HHS Public Access

Author manuscript

*Cell Rep*. Author manuscript; available in PMC 2021 December 27.

Published in final edited form as:

*Cell Rep*. 2021 November 09; 37(6): 109982. doi:10.1016/j.celrep.2021.109982.

# Klf5 establishes bi-potential cell fate by dual regulation of ICM and TE specification genes

**Martin Kinisu**[1,7], **Yong Jin Choi**[1,7], **Claudia Cattoglio**[1,2], **Ke Liu**[3], **Hector Roux de Bezieux**[4], **Raeline Valbuena**[1], **Nicole Pum**[1], **Sandrine Dudoit**[4], **Haiyan Huang**[3], **Zhenyu Xuan**[5], **Sang Yong Kim**[6], **Lin He**[1,8,*]

[1]Division of Cellular and Developmental Biology, MCB Department, University of California, Berkeley, Berkeley, CA 94705, USA

[2]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA 94720, USA

[3]Department of Statistics, University of California, Berkeley, Berkeley, CA 94720, USA

[4]Division of Biostatistics, School of Public Health, University of California, Berkeley, Berkeley, CA 94720, USA

[5]Department of Molecular and Cell Biology, University of Texas at Dallas, 800 West Campbell Road, Richardson, TX 75080, USA

[6]Department of Pathology, NYU Grossman School of Medicine, 540 First Avenue, New York, NY 10016, USA

[7]These authors contributed equally

[8]Lead contact

## SUMMARY

Early blastomeres of mouse preimplantation embryos exhibit bi-potential cell fate, capable of generating both embryonic and extra-embryonic lineages in blastocysts. Here we identify three major two-cell-stage (2C)-specific endogenous retroviruses (ERVs) as the molecular hallmark of this bi-potential plasticity. Using the long terminal repeats (LTRs) of all three 2C-specific ERVs, we identify Krüppel-like factor 5 (Klf5) as their major upstream regulator. Klf5 is essential for bi-potential cell fate; a single *Klf5*-overexpressing embryonic stem cell (ESC) generates terminally



*Correspondence: lhe@berkeley.edu.

AUTHOR CONTRIBUTIONS

M.K. conceived, designed, and executed the majority of the experiments reported in Figures 1, 3, and 4. Y.J.C. performed most experiments reported in Figure 2. M.K., K.L., H.R.d.B., and Z.X. provided bioinformatic analyses on preimplantation single-cell RNA-seq analysis, motif enrichment analysis and visualization, ChIP-seq analysis and visualization, and ingenuity pathway enrichment analysis. C.C. prepared ChIP-seq samples and M.K. the MERVL-reporter. ESC data were generated by N.P. Teratoma experiments and analyses were done by Y.J.C., R.V., and M.K. Morula injections to generate chimeric blastocysts as well as all imaging and quantification were performed by S.Y.K., Y.J.C., and M.K. Depletion experiments by RNAi and CRISPR-EZ as well as downstream qPCR and immunofluorescence were done by M.K. Data mining of Cistrome DB was done by K.L. and M.K. L.H. supported the conception and design of experiments as well as manuscript writing. S.D. and H.H. proofread the prepared manuscript.

SUPPLEMENTAL INFORMATION

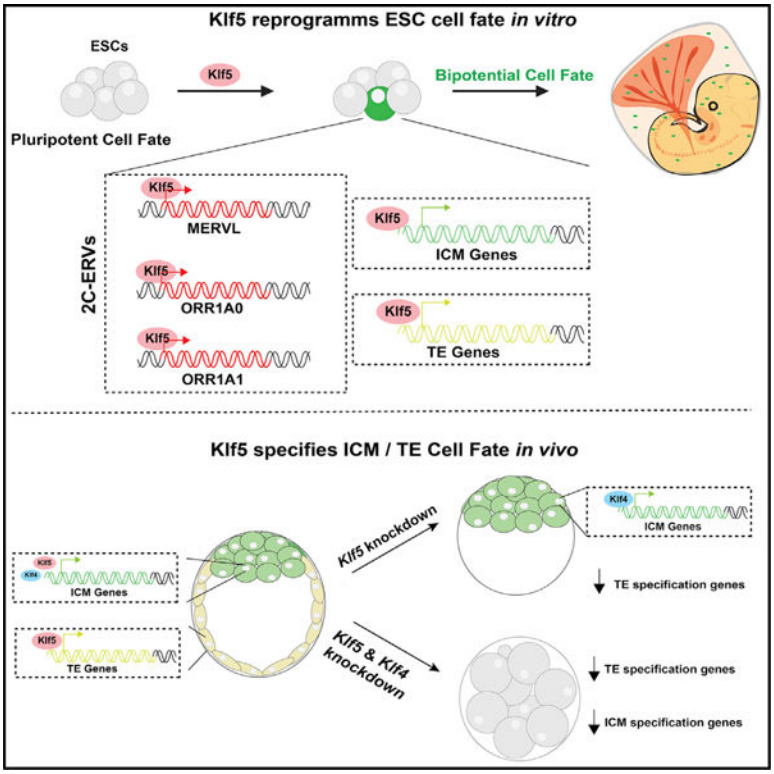
Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2021.109982.

DECLARATION OF INTERESTS

The authors declare no competing interests.

differentiated embryonic and extra-embryonic lineages in chimeric embryos, and Klf5 directly induces inner cell mass (ICM) and trophectoderm (TE) specification genes. Intriguingly, *Klf5* and *Klf4* act redundantly during ICM specification, whereas *Klf5* deficiency alone impairs TE specification. *Klf5* is regulated by multiple 2C-specific transcription factors, particularly Dux, and the Dux/Klf5 axis is evolutionarily conserved. The 2C-specific transcription program converges on Klf5 to establish bi-potential cell fate, enabling a cell state with dual activation of ICM and TE genes.

## Graphical abstract




## In brief

Using multiple 2C-specific ERV cell fate markers, Kinisu et al. identify Klf5 as a key transcription factor that confers a 2C-like developmental potential and activates ICM and TE specification genes. Klf5 and Klf4 act redundantly for ICM and TE specification in mouse preimplantation embryos.

## INTRODUCTION

Mammalian preimplantation development is initiated by maternally inherited factors and zygotic genes transcribed during zygotic genome activation (ZGA) (Deng et al., 2014). Mouse zygotes and two-cell-stage (2C) blastomeres are totipotent, capable of generating all cell types required for a fertile adult organism (Casser et al., 2017). Totipotency is gradually restricted in subsequent developmental stages (4C to 8C stage; Wu et al., 2017),

but cleavage-stage blastomeres retain bi-potential cell fate, generating the inner cell mass (ICM), which largely forms the embryo proper, and trophectoderm (TE), which gives rise to extra-embryonic placental tissues (Korotkevich et al., 2017; Fujimori et al., 2003; Tabansky et al., 2013; Wigger et al., 2017).

A prominent molecular hallmark of bi-potential cell fate is a strong but transient induction of MERVL endogenous retroviruses (ERVs) (Choi et al., 2017; Macfarlan et al., 2011, 2012; Ishiuchi et al., 2015). MERVL transcripts are among the most highly expressed transcripts in the transcriptomes of 2C–4C blastomeres (Franke et al., 2017); they quickly decrease in level as the developmental plasticity of cleavage-stage blastomeres narrows during development. In pluripotent mouse embryonic stem cells (ESCs), which generate all embryonic cell types but rarely extra-embryonic lineages (Beddington and Robertson, 1989), MERVL induction in rare cell populations often correlates with expanded cell fate plasticity, enabling differentiation toward embryonic and extra-embryonic lineages (Choi et al., 2017; Ishiuchi et al., 2015; Macfarlan et al., 2011, 2012; Schoorlemmer et al., 2014; Zhao et al., 2018; Hu et al., 2020; Yan et al., 2019). However, such MERVL$^+$ ESCs are not equivalent to 2C blastomeres (Choi et al., 2017), possessing neither totipotent cell fate potential nor a 2C transcriptome. Rather, MERVL$^+$ ESCs, designated bi-potential ESCs in our study, exhibit both embryonic and extra-embryonic potency, induce MERVL at a modest level, and functionally resemble bi-potential blastomeres, which are more restricted in developmental potential than 2C blastomeres.

Because MERVL is a molecular hallmark of 2C blastomeres, transcription factors that directly regulate MERVL and transiently peak at ZGA have been speculated to establish a transcriptional landscape that enables bi-potential cell fate (Alda-Catalinas et al., 2020; Eckersley-Maslin et al., 2019; Hendrickson et al., 2017; De Iaco et al., 2017; Yan et al., 2019). The double homeo-domain transcription factor Dux is one such candidate, induced at the onset of ZGA to directly promote MERVL expression in 2C blastomeres (Eckersley-Maslin et al., 2019; Hendrickson et al., 2017; De Iaco et al., 2017). Zygotically expressed Dux as well as its maternally inherited upstream regulators Dppa2, Dppa4, Nelfa, and Smarca5 have been speculated to act at the top of the transcriptional hierarchy that governs onset of ZGA, induction of MERVL, and regulation of 2C-specific cell fate potency (Alda-Catalinas et al., 2020; Eckersley-Maslin et al., 2019; Hu et al., 2020). However, deficiency of *Dux*, *Smarca5*, or *Nelfa* alone or *Dppa2* and *Dppa4* in combination, fails to impair ICM or TE specification in mice, suggesting that these factors are not essential to establish/maintain bi-potential cell fate (Chen and Zhang, 2019; Nakamura et al., 2011; Koscielny et al., 2014; Stopka and Skoultchi, 2003). Hence, using MERVL as the sole molecular marker for bi-potential cell fate may not be sufficient to identify the key regulator(s) for bi-potential developmental cell fate.

Using three 2C-specific ERVs as the hallmark of bi-potential cell fate, we identified Krüppel-like factor (Klf5) as an essential regulator that confers developmental potency to embryonic and extra-embryonic lineages. Although previous studies have characterized the preimplantation defects of *Klf5* knockout embryos (Azami et al., 2017; Ema et al., 2008; Lin et al., 2010), the functional importance of Klf5 in bi-potential cell fate and the functional interaction between Klf5 and other Klf transcription factors in the context of preimplantation

development remain largely unknown. Our study shows that *Klf5* overexpression in a single, pluripotent ESC confers bi-potential cell fate in chimeric embryos. Our mouse genetics studies demonstrated the essential role of *Klf5* in enabling TE specification and the redundant role of *Klf5* and *Klf4* in conferring potency for the ICM cell fate. Because Klf5 directly induces ICM and TE specification genes (Lin et al., 2010), our data suggest that the molecular nature of bi-potential developmental potency is a cell state that co-expresses ICM and TE genes.

## RESULTS

### Identification of 2C-specific ERV families

We set out to comprehensively identify 2C-specific retrotransposon markers in search of a master transcriptional regulator of bi-potential cell fate (Figure 1A). Using published single-cell RNA sequencing (RNA-seq) data on preimplantation embryos, we observed a dynamic and tightly regulated expression pattern of retrotransposons (Figure 1B; Figure S1A). Although MERVL induction is a prominent molecular hallmark of 2C blastomeres, two additional ERVs, ORR1A0 and ORR1A1, also constitute the major retrotransposon families with a 2C-specific expression pattern (Figures 1B and 1C; Figure S1A). At the peak of their expression, MERVL, ORR1A0, and ORR1A1 each account for 5.3%, 1.5%, and 1.1% of all mapped reads in 2C blastomeres, respectively (Figure S1A). Subsequently, the expression of these ERV families declined rapidly by the 8C stage and was silenced completely in blastocysts (Figures 1B and 1C; Figure S1A). Their coordinated induction was confirmed in a subset of bi-potential, MERVL$^{+}$ ESC lines (Figure 1D). Specifically, *Lsd1* deletion and knockdown of the Caf-1 subunits *P60* and *P150* in ESCs yield coordinated induction of these three 2C-specific ERVs (Figure 1D). Because MERVL, ORR1A0, and ORR1A1 collectively marked the transcriptional state of 2C blastomeres and bi-potential ESCs, we hypothesized that a transcription factor(s) capable of inducing all three 2C-specific ERVs could functionally confer 2C-like, bi-potential cell fate.

The transcriptional regulatory sequences of ERVs are contained within their 5′ long terminal repeats (LTRs), often ~300– 500 bp in length (McCarthy and McDonald, 2004). Alignment of the consensus LTRs of the three 2C-specific ERVs revealed homology between ORR1A0 and ORR1A1 (96% identity), which shared 42% and 41% sequence identity, respectively, with MERVL LTR (Figure S1B). This finding suggests coordinated and distinct transcriptional regulation of these 2C-specific ERVs.

### Identification of *Klf5* as a driver of 2C/4C-ERVs

To identify putative upstream regulators of this 2C cohort, we employed transcription factor motif enrichment analysis using all annotated LTR elements for each ERV family in the DFAM repository (Hubley et al., 2016), (Benner et al., 2017). A number of experimentally validated MERVL regulators emerged from these analyses, such as Dux, Gata2, retinoic acid receptor alpha (Rara), Zfp281, and Tbx-family factors, indicating the power of this approach (Choi et al., 2017; Dai et al., 2017; Dan et al., 2013; Hendrickson et al., 2017; Tagliaferri et al., 2020). Interestingly, motifs for Klf transcription factors were the most enriched in the LTRs of MERVL, ORR1A0, and ORR1A1 (Figure 1E; Table S1), each of

which harbors four predicted Klf binding motifs. In contrast, Dux and Gata2 exhibit binding motif enrichment only within the LTRs of MERVL LTRs but not ORR1A0/ORR1A1 (Figure 1E; Table S1). Consistently, in published ESC chromatin immunoprecipitation sequencing (ChIP-seq) data (Hendrickson et al., 2017), no direct Dux binding to ORR1A0 or ORR1A1 elements was detected (Figure S1C).

Klfs are evolutionarily conserved zinc-finger transcription factors that have a pivotal role in embryonic development and pluripotent stem cells (Presnell et al., 2015). The mouse genome encodes 17 annotated Klf transcription factors, all of which possess a highly conserved C-terminal DNA binding domain that recognizes guanine-cytosine rich regions and CACCGT box motifs (Presnell et al., 2015). *Klf5*, *Klf4*, and *Klf17* are the most highly expressed Klf transcription factors in mouse preimplantation embryos (Figure 1F; Figure S1D). Although several Klf factors, including Klf4, Klf5, and Klf2, have been shown to function redundantly to sustain pluripotency in ESCs (Yamane et al., 2018), Klf5 is the only factor whose deficiency in mouse embryos impairs preimplantation cell fate decisions (Azami et al., 2017; Ema et al., 2008; Lin et al., 2010; Presnell et al., 2015). In comparison, *Klf4* or *Klf2* individual knockout yield no obvious preimplantation defects (Katz et al., 2002).

