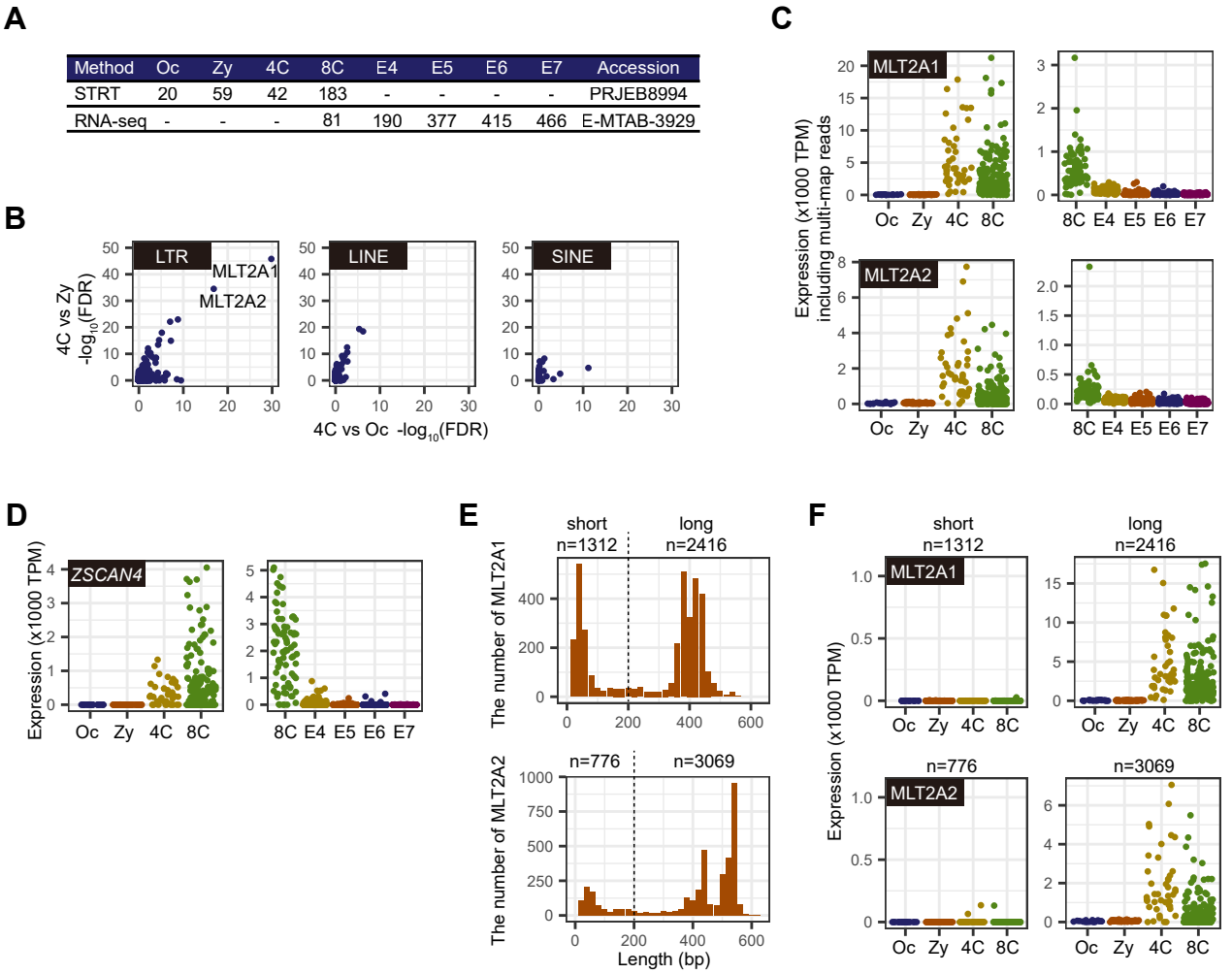
