Investigation of the Regulation and Functional Roles of MuERV-L in Murine Preimplantation Development

By

Martin Kinisu

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Comparative Biochemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Lin He, Chair
Professor Fenyong Liu
Professor Rebecca Heald

Summer 2021

Abstract

Investigation of the Regulation and Functional Roles of MuERV-L in Murine Preimplantation Development

by

Martin Kinisu

Doctor of Philosophy in Comparative Biochemistry

University of California, Berkeley

Professor Lin He, Chair

Murine preimplantation embryogenesis is characterized by the onset of totipotency, and the gradual restriction of cellular developmental potential. Intriguingly, due to the erasure of repressive epigenetic marks and activation of specific transcriptional regulators, endogenous retroviruses (ERVs) are transcriptionally active in a manner that depends on the ERV family and developmental stage being considered. It has been accepted that the ERV, MERVL, is associated with the expansion of cell fate as its expression marks totipotent blastomeres in vivo and correlates with the augmented developmental potential observed in MERVL+ embryonic stem cells (ESCs). Additionally, ERVs ORR1A0 and ORR1A1 share a similar expression profile with MERVL, both in vivo and in vitro, establishing a cohort of markers of expanded cell fate. By applying motif enrichment analyses and overexpression and differentiation assays, I found that *Klf5* can act upstream of this ERV cohort and impart totipotent-like attributes to ESCs. *Klf5*'s motif is highly enriched in all three of these ERVs, and consistently, overexpression of *Klf5* in ESCs expands the scope of their differentiation to include extraembryonic lineages. Remarkably, a single ESC overexpressing *Klf5* can contribute to placenta, yolk sac and the embryo proper, whereas control ESCs invariably only contribute to embryonic tissues. Further, *Klf5* is required for lineage formation in vivo through its direct regulation of extraembryonic trophectoderm specification genes and its cooperative action with *Klf4* during the establishment of the embryonic inner cell mass. Beyond simply marking totipotent cells, I also describe a role for MERVL's encoded Gag protein during the one cell to two cell transition. Depletion of MERVL by RNAi leads to developmental arrest preceding nuclear envelope breakdown (NEBD) after the congress of the maternal and paternal pronuclei. This arrest can be rescued by co-introduction of RNAi resistant MERVL-*Gag* mRNA, ascertaining that the observed defects are, in fact, due to the absence of MERVL-*Gag*. I also show that MERVL-*Gag* binds to, and may functionally cooperate with, ESCRT component Tsg101 and that depletion of Tsg101 phenocopies depletion of MERVL-*Gag*, though at an incomplete penetrance. Altogether, this work describes a novel regulator of totipotent-like developmental potential, as well as the ERVs that mark it, and describes the domestication of MERVL-Gag during cellular abscission in murine preimplantation development.

Table of Contents

Acknowledgements

This dissertation is dedicated to my family and friends for their ever-present support and encouragement. I would like to specifically thank my parents, Mary, and Philip Kinisu, for their advice and perspective throughout the challenges faced while conducting the research described in this work. Additionally, I would like to thank all members of the He lab that engaged in thoughtful discussions and critiques of the work. I would also like to thank all of the support staff in the He lab and at UC-Berkeley at large for their parts in facilitating this research. Finally, I would like to thank the committee in charge for their time and effort in guiding and revising this work

Chapter 1: Introduction

Abstract:

Preimplantation development in the mouse is a complex and dynamic process, involving substantial genomic reprogramming. This reprogramming enables both the acquisition of totipotency and the stage specific expression of ERVs. Commensurate with development is the specification of different lineages, which involves activation of specific transcriptional regimes by lineage specific transcription factors. Intriguingly, ERVs themselves have been domesticated and, through different modalities, facilitate host developmental processes. This review examines the dynamic process of murine preimplantation development and highlights a key ERV, MERVL, as a marker of totipotent cells both in vivo and in vitro. Further, this review discusses examples of ERV domestication, highlighting contexts where ERVs, at the DNA, RNA, and protein level, have become entwined in fundamental host processes.

Preimplantation Embryogenesis: A Gradual Restriction of Developmental Potency

Mammalian preimplantation development initiates with the unification of the terminally differentiated male and female gametes, the sperm, and the oocyte. Shortly after sperm entry, the oocyte is released from the meiotic MII arrest checkpoint and completes the second meiotic division that causes extrusion of the second polar body (Deng and Li, 2009; Wang et al., 2011). Commensurate with polar body II extrusion is the decondensation of the parental chromosomes leading to the formation of the maternal and paternal pronuclei (van der Heijden et al., 2006; Maeda et al., 1998; Santos et al., 2005). Typically, the male pronucleus forms within the fertilization cone near the surface of the zygote, whereas the maternal pronucleus forms in a more central location. This close proximity to the cell surface causes the male pronucleus to rapidly migrate towards the center due to Rab11a dependent recruitment of actin nucleation factors, Spire2 and Formin-2. This fast phase of pronucleus migration is succeeded by a slow phase mediated by microtubule networks polymerizing into cortical actin networks, generating additional inward movement of both maternal and paternal pronuclei (Scheffler et al., 2021). Once the maternal and paternal pronuclei meet in the central region of the zygote, the first mitotic division of life begins.

Murine preimplantation embryos are unique in that they do not initially have centrosomes as maternal gametes degrade their centrosomes and sperm do not deposit basal bodies required for centrosome establishment (Manandhar et al., 1998, 2005; Schuh and Ellenberg, 2007). Therefore, murine preimplantation embryos have a noncanonical mechanism of nucleating and establishing meiotic and mitotic spindles. Microtubule organizing centers (MTOCs) are initially dispersed throughout the cytoplasm and then coalesce at the poles of the spindle as it forms (Schuh and Ellenberg, 2007). It is not until the centrosomes appear *de novo* at the end of preimplantation development that the embryo organizes microtubules in the well-characterized, centrosome dependent fashion (Courtois et al., 2012).

Embryos in the one and two-cell stages of mouse preimplantation development are considered to be totipotent (Casser et al., 2017; Sotomaru et al., 1998; TOGASHI et al., 1987; Tsunoda and McLaren, 1983; Wang et al., 1997). That is, the zygote itself and each individual blastomere of the two-cell embryo are capable of generating all tissues required for the formation and *in utero* support of what will be a fully fertile adult organism. The totipotency of two cell blastomeres is demonstrated by the phenomenon of dichorionic diamniotic monozygotic twinning in mammals. Though natural occurrences of such twinning in mice are rare (McLaren et al., 1995), it can be induced by manual separation of the sister blastomeres and

reintroduction into a supportive uterine environment. This so-called bisection technique has been demonstrated to yield live births of twins in four studies that showed efficiencies of 1 of 26 pairs (TOGASHI et al., 1987), 6 of 23 pairs (Tsunoda and McLaren, 1983), 9 of 26 pairs and 4 of 10 pairs (Sotomaru et al., 1998; Wang et al., 1997), respectively. More recent and ambitious attempts at characterizing two cell totipotency demonstrated that from 66 pairs of twin blastomeres that survived the bisection technique, 27 were able to synchronously complete preimplantation development (Katayama et al., 2010). In sum, over the past 35 years of investigation, it has been shown that 82 of 107 embryos tested reach multiple developmental milestones in only one member of the pair and not the other (Casser et al., 2017). This leads to the speculation that totipotency may not be equally allocated in sister blastomeres of all two cell embryos.

Despite the segregation of totipotency at the two cell stage, recent reports have shown that totipotency in the mouse extends to blastomeres of the 4 cell stage as well (Maemura et al., 2021). Utilizing a similar bisection technique (termed quartering for convenience), all sister blastomeres of four cell stage embryos were able to reach the terminal stage of preimplantation development in approximately 25% of all parental embryos tested. While the preimplantation development regime remained largely intact in embryos derived from quartered embryos, very few of the embryos developed to term when transferred into the uterus of pseudo pregnant recipient females. A total of 2 pups were obtained by cesarean section out all 150 embryos transferred, a rate of 0.01. It was then shown that adults derived from individual four cell stage blastomeres did not display gross physical or metabolic abnormalities. Importantly, the study concluded that adult mice derived from individual four cell stage blastomeres displayed fecundity levels similar to adults derived from unmanipulated one cell stage embryos. It was also concluded that parental eight cell stage embryos do not possess bona fide totipotency as there were no instances where all blastomeres of the parental embryo successfully completed the preimplantation development regime, nor led to the birth of a viable pup.

