Science Advances
AAAS

Supplementary Materials for

Identification of region-specific gene isoforms in the human brain using long-read transcriptome sequencing

Mihoko Shimada *et al.*

Corresponding author: Mihoko Shimada, mihshimada@ri.ncgm.go.jp

Sci. Adv. **10**, eadj5279 (2024)
DOI: 10.1126/sciadv.adj5279

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

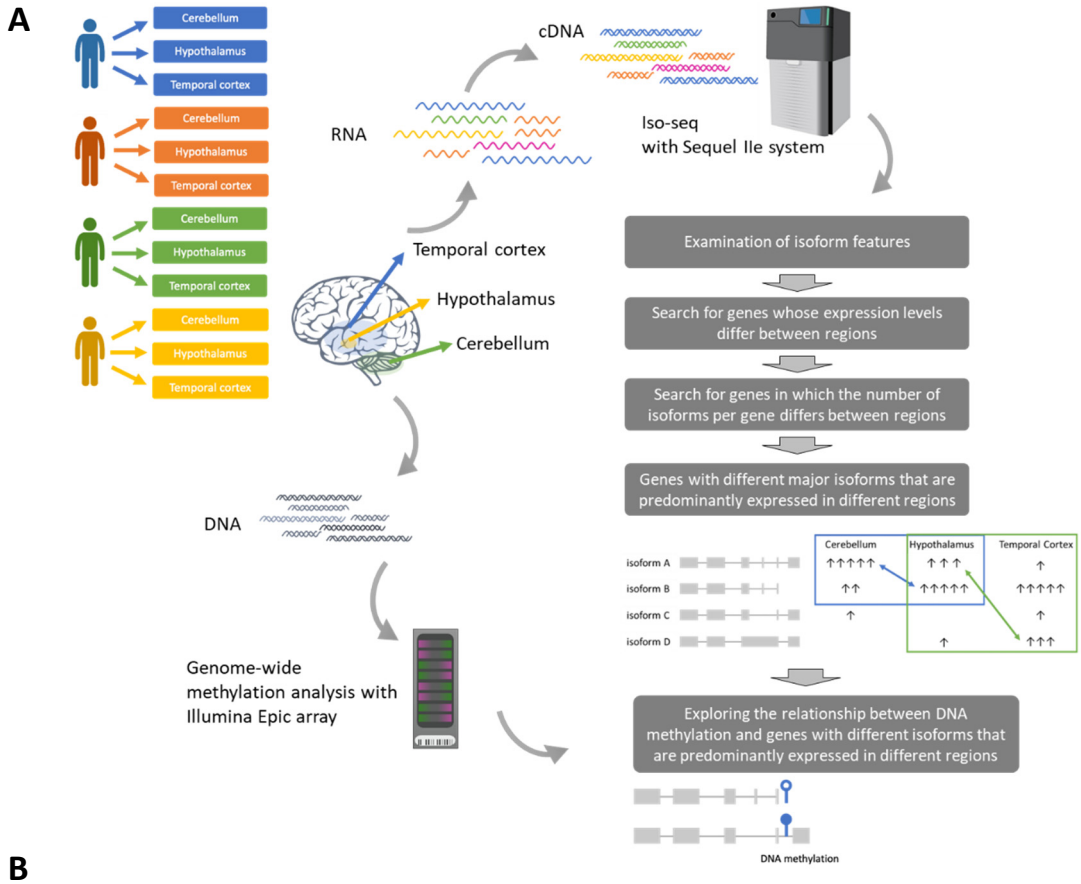


Fig. S1. Graphical abstract and analytical flow of Iso-seq data. A Tissues were excised from three brain regions, cerebellum (Ce), hypothalamus (Hy), and temporal cortex (Tc) from the same samples, RNA was extracted and Iso-seq was performed. DNA methylation was evaluated using DNA from the same tissues. **B** Analysis flow of Iso-seq data and software used.

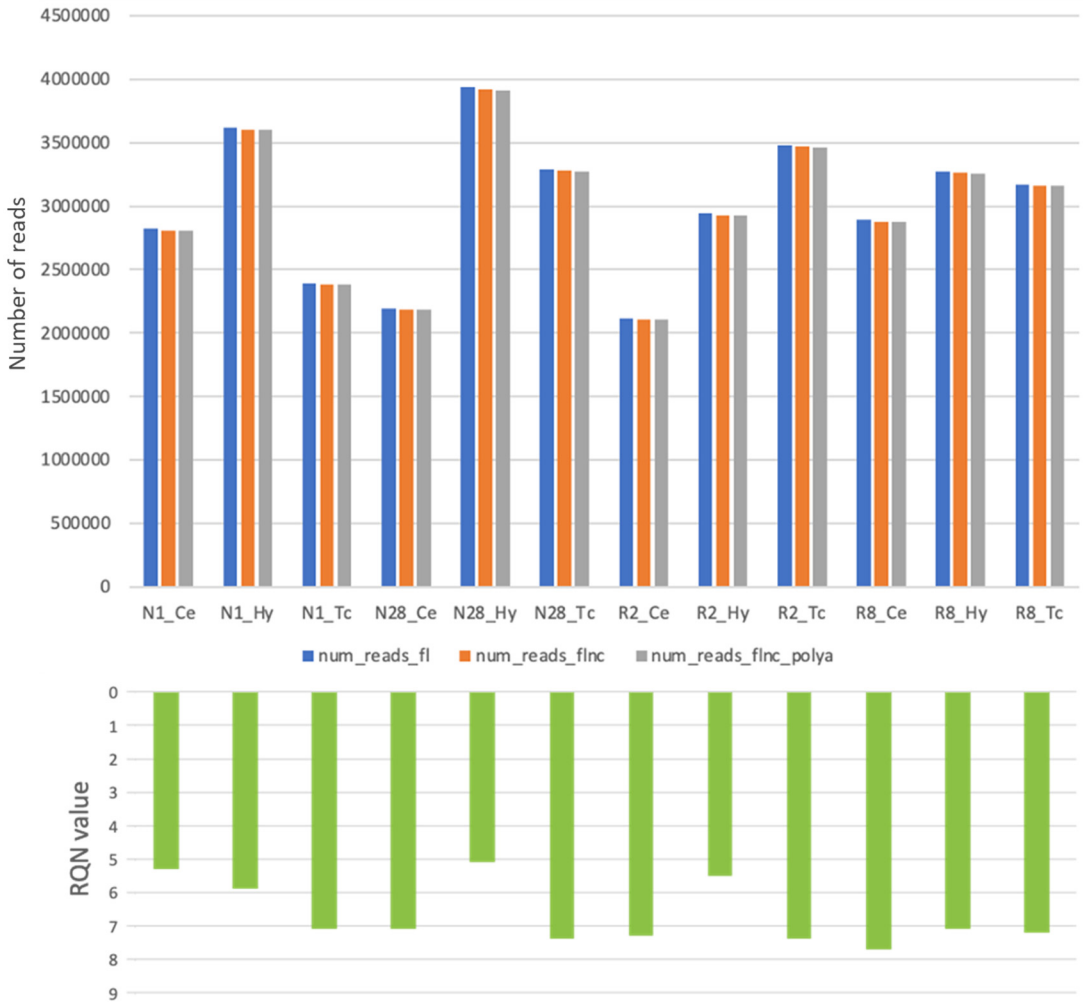


Fig. S2. Quality check results for Iso-seq raw data. Numbers of full-length (fl) reads, full-length non-concatemer (flnc) reads, and flnc reads with polyA tails. The relationship between the number of each read and the RQN value for RNA quality is also shown; no relationship is observed between the RQN and the number of detected reads.