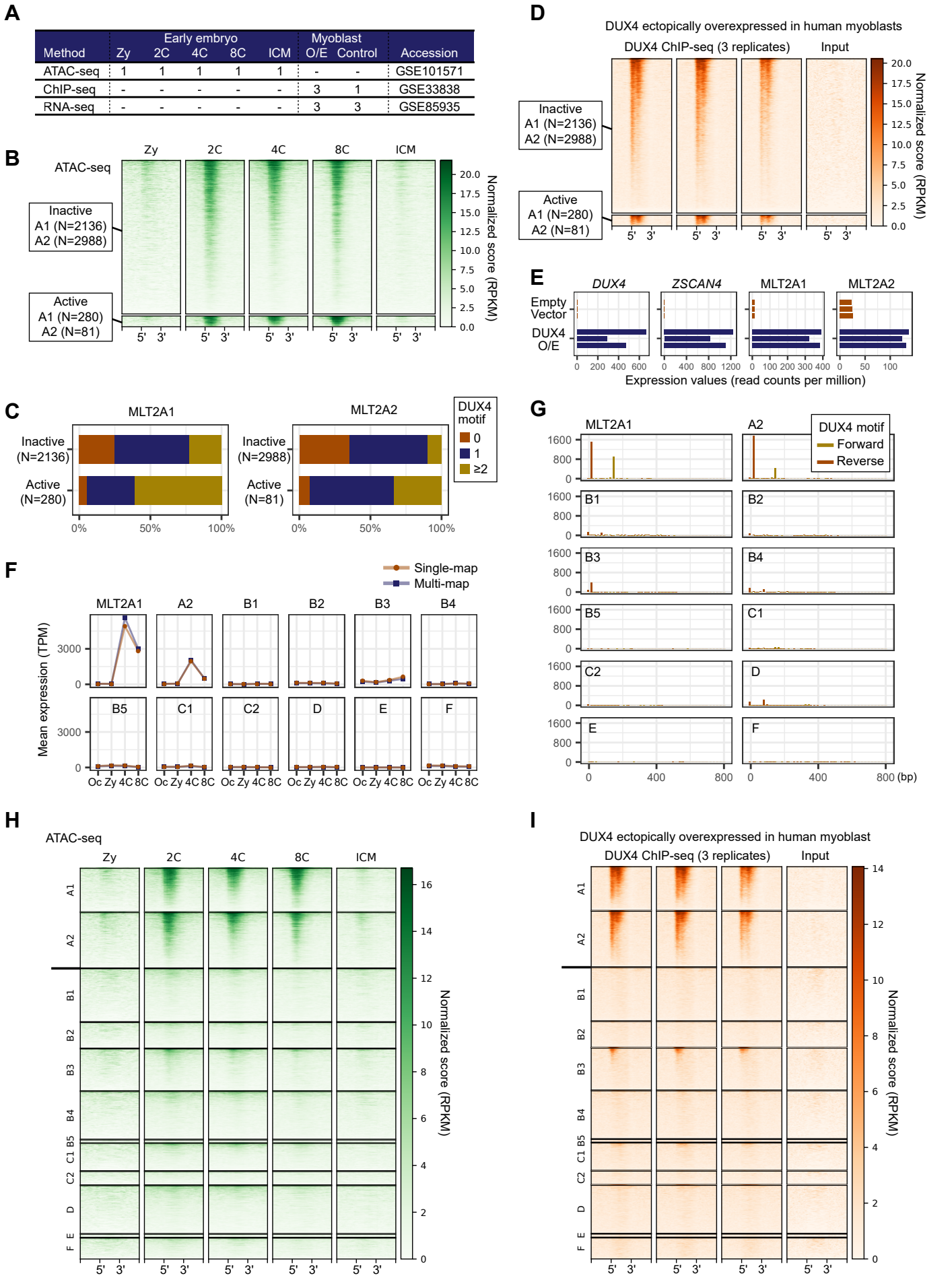
A

