nature genetics

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41588-019-0459-y

In the format provided by the authors and unedited.

Long-read sequencing identifies GGC repeat expansions in *NOTCH2NLC* associated with neuronal intranuclear inclusion disease

Jun Sone[1,2,26], Satomi Mitsuhashi[3,26], Atsushi Fujita[3,26], Takeshi Mizuguchi[3], Kohei Hamanaka[3], Keiko Mori[4], Haruki Koike[1], Akihiro Hashiguchi[5], Hiroshi Takashima[5], Hiroshi Sugiyama[6], Yutaka Kohno[7], Yoshihisa Takiyama[8], Kengo Maeda[9], Hiroshi Doi[10], Shigeru Koyano[10], Hideyuki Takeuchi[10], Michi Kawamoto[11], Nobuo Kohara[11], Tetsuo Ando[12], Toshiaki Ieda[13], Yasushi Kita[14], Norito Kokubun[15], Yoshio Tsuboi[16], Kazutaka Katoh[17,20], Yoshihiro Kino[21], Masahisa Katsuno[1], Yasushi Iwasaki[18], Mari Yoshida[18], Fumiaki Tanaka[10], Ikuo K. Suzuki[19], Martin C. Frith[20,22,23], Naomichi Matsumoto[3,27*] and Gen Sobue[1,24,25,27*]

[1]Department of Neurology, Nagoya University Graduate School of Medicine, Nagoya, Japan. [2]Department of Neurology, National hospital organization Suzuka National Hospital, Suzuka, Japan. [3]Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Japan. [4]Department of Neurology, Oyamada Memorial Spa Hospital, Yokkaichi, Japan. [5]Department of Neurology and Geriatrics, Kagoshima University Graduate School of Medical and Dental Sciences, Kagoshima, Japan. [6]Department of Neurology, National Hospital Organization Utano National Hospital, Kyoto, Japan. [7]Department of Neurology, Ibaraki Prefectural University of Health Sciences, Ibaraki, Japan. [8]Department of Neurology, University of Yamanashi, Chuo, Yamanashi, Japan. [9]Department of Neurology, National hospital organization Higashi-Ohmi General Medical Center, Higashi-Ohmi, Japan. [10]Department of Neurology and Stroke Medicine, Yokohama City University Graduate School of Medicine, Yokohama, Japan. [11]Department of Neurology, Kobe City Medical Center General Hospital, Kobe, Japan. [12]Department of Neurology, Anjo Kosei Hospital, Anjo, Japan. [13]Department of Neurology, Yokkaichi Municipal Hospital, Yokkaichi, Japan. [14]Department of Neurology, Hyogo Brain and Heart Center, Himeji, Japan. [15]Department of Neurology, Dokkyo Medical University, Tochigi, Japan. [16]Department of Neurology, Fukuoka University, Fukuoka, Japan. [17]Research Institute for Microbial Diseases, Osaka University, Suita, Japan. [18]Department of Neuropathology, Institute for Medical Science of Aging, Aichi Medical University, Nagakute, Japan. [19]Department of Biological Sciences, Graduate School of Science, The University of Tokyo, Tokyo, Japan. [20]Artificial Intelligence Research Center, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan. [21]Department of Bioinformatics and Molecular Neuropathology, Meiji Pharmaceutical University, Tokyo, Japan. [22]Graduate School of Frontier Sciences, University of Tokyo, Chiba, Japan. [23]Computational Bio Big-Data Open Innovation Laboratory, National Institute of Advanced Industrial Science and Technology, Tokyo, Japan. [24]Department of Neurology, and Brain and Mind Research Center, Nagoya University Graduate School of Medicine, Nagoya, Japan. [25]Aichi Medical University, Nagakute, Aichi, Japan. [26]These authors contributed equally: Jun Sone, Satomi Mitsuhashi, Atsushi Fujita. [27]These authors jointly supervised this work: Naomichi Matsumoto, Gen Sobue. *e-mail: naomat@yokohama-cu.ac.jp; sobueg@med.nagoya-u.ac.jp

Supplementary Notes

RNA-seq analysis of fibroblasts

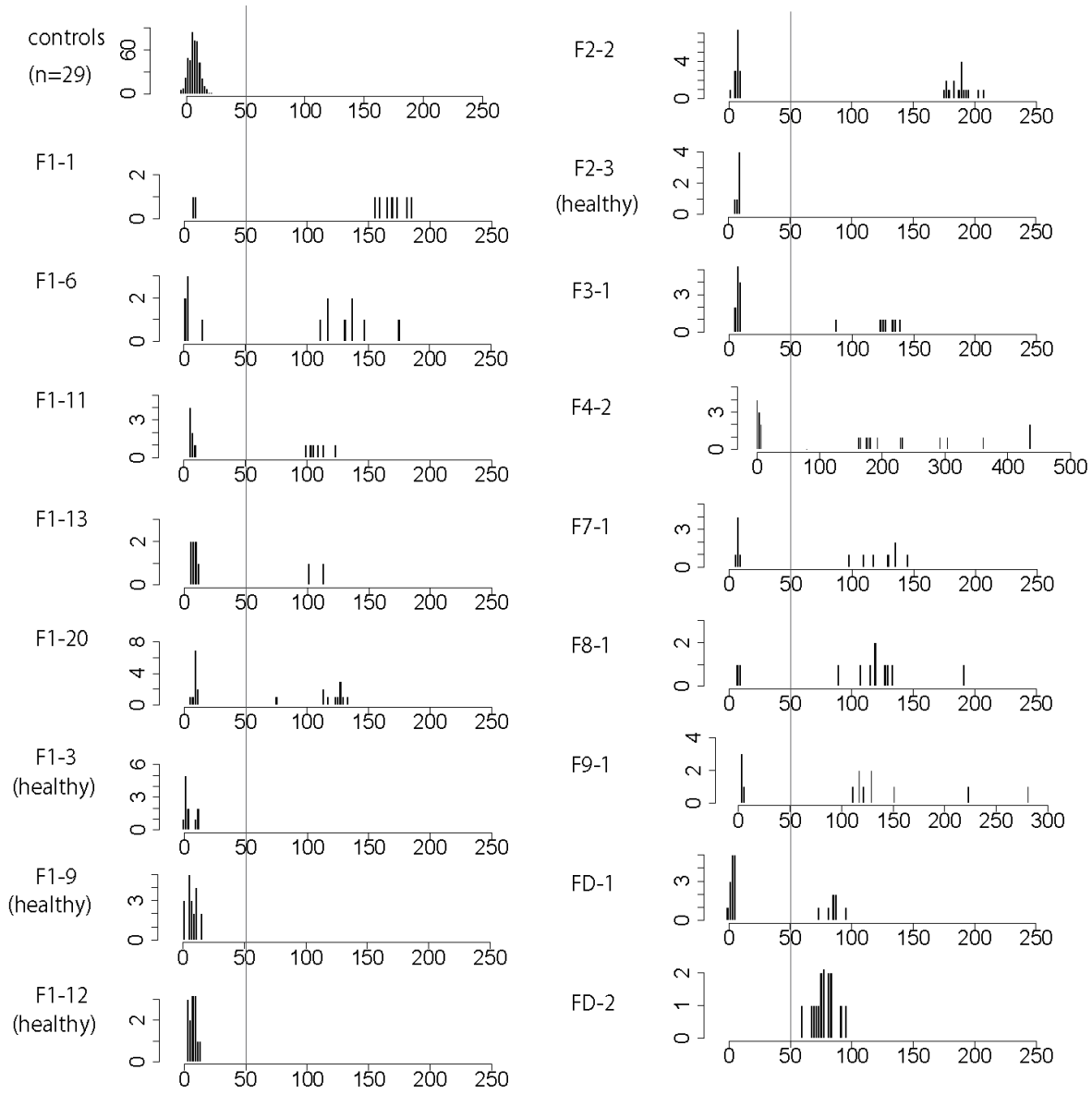
To characterize *NOTCH2NLC* expression in NIID, we performed RNA-seq of fibroblasts from two unaffected (F1-7 and F1-18) and two affected (F1-14 and F1-16) members from Family 1. Each sample had two technical replicates. PCA showed that the technical replicates were closely located in each sample, suggesting that technical variation was small (Supplementary Fig. 8a). *NOTCH2NLC* expression level in patients was not different from control with statistical significance (Fig. 5a). Manual inspection of reads mapped to *NOTCH2NLC* showed reads upstream of exon 1 in patients (Fig. 5b). The reads were from the anti-sense strand of *NOTCH2NLC* and not observed upstream of exon1 in *NOTCH2NL* paralogs (Fig.5c). When multiply mapped reads were visualized, the anti-sense reads extended continuously from the beginning site of or inside the GGC tandem repeat approximately 3.7-kb to a poly-A signal: ATTAAA (chr1: 149,387,121-149,387,126) (Fig. 5d). The results suggested that GGC expansion induced 3.7kb-long antisense transcript possibly starting from the beginning site of or inside the GGC repeat (*NOTCH2NLC-AS*).

