Article

Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-implantation embryos

Graphical abstract

Highlights

- Stage-specific H3K9me3 establishment silences LTRs in human early embryos

- 8-cell-specific H3K9me3 domains may serve as enhancer-like regions at later stages

- Dux and Zfp51 are important for stage-specific H3K9me3 deposition in mouse embryos

- Both K4/K27 and K4/K9 bivalent domains are abundant in human blastocysts

Authors

Ruimin Xu, Sen Li, Qiu Wu, ..., Xiaoyu Liu, Xianghong Ou, Shaorong Gao

Correspondence

08chenfeiwang@tongji.edu.cn (C.W.),
liuxiaoyu@tongji.edu.cn (X.L.),
ouxianghong2008@163.com (X.O.),
gaoshaorong@tongji.edu.cn (S.G.)

In brief

Xu et al. profiled the genome-wide H3K9me3 distribution in human oocytes and early embryos. They found the function and mechanism of H3K9me3 modification on long terminal repeat (LTR) regulation and provide a detailed map of the sequential reprogramming of histone modifications in human pre-implantation embryo development.

CellPress

Cell Stem Cell

Article

Stage-specific H3K9me3 occupancy ensures retrotransposon silencing in human pre-implantation embryos

Ruimin Xu,[1,3,6,9] Sen Li,[2,5,9] Qiu Wu,[3,6,9] Chong Li,[3,6,9] Manxi Jiang,[2,9] Lei Guo,[2] Mo Chen,[3,6] Lingyue Yang,[3,6] Xin Dong,[4,6] Hong Wang,[3,6] Chenfei Wang,[4,6,*] Xiaoyu Liu,[1,*] Xianghong Ou,[2,5,7,8,*] and Shaorong Gao[1,3,6,10,*]

[1]Institute for Regenerative Medicine, Shanghai East Hospital, Shanghai Key Laboratory of Signaling and Disease Research, School of Life Sciences and Technology, Tongji University, Shanghai 200120, China
[2]Guangdong-Hong Kong Metabolism & Reproduction Joint Laboratory, Guangdong Second Provincial General Laboratory, Guangzhou 510317, China
[3]Clinical and Translation Research Center of Shanghai First Maternity & Infant Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China
[4]Department of Urology, Tongji Hospital, School of Life Science and Technology, Tongji University, Shanghai 200092, China
[5]Reproductive Medicine Center, Guangdong Second Provincial General Hospital, Guangzhou 510317, China
[6]Frontier Science Center for Stem Cell Research, Tongji University, Shanghai 200092, China
[7]The Second School of Clinical Medicine, Southern Medical University, Guangzhou, 510515, China
[8]Bioland Laboratory (Guangzhou Regenerative Medicine and Health Guangdong Laboratory), Guangzhou 510320, China
[9]These authors contributed equally
[10]Lead contact
*Correspondence: 08chenfeiwang@tongji.edu.cn (C.W.), liuxiaoyu@tongji.edu.cn (X.L.), ouxianghong2008@163.com (X.O.), gaoshaorong@tongji.edu.cn (S.G.)
https://doi.org/10.1016/j.stem.2022.06.001

SUMMARY

H3K9me3, as a hallmark of heterochromatin, is important for cell-fate specification. However, it remains unknown how H3K9me3 is reprogrammed during human early embryo development. Here, we profiled genome-wide H3K9me3 in human oocytes and early embryos and discovered stage-specific H3K9me3 deposition on long terminal repeats (LTRs) at the 8-cell and blastocyst stages. We found that 8-cell-specific H3K9me3 was temporarily established in enhancer-like regions, whereas blastocyst-specific H3K9me3 was more stable. DUX and multiple Krüppel-associated box domain zinc finger proteins(KRAB-ZNFs) were identified as potential factors for establishing 8C- and blastocyst-specific H3K9me3, respectively. Intriguingly, our analysis showed that stage-specific H3K9me3 allocation was attenuated by either *Dux* knockout or *Zfp51* knockdown in mouse early embryos. Moreover, we observed the existence of H3K4me3/H3K9me3 and H3K4me3/H3K27me3 bivalent chromatin domains in human blastocysts, priming for lineage differentiation. Together, our data unveil that the epigenetic switch from DNA methylation to H3K9me3 ensures the precise regulation of retrotransposons in human pre-implantation embryos.
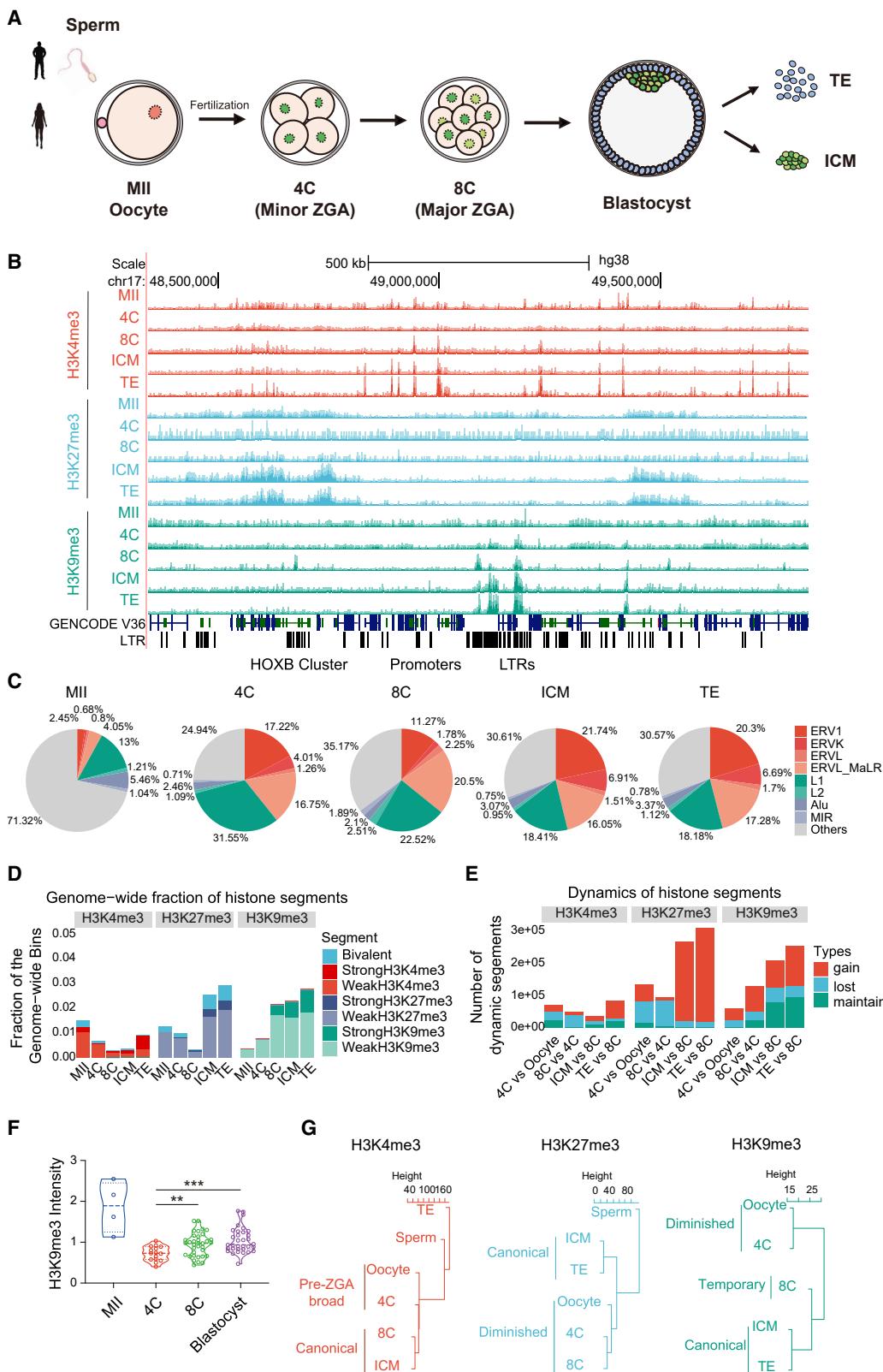
INTRODUCTION

Fertilization is regarded as one of the greatest feats of nature, beginning with the fusion of two specialized gametes, followed by major epigenetic remodeling leading to the formation of a totipotent embryo (Canovas and Ross, 2016; Rivera and Ross, 2013; Xu and Xie, 2018). A precise regulatory network must function appropriately to support a series of pivotal biological events, including oocyte activation, maternal-to-zygotic transition coordinated with zygotic genome activation (ZGA), and the first cell lineage differentiation (Amdani et al., 2015; Bonte et al., 2018; Eckersley-Maslin et al., 2018; Jukam et al., 2017; Mihajlović and Bruce, 2017; Minami et al., 2007; Schulz and Harrison, 2019; Tadros and Lipshitz, 2009; Yao et al., 2019; Yeste et al.,

2017; Zhang et al., 2018). Histone modifications are critical for regulating the spatiotemporal transcriptome during mammalian embryo development (Bannister and Kouzarides, 2011; Sadakierska-Chudy and Filip, 2015; Xu et al., 2021).

H3K9me3, as a hallmark of repressive chromatin states, is regarded as a barrier to cell-fate change (Becker et al., 2016; Burton and Torres-Padilla, 2014). Recent studies have revealed that incomplete demethylation of H3K9me2/3 severely blocks ZGA and early embryo development (Matoba et al., 2014; Wang et al., 2007). Overexpression of H3K9me3 demethylases in both mouse (Kdm4b/d) and human (KDM4A) somatic cell nuclear transferred (SCNT) embryos significantly improves cloning efficiency (Chung et al., 2015; Liu et al., 2016b; Matoba et al., 2014). Moreover, *de novo* H3K9me3 also exists in mouse early

Figure 1. Dynamics and features of histone modifications in human early embryos

(A) Schematic of sample preparation and CUT&RUN of histone modifications in human pre-implantation embryos. 4C, 4-cell; 8C, 8-cell; ZGA, zygotic genome activation; ICM, inner cell mass; and TE, trophectoderm.

embryo development. Either disrupting H3K9me3 writers (Setdb1 and Suv39h1/2) or overexpressing *Suv39h1*, enforcing precocious acquisition of constitutive H3K9me3-marked heterochromatin, impedes mouse embryo development and epigenetic reprogramming (Burton et al., 2020; Dodge et al., 2004; Peters et al., 2001). These results demonstrate that precise regulation of H3K9me3-mediated heterochromatin remodeling is pivotal for mammalian embryo development.

During mouse and human ZGA, large amounts of transposable elements, including murine/human endogenous retrovirus type L (MERVL/HERVL) and long terminal repeats (LTRs), are actively transcribed (Hendrickson et al., 2017; Leidenroth et al., 2012; Macfarlan et al., 2012). The stage-specific activation of MERVL/HERVL depends on Dux, an eutherian-specific transcription factor (TF), driving the entrance into a totipotent state (De Iaco et al., 2017; Hendrickson et al., 2017; Whiddon et al., 2017). LTRs must be properly regulated because they pose a risk to genome integrity through their potential for illicit recombination and self-duplication. Due to genome-wide DNA demethylation, the silencing of LTRs requires a switch from DNA methylation to repressive histone modifications, including H3K9me3 and H3K27me3 in mouse early embryos (Becker et al., 2016; Wang et al., 2018). Moreover, previous studies have highlighted the significance of H3K9me3 present at the promoters of lineage-specific genes, directing future cell-fate determination and lineage differentiation (Nicetto et al., 2019; Wang et al., 2018).

However, how H3K9me3 is reprogrammed in early human development remains completely elusive. In this study, we generated a high-resolution map of H3K9me3 modification in human pre-implantation embryos, utilizing cleavage under targets and release using nuclease (CUT&RUN). Our data revealed the gradual establishment of H3K9me3-dependent heterochromatin during human pre-implantation development, which displayed stage-specific allocation at the 8-cell and blastocyst stages and was mediated by diverse regulatory factors. We found both unique and conserved mechanisms of H3K9me3-mediated silencing of LTRs between human and mouse early embryos, deepening our understanding of the relationship between histone modifications and spatiotemporal gene regulation.

RESULTS

Genome-wide establishment of H3K9me3 during human pre-implantation embryo development

To investigate the dynamics and features of H3K9me3 modification during human early embryogenesis, CUT&RUN was performed for H3K9me3 in human meiosis II (MII) oocytes and at the 4-cell (4C, pre-ZGA), 8-cell (8C, peri-ZGA), and blastocyst

stages (Figures 1A and 1B). Blastocysts were micro-manipulated into the inner cell mass (ICM) and trophectoderm (TE) (Figures 1A and 1B). For conjoint analysis, we also profiled other core histone modifications, including H3K4me3 and H3K27me3 (Figure 1B), as well as the transcriptome and DNA methylome, correspondingly (Table S1). The high quality of CUT&RUN data was verified in mouse embryonic stem cells (mESCs) (Figure S1A) and human early embryos (Figures S1B–S1F).

As expected, both H3K4me3 and H3K27me3 modifications in human embryos were dominantly allocated to genic regions (Xia et al., 2019), whereas H3K9me3 largely occupied LTR and long interspersed nuclear element (LINE) regions (Figures 1B, 1C, S1D, and S1J), suggesting its critical role in regulating retrotransposons (Leung and Lorincz, 2012). Notably, unlike the drastic resetting of H3K4me3 and H3K27me3 (Xia et al., 2019), H3K9me3 was gradually established during human pre-implantation development, consistent with our immunostaining analysis, which showed high H3K9me3 signals in MII stage that may be due to the condensed chromatin state (Figures 1D–1F and S1G). Constitutive heterochromatin with strong H3K9me3 signals, which steadily extended during embryo development, was not established until the 8C stage (Figure 1D). We then performed an overall cluster of the three core histone modifications (Figures 1G and S1H–S1J). Pre-ZGA H3K4me3 showed a different pattern from peri-ZGA and post-ZGA stages, which may be explained by the noncanonical forms of H3K4me3 at the 4C stage (Xia et al., 2019; Figure 1G). Global H3K27me3 modification at the early stages of human embryos was divided into two clusters: (1) a parental-specific pattern, which gradually diminished from MII oocytes to 8C embryos, and (2) a canonical pattern, which dominated ICMs and TEs (Figure 1G). Distinct from H3K27me3, MII oocytes possessed much less H3K9me3, and 8C embryos gained temporary H3K9me3 deposition, which distinguished itself from the canonical state in blastocysts (Figure 1G).

Human early embryos exhibit conserved epigenetic switching from DNA methylation to H3K9me3 for regulating LTRs

To further investigate how H3K9me3 modification is reprogrammed during human early embryo development, we classified H3K9me3 domains into four clusters, from MII oocytes to blastocysts (ICMs and TEs) (Figure 2A). In general, maternal-specific (oocyte-specific) H3K9me3 possessed the least portion, enriched in gene-dense regions, and quickly eliminated upon fertilization; cleavage-specific (4C- and 8C-specific) H3K9me3 was allocated temporarily mainly in LTR regions, with a tiny fraction inherited to the blastocyst stage; and blastocyst-specific (ICM- and TE-specific) H3K9me3 was deposited in both genic and

(B) The UCSC Genome Browser snapshot showing the enrichment of H3K4me3, H3K27me3, and H3K9me3 signals in human MII oocytes and early embryos. Signal ranges: H3K4me3 (0–3), H3K27me3 (0–1), and H3K9me3 (0–1).