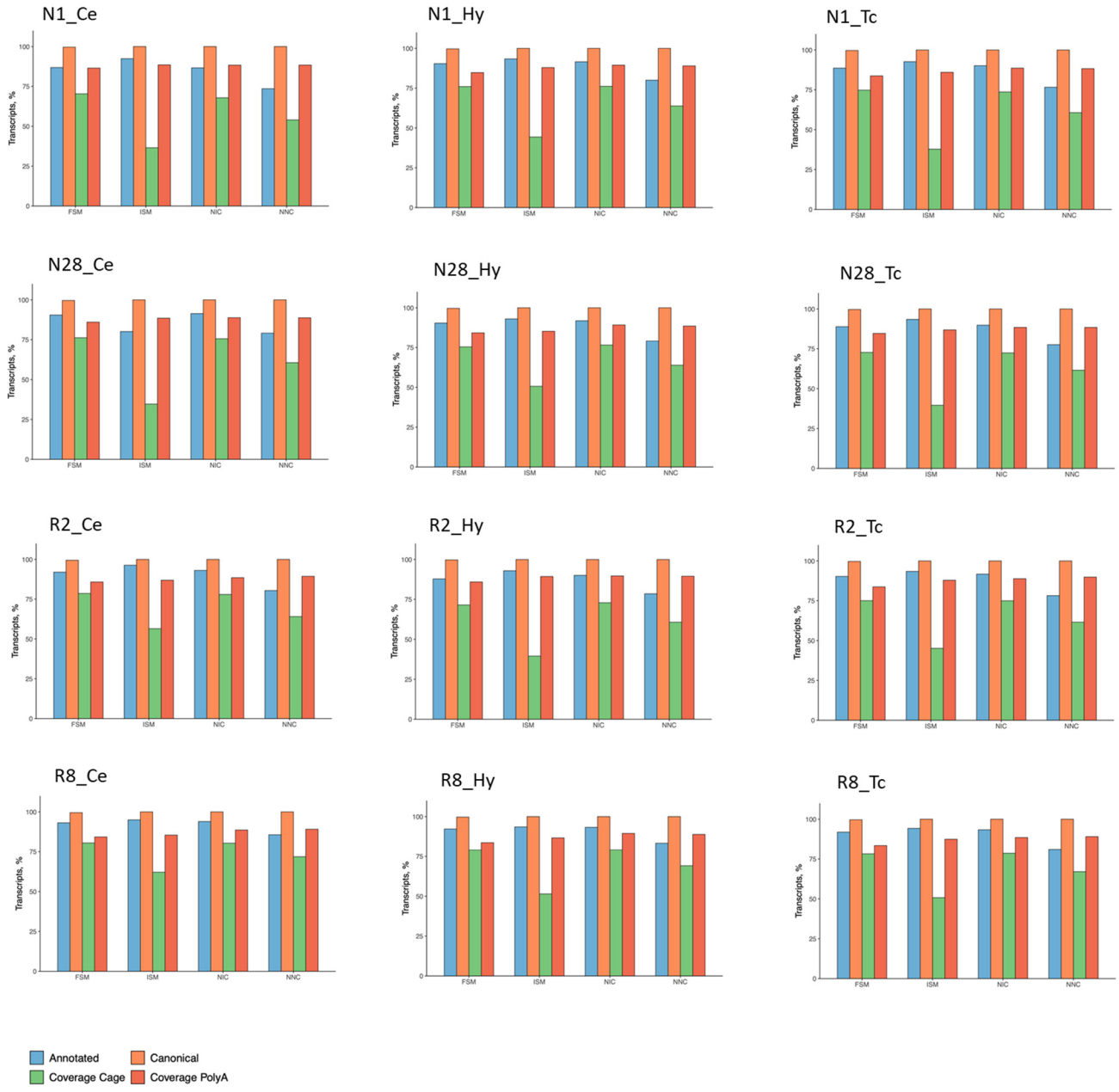


Fig. S3. Indicators of good quality in the Iso-Seq analysis. "Annotated" indicates that the transcription start site (TSS) of each isoform is less than 50 bp from the TSS of a known gene. "Canonical" indicates isoforms with canonical junctions using the following splice donor and acceptor sites: GT/AT, GT/AG, GC/AG, and AT/AC. "Coverage Cage" indicates whether there is a cage peak within 10 kb of the TSS. "Coverage PolyA" indicates whether there is a polyA tail. Although the "Coverage Cage" of ISM is low, the remainder of the indicators are generally high.

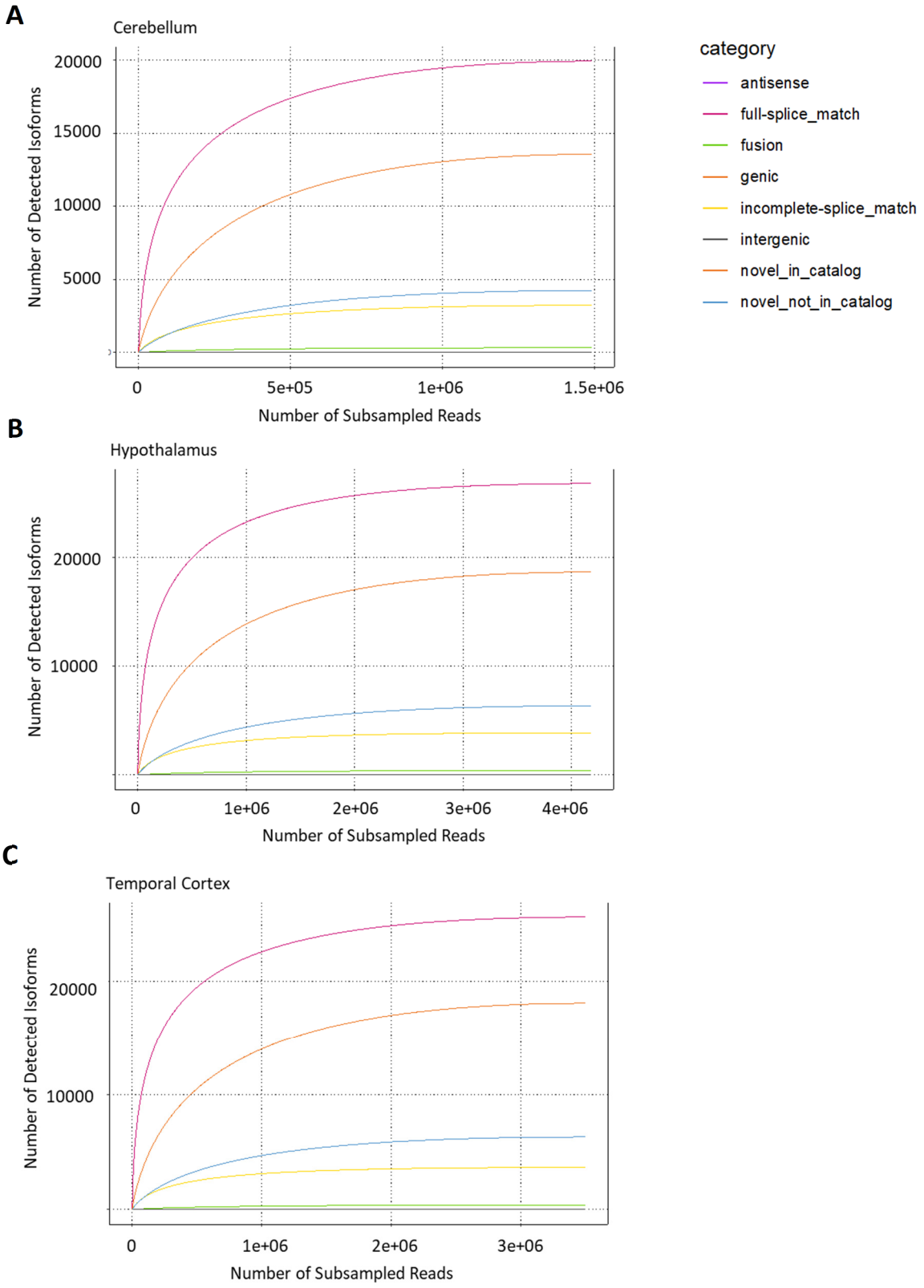


Fig. S4. Rarefaction curves for each isoform category by brain region. A Cerebellum. **B** Hypothalamus. **C** Temporal cortex.

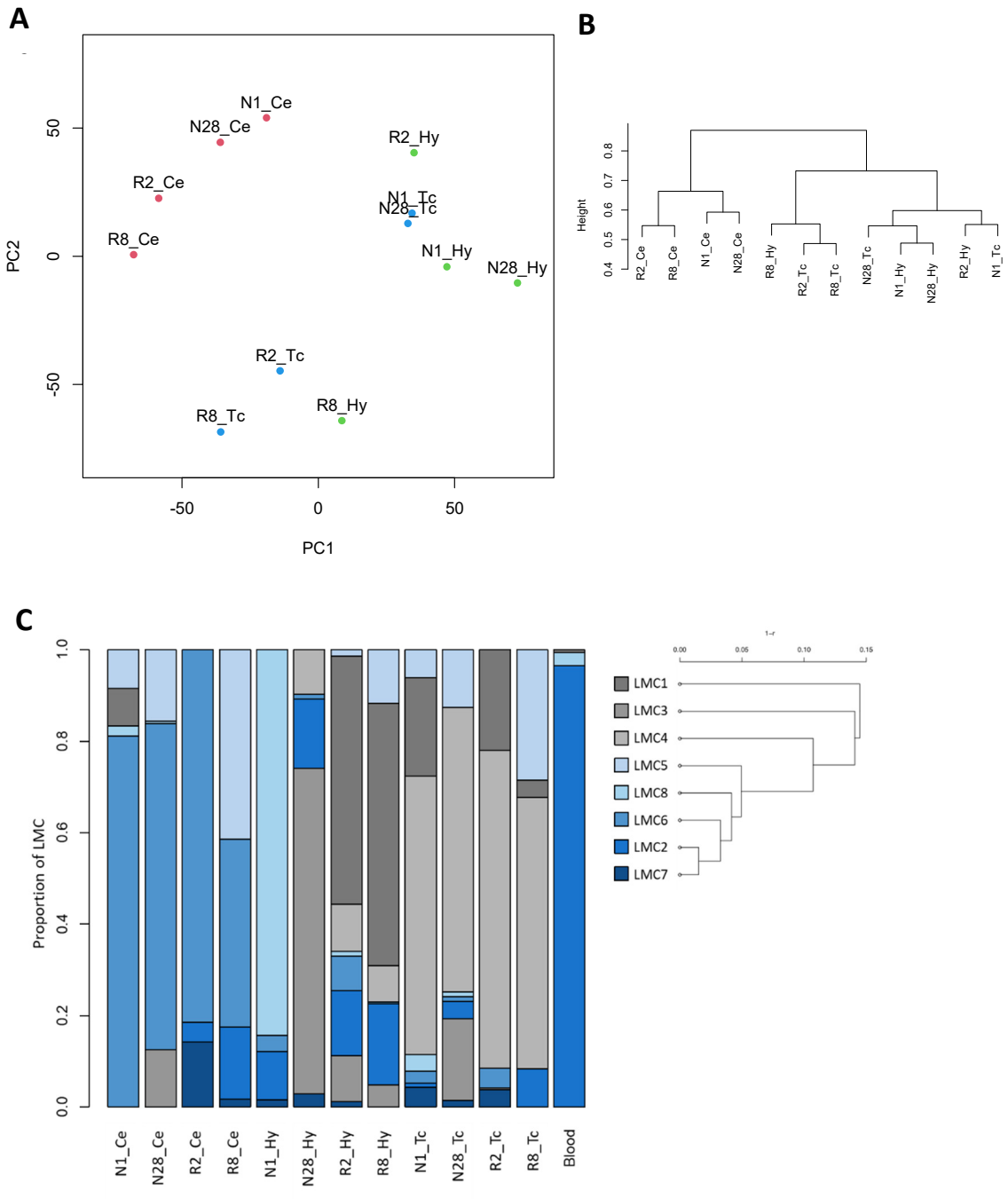


Fig. S5. Relationships between samples inferred from Iso-seq and genome-wide methylation data. **A** Principal component analysis based on isoforms identified by Iso-Seq. **B** Cluster analysis based on Iso-Seq results. **C** Reference-free estimates of cell composition ratios using genome-wide methylation data. Latent methylation component (LMC) corresponds to the cell type assumed to be included in the analysis. One example of peripheral blood-derived DNA methylation data is included as a control, and the ratio of LMC in each sample is shown (by color) in order of proximity of the branches according to the cluster analysis on each LMC.

