Nucleic Acids Research, 2026, **54**, gkag577
https://doi.org/10.1093/nar/gkag577
Gene regulation, Chromatin and Epigenetics

IAPEz retrotransposons regulate the first lineage segregation in mouse pre-implantation development

Rujin Huang[1,2,3,4,†], Zhaoya Han[1,4,†], Zhongqi Liufu[5,†], Xiongzhi Quan[2], Danfeng Li[2], Hongzhi Dong[2], Guangliang Hong[6], Ruhai Chen[2], Man Zhang[4], Guangdun Peng[2], Duanqing Pei[7], Jiekai Chen[1,2,*], Jiangping He[3,4,6,*]

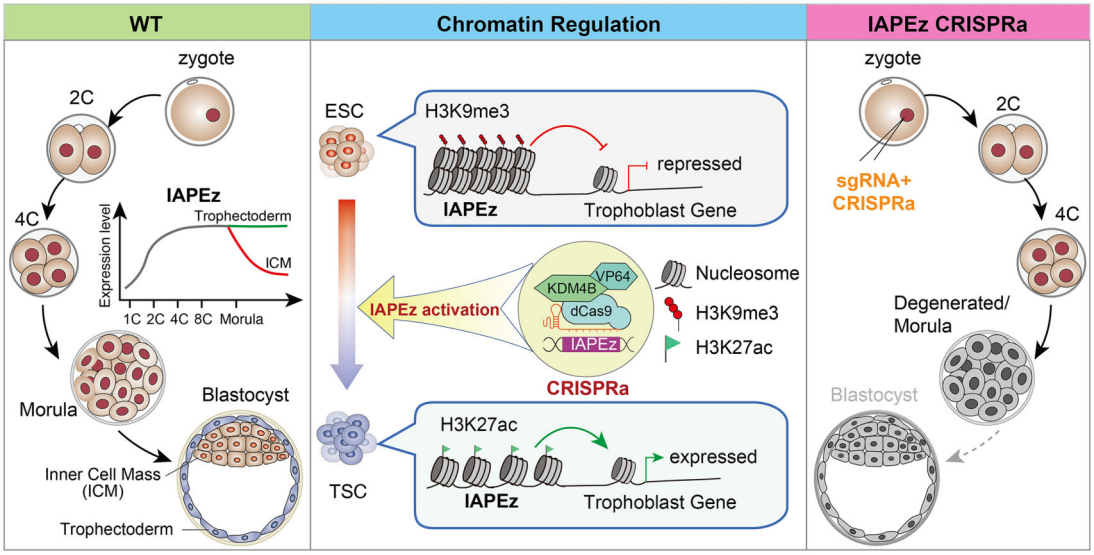
[1]Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China; Guangzhou Medical University, Guangzhou 511436, China
[2]Center for Biomedical Digital Science, Guangdong Provincial Key Laboratory of Stem Cell and Regenerative Medicine, Guangdong-Hong Kong Joint Laboratory for Stem Cell and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China
[3]The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, China
[4]Guangzhou National Laboratory, Guangzhou, China
[5]State Key Laboratory of Genetic Resources and Evolution and Yunnan Key Laboratory of Biodiversity Information, Kunming Institute of Zoology, The Chinese Academy of Sciences, Kunming, China
[6]Faculty of Life and Health Sciences, Shenzhen University of Advanced Technology, 518107 Shenzhen, China
[7]School of Life Sciences, Westlake University, Hangzhou, China

*To whom correspondence should be addressed. Email: chen_jiekai@gibh.ac.cn
Correspondence may also be addressed to Jiangping He. Email: hejiangping@suat-sz.edu.cn
†These authors contributed equally to this work.

Abstract

Transposable elements (TEs) comprise nearly half of mammalian genomes and drive species-specific regulatory innovation, but their contributions to the first lineage segregation—establishing trophectoderm versus inner cell mass (ICM)—remain largely unexplored. Here, we identify IAPEz, a rodent-specific retrotransposon, as a key regulator of this process in mouse pre-implantation embryos. IAPEz is highly expressed in the extraembryonic lineage but silenced in the ICM via H3K9me3-dependent heterochromatin. Genome-wide interaction analyses reveal that IAPEz physically contacts hundreds of trophectoderm-associated genes, repressing trophoblast programs in the ICM to ensure proper lineage allocation. CRISPR-mediated activation of IAPEz in embryonic stem cells accelerates direct conversion to trophoblast stem cells, while zygotic activation induces defective lineage specification at the morula stage. These findings highlight how repression of lineage-specific TEs orchestrates early embryonic fate decisions, offering insights into genome evolution and species-specific development.

Graphical abstract



Received: April 16, 2026. **Revised:** May 11, 2026. **Accepted:** May 19, 2026

Introduction

Transposable elements (TEs) are the most abundant components of both the human and mouse genomes. During early embryonic development in mice, TEs exhibit stage-specific dynamic expression patterns that are closely associated with embryonic development. In two-cell (2C) embryos and two-cell-like cells (2CLCs), the retrotransposons MT2/MERVL—a type of endogenous retrovirus (ERV)—are highly expressed and are considered one of the earliest transcripts activated during zygotic genome activation (ZGA) [1, 2]. Their activation is often used as a hallmark of the activation of the totipotency transcriptional regulatory network [3]. Repression of MT2/MERVL leads to embryonic lethality during the pre-implantation stage [4, 5]. In mouse embryonic stem cells (mESCs), direct activation of the expression of MT2/MERVL or its regulatory factor Dux can induce reprogramming of pluripotency into a totipotent-like state [6, 7]. Additionally, LINE1 (long interspersed nuclear element-1) exhibits dynamic expression during early embryonic development, and either overactivation or suppression of LINE1 expression *in vivo* is detrimental to development beyond the two-cell stage and into the blastocyst [8]. Conversely, knockdown of LINE1 *in vitro* promotes the transition of mESCs toward a totipotent-like state [9].

In primates, analogous species-specific TEs are reactivated during early development [10]. For example, human ERVL (HERVL, also known as MLT2A1/HERVL-int) is specifically expressed in eight-cell (8C) embryos and eight-cell-like cells (8CLCs), suggesting its potential role in human ZGA and totipotency [10, 11]. Furthermore, transient activation of the primate-specific non-long terminal repeat (LTR) retrotransposon subtype SINE-VNTR-Alu (SVA), which functions as an enhancer during the 8C stage, leads to activation of totipotency-related genes in humans [12]. In human pluripotent stem cells (hPSCs), retrotransposons such as LTR7/HERVH, LTR5_Hs/HERVK, and SVA have also been reported to play critical roles in the regulation of naïve pluripotency and early embryonic development [13–16]. Beyond pluripotency, primate ERVs function as lineage-specific enhancers; for example, LTR6B drives TET1-mediated transcription of definitive endoderm genes [17], and ERV-derived transcripts such as SUPYN and BANCR regulate cardiomyocyte specification [18, 19]. These examples underscore how TEs, through co-option as regulatory elements, shape lineage speciation and highlight their evolutionary role in adapting developmental programs to species-specific needs.

The first lineage segregation in mammals establishes the embryonic inner cell mass (ICM), which gives rise to the fetus, and the extraembryonic trophectoderm, essential for implantation and placentation [20]. This process differs markedly between species: in mice, blastocysts form the ICM and trophectoderm, with the trophectoderm differentiating into the placenta and supporting the egg-cylinder morphology [21]; in humans, three lineages emerge [ICM, trophectoderm, and primitive endoderm (PrE)], leading to disc-shaped embryos [21, 22]. Transcription factors (e.g. Oct4, Sox2, Cdx2, Eomes, Tead4, and Elf5) and pathways such as Hippo signaling govern this segregation [23–28], yet these are highly conserved and cannot fully account for interspecies differences. TEs, with their rapid evolution and species-specific insertions, may bridge this gap by providing divergent regulatory landscapes.

In this study, we systematically explored the functional contributions of TEs to the first lineage segregation in humans and mice. Leveraging single-cell transcriptomics, we uncovered lineage-specific TE expression patterns in the ICM and trophectoderm of blastocysts from both species. Notably, the rodent-specific retrotransposon IAPEz (including the internal sequence IAPEz-int and its flanking LTRs, IAPLTR1_Mm and IAPLTR1a_Mm) emerges as a key player: it is robustly expressed from the 2C to morula stages prior to lineage segregation, persists at high levels in the trophectoderm lineage post-segregation, and is selectively silenced in the ICM through H3K9me3-mediated heterochromatin. Mechanistically, IAPEz physically interacts with numerous genes associated with extraembryonic development, repressing their expression in the ICM via H3K9me3-dependent mechanisms—higher at IAPEz loci in embryonic stem cells (ESCs) than in trophoblast stem cells (TSCs). Clustered regularly interspaced palindromic repeats (CRISPR)-mediated activation (CRISPRa) of IAPEz accelerates ESC-to-TSC transition *in vitro* and disrupts trophectoderm allocation *in vivo*, culminating in defective lineage specification at the morula stage. Collectively, these findings establish IAPEz as a critical regulator of mouse pre-implantation development, shedding light on how species-specific TEs enforce lineage fidelity and highlighting their broader evolutionary impact on mammalian embryogenesis.

Materials and methods

Animals

All mice were housed in isolated ventilated cages (maximum six mice per cage) at the barrier facility at Guangzhou Regenerative Medicine and Health. The mice were maintained on a 12/12 h light/dark cycle, 22–26°C with sterile pellet food and water *ad libitum*. All animal protocols used in this study were approved by the IACUC (Institutional Animal Care and Use Committee) of Guangzhou Regenerative Medicine and Health.

Plasmid construction, mESC culture, and transfection

Four single-guide RNAs (sgRNAs) targeted to the consensus sequence of IAPEz-int were designed using the GPP sgRNA Designer and Benchling website (https://www.benchling.com/) and inserted into the vector in tandem under the control of the U6 promoter. dCas9–Kdm4b–VP64 was constructed by fusing the nuclease-deficient CRISPR-associated protein 9 (Cas9) with H3K9me3 histone demethylase Kdm4b and the transcription activator VP64 in tandem. dCas9–Kdm4b–VP64-HA was cloned into a modified PiggyBac (PB) plasmid under the control of the CAG promoter. The primers and sgRNA sequences are listed in Supplementary Tables S1 and S2.

mESCs were cultured on feeder or gelatin-coated plates under serum/leukemia inhibitory factor (LIF) conditions at 37°C with 5% CO_2 [Dulbecco's modified Eagle's medium (DMEM), 2 mM Glutamax, 15% fetal bovine serum (FBS), 0.1 mM 2-mercaptoethanol, 1× non-essential amino acids (NEAA), and LIF). mESCs were tested for mycoplasma every month. Before transfection, mESCs have to be in an early growth phase in well-formed colonies. Cultures were fed 4–12 h prior to transfection. Nucleofection was performed using the Mouse ES Cell Nucleofector® Kit (Lonza, Cat. DPH-1001)

Embryo culture, digestion, and single-cell preparation

For embryo culture, KSOM medium (potassium simplex optimized medium; Millipore) was equilibrated prior to embryo collection. M2 medium (Millipore) was used for embryo handling. Mouse embryos at pre-implantation stages are fragile and react differently to enzyme digestion. For efficient single-cell preparation, in the E1.5–E2.5 stage, embryos were treated in phosphate-buffered saline (PBS) without calcium and magnesium, aspirated repeatedly through a glass needle, and re-suspended in PBS containing 0.04% bovine serum albumin (BSA). For the embryonic day 3.0 (E3.0)–E4.5 stage, embryos were digested with a solution containing 10 U ml^{-1} papain, followed by multiple washes and resuspension in PBS.