(C) The distribution of H3K9me3 domains in SINE, LINE, and LTR subfamilies during human early embryo development.
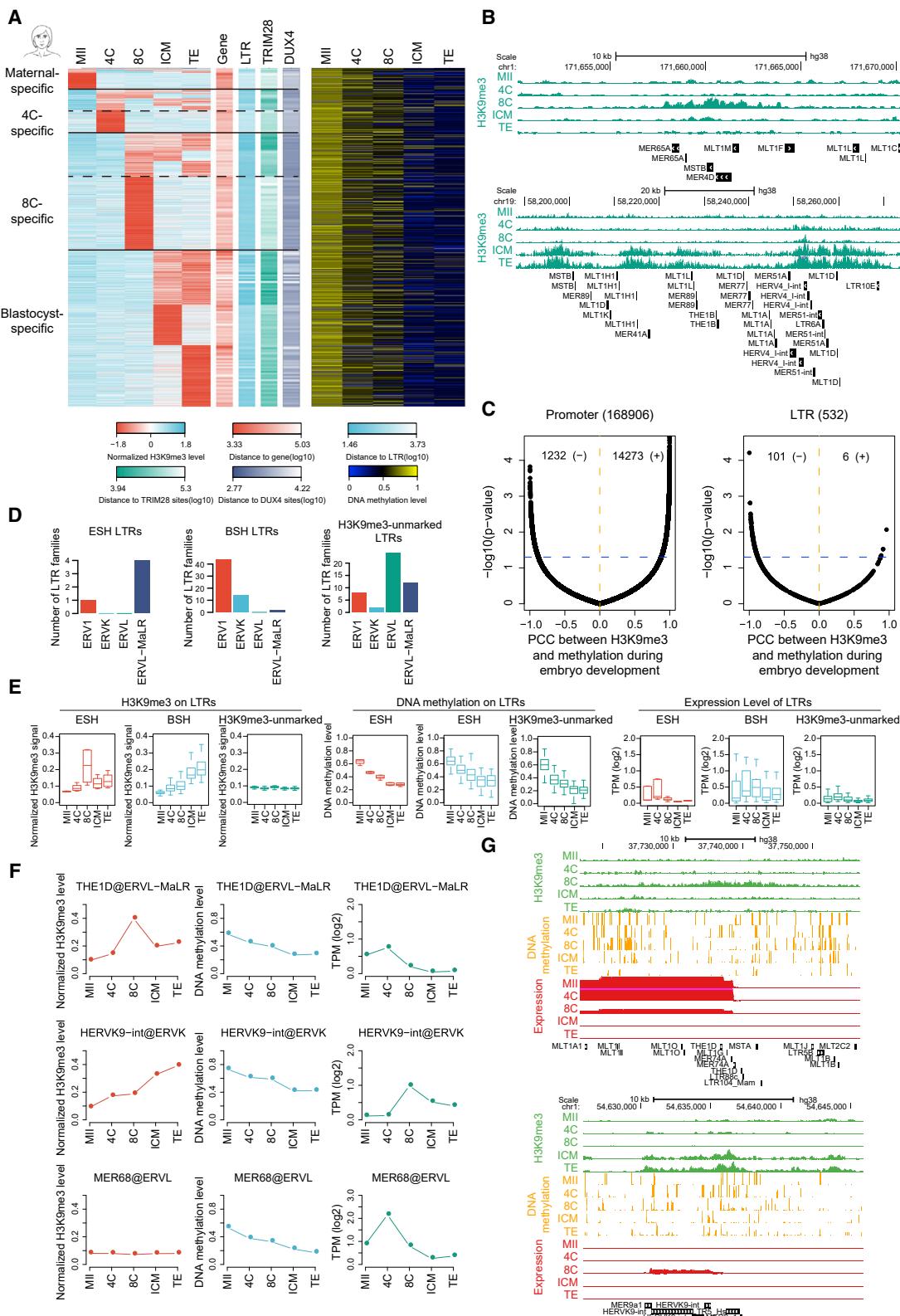
(D) The fraction of genome-wide coverage of histone domains. Coverage was evaluated based on 200-bp consecutive bins.

(E) The dynamics of gained, lost, and maintained histone domains in adjacent stages during human embryonic development.

(F) Quantification of H3K9me3 immunofluorescence signal in human oocytes, 4C, 8C, and blastocysts. Data are presented as the mean ± SEM. Each dot represents one nucleus. Oocytes (n = 4), 4C (n = 4), 8C (n = 4), and blastocysts (n = 3). **p < 0.01 and ***p < 0.001 by Student's t test for comparison.

(G) Hierarchical clustering of embryo stages based on global H3K4me3, H3K27me3, and H3K9me3 signals. Distances were measured by Pearson correlation coefficients.

See also Figure S1 and Table S1.

(legend on next page)

distal regions, configuring mature heterochromatin and priming for lineage segregation (Figures 2A and 2B).

Our previous study in mouse pre-implantation embryos discovered an epigenetic switch from DNA methylation- to H3K9me3-mediated silencing of certain repeat elements (Wang et al., 2018). Given that the genomic DNA methylome is progressively diminished throughout mammalian early embryogenesis, we wondered whether such epigenetic transition in mouse embryos is also conserved in humans (Bird, 2002; Hackett and Surani, 2013; Smith and Meissner, 2013; Zhu et al., 2018; Figure 2A). To test this hypothesis, we first investigated the Pearson's correlation coefficients (PCCs) between DNA methylation and H3K9me3 at promoters and repeat elements in human embryos (Figures 2C and S2A). Abundant protein-coding genes appeared to be co-marked by H3K9me3 and DNA methylation (14,273 positively correlated versus 1,232 negatively correlated), whereas H3K9me3 and DNA methylation on LTRs were mostly anticorrelated (101 versus 6), suggesting that the gradual substitution of DNA methylation with H3K9me3 on LTRs is similar to that in mice (Figures 2C and S2B; Table S2). Notably, such anticorrelation between DNA methylation and H3K9me3 seemed to be confined to LTRs, as very few non-LTR retroelements (0 in SVAs, 4 in short interspersed nuclear elements [SINEs], and 15 in LINEs) satisfied such transition (Figures S2A and S2B). Moreover, to interrogate whether H3K27me3 is also involved in the silencing of a distinct set of LTRs, we also performed the PCC analysis between DNA methylation and H3K27me3. Unlike mice, where H3K27me3 also contributes to the silencing of LTRs, only 7 of 532 targets were defined as H3K27me3-marked LTRs (Wang et al., 2018; Figure S2C). Furthermore, only 2 LTRs were overlapped between H3K9me3- and H3K27me3-marked LTRs (Figure S2D). These results demonstrated a conserved switching pattern from DNA methylation to H3K9me3 on LTRs during human early embryo development. H3K27me3 plays a minor role in LTR silencing during human pre-implantation embryo development, which is possibly due to the dramatic resetting of H3K27me3 that only occurs in human early embryos (Xia et al., 2019).

Stage-specific H3K9me3-dependent silencing of LTRs in human 8-cell and blastocyst

Despite the gradual establishment of H3K9me3-dependent heterochromatin during human early embryogenesis, an appreciable fraction of H3K9me3 domains were temporarily established at the 8C stage and lost in later development (Figure 2A). To investigate whether the stage-specific H3K9me3 domains have distinct impacts on LTR regulation, we designed a workflow to separate 8C-specific H3K9me3-marked (ESH) LTRs (5 types) and blastocyst-specific H3K9me3-marked (BSH) LTRs (60 types) (Figure S2E; Table S2; see STAR Methods). We also identified H3K9me3-unmarked LTRs for comparison based on their H3K9me3 and DNA methylation status (Figure S2F; Table S2). ESH LTRs mainly contained ERV1 and ERVL-MaLR families, which were transiently activated at the 4C stage and silenced by temporarily allocated H3K9me3 (e.g., THE1D) (Figures 2D–2G). By contrast, BSH LTRs included abundant ERV1 and ERVK families, which were actively transcribed at the 4C and 8C stages and silenced by gradually strengthened H3K9me3 deposition (e.g., HERVK9-int) (Figures 2D–2G). In addition, H3K9me3-unmarked LTRs were mainly enriched with ERVL families, which were briefly activated at the 4C stage and probably silenced by other modifications, such as posttranscriptional modifications (e.g., MER68) (Figures 2D–2F). In summary, these results suggested that human 8C and blastocysts initiate stage-specific *de novo* H3K9me3 deposition on LTR regions, and the establishment of H3K9me3 ensures stage-specific silencing of different LTR families.

The transient establishment of H3K9me3 in 8C embryos indicated that there might exist an alternative function of the ESH domains. To test this hypothesis, we analyzed the potential histone status in ESH and BSH domains (Figure S2G). Intriguingly, histone modifications in ESH domains were more likely to be converted to H3K4me1 and H3K27ac, hallmarks of enhancers, in the subsequent developmental stages, while BSH domains tended to be much more stabilized (Figure S2G; Calo and Wysocka, 2013). Indeed, ESH domains possessed enlarged chromatin accessibility and stronger H3K27ac signals in ICMs compared with BSH domains (Wu et al., 2018; Xia et al., 2019; Figures S2H and S2I). For further validation, both H3K4me1 and H3K27ac signals between ESH and BSH domains in 12 kinds of adult tissues were analyzed, all verifying a more accessible chromatin state in ESH ones, which is consistent with a recent study showing that H3K9me3 at the early embryonic stage is compatible with gene expression, referring to a relatively open chromatin state (Burton et al., 2020; Figure S2J). Our data revealed a noncanonical pattern of H3K9me3 domains at the human peri-ZGA stage, which may be transformed into enhancers at later embryonic stages and in somatic tissues.

Figure 2. Combinatorial regulation of LTR expression by H3K9me3 and DNA methylation in human early embryos

(A) Heatmaps showing the dynamics of H3K9me3 (left) and DNA methylation on H3K9me3 domains (right) during human early embryo development. The colors represent the log$_2$-transformed H3K9me3 signal scaled by row. For each cluster, the averaged distances to the TSS, LTR regions, hWIBR3 TRIM28-binding sites (GSE75868), and DUX4-binding sites (GSE33838) are also plotted. The colors represent the genomic distance.

(B) The UCSC browser views showing 8C-specific (upper panel) and blastocyst-specific (lower panel) H3K9me3 domains. Signal range: H3K9me3 (0–1).

(C) The association between normalized H3K9me3 and DNA methylation levels during early embryo development in promoter and LTR regions. The x axis represents Pearson's correlation coefficients (PCCs), and the y axis represents the p values of the two-sided association test (n = 5 biologically independent samples). The blue horizontal line corresponds to p = 0.05. The orange dashed line separates the positive and negative associations. The total numbers of significant positive (+) and negative (−) correlations (p < 0.05) are shown at the top of each plot.

(D) The number of LTR families in ESH, BSH, and H3K9me3-unmarked LTRs.

(E) Normalized H3K9me3, DNA methylation, and expression level of ESH, BSH, and H3K9me3-unmarked LTRs during human early embryo development.

(F) H3K9me3, DNA methylation, and expression level of THE1D, HERVK9-int, and MER68 during human early embryo development.

(G) H3K9me3 signal, absolute DNA methylation level, and normalized RNA-seq read counts on THE1D and HERVK9-int regions. Signal ranges: H3K9me3 (0–1), DNA methylation (0–1), and expression (0–400).

See also Figure S2 and Table S2.
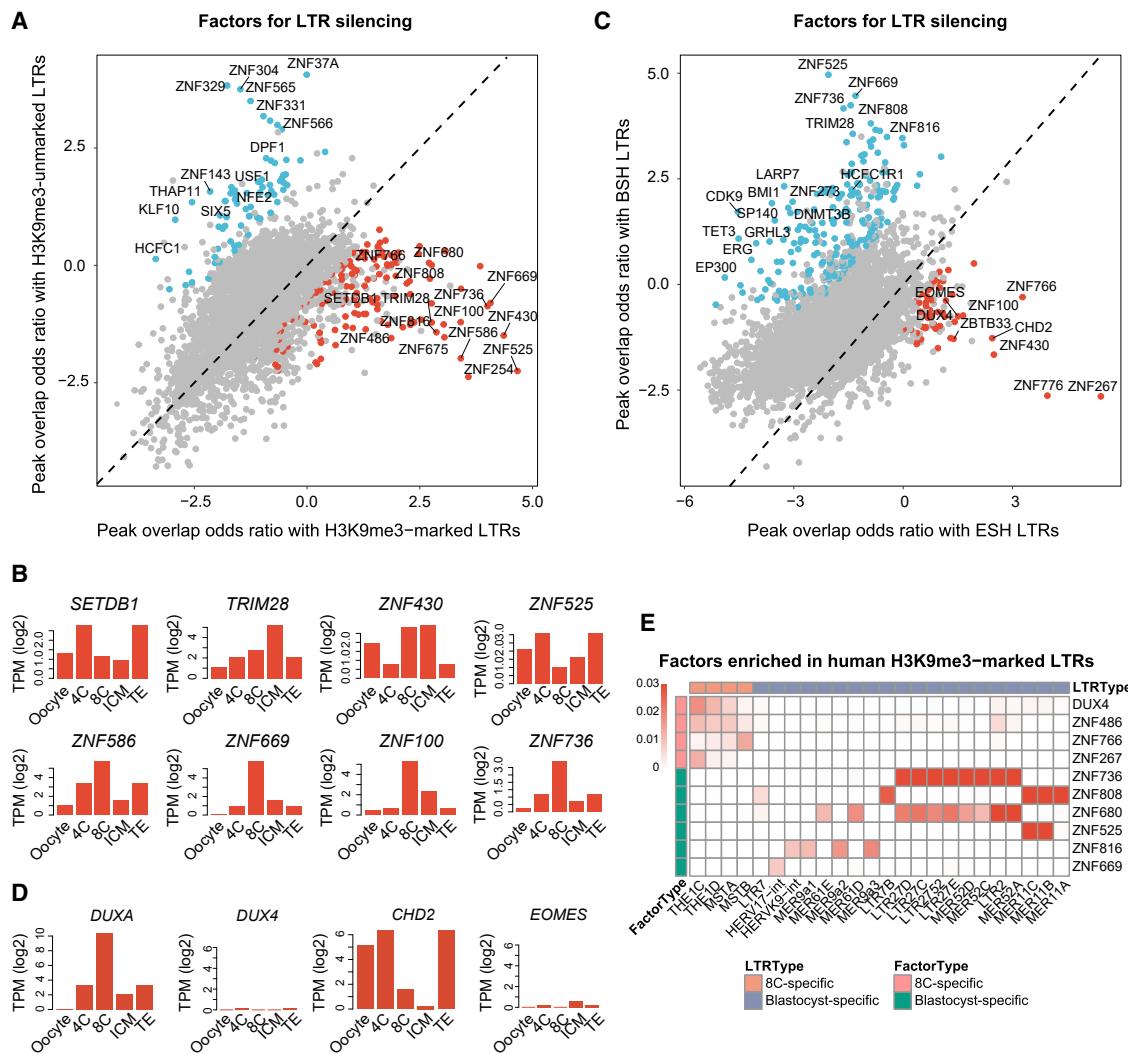
Figure 3. Putative regulators for 8C-specific and blastocyst-specific H3K9me3-marked LTRs

(A) The odds ratio of the peak overlap between transcriptional regulators and H3K9me3-marked or H3K9me3-unmarked LTR types. The x axis represents H3K9me3-marked LTR types, and the y axis represents H3K9me3-unmarked LTR types.

(B) Average gene expression levels of *SETDB1*, *TRIM28*, *ZNF430*, *ZNF525*, *ZNF586*, *ZNF669*, *ZNF100*, and *ZNF736* in human embryos.

(C) The odds ratio of the peak overlap between transcriptional regulators and ESH or BSH LTRs. The x axis represents ESH LTRs and the y axis represents BSH LTRs in human early embryos.

(D) Average gene expression levels of *DUXA*, *DUX4*, *CHD2*, and *EOMES* in human embryos.