Additionally, we characterized dysregulated pathways in NIID. PCA also showed that the samples were not clustered in accordance with the disease status (Supplementary Fig. 8a), implying that the transcriptome-wide effects of NIID were scanty. We performed DEG analysis and identified 54 DEGs (39 upregulated and 15 downregulated genes, Supplementary Fig. 8b). To characterize the 54 DEGs, we analyzed enrichment of BPO and MPO terms among the 54 DEGs. We identified 11 BPO and 5 MPO terms with statistical significance, all of which except one was related to neuronal function (Supplementary Fig. 8c). By clustering the terms sharing the same genes with each other, we identified four BPO clusters including brain development, memory/cognition, and synapse, and one MPO cluster: abnormal memory/cognition (Supplementary Fig. 8d). The results suggested that genes related to neuronal function were dysregulated in NIID. The genes might be downstream targets of GGC expansion/*NOTCH2NLC-AS* (Supplementary Table 5).

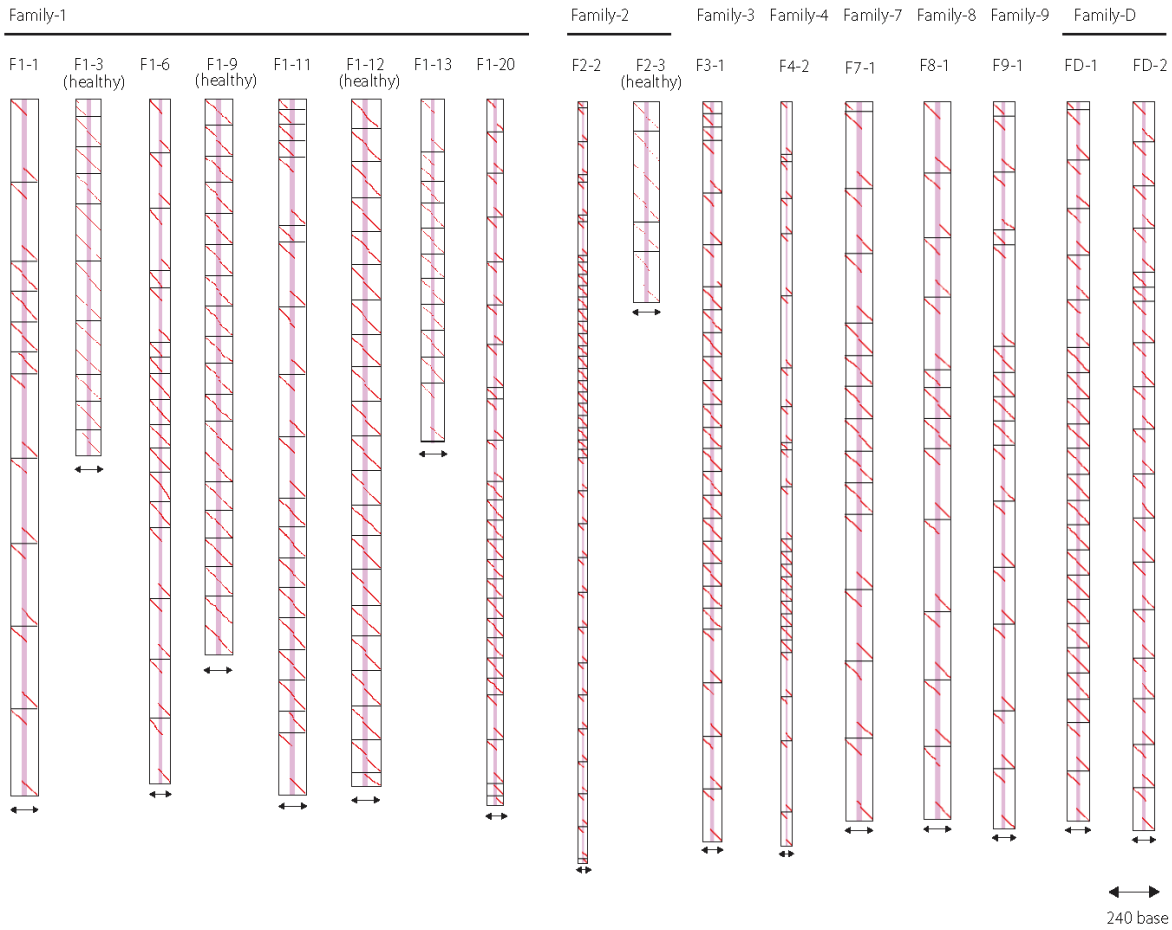
Methylation analysis

It is not clear whether expanded GGC repeats become highly methylated, which might possibly repress *NOTCH2NLC* transcripts from the expansion allele. The 5-methylcytosine (5mC) modification in both expanded and non-expanded alleles was examined from nanopore sequencing signal data, using a recently available nanopore basecaller, `flappie v1.1.0` (https://github.com/nanoporetech/flappie) (Supplementary methods). There was no difference in the rate of 5mC between expanded and non-expanded repeats (Supplementary Fig. 9). Comparing 5mC within the repeat may be difficult because the repeat sequences of expanded and non-expanded alleles are different. Therefore we also examined the CpG island (chr1:149390845-149391541) immediately downstream of the repeat and compared the number of 5mC on the expanded and non-expanded repeat alleles. There is no difference between the alleles (Supplementary Fig. 10). Our 5mC measurement did not support enhanced DNA methylation of the 5'UTR and intron 1 of *NOTCH2NLC* in patients, however, it does not exclude a possibility of allelic repression due to mRNA instability or reduced efficiency of translation due to repeat expansion in the 5'UTR of *NOTCH2NLC*. We also checked the methylation in the CpG islands of *NOTCH2NLC* by Southern analyses with methyl-sensitive restriction enzyme, but could not see any difference between patient and control DNA, supporting the sequencing finding of no specific methylation change (Supplementary Fig. 11).

Supplementary figures

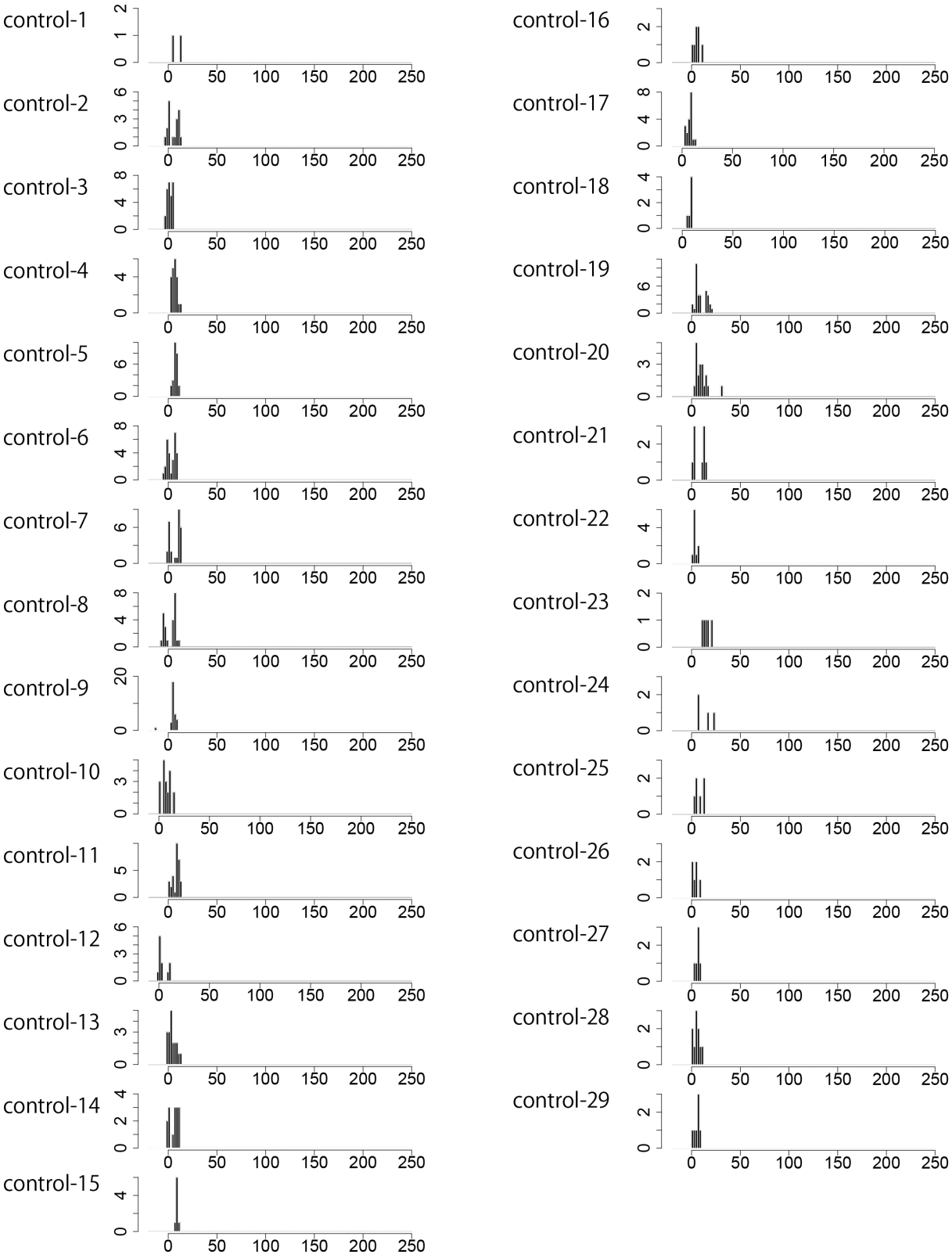


Supplementary Figure 1. Expanded repeats detected by tandem-genotypes tandem-genotypes prediction shows that all affected individuals from 8 families have expanded *NOTCH2NLC* repeat. The *NOTCH2NLC* repeat is at chr1:149390802-149390842, containing $(GGC)_9(GGA)_2(GGC)_2G$ according to simpleRepeat.txt from UCSC. y-axis: read count, x-axis: change in copy number relative to the reference human genome. The +50 copy number change is marked by a vertical line: none of the control and unaffected individuals has repeat expansion beyond this line. Repeat copy number changes from the reference in controls ranges from -13 to 32 (median: 7), and that of NIID has bimodal distribution (predicted normal allele: 0-16 (median: 7) and predicted expanded allele: 60-436 (median: 128)). Note the x-axis ranges in F4-2 and F9-1 are different from others. For controls, combined repeat copy number changes from 29 non-NIID controls are plotted in the same histogram (see Supplementary Fig. 3).