At the eight cell stage, blastomeres situated on the periphery of the embryo eventually acquire an apical domain, making these blastomeres polar and those situated on the interior of the embryo remain apolar (Hirate et al., 2015; Johnson and Ziomek, 1981; Sasaki, 2015). This differential polarization is among the earliest determinants that, essentially, fixes the developmental trajectory of these blastomeres and their progeny into one of two differentiation programs (Posfai et al., 2017; Sasaki, 2015). Blastomeres that have acquired an apical domain will go on to contribute to the extraembryonic trophectoderm (TE) lineage, whereas the apolar cells will contribute to the inner cell mass (ICM) (Posfai et al., 2017; Sasaki, 2015). Mechanistically, the apical domain acts as a sequestration point for angiomotin, prohibiting angiomotin dependent recruitment and activation of Lats kinases (Korotkevich et al., 2017; Posfai et al., 2017; Sasaki, 2015). Due to this inhibition of angiomotin activity, the presence of the apical domain inactivates hippo signaling in outer blastomeres. Lats kinases target transcription factor Yap-1, phosphorylating it. (Lorthongpanich et al., 2013) In its phosphorylated form, Yap-1 cannot undergo cytoplasmic to nuclear translocation (Nishioka et al., 2009). In the context of murine preimplantation development, Yap-1's targets are genes, such as *Tead4, Cdx2* and *Gata3*, which work to establish the TE identity (Home et al., 2009; Kumar et al., 2018; Nishioka et al., 2008, 2009; Strumpf et al., 2005). Consistently, blastomeres on the interior of the eight cell stage embryo display cytoplasmic Yap-1 localization, whereas those on the periphery of the embryo have nuclear Yap-1 localization as well as activation of Yap-1 target genes (Lorthongpanich et al., 2013).

Despite the apparent loss of totipotency upon progression beyond the four cell stage, blastomeres of eight cell embryos retain marked developmental plasticity. Studies on the

potency of eight cell blastomeres have involved quartering four cell stage embryos to obtain single blastomeres and then allowing those blastomeres to divide, forming so called “octet” embryo pairs. These octet embryo pairs were then dissociated again and paired with four individual eight cell stage blastomeres, which served as carriers that are genetically distinct from the octet embryo. These experiments revealed that octet embryos were able to contribute partially to both embryonic and extraembryonic tissues, thus defining the broad developmental potency of eight cell blastomeres (Kelly, 1977).

Intriguingly, recent work has shown that manipulating the apical domain is sufficient to alter the fate of eight cell blastomeres (Korotkevich et al., 2017). It was demonstrated that individual blastomeres of eight cell embryos, cell autonomously, acquire apico-basal polarity marked by enrichment of membrane markers at the blastomere surface. Further, knockout of apical domain component genes, *Cdc42, Prkczi* and *Prkcz,* led to significant decrease in TE cells by the end of preimplantation development. Strikingly, whereas transplantation of cell fragments, not from the apical domain, from polarized eight cell blastomeres had no impact on the polarization of recipient apolar eight cell blastomeres, transplantation of the apical domain induced polarization in recipient blastomeres. Further, these recipients of the apical domain went on to take on a TE fate, establishing the sufficiency of the apical domain in the commitment to the TE. More recent studies on the developmental potency of individual eight cell blastomeres verified these previous findings (Maemura et al., 2021). It was shown that individual blastomeres of parental eight cell embryos could complete the preimplantation developmental regime, though this potency was not equal for all blastomeres. In approximately 70% of cases tested, between 50% and 100% (not inclusive) of the blastomeres from a parental eight cell embryo made it to the terminus of preimplantation development (Maemura et al., 2021). This heterogeneity in developmental potency is presumably due to the acquisition of the apical domain at the eight cell stage. While eight cell blastomeres possess striking plasticity in their developmental potency, no viable pups could ever be recovered following uterine transfer into pseudo pregnant recipient females. Due to this, blastomeres of eight cell embryos are termed bi-potential, rather than totipotent.

By final stage in preimplantation, termed the blastocyst stage, the embryo is marked by a fully segregated TE and ICM as well as an expanded blastocoel cavity. This stage is also characterized by the second cell fate decision of preimplantation development, the segregation of the epiblast (Epi) and the primitive endoderm (PrE) (Mihajlovic et al., 2015; Plusa et al., 2008). The Epi lineage will go onto form the embryo proper, whereas the PrE lineage will go onto form the extraembryonic yolk sac (Kadokawa et al., 1987; Lawson et al., 1991). At this stage, the cells of the Epi and TE lineage are termed pluripotent as they can only differentiate towards embryonic or extraembryonic lineages, respectively.

The establishment of the Epi lineage versus the PrE relies on sequential activation of transcription factor regimes. While their expression is detectable in all blastomeres in eight cell embryos, Nanog and Gata6 are the earliest markers for the Epi and PrE lineages, respectively (Bessonnard et al., 2014; Schrode et al., 2014; Xenopoulos et al., 2015). Cells of the ICM eventually either exclusively express Nanog, and thus take on an Epi fate, or express Gata6, restricting themselves to the PrE compartment (Schrode et al., 2014; Xenopoulos et al., 2015). These two transcriptional regimes also directly repress one another as shown by the observation that Gata6 null embryos express Nanog in all ICM cells (Schrode et al., 2014), and conversely Gata6 is expressed uniformly in the ICM of Nanog null embryos (Mitsui et al., 2003).

It has also been demonstrated that FGF signaling plays an important role in the segregation of the PrE and Epi lineages (Kang et al., 2013). *Fgf4* is specifically expressed in the

Epi compartment, whereas its receptor *Fgfr2* is restricted to PrE cells by the blastocyst stage (Kang et al., 2017). Blocking FGF signaling leads to a complete loss of the PrE lineage, and all cells of the ICM take on an Epi fate. Consistently, the addition of excess FGF4 causes all cells of the ICM to take on a PrE fate (Bessonnard et al., 2017).

Transcriptional Regulation of Totipotency and MERVL

As previously stated, one cell embryos, and blastomeres up through the four cell stage retain the property of totipotency. Intriguingly, it has been shown that mouse embryonic stem cells (mESCs) derived from the ICM of the blastocyst are heterogenous, containing cells with augmented cell fate as compared to otherwise normal pluripotency (Macfarlan et al., 2012). Fascinatingly, these subpopulations are marked by the expression of particular transposable elements, specifically those of the Murine Endogenous Retrovirus that uses the Leucine tRNA to prime reverse transcription, or MERVL, family (Macfarlan et al., 2011, 2012). During the course of normal preimplantation development in the mouse, MERVL is robustly expressed from the one cell to the four cell stage, precisely marking totipotent stages (Kigami et al., 2003). Experimentally, MERVL+ ESCs have been shown to contribute to both the TE and ICM in chimeric blastocysts (Choi et al., 2017; Hu et al., 2020; Macfarlan et al., 2012; Yan et al., 2019). Further, MERVL+ ESCs, when introduced at the blastocyst stage have been shown to contribute to the embryo proper, the yolk sac and the placenta in the resulting chimeric conceptus at ~E12.5 (Choi et al., 2017; Macfarlan et al., 2012). These observations establish the bi-potential cell fate of MERVL+ cells, but do not satisfy the definition of *bona fide* totipotency. Of note, the Human Endogenous Retrovirus that uses the Leucine tRNA to prime reverse transcription (HERVL) is highly similar in sequence to MERVL and shares a common expression profile in human preimplantation development, after accounting for the difference in embryonic genome activation in human and zygotic genome activation in the mouse (Bénit et al., 1997; Göke et al., 2015). However, no investigation into the developmental potency of HERVL+ human embryonic stem cells (hESCs) has yet been described.

Due to its correlation with augmented cell fate in mESCs, the transcriptional regulation of MERVL has recently been of great interest. Hendrickson et al and De Iaco et al were among the first to describe a direct upstream regulator for MERVL (Hendrickson et al., 2017; Iaco et al., 2017). They began by identifying transcription factor motifs enriched in gene expression clusters during the course of human oogenesis and preimplantation development. They identified a transient cluster of genes whose expression was restricted to the cleavage stages of human preimplantation development. By applying motif enrichment analyses, researchers were able to identify DUX4 as a putative regulator for this gene set, and further showed that DUX4 overexpression in hESCs activates cleavage-stage-specific genes and repetitive elements such as HERVL. Strikingly, enforced DUX4 expression in hESCs is cytotoxic (Ashoti et al., 2021), emblematic of DUX4's ability to drive facioscapulohumeral muscular dystrophy (Ashoti et al., 2021; Bosnakovski et al., 2020). Recent work has shown that DUX4 initiates a deterministic set of events, leading to the expression of early pluripotency genes and genes typically expressed early in development (Ashoti et al., 2021). It is believed that the activation of these genes underlies DUX4 cytotoxicity in somatic cells, however the applicability to DUX4 mediated apoptosis in hESCs has not been thoroughly investigated. It is apparent that DUX4 activates several markers of naïve pluripotency, therefore culturing DUX4 overexpressing cells in naïve conditions could be beneficial in probing the impacts on developmental potency DUX4 has on hESCs. Hendrickson et al then noticed that mouse Dux is a weakly conserved DUX4 ortholog with the greatest identity shared in their homeobox domains 1 and 2 at 35% and 58%, respectively (Hendrickson et al., 2017). They demonstrated that Dux overexpression in mESCs activated early cleavage stage genes as well as MERVL and did so in a direct manner. These

observations established Dux and DUX4 as conserved upstream regulators of MERVL and HERVL, respectively. Despite this, no attempt was made to score changes in stem cell potency downstream of Dux of DUX4 expression.