To determine the Klf(s) capable of directly regulating 2C-specific ERVs and, possibly, bi-potent cell fate, we constructed luciferase (Luc) reporters driven by the MERVL LTR (MERVL-Luc), which faithfully recapitulates MERVL expression (Choi et al., 2017; Macfarlan et al., 2011). Although Klf5 was able to activate MERVL-Luc in HEK cells, neither Klf4 nor Klf17 had a similar effect (Figure 1G). The MERVL LTR contains four predicted Klf5 binding motifs, and transversion mutations of these sites abolished Klf5-dependent luciferase induction (Figure 1G). To ascertain whether Klf5 directly mediated MERVL-Luc induction, we performed ChIP using *Klf5*-overexpressing ESCs and confirmed specific Klf5 occupancy on the LTRs of MERVL, ORR1A0, and ORR1A1 (Figure 1H). Consistently, *Klf5* overexpression specifically upregulated MERVL, ORR1A0, and ORR1A1, along with 2C-specific, MERVL-driven 2C gene isoforms (Figures 1I and 1J). No induction of other retrotransposon families, such as IAP, LINE1, and SINE B1, was observed (Figure 1I). *Klf5* overexpression in ESCs also activated a tdtomato reporter driven by the MERVL LTR (Figure S1F). MERVL induction by Klf5 in ESCs was further confirmed by immunostaining for MERVL-Gag protein, with 15% of ESC colonies containing MERVL $Gag^+$ cells and 5%–8% MERVL $Gag^+$ cells overall (Figure 1K). In comparison, control ESCs contained less than 0.1% $MERVL^+$ cells (Figure 1K), indicating that Klf5 overexpression shifted the equilibrium in ESC culture in favor of metastable $MERVL^+$ cells. Similar to other reported $MERVL^+$ ESCs (Choi et al., 2017; Hendrickson et al., 2017; Macfarlan et al., 2012), overexpression of *Klf5* in ESCs activated MERVL at the expense of Oct4 protein expression because MERVL-Gag and Oct expression was mutually exclusive (Figure 1K). Importantly, *Klf5* overexpression levels in ESCs were comparable with the level of *Klf5* in blastocysts within a physiologically relevant range (Figure S1G), and RNA-seq in *Klf5*-overexpressing ESCs further confirmed the ability of Klf5 to upregulate 2C-specific genes (Figure S1H). Our data establish Klf5 as a direct regulator of all three 2C-specific ERVs: MERVL, ORR1A0, and ORR1A1.

### *Klf5* confers bi-potential cell fate *in vitro* and *in vivo*

Because Klf5 induces MERVL, ORR1A0, and ORR1A1 ERVs, which collectively marked bi-potent blastomeres and ESC populations (Figures 1B–1D), we hypothesized that *Klf5* could functionally confer bi-potential cell fate in ESCs. In line with this hypothesis, teratomas generated from *Klf5*-overexpressing ESCs contained cells expressing markers of TE (*Cdx2* and *Elf5*), primitive endoderm (PrE) (*Gata4*, *Gata6*, and *Sox17*), and all three embryonic germ layers (*Pax6*, *Brachyury*, and *Foxa2*) (Figures 2A and 2B; Figures S2A and S2B). In comparison, control teratomas only induced molecular markers of embryonic lineages (Figure 2B). In particular, we identified teratoma cells reminiscent of placental trophoblast giant cells, with strong PL-1 (placental lactogen 1) expression, large cell volume, enlarged nuclei, and proximity to internal hemorrhages (Figure 2A). Similarly, embryoid bodies (EBs) generated from *Klf5*-overexpressing ESCs, but not control ESCs, showed induction of extra-embryonic markers of the TE, PrE, and placental trophoblast lineages (Figure 2C; Figure S2C).

In teratomas and EBs, multiple MERVL$^+$ cells collectively contribute to embryonic and extra-embryonic cell types, making it unclear whether MERVL$^+$ cells have *bona fide* bi-potential cell fate. The bi-potential cell fate of early blastomeres is strictly defined by the capability of a single cell to contribute to both embryonic and extra-embryonic lineages (Wigger et al., 2017; Wu et al., 2017). Hence, we microinjected single, GFP-labeled control or *Klf5*-overexpressing ESCs into C57BL/6J recipient 8C embryo and analyzed the resulting chimeric blastocysts (Figure 2D). Although control ESCs invariably contributed to the ICM of chimeric blastocysts (Figure S2D), individual *Klf5*-overexpressing ESCs colonized the ICM, TE, or both in chimeric embryos (Figures 2D and 2E). Particularly, nearly a third of chimeric blastocysts contained GFP$^+$ progenies from a single *Klf5*-overexpressing ESC in both ICM and TE (Figures 2D and 2E). Intriguingly, GFP$^+$ ICM cells derived from *Klf5*-overexpressing ESCs strongly express Nanog in chimeric blastocysts, similar to their neighboring ICM cells. However, GFP$^+$ TE cells exhibited an intermediate phenotype: they localized to the TE compartment, took on a typical TE morphology, and silenced Nanog expression but failed to robustly express Cdx2 as compared to neighboring TE cells (Figure S2E). This is likely due to the delayed kinetics of *Cdx2* activation upon withdrawl of LIF from the culture medium. TE cells derived from *Klf5*-overexpressing ESCs clearly possess extra-embryonic differentiation capacity but have a reduced efficiency for extra-embryonic differentiation compared to normal TE cells.

We next generated embryonic day 12.5 (E12.5) chimeric embryos by injecting single *Klf5*-overexpressing ESCs into recipient blastocysts. These E12.5 chimeric embryos contained ESC-derived cell lineages in both the embryo proper and extra-embryonic placental and yolk sac lineages (Figures 2F–2H; Figures S2F; Table S2). *Klf5*-overexpressing ESCs gave rise to terminally differentiated trophoblast giant cells, spongiotrophoblasts of the placenta, as well as the yolk sac visceral endoderm (Figures 2F and 2G). In addition, GFP$^+$ cells with a large cytoplasmic to nuclear ratio were found proximal to Mtp1$^+$ syncytiotrophoblast II (SynII) cells, which morphologically resemble sinusoidal trophoblast giant cells (s-TGCs) (Figure 2F). Our findings suggest that, in each E12.5 chimeric embryo, a single injected *Klf5*-overexpressing ESC underwent substantial proliferation prior to terminal lineage

commitment during normal development. E12.5 chimeric embryos generated by injecting 10–15 *Klf5*-overexpressing ESCs yielded similar results, generating terminally differentiated embryonic lineages as well as extra-embryonic placental and yolk sac lineages (Figures S2G and S2H). In all of our Klf5 overexpression studies, *Klf5*-overexpressing ESCs have a level of *Klf5* expression comparable with that of blastocysts (Figure S1G), suggesting that the *Klf5* overexpression phenotype we observed in ESCs is caused by a physiologically relevant level of *Klf5* expression.

### *Klf5* regulates both ICM and TE specification genes

To understand the molecular basis of Klf5-induced bi-potential cell fate in ESCs, we performed ChIP-seq using *Klf5*-overexpressing ESCs and applied Ingenuity Pathway Analysis (IPA) on Klf5-bound genes (Figure 3A). Interestingly, Klf5 occupancy was enriched significantly for genes regulating ESC pluripotency, Hippo signaling, and blastocyst development (Figure 3A; Table S3), many of which promoted cell fate specification of the ICM or TE and displayed dynamic expression in cleavage-stage embryos (Figure 3B). Specifically, ICM-specific transcription factors (*Nanog* and *Klf4*) and TE-specific transcription factors (*Tead4* and *Cdx2*) exhibited Klf5 binding in their putative promoter/enhancer regions (Figure 3C); the corresponding Klf5 peaks invariably harbored multiple predicted Klf binding motifs (Figure S3A). Additionally, Klf5 peaks were also observed in 2C ERV LTRs that drove 2C-specific gene isoforms (Figure S3B).

Although Klf5 has the ability to induce ICM and TE genes based on our ChIP-seq data, Klf5 mediates gene transcription in a cell-type- and context-specific manner. In pluripotent ESCs, *Klf5* overexpression induced multiple pluripotency transcription factors, including *Nanog*, *Klf4*, and *Esrrb* (Figure 3D), but had little effect on TE genes, possibly because of strong epigenetic silencing of TE genes in standard ESC cultures (Niwa et al., 2005). In differentiating EBs derived from *Klf5*-overexpressing ESCs, we observed distinct $Nanog^+$ or $Cdx2^+$ cell populations that were absent in control EBs (Figure 3E). The strong induction of the extra-embryonic TE markers *Cdx2* and *Elf5* was confirmed by real-time PCR (Figure 3E). Hence, *Klf5*-overexpressing ESCs yielded progenies expressing ICM or TE genes within an EB (Figure 3E), consistent with the bi-potential developmental plasticity of Klf5-expressing ESCs.

Prior to cell fate specification during the morula-to-blastocyst transition, modest co-expression of ICM- and TE-specific transcription factors in early blastomeres is essential for establishing a bi-potential cell fate (Hirate et al., 2013; Korotkevich et al., 2017; Strumpf et al., 2005; Pfeffer, 2018). These early blastomeres provide a unique cellular context where ICM and TE genes can be co-expressed at a modest level. Dual induction of ICM and TE specification genes by Klf5 likely plays a role in conferring a cell state that allows differentiation into ICM or TE lineages.

*Klf5* expression initiates in bi-potent blastomeres of cleavage-stage embryos (Lin et al., 2010; Figure 1F). Subsequently, Klf5 exhibits ICM and TE expression in blastocysts, albeit with stronger TE enrichment (Figure S3C). To investigate the role of *Klf5* in cell fate potency during preimplantation development, we efficiently knocked down *Klf5* using RNA interference (RNAi) in zygotes (Figure S3D). Morphologically, *Klf5* knockdown

embryos appeared normal until the morula stage; obvious defects arose with high penetrance during blastocoel formation. By E4.5, control embryos had typical blastocyst morphology, whereas *Klf5* knockdown embryos failed to robustly form a blastocoel cavity (Figure 3F). Immunofluorescence staining revealed a marked reduction of Cdx2 in E4.5 *Klf5* knockdown embryos (Figure 3G), consistent with a significant decrease in major TE specification genes (*Cdx2*, *Elf5*, and *Tead4*) and a placenta early development gene (*Esrrb*), as shown by real-time PCR in control versus knockdown embryos (Figure S3D). In contrast, *Klf5* knockdown alone failed to affect ICM specification genes or pluripotency genes *in vivo* (*Sox2*, *Oct4*, and *Nanog*) (Figure S3D). RNA-seq in control and *Klf5* knockdown embryos further confirmed the different effects of *Klf5* on TE versus ICM *in vivo*. *Klf5* knockdown significantly decreased the level of TE specification genes in E4.5 blastocysts while leaving the ICM specification genes largely unperturbed (Figure 3H).

To confirm these findings, we generated blastocyst embryos with complete *Klf5* disruption using CRISPR-mediated *Klf5* editing (Figure S3E). In *Klf5* knockout embryos, we detected significant loss of Cdx2, but Oct4 expression was relatively intact (Figure S3E). Hence, although *Klf5* has the ability to induce ICM and TE specification genes, *Klf5* deficiency preferentially impairs TE cell fate, consistent with strong Klf5 enrichment in the TE compartment (Figure S3C). Our findings seemingly differ from a previous study that described impaired ICM and TE specification caused by *Klf5* deficiency (Lin et al., 2010). Nevertheless, a closer examination of their data indicates that the TE defects are the predominant preimplantation phenotype in *Klf5* knockout embryos and that the ICM defects are rather mild with incomplete penetrance (Azami et al., 2017; Lin et al., 2010). Hence, although Klf5 can induce ICM and TE genes, it is essential for TE specification but not for ICM specification.

Given the ability of *Klf5* to co-induce ICM and TE genes, the lack of an obvious ICM defect in Klf5-deficient embryos suggests that additional Klf transcription factor(s) could act redundantly to specify ICM. Interestingly, Klf5 is expressed in ICM and TE, with strong TE enrichment; *Klf4* is specifically enriched in the ICM, and its ICM expression level is comparable with that of *Klf5* (Figure 3I). Consistent with this expression pattern, *Klf4* knockout in mice exhibits no obvious defects in preimplantation development (Katz et al., 2002), but knocking down *Klf5* and *Klf4* impairs ICM and TE cell fate during the morula-to-blastocyst transition, as shown by defective Nanog and Cdx2 expression (Figure 3J; Figure S3F). Interestingly, *Klf5* knockdown failed to alter Yap-1 staining in embryos, suggesting that Hippo signaling is not regulated by *Klf5* during preimplantation development (Figure S3G). This is consistent with Klf5 enabling cell fate potency for the ICM and TE lineages rather than specifying the TE lineages. Klf transcription factors constitute a robust transcriptional network for bi-potential cell fate, with Klf5 being an essential regulator for enabling TE specification and Klf4 and Klf5 being functionally redundant during ICM specification.

Interestingly, ICM expression of Klf5 is observed in the epiblast as well as the PrE lineage in E4.0 blastocysts because Klf5 and Sox17 are co-expressed in cells of the PrE compartment (Figure S3H). Real-time PCR analyses of *Klf5* knockdown embryos also exhibited marked downregulation of the PrE specification genes *Gata6*, *Gata4*, and

*Sox17* (Figure 3H). Although this is consistent with induction of PrE gene markers in *Klf5*-overexpressing EBs and teratomas (Figure 2B), the decreased PrE gene expression in *Klf5* knockdown embryos can also be caused by their developmental arrest/delay. Our data are consistent with Klf5 acting upstream of induction of PrE genes. Our findings contrast a previous study that reported Klf5 as a suppressor of PrE specification (Azami et al., 2017). This study showed an increased percentage of PrE cells in *Klf5* knockout blastocyst embryos. However, the significant decrease of the total cell number and the developmental arrest of *Klf5* knockout blastocysts have complicated interpretation of this result, making it difficult to identify Klf5 as a suppressor of PrE cell fate (Azami et al., 2017). In addition, the similar expression of *Klf5* in PrE and epiblast cells is inconsistent with a suppressive role of *Klf5* in the PrE cell fate decision. It is clear that the strongest and most direct phenotype of Klf5-deficient embryos is impaired TE specification.

