCGGCGG-3’
fam-pGEX3’	180 nM	5’-(Fam)-CCGGGAGCTGCATGTGTCAGAGG-3’
LOC642361/*NUTM2B-AS1*	GCI	
LOC642361-R2	180 nM	5’-CGCTAGAAGGAGTGTGGTCCACC-3’
pGEX3’-$(CGG)_5$	9 nM	5’-CCGGGAGCTGCATGTGTCAGAGGCGGCGGCGGCGGCGG-3’
fam-pGEX3’	180 nM	5’-(Fam)-CCGGGAGCTGCATGTGTCAGAGG-3’
LRP12	GCII	
LRP12-R5	180 nM	5’-GGAGGGAGGAGAAGCTGGAGGTAGACG-3’
pGEX3'-$(CGG)_5$	9 nM	5’-CCGGGAGCTGCATGTGTCAGAGGCGGCGGCGGCGGCGG-3’
fam-pGEX3’	180 nM	5’-(Fam)-CCGGGAGCTGCATGTGTCAGAGG-3’

Supplementary Table 4 Primer sequences used for repeat-primed PCR analysis

GCI buffer or GCII buffer was obtained from TaKaRa (TaKaRa LA taq with GC buffer).

Primer	Buffer/ Final conc.	Sequence
FMR1	GCII	
primer-R	800 nM	5'-TCAGGCGCTCAGCTCCGTTTCGGTTTCA-3'
pGEX3'-$(CCG)_5$	80 nM	5'-CCGGGAGCTGCATGTGTCAGAGGCCGCCGCCGCCGCCG-3'
fam-pGEX3'	800 nM	5'-(Fam)-CCGGGAGCTGCATGTGTCAGAGG-3'

Supplementary Table 5 Primer sequences used for the repeat-primed PCR analysis of *FMR1*

We used deaza-dGTP in place of dGTP. PCR reaction was conducted as follows; initial denaturation at 94° C for 1 min, followed by 30 cycles of 94° C for 30 s, 60° C for 30 s, and 72° C for 80 s or slow down PCR protocol shown in online methods. GCII buffer was obtained from TaKaRa (TaKaRa LA taq with GC buffer).

NBPF19 (SacI digestion)

Template/ Probe	Primer	Length of template/probe	Sequence
Template	NBPF19_1	801 bp	5'-ACCGAGAAGATGCCCGCCCTGC-3'
	NBPF19_4		5'-CGCGCCTCGGAAAGAATAACAG-3'
Probe 1	NBPF19_1	419 bp	5'-ACCGAGAAGATGCCCGCCCTGC-3'
	NBPF19_1R		5'-AACTGCCCACCTCCCTGCACC-3'
Probe 2	NBPF19_4F	442 bp	5'-CGGCAGCAAGTCTCAGAAACT-3'
	NBPF19_4		5'-CGCGCCTCGGAAAGAATAACAG-3'

NBPF19 (NheI digestion)

Template/ Probe	Primer	Length of template/probe	Sequence
Template	NBPF19_2aF2	1217 bp	5'-GTGTGCTGCTCGCGTCTTTG-3'
	NBPF19_2cF3		5'-CTACAATTCTCTAAAGCAGG-3'
Probe 3	NBPF19_2aF2	365 bp	5'-GTGTGCTGCTCGCGTCTTTG-3'
	NBPF19_2aR2		5'-GTGTGGGTGGGATGGGGAAG-3'
Probe 4	NBPF19_2bF2	424 bp	5'-TATTAAACGGATGACACTCC-3'
	NBPF19_2bR2		5'-CTGGTCCACTTCTGAAATTC-3'
Probe 5	NBPF19_2cF2	290 bp	5'-GAATTTCAGAAGTGGACCAG-3'
	NBPF19_2cF3		5'-CTACAATTCTCTAAAGCAGG-3'

LOC642361 (XspI digestion)

Template/ Probe	Primer	Length of template/probe	Sequence
Template	2107F	946 bp	5'-TTGGAGTGTGCAGAGGGATAAGG-3'
	3052R		5'-CGCAGGCCAGCTTCTCTCG-3'
Probe 6	2107F	425 bp	5'-TTGGAGTGTGCAGAGGGATAAGG-3'
	2531R		5'-TGGATTCCACCCCCGCGGCTC-3'

LRP12 (XspI digestion)