(E) Heatmap showing the relative enrichment of DUX4 and ZNFs peaks in ESH LTRs and BSH LTRs. Enrichment was calculated using Jaccard indices between LTR regions and DUX4 and between LTR regions and ZNFs peaks.

See also Figure S3.

Distinct regulatory factors take part in establishing the stage-specific *de novo* H3K9me3 modification

We then questioned what regulatory factors might be responsible for such precise silencing of LTRs. The odds ratio of the chromatin immunoprecipitation sequencing (ChIP-seq) peaks overlap between the transcriptional regulators (TRs, including TFs and chromatin regulators) and H3K9me3-marked/unmarked LTRs were examined to predict the factors responsible for LTR silencing (Figure 3A; see STAR Methods for details). As expected, the top-scoring factors for silencing H3K9me3-marked LTRs included typical H3K9me3-related factors, such as SETDB1, TRIM28, and KRAB-ZNFs (e.g., ZNF525, ZNF430,

ZNF586, and ZNF669) (Ecco et al., 2017; Rowe et al., 2010; Schultz et al., 2002; Stoll et al., 2019; Thomas and Schneider, 2011; Figure 3A). These factors were actively transcribed in either human 8C or blastocyst (Figure 3B). By contrast, the factors predicted to silence H3K9me3-independent LTRs were associated with other KRAB-ZNFs (e.g., ZNF329, ZNF331, and ZNF565), as well as TFs involved in the negative regulation of transcription, including SIX5 and KLF10 (Cook et al., 1998; Liu et al., 2016a; Figure 3A).

To further dissect the potential regulators for the stage-specific H3K9me3 deposition on LTR regions, we examined the odds ratio of the peak overlap between TRs and ESH/BSH
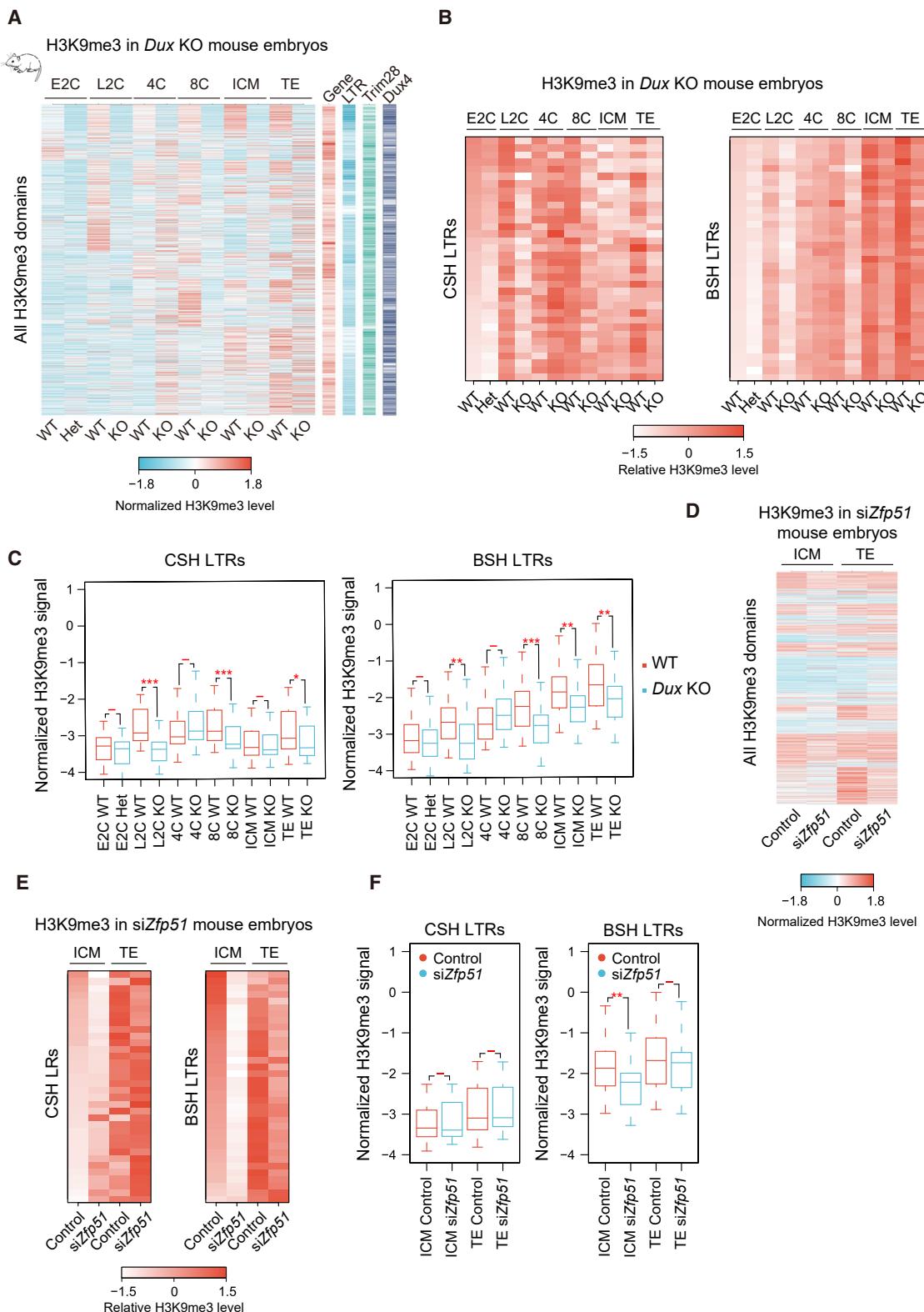
Figure 4. Loss of Dux and Zfp51 in mouse embryos hindered the establishment of stage-specific H3K9me3

(A) Heatmap showing normalized H3K9me3 signal on all H3K9me3 domains in WT and *Dux* KO embryos. The colors represent the normalized H3K9me3 level scaled by row. For each cluster, the averaged distance to the TSS, LTR regions, mESC Trim28-binding sites (GSE77440), and Dux-binding sites (GSE33838) are also plotted. The colors represent the genomic distance.

LTRs (Figure 3C). In general, typical H3K9me3-related co-factors, including TRIM28, SUMO2, and KRAB-ZNFs, were predicted to regulate BSH LTRs; activating TFs, including DUX4, EOMES, and CHD2, were predicted to be involved in H3K9me3 deposition at the 8C stage (Cui and Mager, 2018; Geng et al., 2012; Hendriks et al., 2014; Liu et al., 2020; Nacerddine et al., 2005; Woodage et al., 1997; Young et al., 2013; Figures 3C and 3D). To further investigate whether different KRAB-ZNF families are involved in stage-specific H3K9me3 establishment, we checked their binding enrichment on LTR regions (Figure 3E). Reassuringly, different ZNFs take charge of the stage-specific deposition of H3K9me3 on different LTR families, with ZNF766 and ZNF267 enriched in ESH LTRs, while ZNF736, ZNF808, and ZNF525 enriched in BSH LTRs (Figure 3E). Besides, BSH LTRs, compared with the ESH ones, showed significantly stronger overlap with the binding sites of ZNFs, such as ZNF736 and ZNF525 (Figure 3E), suggesting typical KRAB-ZNF-mediated silencing of LTRs in human blastocysts (Senft and Macfarlan, 2021). The fairly weak occupancy of ZNFs at ESH regions might explain why H3K9me3 vanished at later stages. Taken together, our analyses suggested that different KRAB-ZNF factors might be responsible for the establishment of stage-specific H3K9me3 domains.

Dux and Zfp51 play important roles in H3K9me3 establishment on stage-specific H3K9me3-targeted LTRs

Loss-of-function experiments of the predicted TRs are needed for validating their function on H3K9me3 establishment. Since manipulating the TRs is infeasible in human early embryos, we asked whether the stage-specific regulation of H3K9me3 is conserved in mice. H3K9me3 CUT&RUN data were then generated from mouse pre-implantation embryos. Comfortingly, we also identified stage-specific H3K9me3 domains during mouse embryogenesis, similar to what were found during human embryogenesis. H3K9me3 domains enriched in mouse 2C, 4C, and 8C embryos were defined as cleavage-specific H3K9me3 (CSH), and those enriched in mouse ICM and TE were defined as blastocyst-specific H3K9me3 (BSH) (Figure S3A). The LTR families and expression level were generally conserved in mice (Figures S3B–S3E). Next, we predicted factors that were responsible for establishing the H3K9me3 domains using published mouse ChIP-seq data of TRs (Figure S3F). Binding sites of Dux were predicted to be enriched in CSH domains in mice (Figures S3A and S3G). However, despite the fact that few mouse Zfp ChIP-seq data were available, the binding motifs of numerous TFs are remarkably similar between humans and mice. Thus, we used the binding motifs of human ZNFs to scan the potential binding enrichment in mouse LTRs (Figure S3H). Interestingly, many human 8C-specific ZNFs and blastocyst-specific ZNFs also showed biased enrichment on CSH and BSH LTRs in mice, respectively, indicating their orthologs in mice might play similar roles (Figure S3H). Considering this, we turned to manipulating the mouse orthologs of human ZNFs for validating their role in H3K9me3 establishment.

We first focused on DUXA (a homolog of DUX4) that is transiently expressed in human 4C and 8C embryos, which is highly consistent with the appearance of human ESH domains (Figure 3D) (Leidenroth et al., 2012). Mouse Dux (also known as Duxf3) and its human ortholog DUX4 are deemed key inducers of the 2C state and ZGA, activating retroviral elements (MERVL/HERVL family) (Chen and Zhang, 2019; De Iaco et al., 2017; Guo et al., 2019; Hendrickson et al., 2017; Whiddon et al., 2017; Figure S3G). To investigate the regulation of Dux upon the establishment of H3K9me3, we examined the changes in transcriptome and H3K9me3 modification in *Dux* knockout (KO) mouse early embryos (Guo et al., 2019; Figures S4A–S4C). Unexpectedly, *Dux* KO led to a genome-wide defect on H3K9me3 establishment, as most H3K9me3 domains in L2C, 8C, and ICM exhibited strikingly decreased H3K9me3 signal compared with wild-type (WT) embryos (Figure 4A). For LTR regions, *Dux* KO embryos grossly failed to establish H3K9me3 on CSH LTRs, with BSH LTRs also being attenuated (Figures 4B and 4C). Surprisingly, we identified both up- and down-regulated LTRs in *Dux* KO embryos, suggesting a complex mechanism of Dux on LTR regulation, probably due to its dual effect on H3K9me3 establishment and ZGA promotion (Figures S4D–S4G). Moreover, given that numerous *Zfps* are down-regulated by *Dux* KO, Dux might regulate H3K9me3 deposition on LTRs directly and/or indirectly (through regulating *Zfps*) (Figure S4H). We identified Dux-up-regulated and Dux-down-regulated *Zfps* by *Dux* KO. In general, the two groups of *Zfps* did not show any dominant difference in terms of potentially targeted LTRs that have decreased H3K9me3 and increased expression, indicating that Dux could function either in a direct or indirect way (Figure S4I). Interestingly, we noticed that for Dux-up-regulated *Zfps* (direct group), the percentage of LTRs with decreased H3K9me3 and increased expression level was higher in E2C and L2C embryos, indicating a stronger direct effect of Dux at early stages (Figure S4I). By contrast, at later stages (4C, 8C, ICM, and TE), Dux-down-regulated *Zfps* (indirect group) targeted more LTRs with decreased H3K9me3 and increased expression level, indicating the indirect effect might last longer since Dux is not expressed at the later stages (Figure S4I). Taken together, our data demonstrated that Dux in mice is an important regulator for H3K9me3-marked LTRs during early embryo development, either in a Zfps-dependent or -independent manner.

Zfp51, the ortholog of human ZNF808 (a potential regulator for BSH LTRs), was efficiently knocked down (KD) in mouse embryos by small interfering RNA (siRNA) injection (Figures S3H, S3J, S4J, and S4K). Globally, H3K9me3 modification was moderately affected in si*Zfp51* blastocysts, which was in stark

(B) Heatmap showing relative H3K9me3 signal between WT and *Dux* KO embryos on CSH and BSH LTRs.
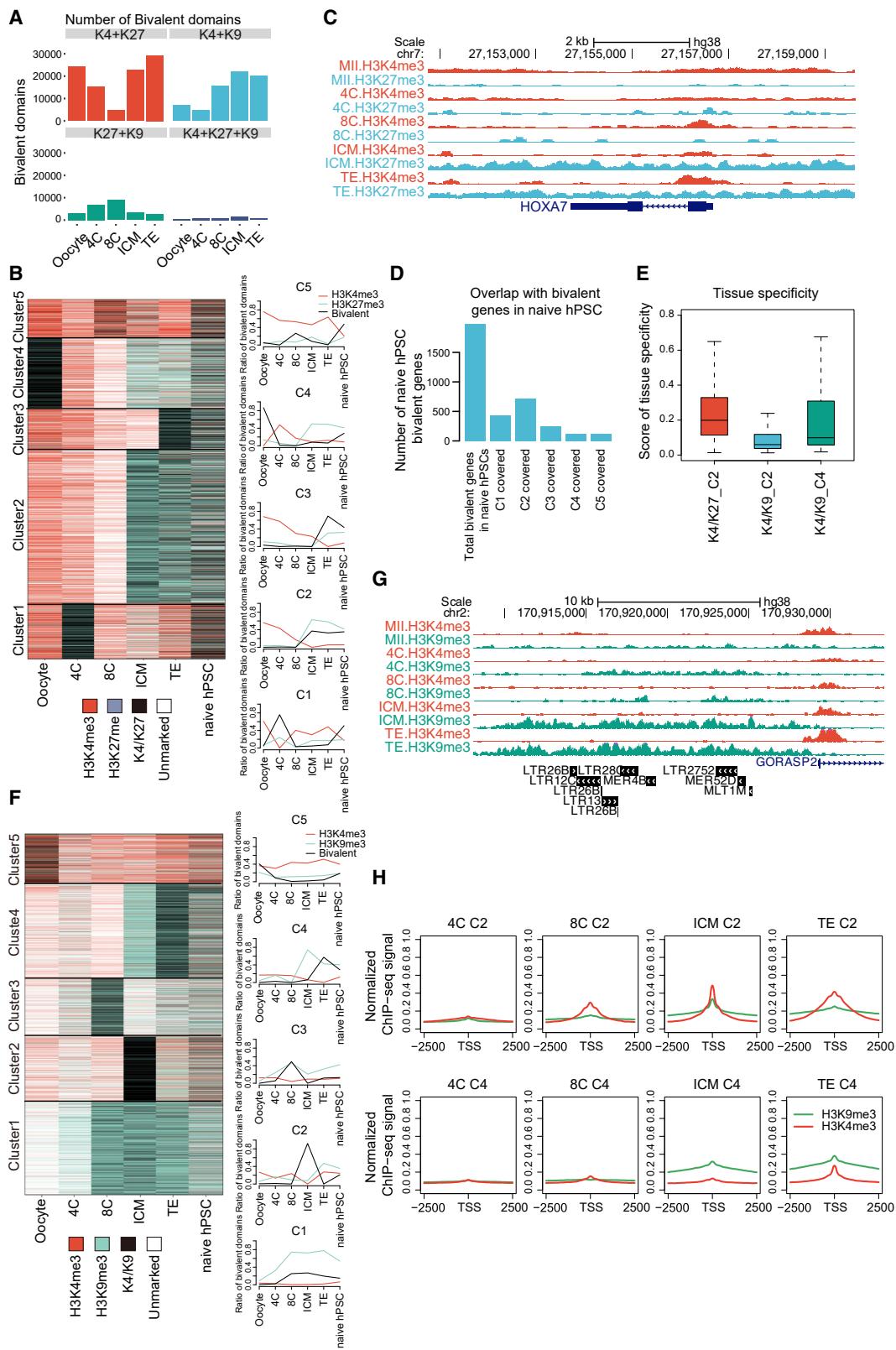
(C) Relative H3K9me3 signal between WT and *Dux* KO embryos on CSH and BSH LTRs.