Method	Oc	Zy	4C	8C	E4	E5	E6	E7	Accession
STRT	20	59	42	183	-	-	-	-	PRJEB8994
RNA-seq	-	-	-	81	190	377	415	466	E-MTAB-3929



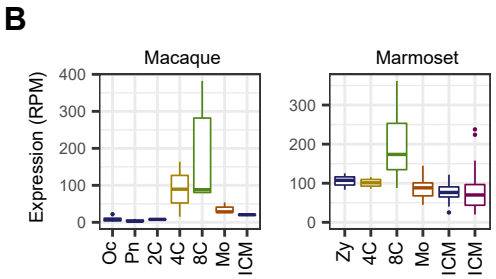
Supplemental Figure S1. Activation of MLT2A1 and MLT2A2 copies.
(A) Details of datasets used in the expression analysis. Oc, oocyte; Zy, zygote; 4C, 4-cell embryo; 8C, 8-cell embryo; E4 to E7, embryo on day 4 to 7. (B) Scatter plots of -$\log_{10}$(FDR) values for differential expression between 4-cell embryo and oocyte (x-axis) and between 4-cell embryo and zygote (y-axis) including multi-map reads; the plots excluding multi-map reads are shown in Fig. 1a. Each dot represents a family of LTR, LINE, or SINE. (C) Normalized expression values (TPM, tags per million mapped tags) of MLT2A1 and MLT2A2 elements in each single-cell including multi-map reads; the expression values excluding multi-map reads are shown in Fig. 1b. STRT data from oocyte to 8-cell embryo are shown in the left panels and RNA-seq data from 8-cell embryo on day 7 are shown in the right panels. (D) Normalized expression values of ZSCAN4 in each single-cell excluding multi-map reads. (E) Distribution of the length of MLT2A1 and MLT2A2 elements defined by RepeatMasker. (F) Normalized expression values of MLT2A1 and MLT2A2 elements separated into two groups according to their length: short truncated elements are ≤ 200 bp and long elements are > 200 bp.

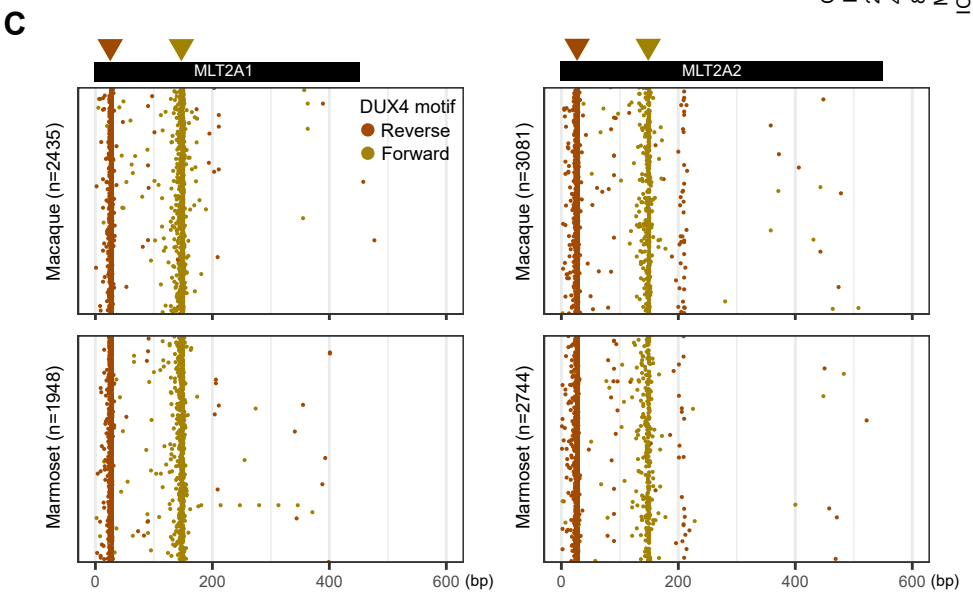


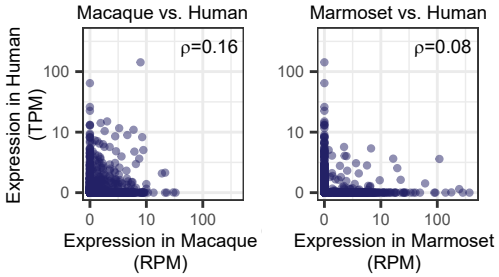
Supplemental Figure S2. MLT2A1 and MLT2A2 elements regulated by DUX4.
(A) Details of datasets used in the chromatin, DUX4 binding, and RNA-seq analyses. Zy, zygote; ICM, inner cell mass; O/E, overexpression of DUX4 (B) Chromatin states of inactive (upper subpanel) and active (lower subpanel) MLT2A elements and flanking regions (+/- 1 kb) at 5 stages. All elements are scaled to the same size, with 5' and 3' denoting their ends. In each subpanel, the elements are sorted from top to bottom according to ATAC-seq signal intensity. (C) Percentages of inactive and active MLT2A elements with 0, 1, or 2 or more DUX4 motifs. (D) DUX4-binding states of inactive and active elements and their flanking regions (+/- 1 kb). In each subpanel, the elements are sorted from top to bottom according to DUX4 ChIP-seq signal intensity. (E) Expression values of DUX4, ZSCAN4, MLT2A1, and MLT2A2 in human myoblasts with or without DUX4 overexpression. (F) Average expression patterns from oocyte (Oc) to 8-cell embryo for 12 different MLT2 families. (G) Number of DUX4 motifs in the reverse and forward orientations at different positions for 12 MLT2 families. (H) Chromatin states of all elements and their flanking regions (+/- 1 kb) for 12 MLT2 families. (I) DUX4 binding states of all elements and their flanking regions (+/- 1kb) for 12 MLT2 families.