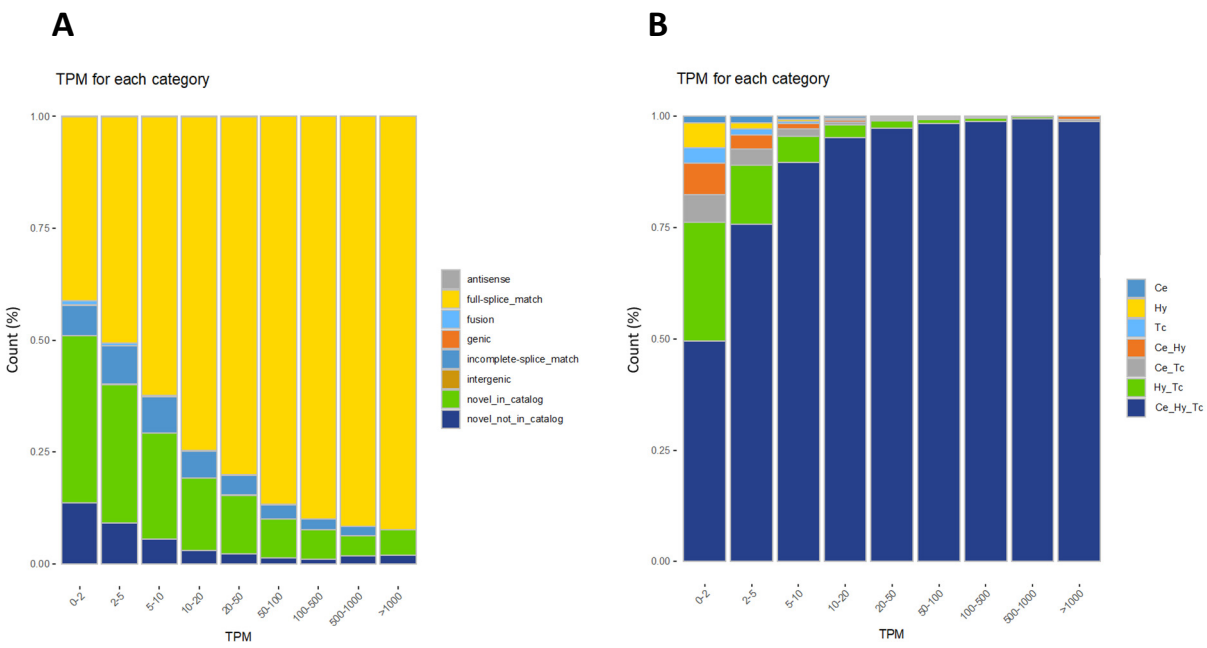


Fig. S6. Examination of expression (TPM) features. A Percentage of isoform categories in each TPM category. While the proportion of FSM increases as TPM increases, the proportion of unregistered isoforms decreases. **B** Proportion of isoforms expressed in different brain regions in each TPM category. While the proportion of isoforms expressed in all the three regions increases as TPM increases, the proportion of isoforms expressed in up to two regions decreases.

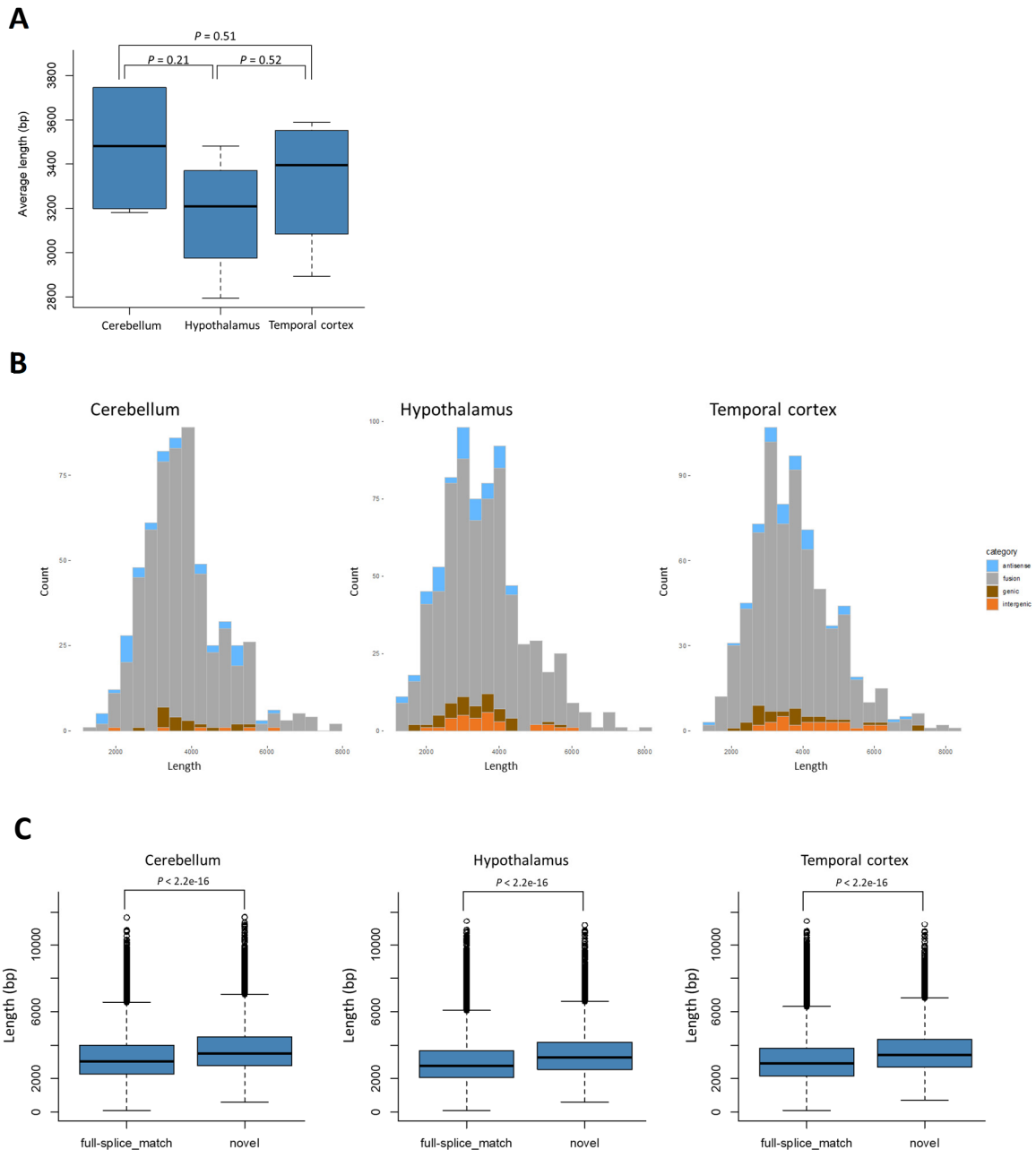


Fig. S7. Results for isoform length. A Comparison of lengths per sample between the brain regions. The p-value was calculated using a two-tailed t-test. **B** Length distribution of isoform categories with few detections. **C** Comparison of FSM and unregistered isoform lengths by region. The p-value was calculated using a two-tailed t-test.

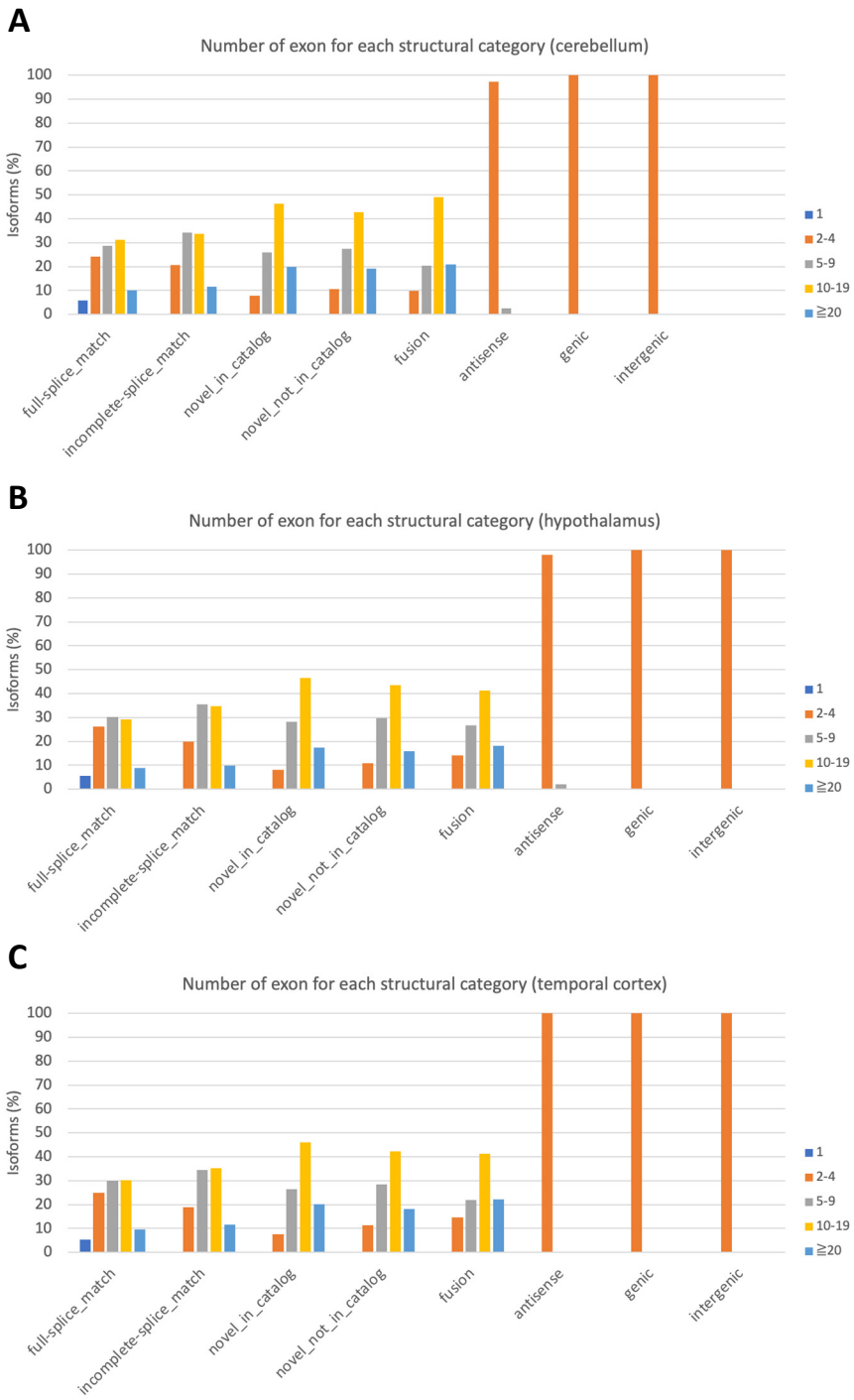
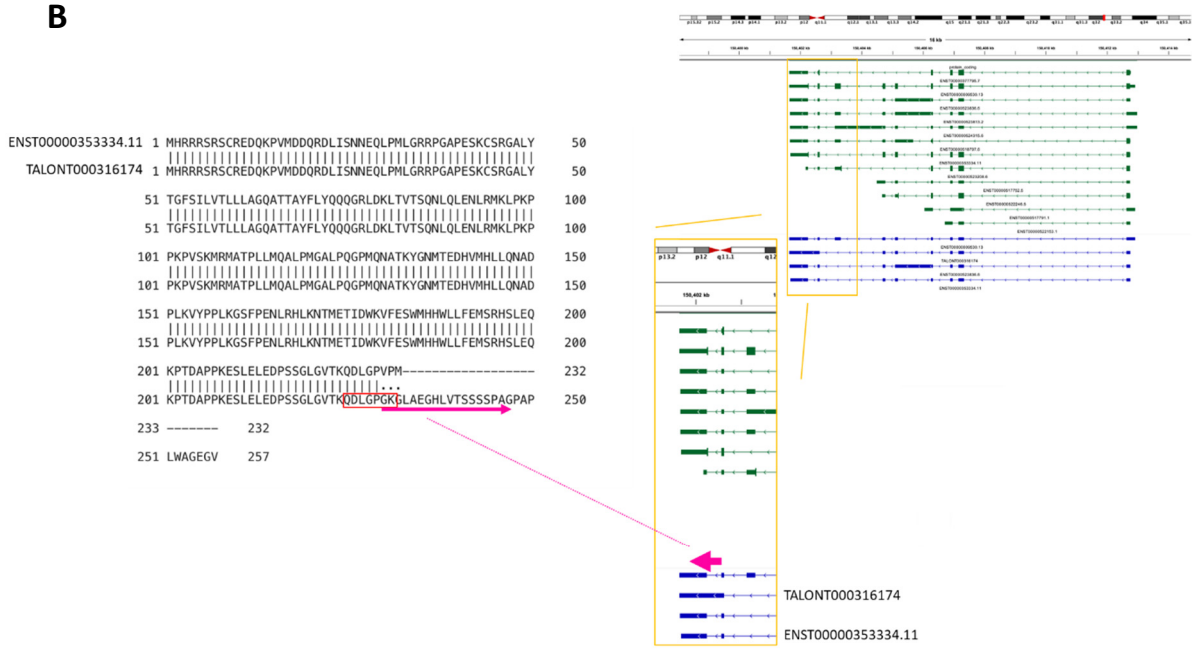