(D) Heatmap showing normalized H3K9me3 signal on all the H3K9me3 domains in the ICM and TE of control and si*Zfp51* blastocysts.

(E) Heatmap showing relative H3K9me3 signal on the CSH and BSH LTRs in the ICM and TE of control and si*Zfp51* blastocysts.

(F) Normalized H3K9me3 signal on the CSH and BSH LTRs in the ICM and TE of control and si*Zfp51* blastocysts.

See also Figure S4.

contrast to that in *Dux* KO embryos (Figures 4A, 4D, and S4L). However, the impact of lacking Zfp51 showed specificity in regulating BSH LTRs, where the establishment of H3K9me3 was dramatically hindered, whereas the other LTRs were barely affected (Figures 4E and 4F). Moreover, a subset of LTR families (N = 9) was abnormally activated in absence of Zfp51-dependent H3K9me3, while others remained unchanged (Figures S4M and S4N). In addition, the dysregulated deposition of H3K9me3 and expression level of LTRs slightly impeded blastocyst formation in the lack of Zfp51 (Figures S4O and S4P).

In summary, our data suggested that Dux and Zfp51 could regulate the establishment of stage-specific H3K9me3 at LTR regions of mouse early embryos, and the precise regulation of LTR silencing could be important for normal embryo development.

Establishment of H3K4me3/H3K9me3 bivalency at lineage-specific genes for future differentiation

"Bivalent domains," which generally refers to the co-occurrence of H3K4me3 and H3K27me3 modifications, are deemed to indicate developmental genes in ESCs for future activation (Bernstein et al., 2006). Our previous study showed an infrequent and unstable pattern of bivalent H3K4me3 and H3K27me3 (K4/K27) in mouse early embryos (Liu et al., 2016c). Nevertheless, how bivalent domains are established during human pre-implantation development remains unknown. To address this, we first examined the dynamics of bivalent and trivalent domains in human MII oocytes and early embryos (Figure 5A; Table S3). Notably, we observed quite a number of K4/K27 bivalent domains in human MII oocytes, ICMs, and TEs, decreasing first and increasing later, which was highly consistent with the dynamic feature of H3K27me3 (Figures 1D, 1E, and 5A). K4/K27 bivalent domains in human embryos were classified into 5 clusters, where cluster 2 (C2) dominated and showed canonical K4/K27 co-occupied at promoter regions of developmental genes (Figures 5B, 5C, and S5A–S5C). Meanwhile, C2 K4/K27 bivalent domains showed the highest overlap with bivalent genes in naive human pluripotent stem cells (hPSCs) (Figure 5D). Moreover, all of the five clusters showed moderate overlap with the lineage-specific genes in human D5-D7 embryos, which indicated the K4/K27 bivalency could be involved in lineage differentiation (Figure S5D; Petropoulos et al., 2016a). Notably, compared with epiblast (EPI)- and primitive endoderm (PE)-specific genes, both H3K4me3 and H3K27me3 were most differently deposited between ICM and TE in D5 blastocysts on TE-specific genes,

despite most of the genes being silenced in this stage, suggesting that the determination of the TE lineage might initiate at quite an early stage (Figure S5E). Interestingly, a previous study indicates that TE maturation is initiated at the polar side, with NR2F2 marking mature TE after implantation (Meistermann et al., 2021). In our data, we observed a more distinct epigenetic bias (especially H3K4me3) in the mural TE genes than in the polar TE genes, which was also stronger in the GATA2-targeted genes (early TE factor) than in the NR2F2-targeted genes (mature TE factor) (Figures S5F and S5G). This indicated that the priming effect might be more obvious in genes that are activated at just the following stage. In brief, these results demonstrated that the K4/K27 bivalent domains are abundant in human early embryos and are involved in lineage differentiation.

In addition, another bivalent histone methylation signature, H3K4me3/H3K9me3 (K4/K9) bivalency, also plays a role in poising differentiation master regulatory genes in trophoblast stem cells (TSCs), extraembryonic endoderm stem cells, and preadipocytes (Matsumura et al., 2015; Rugg-Gunn et al., 2010). In human embryos, we also observed an appreciable number of K4/K9 bivalent domains, which were gradually established during development (Figure 5A). Notably, this trend echoed the reprogramming progress of H3K9me3 modification (Figures 1D and 1E). Similar to K4/K27 bivalency, the dynamic trend of K4/K9 bivalency also largely depends on the deposition or depletion of H3K9me3 (Figures 5F and 5G). From 5 clusters of K4/K9 bivalent domains, we discovered that C2 K4/K9 tended to be ICM specific, whereas C4 K4/K9 was prone to be more TE specific (Figures 5F, 5H, and S5H). Beyond that, C4 K4/K9 bivalency showed stronger tissue specificity than C2, yet still lower than C2 K4/K27 bivalent domains (Figure 5E). Besides, in contrast to K4/K27, all five clusters of K4/K9 bivalent genes were much less overlapped with EPI/PE/TE-specific genes defined in human D5-D7 embryos (Figure S5D; Petropoulos et al., 2016a). Moreover, K4/K9 bivalent genes function mainly in various metabolic processes and were barely activated during early embryo development (Figures S5I–S5K). It is worth mentioning that, given the distribution feature of H3K9me3, nearly 30% of LTRs were co-occupied by H3K4me3 and H3K9me3 modifications, which is dramatically higher than that of K4/K27-marked LTRs (Figure S5C). In conclusion, our data unveiled that K4/K9 bivalent domains are progressively deposited during human pre-implantation development, probably acting to regulate lineage specification, similar to K4/K27 bivalency.

Figure 5. Establishment of bivalent domains in human early embryos

(A) The number of different bivalent and trivalent domains in human MII oocytes and early embryos.

(B) Heatmaps showing the dynamics of K4/K27 bivalent domains in human early embryos and naive hPSC (left panel). The ratios of bivalent domains are also shown (right panel). The red line indicates the fraction of domains that is marked only by H3K4me3, the blue line represents the fraction of domains that is marked by H3K27me3, and the black line indicates the fraction of domains in a bivalent state.

(C) The genome browser view of K4/K27 bivalent domains. Signal ranges: H3K4me3 (0–3) and H3K27me3 (0–1).

(D) The overlapped gene number between naive hPSC bivalent genes and genes in each K4/K27 cluster in (B).

(E) Tissue specificity scores for C2 K4/K27, C2 K4/K9, and C4 K4/K9.

(F) Heatmaps showing the dynamics of K4/K9 bivalent domains in human early embryos and naive hPSC (left panel). The ratios of bivalent domains are also shown (right panel). The red line indicates the fraction of domains that is marked only by H3K4me3, the green line represents the fraction of domains that is marked by H3K9me3, and the black line indicates the fraction of domains in a bivalent state.

(G) The genome browser view of K4/K9 bivalent domains. Signal ranges: H3K4me3 (0–3) and H3K9me3 (0–1).

(H) Normalized H3K4me3 and H3K9me3 signals of C2 and C4 in (F) during human early embryo development.

See also Figure S5 and Table S3.

CellPress

DISCUSSION

H3K9me3 modification is generally regarded as a hallmark of heterochromatin regions and plays an important role in regulating the expression of both protein-coding genes and transposable elements. H3K9me2/3 establishment is pivotal for mouse embryo development since the lack of G9a/GLP (catalyzing H3K9me1/2) and Setdb1/Suv39h1/Suv39h2 (catalyzing H3K9me2/3) causes various degrees of embryonic lethality (Becker et al., 2016). Our previous research revealed the dynamic pattern of H3K9me3-mediated heterochromatin during mouse early embryo development and post-implantation embryonic differentiation. However, deciphering how H3K9me3 is reprogrammed and reorganized during human early embryo development and initial lineage segregation remains a major challenge, limited by the scarcity of materials. In this study, we utilized the CUT&RUN method to delineate a high-resolution reprogramming map of H3K9me3-mediated heterochromatin in human oocytes and early embryos. Similar to mouse embryos, *de novo* H3K9me3 in human embryos is progressively established and strengthened, enriched in a large number of LTR families. In mice, we previously found that LTRs could also be silenced dependent on H3K27me3, but only 7 of the LTR families are H3K27me3-marked in human embryos. Coincidently, H3K27me3 is massively removed at the peri-ZGA 8C stage, and the core components of PRC2 will not be activated until post-ZGA (Schwartz and Pirrotta, 2007; Xia et al., 2019). This might explain why H3K27me3 is not competent in mediating the silencing of repeat elements and, in other words, emphasizes the importance of H3K9me3 establishment for ensuring precise regulation of LTR expression in human embryos.

During ZGA, many of the MERVL elements are actively transcribed. The LTRs of MERVLs can serve as alternative promoters of protein-coding genes (Macfarlan et al., 2012). Mouse Dux and human DUX4 can activate ERVLs and induce mammalian ZGA (Hendrickson et al., 2017; Leidenroth et al., 2012). However, our data indicated that Dux may also contribute to the establishment of 8C-specific H3K9me3, which might repress the expression of ERVLs. $Dux^{-/-}$ mouse embryos showed significantly reduced establishment of CSH, whereas the expression of CSH LTRs was moderately affected. We noticed that the blastocyst-specific H3K9me3 was also reduced by *Dux* KO, and the *Zfps* that are important for H3K9me3 establishment were down-regulated in *Dux* KO embryos. These results collectively suggest that *Dux* may affect the H3K9me3 establishment in an indirect way by affecting the expression of Zfps, which has a global effect. Another explanation is that the H3K9me3 establishment and LTR silencing are controlled by a negative feedback loop for which the loss of upstream activating signal from *Dux* leads to a failure in establishing the repressive signals. The mechanism of Dux playing such bilateral functions requires further investigation. On the other hand, the function of Dux in mice and DUX4 in humans is not exactly the same, for the *Dux* KO mice are viable but no DUX4-silenced human embryo can develop to the morula stage, which indicates a more critical function of DUX4 in human early embryos (Guo et al., 2019; Liu et al., 2022). Further *in vitro* experiments will undoubtedly help us decipher the mechanisms of H3K9me3 dynamics regulated by DUX4 during human pre-implantation development, based on recently reported human blas-toids or 8-cell-like cells from human naive PSCs (Mazid et al., 2022; Taubenschmid-Stowers et al., 2022; Yu et al., 2021). Intriguingly, we observed resilient H3K9me3 signal in 4C Dux-null embryos, which might be explained by postponed rather than ceased establishment of Zfp-dependent H3K9me3, as the expression of numerous Zfps was not totally ruled out (Figure 4A).

KRAB-associated protein 1 (KAP1, encoded by TRIM28) can induce heterochromatin by recruiting SETDB1, and the KRAB-ZFP-KAP1 complex can further promote DNA methylation (Schultz et al., 2002; Stoll et al., 2019). TRIM28 and various KRAB-ZNFs are potential factors for depositing blastocyst-specific H3K9me3, indicating "constitutive" H3K9me3-dependent heterochromatin. By contrast, 8C-specific H3K9me3 probably forms "facultative" heterochromatin that is compatible with gene expression. Interestingly, 8C-specific H3K9me3 will be converted to H3K4me1 and H3K27ac in somatic tissues, becoming putative enhancers. The rather weak connection between KRAB-ZNFs and 8C-specific H3K9me3 provides a more plastic chromatin state. Further analysis is needed to unveil what kinds of ESH LTRs in ERV1, ERVK, and ERVL-MaLR conform to this transition of epigenetic modification. It is worth mentioning that recent research shows that several enhancer-like regions, located within ERVL-MaLR repeat elements, are activated by DUX4 expression in human early embryos (Vuoristo et al., 2022). Also, a previous study has shown that 80% of putative TSC enhancers are derived from ERVs in mice, and ERV-derived putative enhancers in the human placenta are more abundant than those in other tissues (Chuong et al., 2013). ERV-derived enhancers can bind Cdx2, Eomes, and Elf5 to define the regulatory network of TSCs (Chuong et al., 2013; Sun et al., 2020). However, the mechanism of the epigenetic transition from repressive H3K9me3 marks to active enhancer marks remains unknown. The temporarily deposited H3K9me3 may be compensated by unknown epigenetic marks or could recruit active TFs to promote the transition, which needs further investigation. For the first time, our study reveals that the selection of LTR-derived enhancers silenced by H3K9me3-mediated facultative heterochromatin may begin as early as the zygotic stage and is conserved in humans and mice.

More than 400 KRAB-ZFPs exist in mice and humans, and an "arms race" between retrotransposons and KRAB-ZFPs brings about huge amounts of species-specific Zfps (ZNFs in humans) (Bruno et al., 2019; Cosby et al., 2019; Ecco et al., 2017; Imbeault et al., 2017). In human early embryos, we identified two types of ZNFs that are separated according to their connection with H3K9me3 when regulating LTRs. For KRAB-ZNFs binding H3K9me3-marked LTRs, their preference for families of LTRs varies with the embryonic stage. For example, ZNF766 could bind THE1D, MSTA, and MSTB at the 8C stage, whereas ZNF736 and ZNF525 only exhibited an affinity for LTRs in blastocysts (Figure 3E). These results suggested an intricate regulatory network of KRAB-ZNFs for establishing stage-specific H3K9me3 domains in human pre-implantation embryos. Since imprinting control regions (ICRs), resistant to genome-wide DNA demethylation after fertilization, usually overlap with retrotransposons, these KRAB-ZNFs may also take part in maintaining genomic imprinting (Kaneko-Ishino and Ishino, 2010; Kobayashi et al., 2012). Further studies are required to reveal the distinct functions of the two types of ZNFs in human embryos.

Silencing of retroviral elements is a multi-layered genetic and epigenetic mechanism in PSCs, which includes localization of critical TFs to LTR regions, removal of activating histone marks, and establishment of repressive epigenetic marks (Yang et al., 2015). H3K9me3 is identified as a key repressive epigenetic mark on LTR regions in both mouse and human early embryos (Wang et al., 2018). In this study, we found that the KO/KD of *Dux* (the ortholog of human DUX4) and *Zfp51* (the ortholog of human ZNF808) disturbed the deposition of H3K9me3, but only a fraction of LTRs were activated. Our observation is generally consistent with our previous results in *Chaf1a, Sumo2*, *Setdb1*, and *Trim28* KD mouse embryos in which only *Chaf1a* KD led to reduced H3K9me3 level, re-activated LTRs, and severe embryonic lethality (Wang et al., 2018). Other KDs resulted only in significant H3K9me3 reduction and had a moderate effect on LTR re-activation and embryo development. This could be explained from two aspects. Firstly, *Sumo2, Setdb1, Trim28, and Zfp51* might only reduce the establishment of H3K9me3 on LTRs. For some LTRs, the residual H3K9me3 level on the LTRs is enough to recruit alternative silencing machinery to ensure the silence status of LTRs. Secondly, the regulation of KRAB-ZNFs to different LTR families is highly specific, and *Zfp51* is only responsible for regulating a subset of LTRs at the blastocyst stage, the mild influence of *Zfp51* could be anticipated. Last, given the post-implantation lethality of *Trim28, Sumo2*, and *Setdb1* KO mice, how the aberrant Zfp51-dependent H3K9me3 influences embryo development after implantation needs further investigation (Chen and Zhang, 2019; Dodge et al., 2004; Guo et al., 2019; Messerschmidt et al., 2012; Wang et al., 2014).

In summary, our study unveils both unique and conserved reprogramming of H3K9me3 during human ZGA and first lineage segregation and unlocks the black box of heterochromatin reorganization in human pre-implantation development.

