**Supplemental information**

De-erosion of X chromosome dosage compensation

by the editing of *XIST* regulatory regions

restores the differentiation potential in hPSCs
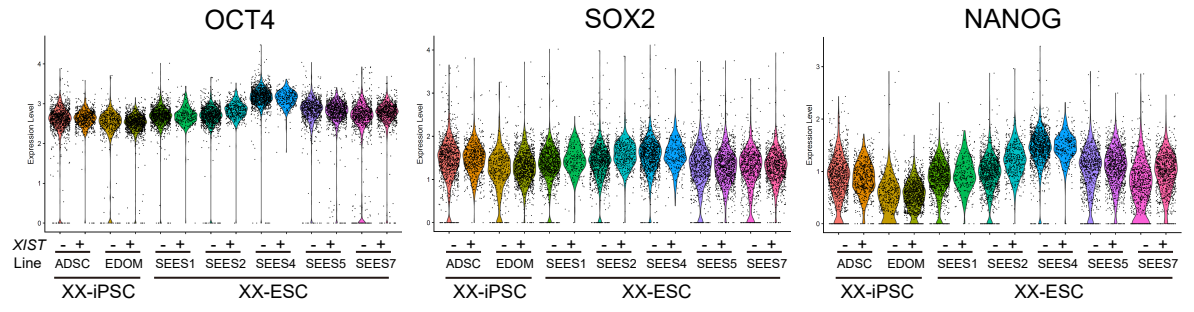
Nami Motosugi, Akiko Sugiyama, Chisa Okada, Asako Otomo, Akihiro Umezawa, Hidenori Akutsu, Shinji Hadano, and Atsushi Fukuda
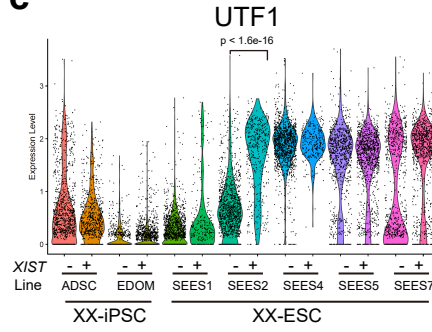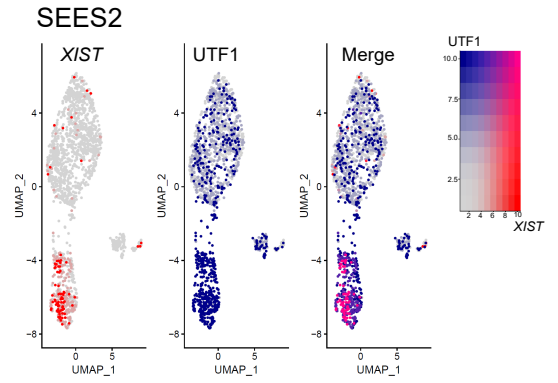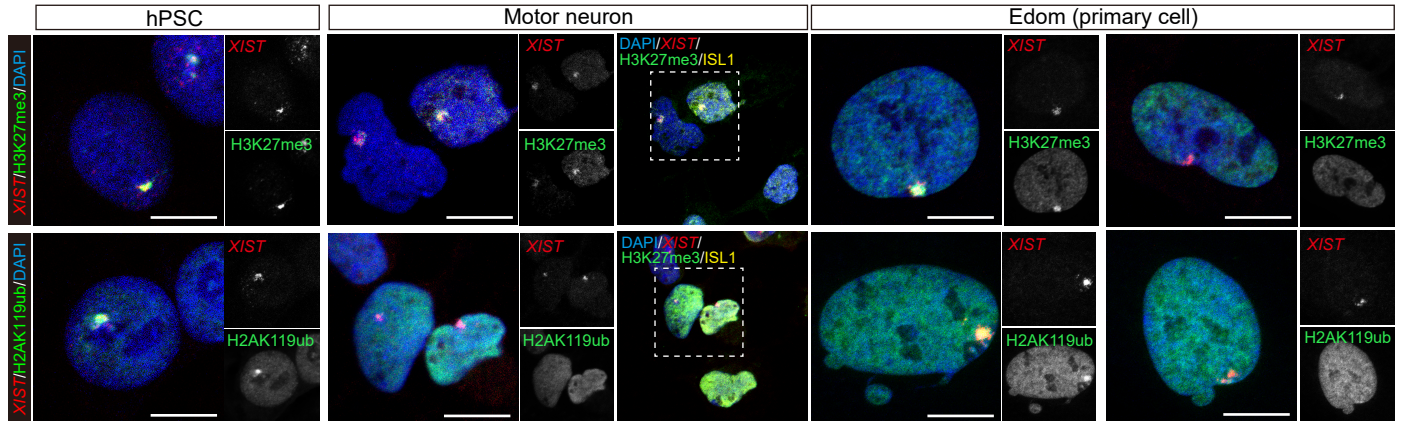
Figure S1