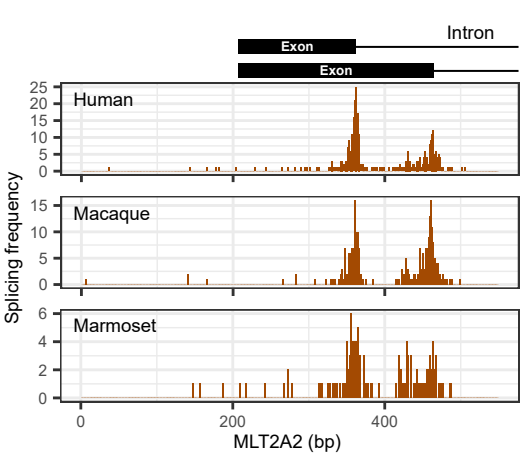
A

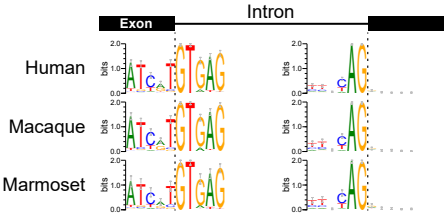
	Oc	Zy	2C	4C	8C	Mo	ICM	Accession
Macaque	6	3	3	2	5	3	4	GSE86938
Marmoset	0	3	0	4	15	53	42+79	E-MTAB-7078