Further investigation into the transcriptional regulation of MERVL and Dux conducted in mESCs revealed that maternal factors Dppa2 and Dppa4 act upstream of Dux, and together these three factors establish a positive feedback loop said to act at the top of the transcriptional hierarchy governing the onset of ZGA, and potentially totipotency (Eckersley-Maslin et al., 2019; De Iaco et al., 2019; Yan et al., 2019). In mESCs, overexpression of Dppa2 and Dppa4 leads to a significant increase in the expression of ZGA associated transcripts, including MERVL (Eckersley-Maslin et al., 2019; De Iaco et al., 2019). Further, single and double knockout ESCs of Dppa2 and Dppa4 revealed their necessity in the activation of these transcripts in ESCs that naturally cycle between MERVL- and MERVL+ states (De Iaco et al., 2019). Finally, ChIP-seq for Dppa2/4 revealed that they directly occupy the Dux promoter region (Eckersley-Maslin et al., 2019). Additional studies on Dppa2 revealed that this factor is regulated by sumoylation, a modification that switches its activity from transcriptional activator to transcriptional repressor (Yan et al., 2019). The SUMO E3 ligase, Pias4, was found to associate with Dppa2 and target K31 and K108 for sumoylation. Further, a SUMO2ΔGG-Dppa2 fusion, made to mimic sumoylated Dppa2, efficiently repressed Dux expression in mESCs (Yan et al., 2019). Finally, it was demonstrated that overexpression of Dppa2 not only induces MERVL and Dux expression in mESCs, but also enhances their potency. Dppa2 overexpressed mESCs are able to contribute to the TE and ICM of chimeric blastocysts following injection into recipient eight cell embryos (Yan et al., 2019). However, the mechanism underlying this expansion of cell fate potential remains elusive, but speculatively is the same mechanism underlying the bi-potential cell fate of MERVL+ cells.

Further work on identifying additional regulators of ZGA and totipotency revealed a role for Smarca5 in the regulation of ZGA transcripts such as MERVL. SMARCA5 expression at that one cell stage is present in both pronuclei and remains nuclear at least through the two cell stage. Intriguingly, it has been shown that Smarca5 and Dppa2 directly associate in mESCs. Consistent with the notion that interdependencies exist between the two factors, Dppa2 overexpression in Smarca5 KO cells partially rescues the reduced expression of ZGA genes. Finally, Smarca5 overexpression in Dppa2 knockout mESCs fails to induce the expression of MERVL and other ZGA transcripts, ascertaining that Smarca5 and Dppa2 cooperate in the regulation of MERVL and other ZGA transcripts (Alda-Catalinas et al., 2020).

A recent studies describes the role of the maternal factor Nelfa in driving Dux expression in mESCs (Hu et al., 2020; Wang et al., 2021). Using an eGFP knock-in tag towards the 3’ end of the *Nelfa* gene, researchers found that Nelfa is expressed heterogeneously in ESCs and overlaps with MERVL expression (Hu et al., 2020). However, this analysis may be confounded by the presence of a MERVL insertion proximal to where their eGFP tag was inserted. Regardless, overexpression of Nelfa in trans was shown to activate both Dux and MERVL, as well as several ZGA transcripts. The activation of MERVL and ZGA transcripts by Nelfa was shown to depend on Dux, as no such activation could be observed when Nelfa was overexpressed in Dux knockout cells. Further, researchers also showed that Nelfa directly associates with Top2a and that knockdown of Top2a in Nelfa overexpressing mESCs abolishes Nelfa mediated Dux upregulation (Hu et al., 2020). These observations establish Nelfa, in collaboration with Top2a, as a novel upstream regulator of Dux and MERVL, but an investigation into the interdependencies of Nelfa/Top2a, Dppa2/4 and Smarca5 has not been described. Therefore, any possible redundancy between these regimes remains ambiguous.

Recently, *Zscan4c*, a gene typically high in MERVL+ cells as well as two cell stage embryos was shown to directly activate MERVL expression (Zhang et al., 2019). *Zscan4c* is also a direct Dux target and is upregulated following Dux overexpression (Choi et al., 2017; Hendrickson et al., 2017; Macfarlan et al., 2012). Presumably, the difference between these two factors on MERVL expression is that the early factor, Dux, is responsible for MERVL initiation, whereas factors like Zscan4c are important for MERVL expression maintenance, although this has not been systematically investigated. Herein lies a discrepancy as previous work revealed that *Zscan4c* localizes specifically to telomeres in mESCs (Zalzman et al., 2010). Therefore, the observation of direct binding of MERVL by *Zscan4c* needs to be elaborated in the context of *Zscan4c* telomeric localization. Intriguingly, transcription factor Tbx3 has also been shown to have elevated expression in MERVL+ ESCs (Dan et al., 2013). Further, overexpression of Tbx3 leads to a significant increase in *Zscan4c* expression. Luciferase assays demonstrated that Tbx3 can directly regulate the *Zscan4c* promoter (Dan et al., 2013). However, no measurement of MERVL nor changes to developmental potency following Tbx3 overexpression were described. Further, it is not clear how Tbx3 is integrated into the Smarca5/Dppa2/4, Dux axis nor how Tbx3 interdepends, if at all, on Nelfa activity.

While activators of MERVL, Dux and bi-potential cell fate are continuing to be identified, several repressors of MERVL have already been described. Among the first repressors of MERVL reported is *Kdm1a* (Macfarlan et al., 2011). Microarray analyses on *Kdm1a* mutant ESCs revealed that several ZGA transcripts, such as MERVL, were significantly upregulated. Strikingly, it was shown that MERVL was also de-repressed in *Kdm1a* mutant blastocysts. Importantly, *Kdm1a* mutant ESCs retained expression of DNA methyltransferase Dnmt1 and had normal levels of global DNA methylation. This is in line with a more recent study that showed, via DNMT triple-knockout, that MERVL is not sensitive to DNA methylation (Eckersley-Maslin et al., 2016). Intriguingly, it was found that *Kdm1a* directly associates with co-repressor Kap1. Consistently, deletion of Kap1 from mESCs leads to the upregulation of IAP and MERVL elements (Rowe et al., 2010). However, the significance of this interaction with respect to MERVL silencing has not been thoroughly described. Further, it is not clear whether Kap1 and *Kdm1a* even function together for MERVL silencing, or if they constitute separate restriction mechanisms. Consistent with MERVL associated bi-potential cell fate, *Kdm1a* mutant ESCs were found to be capable of generating extraembryonic tissues when allowed to spontaneously differentiate in embryoid body assays (Macfarlan et al., 2011).

Investigation of the histone methyltransferase heterodimer G9a/GLP revealed that, in G9a knockout and GLP knockout mESCs, MERVL was significantly upregulated approximately 8-fold and 13-fold, respectively (Maksakova et al., 2013). Utilizing catalytic mutants for G9a and GLP, it was demonstrated that only the catalytic activity of G9a was important for MERVL silencing, despite the observation that depletion of GLP alone leads to robust MERVL de-repression. Further, it was shown that the G9a/GLP heterodimer directly binds to MERVL loci. Importantly, depletion of neither G9a nor GLP causes significant changes in the expression of other known MERVL regulators, ascertaining an independent mechanism of MERVL/ZGA transcript silencing in mESCs (Maksakova et al., 2013).

HP1 proteins are known to play an important role in the regulation of heterochromatin spreading via binding to methylated H3K9 (Al-Sady et al., 2013; Canzio et al., 2013). Importantly, it has been shown that MERVL is sensitive to repression via methylated H3K9me2 (Schoorlemmer et al., 2014). To investigate whether HP1 plays a role in MERVL silencing, researchers generated knockouts of HP1a or HP1b in mESCs (Maksakova et al., 2013). They found that in both knockouts, MERVL elements were among the most highly de-repressed transposable elements. It is unknown whether HP1 directly represses MERVL as ChIP

experiments demonstrated only modest enrichment of HP1 on MERVL elements. In addition, no assays were conducted to assess the potency of HP1 knockout ESCs to verify that these MERVL+ ESCs also have the characteristic of MERVL associated bi-potential cell fate.

Polycomb repressive complexes (PRC) have also been shown to be involved in transposable element silencing in the mouse (Casa and Gabellini, 2012). PRCs typically exist in one of two forms, termed PRC1 and PRC2. The accepted mechanistic model of the action of PRCs firstly involves chromatin associated PRC2, which facilitates H3K27me3. This mark is then bound by PCR1 via its CBX constituent, leading to the deposition of H2AK119Ub1. These marks are inhibitory to transcription, silencing the expression of nearby genes (Kahn et al., 2016). Rybp, a zinc finger protein that associates with PRCs (Rose et al., 2016), has been shown to be essential for mESC pluripotency (Kovacs et al., 2016; Li et al., 2017; Ujhelly et al., 2015). Additionally, in mESCs, Rybp is essential for the repression of MERVL and other ZGA associated transcripts (Schoorlemmer et al., 2014). Additional zinc finger proteins required for MERVL silencing in mESCs have also been described. The factor Rex1, which is expressed throughout preimplantation development and in mESCs, has been shown to directly bind to MERVL loci and depletion of Rex1 in mESCs causes MERVL de-repression (Schoorlemmer et al., 2014). Interestingly, Rex1 shares homology within its DNA-binding zinc fingers with Yy1 and Yy2 and both Yy1 and Yy2 binding motifs were enriched within MERVL in mESCs and mouse trophectoderm stem cells (TSCs), respectively (Kim et al., 2007; Pérez-Palacios et al., 2016). Mechanistically, Rex1 may cooperate with histone deacetylases (HDAC), as Rex1 depletion attenuated the observed de-repression of MERVL in cells treated with HDAC inhibitor, TSA. Additionally, it has been shown that Rex1 directly interacts with Kdm1a, likely cooperating with the factor for MERVL silencing (Schoorlemmer et al., 2014). Consistently, depletion of Rex1 in the context of preimplantation development led to an aberrant increase in MERVL expression shortly after the eight cell stage (Schoorlemmer et al., 2014). While Rybp and Rex1 have been demonstrated as repressors of MERVL, the impact of their depletion on Dux expression and stem cell potency has not been documented.