Template/ Probe	Primer	Length of template/probe	Sequence
Template	2243F	752 bp	5'-GGAGTCAGGACAGATGTGTACAC-3'
	2995R		5'-GTGGTTATGGCCTGTCGCTGG-3'
Probe 7	2243F	319 bp	5'-GGAGTCAGGACAGATGTGTACAC-3'
	2562R		5'-GATGCTTGACTGTGAGAAAGCAGAG-3'
Probe 8	2538F	458 bp	5'-CTCTGCTTTCTCACAGTCAAGCATC-3'
	2995R		5'-GTGGTTATGGCCTGTCGCTGG-3'

Supplementary Table 6 Primer sequences used for preparation of template and probes for Southern blot analysis

Genomic DNA segments flanking the CGG repeats were amplified using the primer pairs (NBPF19_1/NBPF19_4, NBPF19_2aF2/NBPF19_2cF3, 2107F/3052R, and 2243F/2995R) and subcloned into plasmids. Probes for Southern blot hybridization analysis were prepared by digoxigenin (DIG) labeling using primer pairs (NBPF19_1/NBPF19_1R for Probe 1 and NBPF19_4F/NBPF19_4 for Probe 2, NBPF19_2aF2/NBPF19_2aR2 for Probe 3, NBPF19_2bF2/NBPF19_2bR2 for Probe 4, and NBPF19_2cF2/NBPF19_2cR2 for Probe 5 [*NBPF19*], 2107F/2531R for Probe 6 [*LOC642361*/*NUTM2B-AS1*], and 2243F/2562R for Probe 7 and 2538F/2995R for Probe 8 [*LRP12*]).

Primer	Buffer/ Final conc.	Sequence
NBPF19 (*NOTCH2NLC*)	GCII	
NBPF19-5R2	180 nM	5’-TACTCACCATGCGCGGGGGT-3’
pGEX3'-NBPF19-6F	180 nM	5’-CCGGGAGCTGCATGTGTCAGAGG GCCTGTGCTTCGGAC-3’
fam-pGEX3’	180 nM	5’-(Fam)-CCGGGAGCTGCATGTGTCAGAGG-3’

Primer	Buffer/ Final conc.	Sequence
LOC642361/*NUTM2B-AS1*	GCII	
LOC642361_PCR-F3	180 nM	5’-CGCAGCCCGAGTTTCCCACCTTTTA-3’
pGEX3'-LOC642361_PCR-R	180 nM	5’-CCGGGAGCTGCATGTGTCAGAGG CTCGCTAGAAGGAGTGTGGTCCACC-3’
fam-pGEX3’	180 nM	5’-(Fam)-CCGGGAGCTGCATGTGTCAGAGG-3’

Primer	Buffer/ Final conc.	Sequence
LOC642361/*NUTM2B-AS1*	GCII	
LRP12-F6	180 nM	5’-GCCACCCTCTCGTCTCGCGCTG-3’
pGEX3'-LRP12-F	180 nM	5’-CCGGGAGCTGCATGTGTCAGAGG CGAGGAAAAGCAAGAGCAAC-3’
fam-pGEX3’	180 nM	5’-(Fam)-CCGGGAGCTGCATGTGTCAGAGG-3’

Supplementary Table 7 Primer sequences used for the fragment analysis in controls subjects

PCR reaction was conducted as follows; initial denaturation at 98˚ C for 1 min, followed by 35 cycles of 98˚ C for 10 sec, 58˚ C for 30 sec, and 68˚ C for 30 sec for *NBPF19*, initial denaturation at 95˚ C for 1 min, followed by 30 cycles of 94˚ C for 30 s, 50˚ C for 30 s, and 72˚ C for 60 s for *LOC642361*/*NUTM2B-AS1*, and .initial denaturation of 98˚ C for 1min, followed by 35 cycles of 98˚ C for 10 sec, 60˚ C for 30 sec, and 68˚ C for 30 sec for *LRP12*. GCII buffer was obtained from TaKaRa (TaKaRa LA taq with GC buffer).

Primer	Buffer/ Final conc.	Sequence
NBPF19 (*NOTCH2NLC*)	GCII	
Barcode 1-NBPF19-6F	360 nM	5'-X_{16}-TTGCGCCTGTGCTTCGGAC-3'
Barcode 2-NBPF19-5R2	360 nM	5'-Y_{16}-TACTCACCATGCGCGGGGGT-3'

Barcode 1
>0061_Forward
ATGCTGATGACGCGCT
>0062_Forward
GACAGCATCTGCGCTC
>0063_Forward
AGCGTCTGACGTGAGT
>0064_Forward
TCGATATACGACGTGC
>0065_Forward
TCGTCATACGCTCTAG
>0066_Forward
CGACTACGTACAGTAG
>0067_Forward
GCGTAGACAGACTACA
>0068_Forward
ACAGTATGATGTACTC
>0069_Forward
GTCTGATAGATACAGA
>0070_Forward
CTGCGCAGTACGTGCA