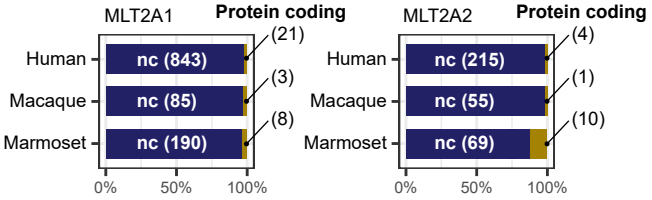








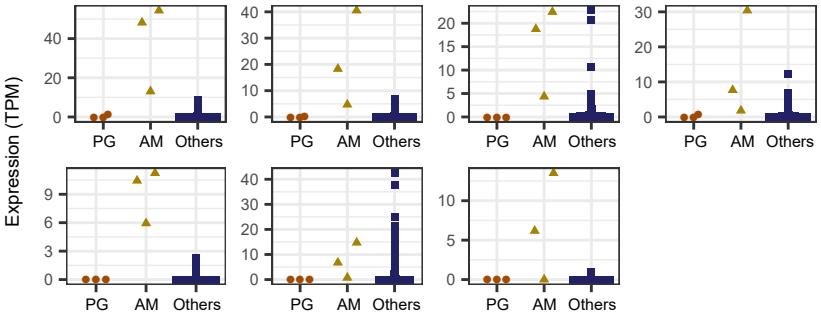




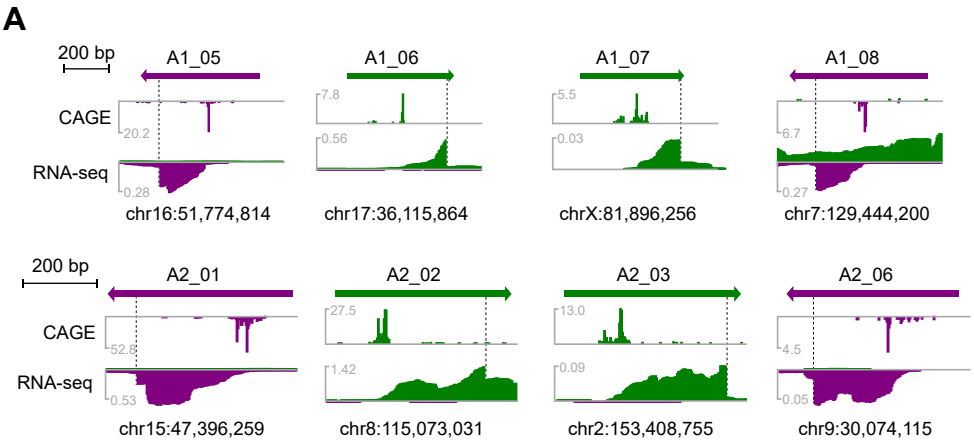
H

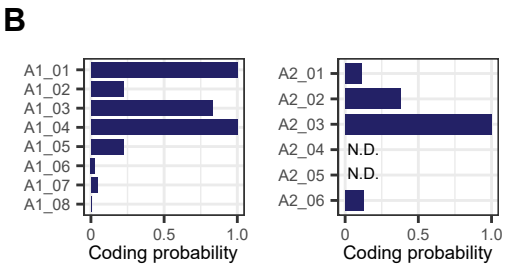
Human		Macaque	Marmoset	
ACSM2A	*SATB2*	*GNRH1*	*CGA*	*SH3BGRL*
ACSM2B	*SGCG x 2*	*LOC100428646*	*DAB2*	*SLC24A5*
ALG13	*SH3BGRL*	*PDK1*	*FAM168A*	*SPAG1*
C12orf56 x 2	*SHC4*	*SH3BGRL*	*FBXO30*	*TMEM132B*
EEF1E1	*SPAG1*		*LOC108589791*	*TPK1*
GBE1	*SPATA5*		*METAP1*	*TTLL11*
GRM8	*SSH2*		*NEK10*	*USH2A*
MTAP	*SUMF1*		*PANX1*	*YPEL2*
NDUFAF6	*TFB1M*		*PDE7A*	*ZNF277*
PDHX	*TFIP11*			
PLSCR4	*UBE3D x 2*			

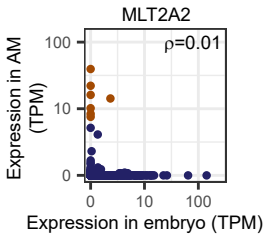
Supplemental Figure S3. Activation of MLT2A elements in primate embryos.
(A) Details of the datasets used in the analysis of splice sites for macaque and marmoset embryos. Oc, oocyte; Zy, zygote; 2C, 2-cell embryo; 4C, 4-cell embryo; 8C, 8-cell embryo; Mo, morula; ICM, inner cell mass. (B) Boxplots of normalized expression values (RPM, reads per million mapped reads) of MLT2A2 elements in oocyte (Oc), pronuclei (Pn), 2-cell, 4-cell, and 8-cell embryos (2C, 4C, and 8C, respectively), morula (Mo), and inner cell mass (ICM). (C) Distributions of DUX4-binding motifs (MA0468.1) in all MLT2A elements; motifs were identified by using MAST (Motif Alignment and Search Tool). Motifs in the reverse and forward orientations are denoted as brown and yellow circles, respectively. (D) Normalized expression values of individual MLT2A2 elements in macaque vs. human and marmoset vs. human embryos. The numbers of elements robustly transcribed in two species (more than 5 TPM and 5 RPM) are 4 for macaque vs. human and 1 for marmoset vs. human. ρ, Spearman's correlation coefficients. (E) Frequency of splicing events in MLT2A2 elements of human, macaque, and marmoset. (F) Sequence logos of splice sites in MLT2A2 found in 4-cell and 8-cell embryos of human, macaque, and marmoset. Left side shows donor sites inside MLT2A2 and right side shows acceptor sites outside MLT2A2. (G) Percentages of MLT2A elements that are connected to non-coding (nc) regions and protein-coding exons in human, macaque, and marmoset embryos, based on the analysis of spliced reads from all 4- and 8-cell embryos. (H) Genes that have spliced junctions with MLT2A elements.