Limitations of the study

This study indicates a spatiotemporal regulation of KRAB-ZNFs on repressing retrotransposons during human early embryogenesis through H3K9me3-mediated heterochromatin. The loss-of-function of numerous ZNFs is realizable in human stem cells and somatic tissues but is an enormous challenge in human early embryos (Senft and Macfarlan, 2021). In addition, since human KRAB-ZNFs largely evolved from mice, we could not completely rely on the regulatory mechanisms in mice. Moreover, to satisfy the lowest input number of 50 cells per sample for CUT&RUN, MII oocytes and sperm were gathered from different donors, which leads to complicated genetic backgrounds. Therefore, this set a barrier for us in dissecting the parental allele-specific reprogramming of H3K9me3 and genomic imprinting. Given the significantly higher heterogeneity in humans than in mice, single-cell-based ChIP-seq technology will inevitably expand our understanding of epigenetic reprogramming in early human development.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE

- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Ethics statement
- METHOD DETAILS
  - Human oocytes and early embryos collection
  - Collection of mouse pre-implantation embryos
  - Mouse *Dux* KO embryos collection
  - Knockdown of and *Zfp51* in mouse embryos
  - Reverse transcription and quantitative RT-PCR analysis
  - Immunofluorescent staining
  - Sample harvest for CUT&RUN, RNA-seq and BS-seq
  - Cell culture
  - CUT&RUN
  - Smart-Seq2
  - WGBS
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - CUT&RUN, RNA-seq and BS-seq data processing, quality control and normalization
  - CUT&RUN peak identification and genome segmentation analysis
  - Genomics enrichment of histone modification domains and CpG density calculation
  - Hierarchical clustering of H3K4me3, H3K27me3 and H3K9me3 data
  - Differential gene expression analysis and definition of minor ZGA and major ZGA genes
  - Gene ontology analysis
  - Clustering analysis of stage-specific H3K9me3 domains
  - Expression, DNA methylation and histone modification level quantification of repeats elements
  - Association analysis of histone modification and methylation level on repeats
  - Enrichment of transcription factor binding sites and histone modifications on LTRs and H3K9me3 domains
  - Analyses of the future chromatin status using public histone modification data
  - Regulation network analyses
  - Identification and clustering of bivalent domains
  - EPI, PE and TE-specific gene expression analyses
  - Tissue specificity score
  - Public hPSC bivalent gene and bivalent domains analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stem.2022.06.001.

Cell Stem Cell
Article

CellPress

AUTHOR CONTRIBUTIONS

X.L. and S.G. conceived and designed the experiments. R.X. and X.L. performed the CUT&RUN experiments. S.L., X.O., M.J., and L.G. collected and prepared the human oocytes and embryos. Q.W. and C.W. designed and performed the data analysis with the help of X.D. C.L. performed the dissection of blastocyst embryos. R.X. performed the RNA-seq and whole genome bisulfite sequencing (WGBS) experiments. H.W., M.C., and L.Y. assisted with the sample preparation. R.X., X.L., Q.W., C.W., and S.G. wrote the manuscript.

DECLARATION OF INTERESTS

S.G. is a Cell Stem Cell Advisory Board member.

INCLUSION AND DIVERSITY

We worked to ensure gender balance in the recruitment of human subjects. We worked to ensure ethnic or other types of diversity in the recruitment of human subjects. We worked to ensure sex balance in the selection of non-human subjects. While citing references scientifically relevant for this work, we also actively worked to promote gender balance in our reference list.

REFERENCES

Amdani, S.N., Yeste, M., Jones, C., and Coward, K. (2015). Sperm factors and oocyte activation: current controversies and considerations. Biol. Reprod. 93, 50. https://doi.org/10.1095/biolreprod.115.130609.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq–a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. https://doi.org/10.1093/bioinformatics/btu638.

Bannister, A.J., and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. Cell Res 21, 381–395. https://doi.org/10.1038/cr.2011.22.

Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-dependent heterochromatin: barrier to cell fate changes. Trends Genet 32, 29–41. https://doi.org/10.1016/j.tig.2015.11.001.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125, 315–326. https://doi.org/10.1016/j.cell.2006.02.041.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. 16, 6–21. https://doi.org/10.1101/gad.947102.

Bonte, D., Reddy Guggilla, R., Stamatiadis, P., De Sutter, P., and Heindryckx, B. (2018). Chapter 14. Unraveling the causes of failed fertilization After intracytoplasmic sperm injection due to oocyte activation deficiency. In Reproductomics, J.A. Horcajadas and J. Gosálvez, eds. (Academic Press), pp. 243–277. https://doi.org/10.1016/B978-0-12-812571-7.00015-0.

Bruno, M., Mahgoub, M., and Macfarlan, T.S. (2019). The arms race Between KRAB-zinc finger proteins and endogenous retroelements and its impact on mammals. Annu. Rev. Genet. 53, 393–416. https://doi.org/10.1146/annurev-genet-112618-043717.

Burton, A., Brochard, V., Galan, C., Ruiz-Morales, E.R., Rovira, Q., Rodriguez-Terrones, D., Kruse, K., Le Gras, S., Udayakumar, V.S., Chin, H.G., et al. (2020). Heterochromatin establishment during early mammalian development is regulated by pericentromeric RNA and characterized by non-repressive

H3K9me3. Nat. Cell Biol. 22, 767–778. https://doi.org/10.1038/s41556-020-0536-6.

Burton, A., and Torres-Padilla, M.E. (2014). Chromatin dynamics in the regulation of cell fate allocation during early embryogenesis. Nat. Rev. Mol. Cell Biol. 15, 723–734. https://doi.org/10.1038/nrm3885.

Calo, E., and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? Mol. Cell 49, 825–837. https://doi.org/10.1016/j.molcel.2013.01.038.

Canovas, S., and Ross, P.J. (2016). Epigenetics in preimplantation mammalian development. Theriogenology 86, 69–79. https://doi.org/10.1016/j.theriogenology.2016.04.020.

Chatot, C.L., Ziomek, C.A., Bavister, B.D., Lewis, J.L., and Torres, I. (1989). An improved culture medium supports development of random-bred 1-cell mouse embryos in vitro. J. Reprod. Fertil. 86, 679–688. https://doi.org/10.1530/jrf.0.0860679.

Chen, Z., and Zhang, Y. (2019). Loss of DUX causes minor defects in zygotic genome activation and is compatible with mouse development. Nat. Genet. 51, 947–951. https://doi.org/10.1038/s41588-019-0418-7.

Chung, Y.G., Matoba, S., Liu, Y., Eum, J.H., Lu, F., Jiang, W., Lee, J.E., Sepilian, V., Cha, K.Y., Lee, D.R., and Zhang, Y. (2015). Histone demethylase expression enhances human somatic cell nuclear transfer efficiency and promotes derivation of pluripotent stem cells. Cell Stem Cell 17, 758–766. https://doi.org/10.1016/j.stem.2015.10.001.

Chuong, E.B., Rumi, M.A., Soares, M.J., and Baker, J.C. (2013). Endogenous retroviruses function as species-specific enhancer elements in the placenta. Nat. Genet. 45, 325–329. https://doi.org/10.1038/ng.2553.

Cook, T., Gebelein, B., Mesa, K., Mladek, A., and Urrutia, R. (1998). Molecular cloning and characterization of TIEG2 reveals a new subfamily of transforming growth factor-beta-inducible Sp1-like zinc finger-encoding genes involved in the regulation of cell growth. J. Biol. Chem. 273, 25929–25936. https://doi.org/10.1074/jbc.273.40.25929.

Cosby, R.L., Chang, N.C., and Feschotte, C. (2019). Host-transposon interactions: conflict, cooperation, and cooption. Genes Dev. 33, 1098–1116. https://doi.org/10.1101/gad.327312.119.

Cui, W., and Mager, J. (2018). Transcriptional regulation and genes involved in first lineage specification During preimplantation development. Adv. Anat. Embryol. Cell Biol. 229, 31–46. https://doi.org/10.1007/978-3-319-63187-5_4.

De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. Nat. Genet. 49, 941–945. https://doi.org/10.1038/ng.3858.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dodge, J.E., Kang, Y.K., Beppu, H., Lei, H., and Li, E. (2004). Histone H3-K9 methyltransferase ESET is essential for early development. Mol. Cell. Biol. 24, 2478–2486. https://doi.org/10.1128/MCB.24.6.2478-2486.2004.

Ecco, G., Imbeault, M., and Trono, D. (2017). KRAB zinc finger proteins. Development 144, 2719–2729. https://doi.org/10.1242/dev.132605.

Eckersley-Maslin, M.A., Alda-Catalinas, C., and Reik, W. (2018). Dynamics of the epigenetic landscape during the maternal-to-zygotic transition. Nat. Rev. Mol. Cell Biol. 19, 436–450. https://doi.org/10.1038/s41580-018-0008-z.

Geng, L.N., Yao, Z., Snider, L., Fong, A.P., Cech, J.N., Young, J.M., van der Maarel, S.M., Ruzzo, W.L., Gentleman, R.C., Tawil, R., et al. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. Dev. Cell 22, 38–51. https://doi.org/10.1016/j.devcel.2011.11.013.

Guo, M., Zhang, Y., Zhou, J., Bi, Y., Xu, J., Xu, C., Kou, X., Zhao, Y., Li, Y., Tu, Z., et al. (2019). Precise temporal regulation of Dux is important for embryo development. Cell Res 29, 956–959. https://doi.org/10.1038/s41422-019-0238-4.

Hackett, J.A., and Surani, M.A. (2013). DNA methylation dynamics during the mammalian life cycle. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20110328. https://doi.org/10.1098/rstb.2011.0328.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. *49*, 925–934. https://doi.org/10.1038/ng.3844.

Hendriks, I.A., D'Souza, R.C., Yang, B., Verlaan-de Vries, M., Mann, M., and Vertegaal, A.C. (2014). Uncovering global SUMOylation signaling networks in a site-specific manner. Nat. Struct. Mol. Biol. *21*, 927–936. https://doi.org/10.1038/nsmb.2890.

Imbeault, M., Helleboid, P.Y., and Trono, D. (2017). KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. Nature *543*, 550–554. https://doi.org/10.1038/nature21683.

Jukam, D., Shariati, S.A.M., and Skotheim, J.M. (2017). Zygotic genome activation in vertebrates. Dev. Cell *42*, 316–332. https://doi.org/10.1016/j.devcel.2017.07.026.

Kaneko-Ishino, T., and Ishino, F. (2010). Retrotransposon silencing by DNA methylation contributed to the evolution of placentation and genomic imprinting in mammals. Dev. Growth Differ. *52*, 533–543. https://doi.org/10.1111/j.1440-169X.2010.01194.x.

Kimura, Y., and Yanagimachi, R. (1995). Intracytoplasmic sperm injection in the mouse. Biol. Reprod. *52*, 709–720. https://doi.org/10.1095/biolreprod52.4.709.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., et al. (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet *8*, e1002440. https://doi.org/10.1371/journal.pgen.1002440.

Layer, R.M., Pedersen, B.S., DiSera, T., Marth, G.T., Gertz, J., and Quinlan, A.R. (2018). GIGGLE: a search engine for large-scale integrated genome analysis. Nat. Methods *15*, 123–126. https://doi.org/10.1038/nmeth.4556.

Leidenroth, A., Clapp, J., Mitchell, L.M., Coneyworth, D., Dearden, F.L., Iannuzzi, L., and Hewitt, J.E. (2012). Evolution of DUX gene macrosatellites in placental mammals. Chromosoma *121*, 489–497. https://doi.org/10.1007/s00412-012-0380-y.

Leung, D.C., and Lorincz, M.C. (2012). Silencing of endogenous retroviruses: when and why do histone marks predominate? Trends Biochem. Sci. *37*, 127–133. https://doi.org/10.1016/j.tibs.2011.11.006.

Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595. https://doi.org/10.1093/bioinformatics/btp698.

Liu, Q., Li, A., Tian, Y., Liu, Y., Li, T., Zhang, C., Wu, J.D., Han, X., and Wu, K. (2016a). The expression profile and clinic significance of the SIX family in non-small cell lung cancer. J. Hematol. Oncol. *9*, 119. https://doi.org/10.1186/s13045-016-0339-1.

Liu, W., Liu, X., Wang, C., Gao, Y., Gao, R., Kou, X., Zhao, Y., Li, J., Wu, Y., Xiu, W., et al. (2016b). Identification of key factors conquering developmental arrest of somatic cell cloned embryos by combining embryo biopsy and single-cell sequencing. Cell Discov *2*, 16010. https://doi.org/10.1038/celldisc.2016.10.

Liu, X., Wang, C., Liu, W., Li, J., Li, C., Kou, X., Chen, J., Zhao, Y., Gao, H., Wang, H., et al. (2016c). Distinct features of H3K4me3 and H3K27me3 chromatin domains in pre-implantation embryos. Nature *537*, 558–562. https://doi.org/10.1038/nature19362.

Liu, Y., Lu, X., Ye, M., Wang, L., Tang, R., Yang, Z., Turathum, B., Liu, C., Xue, Y., Wu, M., et al. (2022). Efficient silencing of the multicopy DUX4 gene by ABE-mediated start codon mutation in human embryos. J. Genet. Genomics. https://doi.org/10.1016/j.jgg.2022.02.010.

Liu, Z., Tardat, M., Gill, M.E., Royo, H., Thierry, R., Ozonov, E.A., and Peters, A.H. (2020). SUMOylated PRC1 controls histone H3.3 deposition and genome

integrity of embryonic heterochromatin. EMBO J *39*, e103697. https://doi.org/10.15252/embj.2019103697.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The Genotype-Tissue Expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.

Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature *487*, 57–63. https://doi.org/10.1038/nature11244.

Matoba, S., Liu, Y., Lu, F., Iwabuchi, K.A., Shen, L., Inoue, A., and Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. Cell *159*, 884–895. https://doi.org/10.1016/j.cell.2014.09.055.

Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., Tanaka, T., Tsutsumi, S., Nakao, M., Doi, T., et al. (2015). H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. Mol. Cell *60*, 584–596. https://doi.org/10.1016/j.molcel.2015.10.025.

Mazid, M.A., Ward, C., Luo, Z., Liu, C., Li, Y., Lai, Y., Wu, L., Li, J., Jia, W., Jiang, Y., et al. (2022). Rolling back human pluripotent stem cells to an eight-cell embryo-like stage. Nature *605*, 315–324. https://doi.org/10.1038/s41586-022-04625-0.

Mei, S., Qin, Q., Wu, Q., Sun, H., Zheng, R., Zang, C., Zhu, M., Wu, J., Shi, X., Taing, L., et al. (2017). Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res *45*, D658–D662. https://doi.org/10.1093/nar/gkw983.

Meistermann, D., Bruneau, A., Loubersac, S., Reignier, A., Firmin, J., François-Campion, V., Kilens, S., Lelièvre, Y., Lammers, J., Feyeux, M., et al. (2021). Integrated pseudotime analysis of human pre-implantation embryo single-cell transcriptomes reveals the dynamics of lineage specification. Cell Stem Cell *28*, 1625–1640.e6. https://doi.org/10.1016/j.stem.2021.04.027.

Messerschmidt, D.M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B.B. (2012). Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. Science *335*, 1499–1502. https://doi.org/10.1126/science.1216154.

Mihajlović, A.I., and Bruce, A.W. (2017). The first cell-fate decision of mouse preimplantation embryo development: integrating cell position and polarity. Open Biol *7*, 170210. https://doi.org/10.1098/rsob.170210.

Minami, N., Suzuki, T., and Tsukamoto, S. (2007). Zygotic gene activation and maternal factors in mammals. J. Reprod. Dev. *53*, 707–715. https://doi.org/10.1262/jrd.19029.