Fig. S8. Frequency distribution of exons in each isoform category. A Cerebellum, **B** Hypothalamus, and **C** Temporal cortex. In all regions, NIC and NNC had more exons than FSM and ISM, with more isoforms in the 10–19 category; antisense, genic, and intergenic isoforms had fewer exons and were mostly classified in the 2–4 category.

A


ENST00000221485.8 1 MEFRQEEFRKLAGRALGKLHRLLEKRQEGAETLELSADGRPVTTQTRDPP 50
TALONT000430941 1 ---------- 0
51 VVDCTCFGLPRRYIIAIMSGLGFCISFGIRCNLGVAIVSMVNNSTTHRGG 100
1 ---------- 0
101 HVVVQKAQFSWDPETVGLIHGSFFWGYIVTQIPGGFICQKFAANRVFGFA 150
1 ---------- 0
151 IVATSTLNMLIPSAARVHYGCVIFVRILQGLVEGVTYPACHGIWSKWAPP 200
1 ---------- 0
201 LERSRLATTAFCGSYAGAVVAMPLAGVLVQYSGWSSVFYVYGSFGIFWYL 250
1 ---------- 0
251 FWLLVSYESPALHPSISEEERKYIEDAIGESAKLMNPLTKFSTPWRRFFT 300
1 ---------- 0
301 SMPVYAIIVANFCRSWTFYLLLISQPAYFEEVFGFEISKVGLVSALPHLV 350
1 ---------- 0
351 MTIIVPIGGQIADFLRSRRIMSTTNVRKLMNCGGFGMEATLLLVVGYSHS 400
1 MTIIVPIGGQIADFLRSRRIMSTTNVRKLMNCG---------- 33
401 KGVAISFLVLAVGFSGFAISGFNVNHLDIAPRYASILMGISNGVGTLSGM 450
34 ----------GFNVNHLDIAPRYASILMGISNGVGTLSGM 63
451 VCPIIVGAMTKHKTREEWQYVFLIASLVHYGGVIFYGVFASGEKQPWAEP 500
64 VCPIIVGAMTKHKTREEWQYVFLIASLVHYGGVIFYGVFASGEKQPWAEP 113
501 EEMSEEKCGFVGHDQLAGSDDSEMEDEAEPPGAPPAPPPSYGATHSTFQP 550
114 EEMSEEKCGFVGHDQLAGSDDSEMEDEAEPPGAPPAPPPSYGATHSTFQP 163
551 PRPPPPVRDY 560
164 PRPPPPVRDY 173
p13.2
p13.13
p13.11
49,430 kb
49,432 kb
ENST00000221485.8
TALONT000430941

B


ENST00000353334.11 1 MHRRRSRSCREDQKPVMDDQRDLISNNEQLPMLGRRPGAPESKCSRGALY 50
TALONT000316174 1 MHRRRSRSCREDQKPVMDDQRDLISNNEQLPMLGRRPGAPESKCSRGALY 50
51 TGFSILVTLLLAGQATTAYFLYQQQGRLDKLTVTSQNLQLENLRMKLPKP 100
51 TGFSILVTLLLAGQATTAYFLYQQQGRLDKLTVTSQNLQLENLRMKLPKP 100
101 PKPVSKMRMATPLLMQALPMGALPQGPMQNATKYGNMTEDHVMHLLQNAD 150
101 PKPVSKMRMATPLLMQALPMGALPQGPMQNATKYGNMTEDHVMHLLQNAD 150
151 PLKVYPPLKGSFPENLRHLKNTMETIDWKVFESWMHHWLLFEMSRHSLEQ 200
151 PLKVYPPLKGSFPENLRHLKNTMETIDWKVFESWMHHWLLFEMSRHSLEQ 200
201 KPTDAPPKESLELEDPSSGLGVTKQDLGPVPM---------- 232
201 KPTDAPPKESLELEDPSSGLGVTKQDLGPGKGLAEGHLVTSSSSPAGPAP 250
233 -------- 232
251 LWAGEGV 257
TALONT000316174
ENST00000353334.11

C

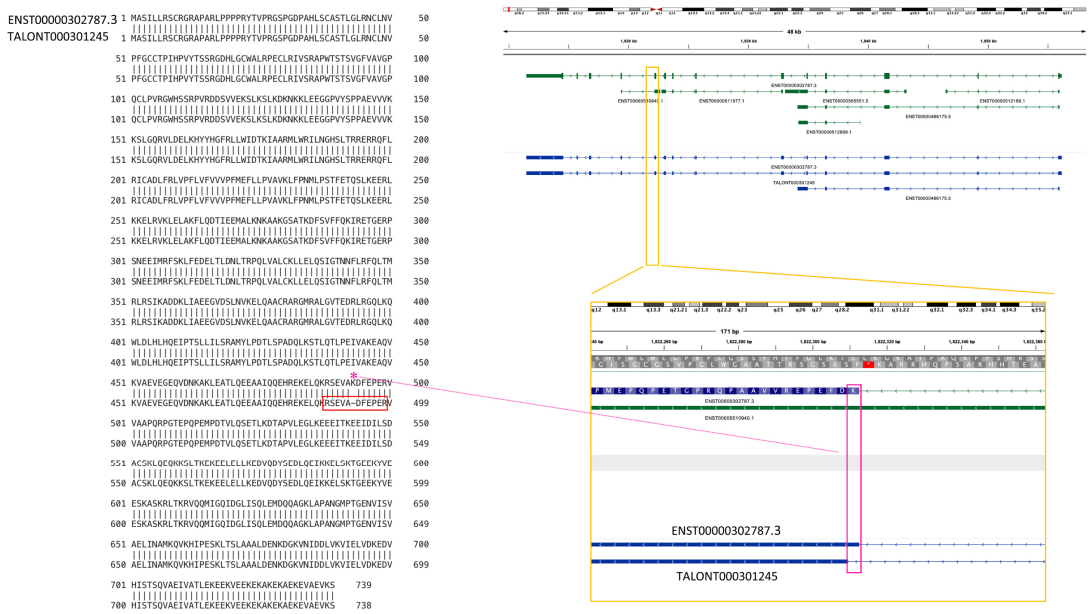
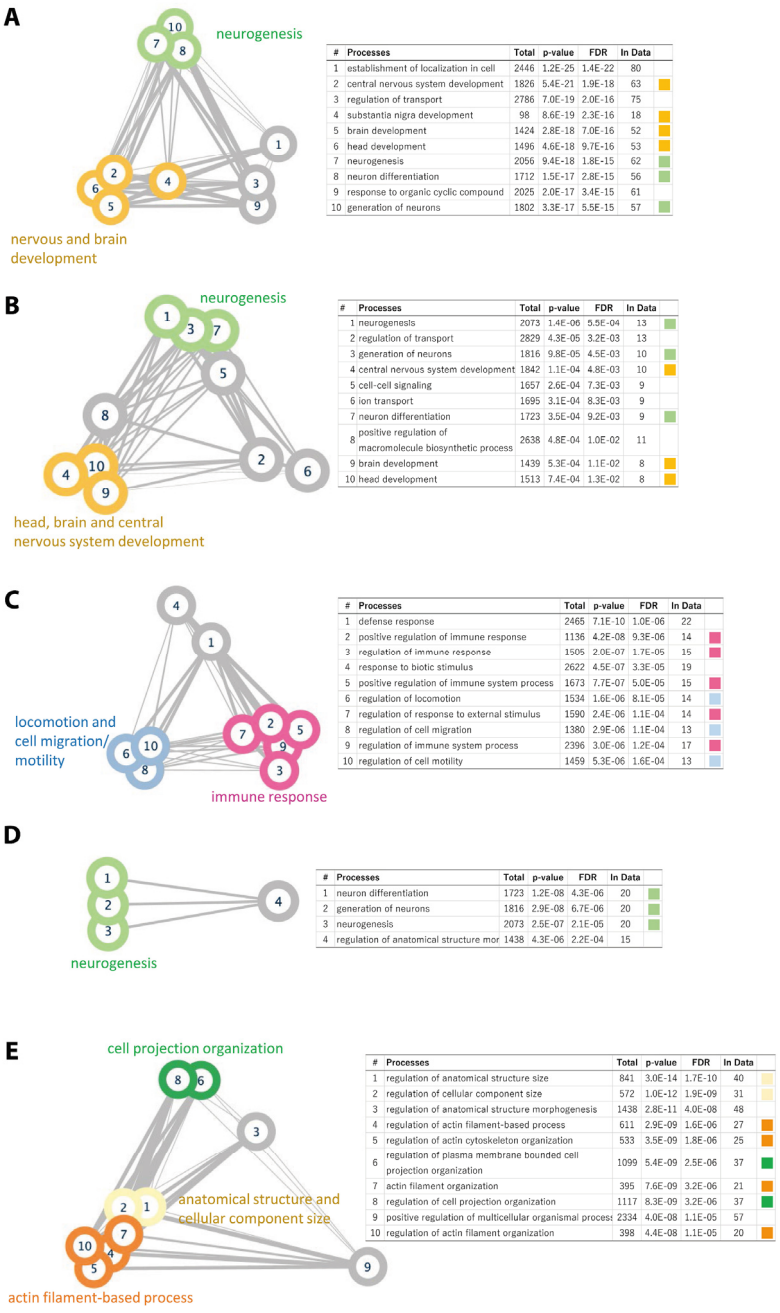