### *Klf5* is regulated by multiple 2C transcription factors

Using all ESC ChIP-seq data in the Cistrome database (Cistrome DB) (Mei et al., 2017), we identified a number of transcription regulators with enriched occupancy proximal to *Klf5* (Figure 4A; Table S4). These putative Klf5 regulators were subjected to IPA, and an enrichment emerged for genes regulating stem cell biology and preimplantation development (Figure S4A). In particular, multiple 2C-specific transcription factors, including Dux, Dppa2, and Tbx3, bound to putative Klf5 regulatory regions, either a region immediately upstream of the transcription start site (TSS) or a region within the intron 1 (Figure 4A). In particular, Dux, the key regulator for the 2C-like transcriptome (Eckersley-Maslin et al., 2019; Hendrickson et al., 2017; Hu et al., 2020; De Iaco et al., 2017; Ishiuchi et al., 2015), showed a strong ChIP-seq peak in *Klf5* intron 1 (Figure 4A). Using published RNA-seq data on *Dux*-overexpressing ESCs (Hendrickson et al., 2017), we showed that *Dux* overexpression upregulates *Klf5* (Figure 4B). In addition to *Dux*, *Dppa2* also induced *Klf5* expression and MERVL expression in ESCs (Figure 4B), and this induction was independent of Dux. Finally, *Tbx3* overexpression in ESCs also induced *Klf5* (Figure 4B). Hence, multiple 2C-specific transcription factors converge their direct regulation on induction of *Klf5*.

In most bi-potential ESCs with expanded cell fate potential, an aberrant increase in *Dux* constitutes the key mechanism underlying the MERVL induction (Choi et al., 2017; Hu et al., 2020; Macfarlan et al., 2012; Yan et al., 2019). In contrast, Klf5-induced MERVL induction and bi-potential cell fate in ESCs likely acts downstream of *Dux*. *Klf5* overexpression invariably failed to induce *Dux* (data not shown); MERVL induction by *Klf5* was preserved even in a *Dux* knockout background (Figure 4C). More importantly, EBs derived from *Klf5*-overexpressing $Dux^{-/-}$ ESCs, but not control $Dux^{-/-}$ ESCs, induced markers for extra-embryonic lineages (the TE marker *Cdx2* and PrE marker *Gata4*) and embryonic lineages (the ectoderm marker *Pax6*, mesoderm marker *Brachyury*, and endoderm marker *Foxa2*) (Figure 4D; Figure S4C). A single *Klf5*-overexpressing $Dux^{-/-}$ ESC was able to colonize the ICM, TE, or both in 44%, 22%, and 33% of chimeric blastocysts, respectively (Figure 4E). However, it is possible that other interactions between *Dux* and *Klf5* exist because the level of MERVL induction and induction of *Cdx2*

upon EB differentiation were somewhat dampened in comparison with wild-type ESCs overexpressing *Klf5.*

The functional importance of Klf5 and its regulation by Dux prompted us to investigate the evolutionary conservation of the Dux/Klf5 axis between mouse and human. Mouse *Dux* and human *DUX4* are induced at the onset of ZGA to govern induction of early zygotic genes (Hendrickson et al., 2017; De Iaco et al., 2017). We then investigated to what extent Dux regulation on *Klf5* was conserved between mouse and human. We analyzed published RNA-seq and ChIP-seq data for *DUX4*-overexpressing human ESCs (hESCs) (Hendrickson et al., 2017). Intriguingly, human *KLF5* was upregulated significantly by *DUX4* in the RNA-seq analysis, but a subset of bi-potency regulators and known Dux targets in the mouse, such as *NELFA* and *DPPA2*, were not affected by enforced *DUX4* expression (Figure 4F). This is consistent with ChIP-seq data because DUX4 demonstrated enriched occupancy in multiple regulatory regions proximal to the *KLF5* transcriptional start site, whereas no DUX4 occupancy could be detected near *NELFA* and *DPPA2* (Figure 4G). The evolutionary conservation of the Dux/Klf5 axis, but not the Dux/Dppa2 or Dux/Nelfa axis, highlights the functional importance of Klf5 in establishing a bi-potential cell fate in early blastomeres.

Although Klf transcription factors are essential for bi-potential cell fate, many 2C-specific transcription factors likely act redundantly, and the individual disruption of *Dux*, *Dppa2/4*, *Tbx3*, and *Nelfa* in knockout mice failed to yield a preimplantation phenotype (Koscielny et al., 2014; Chen and Zhang, 2019; Nakamura et al., 2011). This contrasts the strong ICM and TE specification defects when *Klf5* and *Klf4* are knocked down (Figures 3I and 3J). Hence, it is likely that multiple 2C-specific transcription factors converge on Klf5 via evolutionarily conserved mechanisms, which promotes major 2C-specific ERV expression and establishes bi-potential cell fate (Figure 4H).

## DISCUSSION

Bi-potential blastomeres exhibit an intrinsic transcriptional program co-expressing ICM and TE genes at a modest level; subsequent extrinsic signaling then acts as the decisive factor to promote one lineage while repressing the other (Pfeffer, 2018). Hence, the molecular nature of the bi-potential plasticity is likely a cell state with the developmental plasticity to activate ICM or TE specification genes in a context-dependent manner. To elucidate the key molecular pathway regulating the bi-potent cell fate, we identified three major ERV families as the molecular hallmarks for bi-potential blastomeres. The LTR sequences of all three 2C ERVs likely share critical regulatory sequences for key bi-potential transcription factor(s). We performed motif analyses on these LTRs and identified Klf5 as an important regulator for 2C-specific ERVs and 2C-like bi-potential cell fate.

Upon overexpression, Klf5 establishes a robust bi-potential cell fate in single ESCs to yield multiple terminally differentiated embryonic and extra-embryonic lineages in chimeric mid-gestation embryos. This effect of *Klf5* is highly dependent on its expression threshold because endogenous *Klf5* expression in pluripotent ESCs is not sufficient to confer bi-potential cell fate but an elevated expression Klf5 level is. Although our study lacks a systematic comparison of genes bound by *Klf5* at endogenous levels, we speculate that

the expansion of ESC potency by *Klf5* is mediated by novel targets or targets bound more strongly, following overexpression. Additionally, the functional importance of 2C ERV expression with respect to the expansion of ESC cell potency is ambiguous and remains a limitation of this study. In this study, we treat the 2C ERV cohort as cellular markers. Further work screening *Klf5* targets during ESC differentiation would comprehensively clarify genes downstream of *Klf5* involved in bi-potential cell fate. Such a screen would also highlight 2C ERV-regulated genes that are essential for bi-potential cell fate. The ability of Klf5 to establish a bi-potential cell fate in ESCs is consistent with its functional importance in preimplantation cell fate decisions. *Klf5* is expressed in ICM and TE, with strong enrichment in TE. *Klf5* deficiency alone impairs TE specification, and *Klf5* and *Klf4* deficiency in combination impairs TE and ICM specification (Figure 4H).

Although functional redundancy exists among preimplantation-specific Klf transcription factors, Klf5 is the most important Klf for bi-potential cell fate. Klf5 is the only known Klf factor expressed in ICM and TE, with a knockout phenotype impairing blastocyst cell fate specification (Azami et al., 2017; Ema et al., 2008). Klf5 alone is necessary and sufficient to promote TE specification, whereas Klf5 acts redundantly with other Klf factors, such as Klf4, to promote ICM specification in preimplantation embryos. Hence, Klf5 acts at the core of a robust Klf transcription network, which promotes dual induction of ICM and TE specification genes to establish bi-potential developmental plasticity in cleavage-stage blastomeres.

Findings from our *Klf5* loss-of-function studies are seemingly different from the conclusion of a previous study, which reported defects in ICM and TE cell fate decisions in *Klf5* knockout embryos (Ema et al., 2008). Nevertheless, the actual data in this study as well as those from a later study (Lin et al., 2010) clearly indicated that *Klf5* knockout caused a strong and fully penetrant TE defect but a mild, lowly penetrant ICM defect. The small difference in the extent of ICM defects in different *Klf5* loss-of-function studies can be attributed to the different genetic backgrounds. These observations are intriguing because Klf5 induces ICM and TE genes, but its deficiency preferentially affects the TE cell fate. As is clear from our studies, Klf5 acts alone to promote the TE cell fate, and Klf5 functions redundantly with Klf4 to regulate ICM cell fate.

Klf5 induction is regulated by multiple transcription factors that regulate the 2C-specific transcriptome, including Dux, Dppa2, and Tbx3. Unlike *Klf5*, none of these 2C transcription factors exhibit a strong preimplantation phenotype in knockout studies (Koscielny et al., 2014; Chen and Zhang, 2019; Nakamura et al., 2011). It is likely that they act redundantly in promoting bi-potential cell fate and that their collective regulation of developmental plasticity converges on Klf5 and possibly other Klf genes. In particular, Dux regulation of Klf5 is evolutionarily conserved between mouse and human, but a number of well-characterized, 2C-specific Dux targets in the mouse are not regulated by human *DUX4*. Our findings suggest an evolutionarily conserved functional importance of the *Dux-Klf5* axis in regulating the developmental plasticity of early blastomeres in mammals.

Bi-potential cell fate is a complex biological state that may not be always associated with 2C ERV induction. Expanded pluripotent stem cells (EPCs), derived from mouse 8C

blastomeres or from treatment of a chemical cocktail, yield embryonic and extra-embryonic potency in chimeric embryos without inducing MERVL (Yang et al., 2017a, 2017b). Future studies will likely identify additional pathways acting in parallel to Klf5 to promote this developmental plasticity.

### Limitations of the study

The present article describes the role of the transcription factor *Klf5* in conferring bi-potential cell fate, not only to otherwise pluripotent ESCs but also during preimplantation development in the mouse. The investigation is, however, limited in that we have not been able to fully characterize the functional role of 2C ERVs during acquisition and maintenance of bi-potential cell fate. We utilized the cohort of ORR1A0, ORR1A1, and MERVL, all of which are most highly expressed from the 2C–4C stages and in documented ESCs with expanded cell fate potential (Figures 1B–1D), and this cohort guided us in identifying *Klf5* as an upstream regulator of this cohort and, by proxy, expanded cell fate potential. However, we have not identified genes that may be regulated by these ERVs and facilitate bi-potential cell fate acquisition. This is in part due to the various ways in which ERVs can affect host gene expression from behaving as enhancers (Judd et al., 2021), alternative promoters (Modzelewski et al., 2018), and even regulating chromatin architecture (Kruse et al., 2019). This ambiguity is exacerbated by the observation that ERV-derived RNA can also regulate cellular identity (Lu et al., 2014).

Additionally, we showed that *Klf5* confers bi-potential cell fate upon overexpression in ESCs despite there already being expression of *Klf5* in ESCs. We also conducted our ChIP-seq analyses on overexpressed *Klf5* and not at endogenous levels. Although this is consistent with the observation that only elevated levels of *Klf5* are sufficient to confer bi-potential cell fate, a systematic comparison of *Klf5* targets between endogenous expression levels and overexpressed levels of *Klf5* may reveal downstream targets that are essential for expansion of cell fate by *Klf5.* We speculate that transcription factor thresholds are critical for expansion of ESC cell fate and activation of 2C ERVs because other regulators of this state, such as Dppa2/Dppa4, also need to be overexpressed to activate MERVL despite already being expressed in ESCs (Eckersley-Maslin et al., 2019; Watabe, 2012).

## STAR★METHODS

### RESOURCE AVAILABILITY

**Lead contact—**Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Lin He (lhe@berkeley.edu)

**Materials availability—**Plasmids generated in this study will be made available and deposited to AddGene.

### Data and code availability

- The ChIP-seq and RNA-seq data generated for this publication have been deposited in NCBI’s Gene Expression Omnibus and are accessible through GEO:

GSE137036 (ChIP-seq) and GEO: GSE186005 (RNA-seq). Accession codes of published data-sets used in this study are available in the Key resources table.

- This paper reports no original code.
- Any additional information required to reanalyze the data reported in this work paper is available from the Lead Contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

**Animals—**The embryo experiments reported in this paper involved the use of C57BL/6J or C57BL/6J C3H/HeJ F1 female mice all aged approximately 4 weeks. All mouse studies have appropriate authorizations acquired from institutional and federal regulatory agencies prior to the beginning of any experiment. Our animal care and use protocol (AUP-2015–04-7485–1) has been reviewed and approved by our IACUC for this project. All mouse usage is in accordance with the Animal Welfare Act, the AVMA Guidelines on Euthanasia and are in compliance with the ILAR *Guide for Care and Use of Laboratory Animals* and the UC-Berkeley IACUC.

**Cell lines—**Cell lines used in this study were generated from ESCs derived from wild-type C57BL/6J mice cultured on irradiated MEFs derived also from C57BL/6J mice at E13.5. ESCs were established and cultured as described in the Method details section.

## METHOD DETAILS

**RNA-seq preparation—**500 ng to 1 μg of purified RNA were added of UltraPure Water (ThermoFisher Scientific # 10977–023) to bring the total volume to 50 ml. mRNA purification, RNA fragmentation, first and second strand cDNA synthesis were performed according to the TruSeq RNA Sample Preparation v2 Kit using Superscript III for reverse transcription (incubation: 50°C for 50 minutes). cDNA was purified with AMPure XP beads diluted 1:2 with 20% PEG, 1.25M NaCl, and eluted in 27.5 μL 10 mM Tris-HCl pH 7.9, 25 μL of which were used for the library preparation, with the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB E7645) following manufacturer instructions with a few modifications. The recommended reagent volumes were cut in half. The NEBNext Adaptor for Illumina was diluted 1:5 in Tris/NaCl, pH 8.0 (10 mM Tris-HCl pH 7.9, 10 mM NaCl) and the ligation step extended to 30 min. After ligation, a single purification step with 0.9X volumes of Agencourt AMPure XP PCR purification beads (Beckman Coulter A63880) was performed, eluting DNA in 22 μL of 10 mM Tris-HCl pH 7.9. 20 μL of the eluted DNA was used for the library enrichment step, performed with the KAPA HotStart PCR kit (Roche Diagnostics KK2502) in 50 μL of total reaction volume (10 μL 5X KAPA buffer, 1.5 μL 10 mM dNTPs, 0.5 μL 10 μM NEB Universal PCR primer, 0.5 μL 10 μM NEB index primer, 1 μL KAPA polymerase, 16.5 μL nuclease-free water and 20 μL sample). Samples were enriched with 8 PCR cycles (98°C, 45 s; [98°C, 15 s; 60°C, 10 s] x 8; 72°C, 1 min; 4°C, hold), purified with 0.9 volumes of AMPure XP PCR purification beads and eluted with 33 μL of 10 mM Tris-HCl pH 7.9. Library concentration, quality and fragment size were assessed by Qubit fluorometric quantification (Qubit dsDNA HS Assay Kit, Invitrogen Q32851), qPCR and Fragment analyzer. 9 multiplexed libraries were pooled and sequenced in one lane on the Illumina HiSeq4000 sequencing platform (50-bp, single end-reads) at the

Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant.