A role for an additional chromatin modifying enzyme in MERVL regulation has also been described. Chd5, a chromatin modifier that recognizes H3K27me3 and H3K4un has been shown to be involved in MERVL silencing as mESCs depleted for Chd5 have elevated levels of MERVL as well as upregulated expression other ZGA transcripts (Egan et al., 2013; Hayashi et al., 2016). Depletion of Chd5 has minimal impacts on global gene expression. However, it was observed that at discernably upregulated MERVL loci, the deposition of H3.1/H3.2 was increased in Chd5 knockout mESCs (Hayashi et al., 2016). Additionally, H3K27me3 was downregulated proximal to specific MERVL loci. Mechanistically, Chd5 removes H3.1/H3.2, increasing the relative amount of H3.3. Consistently, H3.3 will be targeted for H3K27me3, contributing to MERVL repression (Wang et al., 2018). It is not known whether loss of Chd5 results in an increase in mESC developmental potency, nor how the activity of Chd5 is integrated with other known repressors of MERVL.

Chromatin assembly has also been implicated in the regulation of MERVL in mESCs. Caf-1, a trimeric complex responsible for the placement of H3 and H4 during DNA synthesis (Verreault et al., 1996), has also been shown to be required for MERVL silencing (Hatanaka et al., 2015; Ishiuchi et al., 2015). Depletion of Caf-1 in mESCs led to a strong activation of a MERVL::Gfp reporter. Moreover, this effect was specific to MERVL loci and major satellite repeats. This repression of MERVL was shown to be dependent of Caf-1's chromatin assembly activity via the utilization of several mutant constructs. Caf-1's p150 domain (depletion of which causes MERVL de-repression) contains two noncanonical PIP domains that interact with PCNA, and HP1 interaction motif and an ED domain (Quivy et al., 2008). The first PIP domain (PIP1) is

not required for chromatin assembly, whereas the second PIP domain (PIP2) is required for faithful chromatin assembly while not being involved in DNA synthesis (Ben-Shahar et al., 2009). Finally, the ED domain is required for assembly activity (Ben-Shahar et al., 2009). Knockdown of endogenous Caf-1 p150 while co-delivering the individual mutant constructs revealed that the PIP1 and HP1 domains were dispensable for MERVL silencing (Ishiuchi et al., 2015). Consistent with their known roles, the ED and PIP domains were both required for MERVL repression, ascertaining that Caf-1 mediated silencing of MERVL relies on chromatin assembly. Finally, it was demonstrated that MERVL+ and Caf-1 knockdown MERVL+ cells show greater efficiency when subjected to somatic cell nuclear transfer. While interesting, this observation is slightly cyclic as the Caf-1 knockdown cells used in the assay were sorted based on MERVL expression prior to SCNT and knocking down Caf-1 did not seem to further augment MERVL associated gains in cell potency (Ishiuchi et al., 2015). Intriguingly, it was recently shown that Dux is required for Caf-1 mediated conversion of mESCs to the MERVL+ state as co-depletion of Dux along with Caf-1 does not lead to MERVL de-repression (Hendrickson et al., 2017).

While many epigenetic and transcriptional regulators of transposable elements have been, and will likely continue to be, described, few examples exist of transposable elements mediating silencing of other transposable elements. Recent research has characterized a role for LINE1 in repressing expression of MERVL (Chen et al., 2021; Percharde et al., 2018). In situ hybridization of LINE1 RNA in mESCs, two cell, and blastocyst embryos revealed that LINE1 is enriched in the nucleus. Further, knockdown of LINE1 with antisense oligonucleotides leads to a loss of self-renewal in mESCs. RNA sequencing in LINE1 knockdown mESCs led to the upregulation of genes typically expressed in two cell embryos as well as MERVL. Expectedly, it was shown that Dux targets are well enriched in the genes that are upregulated following LINE1 knockdown and that co-depletion of Dux rescues the repression of these genes. To gain insight as to the reason behind the extensive overlap between Dux targets and genes upregulated following LINE1 knockdown, the investigators performed chromatin immunoprecipitation by RNA purification (ChIRP). Interestingly, this revealed that LINE1 RNA is enriched at the *Dux* locus. Further, RNA immunoprecipitation (RIP)-qPCR showed that LINE1 RNA bound directly to Nucleolin, a gene involved in rDNA transcriptional regulation. Consistently, knockdown of Nucleolin led to similar transcriptional changes as LINE1 depletion. Intriguingly, it was also shown that Nucleolin associates with Kap1, a direct repressor of Dux in mESCs, and thus must cooperate with Kap1 and LINE1 RNA to achieve Dux repression. It was also shown that LINE1 is functional in vivo. Knockdown of LINE1 at the one cell stage caused significant developmental arrest by the two cell stage, with no embryos making it to the blastocyst stage.

It has recently been shown that m^6A modified LINE1 RNA can be bound by the m^6A binding protein YTHDC1 and that depletion of YTHDC1 leads to similar transcriptomic changes when compared to LINE1 knockdown mESCs, with upregulation of LINE1 target and associated genes (Chen et al., 2021). Incredibly, co-immunoprecipitation experiments demonstrated that YTHDC1 directly associates with Nucleolin and Kap1, though the published data for Kap1 only shows very weak enrichment (Chen et al., 2021). This suggests that YTHDC1 is involved in the LINE1-Nucleolin-Kap1 complex. Consistently, Nucleolin association with LINE1 RNA was significantly hampered following depletion of RNA was significantly hampered following depletion of *Ythdc1*. Finally, ChIP seq for Kap1 in *Ythdc1* depleted cells showed that Kap1 occupancy on LINE1 targets was significantly decreased. However, low levels of Kap1 enrichment at MERVL loci was still observed and the significance, if any, of this modest Kap1 enrichment on MERVL has not been well described.

Another modality described for the repression of transposable elements are small, noncoding RNAs. Recent research has shown that MERVL expression is restricted, indirectly, by mir34a (Choi et al., 2017). Previously, it had been observed that mouse embryonic fibroblasts (MEFs) deficient for mir34a have enhanced reprogramming efficiencies compared to wildtype controls (Choi et al., 2011). This is consistent with a recent observation that promoting the two cell transcriptome during chemical reprogramming enhanced efficiencies (Zhao et al., 2018). ESCs deficient for mir34a demonstrated elevated expression of MERVL in a subset of cells. Further, spontaneous differentiation assays, eight cell ESC injection assays and assays looking at ESC contribution to terminally differentiated tissues in ~E12.4 conceptuses ascertained that mir34a knockout ESCs possess bi-potential cell fate. Importantly, there was no need to sort for MERVL+ cells in order to see this expansion of cell fate potential. Researchers stated that approximately 10-20% of cells become MERVL+, yet it was observed that 60% and 25% of mir34a deficient ESCs and iPSCs, respectively, contributed to both embryonic and extraembryonic compartments in chimeric blastocysts. The percentage contribution to post implantation conceptuses is difficult to estimate as multiple ESCs were injected into each recipient blastocyst. The study identified TE marker. Gata2, as a mir34a target that appeared to be required for MERVL de-repression as knockdown of Gata2 in mir34a deficient ESCs attenuated MERVL expression. However, enforced expression of Gata2 was not sufficient for MERVL de-repression (Choi et al., 2017). To date, no additional mir34a targets involved in neither MERVL expression nor bi-potential cell fate has been described. Moreover, it is not clear to what extent other activators of MERVL, such as Dux, depend on Gata2 for MERVL activation and vice versa.