a

b OCT4   SOX2   NANOG

c UTF1

d SEES2

e hPSC   Motor neuron   Edom (primary cell)
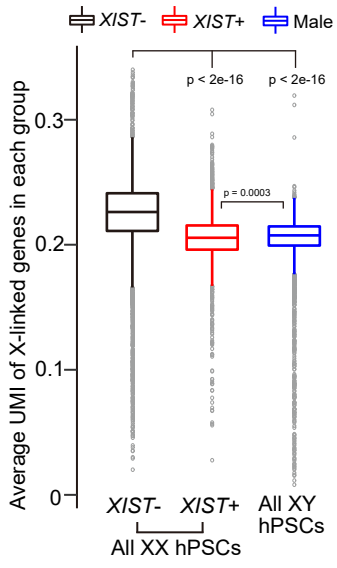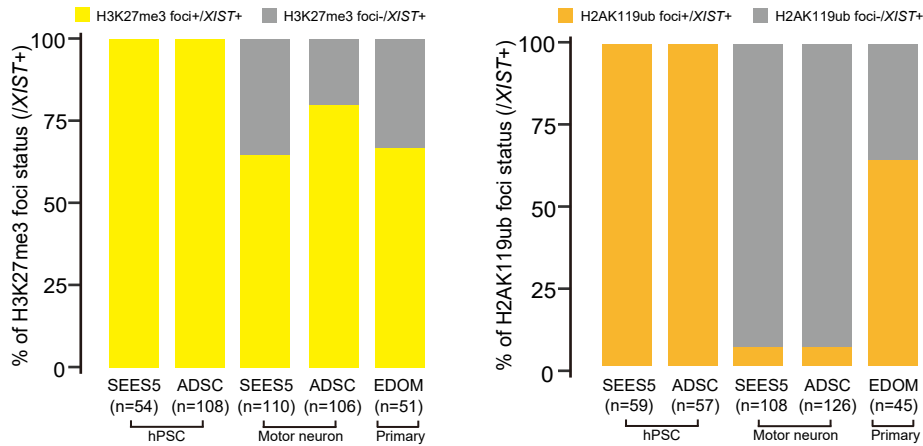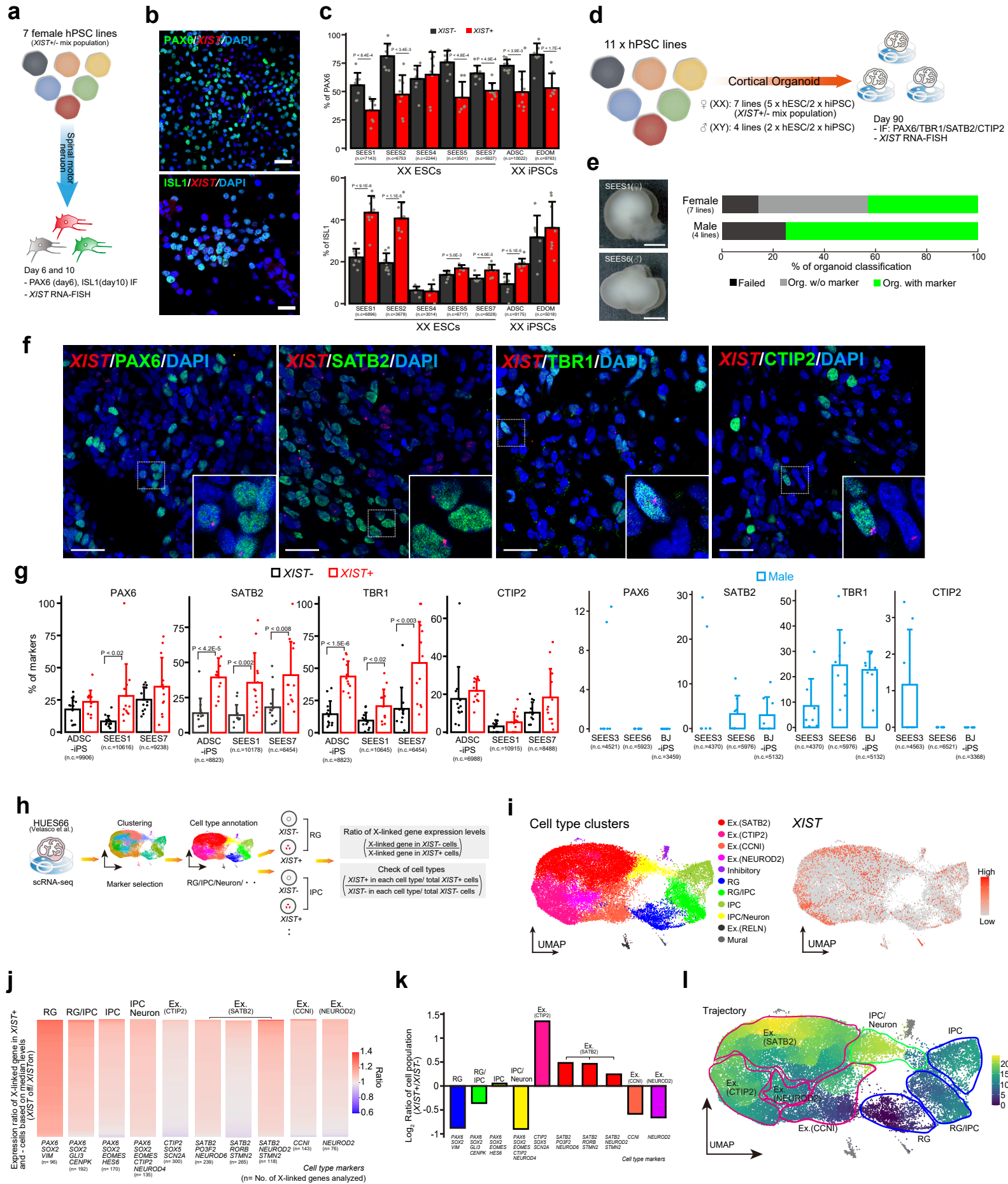
f