CRISPR-mediated activation

The CRISPRa system was used in this study to activate multiple copies of IAPEz-int in mESCs and mouse embryos, respectively. (i) The mESC line with stable expression of dCas9–Kdm4b–VP64 was constructed, and then the tandem sgRNA plasmid targeting the IAPEz-int consensus sequence was transferred to the experimental group, and the sgRNA plasmid targeting luciferase was transferred to the control group. After 4 days, samples were collected for dCas9 CUT&Tag, H3K9me3 ChIP-seq, RNA-seq, and quantitative reverse transcription–PCR (RT–qPCR) experiments, respectively, to verify the H3K9me3 erase effect and IAPEz-int activation effect of the system. CUT&Tag libraries were prepared using the Hyperactive in-situ ChIP Library Prep Kit for Illumina (Vazyme). (ii) In mouse embryos, 25 ng μl^{-1} of each purified sgRNA and 10 ng μl^{-1} of dCas9-VP64-KDM4B plasmid were injected into the male pronuclei of isolated zygotes from a C57BL/6J mouse (RRID: IMSR_JAX:000664), and cultured *in vitro* until the blastocyst stage. During this period, the embryonic development of the experimental group and the non-injected group was observed regularly every day for developmental stage comparison. The RNA-seq libraries were prepared using the Smartseq2 method [29].

H3K9me3 chromatin immunoprecipitation

mESCs were trypsinized, counted, and re-suspended in 10% FBS medium containing 1% formaldehyde. After 10 min of cross-linking, the cells were quenched with 0.125 M glycine in 5 min. After termination, the cells were centrifuged at 3000 rpm at 4°C for 5 min. The precipitates were washed twice with pre-cooled PBS and centrifuged under the same conditions to remove the supernatant, and an equal proportion of S2 cells was directly combined with lysed cells for sonication using a bioruptor sonicator (Diagenode). Sonicated chromatin was centrifuged at 13 500 rpm at 4°C for 5 min, and the supernatant was diluted 10 times with IP buffer [50 mM HEPES/KOH (pH 7.5); 500 mM NaCl; 0.5% Triton X-100; 0.5% NP-40; 0.1% Doc; 1 mM EDTA]. The chromatin samples were incubated with H3K9me3/IgG antibodies at 4°C for >2 h followed by a further overnight incubation with magnetic Dynabeads Protein G; they were then washed three times with IP buffer and wash buffer [50 mM Tris–HCl (pH 7.4); 300 mM LiCl; 2 mM $MgCl_2$; 0.5% NP-40]. Treatment with high temperatures (4 h at 65°C) and enzymes (protease K and RNase A, at 50°C for 1 h) removes the cross-linked material and releases the DNA, which was recovered using a DNA miniElute kit (Qiagen). The recovered DNA was evaluated by qPCR to determine whether the binding of positive and negative sites of H3K9me3 was normal, and whether the binding of the IAPEz-int locus was decreased. Primers used in qPCR are listed in Supplementary Table S1.

Single-cell multi-omics library preparation for blastocysts

Single-cell multi-omics libraries from mouse blastocysts were prepared through embryo dissociation, nuclei isolation, and library construction. Briefly, blastocysts were first treated with acidic Tyrode's solution to remove the zona pellucida. Zona-free embryos were then mechanically disrupted using a 50 μm glass needle and transferred into 50 μl digestion droplets containing an optimized mixture of Accutase and papain (10 μg ml^{-1}). Digestion was performed at 37°C for 10–15 min with gentle pipetting, and this step was repeated three times until a single-cell suspension was obtained, as confirmed under a stereomicroscope. Digestion was terminated by transferring the cells into ice-cold M2 medium. After washing, nuclei were isolated using the 10× Genomics Nuclei Isolation Kit, with the optimal lysis time determined to be 2.5 min by trypan blue staining. Single-cell RNA sequencing (scRNA-seq) and single-cell assay for transposase-accessible chromatin sequencing (scATAC-seq) libraries were subsequently constructed using the Chromium Next GEM Single Cell Multiome ATAC + Gene Expression kit (PN-1000283, 10× Genomics). Library quality was assessed by measuring the concentration with a Qubit fluorometer and fragment size distribution with a qSep system. Only high-quality libraries were sequenced on an Illumina platform.

Single-cell microinjection and chimeric assay

Chimeric embryos were generated by single-cell microinjection into 8C embryos. Briefly, a single green fluorescent protein (GFP)-labeled mESC (or up to three cells) from either control GFP mESCs or IAPEz-activated GFP mESCs was injected into each 8C C57BL/6J recipient embryo using a micromanipulator. After injection, embryos were cultured in KSOM medium for 24 h in a humidified incubator at 37°C with 5% CO_2. Chimeric blastocysts at E3.5 were examined under a fluorescence stereomicroscope to assess the contribution and localization of GFP^+ donor cells in the ICM and trophectoderm.

Immunofluorescence staining of embryos

For immunofluorescence analysis of cultured pre-implantation stage embryos, embryos were fixed with 4% w/v paraformaldehyde (PFA) (Meilunbio, MA0192) in PBS for 30 min at room temperature. Samples were washed three times with 0.1% Triton X-100 (CAS: 9002-93-1) diluted in PBS (PBST). For permeabilization, embryos were incubated for 30 min in 0.3% PBST before samples were washed three times with 0.1% PBST. Embryos were incubated for 2 h in blocking buffer (0.1% PBST + 1% BSA + 3% donkey serum), followed by overnight incubation with primary antibodies diluted in blocking buffer (1:200). On the next day, embryos were incubated with primary antibodies for another hour at room temperature, then washed four times for 5 min in 0.1% PBST before incubation for 2 h in secondary antibodies diluted in blocking buffer. Afterward, embryos were washed three times for 15 min in 0.1% PBST with 4′,6-diamidino-2-phenylindole (DAPI). Embryos were imaged using an Olympus IXplore SpinSR confocal microscope.

Image acquisition was performed with a ×100 oil objective. For illumination, 405, 488, 561, and 647 nm lasers were used. Images were processed with Fiji software.

Stable integration of single-guide RNAs using the PiggyBac system

The mESC line used in this study is the OCT4::GFP mESC strain (mouse ESC line OG2 with GFP controlled by the Oct4 promoter), modified to stably express dCas9a. ESCs were maintained in serum/LIF medium containing high-glucose DMEM (HyClone, SH30022.01), 15% FBS (BVS500), 1× GlutaMAX (Gibco, 35050079), 1× NEAA (Gibco, 11140076), 1× sodium pyruvate (Gibco, 11 360070), 100 μM 2-mercaptoethanol (Gibco, 21 985023), and 1000 U ml^{-1} LIF (Novoprotein, C690). Cells were cultured in a humidified incubator at 37°C with 5% CO_2.

To establish stable cell lines targeting specific genomic regions, sgLuci (control) and sgIAP were introduced into the dCas9a mESC utilizing the PB transposon system. Briefly, ESCs were dissociated into single cells and co-transfected with the respective PB-sgRNA plasmid (PB-sgLuci or PB-sgIAP) and the PBase helper plasmid. Following electroporation, cells were plated onto 0.1% gelatin-coated dishes in ESC medium to recover. To isolate stable integrants, dual antibiotic selection was subsequently applied to the culture medium using 10 μg ml^{-1} blasticidin and 200 μg ml^{-1} hygromycin.

Trophoblast stem-like cells induction

The reprogramming of mESCs into TSLCs was performed following a modified protocol based on the method established by the Zhang laboratory [49]. Initially, 300 000 cells were plated into one well of the 6-well culture dish with the addition of 0.5 mM sodium butyrate (NaB) for 2 days. Following this incubation, the cells were dissociated using 0.05% trypsin. Subsequently, 50 000 cells were resuspended in 1× TSC medium and transferred to a new 6-well plate pre-coated with mitomycin C-inactivated CF-1 feeder cells. The 1× TSC medium (RPMI 1640; 11875-093, Gibco) supplemented with 1× sodium pyruvate, 20% FBS, 100 μM 2-mercaptoethanol, 25 ng ml^{-1} fibroblast growth factor 4 (FGF4; 100-31-100UG, Peprotech), 1 μg ml^{-1} heparin (H3149, Sigma), and 0.5 mM NaB (B5887, Sigma). The culture medium was refreshed every 2 days. After 4 days of induction, 40 000 cells were resuspended in 1.5× TSC medium (containing 37.5 ng ml^{-1} FGF4 and 1.5 μg ml^{-1} heparin) supplemented with 0.5 mM NaB. These cells were then plated onto a 6 cm dish containing mitomycin C-inactivated CF-1 feeders and cultured for an additional 4–6 days.

Flow cytometry analysis and cell sorting

Cells were dissociated into a single-cell suspension using 0.05% trypsin and neutralized in MEF medium (high-glucose DMEM supplemented with 10% FBS, 1× GlutaMAX, 1× sodium pyruvate, and 1× NEAA). Following centrifugation, the cell pellets were resuspended in 100 μl of Dulbecco's PBS (DPBS) containing 10% FBS and stained with an Alexa Fluor® APC-conjugated anti-CD40 antibody (1:100 dilution; Invitrogen, 17-0401-82). After a 30 min incubation on ice, the cells were washed twice with DPBS containing 2% KnockOut Serum Replacement (KSR, Gibco, 10828028). The cells were then resuspended in DPBS/2% KSR solution, filtered through a 35 μm cell strainer cap, and subjected to flow cytometric analysis and sorting on a BD FACS Aria IIU instrument. Specifically, the sorter was utilized to analyze the CD40 expression profile and to isolate the target cells, effectively depleting the CF-1 feeder cell population. The purified, feeder-free sorted cells were subsequently collected for RNA extraction and qPCR analysis. Data analysis was performed using FlowJo 10.0 software.

RNA-seq data analysis

RNA-seq reads were mapped to the mouse mm10 genome and transcriptome (GENCODE M30 annotations) using STAR (v2.7.1a) [30]. Read quantification of TEs was performed with scTE [31], configured as "-CB False -UMI False". Sample normalization and differential gene expression analysis were conducted using DESeq2 [32]. Gene Ontology (GO) analysis was carried out with clusterProfiler [33].

scRNA-seq data analysis

For analysis of C1/SMART-seq-style data, scRNA-seq data were processed using the scTE [31] C1/SMART-seq pipeline. In brief, reads were aligned to the genome with STAR [30], configured as "–winAnchorMultimapNmax 100 –outSAMmultNmax 1". The default scTE settings for C1/SMART-seq were applied to generate the molecule count matrix. Cells were filtered to exclude those with <10 000 counts or <2000 expressed genes, as well as those with mitochondrial counts >20%. Subsequent analyses followed the same workflow as the 10× data pipeline.

For analysis of 10×-style data, scRNA-seq data were processed using the scTE 10× pipeline. In brief, reads were aligned to the genome with STARsolo [30] using the parameters: "–outSAMattributes NH HI AS nM CR CY UR UY –readFilesCommand zcat –outFilterMultimapNmax 100 –winAnchorMultimapNmax 100 –out-MultimapperOrder Random –runRNGseed 777 –outSAMmultNmax 1". The default scTE settings for 10× data were employed to generate the molecule count matrix. This matrix was minimally filtered to remove cell barcodes with low counts: cells with <1000 unique molecular identifiers (UMIs) or <500 detected genes were excluded, and only the top 10 000 cells with the highest gene counts were retained. Downstream analyses were conducted using SCANPY [34]. Single-cell trajectory was analyzed by Harmony [35]. Dynamic features and enrichment analysis were performed by SCP (https://github.com/zhanghao-njmu/SCP).