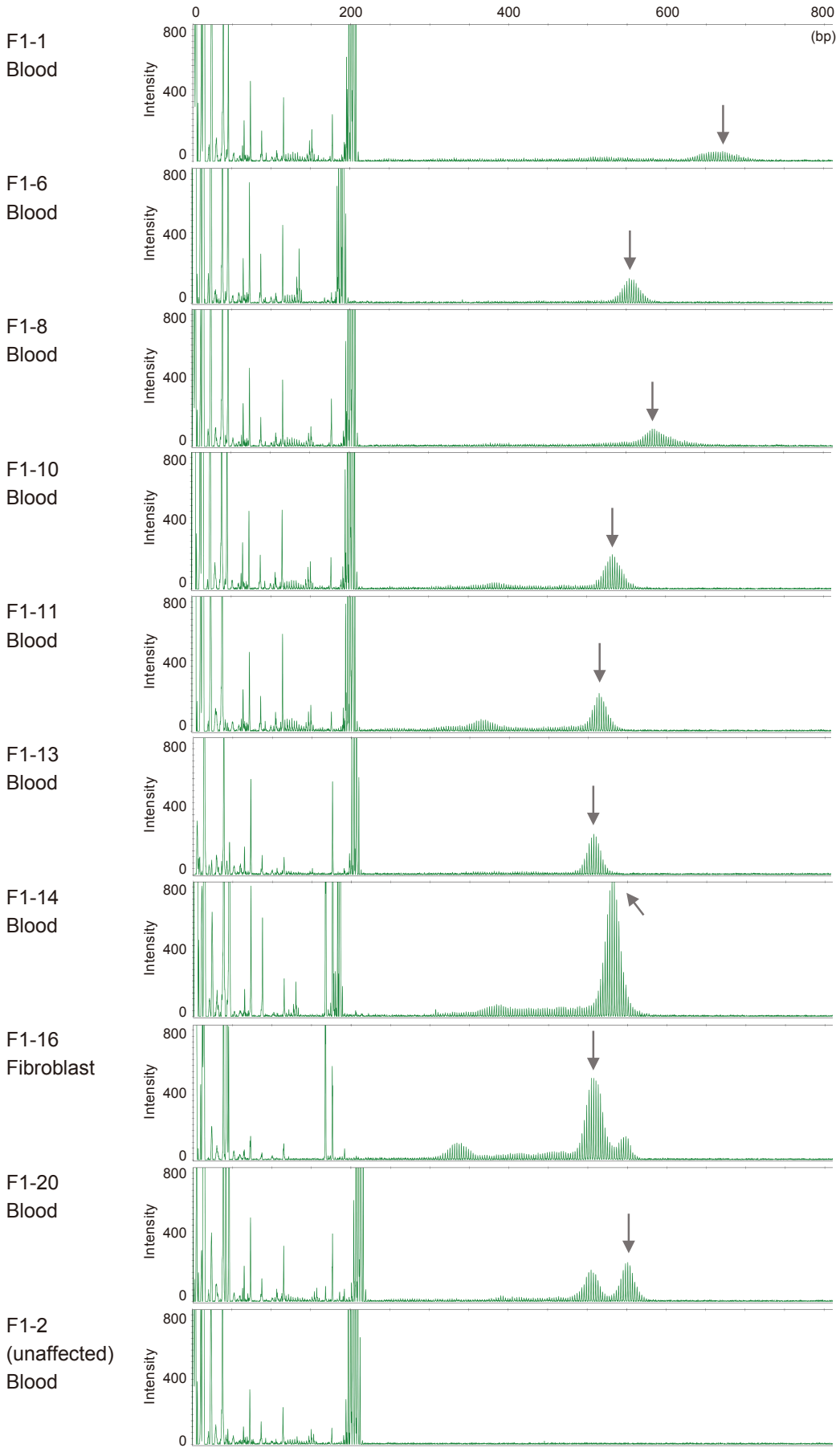


Supplementary Figure 2. Dot-plot presentation of expanded and non-expanded repeats in familial NIID patients

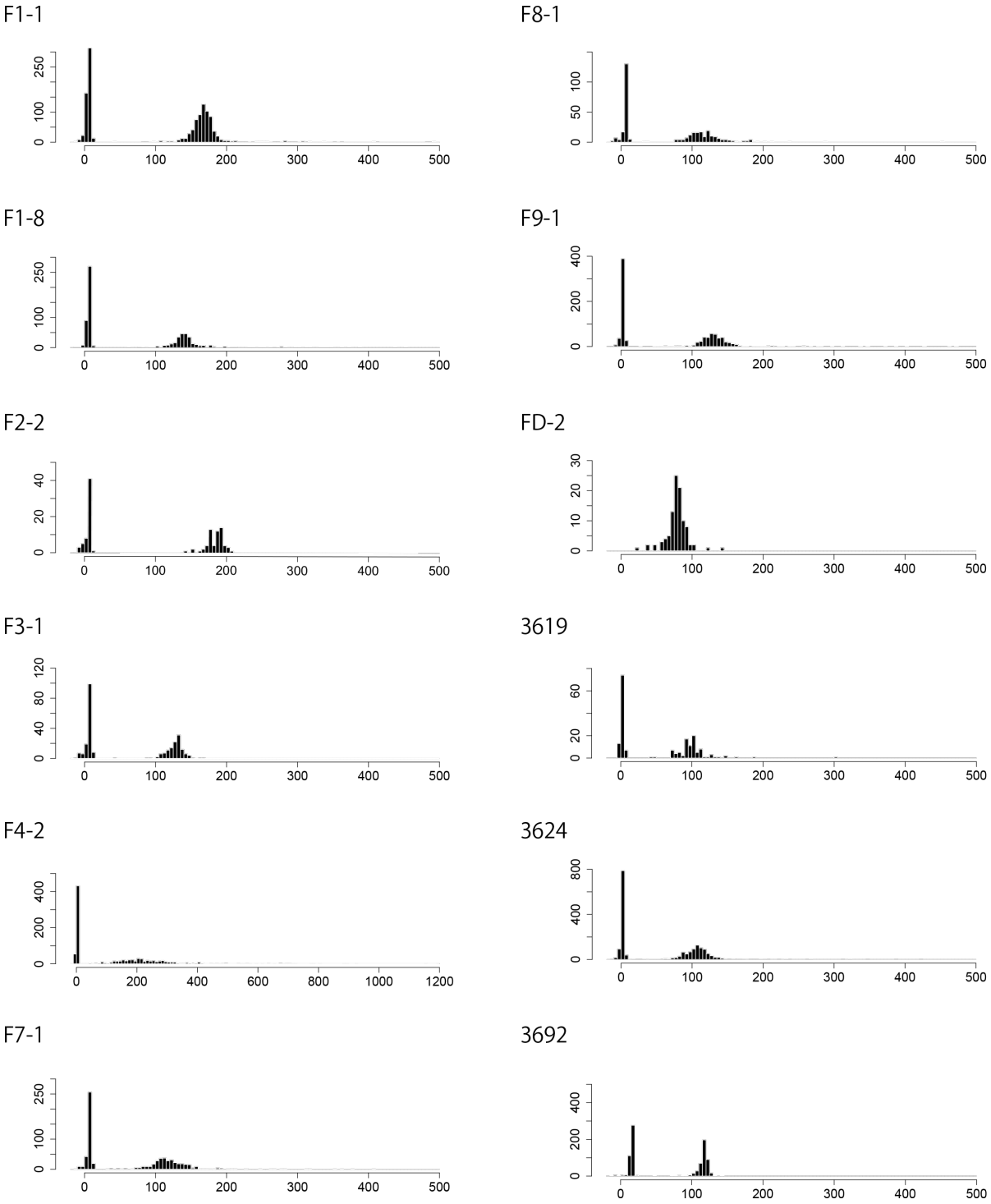
Dot-plot pictures of *NOTCH2NLC* repeat expansions from all members of NIID families, showing alignments of DNA reads (vertical) to the reference human genome (horizontal). Horizontal two-headed arrow shows the length of the reference (240 bases in all dot-plots). The vertical pink-gray line represents the *NOTCH2NLC* repeat (chr1:149390802-149390842). Each dotplot shows one read aligned to the *NOTCH2NLC* repeat with flanking region. A diagonal line was drawn where read sequence aligned to the reference sequence. The horizontal black lines are boundaries between different dotplots showing different reads.