**RNA-seq Analysis—**Preimplantation RNA-seq analyses were performed on previously published data (GSE45719) (Deng et al., 2014), specifically the single cell samples from the zygote to blastocyst developmental stages. Bulk RNA-seq analyses were also performed on previously published data (GSE85632) (Hendrickson et al., 2017). Fastq files were processed using Kallisto, and the resulting gene count matrix was generated (Bray et al., 2016). Genes were filtered, keeping only those with at least 300 counts across 9 samples. We then used the limma *R* package to perform full-quantile normalization (Ritchie et al., 2015). Or DESeq2 to test for differential expression (Anders and Huber, 2010). One sample was filtered out based on the PCA plot, leaving 258 samples and 12909 genes. Pseudotime was then computed using the *R* package slingshot (Street et al., 2018). Trends of gene expression were obtained using tradeSeq (Van den Berge et al., 2020).

For quantification of retrotransposons, Fastq files were analyzed by STAR with options '–outSAMtype BAM SortedByCoordinate– outSAMattributes XS–outFilterMultimapNmax 100000000' (Dobin et al., 2013). The repeatmasker track was downloaded from the UCSC Table browser, using the GRCm38/mm10 reference genome and filtered for short repeats. The genome annotation was downloaded from Gencode (Basic gene annotation, release M22) (Harrow et al., 2012). The bam files were then compared with the repeatmasker or the basic annotation using FeatureCounts, with options '-O -p -B -C -M' (Liao et al., 2014). Using the pseudotime obtained on the as described above, the RT expression of a few selected families across time was then plotted.

**Motif Enrichment Analyses—**All annotated LTR sequences for MERVL, ORR1A0 and ORR1A1 in the C57B/6 mouse genome were downloaded from DFAM (Hubley et al., 2016). Subsequently, HOMER was used to generate scrambled control sequences via the HOMER's scrambleFasta.pl script. And the HOMER known module was used to compute motif enrichment within the LTRs of the tested ERVs for all known transcription factor motifs with the -opt flag to optimize the degeneracy threshold to get the best enrichment (Benner et al., 2017). The analyses were performed on each 2C-ERV family individually as well as a merged cohort. For visualization of specific motif instances within gene bodies, genomic sequences were obtained from the UCSC mm10 genome browser and scanned with the JASPAR database web client with default parameters (Mathelier et al., 2016).

**Luciferase Assays—**For MERVL-luciferase reporter assays, we used the pGL3 luciferase reporter vectors (Promega, Cat. # E1751) that harbors the $MERVL_{125–375}$-fragment previously described (MERVL-Luc) (Choi et al., 2017). MERVL-Luc reporters and control Renilla luciferase reporter pRL-TK (Promega, Cat. # E2241) were co-transfected into ESCs (600 ng and 150 ng per well of a 12-well plate, respectively), using Lipofectamine 2000 (Life Technologies, Cat. # 11668027). Transfection complexes containing the reporter constructs were prepared in Opti-MEM Reduced-Serum Medium (Life Technologies, Cat. # 31985062). After trypsinization with 0.25% Trypsin + EDTA (Life Technologies, Cat. # 25200–056), 100,000 cells were resuspended in ES media lacking Pen Strep, incubated with transfection complexes for 10 minutes at 37°C, and then transferred to one well of

a 12-well plate containing irradiated MEF feeders. After 48 hours, transfected ESCs were trypsinized, plated onto gelatin-coated plates for 1 hour to remove feeders, and then assayed for luciferase activity by Dual-Luciferase Reporter Assay System (Promega, Cat. # E1910) using a Glomax 20/20 Luminometer (Promega).

**Derivation and culture of mouse ESCs—**Mouse ESCs were isolated based on published protocols with slight modifications (Bryja et al., 2006). Uteri containing E3.5 wild-type embryos were isolated from timed pregnancies, and subsequently put in Knockout DMEM (Life Technologies, Cat. # 10829–018) supplemented with 10mM HEPES (Life Technologies, Cat. # 15630–080). E3.5 blastocysts were flushed with 1ml syringes with 18G needles, and individually transferred to a 12-well plate seeded with irradiated MEF (mouse embryonic fibroblasts) feeders in 1 mL N2B27 medium containing 100 U/ml LIF (EMD Millipore, Cat. # ESG1107), 1 μM PD0325901 (Sigma, Cat. # PZ0162) and 3 μM CHIR99021 (EMD Millipore, Cat. # 361559). After 5 days of incubation, embryo outgrowth was separated from the trophectoderm (TE), picked up by a 10 μl pipette, transferred to 20 μl Accutase (Life Technologies, Cat. # A11105–01) and incubated at 37°C for 20 min to dissociate cells. Dissociated cells were then cultured on irradiated MEF feeder cells with N2B27 medium containing LIF and two inhibitors for one passage. Subsequently, ESCs were passaged with 0.25% Trypsin-EDTA and maintained in regular mouse ES medium.

ESCs were cultured onto irradiated MEF feeder layers in the M15 ESC medium, which contained Knockout DMEM (Invitrogen, catalog no. 10829–018), 15% ES-grade fetal bovine serum (Invitrogen, Cat #. 16141079), 2 mM L-glutamine (Invitrogen, Cat #. 25030–164), $1 \times 10^{-4}$ M MEM non-essential amino acids (Invitrogen, Cat #. 11140–076), $1 \times 10^{-4}$ M 2-mercaptoethanol (Sigma, Cat #. M3148) and 1% 100 × penicillin and streptomycin. ESCs were split every two days.

**ESC transfection—**To overexpress *Klf5, Dux,* or *Dppa2* in ESCs, cells were transfected with *PiggyBac* vectors containing an EF1α-driven *Klf5, Dux,* or *Dppa2* expression cassette and an Ubc-puromycin selection marker. Individual *PiggyBac*-plasmid was mixed with the *PiggyBac* transposase plasmid in a 1:1 ratio, and subsequently transfected into ESCs using Lipofectamine 2000 (Life Technologies, Cat. # 12566014) following the manufacturer's instruction. Cells were selected with 3 μg/ml puromycin for two days on puromycin resistant MEF feeders, and then cultured in puromycin-free M15 ES medium for following analyses. The *PiggyBac*-EF1α -GFP-Ubc-Puro plasmid was used as a negative control.

**Single-Embryo QPCR—**All single-embryo cDNA was prepared according to the Single Cell-to-Ct qRT-PCR kit (Life-Technologies, Cat# 4458236) with slight modifications. Pronuclear, 2-cell, 8-cell and blastocyst stage embryos were isolated and passed through three washes of PBS. Single embryos were then placed into individual PCR tubes and lysed in twice the recommended volume of Lysis/DNase (20 μL) for 15min at room temperature. Then, 2 μL of Stop Solution was added and incubated for 2 min. At this point, half of the reaction was stored in −80°C conditions as a technical replicate and the remaining sample (11 μL) continued through the original Single Cell-to-Ct protocol. All qRT-PCR reactions were performed using SSO Universal SYBR Green SuperMix, as per

manufacturer instructions (Biorad, Cat# 1725275). All QPCR analyses were performed on the StepOnePlus Real Time PCR system (ThermoFisher, Cat# 437660).

**Real-time PCR—**RNA was isolated using Trizol following manufacturer's instruction (Life Technologies, Cat. # 15596). cDNA was reverse-transcribed using iScript Advanced Reverse-Transcriptase (Bio-Rad, Cat. # 1725037). All real-time qPCR analyses were performed using SYBR FAST qPCR Master Mix (Kapa Biosystems, Cat. # KK4604), following manufacturer's protocol. Real time PCR analyses on retrotransposons detect their expression at the family level, using primers designed from the corresponding consensus sequences. *Actin* was used as a reference for both mRNA and retrotransposon quantitation in real time PCR analyses. All real time PCR primers used in our studies are listed in Table S5.

**Teratoma generation and histological analyses—**$1\times10^6$ of WT or *Klf5*-overexpressing ESCs were injected into the dorsal flanks of 6–7-week-old immune-deficient NCr-nu/nu female mice (Taconic, Cat# NCRNU). After 4–5 weeks, resulting teratomas were collected by surgical removal, fixed overnight in 10% buffered formalin (Fisher Scientific, Cat. # SF100–4), dehydrated in a graded series of ethanol solutions, embedded in Paraplast X-TRA paraffin (Fisher Scientific, Cat. # 23–021-401), sectioned at 6 mm thickness, and stained with hematoxylin and eosin (H&E) using standard procedures (Choi et al., 2011). These paraffin sections will be subjected to immunohistochemistry

**Embryoid body (EB) differentiation—**For EB differentiation, ESCs were plated in 10cm Petri dish (150,000 cells/ml) in ESC M15 medium without LIF and were gently cultured on a rotator after removal of feeder cells. Samples were collected at day 0, 3, 6 and 9 post EB differentiation for real-time PCR analyses and for immunofluorescence staining (see below).

For hanging drop EB formation, ESCs colonies were removed from feeders and dissociated to near single cell suspension using trypsin (Life Technologies, Cat # 25200114). Small clumps of cells (10–20 cells/clump) were then mouth pipetted into 25uL drops of M15 medium without LIF and placed on the underside of the lid to a 6cm Petri dish. 2mL of PBS was added to the 6mL Petri dish to prevent evaporation of the EB culture drops. EBs were cultured for 72h before subjected to fixation for immunofluorescence staining as described below.

**Generation of chimeric blastocysts and chimeric embryos from ESCs—**Klf5 overexpressing wild-type ESCs, *Klf5*-overexpressing D*ux*$^{-/-}$ ESCs and the corresponding control cells were all engineered to express GFP from a piggybac vector. To generate chimeric blastocysts, single ESCs was injected into each E2.5 8 cell C57BL/6N wild-type recipient morulae. Injected embryos were then cultured overnight in KSOM (Millipore, Cat # MR-106-D) to obtain chimeric blastocysts. GFP positive cells were scored in the ICM, TE or both in the chimeric blastocyst based on their morphology and location.

To generate chimeric mid-gestation embryos, we initially injected 10–15 GFP labeled, *Klf5*-overexpressing wild-type ESCs into C57BL/6N wild-type recipient blastocysts, followed by a uterus transfer into the CD1 pseudo-pregnant mothers. Chimeric embryos were

then collected at E12.5 for immunofluorescence analyses (see below). Subsequently, we generated E12.5 chimeric embryos by injecting single, GFP labeled ESCs into each recipient blastocysts for the same analyses.

**Immunohistochemistry and immunofluorescence analyses—**For immunohistochemistry (IHC) analyses on teratomas, 6μm paraffin sections were deparaffinized, dehydrated, and subjected to heat-induced antigen retrieval in a pressure cooker using Target Retrieval solution (DAKO, Cat. # S1699). Slides were incubated for 10 minutes with 3% H2O2, blocked for 3 hours with PBS containing 5% BSA and 0.3% Triton X-100, and incubated with primary antibodies against PL-1 (1:75, Santa Cruz Biotechnology, Cat. # sc-34713) overnight in PBS buffer containing 1% BSA and 0.3% Triton X-100. Slides were then incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies for 2 hours at room temperature, and then subjected to 3,3′-Diaminobenzidine (DAB) staining (Life Technologies, Cat. # 00–2014) followed by a counterstain with Mayer's hematoxylin (Electron Microscopy Sciences, Cat. # 26503–04). The sinusoidal trophoblast giant cells (s-TGCs) were identified by their enlarged nuclei and adjacent location in the maternal blood sinusoid space.

For immunofluorescence (IF), ESC colonies, differentiated EBs and blastocysts were fixed with 4% paraformaldehyde (Electron Microscopy Sciences, Cat # 19202) for 10 min at room temperature and incubated with blocking solution (0.1%Triton X-100 and 5% normal goat serum in PBS) for 1 hour at room temperature. EBs or ESCs were incubated overnight at 4°C with appropriate antibodies, including MERVL-Gag, 1:100, Epigentek, Cat # A-2801–100, Oct4, 1:100, Santa Cruz Biotechnology, Cat. # sc-5279, Cdx2, 1:100, Abcam, Cat. # ab76541, Nanog, 1:100, Cosmo Bio, Cat # REC-RCAB0002PF, Klf5, 1:100, Protein tech, Cat # 21017–1-AP). Subsequently, samples were stained with goat anti-rabbit IgG (H+L) secondary antibody, Alexa Fluor 594-conjugated secondary antibody (1:500, Life Technologies, Cat. # A11037) for 1 hour at room temperature (ESCs) or overnight at 4°C (EBs). Samples were then stained with DAPI (300 nM, Sigma, Cat. # D9564) and subjected to imaging analyses using spinning disk confocal microscopy (Andor CSU-X on Nikon Eclipse TE200-E). ImageJ was used to analyze mean fluorescence intensity of acquired images.

For IF staining of E12 chimeric mouse embryos, samples were fixed with 4% paraformaldehyde (Electron Microscopy Sciences, 19202) for 2 hours, incubated in 30% sucrose (Fisher, Cat # S5–500) overnight at 4°C, embedded in Tissue-Tek O.C.T. compound (VWR, Cat. #25608–930), and cryo-sectioned at 8 μm. These sections were subsequently subjected to IF analyses using antibodies against GFP (1:100, Abcam, Cat. # ab38689), Tpbpa (1:200, Abcam, Cat. # ab104401), and Mtp1 (1:150, Alpha Diagnostic, Cat. # MTP11-A). Trophoblast giant cells were identified based on their unique location in the placenta and their distinct morphology of enlarged nuclei. Spongiotrophoblasts were identified based on the staining of Tpbpa; syncytiotrophoblasts were identified based on the staining of Mtp1. The bilaminar structure of the yolk sac is identified by DAPI staining, and the visceral endoderm is identified by its columnar, epithelial morphology.