A poorly characterized class of regulatory small RNAs are fragments derived from mature tRNAs, referred to as tRNA fragments (tRFs) (Sharma et al., 2016). These tRFs have been reported to complex with Argonaute proteins, and in that way serve a miRNA-like function (Haussecker et al., 2010). However, Argonaute independent functions have also been described. It was shown that inhibition of the tRF tRNA-Gly-GCC, using antisense LNA oligos, led to the de-repression of MERVL and other ZGA associated transcripts in mESCs (Boskovic et al., 2020). Unexpectedly, there is little homology between the sequence of tRNA-Gly-GCC and the primer binding site within the MERVL LTR, suggesting an indirect mechanism underlying MERVL restriction. Investigation into chromatin architecture in mESCs lacking tRNA-Gly-GCC showed that chromatin accessibility was significantly enhanced in mESCs lacking tRNA-Gly-GCC. Additionally, it was observed that suppression of tRNA-Gly-GCC leads to a commensurate repression of histone mRNAs and that this effect can be reversed by overexpression of tRNA-Gly-GCC in mESCs. Luciferase assays demonstrated that a functional histone 3’ UTR was required for the observed tRNA-Gly-GCC regulation. Intriguingly, another impact of tRNA-Gly-GCC depletion was an observed downregulation of noncoding U7 RNA, that can bind to the 3’ UTR of histone mRNA via and RNA feature termed the histone downstream element (HDE). This is critical as restoration of U7 RNA following depletion of tRNA-Gly-GCC reverses the de-repression of MERVL. Biogenesis of U7 RNA relies on a subnuclear feature termed Cajal bodies, that are a hub for nuclear RNA processing. Additionally, investigators identified hnRNPF/H as direct binding partners of tRNA-Gly-GCC. Consistent with hnRNPF/H and tRNA-Gly-GCC sharing overlapping functional mechanisms, depletion of hnRNPF/H also resulted in de-repression of MERVL as well as disruption of Cajal Body architecture (Boskovic et al., 2020). Taken together, these observations describe a novel mechanism of MERVL expression regulation, however it is ambiguous how the depletion of tRNA-Gly-GCC impacts other known MERVL repressors or activators and if any of these factors have interdependencies with tRNA-Gly-GCC, U7 RNA and hnRNPF/H.

Domestication of Transposable Elements for Host Processes

The evident interconnectedness between MERVL and the factors that confer totipotency in the mouse beget the generalized notion that transposable elements can be co-opted for fundamental host processes. Indeed, reports detail the utilization of transposable element derived promoters, enhancers, RNA, and proteins.

A recent report comprehensively profiled single cell RNA-seq data performed across a variety of species and categorized instances of transposable element – host gene splice fusion events. Particularly, cell cycle regulator *Cdk2ap1* possesses an upstream MT2B2 promoter that drives the expression of a N-terminally truncated isoform that makes up the majority of total *Cdk2ap1* expression during mouse preimplantation development (Modzelewski et al., 2021). The function of the longer, canonical, isoform is to suppress cell proliferation via the degradation of Cdk2 as well as direct repression of Cdk2 kinase activity. Fascinatingly, specific deletion of the shorter, MT2B2 derived isoform of *Cdk2ap1* leads to postnatal lethality with a penetrance of 50-55%. Additionally, mutant embryos showed abnormal morphology, specifically a substantial reduction in total cell number and compromised blastocoel cavities. These aberrancies establish the MT2B2 derived *Cdk2ap1* as the first example of a transposable element promoter with an essential function in mammalian development. Strikingly, specific deletion of the longer, canonical *Cdk2ap1* isoform did not lead to any defects in embryo morphology or cell number. Importantly, deletion of the canonical isoform also leads to postnatal lethality, however this is more likely due to post implantation function of canonical *Cdk2ap1*, as this is the time during development that the canonical isoform is robustly expressed. Further work on the functional importance of other transposable element – host gene fusion events will likely reveal emergent properties of what makes a transposable element suitable for host co-option in this way.

The cis regulatory potential of transposable elements extends to their observed functions as enhancers (Judd et al., 2020). Investigation of circadian regulator (CR) ChIP-seq data sets revealed that between 8.3 and 14.3% of CR peaks overlapped with annotated transposable elements. Interestingly, while several transposable element families demonstrated enrichment among CR bound sites, only RSINE1 displayed significant enrichment across a large number of RSINE1 instance loci, likely mediated by their presence in open chromatin configurations. Further analysis of RSINE1 revealed that its consensus (ancestral) sequence harbors motifs recognized by CRs, and these motifs are specific to the RSINE1 family, likely predisposing it to circadian regulatory function. While RSINE1 appeared to have enhancer function capability, luciferase assays on RSINE1 insertions that overlapped with a published set of enhancer RNAs (eRNAs) revealed that only a minority of the CR bound RSINE1 insertions actually behave as enhancers. Altogether, transposable elements appear to gain transcription factor binding sites over evolutionary time through point mutations that transit precursor motifs present in more ancestral insertions into *bona fide* transcription factor target sites that may then be functionally utilized as enhancers (Judd et al., 2020). However, the discrepancy between transposable elements with specific transcription factor binding sites that are competent to act as enhancers and those that do not despite retaining the same sites requires further exploration.

Beyond acting in cis, transposable elements have been shown to regulate cellular identity in trans via their transcribed noncoding RNA. HERVH is a primate specific transposable element that has been shown to be highly expressed in pluripotent hESCs (Göke et al., 2015; Lu et al., 2014; Santoni et al., 2012; Wang et al., 2016). Particularly, it was found that 231 HERVH insertions are highly expressed in H1 hESCs (Lu et al., 2014). To probe the potential functionality of expressed HERVH RNA, Lu et al. utilized shRNAs against HERVH to mediate its depletion. shRNA mediated downregulation of HERVH resulted in dramatic changes to cellular morphology consistent with differentiation. Consistently, pluripotency genes OCT4, SOX2 and

NANOG were all downregulated and differentiation markers GATA6 and RUNX1 were transcriptionally upregulated. This led to an exploration as to the requirement of HERVH RNA during the acquisition of pluripotency during cellular reprogramming to induced pluripotent stem cells (iPSCs). During the course of normal reprogramming for fibroblasts, HERVH expression was observably high after day 6 of introduction of reprogramming factors. Further, depletion of HERVH during reprogramming led to a decrease in the number of iPSC colonies generated. To elucidate the mechanism through which HERVH RNA promotes pluripotency, researchers performed RNA-CLIP to identify and characterize proteins associated with HERVH RNA. Interestingly, it was found that HERVH RNA associates with transcriptional coactivators. Finally, depletion of HERVH via shRNA had a commensurate suppressive impact on genes that are proximal to HERVH, suggesting that HERVH RNA is required for HERVH enhancer activity. Further work on the requirement of HERVH during the course of preimplantation development, and whether such mechanism exist in other species will be required to comprehensively assess the scope of transposable element derived RNA function.

It is well documented that mammalian viruses co-opt genes from their host due to the selective advantage these host genes confer to the virus. However, this phenomenon can also work in the opposite direction, with mammalian hosts acting as the beneficiary of domesticated viral proteins (Dupressoir et al., 2009, 2011; Levis et al., 1993; Ono et al., 2006; Pastuzyn et al., 2018; Sekita et al., 2008). A landmark study demonstrated that the placental protein, Syncytin, is an envelope protein of the HERVW family (Dupressoir et al., 2011). As part of an attempt to identify novel secreted proteins, a cDNA fragment of the *SYNCITIN* gene was isolated. Protein homology analyses revealed that the encoded protein shared similarity to many retroviral envelope proteins, particular that encoded by HERVW. *In situ* hybridization experiments revealed that *SYNCITIN* is specifically expressed to syncytiotrophoblasts, the products of fusion between cytotrophoblasts of the placenta. This raised the possibility that, due to envelope proteins function in driving cell fusion, *SYNCITIN* may itself mediate cell fusion. Consistently, expression of *SYNCITIN* in COS cells led to cell fusion, with syncytia containing over 30 nuclei. Fascinatingly, inclusion of anti-Syncytin antiserum into cells expressing *SYNCITIN* led to the reduction of cell fusion. Finally, Syncytin gene products are well conserved across Muridae and appear to have similar function in other hosts. For example, Syncytin-A and Syncytin-B in the mouse are also specifically expressed in the placenta and show fusogenic capabilities with Syncytin-A being shown to be required for placentation in the mouse (Dupressoir et al., 2009).

Closing Remarks

The now outdated model of transposable elements and other repetitive genomic elements serving only as junk DNA has now evolved into a model where transposable elements have become entwined in transcriptional networks required for host cell identity. Further, their gene products have demonstrable function, whether as RNA from transposable element derived lncRNAs, or the translated protein products of transposable element RNA. The repetitive nature of transposable elements and their abundance constitute two major hurdles one must overcome when investigating their functionality. However, with advances in high throughput sequencing technologies and genome engineering strategies, transposable element study has never been more tractable. The literature describes several anecdotal instances of transposable element loci having impacts on gene regulation and chromatin architecture, but as of now the emergent property of what makes certain loci or families competent for these functions, and others not, has not been comprehensively investigated.

Additionally, the observation that certain transposable elements can serve as faithful markers of certain cell states has led to systematic application of integrated genomics datasets. ChIP and RNA-seq datasets have provided a framework where the upstream transcriptional

regulators of, really any, transposable element family can be identified. Coupling these with recent large scale and genome wide perturbation screens will be instrumental in comprehensively detailing the various transcriptional networks that harbor transposable elements as one of their nodes.

The greatest challenge will be the study of proteins derived from transposable elements. In the mouse, LINE1, IAP and MERVL are just examples of transposable element families that are still competent to express all, or at least a subset, of their internal proteins. Just as systematic work has been done in understanding the interactome of transposable element derived RNAs, so to must similar approaches be taken on the proteins they encode. Having a comprehensive understanding of what transposable element derived proteins bind to, the cell types they are typically expressed in and access to RNAi or CRISPR based depletion tactics will enable the investigation of more, possibly essential, transposable element derived proteins.