Supplementary Figure 3. GGC repeats in normal controls
`tandem-genotypes` prediction of normal controls in *NOTCH2NLC* repeat. y-axis: read count, x-axis: change in repeat copy number relative to the reference human genome (N=13)


0
200
400
600
800
(bp)
F1-1
Blood
F1-6
Blood
F1-8
Blood
F1-10
Blood
F1-11
Blood
F1-13
Blood
F1-14
Blood
F1-16
Fibroblast
F1-20
Blood
F1-2
(unaffected)
Blood
Intensity
800
400
0

Supplementary Figure 4. Expanded repeat length of *NOTCH2NLC* in affected individuals of family 1 detected by fluorescence amplicon length analysis. Expanded alleles are shown with more than 300 bp in length as unusual peaks. Length of the highest fluorescent peak in expanded alleles was regarded as the expanded allele length (arrows). The entire peaks of non-expanded alleles at around 200 bp were not shown as they were much larger than those of expanded alleles. Non-expanded allele in F1-16 was not seen by amplicon length analysis except for a few peaks likely to be artifacts.

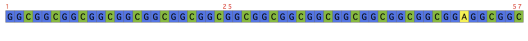


Supplemental Figure 5. tandem-genotypes prediction of Cas9-mediated enrichment sequencing of the *NOTCH2NLC* repeat

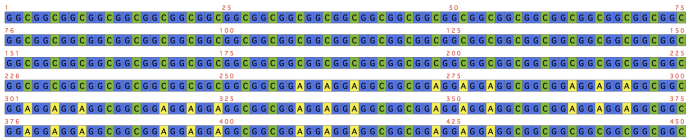
`tandem-genotypes` prediction of Cas9-mediated enrichment sequencing of *NOTCH2NLC* repeat in NIID patients. y-axis: read count, x-axis: change in copy number relative to the reference human genome (N=13).

F1-8*

normal

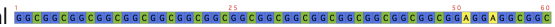


expansion

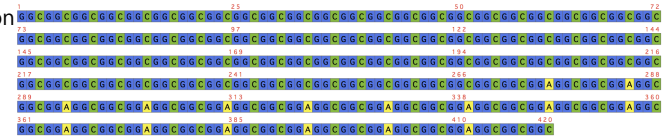


F3-1*

normal

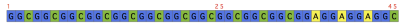


expansion

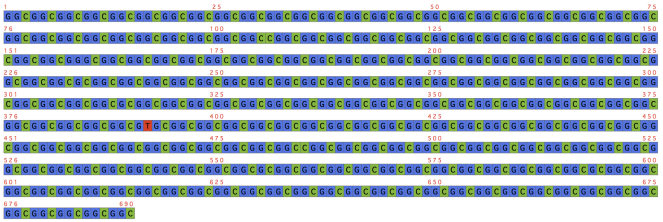


F4-2

normal

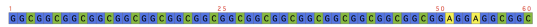


expansion

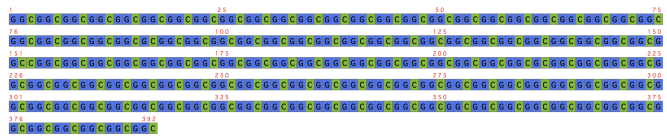


F7-1

normal



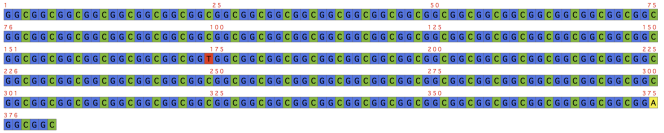
expansion



F8-1

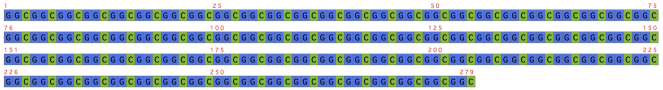
normal

expansion



FD-2

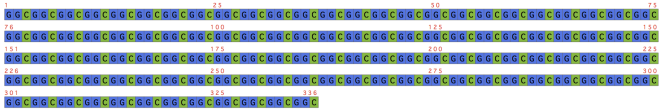
expansion



3619

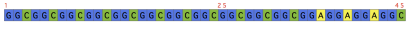
normal

expansion

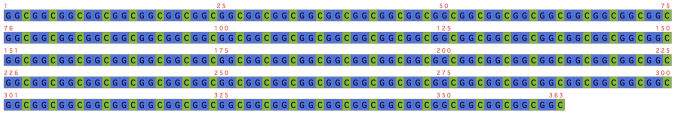


3624

normal



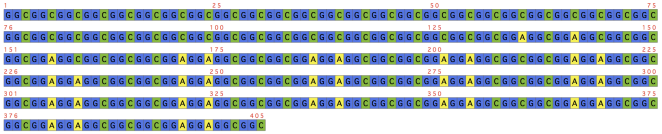
expansion



3692*

normal

expansion

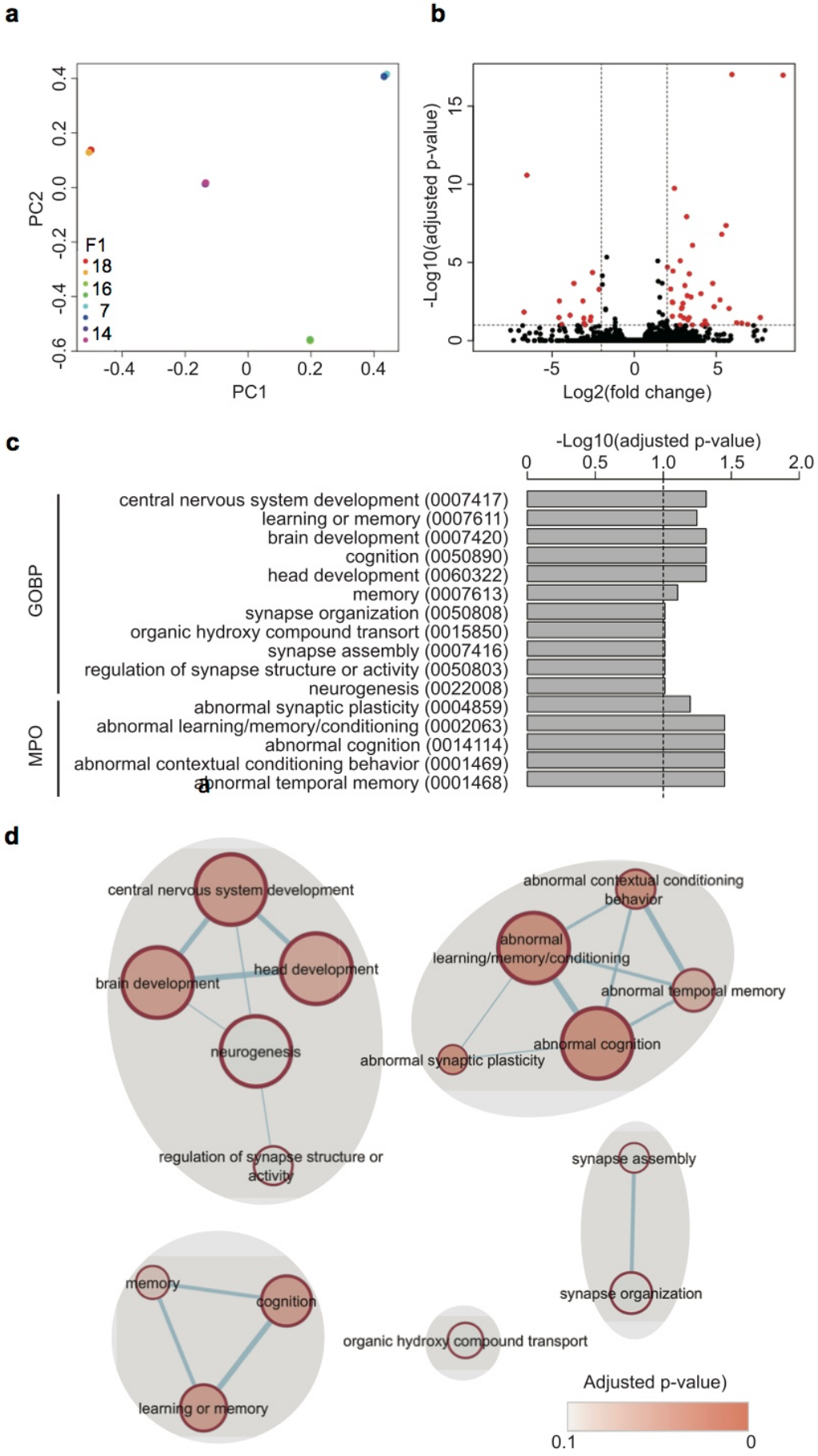


Supplementary Figure 6. Consensus sequence of *NOTCH2NLC* repeat expansion
NOTCH2NLC repeat consensus sequence of multiple alignment using `mafft`. F1-8, F3-1, F4-2, F7-1, F8-1, and FD-2 are familial patients. Patients ID 3619, 3624 and 3629 is are sporadic patients. Individual ID with an asterisk indicates weakness-dominant phenotype in addition to dementia. Individual ID without an asterisk has dementia-dominant phenotype. Even with higher coverage of nanopore sequencing, FD-2 did not have a non-expanded allele. normal: normal allele, expansion: allele with repeat expansion.