Nacerddine, K., Lehembre, F., Bhaumik, M., Artus, J., Cohen-Tannoudji, M., Babinet, C., Pandolfi, P.P., and Dejean, A. (2005). The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. Dev. Cell *9*, 769–779. https://doi.org/10.1016/j.devcel.2005.10.007.

Nicetto, D., Donahue, G., Jain, T., Peng, T., Sidoli, S., Sheng, L., Montavon, T., Becker, J.S., Grindheim, J.M., Blahnik, K., et al. (2019). H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. Science *363*, 294–297. https://doi.org/10.1126/science.aau0583.

Pertea, M., Pertea, G.M., Antonescu, C.M., Chang, T.C., Mendell, J.T., and Salzberg, S.L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat. Biotechnol. *33*, 290–295. https://doi.org/10.1038/nbt.3122.

Peters, A.H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schöfer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., et al. (2001). Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. Cell *107*, 323–337. https://doi.org/10.1016/s0092-8674(01)00542-6.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016a). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human

preimplantation embryos. Cell *165*, 1012–1026. https://doi.org/10.1016/j.cell.2016.03.023.

Petropoulos, S., Edsgärd, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Reyes, A.P., Linnarsson, S., Sandberg, R., and Lanner, F. (2016b). Single-cell RNA-seq reveals lineage and X chromosome dynamics in human preimplantation embryos. Cell *167*, 285. https://doi.org/10.1016/j.cell.2016.08.009.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. *9*, 171–181. https://doi.org/10.1038/nprot.2014.006.

Rivera, R.M., and Ross, J.W. (2013). Epigenetics in fertilization and preimplantation embryo development. Prog. Biophys. Mol. Biol. *113*, 423–432. https://doi.org/10.1016/j.pbiomolbio.2013.02.001.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140. https://doi.org/10.1093/bioinformatics/btp616.

Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P.V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature *463*, 237–240. https://doi.org/10.1038/nature08674.

Rugg-Gunn, P.J., Cox, B.J., Ralston, A., and Rossant, J. (2010). Distinct histone modifications in stem cell lines and tissue lineages from the early mouse embryo. Proc. Natl. Acad. Sci. USA *107*, 10783–10790. https://doi.org/10.1073/pnas.0914507107.

Sadakierska-Chudy, A., and Filip, M. (2015). A comprehensive view of the epigenetic landscape. Part II: Histone post-translational modification, nucleosome level, and chromatin regulation by ncRNAs. Neurotox. Res. *27*, 172–197. https://doi.org/10.1007/s12640-014-9508-6.

Schultz, D.C., Ayyanathan, K., Negorev, D., Maul, G.G., and Rauscher, F.J., 3rd. (2002). SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. Genes Dev. *16*, 919–932. https://doi.org/10.1101/gad.973302.

Schulz, K.N., and Harrison, M.M. (2019). Mechanisms regulating zygotic genome activation. Nat. Rev. Genet. *20*, 221–234. https://doi.org/10.1038/s41576-018-0087-x.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22. https://doi.org/10.1038/nrg1981.

Senft, A.D., and Macfarlan, T.S. (2021). Transposable elements shape the evolution of mammalian development. Nat. Rev. Genet. *22*, 691–711. https://doi.org/10.1038/s41576-021-00385-1.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220. https://doi.org/10.1038/nrg3354.

Stoll, G.A., Oda, S.I., Chong, Z.S., Yu, M., McLaughlin, S.H., and Modis, Y. (2019). Structure of KAP1 tripartite motif identifies molecular interfaces required for retroelement silencing. Proc. Natl. Acad. Sci. USA *116*, 15042–15051. https://doi.org/10.1073/pnas.1901318116.

Sun, D., Xi, Y., Rodriguez, B., Park, H.J., Tong, P., Meong, M., Goodell, M.A., and Li, W. (2014). MOABS: model based analysis of bisulfite sequencing data. Genome Biol *15*, R38. https://doi.org/10.1186/gb-2014-15-2-r38.

Sun, M.-a., Wolf, G., Wang, Y., Ralls, S., Senft, A.D., Jin, J., Dunn-Fletcher, C.E., Muglia, L.J., and Macfarlan, T.S. (2020). Deciphering the evolution of the transcriptional and regulatory landscape in human placenta. Preprint at bioRxiv. https://doi.org/10.1101/2020.09.11.289686.

Tadros, W., and Lipshitz, H.D. (2009). The maternal-to-zygotic transition: a play in two acts. Development *136*, 3033–3042. https://doi.org/10.1242/dev.033183.

Taubenschmid-Stowers, J., Rostovskaya, M., Santos, F., Ljung, S., Argelaguet, R., Krueger, F., Nichols, J., and Reik, W. (2022). 8C-like cells capture the human zygotic genome activation program. Cell Stem Cell *29*, 449–459. https://doi.org/10.1016/j.stem.2022.01.014.

Thomas, J.H., and Schneider, S. (2011). Coevolution of retroelements and tandem zinc finger genes. Genome Res *21*, 1800–1812. https://doi.org/10.1101/gr.121749.111.

Vuoristo, S., Bhagat, S., Hydén-Granskog, C., Yoshihara, M., Gawriyski, L., Jouhilahti, E.M., Ranga, V., Tamirat, M., Huhtala, M., Kirjanov, I., et al. (2022). DUX4 is a multifunctional factor priming human embryonic genome activation. iScience *25*, 104137. https://doi.org/10.1016/j.isci.2022.104137.

Wang, B., Wang, M., Zhang, W., Xiao, T., Chen, C.H., Wu, A., Wu, F., Traugh, N., Wang, X., Li, Z., et al. (2019). Integrative analysis of pooled CRISPR genetic screens using MAGeCKFlute. Nat. Protoc. *14*, 756–780. https://doi.org/10.1038/s41596-018-0113-7.

Wang, C., Liu, X., Gao, Y., Yang, L., Li, C., Liu, W., Chen, C., Kou, X., Zhao, Y., Chen, J., et al. (2018). Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. Nat. Cell Biol. *20*, 620–631. https://doi.org/10.1038/s41556-018-0093-4.

Wang, F., Kou, Z., Zhang, Y., and Gao, S. (2007). Dynamic reprogramming of histone acetylation and methylation in the first cell cycle of cloned mouse embryos. Biol. Reprod. *77*, 1007–1016. https://doi.org/10.1095/biolreprod.107.063149.

Wang, L., Wansleeben, C., Zhao, S., Miao, P., Paschen, W., and Yang, W. (2014). SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. EMBO Rep *15*, 878–885. https://doi.org/10.15252/embr.201438534.

Whiddon, J.L., Langford, A.T., Wong, C.J., Zhong, J.W., and Tapscott, S.J. (2017). Conservation and innovation in the DUX4-family gene network. Nat. Genet. *49*, 935–940. https://doi.org/10.1038/ng.3846.

Woodage, T., Basrai, M.A., Baxevanis, A.D., Hieter, P., and Collins, F.S. (1997). Characterization of the CHD family of proteins. Proc. Natl. Acad. Sci. USA *94*, 11472–11477. https://doi.org/10.1073/pnas.94.21.11472.

Wu, J., Xu, J., Liu, B., Yao, G., Wang, P., Lin, Z., Huang, B., Wang, X., Li, T., Shi, S., et al. (2018). Chromatin analysis in human early development reveals epigenetic transition during ZGA. Nature *557*, 256–260. https://doi.org/10.1038/s41586-018-0080-8.

Xi, Y., and Li, W. (2009). BSMAP: whole genome bisulfite sequence MAPping program. BMC Bioinformatics *10*, 232. https://doi.org/10.1186/1471-2105-10-232.

Xia, W., Xu, J., Yu, G., Yao, G., Xu, K., Ma, X., Zhang, N., Liu, B., Li, T., Lin, Z., et al. (2019). Resetting histone modifications during human parental-to-zygotic transition. Science *365*, 353–360. https://doi.org/10.1126/science.aaw5118.

Xiang, G., Giardine, B.M., Mahony, S., Zhang, Y., and Hardison, R.C. (2021). S3V2-IDEAS: a package for normalizing, denoising and integrating epigenomic datasets across different cell types. Bioinformatics *37*, 3011–3013. https://doi.org/10.1093/bioinformatics/btab148.

Xu, Q., and Xie, W. (2018). Epigenome in early mammalian development: inheritance, reprogramming and establishment. Trends Cell Biol *28*, 237–253. https://doi.org/10.1016/j.tcb.2017.10.008.

Xu, R., Li, C., Liu, X., and Gao, S. (2021). Insights into epigenetic patterns in mammalian early embryos. Protein Cell *12*, 7–28. https://doi.org/10.1007/s13238-020-00757-z.

Yang, B.X., El Farran, C.A., Guo, H.C., Yu, T., Fang, H.T., Wang, H.F., Schlesinger, S., Seah, Y.F., Goh, G.Y., Neo, S.P., et al. (2015). Systematic identification of factors for provirus silencing in embryonic stem cells. Cell *163*, 230–245. https://doi.org/10.1016/j.cell.2015.08.037.

Yao, C., Zhang, W., and Shuai, L. (2019). The first cell fate decision in pre-implantation mouse embryos. Cell Regen *8*, 51–57. https://doi.org/10.1016/j.cr.2019.10.001.

Yeste, M., Jones, C., Amdani, S.N., and Coward, K. (2017). Oocyte activation and fertilisation: crucial contributors from the sperm and oocyte. Results Probl. Cell Differ. *59*, 213–239. https://doi.org/10.1007/978-3-319-44820-6_8.

Young, J.M., Whiddon, J.L., Yao, Z., Kasinathan, B., Snider, L., Geng, L.N., Balog, J., Tawil, R., van der Maarel, S.M., and Tapscott, S.J. (2013). DUX4

binding to retroelements creates promoters that are active in FSHD muscle and testis. PLoS Genet. *9*, e1003947. https://doi.org/10.1371/journal.pgen.1003947.

Yu, L., Duan, J., Schmitz, D.A., Sakurai, M., Wang, L., Wang, K., Zhao, S., Hon, G.C., and Wu, J. (2021). Blastocyst-like structures generated from human pluripotent stem cells. Nature *591*, 620–626. https://doi.org/10.1038/s41586-021-03356-y.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao, Z.A., et al. (2018). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. Nat. Genet. *50*, 96–105. https://doi.org/10.1038/s41588-017-0003-x.

Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.H., Brown, M., Zhang, X., Meyer, C.A., and Liu, X.S. (2019). Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res *47*, D729–D735. https://doi.org/10.1093/nar/gky1094.

Zhu, P., Guo, H., Ren, Y., Hou, Y., Dong, J., Li, R., Lian, Y., Fan, X., Hu, B., Gao, Y., et al. (2018). Single-cell DNA methylome sequencing of human preimplantation embryos. Nat. Genet. *50*, 12–19. https://doi.org/10.1038/s41588-017-0007-6.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Anti-H3K4me3	Cell Signaling Technology	Cat# 9727S; RRID: AB_561095
Anti-H3K27me3	Diagenode	Cat# pAb-069-050; RRID: AB_2616049
Anti-H3K9me3	Active Motif	Cat# 39161; RRID: AB_2532132
Chemicals, peptides, and recombinant proteins		
G-1 PLUS	Vitrolife	Cat# 10128
G-IVF PLUS	Vitrolife	Cat# 10136
Quinn's Advantage Ca/Mg-Free Medium with HEPES	SAGE	Cat# 06611
Pronase E	Sigma-Aldrich	Cat# P8811
Bovine Serum Albumin	Sigma-Aldrich	Cat# A3311
SuperScript II Reverse Transcriptase	Thermo Fisher Scientific	Cat# 18064022
Ambion RNase Inhibitor	Thermo Fisher Scientific	Cat# AM2682
Terminal Deoxynucleotidyl Transferase, recombinant	Thermo Fisher Scientific	Cat# 10533-073
Agencourt AMPure XP beads	Beckman Coulter	Cat# A63880
pA-MNase	a gift from Steven Henikoff lab	
DAPI	Sigma-Aldrich	Cat# 10236276001
Critical commercial assays		
KAPA HyperPrep Kit	KAPA Biosystems	Cat# KK8504
Pico Methyl-Seq Library Prep Kit	Zymo Research	Cat# D5456
Deposited data		
H3K9me3, H3K4me3, H3K27me3 CUT&RUN, WGBS and RNA-seq data (human)	This paper	GSA: HRA001391
H3K9me3 CUT&RUN and RNA-seq data (mouse)	This paper	GSA: CRA005107
Experimental models: Cell lines		
Mouse embryonic stem cell line R1	ATCC	N/A
Experimental models: Organisms/strains		
Mature C57BL/6n female mice	Beijing Vital River Laboratory Animal Technology	Stock No.:213
Mature DBA2 male mice	Beijing Vital River Laboratory Animal Technology	Stock No.: 214
Mature Dux-KO female mice	Shaorong Gao lab, Tongji University, Shanghai, China	(Guo et al., 2019)
Oligonucleotides		
siRNA1 for *Zfp51* KD sense: 5'- AUGCUGUA ACAACGAGGAUTT -3'	This paper	N/A
siRNA1 for *Zfp51* KD anti-sense: 5'- AUCCUC GUUGUUACAGCAUCG -3'	This paper	N/A
siRNA2 for *Zfp51* KD sense: 5'- AGCGUGCA UAUCUUAGAAATT -3'	This paper	N/A
siRNA2 for *Zfp51* KD anti-sense: 5'- UUUCUA AGAUAUGCACGCUGG -3'	This paper	N/A
siRNA3 for *Zfp51* KD sense: 5'- CAGUAUUCU AUCUGCUUAUTT -3'	This paper	N/A

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
siRNA3 for *Zfp51* KD anti-sense: 5'- AUAAG CAGAUAGAAUACUGGA -3'	This paper	N/A
Software and algorithms		
Bsmap v2.89	(Xi and Li, 2009)	http://dldcc-web.brc.bcm.edu/lilab/yxi/ bsmap/bsmap-2.89.tgz
Moabs v1.3.0	(Sun et al., 2014)	http://dldcc-web.brc.bcm.edu/lilab/ deqiangs/ moabs/moabs.html
STAR v2.5.2b	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
MACS2 v2.1.1.20160309	(Zhang et al., 2008)	https://github.com/macs3-project/ MACS/releases/tag/ v2.1.1.20160309
IDEAS v1.11	(Xiang et al., 2021)	https://github.com/guanjue/S3V2_ IDEAS_ESMP
Bwa v0.7.15	(Li and Durbin, 2010)	https://github.com/lh3/bwa
StringTie v1.3.6	(Pertea et al., 2015)	https://ccb.jhu.edu/software/stringtie/
R v3.5.1	https://www.R-project.org/	https://www.R-project.org/
HTSeq v0.6.0	(Anders et al., 2015)	https://htseq.readthedocs.io/en/master/
MEME Suite 5.4.1	(Bailey et al., 2015)	https://meme-suite.org/meme/
HOMER v4.11	(Heinz et al., 2010)	http://homer.ucsd.edu/homer/
Other		
Mouse Embro WGBS data (IVF)	(Wang et al., 2018)	GEO: GSE97778
Mouse ES H3K4me3 ChIP-seq data	(Liu et al., 2016)	GEO: GSE73952
Human Embryo ATAC-seq data	(Wu et al., 2018)	GEO: GSE101571
Human Embryo H3K27ac data	(Xia et al., 2019)	GEO: GSE124718
Human naïve PSC histone data	(Wang et al., 2018)	GEO: GSE89072
Human D5-D7 scRNA-seq data	(Petropoulos et al., 2016)	ArrayExpress: E-MTAB-3929

RESOURCE AVAILABILITY

Lead contact
Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Shaorong Gao (gaoshaorong@tongji.edu.cn).

Materials availability
This study did not generate new unique reagents.