Supplemental Figure S4. Expression of MLT2A elements in amniotic membrane.
Normalized expression values of individual MLT2A2 elements that are actively transcribed in amniotic membranes. Expression values are shown separately for 3 pineal glands (PG), 3 amniotic membranes (AM), and 712 other samples.

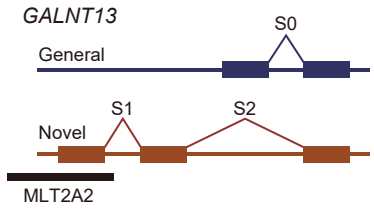


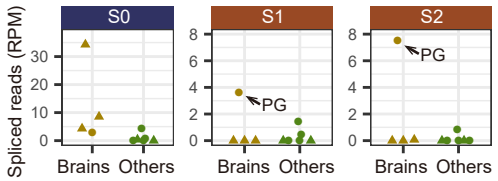


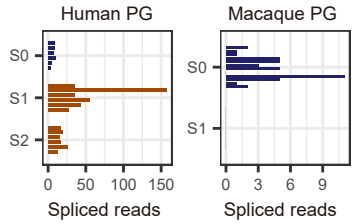


D

	Accession	Tissue	Sample
Human	DRA000991	Pineal gland	1
		Cerebellum	1
		Substantia nigra	1
		Pituitary gland	1
		Testis	1
		Thymus	1
		Epididymis	1
		Parotid gland	1
		Fetal thyroid	1
		Fetal heart	1
Human	GSE100472	Pineal gland	6
Macaque	GSE78165	Pineal gland	12