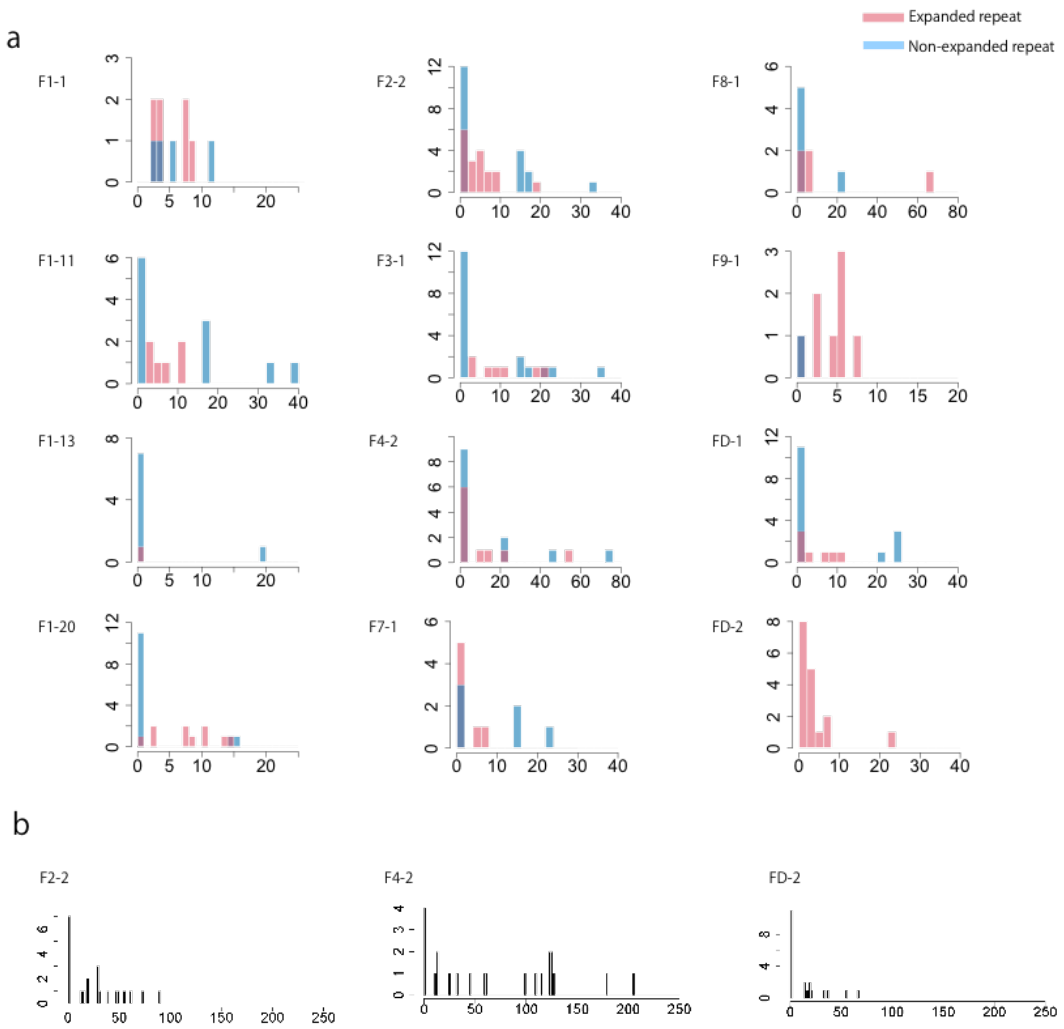
a

b

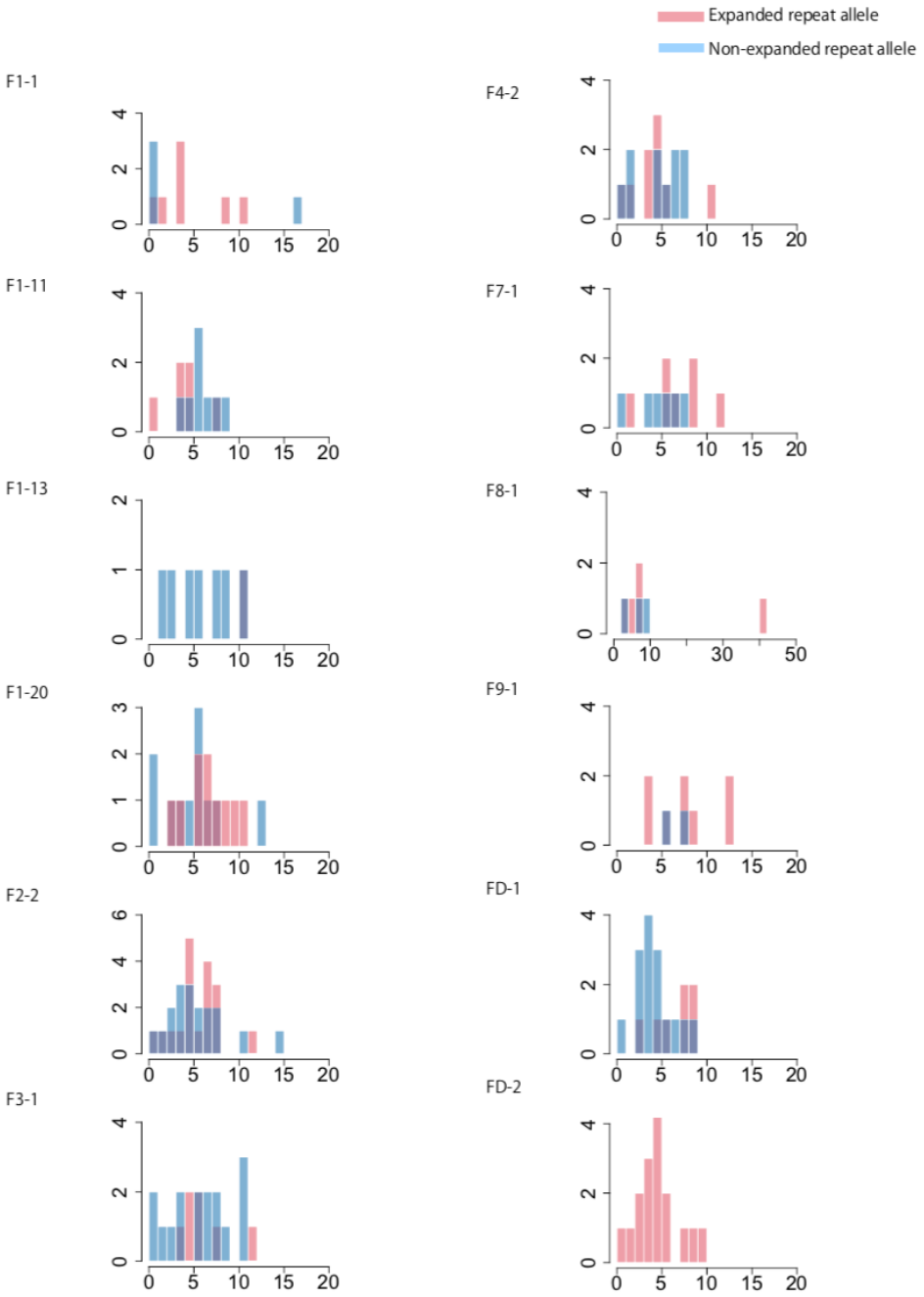
Supplementary Figure 7. Consensus sequence of non-expanded (normal) allele flanking the *NOTCH2NLC* repeat
Two examples of whole Cas9-enrichment target consensus sequences of multiple alignment using mafft. Consensus sequences were aligned to hg38 using LAST. Gaps are only seen in homopolymers besides the *NOTCH2NLC* repeat, supporting the accuracy of the consensus sequences. Note that only normal allele is shown. **a**. individual F1-1. **b**. individual F2-2.


a
b
PC2
PC1
F1
18
16
7
14
-Log10(adjusted p-value)
Log2(fold change)
c
-Log10(adjusted p-value)
GOBP
MPO
central nervous system development (0007417)
learning or memory (0007611)
brain development (0007420)
cognition (0050890)
head development (0060322)
memory (0007613)
synapse organization (0050808)
organic hydroxy compound transport (0015850)
synapse assembly (0007416)
regulation of synapse structure or activity (0050803)
neurogenesis (0022008)
abnormal synaptic plasticity (0004859)
abnormal learning/memory/conditioning (0002063)
abnormal cognition (0014114)
abnormal contextual conditioning behavior (0001469)
abnormal temporal memory (0001468)
d
central nervous system development
brain development
head development
neurogenesis
regulation of synapse structure or activity
abnormal contextual conditioning behavior
abnormal learning/memory/conditioning
abnormal temporal memory
abnormal cognition
abnormal synaptic plasticity
synapse assembly
synapse organization
memory
cognition
learning or memory
organic hydroxy compound transport
Adjusted p-value)