Figure S1 related to Figure 1. Expression status of core pluripotency factors in *XIST*+ and – female hPSCs and changes in epigenetic modifications on Xi in each cell type. (**a**) Average X-linked gene expression status in male and female hPSCs. In females, the individual cells were separated into two groups based on the *XIST* expression status (*XIST*+ or –). *p*-values were calculated using one way ANOVA with post-hoc test. (**b**) Expression status of core pluripotency factors. The expression statuses of OCT4, SOX2, NANOG based on *XIST* expression in each line are shown. (**c**) UTF1 expression status. *p*-values were calculated using the Wilcoxon's rank-sum test. (**d**) UTF1 and *XIST* expression status in SEES2 line. The expression levels of each gene corresponded with the heatmap. (**e** and **f**) Immuno-FISH analysis for H3K27me3, H2AK119ub, and *XIST* in hPSCs, motor neurons, and menstrual blood-derived cells (primary EDOM). Motor neurons were identified by co-staining with ISL1. Representative images (e) and quantifications results (f) are shown. n refers to the number of cells analyzed.
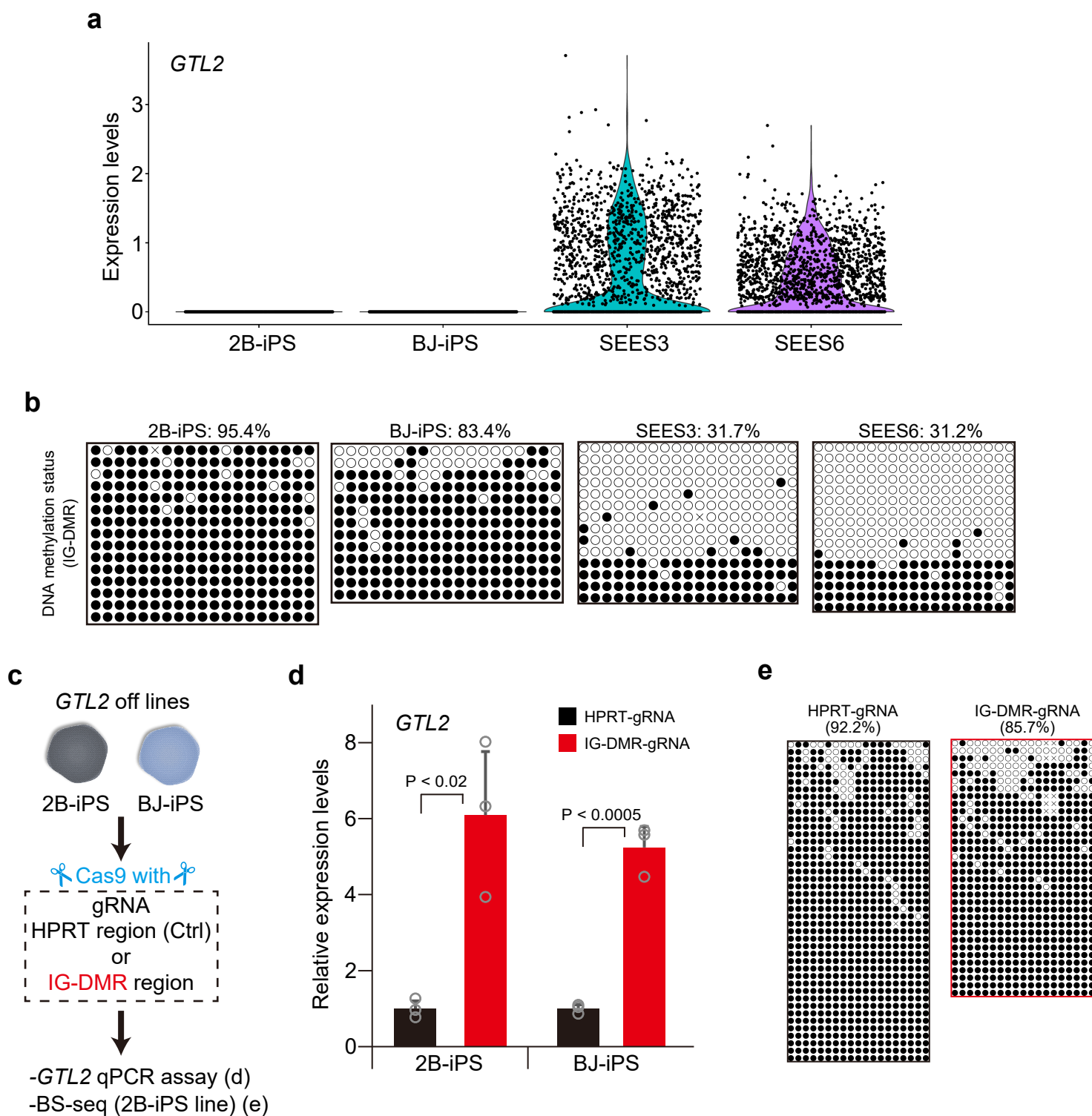
Figure S2

a 7 female hPSC lines (*XIST*+/- mix population)

Spinal motor neuron

Day 6 and 10
- PAX6 (day6), ISL1(day10) IF
- *XIST* RNA-FISH

b PAX6/*XIST*/DAPI

ISL1/*XIST*/DAPI

c ■ *XIST*- ■ *XIST*+

% of PAX6

SEES1 SEES2 SEES4 SEES5 SEES7 ADSC EDOM
(n.c=7143) (n.c=6753) (n.c=2244) (n.c=3501) (n.c=5927) (n.c=10022) (n.c=9763)
XX ESCs / XX iPSCs