Orthologous gene and transposable element analysis

Analysis of human–mouse orthologous genes was performed using the R package orthogene (https://www.bioconductor.org/packages/release/bioc/html/orthogene.html). All human genes were queried to identify their mouse orthologs using the parameters method = homologene and method = babelgene. Orthologous genes identified by either method were combined, with a gene considered conserved if detected by at least one method. Similarly, mouse genes were queried to identify their human orthologs. For TEs, phylogenetic lineage information for each TE family was retrieved from the Dfam database (https://dfam.org/home). The phylogenetic tree indicates that the last common ancestor of human and mouse TEs corresponds to the Glires–Primates divergence. Based on this, TE families were classified as conserved if they originated

before the Glires–Primates split, and non–conserved if they emerged afterward.

ChIP-seq and CUT&Tag data analysis

Cleaned reads were mapped to the mouse mm10 genome using Bowtie2 (v2.2.5) [36] with the parameters: "-p 10 –very-sensitive –end-to-end –no-unal –phred33 –no-mixed -X2000". Multimapped reads were retained for TE analysis, reporting only the best alignment for each; in cases of multiple equally optimal alignments, one was randomly selected. For unique mapping, reads with MAPQ (mapping quality) ≥ 30 were extracted from the bam file. Alignment BAM files were converted to read coverage files (bigWig format) using deepTools (v2.5.4) [37] with RPKM (reads per kilobase per million mapped reads) normalization. TE coordinates and annotations were sourced from the UCSC Genome Browser (GRCm38/mm10) RepeatMasker dataset. Heatmaps and pileup profiles were created with deepTools. Motif discovery was conducted using HOMER [38].

Promoter capture Hi-C data analysis

Raw sequencing reads were analyzed using the HiC-Pro pipeline [39]. Genomic compartments were identified at a 40 kb resolution with HOMER's runHiCpca.pl command [38], where the eigenvector of the first principal component (PC1) delineated compartments: negative values denoted compartment B, and positive values indicated compartment A. Promoter–genome interaction data were obtained from the original study [40]. BEDTools [41] was then employed to identify promoters interacting with IAPEz.

Single-cell multi-omics data analysis

Raw sequencing reads were aligned to the mouse reference genome (mm10) using Cell Ranger ARC (version 2.0.2) with default parameters (cellranger-arc count). Quality control was performed separately for the RNA and ATAC modalities. For the gene expression data, potential doublets were removed using Scrublet. Cells with > 20% mitochondrial reads or < 1000 detected genes were excluded. For the chromatin accessibility data, quality control was conducted using pycisTopic [42]. Only cells meeting the following criteria were retained: transcription start site (TSS) enrichment score > 1, FRIP (fraction of reads in peaks) > 0.4, and $\log_{10}$(unique nuclear fragments) > 3.3. Finally, only cells that passed quality control in both RNA and ATAC modalities were retained for downstream analyses. All subsequent analyses, including dimensionality reduction, clustering, and regulatory network inference, were performed using Scanpy [34] and Scenic+ [42].

Results

Single-cell RNA-seq analysis reveals species-specific TE expression in the first lineage segregation of humans and mice

To investigate the involvement of TEs in the process of the first lineage segregation in humans and mice, we first conducted scRNA-seq on mouse embryos at the E3.5 developmental stage. Utilizing Uniform Manifold Approximation and Projection (UMAP) analysis of gene expression, distinct cellular subpopulations were identified, specifically the ICM and the trophectoderm (Fig. 1A). Remarkably, UMAP based on single-cell TE expression alone effectively distinguished these cell types (Fig. 1A). Subsequent analysis on human data from a previously published scRNA-seq dataset [43] similarly demonstrated the ability to distinguish human ICM and trophectoderm cells using only TE expression information (Fig. 1A). These findings underscore the potential functional significance of TEs in the first lineage segregation process in both humans and mice.

To explore the specific TEs involved in the regulatory mechanisms of this process, we performed a differential expression analysis on ICM and trophectoderm cells. The results revealed prominent expression of hallmark genes such as *Sox2* and *Nanog* in ICM cells, and *Cdx2* and *Gata2* in trophectoderm cells (Fig. 1B; Supplementary Fig. S1A, B), validating the credibility of the dataset. Additionally, numerous TEs with cell-type-specific expression profiles in ICM and trophectoderm cells were identified. In mice, notable TEs such as MMETn-int, RLTRETN_Mm, ETnERV-int, and RLTR9E exhibited heightened expression in ICM cells, while IAPEz-int, IAPLTR1_Mm, RLTR19B, ETnERV3-int, and RLTR25B demonstrated elevated expression in trophectoderm cells (Fig. 1B, C). In humans, HERVH-int, SVA_D, L1HS, and HERVK-int displayed increased expression in the ICM, whereas LTR21A, LTR1D1, LTR35B, HERVE_a-int, and MER63C manifested elevated expression levels in trophectoderm cells (Fig. 1B, C). Specifically, in mice, 45 TE families were found to be up-regulated in ICM cells, with 5 TE families showing specific expression in trophectoderm cells (Fig. 1D). In humans, 26 TE families were highly expressed in ICM cells, while 24 TE families were elevated in trophectoderm cells (Fig. 1D). These TEs predominantly originate from ERV families (Fig. 1E).

To assess evolutionary conservation, we compared orthologous genes and TE families between mouse and human. Orthologous gene analysis (via the orthogene R package) revealed high conservation of lineage–specific genes across species (Fig. 1F). In contrast, TE conservation—determined using phylogenetic lineage data from the Dfam database—was markedly divergent (Fig. 1F). Based on the Dfam phylogeny, the last common ancestor of human and mouse TEs lies near the Glires–Primates split (Supplementary Fig. S1C). Thus, TE families that arose after this split were defined as species–specific. Remarkably, all (100%) TEs specifically enriched in mouse trophectoderm cells originated after this divergence and are absent from humans (Fig. 1F).

Evolutionary age analysis (estimated via Jukes–Cantor divergence) further indicated that lineage–biased TEs in mice are evolutionarily young, with most < 100 million years (Myr) old, whereas human counterparts tend to be much older (> 100 Myr) (Supplementary Fig. S1D). When comparing TE age with lineage bias (average $\log_2$ fold change, trophectoderm versus ICM), IAPEz–int and its associated LTRs (IAPLTR1_Mm) emerged as the youngest and most strongly trophectoderm–biased elements (Fig. 1G), implicating IAPEz as a key, species–specific regulator of trophectoderm identity.

To extend these findings, we isolated E3.5–E4.5 embryos and performed simultaneous single–cell transcriptomic (scRNA–seq) and chromatin–accessibility (scATAC–seq) profiling (Fig. 1H). After quality control, 745 high–quality cells were recovered. Transcriptome–based UMAP clustering resolved three major lineages—trophectoderm, ICM, and PrE—consistent with canonical markers (*Cdx2*, *Gata2*, *Nanog*, *Sox2*, *Gata4*, and *Sox17*) (Fig. 1I; Supplementary Fig. S1E). Notably, IAPEz–int expression