**Chromatin immunoprecipitation (ChIP) and ChIP-seq libraries—**V5 ChIP assays were performed using ESCs overexpressing a V5-tagged Klf5 protein. Wild-type D3 mouse ESCs were used as a control. Cells were cross-linked for 5′ at room temperature with 1% formaldehyde-containing Knockout D-MEM; cross-linking was stopped by PBS-glycine at a final concentration of 0.125 M. Cells were washed twice with ice-cold PBS, scraped, centrifuged for 10 min at 4000 rpm and flash-frozen in liquid nitrogen. Cell pellets were thawed in ice, resuspended in cell lysis buffer (5 mM PIPES, pH 8.0, 85 mM KCl, and 0.5% NP-40, 750 μl/15 cm plate) and incubated for 10 min on ice. During the incubation, the lysates were repeatedly pipetted up and down every 5 minutes. Lysates were then centrifuged for 10 min at 4000 rpm. Nuclear pellets were measured and resuspended in 6 volumes of sonication buffer (50 mM Tris-HCl, pH 8.1, 10 mM EDTA pH 8.0, 0.1% SDS), incubated on ice for 10 min, and sonicated to obtain DNA fragments below 2000 bp in length (Covaris S220 sonicator, 20% Duty factor, 200 cycles/burst, 150 peak incident power, 32 cycles of 20'' on and 40'' off). Sonicated lysates were cleared by centrifugation (20' at 13200 rpm) and 800 μg of chromatin were diluted in RIPA buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA pH 8.0, 0.5 mM EGTA, 1% Triton X-100, 0.1% SDS, 0.1% Na-deoxycholate, 140 mM NaCl), precleared with Protein G Sepharose (GE Healthcare, Cat # GE17–0618-01) for 2 hours at 4°C and immunoprecipitated overnight with 8 μg of normal mouse IgGs (ChromPure mouse normal IgG; Jackson ImmunoResearch), or anti-V5 antibodies (Invitrogen R960–25). 4% of the precleared chromatin was saved as input. After the overnight incubation, samples were incubated with 25 μL Protein G Sepharose beads precleared overnight in RIPA buffer with 0.5% (w/v) BSA and incubated for 2 hours at 4°C. Immunopecipitated samples were washed 5 times with RIPA buffer, once with LiCl buffer (0.5% NP-40, 0.5% Na-deoxicholate, 250 mM LiCl, 1 mM EDTA pH 8.0), and once with TE. After the last wash, immuno-precipitated complexes were eluted from the beads twice with 150 μL of TE with 1% SDS, each time incubating 30 min in a thermomixer set at 37°C and 900 rpm. The 300 μL eluted material was added of 1 μL RNaseA (10 mg/ml) and 18 μL 5M NaCl and incubated at 67°C for 4–5 hours to reverse formaldehyde cross-linking. Inputs were added of elution buffer to 300 μL total volume, and subject to the same treatment. Reverse cross-linked samples were added of 2.5 volumes of ice-cold ethanol and precipitated overnight at −20°C. DNA was pelleted by centrifugation (20min at 13,200 rpm and 4°C), and pellets resuspended in 100 μL TE, 25 μL 5X PK buffer (50 mM Tris-HCl, pH 7.5, 25 mM EDTA pH 8.0, 1.25% SDS), and 1.5 μL proteinase K (20 mg/ml), and incubated 2 hours at 45°C. After proteinase K digestion, DNA was purified with the QIAGEN QIAquick PCR Purification Kit, eluted in 60 μL of water and analyzed by qPCR together with 2% of the input chromatin prior to ChIP-seq library preparation using SYBR FAST qPCR Master Mix (KAPA, Bio-systems Cat. # KK4604).

ChIP-seq libraries were prepared using the Illumina TruSeq DNA sample preparation kit according to manufacturer instructions with few modifications. We used 150 ng of ChIP input DNA (as measured by Nanodrop) and 50 μl of immunoprecipitated DNA as a starting material; library samples were enriched through 12 cycles of PCR amplification. We assessed library quality and fragment size by qPCR and Fragment analyzer and sequenced the multiplexed libraries on one lane the Illumina HiSeq4000 sequencing platform (single

end-reads, 50 bp long) at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

**ChIP-seq analyses—**Fastq files were aligned using STAR with the same options as above. Bam files for Klf5 ChIP samples were merged separately. Peaks were then called using macs2 (Zhang et al., 2008). Coverage of the peaks by the initial bam files was then computed using be-tools_2.25.0. Finally, peaks were annotated using the HOMER_4.10 annotatePeaks function, with option -m klf5.motif to specifically look for peaks containing the Klf5 motif. RNA-Seq data was processed as indicated for the TFs plots. For each peak, we therefore also looked at the mean gene-expression (TPM) between late 2-cell and 16 cells (with log1p transformation afterward) of the gene that is nearest to the peak. Published data were analyzed from GSE85632 (Dux, DUX4), GSM515664 (Nelfa), GSE60066 (Tbx3), GSE117171 (Dppa2).

**Cistrome DB analyses—**Peak bed files from all ChIP-seq experiments performed in mouse ESCs were downloaded from the CistromDB batch repository and subsequently annotated with HOMER2′s annotatePeaks.pl script (Mei et al., 2017). Following annotation, factors with the potential to regulate Klf5 were defined as those that had peaks present within a window 2.5kb upstream and 2.5kb downstream of the TSS.

**Klf5 knockdown in preimplantation embryos using RNA interference (RNAi)—**5IU of PMSG (Prospec Cat# HOR-272) and HCG (Sigma-Aldrich, Cat # CG5–1VL) were injected intraperitoneally into 4–5 week old, wild-type C57BL/6J female or or C57BL/6J C3H/HeJ F1 mice 46–48h apart. Immediately following HCG injection, females were paired with C57BL/6J stud males and pronuclear stage embryos together with cumulus cell clusters were harvested from plugged females 20h after the mating. Pronuclear stage embryos were dissociated from cumulus cells using Hylorondase (Fisher, Cat # MR-0510F). Embryos were then incubated at 37C with 5% CO2 while siRNAs were prepared for injection. Before injection, scrambled siRNA (Thermo Fisher Scientific, Cat# AM4611) and *Klf5* siRNAs (Thermo Fisher Scientific, Cat# 160900) and *Klf4* siRNAs (Thermo Fisher Scientific, Cat# 156021) were prepared at a working concentration of 100uM were spun at 10k rpm for 5 min to clear debris. In experiments when we double knocked down *Klf4* and *Klf5*, we prepared siRNA mixtures containing 50 uM Klf4 siRNAs and 50 uM Klf5 siRNAs. After the microinjection of siRNAs, embryos were cultured *in vitro* to E4.0, and then subjected to IF analyses described above.

**Klf5 knockout in preimplantation embryos using CRISPR-EZ—**C57BL/6J or C57BL/6J C3H/HeJ F1 mice were crossed to C57BL/6J stud males and pronuclear stage embryos were dissociated from cumulus cells using Hylorondase (Fisher, Cat # MR-0510F). After the zona was weakened with acid Tyrode's (Sigma Aldrich, Cat # T1788), the embryos were subsequently washed in M2 buffer. Cas9 RNP complexes were then assembled *in vitro* by combining 40uM of Cas9 protein with 2ug of sgRNAs targeting Klf5 (CCAGACCGUCCAUGCCCACG, AGCACCCGCGUGGGCAUGGA, GGUC AGCACCCGCGUGGGCA, Synthego). Assembled RNPs were then mixed with a cohort of 50–75 embryos, and electroporated with standard parameters (Modzelewski et al.,

2018). Electroporated embryos were then cultured in KSOM until E3.5, at which point they were processed for IF analysis described above.

### QUANTIFICATION AND STATISTICAL ANALYSIS

**Quantification and statistical analysis for experimental data—**For experimental data, statistical analyses were performed in GraphPad Prisim. All statistical details for each experiment are specified in the figure legend. In most experiments, an unpaired Student's t test was used to compare two groups with a significance level of 0.05. For embryo experiments N refers to the total number of individual embryos used. For cell line experiments N refers to independent cell lines generated. For chimeric conceptus experiments, N refers to individual chimeric conceptuses.

**Quantification and statistical analysis for differential expression—**Quantification of gene expression was obtained from Kallisto. The resulting counts matrix was then filtered, keeping genes with at least 300 counts across 9 samples. The filtered matrix was then used as input for DESeq2 for differential expression analyses. For retrotransposon expression analysis, mapping was performed with STAR with the following parameters:'–outSAMtype BAM SortedByCoordinate–outSAMattributes XS–outFilterMultimapNmax 100000000' (Dobin et al., 2013). Retrotransposon subfamily coordinates were obtained from the repeatmasker track from the USCS Table Browser utility. FeatureCounts was then used to quantify retrotransposon expression at the family level with the following parameters: '-O -p -B -C -M' (Liao et al., 2014).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

We thank Nina Xiong, Jessica Mar, Andrew Modzelewski, Joseph Martin, Pratishtha Rawat, and Suifang Mao for technical assistance, discussions, input, and critical evaluation of the assumptions and evidence provided in this study. Further, we express our deepest gratitude to Alberto de Laco and the Trono lab for their generosity in sharing *Dux* knockout and control ESC lines. This study utilized Berkeley's Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley (supported by the NIH S10 OD018174 instrumentation grant) and computational resources provided by Berkeley Statistics, Biostats, and UT-Dallas. L.H. is a Thomas and Stacey Siebel Distinguished Chair Professor and Bakar fellow and is supported by a Howard Hughes Medical Institute (HHMI) faculty scholar award and several grants from the NIH (1R21HD088885, GRANT12095758, and 1R21OD027053–01). M.K. is supported by a CRCC pre-doctoral fellowship.

## REFERENCES

Alda-Catalinas C, Bredikhin D, Hernando-Herraez I, Santos F, Kubinyecz O, Eckersley-Maslin MA, Stegle O, and Reik W (2020). A Single-Cell Transcriptomics CRISPR-Activation Screen Identifies Epigenetic Regulators of the Zygotic Genome Activation Program. Cell Syst 11, 25–41.e9. [PubMed: 32634384]

Anders S, and Huber W (2010). Differential expression analysis for sequence count data. Genome Biol 11, R106. [PubMed: 20979621]

Azami T, Waku T, Matsumoto K, Jeon H, Muratani M, Kawashima A, Yanagisawa J, Manabe I, Nagai R, Kunath T, et al. (2017). Klf5 maintains the balance of primitive endoderm versus epiblast specification during mouse embryonic development by suppression of Fgf4. Development 144, 3706–3718. [PubMed: 28870993]

Beddington RSP, and Robertson EJ (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. Development 105, 733–737. [PubMed: 2598811]

Benner C, Heinz S, and Glass CK (2017). HOMER - Software for motif discovery and next generation sequencing analysis http://homer.ucsd.edu/homer/.

Bray NL, Pimentel H, Melsted P, and Pachter L (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol 34, 525–527. [PubMed: 27043002]

Bryja V, Bonilla S, and Arenas E (2006). Derivation of mouse embryonic stem cells. Nat. Protoc 1, 2082–2087. [PubMed: 17487198]

Casser E, Israel S, Witten A, Schulte K, Schlatt S, Nordhoff V, and Boiani M (2017). Totipotency segregates between the sister blastomeres of two-cell stage mouse embryos. Sci. Rep 7, 8299. [PubMed: 28811525]

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. [PubMed: 31133747]

Choi YJ, Lin CP, Ho JJ, He X, Okada N, Bu P, Zhong Y, Kim SY, Bennett MJ, Chen C, et al. (2011). miR-34 miRNAs provide a barrier for somatic cell reprogramming. Nat. Cell Biol 13, 1353–1360. [PubMed: 22020437]

Choi YJ, Lin C-P, Risso D, Chen S, Kim TA, Tan MH, Li JB, Wu Y, Chen C, Xuan Z, et al. (2017). Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. Science 355, eaag1927.

Dai Q, Shen Y, Wang Y, Wang X, Francisco JC, Luo Z, and Lin C (2017). Striking a balance: regulation of transposable elements by Zfp281 and Mll2 in mouse embryonic stem cells. Nucleic Acids Res 45, 12301–12310. [PubMed: 29036642]

Dan J, Li M, Yang J, Li J, Okuka M, Ye X, and Liu L (2013). Roles for Tbx3 in regulation of two-cell state and telomere elongation in mouse ES cells. Sci. Rep 3, 3492. [PubMed: 24336466]

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. [PubMed: 28459456]

Deng Q, Ramsköld D, Reinius B, Sandberg R, Ramskold D, Reinius B, and Sandberg R (2014). Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science 343, 193–196. [PubMed: 24408435]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Eckersley-Maslin M, Alda-Catalinas C, Blotenburg M, Kreibich E, Krueger C, and Reik W (2019). Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. Genes Dev 33, 194–208. [PubMed: 30692203]

Ema M, Mori D, Niwa H, Hasegawa Y, Yamanaka Y, Hitoshi S, Mimura J, Kawabe Y, Hosoya T, Morita M, et al. (2008). Krüppel-like factor 5 is essential for blastocyst development and the normal self-renewal of mouse ESCs. Cell Stem Cell 3, 555–567. [PubMed: 18983969]

Franke V, Ganesh S, Karlic R, Malik R, Pasulka J, Horvat F, Kuzman M, Fulka H, Cernohorska M, Urbanova J, et al. (2017). Long terminal repeats power evolution of genes and gene expression programs in mammalian oocytes and zygotes. Genome Res 27, 1384–1394. [PubMed: 28522611]

Fujimori T, Kurotaki Y, Miyazaki J, and Nabeshima Y (2003). Analysis of cell lineage in two- and four-cell mouse embryos. Development 130, 5113–5122. [PubMed: 12944430]

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res 22, 1760–1774. [PubMed: 22955987]

Hendrickson PG, Doráis JA, Grow EJ, Whiddon JL, Lim J-WW, Wike CL, Weaver BD, Pflueger C, Emery BR, Wilcox AL, 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. [PubMed: 28459457]

Hirate Y, Hirahara S, Inoue K, Suzuki A, Alarcon VB, Akimoto K, Hirai T, Hara T, Adachi M, Chida K, et al. (2013). Polarity-dependent distribution of angiomotin localizes Hippo signaling in preimplantation embryos. Curr. Biol 23, 1181–1194. [PubMed: 23791731]

Hu Z, Tan DEK, Chia G, Tan H, Leong HF, Chen BJ, Lau MS, Tan KYS, Bi X, Yang D, et al. (2020). Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. Nat. Cell Biol 22, 175–186. [PubMed: 31932739]

Hubley R, Finn RD, Clements J, Eddy SR, Jones TA, Bao W, Smit AFA, and Wheeler TJ (2016). The Dfam database of repetitive DNA families. Nucleic Acids Res 44 (D1), D81–D89. [PubMed: 26612867]

Ishiuchi T, Enriquez-Gasca R, Mizutani E, Bošković A, Ziegler-Birling C, Rodriguez-Terrones D, Wakayama T, Vaquerizas JM, and Torres-Padilla ME (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat. Struct. Mol. Biol 22, 662–671. [PubMed: 26237512]

Judd J, Sanderson H, and Feschotte C (2021). Evolution of mouse circadian enhancers from transposable elements. Genome Biol 22, 193. [PubMed: 34187518]

Katz JP, Perreault N, Goldstein BG, Lee CS, Labosky PA, Yang VW, and Kaestner KH (2002). The zinc-finger transcription factor Klf4 is required for terminal differentiation of goblet cells in the colon. Development 129, 2619–2628. [PubMed: 12015290]

Korotkevich E, Niwayama R, Courtois A, Friese S, Berger N, Buchholz F, and Hiiragi T (2017). The Apical Domain Is Required and Sufficient for the First Lineage Segregation in the Mouse Embryo. Dev. Cell 40, 235–247.e7. [PubMed: 28171747]

Koscielny G, Yaikhom G, Iyer V, Meehan TF, Morgan H, Atienza-Herrero J, Blake A, Chen CK, Easty R, Di Fenza A, et al. (2014). The International Mouse Phenotyping Consortium Web Portal, a unified point of access for knockout mice and related phenotyping data. Nucleic Acids Res. Nucleic Acids Res 42, D802–D809, Epub 2013 Nov 4. 10.1093/nar/gkt977. [PubMed: 24194600]

Kruse K, Diaz N, Enriquez-Gasca R, Gaume X, Torres-Padilla M-E, and Vaquerizas JM (2019). Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. bioRxiv 10.1101/523712.