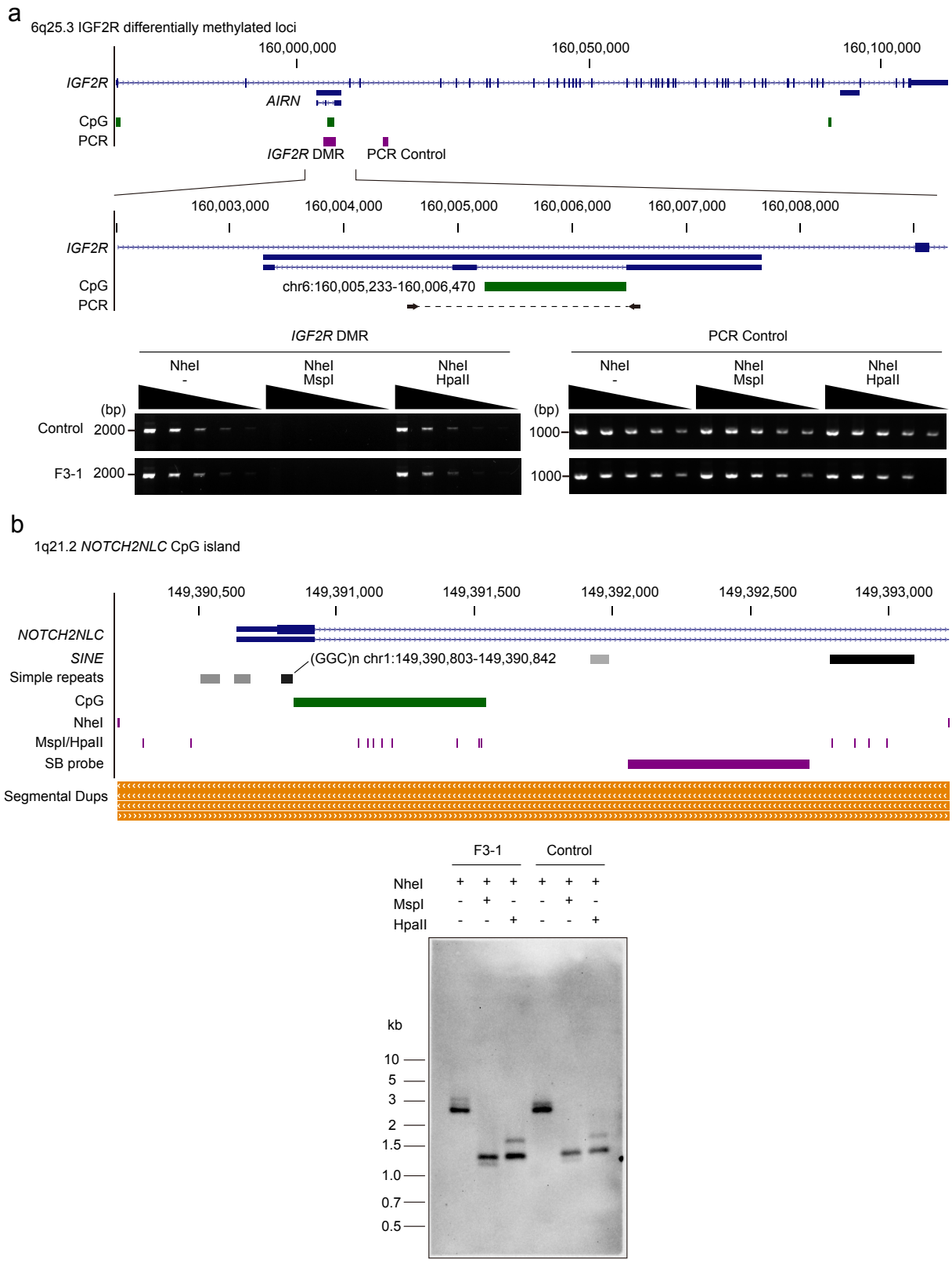
Supplementary Figure 8. RNA-seq analysis of two controls and two patients in Family 1.
a. PCA analysis of gene expression levels of control (N=2) and patient (N=2) samples. Each sample had two technical replicates. The technical replicates were closely located in each sample, suggesting that technical variation was small. **b.** Volcano plot of DEG analysis between control (N=2) and patient (N=2) samples. Thresholds of fold change and statistical significance of differential expression for DEG were shown as dotted lines. The statistical significance was checked with negative binomial generalized model and Wald test using DESeq2. P-value was adjusted using Benjamini-Hochberg method. **c.** GO analysis of DEGs (N=54). Two categories of ontology, biological process ontology (BPO) and mammalian phenotype ontology (MPO), were shown. Threshold for statistical significance was shown as a dotted line. Statistical significance was checked with hypergeometric test using ToppGene Suite. P-value was adjusted using Benjamini-Hochberg method. **d.** Visualization of GO terms sharing genes. Node (white to red circle): GO term; node size: correlated with the number of genes which the GO term contains; node color: correlated with p-value shown as a color bar at the right; edge (blue lines): overlap of contained genes between GO terms. The p-value was calculated in (**c**).



Supplementary Figure 9. The numbers of 5mC in *NOTCH2NLC* repeat.
a. The numbers of 5mC in *NOTCH2NLC* repeats (chr1:149390802-149390842) in both expanded and non-expanded repeats are shown (pale pink: expanded, pale blue: non-expanded). There is no difference between expanded and non-expanded repeats in 5mC modification of cytosine in any of the NIID patients. y-axis: read count, x-axis: number of flappie-called 5mC in 1,000 bases. **b.** Imprinted region in chr15 CpG island [1] from three individuals with largest datasets (F2-2, F4-2 and FD-2) are shown as positive controls for flappie 5mC methylation detection. This region of the maternal allele should be methylated. As expected, it shows some tendency to bimodal distribution.

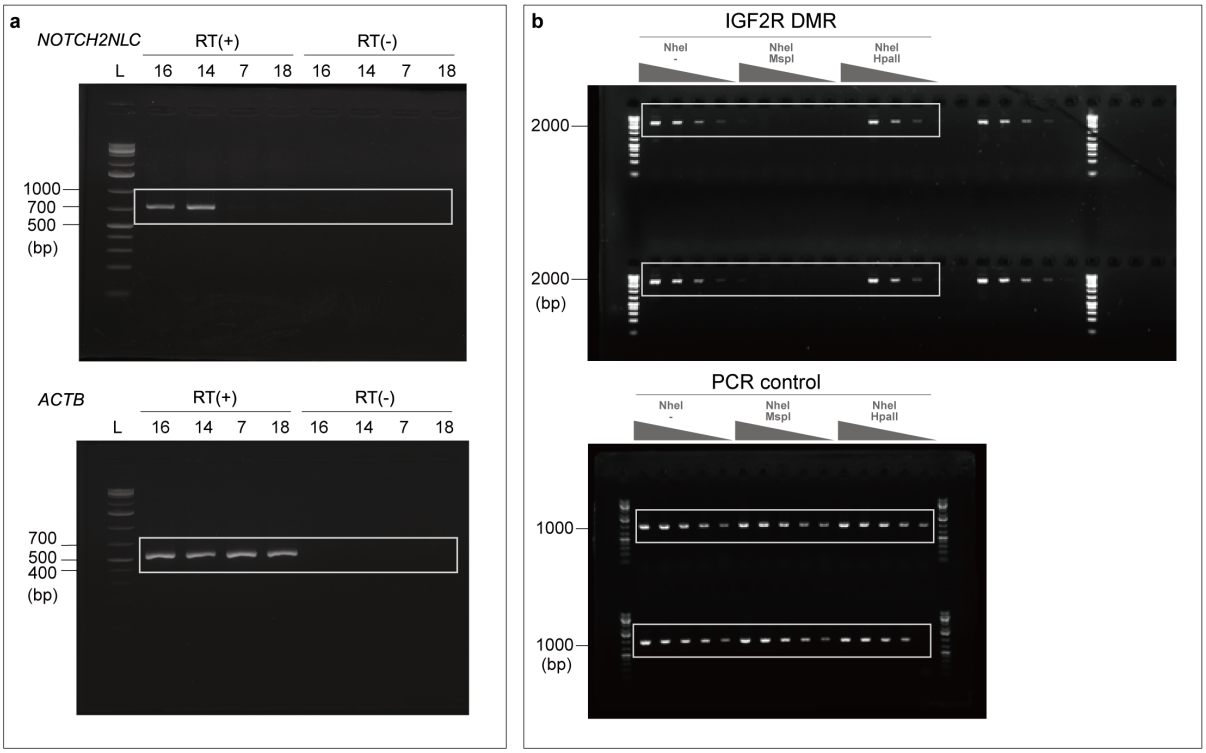


Supplementary Figure 10. The numbers of 5mC downstream of *NOTCH2NLC* repeat.
The numbers of 5mC modifications in a CpG island (chr1:149390845-149391541), which is downstream of the *NOTCH2NLC* repeat. The CpG island methylation of expanded and non-expanded repeat alleles (chr1:149390802-149390842) is shown separately. There is no difference between expanded and non-expanded repeat alleles in 5mC modification in any of the NIID patients. Pale pink: CpG on expanded allele, pale blue: CpG on non-expanded allele. y-axis: read count, x-axis: number of flappie-called 5mC in 1,000 bases.