% of ISL1

SEES1 SEES2 SEES4 SEES5 SEES7 ADSC EDOM
(n.c=6896) (n.c=3678) (n.c=3014) (n.c=6717) (n.c=6028) (n.c=9175) (n.c=5018)
XX ESCs / XX iPSCs

d 11 x hPSC lines

Cortical Organoid

♀ (XX): 7 lines (5 x hESC/2 x hiPSC) (*XIST*+/- mix population)
♂ (XY): 4 lines (2 x hESC/2 x hiPSC)

Day 90
- IF: PAX6/TBR1/SATB2/CTIP2
- *XIST* RNA-FISH

e SEES1(♀) / SEES6(♂)

Female (7 lines) / Male (4 lines)

% of organoid classification
■ Failed ■ Org. w/o marker ■ Org. with marker

f *XIST*/PAX6/DAPI | *XIST*/SATB2/DAPI | *XIST*/TBR1/DAPI | *XIST*/CTIP2/DAPI

g □ *XIST*- □ *XIST*+   □ Male

% of markers

PAX6 | SATB2 | TBR1 | CTIP2

PAX6 | SATB2 | TBR1 | CTIP2

h HUES66 (Velasco et al.)

scRNA-seq

Clustering → Cell type annotation → RG/IPC/Neuron/··

Marker selection

RG: *XIST*-/*XIST*+
IPC: *XIST*-/*XIST*+

Ratio of X-linked gene expression levels
(X-linked gene in *XIST*- cells) / (X-linked gene in *XIST*+ cells)

Check of cell types
(*XIST*+ in each cell type/ total *XIST*+ cells) / (*XIST*- in each cell type/ total *XIST*- cells)

i Cell type clusters / *XIST*

■ Ex.(SATB2) ■ Ex.(CTIP2) ■ Ex.(CCNI) ■ Ex.(NEUROD2) ■ Inhibitory ■ RG ■ RG/IPC ■ IPC ■ IPC/Neuron ■ Ex.(RELN) ■ Mural

UMAP / High Low

j Expression ratio of X-linked gene in *XIST*+ and - cells based on median levels (*XIST* off/ *XIST* on)

RG | RG/IPC | IPC | IPC Neuron | Ex.(CTIP2) | Ex.(SATB2) | Ex.(CCNI) | Ex.(NEUROD2)

Ratio

Cell type markers (n= No. of X-linked genes analyzed)

k Log₂ Ratio of cell population (*XIST*+/*XIST*-)

RG | RG/IPC | IPC | IPC/Neuron | Ex.(CTIP2) | Ex.(SATB2) | Ex.(CCNI) | Ex.(NEUROD2)

Cell type markers (n= No. of X-linked genes analyzed)

l Trajectory

Ex.(SATB2) | IPC/Neuron | IPC | Ex.(CTIP2) | Ex.(NEUROD2) | Ex.(CCNI) | RG/IPC | RG

UMAP / Pseudotime

Figure S2 related to Figure 2. *XIST* **expression status is linked to the stages of neuronal differentiation.** (**a**) Experimental scheme for MN differentiation from hPSCs using a 2D culture protocol. (**b** and **c**) Immunofluorescence combined with *XIST* RNA-FISH analysis in differentiating cells. The picture shows the representative image (b), and the quantification results are shown as bar graphs (c). The percentage of PAX6+ or ISL1+ cells among *XIST*+ or – cells was calculated. Each dot and n.c indicate the percentage in the observed area and the total number of cells analyzed, respectively. p-values are calculated using Student's *t*-test. Scale bar represents 50 µm. (**d**) Experimental scheme for generating cortical organoids from hPSCs. On day 90, the samples were subjected to immuno-*XIST* RNA-FISH analysis. Representatives for progenitor cell (PAX6) and cortical layers (SATB2, CTIP2, and TBR1) were used. (**b**) Representative image of a cortical organoid and developmental summary at day 90. Scale bar represents 1 mm. (**c**) Representative images of immuno-*XIST* RNA-FISH analysis in cortical organoid samples. Scale bar represents 50 µm. (**d**) Quantification results of each marker. In female samples, the marker expression status was based on *XIST* expression. Each dot and n.c indicate the percentage in the observed area and the total number of cells analyzed, respectively. Error bars show the standard deviation. *p*-values are calculated by Student's *t*-test. (**h**) Experimental scheme for analyzing cell types based on the *XIST* expression status in the HUES66 line (Velasco *et al.*, 2019). For analysis of the ratio of X-linked gene expression levels, X-linked genes expressed in at least 20 cells in both *XIST*+ and – groups were used. The ratios were based on the median expression levels. Each cell type annotation was based on a previous report (Bhaduri *et al.*, 2020), and the marker genes used for classifications were identified using MAST. (**i**) UMAP based on cell types (left) and *XIST* expression (right). RG; radial glia, IPC; intermediate progenitor cell, Ex.; excitatory. Ex. Inhibitory; inhibitory neuron. (**j**) Expression ratio of X-linked genes in *XIST*+ and – cells based on median levels. The heatmap shows the ratio of X-linked genes (*XIST*-/*XIST*+), indicating that red is overexpressed in *XIST*- cells. n refers to the number of X-linked genes analyzed. Representative marker genes used for the classifications are shown as cell type markers. SATB2 cells were separated into three types based on the markers. (**k**) The ratio of cell types based on the *XIST* expression status. Log$_2$ ratio (*XIST*+/*XIST*-) was shown in each cell type. (**l**) Trajectory analysis of cortical organoid. Pseudotime was calculated and the cells were highlighted based on the scores.
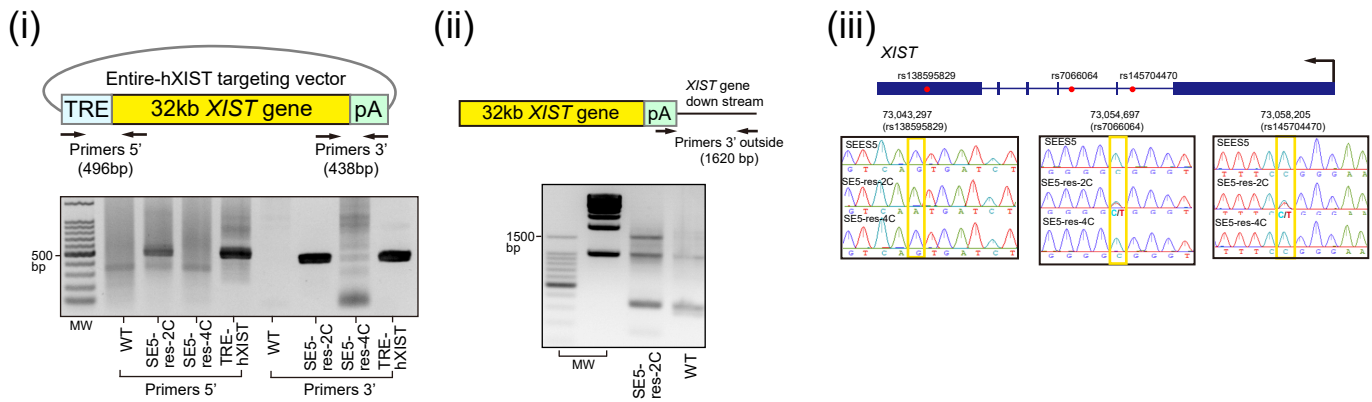
Figure S3