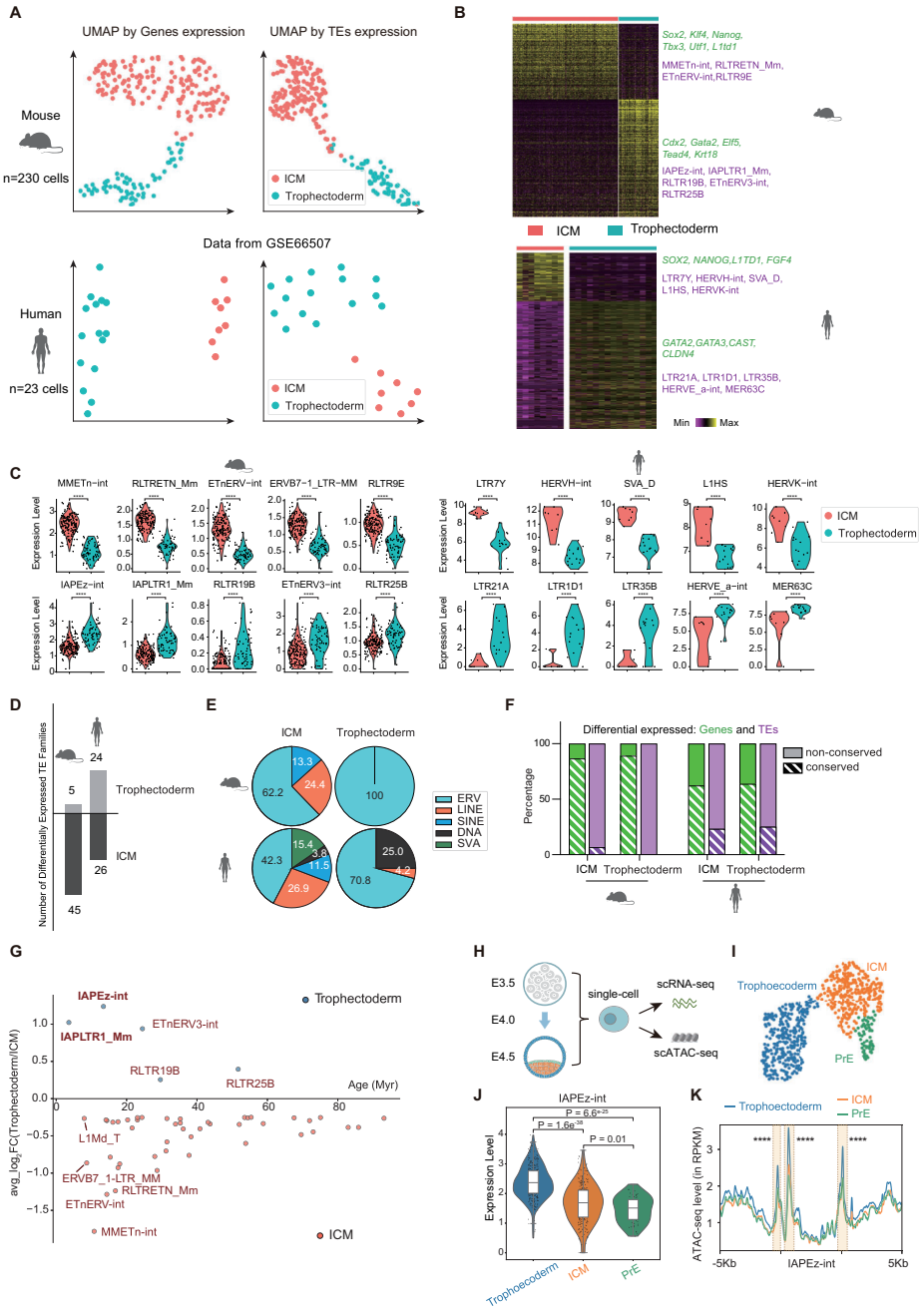


Figure 1. Differential expression profiles of TEs in the ICM and trophectoderm lineages between human and mouse embryos. (**A**) UMAP dimensionality reduction of integrated scRNA-seq data from mouse and human blastocysts, with cells colored by annotated cell-type clusters. (**B**) Heatmap of row-scaled expression levels showing differentially expressed genes (green labels) and TEs (purple labels) between the ICM and trophectoderm lineages in mouse (top) and human (bottom) embryos. Rows represent individual genes/TEs; columns represent cells clustered by lineage. (**C**) Violin plots depicting expression levels of selected differentially expressed TEs in ICM versus trophectoderm lineages for mouse (left panels) and human (right panels). Each violin represents the distribution of expression across individual cells within the respective lineage. ****P-value < 0.0001, Wilcoxon test. (**D**) Bar chart displaying the number of differentially expressed TE families between ICM and trophectoderm lineages in mouse and human embryos. (**E**) Pie charts illustrating the relative proportions of major TE classes among TEs specifically up-regulated in the ICM lineage (left) or trophectoderm lineage (right) in mouse (top) and human (bottom). (**F**) Bar plots showing the percentage of differentially expressed genes (green) and TEs (purple) in the ICM and trophectoderm that are conserved (hatched) versus non-conserved (solid fill). Percentages are calculated relative to the total number of differentially expressed features in each category and lineage. (**G**) Scatter plot of lineage-specific TEs identified from scRNA-seq of E3.5 mouse blastocysts. The *x*-axis shows estimated evolutionary age (million years, Myr) calculated using the Jukes–Cantor model, and the *y*-axis shows average $\log_2$ fold change (trophectoderm/ICM) derived from Seurat analysis. Blue dots represent trophectoderm-enriched TEs, and red dots represent ICM-enriched TEs. (**H**) Schematic overview of the experimental design. Mouse embryos at E3.5–E4.5 were collected and subjected to simultaneous scRNA–seq and scATAC–seq to assess transcriptomic and chromatin accessibility profiles within the same cells. (**I**) UMAP visualization of 745 high–quality single cells following quality filtering, revealing three major embryonic lineages: trophectoderm (blue), ICM (orange), and primitive endoderm (PrE, green). (**J**) Violin and box plots showing the expression level of IAPEz-int across the three major pre-implantation lineages (trophectoderm, ICM, and PrE) in the single-cell multi-omic dataset. Each violin represents the distribution of expression across individual cells within the respective lineage. *P*-value derived from Mann–Whitney U-tests. (**K**) Average ATAC–seq signal (RPKM–normalized) surrounding IAPEz–int elements (± 5 kb) across the three embryonic lineages. Chromatin accessibility is markedly elevated in trophectoderm compared with ICM (****$P < 10^{-4}$, Mann–Whitney U-tests).

was significantly higher in the trophectoderm than in the ICM and PrE (Fig. 1J). Chromatin–level analyses confirmed that accessibility at IAPEz–int loci and trophectoderm lineage marker genes was strongly enriched in trophectoderm cells (Fig. 1K; Supplementary Fig. S1F).

Together, these findings reveal profound species–specific divergence in TE use during the first lineage segregation, and highlight the distinctive regulatory contribution of IAPEz to mouse trophectoderm specification.

IAPEz elements are specifically repressed in the ICM lineage

To explore the dynamic expression patterns of IAPEz during pre-implantation development, we gathered scRNA-seq data from mouse embryos spanning key developmental stages: E1.5 (2C stage), E2.0 (4C stage), E2.5 (8C stage), E3.0 (morula), E3.5 (early blastocyst), and E4.5 (blastocyst), sampled at 0.5 day intervals. This comprehensive dataset enabled us to trace the developmental trajectory from the 2C stage to the establishment of the three primary cell lineages: trophectoderm, epiblast (EPI), and PrE (Fig. 2A, B). Differential expression analysis confirmed the identities of the three primary lineages, as evidenced by the robust expression of their canonical marker genes: *Gata3* (trophectoderm), *Otx2* (EPI), and *Sox17* (PrE) (Fig. 2C; Supplementary Fig. S2A).

Beyond lineage markers, our analysis revealed that certain TEs exhibit stage- or lineage-specific expression patterns during pre-implantation development. Notable examples include ORR1A1/3, predominantly active from the 2C to 4C stages; ERVB7-1-LTR_MM, ETnERV-int, MMETn-int, and RLTRETN_Mm, enriched along the ICM to EPI trajectory; LTRIS2, up-regulated in both PrE and EPI lineages; MMERVK9C_I-int, expressed from the 8C stage to EPI; and IAPEz-int, prominent from the 2C stage to the trophectoderm lineage (Fig. 2D; Supplementary Fig. S2B–E). Focusing on IAPEz, we observed a striking expression profile: it is highly expressed at the 8C stage, prior to the first lineage segregation, suggesting a potential role in early embryonic patterning (Fig. 2E). As development progresses, IAPEz–int expression begins to decline at the morula stage (Fig. 2E, F), which marks the onset of the first lineage segregation in mouse embryos. Following lineage specification, IAPEz-int undergoes specific repression within the ICM lineage, while remaining active in the trophectoderm lineage (Fig. 2E, F). This differential regulation hints at a lineage-specific functional role for IAPEz, possibly in supporting trophectoderm development or silencing within ICM-derived lineages. These findings collectively underscore the intricate stage- and lineage-dependent dynamics of TE expression, with IAPEz emerging as a key player in the molecular landscape of pre-implantation embryogenesis.

IAPEz displays a lower H3K9me3 level in TSCs compared with ESCs

To further investigate the specific silencing mechanism of IAPEz-int in the ICM lineage, we employed ESCs and TSCs as models for the ICM and trophectoderm cell states during *in vivo* development. Transcriptome analysis revealed that TSCs exhibited 4659 up-regulated and 4065 down-regulated genes or TEs compared with ESCs (Supplementary Fig. S3A). ESCs displayed higher expression of classical ICM marker genes (*Nanog* and *Pou5f1*), while TSCs showed high expression of trophectoderm marker genes (*Cdx2* and *Elf5*) (Supplementary Fig. S3B). Additionally, IAPEz-int was specifically expressed in TSCs, as well as its neighboring LTR elements, IAPLTR1_Mm and IAPLTR1a_Mm (Fig. 3A; Supplementary Fig. S3C), aligning with *in vivo* results. Subsequently, we examined the epigenetic modification landscapes of IAPEz in ESCs and TSCs. H3K9me3 is a crucial epigenetic marker for silencing ERV-type retrotransposons [44]. Upon reanalysis of H3K9me3 ChIP-seq data in ESCs and TSCs[45], we observed that H3K9me3 was significantly enrichedat IAPEz elements in ESCs compared with TSCs (Fig. 3B–E).

The above findings suggest that the specificity of IAPEz is repressed by H3K9me3 in ESCs compared with TSCs, prompting further exploration into the activation of IAPEz in ESCs. Reanalysis of published RNA-seq and ChIP-seq data from *Setdb1* conditional knockout (cKO) [46] and TRIM28-degraded ESCs [47] revealed a significant reduction in H3K9me3 enrichment at IAPEz-int elements accompanied by a marked increase in IAPEz expression (Supplementary Fig. S3D, E). Moreover, we identified 460 genes or TEs that were significantly up-regulated in both situations (Fig. 3F). GO annotation indicated that these genes are primarily associated with cell adhesion (Fig. 3G), implying a potential role in extraembryonic development. As a control, we examined the 212 genes or TEs commonly down-regulated in *Setdb1*- or *Trim28*-deficient ESCs (Supplementary Fig. S3F). Using the RAD pipeline to identify region-associated differentially expressed genes [48], we found that the commonly up-regulated differentially expressed genes were significantly enriched within 2–5 kb of IAPEz elements (Supplementary Fig. S3G). This proximity supports the hypothesis that derepression of these genes occurs via the H3K9me3–IAPEz regulatory axis. Consistent with expectations, these commonly up-regulated genes exhibited significantly higher expression in *in vitro* TSCs than in ESCs (Supplementary Fig. S3H, I), as well as higher expression in trophectoderm during *in vivo* development than in the ICM (Fig. 3H), suggesting a potential involvement of IAPEz in embryonic lineage commitment.

Regulation of trophoblast genes through IAPEz interactions in ESCs and TSCs

To further investigate genes directly regulated by IAPEz, we reanalyzed published promoter capture HiC (PCHi-C) data from ESCs and TSCs [40]. Principal component analysis (PCA) revealed a significant increase in the PC1 value of IAPEz in TSCs compared with ESCs (Fig. 4A), indicating a shift of the IAPs from the inactive B compartment to the active A compartment (Fig. 4B). Through our analysis of potential IAPEz-regulated genes in ESCs and TSCs, we identified 2519 genes directly interacting with IAPEz. Among them, 318 genes were significantly up-regulated in TSCs compared with ESCs (Fig. 4C, D). These up-regulated genes were primarily involved in developmental processes and cell adhesion (Fig. 4E), suggesting a potential role in regulating trophoblast lineage development. Notably, several key trophoblast-related genes—such as *Tfap2c, Klf14*, and *Fgfr2*—were identified among the IAPEz-interacting genes (Fig. 4F). These genes are well documented as critical regulators of trophoblast lineage establishment. Although most of these IAPEz interactions were present in both ESCs and TSCs (Supplementary Fig. S4A), elevated H3K9me3 levels at IAPEz anchor sites in ESCs appeared to act as a repressive barrier, restricting the expression of target genes associated with these interactions (Fig. 4G, H;