Fig. S9. The examples of the unregistered isoforms validated by LC-MS/MS. Each blue isoform represents the isoform detected in this Iso-Seq analysis, while the green isoforms represent all isoforms registered in GENCODE v.41. The amino acid sequence alignment was performed using EMBOSS Needle (https://www.ebi.ac.uk/Tools/psa/emboss_needle/). **A** An example of exon skipping. In the unregistered isoform, TALONT000430941, of the *SLC17A7* (Solute Carrier Family 17 Member 7) region, the third exon from the last exon was skipped, and the amino acid sequence resulting from the joining of the exons before and after that exon was detected. **B** An example of intron retention. In the unregistered isoform, TALONT000316174, of the *CD74* (CD74 Molecule, Major Histocompatibility Complex, Class II Invariant Chain) region, the intron between the last exon and the exon before it was retained, and the translated amino acid sequence resulting from the retained intronic region was detected. **C** An example of unregistered splice junction of NNC. In the unregistered amino acid sequence (TALONT000301245) of the *LETM1* (Leucine Zipper And EF-Hand Containing Transmembrane Protein 1) region, a one-amino-acid shift was found in the 5' splice site of the 10th exon. The detection of the amino acid sequence containing this amino acid indicated the presence of an unregistered splice junction.



#	Processes	Total	p-value	FDR	In Data
1	establishment of localization in cell	2446	1.2E-25	1.4E-22	80
2	central nervous system development	1826	5.4E-21	1.9E-18	63
3	regulation of transport	2786	7.0E-19	2.0E-16	75
4	substantia nigra development	98	8.6E-19	2.3E-16	18
5	brain development	1424	2.8E-18	7.0E-16	52
6	head development	1496	4.6E-18	9.7E-16	53
7	neurogenesis	2056	9.4E-18	1.8E-15	62
8	neuron differentiation	1712	1.5E-17	2.8E-15	56
9	response to organic cyclic compound	2025	2.0E-17	3.4E-15	61
10	generation of neurons	1802	3.3E-17	5.5E-15	57

#	Processes	Total	p-value	FDR	In Data
1	neurogenesis	2073	1.4E-06	5.5E-04	13
2	regulation of transport	2829	4.3E-05	3.2E-03	13
3	generation of neurons	1816	9.8E-05	4.5E-03	10
4	central nervous system development	1842	1.1E-04	4.8E-03	10
5	cell-cell signaling	1657	2.6E-04	7.3E-03	9
6	ion transport	1695	3.1E-04	8.3E-03	9
7	neuron differentiation	1723	3.5E-04	9.2E-03	9
8	positive regulation of macromolecule biosynthetic process	2638	4.8E-04	1.0E-02	11
9	brain development	1439	5.3E-04	1.1E-02	8
10	head development	1513	7.4E-04	1.3E-02	8

#	Processes	Total	p-value	FDR	In Data
1	defense response	2465	7.1E-10	1.0E-06	22
2	positive regulation of immune response	1136	4.2E-08	9.3E-06	14
3	regulation of immune response	1505	2.0E-07	1.7E-05	15
4	response to biotic stimulus	2622	4.5E-07	3.3E-05	19
5	positive regulation of immune system process	1673	7.7E-07	5.0E-05	15
6	regulation of locomotion	1534	1.6E-06	8.1E-05	14
7	regulation of response to external stimulus	1590	2.4E-06	1.1E-04	14
8	regulation of cell migration	1380	2.9E-06	1.1E-04	13
9	regulation of immune system process	2396	3.0E-06	1.2E-04	17
10	regulation of cell motility	1459	5.3E-06	1.6E-04	13

#	Processes	Total	p-value	FDR	In Data
1	neuron differentiation	1723	1.2E-08	4.3E-06	20
2	generation of neurons	1816	2.9E-08	6.7E-06	20
3	neurogenesis	2073	2.5E-07	2.1E-05	20
4	regulation of anatomical structure mor	1438	4.3E-06	2.2E-04	15

#	Processes	Total	p-value	FDR	In Data
1	regulation of anatomical structure size	841	3.0E-14	1.7E-10	40
2	regulation of cellular component size	572	1.0E-12	1.9E-09	31
3	regulation of anatomical structure morphogenesis	1438	2.8E-11	4.0E-08	48
4	regulation of actin filament-based process	611	2.9E-09	1.6E-06	27
5	regulation of actin cytoskeleton organization	533	3.5E-09	1.8E-06	25
6	regulation of plasma membrane bounded cell projection organization	1099	5.4E-09	2.5E-06	37
7	actin filament organization	395	7.6E-09	3.2E-06	21
8	regulation of cell projection organization	1117	8.3E-09	3.2E-06	37
9	positive regulation of multicellular organismal process	2334	4.0E-08	1.1E-05	57
10	regulation of actin filament organization	398	4.4E-08	1.1E-05	20

Fig. S10. Pathway analysis of genes associated with gene expression levels. When more than 10 pathways are detected, the top 10 pathways with the smallest FDR are shown. Edge thickness is proportional to the number of related genes shared among pathways. **A** Pathway analysis results for genes with high expression (transcripts per million (TPM) >200) in all brain regions. **B** Pathway analysis results for genes that are upregulated in the cerebellum. **C** Pathway analysis results for genes that are upregulated in the hypothalamus. **D** Pathway analysis results for genes that are upregulated in the temporal cortex. **E** Pathway analysis results for genes that are downregulated in the cerebellum.

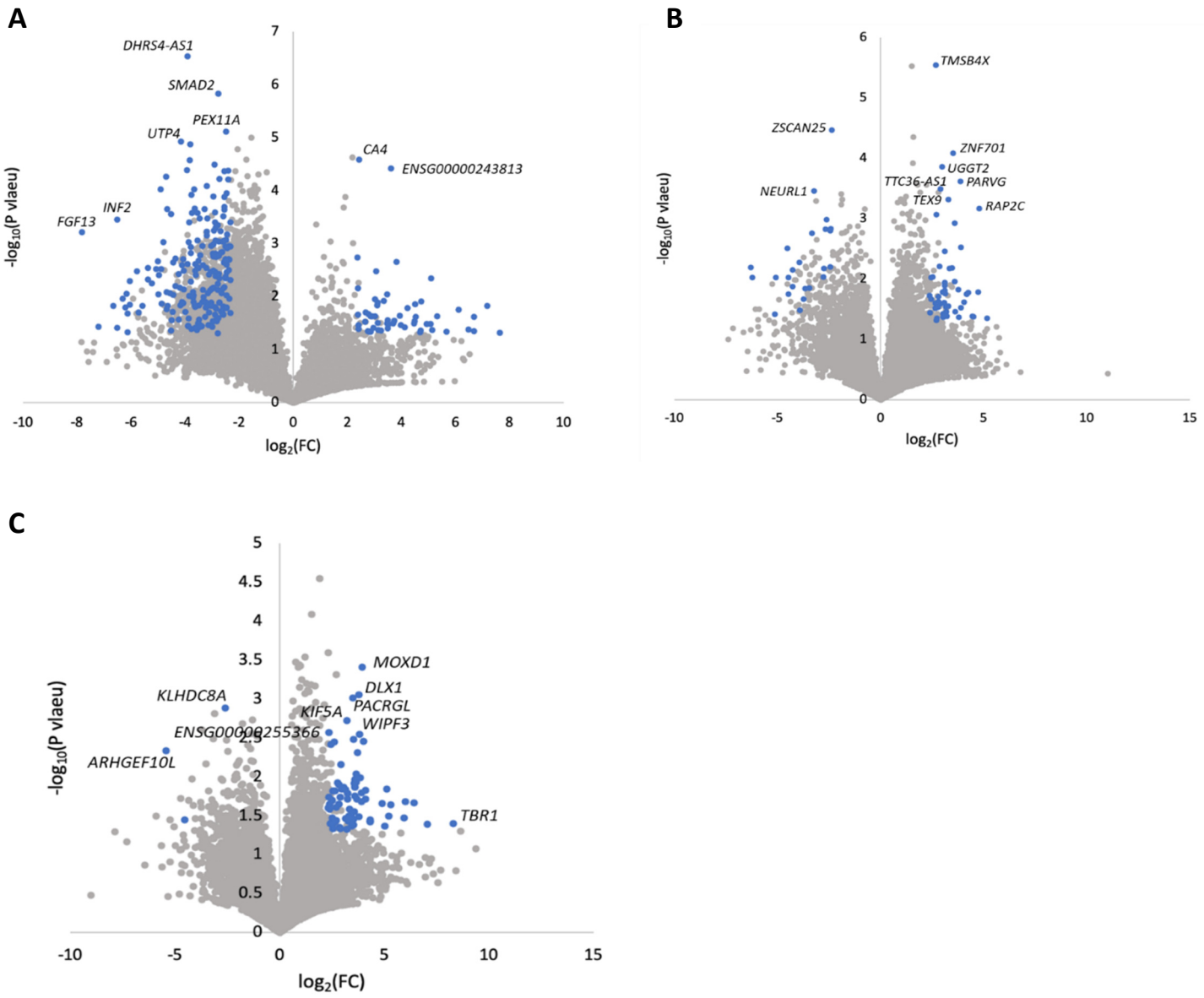
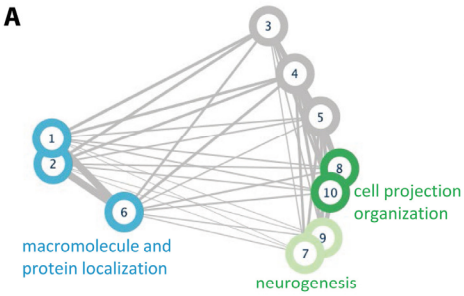
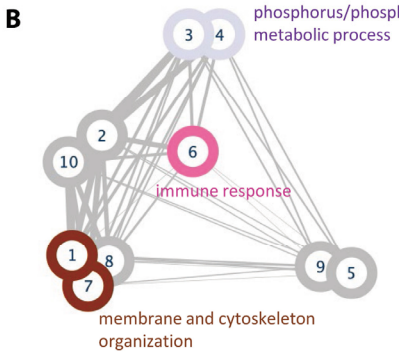