H

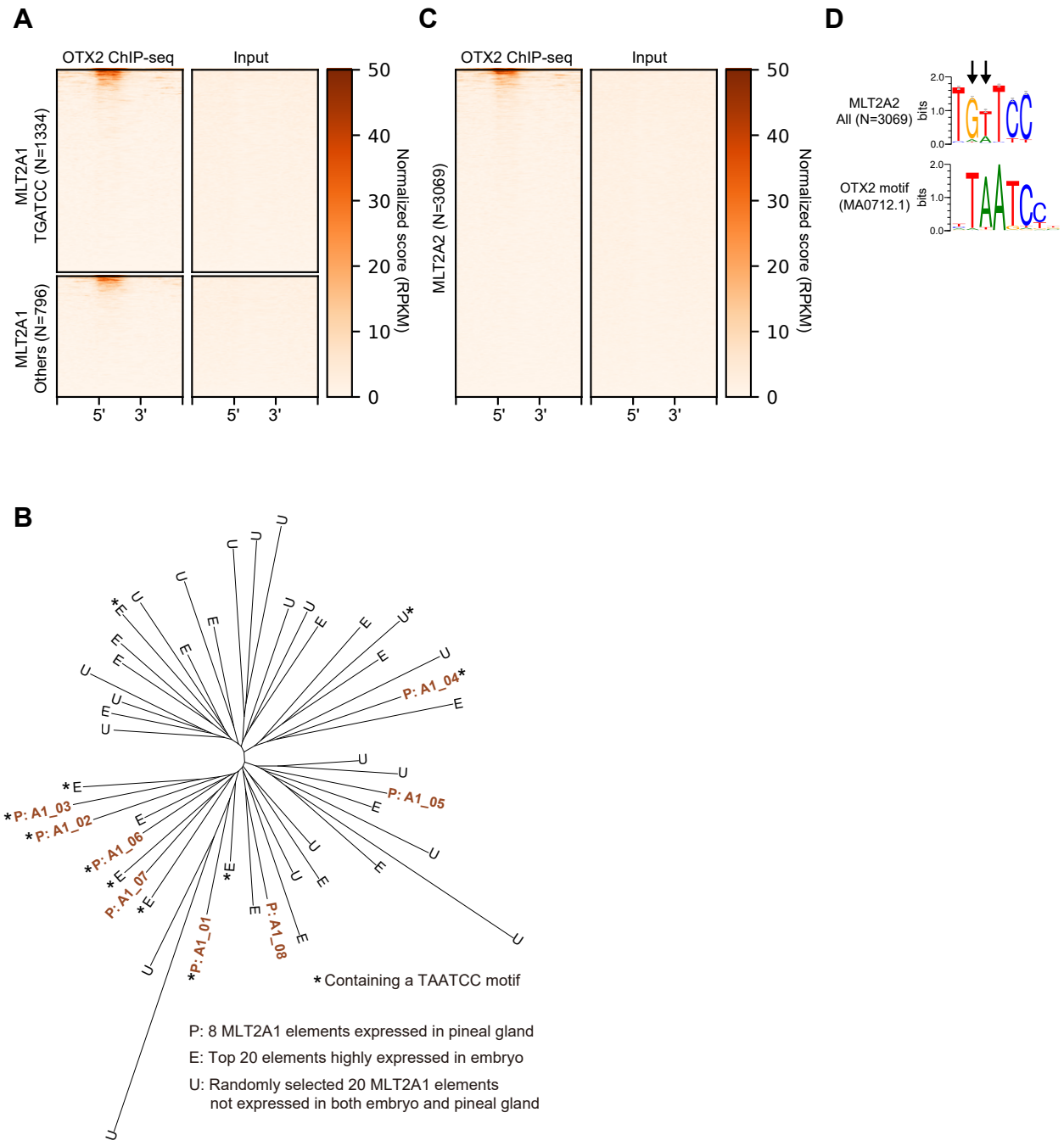
Human

Gene	Splice	Coordinates (hg38)
ABCE1	S0	chr4:145,098,420-145,104,386
	S1	chr4:145,034,496-145,072,916
	S2	chr4:145,073,036-145,104,386
	S2	chr4:145,073,056-145,104,386
	S3	chr4:145,034,496-145,104,386
COL5A1	S0	chr9:134,642,297-134,690,912
	S1	chr9:134,648,265-134,690,912
GALNT13	S0	chr2:153,872,304-153,900,936
	S1	chr2:153,409,242-153,732,769
	S2	chr2:153,732,831-153,900,936

Macaque

Gene	Splice	Coordinates (rheMac8)
ABCE1	S0	chr5:144,961,181-144,967,164
	S1	chr5:144,898,064-144,967,164
COL5A1	S0	chr15:3,650,976-3,704,634
	S1	chr15:3,650,976-3,699,059
GALNT13	S0	chr12:40,473,546-40,501,493
	S1	chr12:39,917,661-40,501,493

Supplemental Figure S5. Pineal gland-specific novel promoters and splice sites.
(A) Expression patterns of individual MLT2A2 elements actively transcribed in pineal glands. CAGE captures the 5' ends of the transcripts, whereas RNA-seq captures entire exons. Spliced sites are denoted by dashed lines. Forward and reverse orientations are indicated by green and purple, respectively. (B) Coding probabilities of MLT2A-derived transcripts predicted by CPAT software. N.D. indicates that the transcripts were detected by CAGE, but not by RNA-seq. (C) Normalized mean expression values (TPM, tags per million mapped tags) of individual MLT2A2 elements in early embryo (N = 225, all 4-cells and 8-cells) and amniotic membrane (AM, N = 3). The 6 elements actively transcribed in amniotic membrane are shown in brown. (D) Details for the datasets used in the analysis of the splice junctions of ABCE1, COL5A1, and GALNT13 genes. (E) Schematic representation of splice sites of GALNT13. Novel splice junctions derived from MLT2A2 are denoted as S1 and S2, whereas a common splice junction is denoted as S0. (F) Normalized splice counts (RPM, spliced reads per million mapped reads) in 4 human brain (yellow) and 6 non-brain (green) tissues for GALNT13. Pineal glands are denoted by yellow circles. (G) Normalized splice counts of GALNT13 in human and macaque pineal gland. The human samples are from 6 different donors (GSE100472), and the macaque samples are from 12 different individuals (GSE78165). (H) Coordinates of splice junctions for ABCE1, COL5A1, and GALNT13 in human and macaque genomes.



Supplemental Figure S6. OTX2 as a potential regulator of MLT2A elements in pineal gland.
(A) OTX2 binding states in MLT2A1 elements and flanking regions (+/- 1 kb) for ChIP-seq and input samples. (B) Phylogenetic tree of 48 MLT2A1 elements. (C) OTX2 binding states in MLT2A2 elements and flanking regions (+/- 1 kb) for ChIP-seq and input samples. (D) Sequence logos of MLT2A2 and the original OTX2 motif.