Data and code availability
All the ChIP-seq, RNA-seq and BS-seq data generated in this study were summarized in Table S1 and have been deposited to the GSA database under the accession number HRA001391 and CRA005107. Cistrome ChIP-seq data were downloaded from the Cistrome database. Human embryo ATAC-seq and H3K27ac dataset were downloaded from GSE101571 and GSE124718. Mouse ES H3K4me3 data were downloaded from our previous publications GSE73952. Mouse embryo ATAC data were downloaded from GSE66581. This paper does not report original code. All the other data supporting the findings of this study are available from the corresponding author upon reasonable request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

The specific pathogen-free grade mice (SPF) grade mice, including C57BL/6n, DBA2 and BDF1 mice were housed in the animal facility at Tongji University, Shanghai, China. The BDF1 hybrid mice (8-10 weeks old) were obtained from mating female C57BL/6n mice with male DBA/2 mice. Dux-KO mice were generated and maintained as previously described (Guo et al., 2019). All the mice had free access to food and water. All experiments were performed in accordance with the University of Health Guide for the Care and Use of Laboratory Animals and were approved by the Biological Research Ethics Committee of Tongji University.

Ethics statement
This study was approved by the Institutional Review Board (IRB) of Guangdong Second Provincial General Hospital (2018-SZLL-010), China. In accordance with the measures of the People's Republic of China on the administration of Human Assisted

Reproductive Technology, the ethical principles of the Human Assisted Reproductive Technology and the Helsinki declaration. The research followed the guiding principles of the Human Embryonic Stem Cell Ethics issued by the MOST and MOH and was regularly reviewed by the Medical Ethics Committee of Guangdong Second Provincial General Hospital. All oocytes and embryos were donated by volunteers after signing informed consent at the Center for Reproductive Medicine in Guangdong Second Provincial General Hospital. Volunteers were informed that their oocytes and embryos would be only used to study the histone modifications during human embryo development.

METHOD DETAILS

Human oocytes and early embryos collection

The discarded oocytes and embryos were from patients who underwent assisted reproductive technology due to tubal factor infertility. The patients received the controlled ovarian stimulation using GnRH agonist or GnRH combined with human menopausal gonadotrophins or recombinant follicle stimulating hormone (FSH) for pituitary desensitization. Human chorionic gonadotrophin (hCG) was administered to trigger ovulation. Transvaginal ultrasound-guided oocyte retrieval was scheduled 36 hours after hCG administration. MI oocytes were identified by microscopy according to the absence of nucleus. We obtained MII oocytes after *in vitro* maturation (IVM) of MI oocytes. Embryos were derived from *in vitro* fertilization (IVF) of fresh MII oocytes and sperms, and fertilization was verified 16-19 h post fertilization (h.p.f). For the consistency of sample collection, all the donated oocytes and embryos went through vitrification and thawing program before performing CUT&RUN, RNA-seq and BS-seq. MII oocytes, 4-cell (43-45 h.p.f), 8-cell (67-69 h.p.f) and blastocysts were vitrified in liquid nitrogen for storage and sample pooling. 4-cell embryos from zygotes with three pronuclei were also collected, which exhibited morphologically high quality without developmental arrest. Blastocysts at D5 (114-118 h.p.f) with an expansion grade from B4 to B6 were selected as blastocysts samples (B4 blastocysts at D6 were also accepted).

Collection of mouse pre-implantation embryos

To get mouse fertilized embryos, 6∼8-week-old B6D2F1 (C57BL/6n × DBA2, BDF1) female mice were super-ovulated by injection with 7 IU each of pregnant mare serum gonadotropin (PMSG), followed by injection of 5 IU of human chorionic gonadotropin (hCG) (San-Sheng Pharmaceutical) 48 h later. The super-ovulated BDF1 female mice were mated with B6D2F1 male mice. Then, the zygotes were collected from the oviducts of the female mice at 20 h post hCG (h.p.h) injection and were cultured in G-1 PLUS medium until blastocysts. Late 1-cell (26-28 h.p.h), early 2-cell (30-32 h.p.h), late 2-cell (47-48 h.p.h), 4-cell (56 h.p.h), 8-cell (67 h.p.h) and ICM/TE (96 h.p.h) were harvested for Smart-Seq2 and H3K9me3 CUT&RUN.

Mouse *Dux* KO embryos collection

Dux$^{-/-}$MII oocytes were retrieved from super-ovulated 6∼8-week-old *Dux*$^{-/-}$female mice, followed by intracytoplasmic sperm injection (ICSI). The sperm head generated from *Dux*$^{-/-}$ male mice was then injected into the oocyte according to the method described previously (Kimura and Yanagimachi, 1995). HEPES-buffered CZB (HCZB) medium was used for gamete handling and ICSI in air. G-1 PLUS medium was used for embryo culture in an atmosphere of 5% CO_2. Late 1-cell (14-16 h.p.f), early 2-cell (18-20 h.p.f), late 2-cell (35-36 h.p.f), 4-cell (44 h.p.f), 8-cell (55 h.p.f) and ICM/TE (84 h.p.f) were harvested for Smart-Seq2 and H3K9me3 CUT&RUN (except for late 1-cell). Notably, sample of early-2cell for H3K9me3 CUT&RUN was generated by *Dux*$^{-/-}$ female mice and WT BDF1 male mice.

Knockdown of and *Zfp51* in mouse embryos

Three siRNAs against *Zfp51* were diluted and mixed in nuclease-free water to a working concentration of 20 μM per siRNA. Isolated mouse zygotes were injected with siRNAs and cultured in G-1 PLUS medium at 37°C under 5% CO_2. Knocking-down (KD) efficiency was examined at 2-cell stage. ICM and TE of E3.5 blastocysts were then isolated and collected for Smart-seq2 and H3K9me3 CUT&RUN.

Reverse transcription and quantitative RT-PCR analysis

For quantitative RT-PCR analysis of *Zfp51*-KD efficiency, cDNA of 10 blastomeres were synthesized and amplified using Smart-Seq2. Quantitative RT–PCR was performed using a SYBR Premix Ex Taq and signals were detected with the ABI7500 Real-Time PCR System. *H2afz* was used as an endogenous control.

Immunofluorescent staining

Human MII oocytes, 4-cell and 8-cell embryos and blastocysts were fixed with 4% paraformaldehyde (Sigma) for 1 hour at room temperature and then permeabilized with 0.2% Triton X-100 for 1 hour at room temperature. the samples were incubated with the primary antibodies against H3K9me3 for 2 hours at room temperature. after washing three times with 1% bovine serum albumin (BSA) (Sigma) in PBS, the samples were incubated with the appropriate secondary antibodies for 1 hour at room temperature. the nuclei were stained with 4',6-diamidino-2-phenylindole (DAPI). All stained samples were observed using a Zeiss LSM880 confocal microscope. Images were processed and quantified in ImageJ software.

Sample harvest for CUT&RUN, RNA-seq and BS-seq

For MII oocytes and cleavage-stage embryos in human, the zona pellucidae of the embryos were removed with 0.5% pronase E (Sigma). Polar bodies were removed by gentle pipetting using a fire-polished glass needle. For ICM and TE isolation, the zona pellucidae of blastocysts were removed with 0.5% pronase E (Sigma). To eliminate tight cell-cell junctions, the zona pellucidae removed blastocyst were incubated in Ca^{2+}-free CZB (for mouse) or Ca^{2+}/Mg^{2+}-free medium (Quinn's Advantage Medium with HEPES Ca and Mg Free) with 0.005% Trypsin/EDTA and 0.5% BSA (for human) for 20 minutes (Chatot et al., 1989). ICMs (12-15 μm) and TEs (18-20 μm) were then separated by micromanipulation using needles with an inner diameter of 20 μm, according to their distinct morphology and spatial position. Notably, given the initiation of lineage segregation of ICMs in D5 blastocysts, the separated ICMs likely contain both EPI and PE cells (Petropoulos et al., 2016a). Spatially, since TE comprises the outer layer of a blastocyst while ICM lies inside, the cells separated at first were basically TE, with much flabbier cell-cell junction than that of ICM. Morphologically, TE possesses more smooth cell membrane surface and mostly larger cell size. To ensure the accuracy of our manual separation to the greatest extent, any cell with indistinct phenotype that might misguide our judgment was discarded.

Cell culture

The R1 ES cells were purchased from the American Type Culture Collection (ATCC) and not further authenticated. All cell lines regularly tested negative for mycoplasm contamination. The R1 ES cells were cultured on mitomycin-C-treated MEFs in ES medium containing DMEM (Merk Millipore) supplemented with 15% (v/v) fetal bovine serum (Hyclone), 1 mM l-glutamine (Merk Millipore), 0.1 mM mercaptoethanol (Merk Millipore), 1% nonessential amino acid stock (Merk Millipore), penicillin/streptomycin (100×, Merk Millipore), nucleosides (100×, Merk Millipore) and 1,000 U/mL LIF (Merk Millipore).

CUT&RUN

CUT&RUN was conducted following the modified published protocol. 50 oocytes or embryos were used per reaction, and one or two replicates were performed for each stage. The samples were resuspended in 600μL room temperature Wash buffer (HEPES pH=7.5, 20mM; NaCl, 150mM; Spermidine, 0.5mM; BSA, 0.1% and Roche complete protease inhibitor) by gently pipetting. 5μL/reactions concanavalin-coated magnetic beads were transferred to a microfuge tube containing 3X volume cold Binding Buffer (HEPES-KOH pH=7.9, 20mM; KCl, 10mM; $CaCl_2$, 1mM; $MnCl_2$, 1mM). Beads were washed twice in 1mL cold Binding buffer and resuspended in 300μL binding buffer. Beads were added to cells with gentle vortexing and incubated for 10 minutes (min) at room temperature. Bead-bound nuclei were blocked with 1mL cold Blocking Buffer (HEPES pH7.5, 20mM; NaCl, 150mM; Spermidine 0.5mM; BSA, 0.1%; EDTA pH=8.0, 2mM, PIC and Digitonin, 0.1%), incubated for 5 min at room temperature. Beads were washed in 1mL cold Dig-Wash Buffer (Digitonin, 0.01%) and resuspended in 250μL Dig-Wash Buffer. 1μg antibody of H3K4me3 (Cell signaling Technology, #9727), H3K27me3 (Diagnode, pAb-069-050) or H3K9me3 (Active Motif, 39161) was added with genetl vortexing of bead-bound cell in 250μL cold Dig-Wash Buffer. Sample were incubated with rotation at 4°C for 2 h or overnight. Samples were washed twice in 1mL cold Dig-Wash Buffer and resuspended in 250ul cold Dig-Wash Buffer with 1:500 deluted pA-MNase (a gift from Steven Henikoff lab). Samples were incubated with rotation at 4°C for 1 h and washed twice in 1mL cold Dig-Wash Buffer. The supernatant was discarded and samples were resuspended in 150μL cold Dig-Wash Buffer. Samples were equilibrated to 0°C on ice for 5 min and 3μL $CaCl_2$ (100mM) was added to initiated cleavage. Reactions were stopped by 150μL STOP Buffer (NaCl, 200mM; EDTA pH=8.0, 20mM; EGTA pH=8.0, 4mM; Digitonin, 0.02%; RNase A, 50ug/mL; glycogen, 20μg/mL) after 15 min digestion. Samples were incubated at 37°C for 20 min to digest RNA and release DNA fragments. Samples were centrifuged at 16,000 g for 5 min and supernatants were transferred to a new micruofge tube while pellets and beads were discarded. 3μL 10% SDS and 2.5μL Proteinase K 2(20mg/mL) were added to wach sample and incubate at 70°C for 10 min. DNA was purified by phenol chloroform followed by ethanol purification. The sequence libraries were generated using KAPA Hyper Prep Kit for the Illumina platform (kk8504), following the manufacture's instructions. Paired-end 150-bp sequencing was performed on a NovaSeq (Illumina) platform in Berry Genomics and Novogene.

Smart-Seq2

For Smart-seq2, 10 blastomeres were used per reaction, and three replicates were performed for each stage. All isolated blastomeres were washed three times in 0.5% BSA-PBS solution to avoid possible contamination. RNA-seq libraries were generated using the Covaris DNA shearing protocol for Smart-seq sequence library generation as previously described. Briefly, RNAs with a polyadenylated tail were captured, reverse transcribed and pre-amplified (Picelli et al., 2014). After fragmentation, the sequence libraries were generated using KAPA HyperPrep Kit for the Illumina platform, following the manufacture's instructions. Paired-end 150-bp sequencing was performed on a NovaSeq (Illumina) platform in Berry Genomics and Novogene.

WGBS

For BS-seq, 10 blastomerese were used per reaction, and three replicates were performed for each stage. All isolated blastomerese were washed three times in 0.5% BSA-PBS solution to avoid possible contamination. The sequencing libraries were generated using the Pico Methyl-Seq Library Prep Kit (D5456, Zymo Research) following the manufacture's instructions. Paired-end 150-bp sequencing was performed on a NovaSeq (Illumina) platform in Berry Genomics and Novogene.

QUANTIFICATION AND STATISTICAL ANALYSIS

CUT&RUN, RNA-seq and BS-seq data processing, quality control and normalization

CUT&RUN sequencing reads were aligned to the human genome build hg38 or mouse genome mm10 using the bwa (v0.7.15) (Li and Durbin, 2010) mem command. Signal tracks for each sample were generated using the MACS2 (v2.1.1.20160309) pileup function and were normalized to 1 million reads (–SPMR option) for visualization (Zhang et al., 2008). Duplicates were removed by setting –keep-duplicate=1 when generating the signal tracks using MACS2. For H3K4me3, H3K27me3, and H3K9me3, to examine the reproducibility of the CUT&RUN experiments, we calculated the correlation of the normalized signal intensity between biological replicates on all Ensembl gene promoters, which were defined as ±2 kb around the TSS. All replicates showed high correlation for different stages. We also confirmed the high correlation between different replicates genome-widely using 10kb bins. We then pooled the biological replicates together for each stage and performed the downstream analysis. The quality information of CUT&RUN data used in this study and the correlation of replicates in each stage are summarized in Tables S1 and S2. The RNA-seq reads were mapped to the hg38 or mm10 reference genome using STAR (v2.5.2b) (Dobin et al., 2013). Expression levels for all Ensembl genes were quantified to fragments per kilobase million (FPKM) using StringTie (v1.3.6) (Pertea et al., 2015), and FPKM values of replicates were averaged. All of the WGBS-seq reads were first processed using TrimGalore (v0.4.2) (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to trim adaptor and low-quality reads. Adaptor-trimmed reads were then mapped to a combined genome with human hg38 and 48052 lambda sequence using bsmap (v2.89) (Xi and Li, 2009). The methylation level of each CpG site was estimated using mcall (v1.3.0) (Sun et al., 2014).

CUT&RUN peak identification and genome segmentation analysis

All the CUT&RUN peaks were called by MACS2 (v2.1.1.20160309) with the parameters –nomodel –nolambda –shiftsize = 73 (Zhang et al., 2008). As the peak number detected at each stage could be affected by the sequencing depths, we used the same number of reads (40 million) when available that were randomly selected from samples of each stage. We performed S3V2-IDEAS analysis to characterize the chromatin states using multiple histone modification CUT&RUN data. The alignment files of H3K4me3, H3K27me3, and H3K9me3 modifications across 5 developmental stages were normalized and denoised by the S3V2 method and binned into 200 bp. Chromatin states of the normalized data were identified and characterized using IDEAS (v1.11) (Xiang et al., 2021). The whole genome was classified into eight different states: strong H3K4me3 domain, weak H3K4me3 domain, strong H3K27me3 domain, weak H3K27me3 domain, strong H3K9me3 domain, weak H3K9me3 domain, bivalent domain and non-marked domain. To remove the low enrichment domains, we used the normalized CUT&RUN signal on each domain for further filtering. The H3K4me3-covered domains were defined as domains with signal greater than max H3K4me3 signal of MII oocyte stage non-marked domains. Similar filters were applied on H3K27me3-covered and H3K9me3-covered domains, with each defined as signal greater than max of MII oocyte stage non-marked domain's H3K27me3 and H3K9me3 signal. The number of H3K4me3, H3K27me3 and H3K9me3 covered 200 bp domains were then plotted using ggplot2 in R (v3.5.1) (http://CRAN.R-project.org/).