Liao Y, Smyth GK, and Shi W (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–930. [PubMed: 24227677]

Lin S-CJ, Wani MA, Whitsett JA, and Wells JM (2010). Klf5 regulates lineage formation in the pre-implantation mouse embryo. Development 137, 3953–3963. [PubMed: 20980403]

Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 550. [PubMed: 25516281]

Lu X, Sachs F, Ramsay LA, Jacques PÉ, Göke J, Bourque G, and Ng HH (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol 21, 423–425. [PubMed: 24681886]

Macfarlan TS, Gifford WD, Agarwal S, Driscoll S, Lettieri K, Wang J, Andrews SE, Franco L, Rosenfeld MG, Ren B, and Pfaff SL (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev 25, 594–607. [PubMed: 21357675]

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, and Pfaff SL (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 487, 57–63. [PubMed: 22722858]

Mathelier A, Fornes O, Arenillas DJ, Chen CY, Denay G, Lee J, Shi W, Shyr C, Tan G, Worsley-Hunt R, et al. (2016). JASPAR 2016: A major expansion and update of the open-access database of transcription factor binding profiles. Nucleic Acids Res 44, D110–D115. [PubMed: 26531826]

McCarthy EM, and McDonald JF (2004). Long terminal repeat retrotransposons of Mus musculus. Genome Biol 5, R14. [PubMed: 15003117]

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 (D1), D658–D662. [PubMed: 27789702]

Modzelewski AJ, Chen S, Willis BJ, Lloyd KCK, Wood JA, and He L (2018). Efficient mouse genome engineering by CRISPR-EZ technology. Nat. Protoc 13, 1253–1274. [PubMed: 29748649]

Nakamura T, Nakagawa M, Ichisaka T, Shiota A, and Yamanaka S (2011). Essential roles of ECAT15–2/Dppa2 in functional lung development. Mol. Cell. Biol 31, 4366–4378. [PubMed: 21896782]

Niwa H, Toyooka Y, Shimosato D, Strumpf D, Takahashi K, Yagi R, and Rossant J (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell 123, 917–929. [PubMed: 16325584]

Pfeffer PL (2018). Building principles for constructing a mammalian blastocyst embryo. Biology (Basel) 7, 41.

Presnell JS, Schnitzler CE, and Browne WE (2015). KLF/SP transcription factor family evolution: Expansion, diversification, and innovation in eukaryotes. Genome Biol. Evol 7, 2289–2309. [PubMed: 26232396]

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43, e47. [PubMed: 25605792]

Schoorlemmer J, Pérez-Palacios R, Climent M, Guallar D, and Muniesa P (2014). Regulation of Mouse Retroelement MuERV-L/MERVL Expression by REX1 and Epigenetic Control of Stem Cell Potency. Front. Oncol 4, 14. [PubMed: 24567914]

Stopka T, and Skoultchi AI (2003). The ISWI ATPase Snf2h is required for early mouse development. Proc. Natl. Acad. Sci. U.S.A 100, 14097–14102. [PubMed: 14617767]

Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, Purdom E, and Dudoit S (2018). Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics 19, 477. [PubMed: 29914354]

Strumpf D, Mao CA, Yamanaka Y, Ralston A, Chawengsaksophak K, Beck F, and Rossant J (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development 132, 2093–2102. [PubMed: 15788452]

Tabansky I, Lenarcic A, Draft RW, Loulier K, Keskin DB, Rosains J, Rivera-Feliciano J, Lichtman JW, Livet J, Stern JNH, et al. (2013). Developmental bias in cleavage-stage mouse blastomeres. Curr. Biol 23, 21–31. [PubMed: 23177476]

Tagliaferri D, Mazzone P, Noviello TMR, Addeo M, Angrisano T, Del Vecchio L, Visconte F, Ruggieri V, Russi S, Caivano A, et al. (2020). Retinoic Acid Induces Embryonic Stem Cells (ESCs) Transition to 2 Cell-Like State Through a Coordinated Expression of Dux and Duxbl1. Front. Cell Dev. Biol 7, 385. [PubMed: 32010697]

Van den Berge K, Roux de Bézieux H, Street K, Saelens W, Cannoodt R, Saeys Y, Dudoit S, and Clement L (2020). Trajectory-based differential expression analysis for single-cell sequencing data. Nat. Commun 11, 1201. [PubMed: 32139671]

Wigger M, Kisielewska K, Filimonow K, Plusa B, Maleszewski M, and Suwińska A (2017). Plasticity of the inner cell mass in mouse blastocyst is restricted by the activity of FGF/MAPK pathway. Sci. Rep 7, 15136. [PubMed: 29123210]

Wu G, Lei L, and Schöler HR (2017). Totipotency in the mouse. J. Mol. Med. (Berl.) 95, 687–694. [PubMed: 28102431]

Yamane M, Ohtsuka S, Matsuura K, Nakamura A, and Niwa H (2018). Overlapping functions of krüppel-like factor family members: Targeting multiple transcription factors to maintain the naïve pluripotency of mouse embryonic stem cells. Development 145, dev162404.

Yan YL, Zhang C, Hao J, Wang XL, Ming J, Mi L, Na J, Hu X, and Wang Y (2019). DPPA2/4 and SUMO E3 ligase PIAS4 opposingly regulate zygotic transcriptional program. PLoS Biol 17, e3000324. [PubMed: 31226106]

Yang J, Ryan DJ, Wang W, Tsang JCH, Lan G, Masaki H, Gao X, Antunes L, Yu Y, Zhu Z, et al. (2017a). Establishment of mouse expanded potential stem cells. Nature 550, 393–397. [PubMed: 29019987]

Yang Y, Liu B, Xu J, Wang J, Wu J, Shi C, Xu Y, Dong J, Wang C, Lai W, et al. (2017b). Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. Cell 169, 243–257.e25. [PubMed: 28388409]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, and Liu XS (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol 9, R137. [PubMed: 18798982]

Zhao T, Fu Y, Zhu J, Liu Y, Zhang Q, Yi Z, Chen S, Jiao Z, Xu X, Xu J, et al. (2018). Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. Cell Stem Cell 23, 31–45.e7. [PubMed: 29937202]

## Highlights

- Three endogenous virus (ERV) families are specifically induced in mouse 2C blastomeres
- Klf5 is a key transcription factor that induces all three 2C-specific ERVs
- Klf5 overexpression confers a 2C-like cell fate in single embryonic stem cells
- Klf5 and Klf4 act redundantly for ICM and TE specification in preimplantation embryos

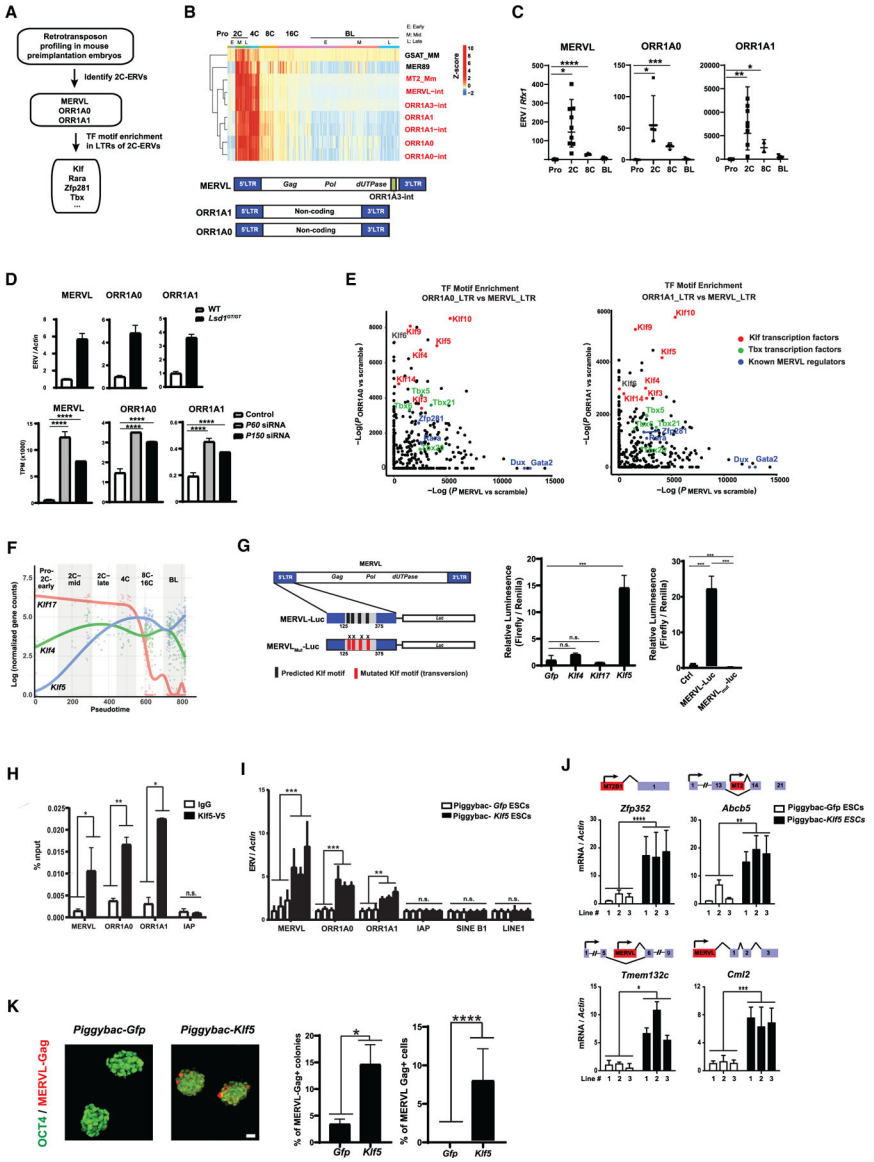


**Figure 1. Klf5 directly regulates three major 2C-specific ERV families in preimplantation embryos**

(A) A flowchart illustrating the strategy to identify key transcription factors that induce 2C-specific ERVs. Retrotransposon profiling using reanalyzed published (GSE45719) RNA-seq data of mouse preimplantation embryos identifies 2C-specific ERVs (Deng et al., 2014). LTR sequences of 2C-specific ERVs were subjected to HOMER motif enrichment analyses to predict transcription factors that regulate these ERVs.

(B and C) MERVL, ORR1A0, and ORR1A1 are three major families of 2C-specific ERVs in preimplantation embryos.

(B) Top: a heatmap illustrates the dynamics and stage-specific expression of a cohort of retrotransposons that peak in the 2C stage in mouse preimplantation development. Red, annotated elements of major 2C-specific ERV families; Pro, pronucleus; 2C, two-cell stage; 4C, four-cell stage; 8C, eight-cell stage; 16C, 16-cell stage; BL, blastocyst. Bottom: gene structures are shown as diagrams for MERVL, ORR1A0, and ORR1A1.

(C) Single-embryo real-time PCR analyses using pronuclear (n = 4), 2C (n = 9), 8C (n = 3), and BL embryos (n = 5) experimentally validated the 2C-specfic induction of MERVL (Pro versus 2C, *p = 0.0116, t = 3.024, degrees of freedom [df] = 11; Pro versus 8C, ****p < 0.0001, t = 13.57, df = 5), ORR1A0 (Pro versus 2C, *p = 0.0295, t = 2.843, df = 6; Pro versus 8C, ***p = 0.0002, t = 9.278, df = 5), and ORR1A1 (Pro versus 2C, **p = 0.0063, t = 3.447, df = 10; Pro versus 8C, *p = 0.0103, t= 4.567, df = 4). Error bars, SD.

(D) MERVL, ORR1A0, and ORR1A1 are coordinately induced in multiple lines of 2C-like ESCs. Left: upon bioinformatics analyses of published RNA-seq data (Ishiuchi et al., 2015), *Caf-1* deficient ESCs, including *P60* knockdown and *P150* knockdown ESCs, exhibited coordinated induction of MERVL, ORR1A0, and ORR1A1 (MERVL, control versus *P60* knockdown, adjusted p = 1.99e−224, control versus *P150* knockdown, adjusted p = 2.14e−212; ORR1A0, control versus *P60* knockdown adjusted p = 9.79e−64, control versus *P150* knockdown, adjusted p = 2.67e−86; ORR1A1, control versus *P60* knockdown, adjusted p = 1.92e−81, control versus *P150* knockdown, adjusted p = 2.68e−86). Error bars, SD. The p value was computed with the DESeq2 package in R. Right: real-time PCR analyses confirmed coordinated induction of ORR1A0, ORR1A1, and MERVL in *Lsd1* $^{GT/GT}$ ESCs.