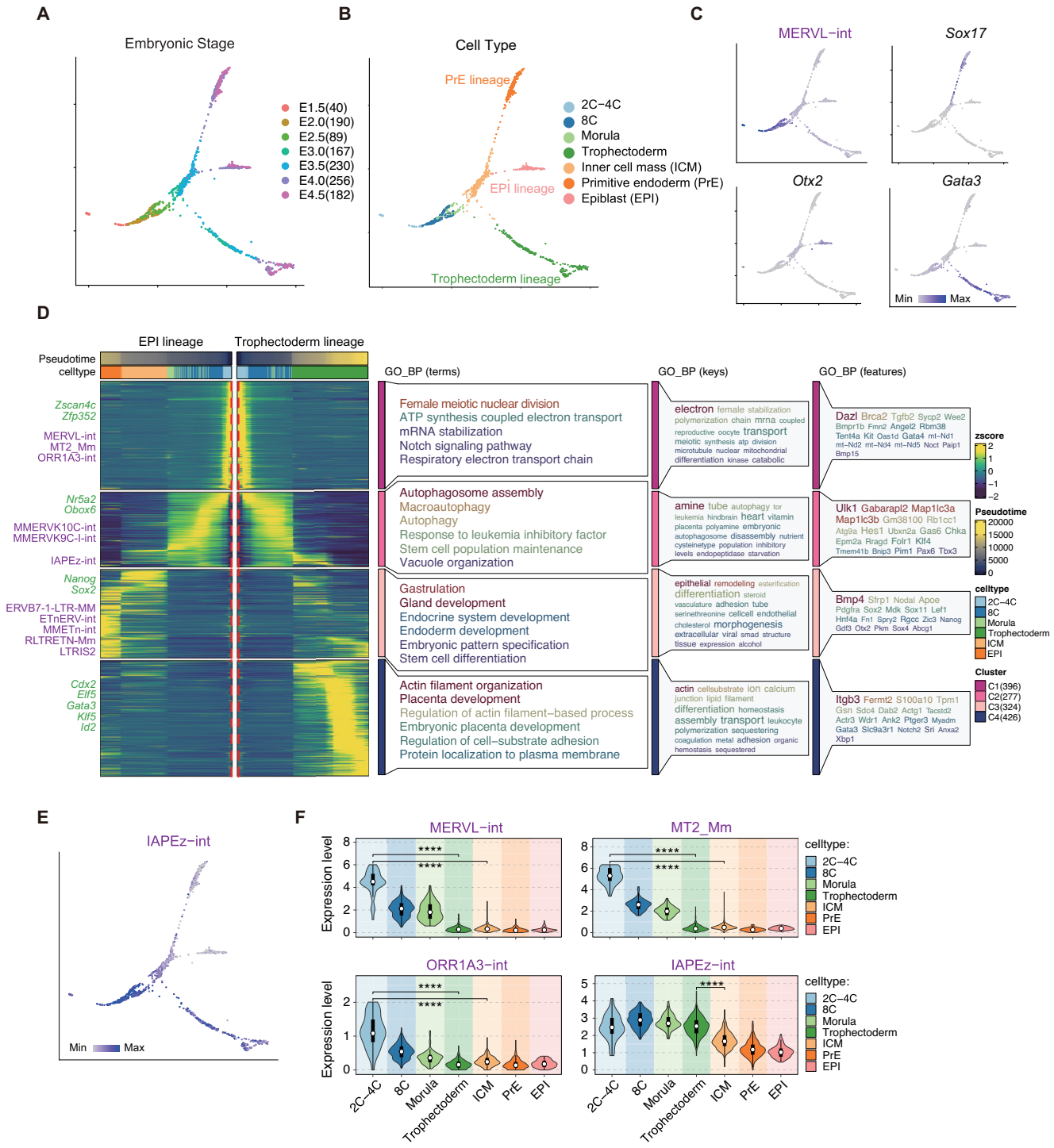


Figure 2. Stage- and lineage-specific expression of TEs in pre-implantation embryos. (**A and B**) Trajectory reconstruction during mouse embryo pre-implantation. (**A**) Cells are color-coded based on the indicated developmental time points, with numbers in parentheses indicating cell counts. (**B**) Cells color-coded by cell type. (**C**) Similar to (A), but cells are colored by the expression of the indicated TEs and genes. (**D**) Left: heatmap of row-scaled expression levels for the top 1423 differentially expressed genes and TEs along the pseudotime trajectory, grouped into four distinct clusters. Representative marker genes and TEs for each cluster are labeled on the left. Right: GO enrichment analysis for differentially expressed genes in each cluster (the top terms are shown). (**E**) Similar to (A), but cells are color-coded based on the expression of IAPEz-int. (**F**) Violin plots depicting normalized expression levels of selected dynamically expressed TEs across different cell types in mouse pre-implantation embryos. Each violin represents the distribution across cells of the indicated cell types. ****P-value < 0.0001, Wilcoxon test.

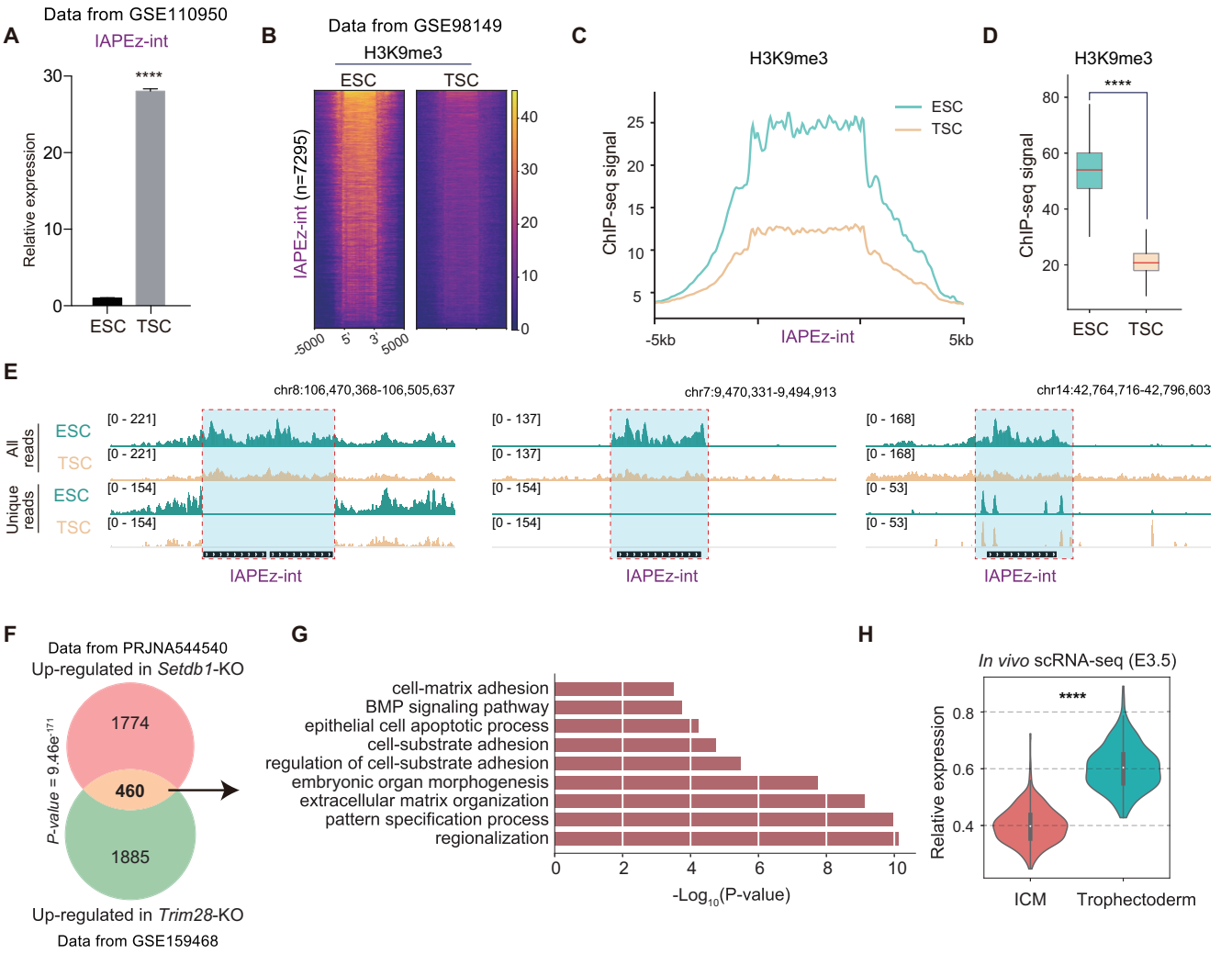


Figure 3. H3K9me3 biased repression of IAPEz-int elements in the embryonic lineage. (**A**) Bar chart depicting the relative expression level of IAPEz-int in ESCs and TSCs. Raw data from GSE110950 with $n = 2$ independent biological replicates. Error bars represent the mean ± standard deviation (SD). Statistical significance is indicated by ****, denoting *a P*-value < 0.0001 (unpaired *t*-test). (**B**) Heatmaps of normalized H3K9me3 ChIP-seq read density (RPKM) across IAPEz-int elements in ESCs (left) and TSCs (right). Elements are scaled to equal length; ±5 kb flanking regions are shown. *n* = number of IAPEz-int loci analyzed. (**C**) Average H3K9me3 ChIP-seq tag density pileup profiles (RPKM) at IAPEz-int elements in ESCs and TSCs; ±5 kb flanking regions are shown as control. (**D**) Boxplot showing the differences in H3K9me3 levels on IAPEz-int elements between ESCs and TSCs. ****$P < 0.0001$, Mann–Whitney U-test. (**E**) Selected genomic views of H3K9me3 ChIP-seq data for three individual IAPEz-int elements. All reads, using all mapped reads; unique reads, using only the unique mapped reads (MAPQ ≥ 30). (**F**) Venn diagram showing the overlap of TEs and genes repressed by Setdb1 and Trim28 in ESCs (determined from differential expression analysis in Setdb1-cKO and Trim28-dTAG datasets). The significance of the overlap was assessed using the hypergeometric distribution test (background set: all expressed genes in ESCs); the resulting *P*-value is indicated on the diagram. (**G**) GO enrichment analysis of genes commonly repressed by SETDB1 and TRIM28 in ESCs. Top enriched terms are shown with $-\log_{10}$ *P*-values. (**H**) Violin plots illustrating expression levels of SETDB1- and TRIM28-repressed genes in ICM versus trophectoderm lineages in mouse blastocysts. Each violin represents the distribution across cells in the respective lineage. ****$P < 0.0001$ (Mann–Whitney U-test)

Supplementary Fig. S4B). These findings suggest that the interplay between IAPEz-mediated interactions and increased H3K9me3 levels contributes to maintaining the silenced state of trophoblast-related genes in ESCs.

To further explore whether IAPEz plays additional roles in TSCs, motif enrichment analysis revealed that IAPEz contains numerous binding sites for transcription factors known to be critical for extraembryonic development, all of which were upregulated in TSCs (Supplementary Fig. S4C). Sequence alignment across the IAPEz family showed a high degree of conservation within regions harboring Tbx20 and Cdx2 binding motifs (Supplementary Fig. S4D). Consistently, ChIP–seq analysis confirmed that Cdx2 and Tbx20 directly bind to IAPEz–int (Supplementary Fig. S4E). Moreover, ATAC–seq analysis revealed increased chromatin accessibility at IAPEz loci in TSCs (Supplementary Fig. S4F), consistent with the *in vivo* data showing that IAPEz loci were also more accessible in the trophectoderm than in the ICM (Fig. 1K), suggesting that in trophectoderm/TSCs, IAPEz provides accessible chromatin hubs serving as binding scaffolds for transcription factors essential to extraembryonic lineage development. Furthermore, epigenetic profiling demonstrated that IAPEz regions in TSCs are enriched for active histone marks (H3K4me3 and H3K27ac) and are bound by the co–activator P300 (Supplementary Fig. S4G), indicating that IAPEz elements may function as enhancers to activate trophectoderm lineage–specific genes in TSCs. Taken together, these findings highlight the dual regulatory role of IAPEz, acting as a silencer in ESCs to repress gene expression through elevated H3K9me3 levels, and as an activator in TSCs by providing ac-