Barcode 2
>0041_Forward
GTACATATGCGTCTGT
>0042_Forward
GAGACTAGAGATAGTG
>0043_Forward
TACGCGTGTACGCAGA
>0044_Forward
TGTCACTCATCTGAGT
>0045_Forward
GCACATACACGCTCAC
>0046_Forward
GCTCGTCGCGCGCACA
>0047_Forward
ACAGTGCGCTGTCTAT
>0048_Forward
TCACACTCTAGAGCGA
>0049_Forward
TCACATATGTATACAT
>0050_Forward
CGCTGCGAGAGACAGT
>0051_Forward
ACACACAGACTGTGAG
>0052_Forward
GCAGACTCTCACACGC

Supplementary Table 8 Primer sequences and barcode sequences used for the circular consensus sequencing (CCS) analysis using a SMRT sequencer

Each forward and reverse primers contained 16-mer barcodes as shown below. PCR reaction was conducted as follows; initial denaturation at 98˚ C for 1 min, followed by 35 cycles of 98˚ C for 10 sec, 58˚ C for 30 sec, and 68˚ C for 30 sec. GCII buffer was obtained from TaKaRa (TaKaRa LA taq with GC buffer).

Gene/region	Chromo-somal band	Strand	Start	End	Qscore	Identity
LOC642361 /NUTM2B-AS1	10q22.3	+	79,825,306	79,827,410	2105	100.0%
LINC00863 /NUTMS1-AS1	10q23.2	+	87,341,355	87,343,458	2041	98.7%
FLJ22063 /AMMECR1L	2q14.3	+	127,884,991	127,887,030	39246	98.5%

Supplementary Table 9 Homologous regions of CGG repeats in *LOC642361*/*NUTM2B-AS1*

The regions of CGG repeats in *LOC642361*/*NUTM2B-AS1* have two homologous sequences with high similarity in the reference genome (hg38). Identity and qscore are calculated using BLAT. The sequence (chr10:79,825,306-79,827,410) that corresponds to 1 kb upstream and downstream of the CGG repeat in *LOC642361*/*NUTM2B-AS1* is used as a query.

Supplementary Note

Summary of clinical presentation of the index patient (III-3) in family F5305 with oculopharyngeal myopathy with leukoencephalopathy (OPML).

The pedigree chart of this family (F5305) is shown in **Fig. 5c**. There are seven affected individuals consistent with autosomal dominant inheritance.

The index patient (III-3, **Fig, 5c**) noticed nasal voice at the age of 15. The progression of her symptom was as follows: at 27 years old (y/o), she began noticing easy fatigability of her extremities; at 30 y/o, ptosis; and at 32 y/o, mild dysphagia. She underwent repeated blepharoplasties at ages 34, 45, and 56. She was examined at another hospital at 35 y/o, where ptosis, dysarthria, dysphagia, and weakness of facial and neck muscles were observed, however, the limb muscles were minimally involved. Needle electromyography revealed motor units with short duration and low voltage, which were considered as myogenic changes. Muscle biopsy revealed no abnormal findings. Motor nerve conduction studies were normal.

Her symptoms gradually progressed. Detailed examinations at 58 y/o at the Department of Neurology, The University of Tokyo Hospital revealed ptosis, nearly complete external ophthalmoplegia, dysarthria with nasal voice, and dysphagia. She also had facial, neck, and diffuse limb muscle weakness, accompanied with diffuse muscular atrophy and generalized areflexia. She had dysuria requiring abdominal pressure to assist urination. Although tube feeding was tried because of dysphagia and repeated aspiration pneumonia, tube enteral feeding was not adequate due to severe gastrointestinal dysmotility. Weakness of respiratory muscles led to hypercapnia. On laboratory examination, serum creatine kinase levels were below the lower limit (29 IU/L), while serum lactate and pyruvate levels were normal. Echocardiography revealed diffuse hypokinesis of the left ventricle (ejection fraction of 44%). Magnetic resonance imaging of the head revealed T2-hyperintensity signals in the white matter accompanied with hyperintensity signals on diffusion-weighted images in the corticomedullary junction (**Fig. 1**). Clinical presentation of other family members are summarized in **Supplementary Table 10**.