(E) HOMER motif analyses predicted the enrichment of Klf binding motifs in the LTRs of 2C-specific ERVs. Binding motifs of Klf-family transcription factors are strongly enriched in LTRs of MERVL, ORR1A0, and ORR1A1 compared with scrambled control sequences. Dux and Gata2 are also labeled, highlighting the specific enrichment of their binding motifs in MERVL LTRs.

(F) *Klf5*, *Klf4*, and *Klf17* are the major Klf transcription factors expressed in mouse early preimplantation embryos. A pseudotime expression plot shows the dynamic expression profiles of *Klf17*, *Klf4*, and *Klf5* across preimplantation development. Klf5 exhibited early induction that peaked at 4C, and its expression persisted throughout preimplantation development.

(G) Overexpression of *Klf5*, but not *Klf4* or *Klf17*, specifically induces the MERVL-luc reporter. Left: a diagram showing the structure of the MERVL-luc and MERVL$_{Mut}$-Luc reporters, which contained four predicted Klf binding motifs and four mutated motifs, respectively. The four predicted Klf binding motifs reside in the minimal MERVL LTR fragment (125–375 bp) required to recapitulate MERVL expression in ESCs (Choi et al., 2017; Macfarlan et al., 2011). Right: overexpression of Klf5, but not Klf4 or Klf17, in HEK cells, along with a MERVL-Luc reporter, induced an increase in luciferase activity. Klf5, n = 3, p = 0.0007, df = 4, t = 9.456; Klf4, n = 2, not significant (n.s.); Klf17, n = 2, n.s. Mutations of all four predicted Klf binding motifs in the MERVL$_{Mut}$-Luc reporter abolished Klf5-dependent regulation. n = 3; error bars, SD; miniP-luc versus MERVL-luc, p = 0.0005, df = 4, t = 10.39; MERVL-luc versus MERVL$_{Mut}$-Luc, p = 0.0004, df = 4, t = 0.73.

(H–J) *Klf5* overexpression in ESCs specifically induces MERVL, ORR1A0, and ORR1A1 and promotes robust activation of the MERVL-associated transcriptome.

(H) Klf5 specifically occupies the LTRs of major 2C ERVs. Chromatin immunoprecipitation (ChIP) of Klf5 in ESCs revealed specific Klf5 association with the LTRs of MERVL, ORR1A0, and ORR1A1. Two independent, passage-matched ESC lines were tested. Error bars, SD; MERVL, *p = 0.0428, df = 2, t = 4.679, ORR1A0, **p = 0.0069, df = 2, t = 12.00; ORR1A1, *p = 0.0264, df = 2, t = 6.030.

(I) Three independent ESC lines were measured using real-time PCR analyses. Error bars, SD. MERVL, **p = 0.0085, df = 4, t = 4.829; ORR1A0, *p = 0.0003, df = 4, t = 11.83; ORR1A1, *p = 0.0033, df = 4, t = 6.258.

(J) Specific MERVL elements can serve as alternative promoters to generate preimplantation-specific gene isoforms. MERVL-dependent gene isoforms were strongly induced in ESCs upon *Klf5* overexpression. *Zfp352*, ****p < 0.0001, df = 4, t = 16.12; *Abcb5*, **p = 0.0030, df = 4, t = 6.408; *Tmem132c*, *p = 0.0147, df = 4, t = 4.110; *Cml2*, **p = 0.0001, df = 4, t = 15.18.

(K) A subset of *Klf5*-overexpressing ESCs exhibited strong expression of MERVL-Gag in immunofluorescence staining. The expression of MERVL-Gag in *Klf5*-overexpressing ESCs was mutually exclusive with that of Oct4 (left). The percentage of MERVL-Gag$^+$ ESC colonies (ESC colonies with 1 or more MERVL-Gag$^+$ cell(s)) and the percentage of MERVL-Gag$^+$ cells in the total population were quantified using fluorescent staining (right). Scale bar, 20 μm. Error bars, SD. Percentage of MERVL Gag$^+$ colonies: *p = 0.0405, df = 4, t = 2.987. Percentage of MERVL-Gag$^+$ cells: ****p < 0.0001, df = 32, t = 6.712. *p < 0.05, **p < 0.01, ***p < 0.001, ****p < 0.0001.

All p values were computed using unpaired, two-tailed Student’s t test unless otherwise indicated. See also Figure S1 and Table S1.

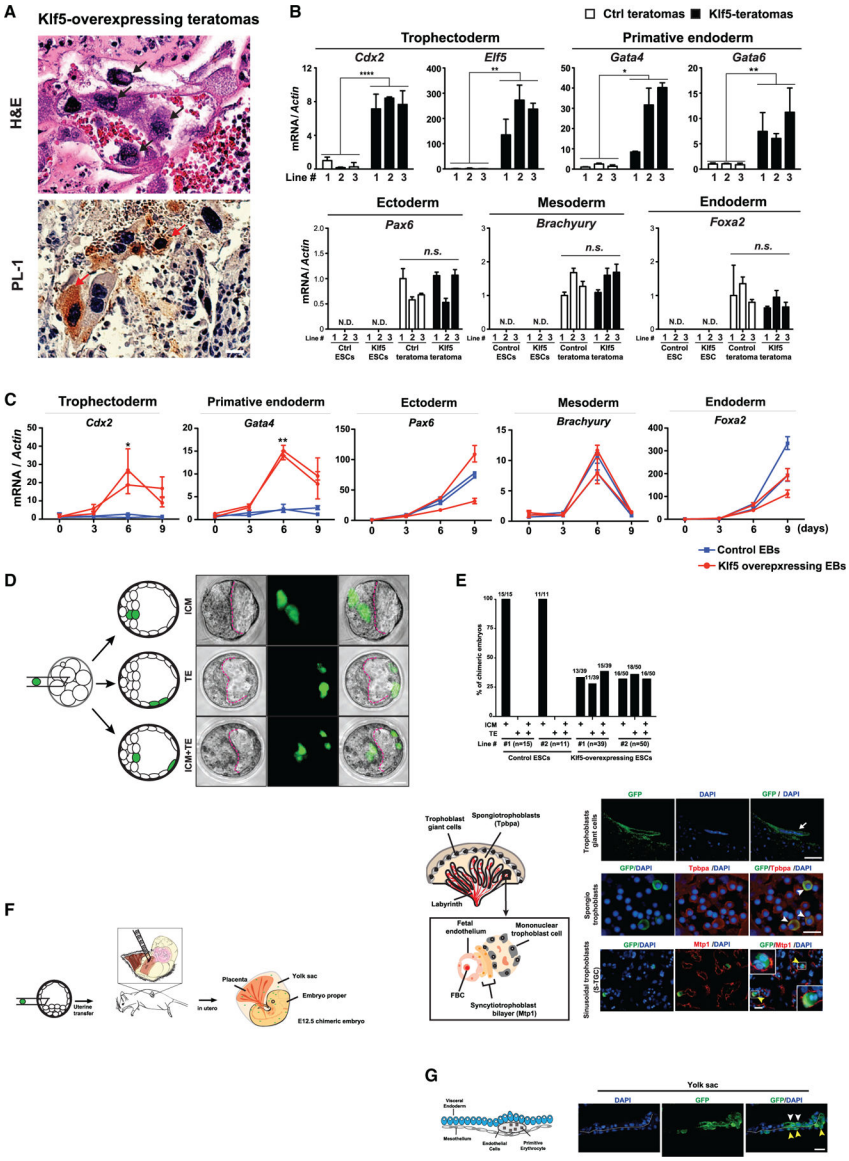



**Figure 2. Klf5 overexpression confers a bi-potential cell fate in ESCs *in vitro* and *in vivo***

(A) Teratomas derived from *Klf5*-overexpressing ESCs contain embryonic and extra-embryonic cell lineages. Teratomas generated from *Klf5*-overexpressing ESCs contain cells with characteristic placental TGC-like morphology (black arrows, top) and strong placental lactogen 1 (PL-1) expression (red arrows, bottom). Scale bars, 50 μm.

(B) The TE markers (*Cdx2* and *Elf5*) and the primitive endoderm (PrE) markers (*Gata4* and *Gata6*) were highly induced in *Klf5*-overexpressing teratomas in real-time PCR analyses but not in control teratomas. In contrast, expression of *Pax6* (an ectoderm marker), *Brachyury* (a mesoderm marker), and *Foxa2* (an endoderm marker) is induced similarly in control and *Klf5*-overexpressing teratomas. Teratomas were generated from three independent pairs of passage-controlled control and *Klf5*-overexpressing ESC lines. Error bars, SD; *Cdx2*, ****p < 0.0001, df = 4, t = 16.05; *Elf5*, **p = 0.0066, df = 4, t = 5.182; *Gata4*, *p = 0.0427, df = 4, t = 2.934; *Gata6*, **p = 0.0092, df = 4, t = 4.713.

(C) *Klf5*-overexpressing EBs, but not control EBs, showed significant induction in the TE markers *Cdx2* and PE markers *Gata4*. In contrast, they similarly induced markers of all three germ layers: *Pax6*, *Brachyury*, and *Foxa2.* EBs were generated from two independent pairs of passage-controlled control and *Klf5*-overexpressing ESC lines. Error bars, SD; *Cdx2* (day 6), *p = 0.0372, df = 2, t = 5.036; *Gata4* (day 6), **p = 0.0019, df = 2, t = 23.20.

(D and E) Single *Klf5*-overexpressing ESCs confer bi-potential cell fate in chimeric BL embryos. Shown are representative images (D) and quantitation (E) of chimeric BL embryos with ESC contribution to the ICM, TE, or both. In (D), a diagram illustrates the experimental strategy to generate chimeric BL embryos by microinjecting a single GFP-labeled, *Klf5*-overexpressing ESC into each C57BL/6N recipient morula. ESC contribution to the ICM and/or TE was determined by localization of $GFP^+$ ESC progenies in fluorescence imaging. Chimeric BLs were generated from two independent pairs of passage-controlled control and *Klf5*-overexpressing ESCs lines. Scale bar, 20 μm.

(F and G) Single *Klf5*-overexpressing ESCs confer bi-potential cell fate in chimeric E12.5 embryos, generating terminally differentiated extra-embryonic placental and yolk sac lineages.

(F) Top: a diagram illustrating the experimental strategy to generate chimeric E12.5 embryos by microinjecting a single GFP-labeled, *Klf5*-overexpressing ESC into each C57BL/6N recipient BLs, followed by embryo transfer into a pseudo-pregnant mother. Bottom left: a diagram illustrating several terminally differentiated cell lineages of placenta. Bottom right: in E12.5 chimeric embryos, progenies from a single GFP-labeled, *Klf5*-overexpressing ESC generated TGCs with a characteristic cellular morphology (white arrows), spongiotrophoblasts with Tpbpa expression (white arrowheads), and syncytio trophoblasts with Mtp1 expression (yellow arrowheads). Scale bars, 100 μm.

(G) Left: a diagram illustrating several terminally differentiated cell lineages of yolk sac. Right: a single GFP-labeled, *Klf5*-overexpressing ESC yield terminally differentiated visceral endoderm cells (white arrowheads) and embryonic mesothelium (yellow arrowheads) in the yolk sac of E12.5 chimeric embryos. Extra-embryonic visceral endoderm cells were identified based on the bilaminar structure of the yolk sac and their characteristic columnar epithelial morphology. Scale bar, 20 μm; *p < 0.05, **p < 0.01, ***p < 0.001. All p values were computed using unpaired, two-tailed Student's t test. See also Figure S2 and Table S2.

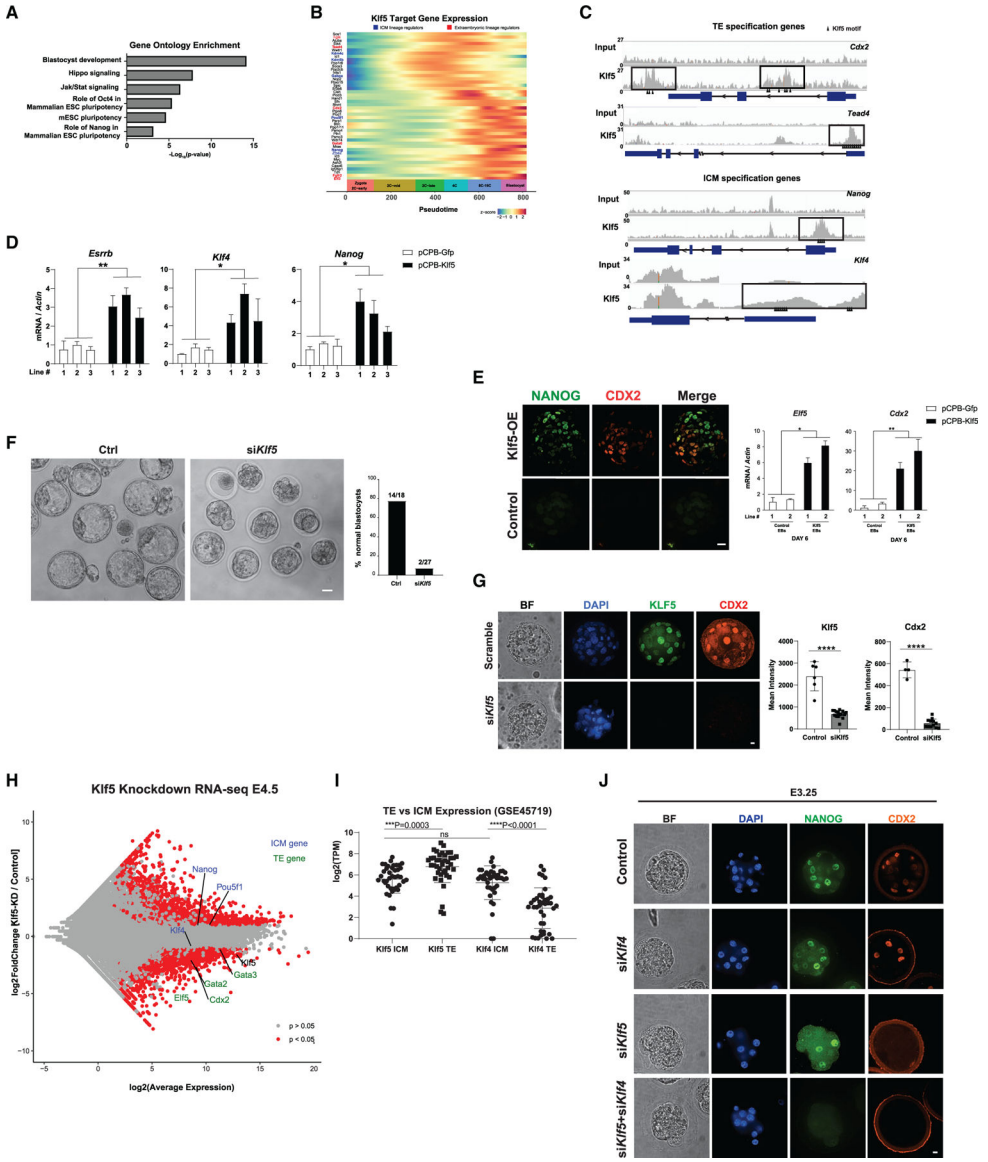



**Figure 3. Klf5 promotes dual induction of cell fate specification genes for embryonic and extra-embryonic commitment**

(A) IPA reveals the enriched functional terms among Klf5 ChIP-seq targets.