a

b

c

GTL2 off lines

2B-iPS   BJ-iPS

✂ Cas9 with ✂

gRNA
HPRT region (Ctrl)
or
IG-DMR region

-*GTL2* qPCR assay (d)
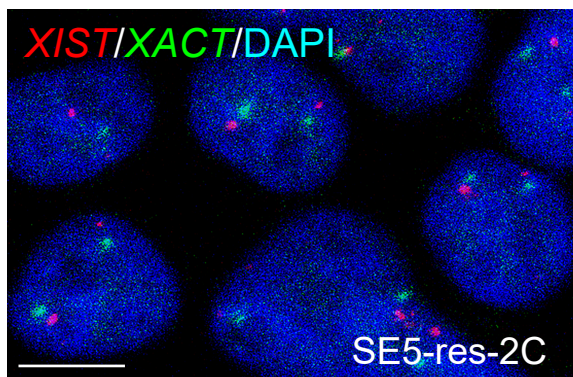-BS-seq (2B-iPS line) (e)

d

e

Figure S3 related to Figure 3. Reactivation of methylated *GTL2* by Cas9-mediated DSB. (**a**) *GTL2* expression in male hiPSC lines. Single-cell RNA-sequencing (scRNA-seq) data are shown. *GTL2* was not detected in hiPSC lines. (**b**) BS analysis of IG-DMR in *GTL2*^{off} and *GTL2*^{on} lines. (**c**) Experimental scheme for *GTL2* reactivation and DNA demethylation by Cas9-mediated DSB. *GTL2*^{off} lines were used for the TaqMan gene expression assays (d) and BS-seq analysis (e). gRNA targeting *HPRT1* was used as a negative control. At 144 h after transfection, the cells were analyzed. (**d**) qPCR results for *GTL2* expression. The average expression level of the HPRT1-gRNA sample was set to 1 in each line. Each dot indicates an independent experiment. Error bars show the standard deviations. *p*-values were calculated using Student's *t*-test. (e) BS-seq analysis of IG-DMR in Cas9/gRNA transfected 2B-iPSCs.
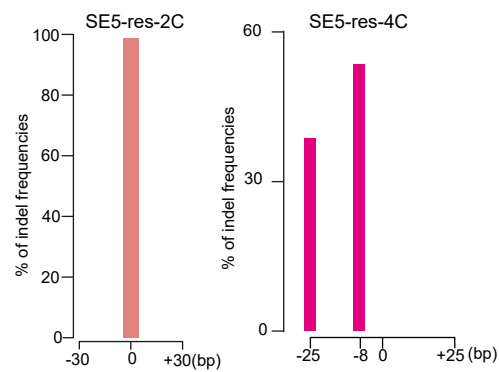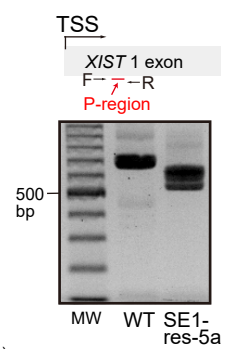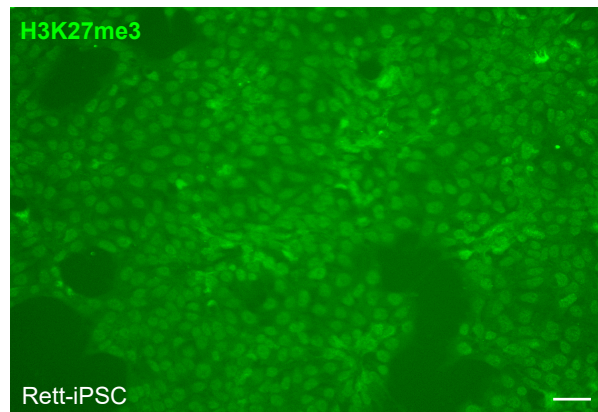
Figure S4

a

(i)