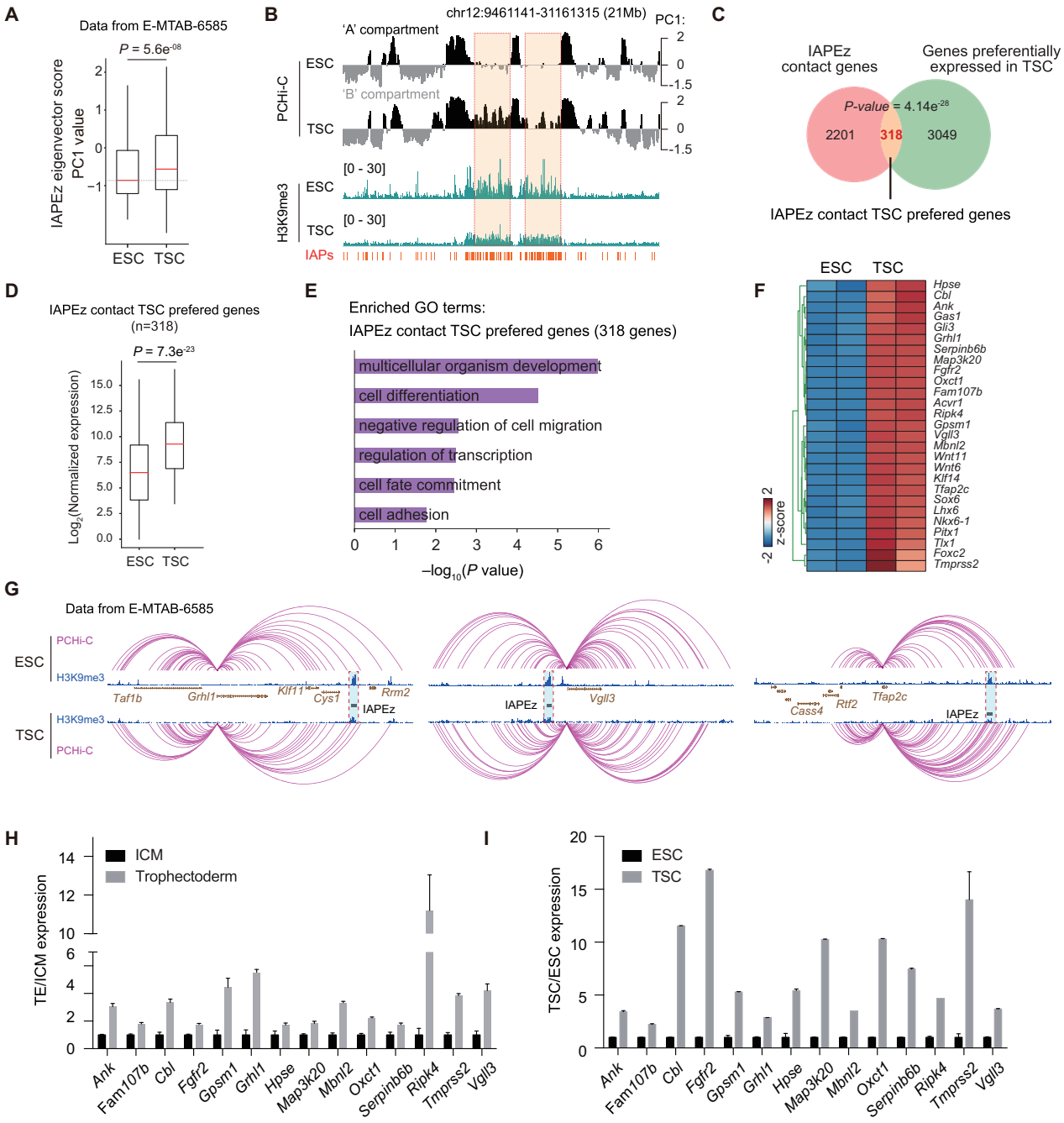


Figure 4. Genomic and transcriptional analysis of IAPEz-associated interactions in ESCs and TSCs. (**A**) PCA of PCHi-C data. The boxplot displays differences in PC1 values between ESCs and TSCs. Statistical significance was determined using the Wilcoxon test. (**B**) Compartmentalization analysis at an IAP-enriched locus on chromosome 12 using PC1 from the PCHi-C correlation matrix. PC1 values are plotted along the genomic coordinate, with positive values indicating the A compartment (euchromatin-like, black) and negative values indicating the B compartment (heterochromatin-like, gray). Red boxes highlight regions exhibiting differential compartmentalization between ESCs and TSCs. H3K9me3 ChIP-seq signal tracks from ESCs and TSCs are shown below for comparison. (**C**) Venn diagram showing the overlap between genes that physically interact with IAPEz-int elements (identified by PCHi-C) and genes preferentially expressed in TSCs compared with ESCs (adjusted $P < 0.05$, $\log_2$ fold change > 2). The significance of the overlap was assessed using the hypergeometric distribution test (background set: all expressed genes in ESCs or TSCs); the resulting *P*-value is indicated on the diagram. (**D**) Boxplot of normalized expression levels of TSC-preferred genes that interact with IAPEz-int elements in ESCs versus TSCs. Statistical significance was assessed using the Mann–Whitney U-test; the *P*-value is indicated above the plot. (**E**) GO enrichment analysis of TSC-preferred genes interacting with IAPEz-int elements. The top enriched terms are shown with $-\log_{10}$ *P*-values. (**F**) Heatmap displaying row-scaled expression patterns of selected genes interacting with IAPEz-int elements in ESCs and TSCs. (**G**) PCHi–C interaction profiles (purple arcs) illustrate chromatin contacts between IAPEz and nearby genomic regions in ESCs (top) and TSCs (bottom). Each purple arc represents a pair of genomic loci exhibiting significant promoter–capture Hi–C interactions. The blue tracks show H3K9me3 ChIP–seq signals, indicating regions enriched for H3K9me3 histone modification. Gene annotations are shown in brown below the tracks. The IAPEz element (highlighted by the dashed red box) serves as the anchor point for interactions, showing distinct contact patterns and H3K9me3 enrichment between ESCs and TSCs. (**H**) Bar charts showing the relative expression levels of selected genes interacting with IAPEz-int elements in *in vivo* ICM and trophectoderm samples. Error bars represent the mean ± SD, with $n = 4$ independent biological replicates. (**I**) Bar charts showing the relative expression levels of selected genes interacting with IAPEz-int elements in *in vitro* ESC and TSC samples. Error bars represent the mean ± SD, with $n = 2$ independent biological replicates.

cessible chromatin regions that promote transcription factor binding, thereby facilitating the regulation of extraembryonic development. This dual functionality underscores the dynamic and context-dependent nature of IAPEz in early embryonic development.

IAPEz activation disrupts the first lineage specification in mouse embryos and enhances TSC reprogramming of ESCs

To investigate the role of IAPEz insertions in the regulation of embryonic gene expression, we designed four sgRNAs that target interspersed IAPEz-int copies (Supplementary Fig. S5A). These sgRNAs collectively recognize ~43% (3163/7295) of IAPEz-int integrations in the mm10 genome, allowing zero mismatches with the IAPEz-int consensus sequence. We then attempted to activate IAPEz using a dCas9 fusion with the H3K9me3 demethylase Kdm4b and VP64 (dCas9–Kdm4b–VP64, CRISPRa) (Fig. 5A; Supplementary Fig. S5B, C). Subsequently, we employed CUT&Tag followed by high-throughput sequencing to characterize the prevalence and specificity of dCas9 targeting to individual IAPEz instances across the genome. Our findings revealed that the majority of dCas9 binding was specific to IAPEz sites (Fig. 5B; Supplementary Fig. S5D–G), and the decreased H3K9me3 level to IAPEz retrotransposons was further verified by ChIP-qPCR and ChIP-seq analysis (Supplementary Fig. S5H–K).

Additionally, RNA-seq analysis revealed robust up-regulation of IAPEz-int in the sgIAPEz group compared with non-transfected (NT) and sgLuc controls, confirming the efficacy of the CRISPRa system (Fig. 5C). In contrast, dCas9–VP64 or dCas9–Kdm4b alone failed to activate IAPEz–int expression (Supplementary Fig. S5L), demonstrating that the full dCas9–Kdm4b–VP64 fusion is required for effective activation, and H3K9me3 demethylation (Kdm4b) and forced transcriptional activation (VP64) act synergistically rather than redundantly to achieve IAPEz-int reactivation. Notably, among the 318 genes previously defined as interacting with IAPEz and significantly up-regulated in TSCs (Fig. 4C), we found that their expression was also enhanced by the CRISPRa system (Fig. 5D). This suggests that the activation of IAPEz could potentially influence the expression of genes related to extraembryonic development.

To test whether IAPEz activation promotes TSC fate *in vitro*, we induced direct reprogramming of ESCs into TSC-like cells using NaB [49] (Fig. 5E). Flow cytometry confirmed the protocol's reliability, with a marked increase in $CD40^+$ TSCs [50] upon NaB treatment (Fig. 5F). Notably, compared with NT and sgLuc controls, CRISPRa targeting IAPEz significantly enhanced conversion of TSC-like cells, as shown by a higher $CD40^+$ fraction (Fig. 5F). RT–qPCR further revealed robust Cdx2 up-regulation and Oct4 suppression upon IAPEz activation (Fig. 5G). Together, these results demonstrate that IAPEz activation actively promotes ESC to TSC transition. To further assess whether IAPEz activation alters the trophectoderm contribution *in vivo*, we injected CRISPRa-modified ESCs into 8C stage host embryos and cultured them *ex vivo* to the blastocyst stage, and the results showed that IAPEz-CRISPRa cells were successfully incorporated into the trophectoderm (Supplementary Fig. S5M, N).

We next sought to explore the role of IAPEz in embryonic development. We injected the CRISPRa system into zygotes (Fig. 5H), and observed that approximately half of the embryos microinjected with IAPEz-int gRNAs and dCas9a fusion exhibited developmental failure at 3.5 dpc (days post-coitum) at the morula stage (Fig. 5I, J). Immunofluorescence staining of arrested embryos revealed disrupted lineage segregation, with aberrant expression patterns of *Sox2* (ICM) and *Cdx2* (trophectoderm) (Supplementary Fig. S6A, B). By 4.5 dpc, control embryos had efficiently progressed to the blastocyst stage, whereas the sgIAPEz group showed a marked failure to reach this stage (Supplementary Fig. S6C, D).

To determine the onset of developmental abnormalities, we performed RNA–seq on 8C and morula–stage embryos (Fig. 5K). At the 8C stage, CRISPRa induction caused minimal transcriptional changes (Pearson correlation = 0.98 compared with no–sgRNA controls) (Fig. 5L) and did not alter IAPEz–int expression (Fig. 5K), probably due to its already high endogenous levels at this stage (Fig. 2F), thus excluding pre-8C contributions to later abnormalities. In contrast, at the morula stage, CRISPRa induction led to profound transcriptional reprogramming (Pearson correlation = 0.56) (Fig. 5L), including strong up-regulation of IAPEz–int and trophectoderm markers (*Cdx2* and *Elf5*), down-regulation of ICM genes (*Pou5f1* and *Klf2*) (Fig. 5M), and differential expression of 5613 genes overall (Fig. 5N). Notably, the 318 IAPEz–interacting and TSC–enriched genes (Fig. 4C) were significantly up-regulated upon IAPEz activation (Supplementary Fig. S6E).

Collectively, these findings demonstrate that ectopic IAPEz activation disrupts the first lineage specification at the morula stage, establishing IAPEz as a key regulator of ICM–trophectoderm segregation by promoting trophectoderm–associated gene networks.

Discussion

Understanding early embryonic development is crucial to unraveling the origins of life. ZGA and the first lineage segregation are two of the most critical events during preimplantation development. Elucidating the molecular mechanisms underlying these processes is not only fundamental to developmental biology but is also of great significance for regenerative medicine. TEs, which are the most abundant components of mammalian genomes, have been shown to play essential roles during ZGA. However, their functional contributions to the first lineage segregation remain poorly explored. Our study elucidates the regulatory functions of TEs in the first lineage segregation, focusing on the rodent-specific retrotransposon IAPEz as a key modulator in mouse embryos. By integrating single-cell transcriptomics, epigenomic profiling, and CRISPR-based perturbations, we demonstrate that IAPEz repression in the ICM via H3K9me3-dependent heterochromatin safeguards embryonic potency, while its expression in the trophectoderm supports extraembryonic programs. Ectopic IAPEz activation disrupts gene networks, leading to developmental arrest and aberrant lineage specification. These findings expand our understanding of the co-option of TEs in developmental transitions and highlight their role in species-specific adaptations.