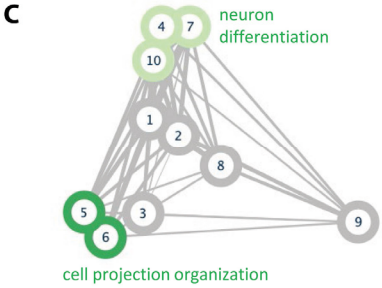
Fig. S11. Volcano plot of the differentially expressed genes between regions. Genes with an expression level of 0 in either of the regions being compared have been excluded from the plot, as most of them have been identified in the analysis of genes expressed only in specific regions (Table S10-12). **A** Genes that are differentially expressed between the cerebellum and other regions (hypothalamus and temporal cortex). **B** Genes that are differentially expressed between the hypothalamus and other regions (cerebellum and temporal cortex). **C** Genes that are differentially expressed between the temporal cortex and other regions (cerebellum and hypothalamus).



#	Processes	Total	p-value	FDR	In Data	Detected in analysis of tissues or cells other than the brain	
1	cellular macromolecule localization	2665	2.4E-24	3.0E-21	126	○	
2	protein localization	2657	6.1E-24	6.2E-21	125	○	
3	regulation of cellular component biogenesis	1374	7.6E-23	5.9E-20	84	○	
4	positive regulation of cellular component organizati	1647	2.6E-22	1.8E-19	92	○	
5	regulation of anatomical structure morphogenesis	1433	2.9E-22	1.8E-19	85	○	
6	establishment of localization in cell	2446	7.4E-21	3.5E-18	113		
7	regulation of neurogenesis	648	5.7E-20	2.5E-17	53		
8	regulation of cell projection organization	1113	1.0E-19	4.2E-17	70		
9	regulation of nervous system development	794	2.0E-19	7.7E-17	58		
10	regulation of plasma membrane bounded cell projection organization	1093	6.5E-19	2.4E-16	68		



#	Processes	Total	p-value	FDR	In Data	
1	membrane organization	1226	2.0E-05	5.8E-03	19	
2	regulation of hydrolase activity	1623	9.8E-05	1.2E-02	21	
3	positive regulation of phosphorus metabolic process	1388	1.1E-04	1.2E-02	19	
4	positive regulation of phosphate metabolic process	1388	1.1E-04	1.2E-02	19	
5	organonitrogen compound biosynthetic process	1748	2.7E-04	2.4E-02	21	
6	activation of immune response	711	3.5E-04	2.7E-02	12	
7	cytoskeleton organization	1696	4.9E-04	2.8E-02	20	
8	organelle localization	739	4.9E-04	2.8E-02	12	
9	carbohydrate derivative metabolic process	1322	5.1E-04	2.8E-02	17	
10	negative regulation of response to stimulus	2384	5.3E-04	2.8E-02	25	



#	Processes	Total	p-value	FDR	In Data	
1	cell morphogenesis	1095	1.6E-12	4.4E-09	22	
2	cell morphogenesis involved in differentiation	869	1.2E-09	6.9E-07	17	
3	cell junction organization	832	3.8E-08	6.1E-06	15	
4	neuron differentiation	1712	2.1E-07	1.9E-05	20	
5	regulation of plasma membrane bounded cell projection organization	1093	2.2E-07	2.0E-05	16	
6	regulation of cell projection organization	1113	2.8E-07	2.5E-05	16	
7	generation of neurons	1802	4.7E-07	3.6E-05	20	
8	cell migration	1326	5.6E-07	4.2E-05	17	
9	regulation of transport	2786	6.7E-07	4.9E-05	25	
10	neuron development	1385	1.0E-06	6.8E-05	17	

Fig. S12. Pathway analysis of genes associated with genes with many isoforms. When more than 10 pathways are detected, the top 10 pathways with the smallest FDR are shown. Edge thickness is proportional to the number of related genes shared among pathways. No significant pathways were detected in the pathway analysis using isoform-rich genes in the cerebellum. **A** Results of isoform-rich genes in all of the brain regions. **B** Results of isoform-rich genes in the hypothalamus. **C** Results of isoform-rich genes in the temporal cortex.

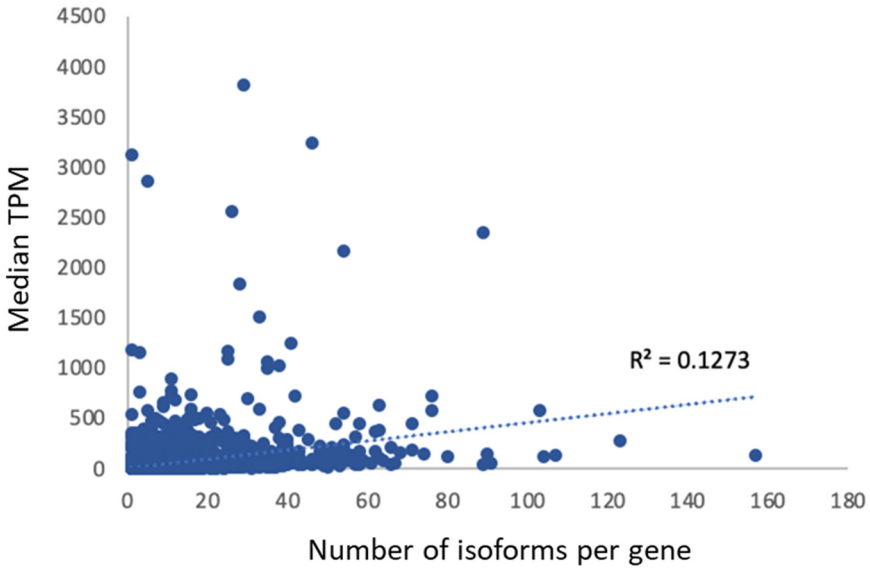


Fig. S13. Relationship between expression level and the number of isoforms per gene. The expression levels (TPM) and the number of isoforms per gene are weakly correlated with a correlation coefficient of 0.127.

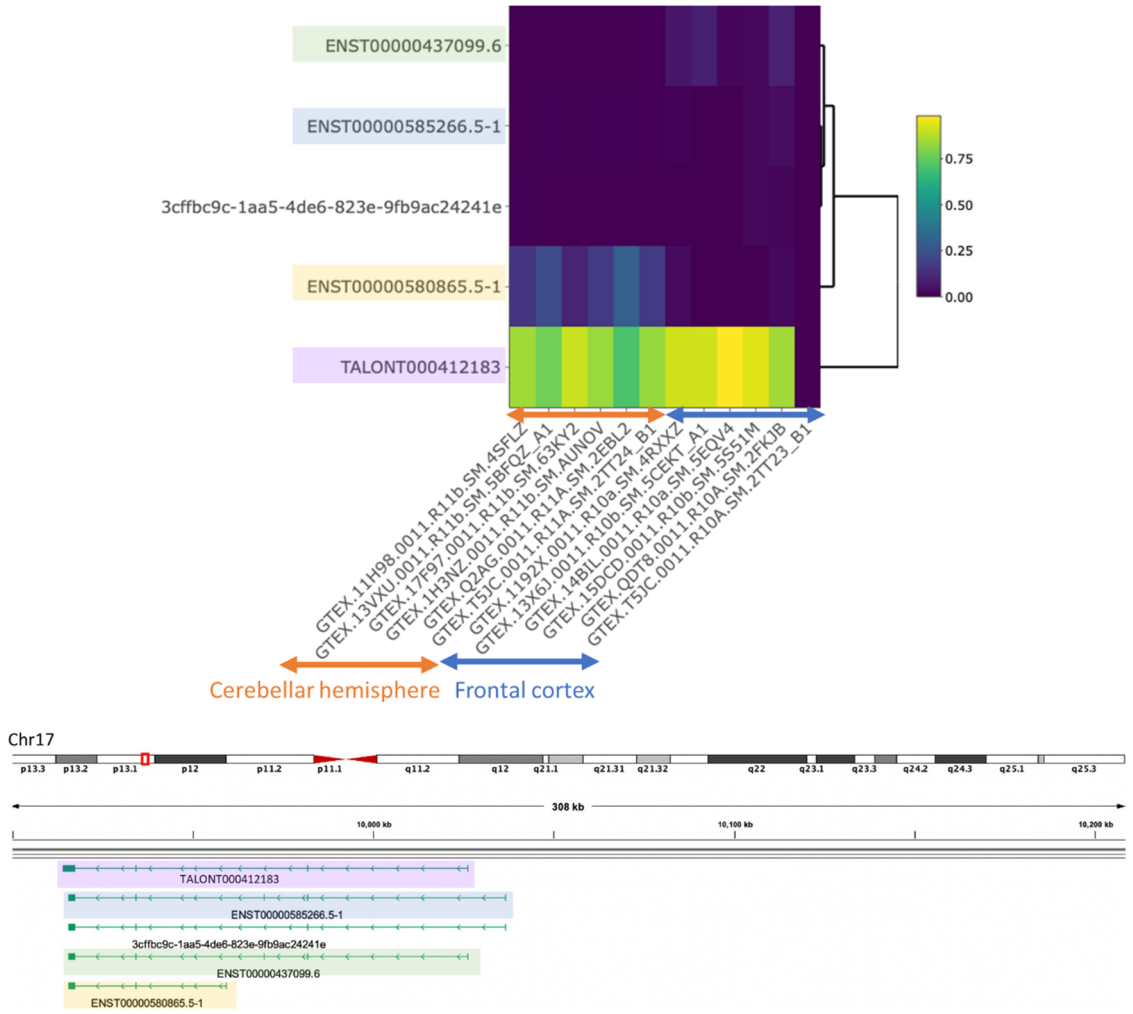


Fig. S14. *GAS7* isoforms in GTEx samples. For the isoforms of *GAS7*, we examined using long-read RNA-seq data of the human brain from GTEx different from this study. The colors of each isoform correspond to the *GAS7* isoforms in Fig3.