Supplementary Figure 11. Analysis of DNA methylation status at *IGF2R* and *NOTCH2NLC* CpG islands.
Genomic DNA from F3-1 and control individuals were initially digested with NheI followed by MspI or HpaII that are isoschizomers with different sensitivity to CpG methylation. Digested DNA was subjected to semi-quantitative PCR (end-point PCR) and Southern blotting analysis for *IGF2R* and *NOTCH2NLC* loci, respectively.

a. Genomic positions of two sets of PCR primers, *IGF2R* DMR (target) and PCR control (control), relative to the known differentially methylated CpG island are shown. Threefold serial dilution of the digested DNA was used as a PCR template. PCR products after 33 PCR cycles were analyzed using 1% agarose gels with ethidium bromide staining. PCR amplicon in MspI digestion was undetectable because MspI cleaved both unmethylated and methylated DNA of CpG island. Amplification efficiency decreased in HpaII digestion as compared with that of uncut DNA (NheI), supporting differential methylation at *IGF2R* locus as described previously[2]. Note that input DNA were approximately equal, as suggested by PCR amplification at control region with no MspI/HpaII sites (PCR control). PCR was performed twice for technical replicates. **b.** Genomic positions of NheI, MspI/HpaII, Southern blotting probe and CpG island are shown relative to the (GGC) expansion variant. NheI fragment were completely but differently digested by either MspI or HpaII, suggesting the presence of methylated CpG within NheI fragment. Nonetheless, F3-1 showed no difference in digestion pattern compared to that of control individual. Due to the limitation in DNA availability for a patient, experiment could be performed only once.



Supplementary Figure 12. Full scan of gel images of Fig. 5e and Supplementary Fig. 11a. a. Full scan gel imanges of Fig. 5e. b. Full scan gel images of Supplementary Fig. 11a. Cropped images are shown with gray squares.

Supplementary Tables

individual ID	disease status	sequencer	number of reads	total base	Median	average read length	predicted coverage	total number of tandem-genotypes call	number of read with repeat expansion	prioritization with single data	prioritization with 4 control data
F1-1	affected	PromethION	2,112,375	19,006,257,320	3520	8,998	6	9	7	1	1
F1-3	unaffected	PromethION	23,303,818	54,535,886,940	691	2,340	17	11	0	NA	NA
F1-6	affected	PacBio RSII	9,064,262	74,975,739,315	7701	8,272	23	14	8	1	1
F1-9	unaffected	PromethION	9,635,261	35,004,315,947	2452	3,633	11	19	0	NA	NA
F1-11	affected	PromethION	10,589,493	57,885,853,963	3298	5,466	18	13	6	3	1
F1-12	unaffected	PromethION	9,091,759	63,874,106,363	5889	7,026	19	19	0	NA	NA
F1-13	affected	PromethION	4,790,029	27,954,971,034	4364	5,836	8	9	2	3	1
F1-20	affected	PromethION	12,830,261	57,451,863,375	1558	4,478	17	22	11	3	1
F2-2	affected	PromethION	15,926,839	82,870,233,437	3520	5,203	25	32	16	2	1
F2-3	unaffected	PromethION	18,339,946	23,268,659,841	865	1,269	7	6	0	NA	NA
F3-1	affected	PromethION	11,109,622	62,195,375,599	2990	5,598	19	21	6	2	2
F4-2	affected	PromethION	10,658,181	72,698,332,750	3736	6,821	22	22	13	1	1
F7-1	affected	PromethION	5,529,417	38,773,128,508	3169	7,012	12	13	7	1	1
F8-1	affected	PromethION	17,969,232	49,838,565,927	1608	2,774	15	11	9	2	1
F9-1	affected	PromethION	6,786,707	55,398,873,781	3520	8,163	17	13	9	1	1
FD-1	affected	PromethION	11,467,114	56,214,013,607	3110	4,902	17	21	7	5	5
FD-2	affected	PromethION	13,655,084	78,995,795,042	3861	5,785	24	16	16	5	3

Supplementary Table 1. Summary of dataset information and tandem-genotypes results. Prioritization ranks of tandem-genotypes for the *NOTCH2NLC* repeat out of all the one million tandem repeats throughout the whole genome are also shown. Four control datasets were used for prioritization are data 3, 19 (public data) and F1-3, -9 (unaffected family members).

data	sequencer	total number of tandem-genotypes call at NOTCH2NLC repeat	number of reads	total base	average length
1	MinION	2	2,239,370	18,159,653,811	8,109
2	MinION	18	5,246,726	45,134,820,953	8,603
3	MinION	27	14,183,584	91,240,120,433	6,433
4	PromethION	21	6,003,887	46,621,512,542	7,765
5	PromethION	25	3,394,312	43,196,945,412	12,726
6	PromethION	28	5,074,319	96,709,785,254	19,059
7	PromethION	28	4,497,556	90,866,959,081	20,204
8	PromethION	24	4,403,236	81,554,440,718	18,522
9	PromethION	32	8,642,604	111,990,048,861	12,958
10	PromethION	19	9,635,261	35,004,315,947	3,633
11	PromethION	30	7,136,355	86,634,317,185	12,140
12	PromethION	11	23,303,818	54,535,886,940	2,340
13	PromethION	19	7,824,636	77,878,925,289	9,953
14	PromethION	15	3,557,359	53,696,135,713	15,094
15	PromethION	8	1,885,376	19,822,791,751	10,514
16	PromethION	7	2,981,694	22,136,031,084	7,424
17	PromethION	19	9,091,759	63,874,106,363	7,026
18	PromethION	6	18,339,946	23,268,659,841	1,269
19	PromethION	34	13,134,890	185,056,318,464	14,089
20	Pacbio Sequel	19	6,174,383	44,835,599,221	7,262
21	Pacbio Sequel	9	4,029,436	39,348,377,910	9,765
22	Pacbio Sequel	10	3,314,492	30,665,693,222	9,252
23	Pacbio Sequel	5	2,218,415	21,275,114,124	9,590
24	Pacbio Sequel	4	2,874,747	25,207,588,869	8,769
25	Pacbio Sequel	6	1,929,367	20,280,159,023	10,511
26	Pacbio Sequel	6	2,705,416	25,484,894,848	9,420
27	Pacbio Sequel	6	8,391,859	56,100,283,182	6,685
28	Pacbio Sequel	10	3,228,957	30,151,182,772	9,338
29	Pacbio Sequel	7	3,650,569	30,427,397,011	8,335

Supplementary Table 2. Sequencing summary of control datasets. Data 3 and 19 are publicly available, rel3 [3] and ERR2585112-5 [4], respectively.

Family No.	Individual No.	DNA	Disease status	Long read seq.	RP-PCR	AL analysis (repeat number) [a]		Individual No.	
						Non-expanded	Expanded	Father	Mother
1	1	Blood	Affected	Expansion	Expansion	6	162		
1	2	Blood	Unaffected	N.P.	No expansion	6, 8	none		
1	3	Blood	Unaffected	No expansion	No expansion	7	none		
1	6	Blood	Affected	Expansion	Expansion	2	123		
		Brain		N.P.	N.P.	2	122		
		Cerebellum		N.P.	N.P.	2	122		
		Pancrease		N.P.	N.P.	2	122		
1	7	Fibroblast	Unaffected	N.P.	No expansion	6	none		
1	8	Blood	Affected	N.P.	Expansion	6	133		
		Brain		N.P.	N.P.	6	132		
1	9	Blood	Unaffected	No expansion	No expansion	2, 9	none		
1	10	Blood	Affected	N.P.	Expansion	6	66, 116 [b]		
1	11	Blood	Affected	Expansion	Expansion	6	60, 110 [b]	1	2
1	12	Blood	Unaffected	No expansion	No expansion	6, 8	none	1	2
1	13	Blood	Affected	Expansion	Expansion	7	107		3
1	14	Blood	Affected	N.P.	Expansion	0	66, 115 [b]		3
		Fibroblast		N.P.	Expansion	0	68, 117 [b]		3
1	16	Fibroblast	Affected	N.P.	Expansion	none	49, 100 [b], 121	6	7
1	17	Blood	Unaffected	N.P.	No expansion	2	none	6	7
1	18	Fibroblast	Unaffected	N.P.	No expansion	2	none	6	7
1	20	Blood	Affected	Expansion	Expansion	9	106, 122 [b]	8	9
2	1	Blood	Unaffected	N.P.	No expansion	5, 7	none		
2	2	Blood	Affected	Expansion	Expansion	7	183		1
2	3	Blood	Unaffected	No expansion	No expansion	7	none		1
3	1	Blood	Affected	Expansion	Expansion	7	127		
4	1	Blood	Affected	N.P.	Expansion	7	96		
4	2	Blood	Affected	Expansion	Expansion	2	none		
4	3	Blood	Affected	N.P.	Expansion	7	103		
4	4	Blood	Affected	N.P.	Expansion	7	100		
7	1	Blood	Affected	Expansion	Expansion	7	105		
7	2	Blood	Affected	N.P.	Expansion	7	none		
8	1	Blood	Affected	Expansion	Expansion	7	90, 106 [b]		
8	2	Blood	Affected	N.P.	Expansion	7	85		
9	1	Blood	Affected	Expansion	Expansion	2	121		
9	2	Blood	Affected	N.P.	Expansion	2	112		
10	1	Blood	Affected	N.P.	Expansion	7	129		
D	1	Blood	Affected	Expansion	Expansion	2	80		
D	2	Blood	Affected	Expansion	Expansion	none	60, 76 [b]		

Supplementary Table 3. Comparison of long read sequencing, RP-PCR and amplicon length analysis results in 9 NIID families.