Significant interspecies differences are evident during early embryonic development [21, 51]. For instance, humans and mice exhibit pronounced morphological distinctions during the implantation and post-implantation stages as previously discussed [21]. Transcription factors, such as Cdx2 and Oct4,

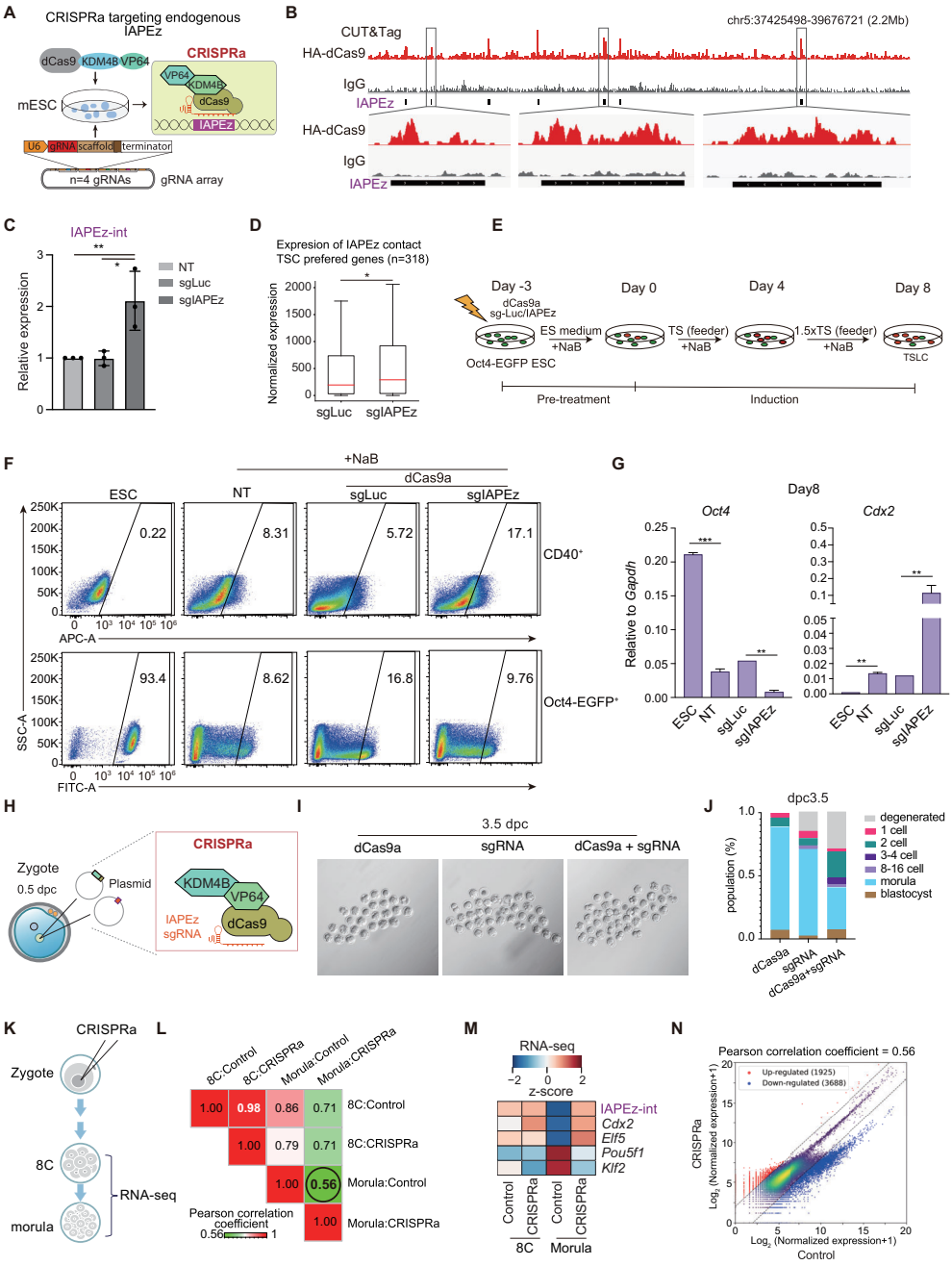


Figure 5. Targeted epigenetic manipulation of IAPEz using dCas9 fusions. (**A**) Schematic illustration of the CRISPRa strategy for endogenous activation of IAPEz-int retrotransposons in mouse ESCs. A dCas9 fusion protein (dCas9–VP64–KDM4B) is recruited to target IAPEz-int loci via an array of four sgRNAs. (**B**) Representative UCSC Genome Browser tracks (mm10 assembly) displaying CUT&Tag profiles for HA-tagged dCas9 at a representative ~2.2 Mb genomic region on chromosome 5. Tracks show enrichment (red) for HA-dCas9 (top) compared with IgG control (bottom), with zoomed views highlighting specific IAPEz-int loci (gray bars, with arrowheads indicating element orientation). (**C**) Bar charts illustrating the relative expression levels of IAPEz-int in ESCs. The no-transfection (NT) and sgRNA targeting luciferase (sgLuc) groups serve as controls, while the sgIAPEz group represents ESCs transfected with sgRNA specifically targeting IAPEz. Data from $n = 3$ independent biological replicates. Error bars indicate the mean ± SD. Statistical significance is denoted by $*$ for $P < 0.05$ and $**$ for $P < 0.01$ (unpaired t-test). (**D**) Boxplot showing expression changes of the 318 IAPEz–contacted, TSC–preferred genes identified in Fig. 4C. Statistical significance is indicated by $*$, denoting P-value < 0.05 (Wilcoxon test). (**E**) Schematic of the experimental workflow for direct reprogramming of ESCs into TSCs using NaB. (**F**) Flow cytometry analysis of the expression of the TSC marker CD40 and pluripotency marker OCT4 at day 8 post-induction. (**G**) RT–qPCR quantification of Cdx2 and Oct4 mRNA levels at day 8 post-induction. Data from $n = 2$ independent biological replicates. Error bars represent the mean ± SD. $**P < 0.01$, $***P < 0.001$, unpaired t-test. (**H**) Schematic representation of the experimental procedure for CRISPRa targeting the IAPEz sequence in a zygote. (**I**) Representative phase–contrast images of 3.5 dpc (days post-coitum) embryos following IAPEz activation. The dCas9a group received dCas9–KDM4B without sgRNAs; the sgRNA group received only sgRNAs targeting IAPEz; and the dCas9a + sgRNA group was injected with both. (**J**) Quantification of the proportion of embryos at different developmental stages following IAPEz CRISPRa manipulation. (**K**) Schematic of the CRISPRa experiment: zygotic injection of CRISPRa constructs into embryos, followed by sample collection at the 8C and morula stages for RNA-seq analysis. (**L**) Heatmap of Pearson correlation coefficients for transcriptomes across samples. (**M**) Heatmap of row-scaled expression (z-score) for selected TEs and lineage marker genes in 8C and morula-stage embryos from control and CRISPRa conditions. (**N**) Scatter plot of differential gene and TE family expression at the morula stage (CRISPRa versus control). Each dot represents one gene or TE family; up-regulated (red) and down-regulated (blue) features are highlighted.

which are key regulators of gene expression and development, are notably conserved between these species [28]. In contrast, TEs, which are major contributors to evolutionary diversity, may critically influence species-specific developmental processes. Previous research has shown that TEs can be co-opted as enhancers in mouse embryonic and trophoblast stem cells. For example, mouse-specific ERVs such as RLTR13D5 and primate-specific ERVs such as LTR10A contribute to the rapid evolution of gene regulatory networks in the trophoblast and placenta [52, 53], while RLTR9s and RLTR13s serve as functional *cis*-regulatory elements in mouse ESCs [54]. Nonetheless, their specific roles in the first lineage segregation have remained elusive, largely due to the scarcity of direct, genome-wide tools capable of manipulating their activity.

Although our study provides comprehensive insights into the functional roles and mechanisms of IAPEz in regulating the first lineage segregation of mouse embryos, several limitations remain. First, there are technical limitations in our approach. We employed CRISPRa to activate IAPEz elements and investigate their roles in embryonic development. However, our sgRNA combination targeted only ∼ 43% of IAPEz-int elements genome-wide, resulting in an activation efficiency of ∼2-fold. Developing a more efficient system for manipulating IAPEz activity could potentially uncover more pronounced developmental phenotypes. Second, although we demonstrated that IAPEz DNA can directly interact with and regulate genes involved in extraembryonic development, it has been reported that transcribed RNA from IAP elements can act *in cis* by "hijacking" RNA polymerase II in ESCs [47]. Our study does not address whether IAPEz RNA plays a role in this process, leaving room for future investigations to explore this potential regulatory mechanism. Third, our study focuses on the bulk-level functionality of IAPEz elements, and does not distinguish whether all, or only a subset, contribute to the development. Furthermore, we did not differentiate the specific regulatory contributions of internal sequences versus LTRs.

Notably, IAPEz is highly expressed from the 2C to 16C stage, suggesting a potential role even earlier during ZGA. Our gain–of–function study defines its impact on the first lineage segregation, yet future loss–of–function approaches will be essential to determine whether IAPEz also contributes to transcriptional activation during ZGA. Furthermore, our investigation focused on mouse trophectoderm–associated TEs. The manner in which species–specific TEs function in the ICM lineage and in humans remains largely uncharted. While our study establishes IAPEz as a critical regulator of ICM–trophectoderm segregation in mouse, analogous mechanisms in human remain to be explored. Future work could leverage emerging human embryo culture systems and stem–cell–based blastoid models to test whether primate–specific young LTRs or Alu elements (e.g. LTR12B or AluYa5) similarly interact with and regulate trophectoderm gene networks via epigenetic silencing in the epiblast–equivalent lineage. Single–cell multi–omics in human blastocysts, combined with CRISPR–based perturbation in blastoids, will be essential to determine whether species–specific TE families play analogous or divergent roles in primate lineage specification.

Acknowledgements

We thank the Guangzhou Branch of the Supercomputing Center of the Chinese Academy of Sciences for its support.

Author contributions: Jiangping He and Jiekai Chen (Conceptualization, Data curation, Funding acquisition, Resources, Supervision, Writing – review & editing). Jiangping He and Rujin Huang (Conceptualization, Formal analysis, Investigation, Funding acquisition, Resources, Supervision, Writing – original draft, Writing – review & editing). Rujin Huang, Zhaoya Han and Zhongqi Liufu (performed most of the experiments); Xiongzhi Quan (performed the microinjection experiments); Danfeng Li (performed the TSC induction experiments); Hongzhi Dong (performed the CUT&Tag experiments); Guangliang Hong and Ruhai Chen (performed the bioinformatics analysis); Man Zhang, Guangdun Peng and Duanqing Pei (provided resources). All authors approved the manuscript.

Supplementary data

Supplementary data are available at NAR online.

Conflict of interest

None declared.

Funding

The National Key R&D Program of China [2021YFA1102200 and 2024YFA1802300]; the National Natural Science Foundation of China [32225012, 32370856, 32570664, 32400440, and 32070794]; the Guangdong Basic and Applied Basic Research Foundation [2023B1515020111]; the Science and Technology Projects in Guangzhou [2023A04J0724]; the Science and Technology Planning Project of Guangdong Province, China [2023B1212060050 and 2023B1212120009]; the Health@InnoHK Program launched by Innovation Technology Commission of the Hong Kong SAR, China; the China Postdoctoral Science Foundation [2025M772652 and 2022M723168]; the postdoctoral fellowship(BSHF23-080, received by Rujin Huang); and the Yunnan Revitalization Talent Support Program Top team [202505AT350003 and 202405AS350022].

Data availability

The raw scRNA-seq data reported in this study have been deposited in the Genome Sequence Archive (GSA) in the National Genomics Data Center, Beijing Institute of Genomics (China National Center for Bioinformation), Chinese Academy of Sciences, under accession number CRA003171. Other sequencing data are available at the Sequence Read Archive (SRA) database under accession number PRJNA1240901 and are publicly available as of the date of publication. The scRNA-seq data for human blastocysts were from GSE66507 [43]. The H3K9me3 ChIP-seq were from GSE98149 [45], GSE184471 [55], and GSE159468 [47]. The CDX2 ChIP-seq data were from GSE42207 [52]. The H3K4me3 ChIP-seq data were from GSE73952 [56]. The ChIP-seq data for TBX20, H3K27ac, and P300, as well as the WT ATAC-seq and RNA-seq data for ESCs and TSCs were from GSE110950[57]. The Trim28-KO RNA-seq data were from GSE159468 [47]. The RNA-seq data of Setdb1-KO mouse ES cells were obtained from the BioProject accession PRJNA544540 [46]. The PCHiC data were obtained

from ArrayExpress with the accession code E-MTAB-6585 [40].