Chapter 2: Klf5 establishes bi-potential cell fate by dual regulation of ICM and TE specification genes

Abstract:

Early blastomeres of mouse preimplantation embryos exhibit bi-potential cell fate, capable of generating both embryonic and extra-embryonic lineages in blastocysts. Here, we identified three major 2 cell (2C) specific endogenous retroviruses (ERVs) as the molecular hallmark of the bi-potential plasticity. Using the LTRs of all three 2C-ERVs, we identified Klf5 as their major upstream regulator. Klf5 is essential for bi-potential cell fate: a single *Klf5-overexpressing* ESC generated terminally differentiated embryonic and extra-embryonic lineages in chimeric embryos, and Klf5 directly induces both ICM and 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. Altogether, 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.

Background

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 2C blastomeres are totipotent, capable of generating all cell types requisite 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); yet cleavage stage blastomeres retain bi-potential cell fate, generating the inner cell mass (ICM) that largely forms the embryo proper, and trophectoderm (TE) that 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 yet transient induction of MERVL endogenous retroviruses (ERVs) (Choi et al., 2017; Macfarlan et al., 2011, 2012), (Ishiuchi et al., 2015a). 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 extraembryonic lineages (Beddington and Robertson, 1989), MERVL induction in rare cell populations often correlates with an expanded cell fate plasticity, enabling differentiation towards both embryonic and extra-embryonic lineages (Choi et al., 2017; Ishiuchi et al., 2015a; 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, the MERVL$^+$ ESCs, designated as bi-potential ESCs in our study, exhibit both embryonic and extraembryonic potency, induce MERVL at a modest level, and functionally resemble bi-potential blastomeres, which are more restricted in developmental potentials than 2C blastomeres.

Since 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 transcriptome landscape of bi-potential cell fate (Alda-Catalinas et al., 2020; Eckersley-Maslin et al., 2019; Hendrickson et al., 2017; Iaco et al., 2017; Yan et al., 2019). The double homeodomain 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; Iaco et al., 2017). Zygotically expressed Dux, as well as its maternally inherited upstream regulators, Dppa2, Dppa4, Nelfa and Smarca5, have all been speculated to

act at the top of the transcriptional hierarchy that governs the onset of ZGA, induces MERVL and regulates 2C-specific cell fate potency (Alda-Catalinas et al., 2020; Eckersley-Maslin et al., 2019; Hu et al., 2020). Yet 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), ((IMPC), 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 Klf5 as an essential regulator that confers a developmental potency to both embryonic and extraembryonic lineages. While 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 remain largely unknown. Here, our studies showed that *Klf5* overexpression in a single, pluripotent ESC confers a 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. Since Klf5 directly induces both ICM and TE specification genes (Lin et al., 2010), our data suggest that the molecular nature of a 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 (Fig. 1A). Using published single cell RNA-seq data on preimplantation embryos, we observed a dynamic and tightly regulated expression pattern of retrotransposons (Fig. 1B, Sup Fig. S1A). While 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 (Fig. 1B, 1C, Sup Fig. 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 (Sup Fig. S1A). Subsequently, all three ERV families rapidly declined by the 8C stage, and were completely silenced in blastocysts (Fig. 1B, 1C, Sup Fig. S1A). Their coordinated induction was confirmed in a subset of bi-potential, MERVL+ ESC lines (Fig. 1D). Specifically, *Lsd1* deletion, and knockdown of Caf-1 subunits *P60* and *P150* in ESCs all yield coordinated induction of these three 2C-specific ERVs (Fig. 1D). Since 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' 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, to MERVL LTR (Sup Fig. S1B). This finding suggests both 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 (Fig. 1E, Sup Table S1), each of which harbors four predicted Klf binding motifs. In contrast, Dux and Gata2 exhibit a binding motif enrichment only within the LTRs of MERVL LTRs, but not ORR1A0/ORR1A1 (Fig. 1E, Sup Table S1). Consistently, in published ESC ChIP-seq data (Hendrickson et al., 2017), no direct Dux binding to ORR1A0 or ORR1A1 elements was detected (Sup Fig. S1C).

Krüppel-like factors (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 (Fig 1F, Sup Fig. S1D). While 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) (Wani et al., 1998).

To determine the Klf (s) capable of directly regulating 2C-specific ERVs, and possibly, the bi-potent cell fate, we constructed luciferase reporters driven by the MERVL LTR (MERVL-Luc), which faithfully recapitulates MERVL expression (Choi et al., 2017; Macfarlan et al., 2011). While Klf5 was able to activate MERVL-luc in HEK cells, neither Klf4 nor Klf17 had a similar effect (Fig. 1G). The MERVL LTR contains 4 predicted Klf5 binding motifs, and transversion mutations of these sites abolished Klf5-dependent luciferase induction (Fig. 1G). To ascertain whether Klf5 directly mediated MERVL-luc induction, we performed chromatin immunoprecipitation (ChIP) using *Klf5-overexpressing* ESCs and confirmed specific Klf5 occupancy on the LTRs of MERVL, ORR1A0, and ORR1A1 (Fig. 1H). Consistently, *Klf5* overexpression specifically upregulated MERVL, ORR1A0, and ORR1A1, along with 2C-specific, MERVL driven 2C gene isoforms (Fig. 1I, 1J). No induction of other retrotransposon families, such as IAP, LINE1, and SINE B1, was observed (Fig. 1I). *Klf5* overexpression in ESCs also activated a tdtomato reporter driven by the MERVL LTR (Sup Fig. S1F). MERVL induction by Klf5 in ESCs was further confirmed by immunostaining for MERVL-Gag protein, with ~15% ESC colonies containing MERVL Gag-positive cells and 5-8% MERVL Gag-positive cells overall (Fig. 1K). In comparison, control ESCs contained <0.1% MERVL+ cells (Fig 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, as MERVL-Gag and Oct expression was mutually exclusive (Fig. 1K). Importantly, *Klf5* overexpression levels in ESCs were comparable to the level of *Klf5* in blastocysts, within a physiologically relevant range (Sup Fig. S1G), and RNA-seq in *Klf5* overexpressing ESCs further confirmed the ability of Klf5 to upregulate 2C-specific genes (Sup Fig. S1H). Altogether, our data establish Klf5 as a direct regulator of all three 2C specific ERVs, including MERVL, ORR1A0 and ORR1A1.

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

Since Klf5 induces MERVL, ORR1A0, and ORR1A1 ERVs, which collectively marked bi-potent blastomeres and ESC populations (Fig. 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*) (Fig. 2A, 2B, Sup Fig. S2A, S2B). In comparison, control teratomas only induced molecular markers of embryonic lineages (Fig. 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 (Fig 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 (Fig 2C, Sup Fig. S2C)

In teratomas and EBs, multiple MERVL+ cells collectively contribute to both embryonic and extraembryonic cell types, making it unclear if 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 extraembryonic lineages (Wigger et al., 2017), (Wu et al., 2017). Hence, we microinjected single, GFP-labeled control or *Klf5-overexpressing* ESCs into each C57BL/6J recipient 8C embryos and analyzed the resulting chimeric blastocysts (Fig. 2D). While control ESCs invariably contributed to the ICM of chimeric blastocysts (Sup Fig. S2D), individual *Klf5-overexpressing* ESCs colonized the ICM, TE, or both in chimeric embryos (Fig. 2D, 2E). Particularly, nearly a third of chimeric blastocysts contained GFP positive progenies from a single *Klf5-overexpressing* ESC in both ICM and TE (Fig. 2D, 2E). Intriguingly, the GFP+ ICM cells derived from *Klf5*-expressing ESCs strongly express Nanog in chimeric blastocysts, similar to their neighboring ICM cells. Yet the GFP+ TE cells exhibit an intermediate phenotype: They localized to the TE compartment, took on a typical TE morphology, lost Nanog expression, but failed to robustly express Cdx2 as their neighboring TE cells (Sup Fig. S2E). This is likely due to the delayed kinetics of *Cdx2* activation upon Klf5-overexpressing ESCs withdrawal from its pluripotent cell fate to take a TE cell fate. These TE cells derived from *Klf5*-overexpressing ESCs, while acquired extra-embryonic potency, have a reduced efficiency for extra-embryonic differentiation compared to normal TE cells.

We next generated 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 (Fig. 2F-2H, Sup Fig. S2F, Sup 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 (Fig. 2F, 2G). In addition, GFP positive cells with large cytoplasmic to nuclear ratio were found proximal to Mtp1 positive syncytiotrophoblast II (SynII) cells, which morphologically resemble sinusoidal trophoblast giant cells (s-TGCs) (Fig. 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 (Sup Fig. S2G-S2H). In all our Klf5 overexpression studies, *Klf5* overexpressing ESCs have a level of *Klf5* expression comparable to that of blastocysts (Sup Fig. S1G), suggesting that the *Klf5* overexpression phenotype we observed in ESCs is caused by a physiologically relevant level of *Klf5* expression. The bi-potential phenotype of *Klf5* over-expressing ESCs documented here is among the best documented, *in*

vivo bi-potential phenotype generated by a single cell, suggesting *Klf5* as an important bi-potential regulator.