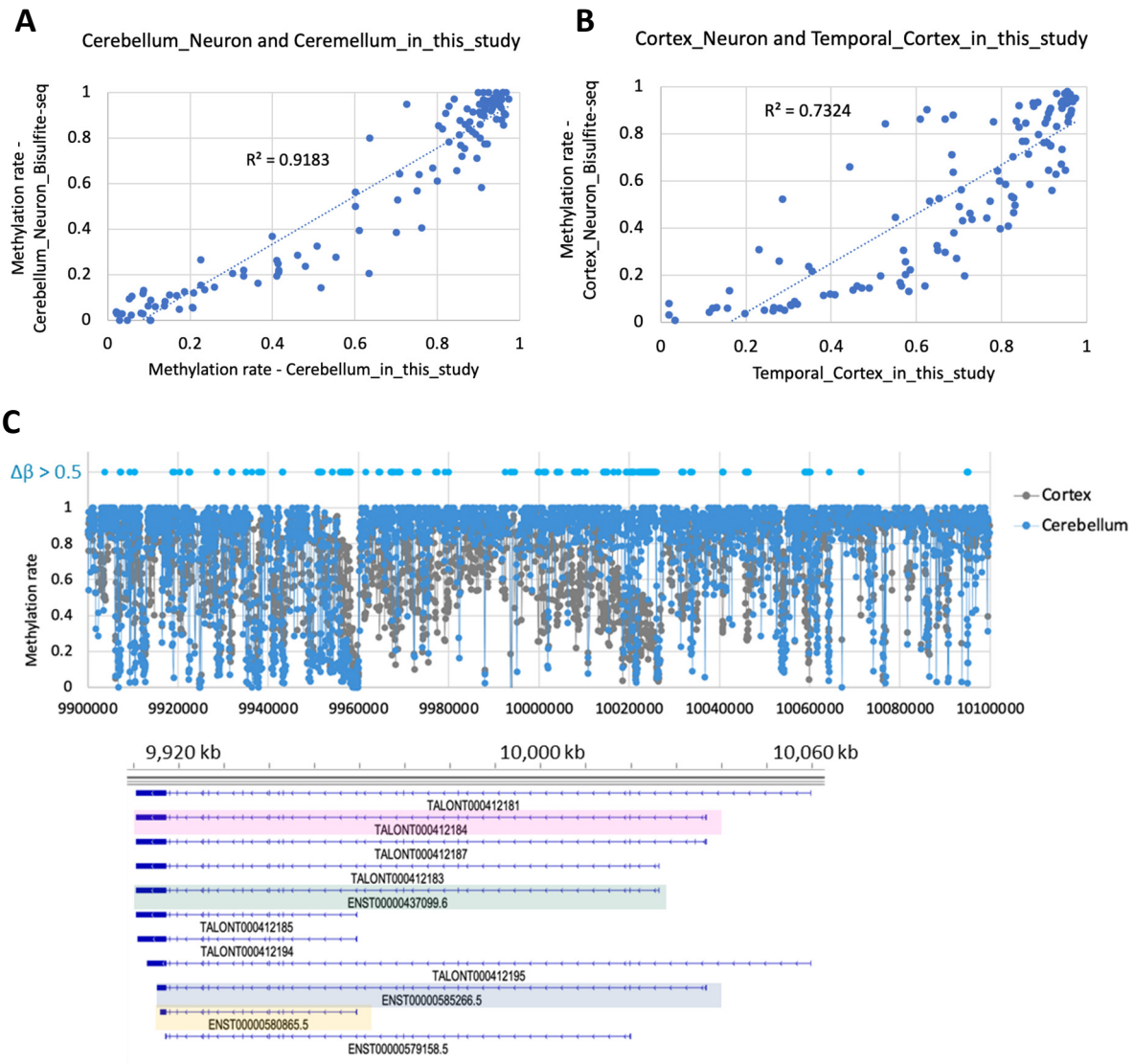


Fig. S15. Correlation with methylation rates determined by bisulfite sequencing, and the *GAS7* region in the other bisulfite-sequencing study. We compared the data from a previous study that examined DNA methylation in cerebellar and cortical-derived neurons using bisulfite sequencing with the methylation data from our current study. **A** Correlation between the methylation rate in cerebellar neurons and the methylation rate in cerebellum samples from this study. **B** Correlation between the methylation rate in cortex neurons and the methylation rate in temporal cortex samples from this study. **C** The relationship between the methylation rate determined by bisulfite sequencing and isoforms in the *GAS7* region.

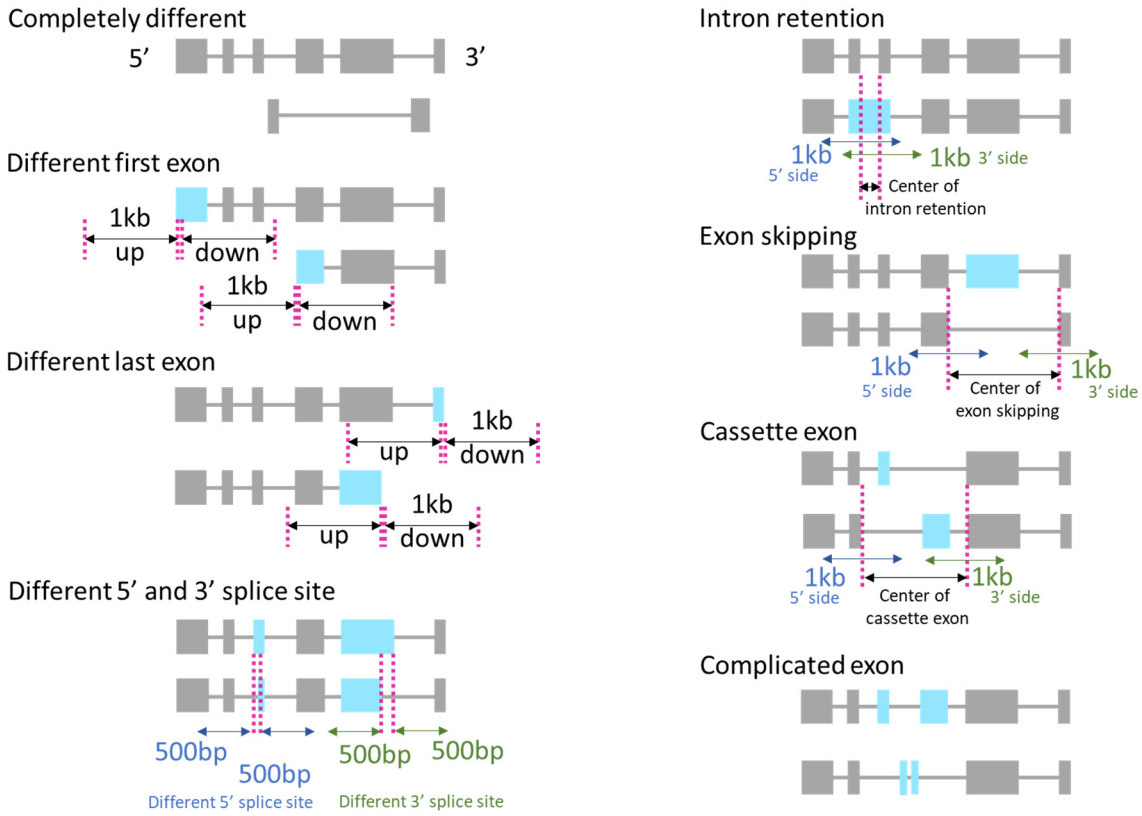


Fig. S16. Definitions of isoform differences and related areas. We examined whether or not there is a difference between each pair of the compared isoforms. “**Completely different**” is a case in which exons do not overlap at all. Here, if the shorter of the two exons to be compared overlapped by more than 50%, they were considered to be the same exon. “**Different first exon**” and “**Different last exon**” indicate that the first exon or the last exon are different, and the region 1 kb upstream and 1 kb downstream from either the TSSs or the transcription termination sites were defined as relevant regions, respectively. “**Different 5’ and 3’ splice site**” are cases where the 3' and/or 5' sides of overlapping exons do not match. The 5' end of the first exon and the 3' end of the last exon were excluded from the analysis. For “**Intron retention**”, the region of the expressed intron and 1-kb areas from both ends of the expressing intron were defined as the relevant region. For “**Exon skipping**”, the 3' to 5' ends of the overlapping exons at both ends of the skipping and the surrounding 1 kb region were defined as the relevant areas, respectively. “**Cassette exon**” were counted only if they used one different exon each; other complex structures were classified as “**Complicated exon**”. The 3' to 5' ends of the overlapping exons at both ends of the cassette exon and a 1 kb area from each end were defined as the relevant region. We have not defined the related area for complicated exon.