(B) A heatmap illustrates the expression profiles of the 70 most dynamically expressed *Klf5* targets in preimplantation development, which include multiple ICM (blue) and TE (red) specification genes.

(C) Read density of Klf5 ChIP-seq reads illustrates the enrichment of Klf5 occupancy on the *cis*-regulatory elements of lineage specification genes, including TE specification genes (*Cdx2* and *Tead4*), as well as ICM specification genes (*Nanog* and *Klf4)*.

(D) *Klf5* overexpression in ESCs elevates the expression of pluripotency-specific transcription factors, including *Nanog*, *Klf4*, and *Errsb*. Nanog: *p = 0.0264, df = 4, t = 3.436; Klf4: *p = 0.0162, df = 4, t = 3.992; Essrb: **p = 0.0035, df = 4, t = 6.153. n = 3 passage-controlled ESC lines. Error bars, SD.

(E) EBs derived from *Klf5*-overexpressing ESCs contain $Cdx2^+$ cells as well as $Nanog^+$ cells, whereas control EBs failed to activate Cdx2 and completely lost Nanog expression. Real-time PCR analyses confirmed induction of TE markers (Cdx2 and Elf5) in *Klf5*-overexpressing EBs. Scale bars, 100 μm (left). Error bars, SD. *Cdx2*, **p = 0.0372, df = 2, t = 5.036; *Elf5*, *p = 0.0339, df = 2, t = 5.289.

(F and G) *Klf5* knockdown stalls preimplantation development before the BL stage and abolishes Cdx2 expression in TE.

(F) Representative bright-field images (left) and quantitation (right) for E4.0 *Klf5* knockdown (siKlf5) or control (scramble small interfering RNA [siRNA]) BL embryos.

(G) Representative Cdx2 immunofluorescence images (left) and relative fluorescence quantitation (right) for control and *Klf5* knockdown BL embryos. Three independent *Klf5* knockdown experiments were performed, using a total of 18 control embryos and 27 *Klf5* knockdown embryos. Scale bar, 20 μM. Klf5, ****p % 0.0001, df = 22, t = 10.62; Cdx2, ****p % 0.0001, df = 14, t = 17.46.

(H) MA plot of RNA-seq data from control versus *Klf5* knockdown E4.5 embryos, confirming that *Klf5* depletion leads to defects in TE specification, as shown by reduced expression of the TE marker genes *Cdx2*, *Elf5*, *Gata2*, and *Gata3*, whereas expression of the ICM genes *Klf4*, *Nanog*, and *Pou5f1* is not affected significantly.

(I) Expression analysis of *Klf4* and *Klf5* in TE and ICM cells, demonstrating that *Klf5* mRNA is enriched in the TE whereas *Klf4* is expressed at levels comparable with *Klf5* in the ICM and at a significantly lower level in the TE.

(J) Immunofluorescent panel from control, *Klf4*, *Klf5*, and *Klf4 + Klf5* knockdown embryos at E3.25, stained for NANOG and CDX2 protein, demonstrating that Klf4 and Klf5 cooperate during ICM specification. Scale bar, 20 μM.

See also Figure S3 and Table S3.

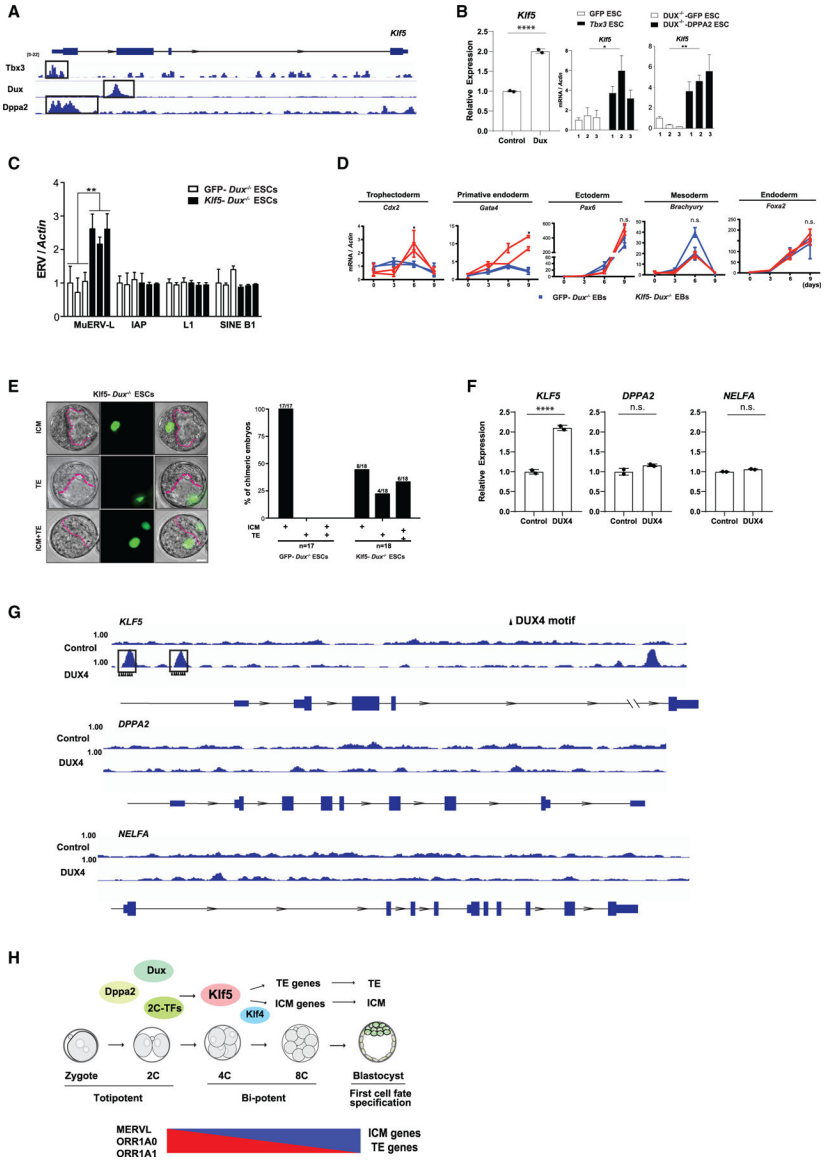



**Figure 4. Klf5 induction is regulated by the 2C-specific transcription factor Dux**

(A) Multiple 2C-specific transcription factors show enrichment of occupancies proximal to *Klf5*. We mined all ESC ChIP-seq experiments in the Cistrome DB and identified specific binding of Klf5 to a number of published 2C-specific transcription factors. The y axis shows the coverage depth of ChIP-seq reads, ranging from 0 to 22 for each transcription factor. (B) *Dux*, *Tbx3*, and *Dppa2* overexpression in ESCs induces *Klf5*. Using RNA-seq data from *Dux*-overexpressing and control ESCs (Hendrickson et al., 2017), we demonstrated robust induction of *Klf5* by *Dux*. Additionally, *Tbx3* and *Dppa2* overexpression also upregulates *Klf5*. Error bars, SD; *Klf5* (Dux overexpression), $p < 0.0001$; *Klf5* (*Tbx3* overexpression), $p = 0.0244$, $df = 4$, $t = 3.522$; *Klf5* (*Dppa2* overexpression), $^{**}p = 0.0027$, $df = 4$, $t = 6.622$. The p values for Dux overexpression were calculated with the DEseq2 package in R; otherwise, the p values were calculated on the basis of an unpaired Student's t test.

(C) *Klf5* acts downstream of *Dux* to induce MERVL. Real-time PCR analyses detected specific MERVL induction following Klf5 overexpression in *Dux* knockout ESCs. MERVL, **p = 0.0011, df = 4, t = 8.407.

(D) EBs derived from *Klf5*-overexpressing $Dux^{-/-}$ ESCs, but not control ESCs, showed significant induction of the extra-embryonic TE marker *Cdx2* and PrE marker *Gata4*. Markers for endoderm, mesoderm, and ectoderm (*Foxa2*, *Brachyury*, and *Pax6*, respectively) were induced similarly in *Klf5*-overexpressing and control EBs. Data were generated from two independent, passage-controlled ESC lines. Error bars, SD. *Cdx2* (day 6), *p = 0.0496, df = 2, t = 4.321; *Gata4* (day 9), *p = 0.0423, df = 2, t = 4.704.

(E) Single GFP-labeled, *Klf5*-overexpressing $Dux^{-/-}$ ESCs exhibit a bi-potential cell fate in chimeric BLs. ESC contribution to the ICM and/or TE was determined by localization of $GFP^{+}$ ESC progenies (left). The percentages of chimeric BLs with ESC contribution to the ICM, TE, or both were quantified (right). Scale bar, 20 μm.

(F and G) Human DUX4 directly induces *KLF5* but not other 2C-transcription factors (*DPPA2* and *NELFA*).

(F) Re-analysis of RNA-seq data from *DUX4*-overexpressing and control ESCs suggests that DUX/KLF5 regulation, but not DUX/DPPA2 or DUX/NELFA reguslation, is conserved in hESCs; *KLF5*, ****p < 0.0001.

(G) Read density plots for DUX4 ChIP-seq data in hESCs (Hendrickson et al., 2017) demonstrate its enriched occupancy proximal to *KLF5* but not *DPPA2* or *NELFA*.

(H) Our proposed model for the role of Klf5 in promoting bi-potential cell fate in preimplantation embryos. 2C-specific transcription factors, such as Dux, are induced in early 2C and converge on transcriptional activation of *Klf5.* Klf5 establishes bi-potential cell fate in *vitro* and in *vivo* through its dual regulation of ICM (in cooperation with *Klf4*) and TE specification genes.

See also Figure S4.

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Antibodies | | |
| PL-1 | Santa Cruz Biotechnology | sc-34713; RRID: AB_65419 |
| MERVL-Gag | Epigentek | A-2801–100 |
| Oct4 | Santa Cruz Biotechnology | sc-5279; RRID: AB_628051 |
| Cdx2 | Abcam | ab76541; RRID: 1523334 |
| Nanog | Cosmo Bio | REC-RCAB0002PF; RRID: AB_567471 |
| Klf5 | Proteintech | 21017–1-AP; RRID: AB_10696447 |
| Gfp | Abcam | ab38689; RRID: AB_732715 |
| Tpbpa | Abcam | ab104401; RRID: AB_10901888 |
| Mtp1 | Alpha Diagnostic | MTP11-A; RRID: AB_1619475 |
| V5 | Invitrogen | R960–25; RRID: AB_2556564 |
| Bacterial and virus strains | | |
| Top10 | Thermofisher | C404010 |
| Chemicals, peptides, and recombinant proteins | | |
| Cas9-NLS purified protein | UC-Berkeley, Macrolab | N/A |
| iScript Advanced Reverse Transcriptase | Bio-Rad | 1725037 |
| Phusion polymerase | UC-Berkeley, Macrolab | N/A |
| Esp3i | Thermofisher | Er0451 |
| Lipofectamine 2000 | Life technologies | 12566014 |
| PMSG | Prospec | HOR-272 |
| HCG | Sigma-Aldrich | CG5–1VL |
| Hyaluronidase | Thermo Fisher | MR-0510F |
| KSOM | Sigma-Aldrich | MR-121-D |
| Critical commercial assays | | |
| CellAmp whole Transcriptome Amplification Kit | Takarabio | 3734 |
| Single Cell-to-Ct qRT-PCR kit | Life technologies | 4458236 |
| SYBR FAST qPCR Master mix | Kapa Biosystems | KK4604 |
| Deposited data | | |
| Preimplantation RNA-seq | Deng et al., 2014 | GSE45719 |
| Dux and DUX4 overexpressed ESCs | Hendrickson et al., 2017 | GSE85632 |
| Klf5 (and control) ChIP-seq | This study | GSE137036 |
| Klf5 overexpression (ESC, bulk) and Knockdown (single embryo) RNA-seq | This study | GSE186005 |
| Experimental models: Cell lines | | |

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| Wild Type ESCs | This study | N/A |
| Klf5 overexpressed ESCs | This study | N/A |
| Tbx3 overexpressed ESCs | This study | N/A |
| Gfp Control ESCs | This study | N/A |
| Dux Knockout ESCs | Gift from Trono Lab | N/A |
| Dux Knockout, Klf5 overexpressed ESCs | This study | N/A |
| Dux Knockout, Dppa2 overexpressed ESCs | This study | N/A |
| Dux Knockout Gfp control ESCs | This study | N/A |
| Experimental models: Organisms/strains | | |
| C57BL/6J females | Jackson Laboratories | 000664 |
| C3H males | Jackson Laboratories | 000659 |
| C57BL/6J C3H/HeJ F1 females | This study | N/A |
| Oligonucleotides | | |
| Klf5 siRNA | Thermofisher | 160900 |
| Klf4 siRNA | Thermofisher | 156021 |
| Scramble siRNA | Thermofisher | AM4611 |
| sgKlf51–3 | Synthego | N/A |
| Recombinant DNA | | |
| Piggybac-Gfp | This study | N/A |
| Piggybac-Klf5v5 | This study | N/A |
| Piggybac-Dppa2 | This study | N/A |
| Piggybac-Tbx3 | This study | N/A |
| Piggybac-Dux | This study | N/A |
| Software and algorithms | | |
| HOMER | Benner et al., 2017 | 2010 |
| STAR | Dobin et al., 2013 | N/A |
| FeatureCounts | Liao et al., 2014 | N/A |
| R | https://www.r-project.org/ | N/A |
| Python | https://www.python.org/ | N/A |
| Kallisto | Pachter lab | N/A |
| DEseq2 | Love et al., 2014 | N/A |
| Other | | |
| FemtoJet 4i | Eppendorf | 5252000021 |
| Nikon Eclipse TE2000-E | Nikon | N/A |