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 (Fig. 3A). Interestingly, Klf5 occupancies were significantly enriched for genes regulating ESC pluripotency, Hippo signaling and blastocyst development (Fig. 3A, Sup Table S3), many of which promoted cell fate specification of the ICM or TE and displayed dynamic expression in cleavage stage embryos (Fig. 3B). Specifically, ICM-specific transcription factors (*Nanog* and *Klf4*) and TE-specific transcription factors (*Tead4* and *Cdx2*) all exhibited Klf5 binding in their putative promoter / enhancer regions (Fig. 3C); the corresponding Klf5 peaks invariably harbored multiple predicted Klf binding motifs (Sup Fig S3A). Additionally, Klf5 peaks were also observed in 2C-ERV LTRs that drove 2C-specific gene isoforms (Sup Fig. S3B).

While Klf5 has the ability to induce both 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* (Fig. 3D), but had little effect on TE genes, possibly due to 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, which were absent in control EBs (Fig. 3E). The strong induction of extra-embryonic TE markers, *Cdx2* and *Elf5,* was confirmed by real time PCR (Fig. 3E). Hence, *Klf5-overexpressing* ESCs yielded progenies expressing either ICM or TE genes within an EB (Fig. 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, 2005), (Pfeffer, 2018). These early blastomeres provide a unique cellular context where ICM and TE genes can be co-expressed at a modest level. The ability of Klf5 for dual induction of ICM and TE specification genes likely plays a role in conferring a cell state that allows differentiation into either ICM or TE lineages.

Klf5 expression initiates in bi-potent blastomeres of cleavage stage embryos (Lin et al., 2010) (Fig. 1F). Subsequently, Klf5 exhibits both ICM and TE expression in blastocysts, albeit with a stronger TE enrichment (Sup Fig. S3C). To investigate the role of *Klf5* in cell fate potency during preimplantation development, we efficiently knocked down *Klf5* by 90% using RNA interference (RNAi) in zygotes (Sup Fig. 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 (Fig 3F). Immunofluorescence staining revealed a marked reduction of Cdx2 in E4.5 *Klf5* knockdown embryos (Fig 3G), consistent with a significant decrease in major TE specification genes (*Cdx2*, *Elf5* and *Tead4*) and placenta early development gene (*Esrrb*) as shown by real time PCR in control versus knockdown embryos (Sup Fig. S3D). In contrast, *Klf5* knock down alone failed to affect ICM specification genes or pluripotency genes *in vivo* (*Sox2*, *Oct4* and *Nanog*) (Sup Fig. 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 (Fig 3H).

To confirm these findings, we generated blastocyst embryos with a complete *Klf5* disruption, using CRISPR-mediated *Klf5* editing (Sup Fig. S3E). In *Klf5* knockout embryos, we detected a significant loss of Cdx2, but Oct4 expression was relatively intact (Sup Fig S3E). Hence, while *Klf5* has the ability to induce both ICM and TE specification genes, *Klf5* deficiency preferentially impairs TE cell fate, consistent with a strong Klf5 enrichment in the TE compartment (Sup Fig. S3C). Our findings seemingly differ from a previous study, which 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 the ICM defects are rather mild with incomplete penetrance (Azami et al., 2017; Lin et al., 2010). Hence, while Klf5 can induce both 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 both ICM and TE, with a strong TE enrichment; *Klf4* is specifically enriched in the ICM, and its ICM expression level is comparable to that of *Klf5* (Fig. 3I). Consistent with this expression pattern, *Klf4* knockout in mice exhibits no obvious defects in preimplantation development (Katz et al., 2002), yet knocking down both *Klf5* and *Klf4* impairs both ICM and TE cell fate during the morula to blastocyst transition, as shown by defective Nanog and Cdx2 expression (Fig. 3J, Sup Fig. S3F). Interestingly, *Klf5* knockdown failed to alter Yap-1 staining in embryos, suggesting that the Hippo signaling is not regulated by *Klf5* during preimplantation development (Sup Fig. S3G). This is consistent with Klf5 enabling cell fate potency for both the ICM and the TE lineages, rather than specifying the TE lineages. Altogether, Klf transcription factors constitute a functional 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 to enable ICM specification.

Interestingly, the ICM expression of Klf5 is observed in the epiblast as well as the PrE lineage in E4.0 blastocysts, as Klf5 and Sox17 are co-expressed in cells of the PrE compartment (Sup Fig. 3H). Real time PCR analyses of *Klf5* knockdown embryos also exhibited a marked downregulation of PrE specification genes *Gata6*, *Gata4* and *Sox17 (*Sup Fig 3H). While this is consistent with the induction of PrE gene markers in *Klf5*-overexpressing EBs and teratomas (Fig. 2B), the decreased PrE gene expression in *Klf5* knockdown embryos can also be caused by their developmental arrest/delay. Taken together, our data are consistent with Klf5 acting upstream of the induction of PrE genes. Our findings contrast a previous study, which reported Klf5 as a suppressor for PrE specification (Azami et al., 2017). This study showed an increased percentage of PrE cells in *Klf5* knockout blastocyst embryos. Yet the significant decrease of the total cell number and the developmental arrest of *Klf5* knockout blastocysts have complicated the interpretation of this result, making it difficult to conclude Klf5 as a suppressor of the 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 the most direct phenotype of Klf5 deficient embryos is the 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* (Fig. 4A, Sup Table S4). These putative Klf5 regulators were subjected to Ingenuity Pathway Analysis (IPA), and an enrichment emerged for genes regulating stem cell biology and preimplantation development (Sup Fig. 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 (Fig. 4A). In particular, Dux, the key regulator for 2C-like transcriptome (Eckersley-Maslin et al., 2019; Hendrickson et al., 2017; Hu et al., 2020; Iaco et al., 2017; Ishiuchi et al., 2015a), showed a strong ChIP-seq peak in *Klf5* intron 1 (Fig. 4A). Using published RNA-seq data on *Dux*-overexpressing ESCs (Hendrickson et al., 2017), we showed that *Dux* overexpression upregulates *Klf5* (Fig. 4B). In addition to *Dux*, *Dppa2* also induced *Klf5* expression and MERVL expression in ESCs (Fig. 4B) and this induction was independent of Dux. Finally, *Tbx3* overexpression in ESCs also induced *Klf5* (Fig. 4B). Hence, multiple 2C-specific transcription factors converge their direct regulation on the 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 (Fig. 4C). More importantly, EBs derived from *Klf5-overexpressing Dux*$^{-/-}$ ESCs, but not control *Dux*$^{-/-}$ ESCs, induced markers for both extraembryonic lineages (TE marker *Cdx2* and PrE marker *Gata4*) and embryonic lineages (Ectoderm marker *Pax6,* Mesoderm marker *Brachyury* and Endoderm marker *Foxa2*) (Fig 4D, Sup. Fig. S4C). And a single *Klf5-overexpressing Dux*$^{-/-}$ ESC was able to colonize the ICM, TE or both in 44%, 22% and 33% chimeric blastocysts, respectively (Fig. 4E). However, it is possible that other interactions between *Dux* and *Klf5* exist as the level of MERVL induction and induction of *Cdx2* upon EB differentiation were somewhat dampened in comparison to wildtype 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 both induced at the onset of ZGA to govern the induction of early zygotic genes (Hendrickson et al., 2017; 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* over-expressing human ESCs (hESCs) (Hendrickson et al., 2017). Intriguingly, human *KLF5* was significantly upregulated by *DUX4* in RNA-seq analysis*,* yet a subset of bi-potency regulators and known Dux targets in mouse, such as *NELFA* and *DPPA2*, were not impacted by enforced *DUX4* expression (Fig. 4F). This is consistent with ChIP-seq data, as 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 (*Fig 4G). The evolutionary conservation of the Dux/Klf5 axis, but not Dux/Dppa2 or Dux/Nelfa axis, highlights the functional importance of Klf5 in establishing a bi-potential cell fate in early blastomeres.

While 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 ((IMPC), 2014; Chen and Zhang, 2019; Nakamura et al., 2011). This contrasts to the strong ICM and TE specification defects when *Klf5* and *Klf4* are both knocked down (Fig. 3I, 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 establish bi-potential cell fate (Fig. 4H).

Discussion

Bi-potential blastomeres exhibit an intrinsic transcriptional program co-expressing both 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 either ICM or TE specification genes, in a context-dependent manner. To elucidate the key molecular pathway regulating the bi-potent cell fate, we identified 3 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, as the endogenous *Klf5* expression in pluripotent ESCs is not sufficient to confer bi-potential cell fate, yet an elevated expression Klf5 level is. While 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 present 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 both ICM and TE, with a strong enrichment in TE. *Klf5* deficiency alone impairs TE specification, and *Klf5* and *Klf4* deficiency in combination impairs both TE and ICM specification (Fig. 4H).