Entire-hXIST targeting vector

TRE | 32kb *XIST* gene | pA

Primers 5' (496bp)   Primers 3' (438bp)

500 bp

MW | WT | SE5-res-2C | SE5-res-4C | TRE-hXIST | WT | SE5-res-2C | SE5-res-4C | TRE-hXIST

Primers 5'   Primers 3'

(ii)

32kb *XIST* gene | pA

XIST gene down stream

Primers 3' outside (1620 bp)

1500 bp

MW | SE5-res-2C | WT

(iii)

XIST

rs138595829   rs7066064   rs145704470

73,043,297 (rs138595829) | 73,054,697 (rs7066064) | 73,058,205 (rs145704470)

SEES5 / SE5-res-2C / SE5-res-4C

b

XIST/*XACT*/DAPI

SE5-res-2C

c

SE5-res-2C

% of indel frequencies

-30   0   +30 (bp)

SE5-res-4C

% of indel frequencies

-25 -8 0   +25 (bp)

d

TSS

XIST 1 exon

F→ ←R
P-region

500 bp

MW | WT | SE1-res-5a

e

Summary of H3K27me3 rescue clones
(> 85% of H3K27me3)

Lines	Recombination	P-region	
		NHEJ	Genotype
SE1-res-5a		◯	Mix
SE5-res-2C	◯	n.d	Homo
SE5-res-4C	n.d	◯	Mix

f

H3K27me3

Rett-iPSC

g

R306H

gDNA

A A G A A G C **G/A** C A A G A C C

cDNA

C A A G A A G C A C A A G A C C

CGC (R): Healthy
CAC (H): Mutant

h

Rett-iPSC:*XIST*^re-on   HDR(Δ0-*XIST*)