[a]Change in repeat copy number relative to the reference human genome. [b]Highest peaks of expanded alleles in fluorescent amplicon length (AL) analysis. N.P. Not performed.

Individual No.	Disease status	Long read seq.	RP-PCR	AL-analysis (repeat number)[a]	
				Non-expanded	Expanded
3614	Affected	N.P.	Expansion	14	108
3615	Affected	N.P.	Expansion	10	86
3616	Affected	N.P.	Expansion	7	105
3618	Affected	N.P.	Expansion	9	123
3619	Affected	N.P.	Expansion	2	73, 93[b]
3623	Affected	N.P.	Expansion	-3	85
3624	Affected	N.P.	Expansion	2	86, 106[b]
3625	Affected	N.P.	Expansion	9	89
3626	Affected	N.P.	Expansion	13	96
3627	Affected	N.P.	Expansion	2	98
3628	Affected	N.P.	Expansion	13	81
3630	Affected	N.P.	Expansion	6	112
3636	Affected	N.P.	Expansion	1	121
3639	Affected	N.P.	Expansion	7	115
3644	Affected	N.P.	Expansion	13	104
3645	Affected	N.P.	Expansion	none	77
3646	Affected	N.P.	Expansion	2	92
3647	Affected	N.P.	Expansion	15	71
3648	Affected	N.P.	Expansion	5	109
3650	Affected	N.P.	Expansion	9	112
3651	Affected	N.P.	Expansion	7	none
3652	Affected	N.P.	Expansion	11	108
3653	Affected	N.P.	Expansion	7	82
3654	Affected	N.P.	Expansion	11	104
3658	Affected	N.P.	Expansion	7	91
3661	Affected	N.P.	Expansion	7	80
3662	Affected	N.P.	Expansion	11	82
3663	Affected	N.P.	Expansion	12	117
3665	Affected	N.P.	Expansion	6	74, 101[b]
3667	Affected	N.P.	Expansion	8	106
3668	Affected	N.P.	Expansion	13	92
3672	Affected	N.P.	Expansion	9	109
3673	Affected	N.P.	Expansion	11	106
3676	Affected	N.P.	Expansion	6	108
3677	Affected	N.P.	Expansion	9	86
3680	Affected	N.P.	Expansion	9	none
3683	Affected	N.P.	Expansion	2	104
3687	Affected	N.P.	Expansion	2	99
3692	Affected	N.P.	Expansion	17	119
S81	Affected	N.P.	Expansion	7	97[b], 125
Father of 3661	Unaffected	N.P.	No expansion	6	none
Mother of 3661	Unaffected	N.P.	No expansion	7	none

Supplementary Table 4. RP-PCR and amplicon length analysis results in 40 sporadic NIID cases.

[a]Change in repeat copy number relative to the reference human genome. [b]Highest peaks of expanded alleles in fluorescent amplicon length (AL) analysis. N.P. Not performed.

Ontology ID	Description	P-value	FDR	DEGs
GO:0007417	central nervous system development	0.0000373	0.0481	PCP4,NEFL,HOXB3,RELN,DRD1,LMX1B,WNT2,ITGA8,NCAM1,SHC3,TFAP2C
GO:0007611	learning or memory	0.000066	0.0481	RELN,DRD1,LMX1B,ITGA8,NCAM1,SHC3
GO:0007420	brain development	0.000114	0.0481	NEFL,HOXB3,RELN,DRD1,LMX1B,WNT2,ITGA8,NCAM1,TFAP2C
GO:0050890	cognition	0.000122	0.0481	RELN,DRD1,LMX1B,ITGA8,NCAM1,SHC3
GO:0060322	head development	0.00018	0.0569	NEFL,HOXB3,RELN,DRD1,LMX1B,WNT2,ITGA8,NCAM1,TFAP2C
GO:0007613	memory	0.0003	0.0788	RELN,DRD1,LMX1B,ITGA8
GO:0050808	synapse organization	0.000517	0.0976	RELN,LRRN3,DRD1,LMX1B,AMIGO2
GO:0015850	organic hydroxy compound transport	0.000554	0.0976	PCP4,DRD1,EMB,CD36,SYT7
GO:0007416	synapse assembly	0.000568	0.0976	RELN,LRRN3,DRD1,AMIGO2
GO:0050803	regulation of synapse structure or activity	0.000657	0.0976	RELN,LRRN3,DRD1,NCAM1,AMIGO2
GO:0022008	neurogenesis	0.000681	0.0976	PCP4,NEFL,HOXB3,RELN,DRD1,LMX1B,KIF26B,MARK1,WNT2,NCAM1,TFAP2C,IL33
MP:0004859	abnormal synaptic plasticity	5.25E-05	3.54E-02	PCP4,DRD1,ITGA8,ADD2
MP:0002063	abnormal learning/memory/conditioning	1.06E-04	3.54E-02	PCP4,RELN,DRD1,SORBS2,LMX1B,ITGA8,NCAM1,CHRM3,ADD2
MP:0014114	abnormal cognition	1.07E-04	3.54E-02	PCP4,RELN,DRD1,SORBS2,LMX1B,ITGA8,NCAM1,CHRM3,ADD2
MP:0001469	abnormal contextual conditioning behavior	1.21E-04	3.54E-02	DRD1,SORBS2,LMX1B,CHRM3,ADD2
MP:0001468	abnormal temporal memory	2.69E-04	6.31E-02	DRD1,SORBS2,LMX1B,CHRM3,ADD2

Supplementary Table 5. Enriched terms among 54 DEGs
Detailed description of GO analysis of DEGs (n=54) in Supplementary Fig. 8c. BPO and MPO terms enriched among DEGs were shown. Statistical significance was analyzed with hypergeometric test using ToppGene Suite. P-value was adjusted using Benjamini-Hochberg method.

Experiments	Primer name	Sequences
Repeat-primed PCR	NOTCH2NLC-RP-F	5'-FAM-GGCATTTGCGCCTGTGCTTCGGACCGT-3'
	M13-$(GGC)_4(GGA)_2$-R	5'-CAGGAAACAGCTATGACCTCCTCCGCCGCCGCCGCC-3'
	M13-linker-R	5'-CAGGAAACAGCTATGACC-3'
Fluorescence amplicon length analysis	NOTCH2NLC-AL-F	5'-VIC-CATTTGCGCCTGTGCTTCGGAC-3'
	NOTCH2NLC-AL-R	5'-AGAGCGGCGCAGGGCGGGCATCTT-3'
Cas9-mediated enrichment*	Alt-R™ CRISPR-Cas9 crRNA 1	5'-UUCUUAGCCCACUUGUACCCAGG-3'
	Alt-R™ CRISPR-Cas9 crRNA 2	5'-GGAGCACUCAAAAGUUUAGAAGG-3'
cDNA amplification of *NOTCH2NLC-AS*	Forward	5'-ATAAAGAAGCGGCAGTGGAA-3'
	Reverse	5'-AATCGCAGAGGGAGACAAAG-3'
cDNA amplification of *ACTB*	Forward	5'-GTGGGGCGCCCCAGGCACCA-3'
	Reverse	5'-CTCCTTAATGTCACGCACGATTTC-3'
Semi-quantitative PCR at IGF2R DMR	Forward	5'-GGGTGGGTCTAGTGAGTGT-3'
	Reverse	5'- AAAGTCAGCCACGATCATCC-3'
PCR control without MspI/HpaII sites at IGF2R region	Forward	5'-GCAGGATAATGGGCACTTGT-3'
	Reverse	5'-TGCTCAGATGGTGAAACTGC-3'
Digoxigenin-labeled DNA probe for *NOTCH2NLC*	Forward	5'-AACGGATGACACTCCAAAGG-3'
	Reverse	5'-TCCTGCTTCATAGGTGAAGAGAC-3'

Supplementary Table 6. Oligonucleotide and guide RNA sequences
*: guide RNA sequences for Cas9 mediated enrichment

Supplementary References

1. Sharp, A.J. *et al.* Methylation profiling in individuals with uniparental disomy identifies novel differentially methylated regions on chromosome 15. *Genome Res* **20**, 1271-8 (2010).
2. Hayward, B.E. *et al.* The human GNAS1 gene is imprinted and encodes distinct paternally and biallelically expressed G proteins. *Proc Natl Acad Sci U S A* **95**, 10038-43 (1998).
3. Jain, M. *et al.* Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nat Biotechnol* **36**, 338-345 (2018).
4. de Coster, W. et al. Structural variants identified by Oxford Nanopore PromethION sequencing of the human genome. Genome Research (2019). *in press*. (doi: 10.1101/gr.244939.118).