Although autosomal dominant mitochondrial diseases exhibiting chronic progressive external ophthalmoplegia were initially considered (**Fig. 5c**) from the pedigree chart, no rearrangements or deletions of mitochondrial DNA were identified by Southern blot hybridization analysis of genomic DNA extracted from the abdominal muscle specimen. Causative mutations in the nuclear genes responsible for autosomal

dominant mitochondrial diseases (*POLG*, *SLC25A4*, *C10ORF2*, *POLG2*, *RRM2B*, *DNA2*, *OPA1*, and *AFG3L2*) were not identified by whole-genome sequence analysis. Oculopharyngeal muscular dystrophy was excluded by the analysis of the CGG repeat in *PABPN1*. Although oculopharyngodistal myopathy (OPDM) was another differential diagnosis, patients with OPDM usually showed muscular weakness with predominance in distal limbs and rimmed vacuoles in muscle biopsy specimens[1], while the patients in this family did not show such findings. Involvement of the gastrointestinal tract[2] or the heart[3] was only infrequently observed in patients with OPDM. Taken together with myopathy of the oculopharyngeal type, diffuse muscular weakness, characteristic brain MRI findings (leukoencephalopathy), and the gastrointestinal involvement, we considered the characteristic clinical presentation in this family constitute a novel clinical entity and designate the disease as OPML.

References

1. Satoyoshi, E. & Kinoshita, M. Oculopharyngodistal myopathy. *Arch. Neurol.* **34**, 89–92 (1977).
2. Amato, A. A., Jackson, C. E., Ridings, L. W. & Barohn, R. J. Childhood-onset oculopharyngodistal myopathy with chronic intestinal pseudo-obstruction. *Muscle Nerve* **18**, 842–847 (1995).
3. Thevathasan, W., et al. Oculopharyngodistal myopathy—a possible association with cardiomyopathy. *Neuromuscul. Disord.* **21**, 121–125 (2011).

Supplementary Table 10 Clinical characteristics of the family (F5305) with oculopharyngeal myopathy with leukoencephalopathy (OPML)

Affected family members	II-3	III-1	III-3	III-5	III-6	IV-4
Age at onset	unknown	unknown	15 y/o	40 y/o	adolescence-20 y/o	18-19 y/o
Age at death	40 y/o	55 y/o	64 y/o	N/A	54 y/o	N/A
Age at examination	N/A	N/A	58 y/o	59 y/o	52 y/o	28 y/o
Sex	Male	Male	Female	Female	Female	Female
Initial symptom	ND	ND	Nasal voice	Ptosis	Dysarthria	Ptosis
Cognitive function	ND	ND	MMSE 29/30 (56 y/o)	MMSE 27/30 FAB 9/18 VIQ 83, PIQ 82, IQ 82 (WAIS-R, 55 y/o)	HDS-R 30/30	MMSE 30/30 FAB 17/18
Ptosis	+	+	+	+	+	+
Blepharoplasty	-	-	+	+	+	+
Dysphagia	-	+	+	+	+	-
Dysarthria (small, nasal voice)	+	+	+	+	+	Mild
Restricted eye movement (external ophthalmoplegia)	ND	ND	+	+	+	-
Weakness Face	ND	ND	+	+	+	+
Limb	ND	ND	Proximal and distal	Proximal and distal	Proximal and distal	Proximal and distal
Respiratory failure	ND	ND	60 y/o	-	52 y/o	-

Other neurological findings	ND	ND	Mild ataxia	Tremor (cerebellar origin suspected)	-	-
White matter change/Brain atrophy	ND	ND	+/+	-/+	+/+	ND
DWI hyperintensity signal	ND	ND	+	ND	ND	ND
Gastrointestinal involvement			+	+	+	-
Cardiac involvement	ND	ND	Dilated cardiomyopathy	Paroxysmal atrial fibrillation	-	ND
Creatine kinase level	ND	ND	29 IU/L	94 IU/L	57 IU/L	ND
Electromyography	ND	ND	Myogenic	Almost normal (39 y/o)	Myogenic	Not performed
Muscle biopsy	ND	ND	Non-specific myopathic changes (64 y/o)	Non-specific myopathic changes	ND	Not performed
Nerve conduction study	ND	ND	Normal (35 y/o)	Normal (39 y/o)	Normal	Not performed

Abbreviation: y/o, years old; ND, not described; N/A: not applicable; MMSE, Mini-Mental State Examination; HDS-R, The Revised Hasegawa dementia scale; WAIS-R, Wechsler Adult Intelligence Scale revised; PIQ, performance intelligence Quotient; VIQ, verbal intelligence quotient; TIQ, total intelligence quotient