100% normal
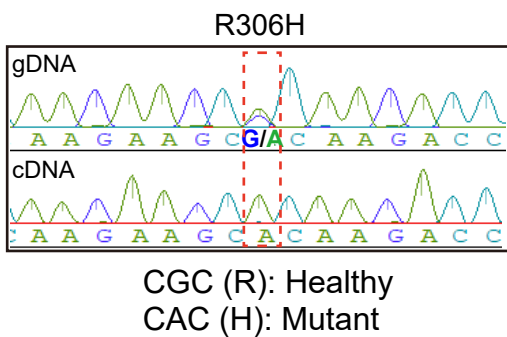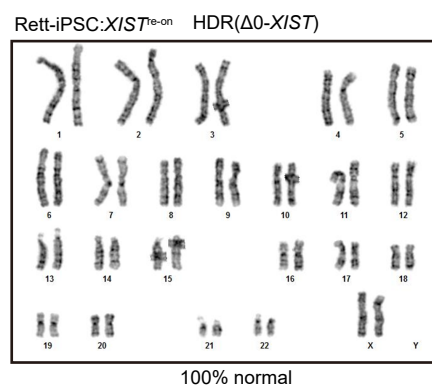
Figure S4 related to Figure 3 and 4. Genotyping and DNAme status in *XIST* re-expressed lines and characterization of Rett-iPSCs. (**a**) Genotyping PCR analysis of *XIST* re-expressing lines in SEES5 cells. TRE and SV40 sequences were included in Entire-hXIST vector (i and ii). Sanger sequencing analysis of SNVs (iii). (**b**) DNA-FISH assay using *XIST* and *XACT* probes in SE5-res-2C. The scale bar represents 10 μm. (**c**)TIDE analysis for checking scars. The mutant allele was not detected in the SE5-res-2C line, whereas the SE5-res-4C line comprised cells with two different alleles. (**d**) Genotyping PCR spanning the P-region in SE1-res-5a. (**e**)Summary of the obtained H3K27me3 foci rescue clones. Mix refers to two different alleles identified by TIDE analysis or Sanger sequences. n.d. refers to not determined. (**f**) H3K27me3 staining of Rett-iPSC. Scale bar represents 50 μm. (**g**) Genotyping and expression allele analysis using gDNA and cDNA. The mutation point is highlighted in the red dashed line. (**h**) G-band analysis of the Rett-iPSC: $XIST^{\text{re-on}}$ (Δ0-*XIST*).
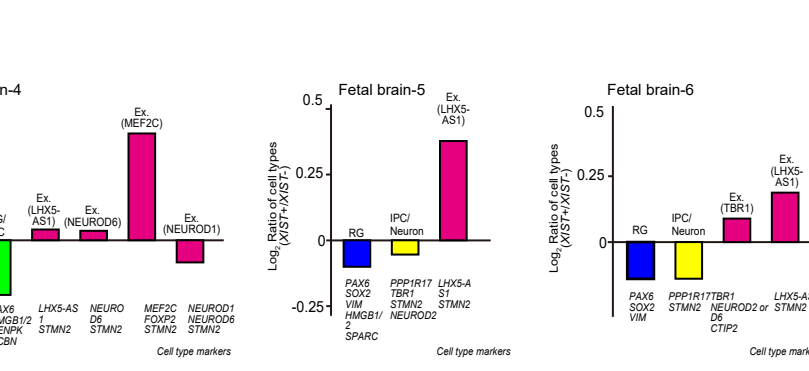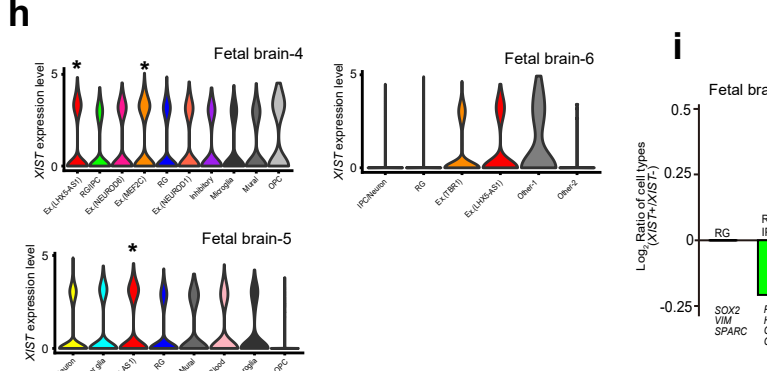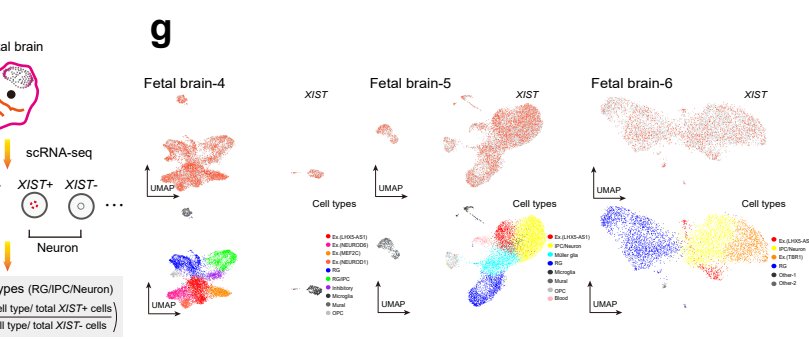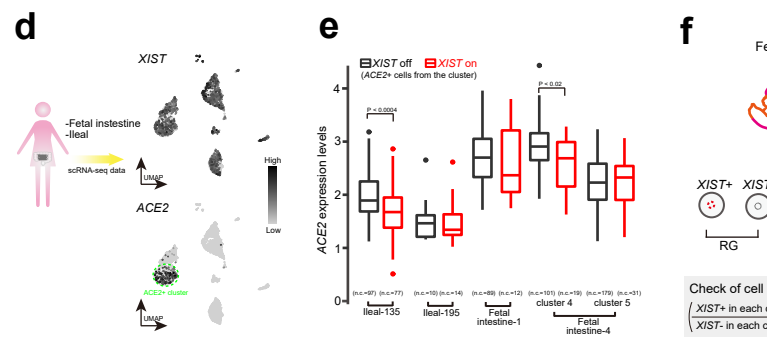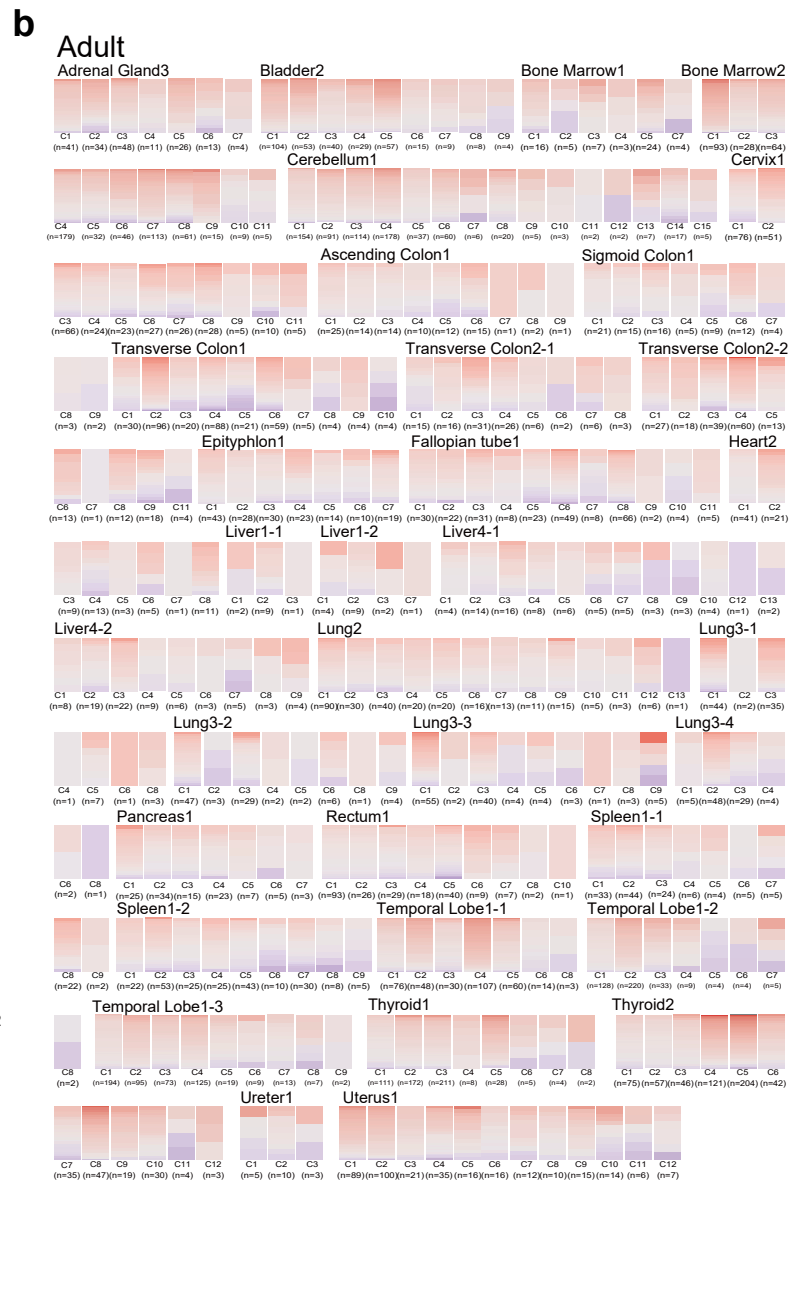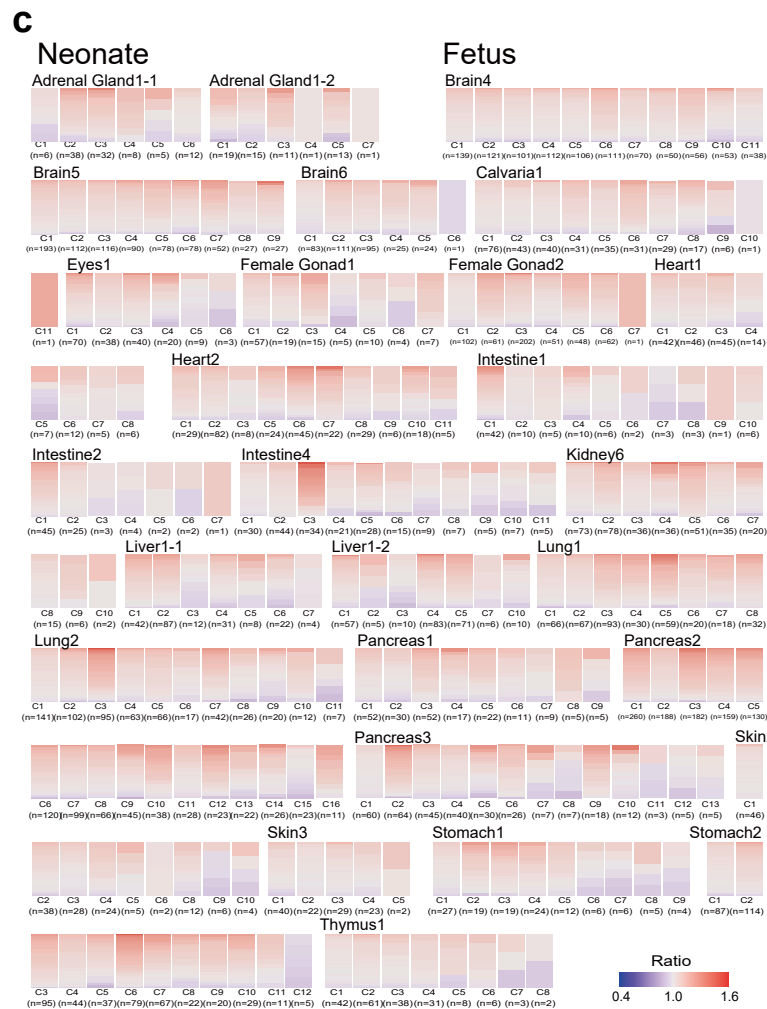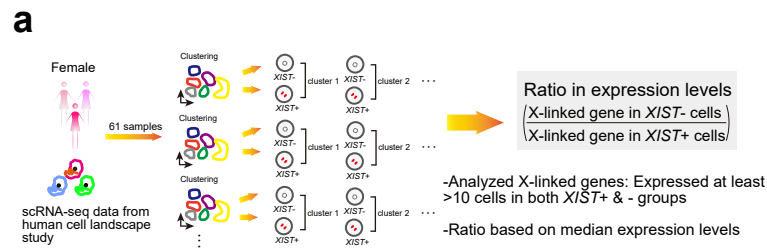
Figure 5