To define the established, disappeared and maintained domains, we first transformed the domain matrices of the developmental stages into 0,1 matrices based on their peak signals, with 1 representing H3K4me3, H3K27me3, H3K9me3 covered regions, and 0 represent non-covered regions. The "established" domains defined as its signals across two adjacent stages are marked as "01" and the $\log_2$ fold change between later and earlier stages are more than 4; The "disappeared" domains defined as its signals across two adjacent stages are marked as "10" and the $\log_2$ fold change between earlier and later stages are more than 4; The "maintained" domains defined as its signals across two adjacent stages are marked as "11".

Genomics enrichment of histone modification domains and CpG density calculation

Genomic enrichment of histone modification domains in promoter, exon, intron, SVA composite retroposons, short interspersed nuclear element (SINE), long interspersed nuclear element (LINE), long terminal repeats (LTR) and simple repeats regions was calculated using observed versus expected probability. The observed probability was calculated using the length of the histone modification domains that covers the related genomic regions versus the length of the total histone modification domains, and the expected probability was calculated using the length of the total related genomic regions versus the length of the human genome. CpG density of each histone modification domains were calculated as the number of CpGs across 100 bp bins.

Hierarchical clustering of H3K4me3, H3K27me3 and H3K9me3 data

To classify the global feature of H3K4me3, H3K27me3, and H3K9me3, we performed hierarchical clustering on genome-wide domains and promoter regions based on normalized H3K4me3, H3K27me3, and H3K9me3 signal in human MII stage oocytes and early embryos. The euclidean distance was used to measure the distance and the complete linkage method was used to identify similar clusters.

Differential gene expression analysis and definition of minor ZGA and major ZGA genes

To perform differential gene expression analysis, we first calculated the read counts of each RNA-seq sample using HTSeq (v0.6.0) (Anders et al., 2015). Then, the results were fed into edgeR (Robinson et al., 2010) to perform differential analysis. Genes with a Benjamini and Hochberg-adjusted $P \leq 0.05$ and a fold change > 1 were defined as differentially expressed between compared stages. The minor ZGA genes were defined as the differentially up-regulated genes between 4-cell and MII oocyte samples, while the major

ZGA genes were the differentially up-regulated genes between 8-cell and 4-cell samples. For mouse, the major ZGA genes were defined as up-regulated genes between late 2-cell embryos and early 2-cell embryos. Similar DE cut-offs were used for identify DE genes between WT and KO, KD groups.

Gene ontology analysis

Functional annotation analysis was performed using the MAGeCK-Flute package (Wang et al., 2019). We only selected the Gene Ontology terms from biological processes to calculate the enrichment. P-values were calculated similar to the online tool of DAVID, which is based on a modified Fisher's exact test.

Clustering analysis of stage-specific H3K9me3 domains

To classify the stage-specific H3K9me3 domains, we first transformed the H3K9me3-covered domains at MII Oocyte, 4-cell, 8-cell, ICM and TE stage into binary matrices, using the same cut-off during the identification of dynamic histone modification domains (greater than the max H3K9me3 signal level of non-marked domains). For each domain, we arranged the column based on the order of developmental stages and transformed the binary 0,1 value into a decimal score. All of the domains were then ranked based on their decimal score and the actual H3K9me3 signal value were used to generate heatmaps. Class 1 was defined as MII Oocyte domains, which were disappeared in later stages; Class 2 was defined as 4-cell domains; Class 3 were defined as 8-cell domains, and the majority of 4-cell and 8-cell domains were lost in the later stages; and Class 4 was defined as blastocyst domains. We further defined domains with state "00100" as "8-cell specific domains", and domains with state "00111" or "00011" as "blastocyst specific domains". The distance between each domain center and gene TSS, LTR center, TRIM28 peak center and DUX4 peak center were calculated and visualized using the same order of H3K9me3-covered domains.

Expression, DNA methylation and histone modification level quantification of repeats elements

To assess the expression level of repeats elements, all the RNA-seq files were re-mapped to the hg38 or mm10 genome using the STAR (Dobin et al., 2013) aligner software, allowing up to three mismatches and filtering out reads that mapped to >500 positions in the genome. Mapped files were then processed using the makeTagDirectory script of HOMER (v4.11) with the -keepOne option (Heinz et al., 2010). The tag directories of the mapped files were analysed using the analyzeRepeats.pl script of HOMER with the option 'repeat' and -noadj. The total read counts of each sample were normalized to 1 million, and replicates were averaged for comparison. To analyse the methylation level and histone modification level of repeat elements, we downloaded the repeat annotations from the UCSC table browser. DNA methylation level and normalized histone modification signals were calculated for each repeat annotations and the values of the same repeat class were summed and then averaged by the number of copies in the genome.

Association analysis of histone modification and methylation level on repeats

We performed association analysis between the DNA methylation level and the normalized H3K9me3 signal on samples from the oocyte stage to the TE stage. Pearson's correlation was calculated for DNA methylation and the normalized H3K9me3 signal on genes, LTRs, LINEs, SINEs and SVAs, and association tests were performed based on weighted Pearson's correlation coefficient. Elements with a negative correlation and a significant association ($P \leq 0.05$) have increased H3K9me3 signal and decreased DNA methylation level along development, and we further filter those elements with max H3K9me3 signal greater than 0.1 to define H3K9me3-marked elements; elements with a positive correlation have both decreased H3K9me3 signal and DNA methylation level along development, and were further filtered by max H3K9me3 signal less than 0.1 to defined H3K9me3-unmarked elements. Similar calculations were made between the DNA methylation level and the normalized H3K27me3 signal.

To further characterize the two wave establishment of H3K9me3 and its regulation on LTRs. For human, we classified 532 different LTR types based on their covered H3K9me3 domain types. We fist calculated overlaps of 8C-specific and blastocyst-specific H3K9me3 domains with LTRs, and defined the overlapped ratio s as,

$$s = \frac{\text{Number of LTR copies overlapped with 8C} - \text{sepcific H3K9me3 domains}}{\text{Number of LTR copies overlapped with blastocyst} - \text{specific H3K9me3 domains}}$$

As each LTR type have multiple copied across the whole genome, we identified the LTR type which had at least 50 copies overlapped with 8C-specific H3K9me3 domains, the ratio s greater than 1.5 and the H3K9me3 signal on this LTR greater than 0.1 as "8C-specific H3K9me3-marked (ESH) LTR types" (5 types for human and 34 types for mouse); the LTR type which had at least 50 copies overlapped with blastocyst-specific H3K9me3 domains, the ratio s less than 0.2 and the H3K9me3 signal on this LTR greater than 0.1 was defined as "blastocyst-specific H3K9me3-marked (BSH) LTR types" (60 types for human and 35 types for mouse). These definitions were also visualized in Figure S2.

Enrichment of transcription factor binding sites and histone modifications on LTRs and H3K9me3 domains

To evaluate the enrichment of transcription regulator (TR) binding sites in LTR regions and identify potential regulators, we obtained 11,349 ChIP-seq peak files of 1,360 human transcription factors and chromatin remodlers from CistromeDB (Mei et al., 2017). We used GIGGLE (v0.6.3) to calculate the overlapped peak odds ratio between TR peaks with H3K9me3-marked LTR types or H3K9me3-unmarked LTR types (Layer et al., 2018). To determine the preference of factors on H3K9me3-marked and H3K9me3-unmarked LTRs, we defined a preference score s, for which,

$$s = \frac{\text{TR Peak Overlap Odds Ratio of H3K9me3} - \text{marked LTRs}}{\text{TR Peak Overlap Odds Ratio of H3K9me3} - \text{unmarked LTRs}}$$

The H3K9me3-marked LTRs enriched factors were those factors with at least 200 peaks overlapped with H3K9me3-marked LTR regions and $s >= 10$; The H3K9me3-unmarked LTRs enriched factors were those factors with at least 200 peaks overlapped with H3K9me3-unmarked LTRs regions and $s <= 0.1$. Similar calculations were performed for ESH and BSH LTR types, for which the minimum peak overlap was defined as 150 peaks, and s were defined as 2 or 0.15, respectively. Only top 10 factors were visualized in the scatter plot.

To identify the enriched histone modifications on ESH and BSH domains, we obtained 11,080 ChIP-seq peak files of 83 human histone modifications. Histone with less 500 peaks were removed from the analyses. And the same method was used to identify the enriched histone modifications on H3K9me3 domain region as TFs on LTR types. To focused on the potential chromatin status of H3K9me3 domains, but not LTRs in the future development, we calculated the relative enrichment of histone modifications using H3K9me3-covered domains but not H3K9me3-covered LTRs. To performed the TF and histone modification enrichment analysis on mouse H3K9me3 domains, we obtained 9,061 ChIP-seq peak files of 704 mouse transcription factors and chromatin remodlers, and 10,945 ChIP-seq peak files of 79 mouse histone modifications (Zheng et al., 2019). Similar calculations were performed to identify the mouse H3K9me3 domains enriched TFs and histone modifications, for which the minimum peak overlap was defined as 150 peaks and s defined as 2 or 0.15 for enriched TFs, and the minimum peak overlap were defined as 200 peaks and s defined as 2 or 0.2 for enriched histone modifications.

Analyses of the future chromatin status using public histone modification data

To analysis the state of ESH LTR domains in ESCs or adult stages. We downloaded the H3K4me1 and H3K27ac histone modification datasets from the ENCODE and ROADMAP epigenomics data portal, we used H1 ES cell line (GSE29611), H9 ES cell line (GSE16256), left ventricle (GSE101357), tibia nerve (GSE100993), gastrocnemius medialis (GSE101276), stomach (GSE101188), spleen (GSE101073), ovary (GSE16256), liver (GSE16256), adipose (GSE19465), lung (GSE16256), and adrenal gland (GSE16256) to evaluate the enhancer binding potency in ESH and BSH LTR domains.

Regulation network analyses

To investigate the regulation of DUX4 and ZNF factors on LTR, we first derived ChIP-seq peaks for 8C-specific regulators (DUX4, ZNF486, ZNF766, ZNF267) and blastocyst-specific regulators (ZNF736, ZNF808, ZNF680, ZNF525, ZNF816, and ZNF669) from Cistrome database. For human, we calculated the Jaccard index between ZNF peaks and LTR regions to investigate their potential binding on those LTRs. While the ZNF peaks are not available for mouse, and considering that many TF binding motifs are in general similar between human and mouse, we used MEME to scan the motifs using the top 1000 peaks from human ZNF peaks, and use the top one GC rich motifs from human to identify the potential ZNF regions in mouse. We applied FIMO analyses to scan motif in mouse LTRs, and calculated the motif occurrences using LTR count with the motif versus total LTR count. The ortholog genes between human and mouse were downloaded from gene card (https://www.genecards.org/). Regulation networks were constructed using ggraph, with node set as ZNF factors and LTRs, and edge set as the Jaccard index between ZNF peaks and LTR regions for human, or motif occurrences in LTR regions for mouse.

To evaluate the effect of Dux on the establishment of H3K9me3, we performed differential expression analyses on 371 Zfps. We evaluated the differential expression by averaging the $\log_2$ fold change between Dux KO and WT embryos. Using a stringent cut-off ($\log_2$ fold change < -0.7 or $\log_2$ fold change > 0.7), we derived 26 down-regulated and 4 up-regulated Zfps. While the up-regulated Zfps were too few for downstream analyses, we used a loose cut-off ($\log_2$ fold change > 0.3) to define 26 up-regulated Zfps. Then we used ortholog motifs of down- and up-regulated ZNFs from human to scan its potential binding on mouse LTRs, and calculated the motif occurrences on mouse LTRs using FIMO, similar to 8C/ blastocyst-specific regulators in human. We defined the enriched LTRs as its motif occurrences more than 90th quantile among all LTRs, and the Zfps were linked to 30-90 LTRs by this definition. Next, we calculated the LTRs with differential H3K9me3 level and expression using their $\log_2$ fold change between Dux KO and WT group. The LTRs with a log fold change > 0.5 and TPM > 0.05 were defined as differential expressed LTRs. The LTRs with a log fold change > 0.5 and normalized H3K9me3 signal > 0.05 were defined as differential H3K9me3-marked LTRs. Finally, we calculated the percentile of differential decreased H3K9me3 LTRs and differential up-regulated LTRs for the down- and up-regulated ZNFs enriched LTRs.

Identification and clustering of bivalent domains

To identify bivalent domains, we first derived the H3K4me3, H3K27me3 and H3K9me3 covered matrix using the same criterion in determine the dynamic histone modification domains. The K4/K27 bivalent domains were those domains with both H3K4me3 and H3K27me3 covered but not H3K9me3 covered domains, similar definition was used for K4/K9 bivalent domains and K9/K27 bivalent domains. The trivalent domains were those domains with both H3K4me3, H3K27me3 and H3K9me3 covered domains. We used k-means clustering to cluster different bivalent domains, with k setting to 5 for both K4/K27 and K4/K9 bivalent analysis.

EPI, PE and TE-specific gene expression analyses

To investigate the bivalent status and its influence on lineage specification, we downloaded the single-cell gene expression data of human D5-D7 embryos from Petropoulos et al. and identified top 500 differential expressed genes from EPI, PE and TE as the

signature gene of each lineage (Petropoulos et al., 2016b). We also derieved the mural vs polar TE signature and GATA2-regulated, NR2F2-regulated genes from Meisterman et al. to investigate the priming effect of histone modification on more detailed lineage gene list (Meistermann et al., 2021).

Tissue specificity score

We downloaded expression data across 54 human tissues from Genotype-Tissue Expression (GTEx) project (v8) (Lonsdale et al., 2013) to calculate tissue specificity scores of different genes. Genes that are not expressed in any of the 54 tissues were removed. We used tspex (https://tspex.lge.ibi.unicamp.br/) module to compute the Jensen–Shannon specicity dispersion (JSS DPM) between the relative expression levels across the tissue types. An extreme condition is defined as a transcript expressed in only one tissue type. The JSS DPM metric provides values from 0 to 1, while the 0 means ubiquitous expression and 1 means tissue-specific expression.

Public hPSC bivalent gene and bivalent domains analysis

To analyse the bivalent status of human naïve pluripotent stem cells, we downloaded the naive hPSC H3K4me3, H3K9me3 and H3K27me3 data from Wang et al. (2018) (GSE89072). The bivalent genes were defined as genes that with both H3K4me3 signal (greater than 60% quantile of all genes) and H3K27me3 signal (greater than 90% quantile of all genes) considering the signal distribution. In total, 1984 genes were defined as K4/K27 bivalent genes in human naïve hPSC and the list were included in the Table S3.

QUANTIFICATION AND STATISTICAL ANALYSIS

Error bars in the graphical data represent the standard deviation (SD). For all the presented boxplots, the center represents the median value, and the lower and upper lines represent the 5% and 95% quantiles, respectively. Significant difference between different groups was determined using the one-sided Wilcoxon rank-sum test adjusted by the FDR, and $p < 0.05$ was considered to be statistically significant. BS-seq and RNA-seq experiments were performed two to six times for each group. Histone modification CUT&RUN experiments were performed one to two times for each group, and the precise numbers of replicates and the data qualities were summarized in Table S1.