References

1. Rodriguez-Terrones D, Torres-Padilla ME. Nimble and ready to mingle: transposon outbursts of early development. *Trends Genet* 2018;34:806–20. https://doi.org/10.1016/j.tig.2018.06.006
2. Macfarlan TS, Gifford WD, Driscoll S *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 2012;487:57–63. https://doi.org/10.1038/nature11244
3. Du P, Wu J. Hallmarks of totipotent and pluripotent stem cell states. *Cell Stem Cell* 2024;31:312–33. https://doi.org/10.1016/j.stem.2024.01.009
4. Sakashita A, Kitano T, Ishizu H *et al.* Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat Genet* 2023;55:484–95. https://doi.org/10.1038/s41588-023-01324-y
5. Yang J, Cook L, Chen Z. Systematic evaluation of retroviral LTRs as cis-regulatory elements in mouse embryos. *Cell Rep* 2024;43:113775. https://doi.org/10.1016/j.celrep.2024.
6. Yang F, Huang X, Zang R *et al.* DUX–miR-344–ZMYM2-mediated activation of MERVL LTRs induces a totipotent 2C-like state. *Cell Stem Cell* 2020;26:234–50. https://doi.org/10.1016/j.stem.2020.01.004
7. Hendrickson PG, Dorais JA, Grow EJ *et al.* Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* 2017;49:925–34. https://doi.org/10.1038/ng.3844
8. Jachowicz JW, Bing X, Pontabry J *et al.* LINE-1 activation after fertilization regulates global chromatin accessibility in the early mouse embryo. *Nat Genet* 2017;49:1502–10. https://doi.org/10.1038/ng.3945
9. Percharde M, Lin CJ, Yin Y *et al.* A LINE1–nucleolin partnership regulates early development and ESC identity. *Cell* 2018;174:391–405. https://doi.org/10.1016/j.cell.2018.05.043
10. Goke J, Lu X, Chan YS *et al.* Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. *Cell Stem Cell* 2015;16:135–41. https://doi.org/10.1016/j.stem.2015.01.005
11. Yu X, Liang S, Chen M *et al.* Recapitulating early human development with 8C-like cells. *Cell Rep* 2022;39:110994. https://doi.org/10.1016/j.celrep.2022. 110994.
12. Yu H, Chen M, Hu Y *et al.* Dynamic reprogramming of H3K9me3 at hominoid-specific retrotransposons during human preimplantation development. *Cell Stem Cell* 2022;29:1031–50. https://doi.org/10.1016/j.stem.2022.06.006
13. Wang J, Xie G, Singh M *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* 2014;516:405–9. https://doi.org/10.1038/nature13804
14. Grow EJ, Flynn RA, Chavez SL *et al.* Intrinsic retroviral reactivation in human preimplantation embryos and pluripotent cells. *Nature* 2015;522:221–5. https://doi.org/10.1038/nature14308
15. Pontis J, Planet E, Offner S *et al.* Hominoid-Specific transposable elements and KZFPs facilitate human embryonic genome activation and control transcription in naive human ESCs. *Cell Stem Cell* 2019;24:724–35. https://doi.org/10.1016/j.stem.2019.03.012
16. Zhang W, Fu H, Huang Y *et al.* METTL3-dependent m6A RNA methylation regulates transposable elements and represses human naïve pluripotency through transposable element-derived enhancers. *Nucleic Acids Res* 2025;53:gkaf349. https://doi.org/10.1093/nar/gkaf349.
17. Wu F, Liufu Z, Liu Y *et al.* Species-specific rewiring of definitive endoderm developmental gene activation via endogenous retroviruses through TET1-mediated demethylation. *Cell Rep* 2022;41:111791. https://doi.org/10.1016/j.celrep.2022. 111791.
18. Wilson KD, Ameen M, Guo H *et al.* Endogenous retrovirus-derived lncRNA BANCR promotes cardiomyocyte migration in humans and non-human primates. *Dev Cell* 2020;54:694–709. https://doi.org/10.1016/j.devcel.2020.07.006
19. Zhang R, Wu M, Xiang D *et al.* A primate-specific endogenous retroviral envelope protein sequesters SFRP2 to regulate human cardiomyocyte development. *Cell Stem Cell* 2024;31:1298–314. https://doi.org/10.1016/j.stem.2024.07.006
20. Shahbazi MN, Zernicka-Goetz M. Deconstructing and reconstructing the mouse and human early embryo. *Nat Cell Biol* 2018;20:878–87. https://doi.org/10.1038/s41556-018-0144-x
21. Pereira Daoud AM, Popovic M, Dondorp WJ *et al.* Modelling human embryogenesis: embryo-like structures spark ethical and policy debate. *Hum Reprod Update* 2020;26:779–98. https://doi.org/10.1093/humupd/dmaa027
22. Petropoulos S, Edsgärd D, Reinius B *et al.* Single-Cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. *Cell* 2016;165:1012–26. https://doi.org/10.1016/j.cell.2016.03.023
23. Niwa H, Toyooka Y, Shimosato D *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* 2005;123:917–29. https://doi.org/10.1016/j.cell.2005.08.040
24. Ng RK, Dean W, Dawson C *et al.* Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nat Cell Biol* 2008;10:1280–90. https://doi.org/10.1038/ncb1786
25. Russ AP, Wattler S, Colledge WH *et al.* Eomesodermin is required for mouse trophoblast development and mesoderm formation. *Nature* 2000;404:95–9. https://doi.org/10.1038/35003601
26. Nishioka N, Inoue K, Adachi K *et al.* The Hippo signaling pathway components Lats and Yap pattern Tead4 activity to distinguish mouse trophectoderm from inner cell mass. *Dev Cell* 2009;16:398–410. https://doi.org/10.1016/j.devcel.2009.02.003
27. Yagi R, Kohn MJ, Karavanova I *et al.* Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* 2007;134:3827–36. https://doi.org/10.1242/dev.010223
28. Rossant J. Genetic control of early cell lineages in the mammalian embryo. *Annu Rev Genet* 2018;52:185–201. https://doi.org/10.1146/annurev-genet-120116-024544
29. Picelli S, Faridani OR, Björklund AK *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 2014;9:171–81. https://doi.org/10.1038/nprot.2014.006
30. Dobin A, Davis CA, Schlesinger F *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15–21. https://doi.org/10.1093/bioinformatics/bts635
31. He J, Babarinde IA, Sun L *et al.* Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nat Commun* 2021;12:1456. https://doi.org/10.1038/s41467-021-21808-x
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550. https://doi.org/10.1186/s13059-014-0550-8
33. Yu G, Wang L-G, Han Y *et al.* clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 2012;16:284–7. https://doi.org/10.1089/omi.2011.0118
34. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 2018;19:15. https://doi.org/10.1186/s13059-017-1382-0
35. Nowotschin S, Setty M, Kuo YY *et al.* The emergent landscape of the mouse gut endoderm at single-cell resolution. *Nature* 2019;569:361–7. https://doi.org/10.1038/s41586-019-1127-1
36. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9. https://doi.org/10.1038/nmeth.1923
37. Ramirez F, Ryan DP, Gruning B *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;44:W160–5. https://doi.org/10.1093/nar/gkw257
38. Heinz S, Benner C, Spann N *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory

elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89. https://doi.org/10.1016/j.molcel.2010.05.004
39. Servant N, Varoquaux N, Lajoie BR *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol* 2015;16:259. https://doi.org/10.1186/s13059-015-0831-x
40. Schoenfelder S, Mifsud B, Senner CE *et al.* Divergent wiring of repressive and active chromatin interactions between mouse embryonic and trophoblast lineages. *Nat Commun* 2018;9:4189. https://doi.org/10.1038/s41467-018-06666-4
41. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–2. https://doi.org/10.1093/bioinformatics/btq033
42. Bravo González-Blas C, De Winter S, Hulselmans G *et al.* SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks. *Nat Methods* 2023;20:1355–67. https://doi.org/10.1038/s41592-023-01938-4
43. Gerri C, McCarthy A, Alanis-Lobato G *et al.* Initiation of a conserved trophectoderm program in human, cow and mouse embryos. *Nature* 2020;587:443–7. https://doi.org/10.1038/s41586-020-2759-x
44. He J, Fu X, Zhang M *et al.* Transposable elements are regulated by context-specific patterns of chromatin marks in mouse embryonic stem cells. *Nat Commun* 2019;10:34. https://doi.org/10.1038/s41467-018-08006-y
45. Wang C, Liu X, Gao Y *et al.* Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat Cell Biol* 2018;20:620–31. https://doi.org/10.1038/s41556-018-0093-4
46. Wu K, Liu H, Wang Y *et al.* SETDB1-mediated cell fate transition between 2C-like and pluripotent states. *Cell Rep* 2020;30:25–36. https://doi.org/10.1016/j.celrep.2019.12.010
47. Asimi V, Sampath Kumar A, Niskanen H *et al.* Hijacking of transcriptional condensates by endogenous retroviruses. *Nat Genet* 2022;54:1238–47. https://doi.org/10.1038/s41588-022-01132-w
48. Guo Y, Xue Z, Yuan R *et al.* RAD: a web application to identify region associated differentially expressed genes. *Bioinformatics* 2021;37:2741–3. https://doi.org/10.1093/bioinformatics/btab075
49. Huang B, Peng X, Zhai X *et al.* Inhibition of HDAC activity directly reprograms murine embryonic stem cells to trophoblast stem cells. *Dev Cell* 2024;59:2101–17. https://doi.org/10.1016/j.devcel.2024.05.009
50. Rugg-Gunn PJ, Cox BJ, Lanner F *et al.* Cell-surface proteomics identifies lineage-specific markers of embryo-derived stem cells. *Dev Cell* 2012;22:887–901. https://doi.org/10.1016/j.devcel.2012.01.005
51. Kuijk EW, Du Puy L, Van Tol HT *et al.* Differences in early lineage segregation between mammals. *Dev Dyn* 2008;237:918–27. https://doi.org/10.1002/dvdy.21480
52. Chuong EB, Rumi MA, Soares MJ *et al.* Endogenous retroviruses function as species-specific enhancer elements in the placenta. *Nat Genet* 2013;45:325–9. https://doi.org/10.1038/ng.2553
53. Frost JM, Amante SM, Okae H *et al.* Regulation of human trophoblast gene expression by endogenous retroviruses. *Nat Struct Mol Biol* 2023;30:527–38. https://doi.org/10.1038/s41594-023-00960-6
54. Sundaram V, Choudhary MN, Pehrsson E *et al.* Functional cis-regulatory modules encoded by mouse-specific endogenous retrovirus. *Nat Commun* 2017;8:14550. https://doi.org/10.1038/ncomms14550
55. Tam PLF, Cheung MF, Chan LY *et al.* Cell-type differential targeting of SETDB1 prevents aberrant CTCF binding, chromatin looping, and cis-regulatory interactions. *Nat Commun* 2024;15:15. https://doi.org/10.1038/s41467-023-44578-0
56. Liu X, Wang C, Liu W *et al.* Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. *Nature* 2016;537:558–62. https://doi.org/10.1038/nature19362
57. Lee BK, Jang YJ, Kim M *et al.* Super-enhancer-guided mapping of regulatory networks controlling mouse trophoblast stem cells. *Nat Commun* 2019;10:4749. https://doi.org/10.1038/s41467-019-12720-6

Received: April 16, 2026. **Revised:** May 11, 2026. **Accepted:** May 19, 2026