Figure S5 related to Figure 1, 2 and 5. Effect of the *XIST* expression status *in vivo*. (a) Analysis scheme using 61 scRNA-seq datasets from human cell atlas (Han *et al.*, 2020). (**b** and **c**) The ratios (*XIST*– cells/*XIST*+ cells) in adult tissues (b) and neonatal and fetal tissues (c). C and n indicate the cluster identity and number of X-linked genes analyzed, respectively. (**d**) Experimental scheme for identifying *ACE2*+ clusters from the fetal intestine and ileal cells. The representative images of UMAP analysis with *XIST* and *ACE2* expression are shown. The *ACE2*+ clusters were identified by MAST analysis. (**e**) *ACE2* expression status based on the *XIST* expression levels. The cells from *ACE2*+ clusters were isolated, and the *ACE2* expression levels were compared based on the *XIST* expression status. n.c. refers to the number of cells analyzed. *p*-values were calculated using the Wilcoxon's rank-sum test. (**f**) Experimental scheme for evaluating neuronal development in vivo based on the *XIST* expression status. Three fetal brain samples were used. The gene sets for cell-type annotation were identified by MAST analysis and classifications were based on the previously identified markers (Bhaduri *et al.*, 2020; Camp et al., 2015). (**g**) UMAP analysis. The *XIST* expression status (upper) and cell types were highlighted by each color (bottom). (**h**) The *XIST* expression levels in each cell type based on scRNA-seq results. The violin plot shows the *XIST* expression status in each cell type. Asterisks indicate statistical significance determined using MAST test with Bonferroni correction. All p-values were < 7.4 E-9. (**i**) The ratio of cell types based on the *XIST* expression status. Log_2 ratio (*XIST*+/*XIST*-) was shown in each cell type. The representative marker genes used for the classification were shown.