Command lines used in this study:

Iso-Seq Analysis Using pbBioConda

#---------- Primer Removal & Demultiplexing

lima --isoseq --dump-clips --peek-guess -j 32 ./$sample/$sample.hifi_reads.bam ./$sample/isoseq_primers.fasta ./$sample/$sample.hifi.demult.bam

#---------- Trimming PolyA Trails & Concatemer Removal

isoseq3 refine --require-polya ./$sample/$sample.hifi.demult.5p--3p.bam ./$sample/isoseq_primers.fasta ./$sample/$sample.hifi.flnc.bam

#---------- Cluster

isoseq3 cluster ./$sample/$sample.hifi.flnc.bam ./$sample/$sample.hifi.polished.bam --verbose --use-qvs

#---------- Mapping

cp /[directory_1]/hg38.fasta [directory_2]/$sample/

pbmm2 align ./$sample/hg38.fasta ./$sample/$sample.hifi.polished.hq.bam ./$sample/$sample.hifi.aligned.bam -j 32 --preset ISOSEQ --sort --log-level INFO

#---------- Collapse

isoseq3 collapse ./$sample/$sample.hifi.aligned.bam ./$sample/$sample.hifi.collapsed.gff

SQANTI3

conda activate SQANTI3.env

#---------- QC

export PYTHONPATH=$PYTHONPATH: [directory_for_cDNA_Cupcake]/

python sqanti3_qc.py $sample.hifi.collapsed.gff
gencode.v41.chr_patch_hapl_scaff.annotation.gtf /

hg38.fasta -d $sample.SQ3 -o $sample.SQ3 --CAGE_peak [directory_3]
/human.refTSS_v3.1.hg38.bed /

--polyA_motif_list [directory_3]/mouse_and_human.polyA_motif.txt --polyA_peak
[directory]/atlas.clusters.2.0.GRCh38.96.bed /

-n 16 --saturation --report both --isoAnnotLite --gff3
Homo_sapiens_GRCh38_RefSeq_78.gff3 -fl $sample.hifi.collapsed.abundance.txt

#---------- Filtering (Rule filter)

python sqanti3_filter.py rules $sample.SQ3_classification.txt -j [jsonfile: filtering.json] -
o $[output_name] -d $[output_directory]

#---------- Filtering (Machine learning filter)_ Used only for the purpose of filtering ISM in the subsequent analysis

python sqanti3_filter.py ML $sample.SQ3_classification.txt -d $[output_directory]

TALON

#---------- Database construction

talon_initialize_database --f [directory]/gencode.v41.annotation.gtf --g hg38 --a
gencode41 --o myDatabase

#---------- Preparation for each sample file for TALON analysis

minimap2 -ax splice -uf --secondary=no -C5 -ax splice:hq [directory]/hg38.fasta -t30
¥ $sample_final.fasta > $sample_final.hg38.sam

samtools calmd $sample_final.hg38.sam [directory]/hg38.fasta/

--output-fmt sam > $sample_final.hg38.MDtagged.sam

#---------- Running TALON

```
talon --f config.csv/
   --db myDatabase.db /
   --build hg38 /
   --cov 0.95 /
   --identity 0.95 /
   --o [output_file_name]
talon_create_GTF --db=myDatabase.db /
--annot=gencode41 /
-b hg38 /
--observed /
--o=[output_gtf_filename]
#-------------------- Output data creation
# R program:
x<-read.table('[output_file_name]_read_annot.tsv',sep='/t',header=T)
x.$sample <- subset(x, dataset=="$sample")
write.table(x.$sample '$[output_filename_2].tsv', quote=F, row.names=F, sep='/t')
#-------------------- Merge with the SQANTI3 output file, the classification file
# R program:
x<-read.table('[output_filename_2].tsv',sep='/t',header=T)
y <- read.table('$sample_classification.txt',sep='/t',header=T)
y$read_name <- y$isoform
y <- y[, colnames(y) != "X"]
y <- y[, colnames(y) != "X.1"]
m <- merge(x, y, by='read_name')
```

```
m$superID <- paste(m$annot_gene_id, m$annot_transcript_id, sep = "_")

write.table(m, '$sample.talon.classification.tsv', quote=F, row.names=F, sep='/t')
```

R code:

```
#-------------------- Cluster analysis
x<-read.table('[data_file]', sep='/t', header=T, row.names=1)
x2 <- t(x)
d <- dist(x2)
ans <- hclust(d, method="ward.D2")
plot(ans)

#-------------------- PCA
prcomp3 <- function(df)
{
   df <- na.omit(df)
   ans <- prcomp(df)
   ans$loadings <- t(t(ans$rotation)*ans$sdev)
   ans$eigenvalues <- ans$sdev^2
   print.default(ans)
   invisible(ans)
}
x<-read.table('[data_file]', sep='/t', header=T, row.names=1)
x2 <- t(x)
ans <- prcomp3(x2)
pos <- c(1,1,1,1,2,2,2,2,3,3,3,3)
```

```
add_data_2 <- data.frame(pos)

names(add_data_2) <- c("pos")

df_out <- as.data.frame(ans$x)

data <- cbind(df_out, add_data_2)

plot(data$PC1,data$PC2,pch = 21, cex=1 ,bg = c(2, 3,
4)[unclass(data$pos)],lwd=0,ylim = c(-40,40),xlim=c(-40,40))

text(data$PC1,data$PC2,colnames(x),pos=3, cex=1)



#---------- t test

x<-read.table('[data_file]', sep='/t', header=T, row.names=1)

t.test(x$value~x$group)


#---------- Wilcoxon rank-sum test

x<-read.table('[data_file]', sep='/t', header=T, row.names=1)

wilcox.test(x$value~x$group)


#---------- Chi-square test, Fisher's exact test and calculation of ORs (Repeat
for multiple lines of data)

x <- read.table("[data_file]",header=TRUE)

nrow <- nrow(x)

sink("[output_filename]")

for (i in 1:nrow){

dat1_1 <- x[i,2:3]

dat1_2 <- x[i,4:5]
```

```
dat1 <- merge(dat1_1, dat1_2)

if ((dat1_1[1,]>0) || (dat1_2[1,]>0)) {

dat2 <- matrix(unlist(dat1), ncol=2, byrow=TRUE)

table <- as.table(dat2)

ans <- prop.test(table)

chi <- ans$statistic

pvalue <- ans$p.value

ans2 <- fisher.test(table)

pvalue_2 <- ans2$p.value

rate1 <- dat1_1[1]/(dat1_1[2]+dat1_1[1])

rate2 <- dat1_2[1]/(dat1_2[2]+dat1_2[1])

#OR

x <- table

odds <- (x[1,1]/x[1,2]) / (x[2,1]/x[2,2])

#OR (95%CI)

y1 <- log((x[1,1]/x[1,2]) / (x[2,1]/x[2,2]))

y2 <- 1/x[1,1] + 1/x[1,2] + 1/x[2,1] + 1/x[2,2]

interval <- exp(y1 + qnorm(c(0.025,0.975)) * sqrt(y2))

result <-
sprintf("%.3e,%.3e,%.3f,%.3e,%.3e,%.3f,%.3f,%.3f,%.1f,%.1f,%.1f,%.1f",rate1,rate2,chi,pval
ue,pvalue_2,odds,interval[1],interval[2],dat1_2[1],dat1_2[2],dat1_1[1],dat1_1[2])

print(result,quote=F,row.names=F)

}

}
```

```
sink()

#-------------------- MeDeCom
library(MeDeCom)
Data <- read.table("[methylation_datafile]")
Data2 <- as.matrix(Data)

medecom.result<-runMeDeCom(Data2, 2:10, c(0,10^(-5:-1)), NINIT=10, NFOLDS=10, ITERMAX=300, NCORES=5)

pdf("[directory]/plotParameters.pdf")
plotParameters(medecom.result)
dev.off()

pdf("[directory]/plotParameters2.pdf")
plotParameters(medecom.result, K=8, lambdaScale="log")
dev.off()

lmcs<-getLMCs(medecom.result, K=8, lambda=0.00001)
str(lmcs)
write.table(lmcs, "[directory]/lmcs.txt")

pdf("[directory]/dendrogram.pdf")
plotLMCs(medecom.result, K=8, lambda=0.00001, type="dendrogram")
```

```
dev.off()


prop<-getProportions(medecom.result, K=8, lambda=0.00001)
str(prop)
write.table(prop, "[directory]/prop.txt")


pdf("[directory]/Proportions.pdf")
plotProportions(medecom.result, K=8, lambda=0.00001, type="barplot")
dev.off()


pdf("[directory]/heatmap.pdf")
plotProportions(medecom.result, K=8, lambda=0.00001, type="heatmap")
dev.off()
```

Supplementary Table Information

Table S27. Isoforms that are differentially expressed as a percentage of overall expression per gene in cerebellum and hypothalamus

Table S28. Isoforms that are differentially expressed as a percentage of overall expression per gene in cerebellum and temporal cortex

Table S29. Isoforms that are differentially expressed as a percentage of overall expression per gene in hypothalamus and temporal cortex

Table S30. Pathway analysis with genes which isoforms were differentially expressed in cerebellum and hypothalamus

Table S31. Pathway analysis with genes which isoforms were differentially expressed in cerebellum and temporal cortex

Table S32. Pathway analysis with genes which isoforms were differentially expressed in hypothalamus and temporal cortex

Table S33. DMPs in relevant areas for each isoform structure difference

Table S34. Results of the statistical analysis of DMPs in relevant areas for each isoform structure difference

Table S35. Pair of isoforms that has a DMP within 1 kb downstream from the TSS and whose first exons are different between the isoforms

Table S36. DMRs in relevant areas for each isoform structure difference

Table S37. Results of the statistical analysis of DMRs in relevant areas for each isoform structure difference

Table S38. Demographic characteristics of the samples and RNA quality

The List of Links to the GitHub Repositories

IsoSeq v3 (https://github.com/PacificBiosciences/IsoSeq)	v3.4.0
pbmm2 (https://github.com/PacificBiosciences/pbmm2)	v1.7.0
SQANTI3 (https://github.com/ConesaLab/SQANTI3)	v5.0
TALON (https://github.com/mortazavilab/TALON)	v5.0
cDNA_Cupcake (https://github.com/Magdoll/cDNA_Cupcake)	v29.0.0
Brain_Iso-Seq (https://github.com/mihshimada/Brain_Iso-Seq)	v1.1