While 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 both ICM and TE, with a knockout phenotype impairing blastocyst cell fate specification (Azami et al., 2017; M. 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 both 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 background. These observations are intriguing, as Klf5 induces both ICM and TE genes, yet its deficiency preferentially impacts 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 2C-specific transcriptome, including Dux, Dppa2 and Tbx3. Unlike *Klf5*, none of these 2C transcription factors exhibit strong preimplantation phenotype in knockout studies ((IMPC), 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 on developmental plasticity converges on Klf5, and possibly other Klf genes. In particular, Dux regulation on Klf5 is evolutionarily conserved between mouse and human, but a number of well characterized, 2C-specific Dux targets in 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, which may not be always associated with 2C-ERV induction. Expanded pluripotent stem cells (EPCs), derived either 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.

Materials and Methods

RNA-seq

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., 2017b). 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 (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 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 $MERVL_{125\text{-}375}$-fragment previously described (MERVL-Luc) (Choi et al., 2017b). 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 37ºC 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, catalogue no. 10829-018), 15% ES-grade fetal bovine serum (Invitrogen, Cat #. 16141079), 2 mM L-glutamine (Invitrogen, Cat #. 25030-164), 1×10^{-4} M MEM non-essential amino acids (Invitrogen, Cat #. 11140-076), 1×10^{-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 □g/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^{o}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 Supplementary Table S5.

Teratoma generation and histological analyses

1x10^{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 µm thickness, and stained with hematoxylin and eosin (H&E) using standard procedures (Choi et al., 2011b). 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 wildtype ESCs, *Klf5*-overexpressing *dux*$^{-/-}$ 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 wildtype ESCs into C57Bl/6N wildtype 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, immunoprecipitated 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, Biosystems 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 □l 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 with the same options as above. Bam files for Klf5 ChIP samples were merged separately. Peaks were then called using (Zhang et al., 2008). Coverage of the

peaks by the initial bam files was then computed using betools_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 afterwards) 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, GGUCAGCACCCGCGUGGGCA, 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.

Supplementary Tables

Supplementary Table S1. Known motif analysis performed using all ORR1A0, ORR1A1 and MT2_Mm LTR sequences compared against scrambled control sequences. HOMER motif analysis results is performed for MERVL-LTR (MT2_MM), ORR1A0 LTR and the ORR1A1 LTR. Analyses on separate families as well as the cohort together are shown.

https://drive.google.com/file/d/1UqD-wLqZkUR4AKaPLDQjx640fcSLGU7_/view?usp=sharing

Supplementary Table S2. Phenotypical analyses of chimeric embryos generated from wild-type and *Klf5*-overexpressing ESCs.

https://drive.google.com/file/d/1WKr6GEHa3XDCsuHfRvLHI4E8SoJRNnXn/view?usp=sharing

Supplementary Table S3. Annotated peaks from Klf5 ChIP-seq experiments using Klf5-overexpressing ESCs. A list of Klf5 bound regions are shown in BED format. The coordinates are annotated with exact feature information, nearest gene transcriptional start site and the number of Klf5 motifs within the peak.

https://drive.google.com/file/d/1zf1hGS-BaccHidXmcETLydlGtQ5vb6vl/view?usp=sharing

Supplementary Table S4. A summary of transcription factors with proximal binding to cis-regulatory sequences of *Klf5*

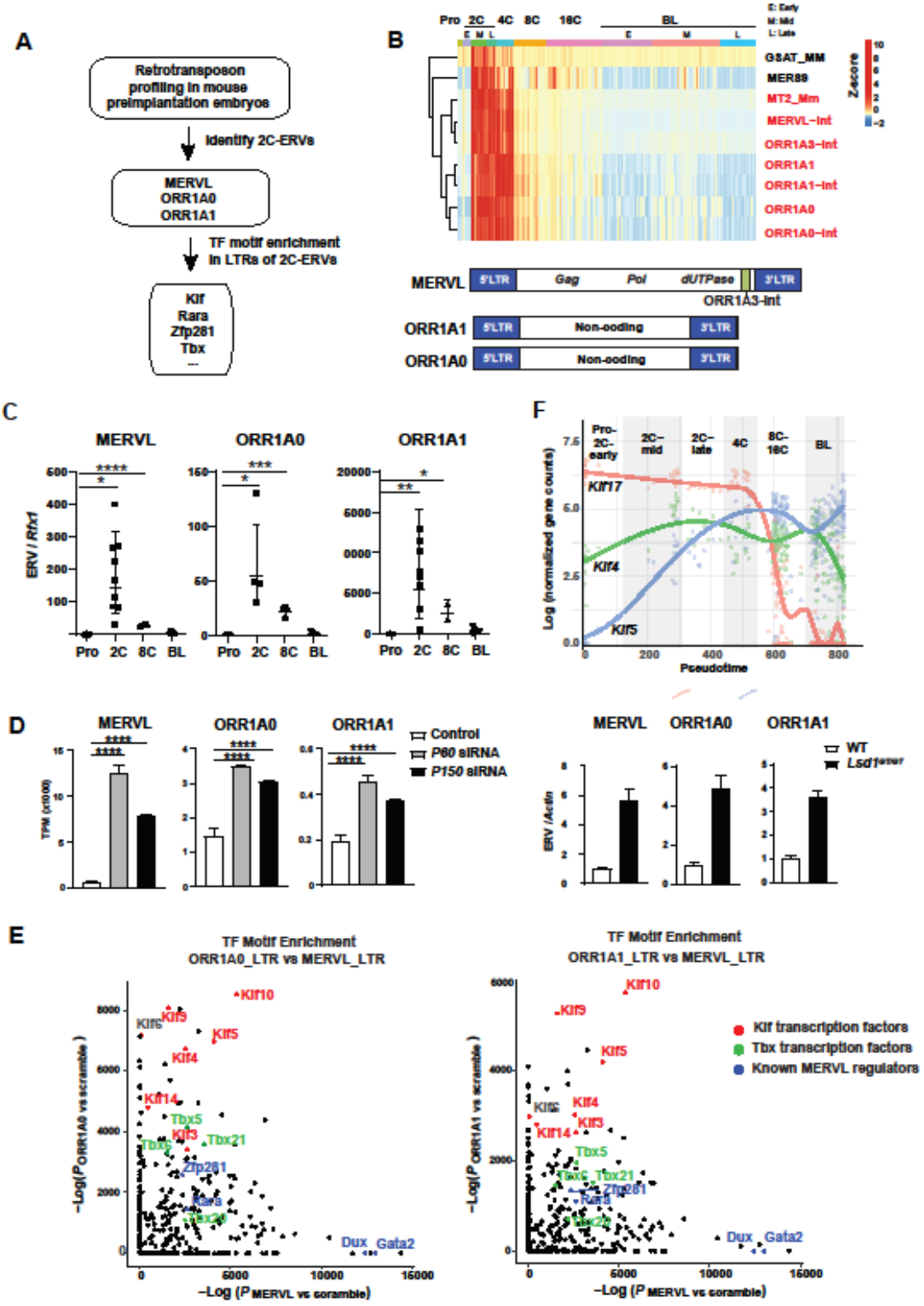
https://drive.google.com/file/d/1UXbxa-nMJNrvKVqMvo_2saXT9uzbXBk4/view?usp=sharing

Supplementary Table S5. Sequences of real-time PCR primers and sgRNAs.

https://drive.google.com/file/d/1htNn4v4AujoAeMHVl9NS67ATdAAlc5cw/view?usp=sharing

Figures

Figure 1


A
Retrotransposon profiling in mouse preimplantation embryos
Identify 2C-ERVs
MERVL
ORR1A0
ORR1A1
TF motif enrichment in LTRs of 2C-ERVs
Klf
Rara
Zfp281
Tbx
B
Pro 2C 4C 8C 16C BL
GSAT_MM
MER88
MT2_Mm
MERVL-int
ORR1A3-int
ORR1A1
ORR1A1-int
ORR1A0
ORR1A0-int
Zscore
MERVL 5'LTR Gag Pol dUTPase 3'LTR
ORR1A3-int
ORR1A1 5'LTR Non-coding 3'LTR
ORR1A0 5'LTR Non-coding 3'LTR
C
MERVL
ORR1A0
ORR1A1
ERV / Rfx1
Pro 2C 8C BL
D
MERVL
ORR1A0
ORR1A1
Control
P60 siRNA
P150 siRNA
TPM (x100)
WT
Lsd1 mut
ERV /Actin
E
TF Motif Enrichment
ORR1A0_LTR vs MERVL_LTR
TF Motif Enrichment
ORR1A1_LTR vs MERVL_LTR
Klf10
Klf9
Klf6
Klf5
Klf4
Klf14
Klf3
Tbx5
Tbx21
Tbx3
Zfp281
Rara
Tbx20
Dux
Gata2
Klf transcription factors
Tbx transcription factors
Known MERVL regulators
−Log(P ORR1A0 vs scramble)
−Log(P ORR1A1 vs scramble)
−Log (P MERVL vs scramble)
F
Pro-2C-early
2C-mid
2C-late
4C
8C-16C
BL
Klf17
Klf4
Klf5
Log (normalized gene counts)
Pseudotime

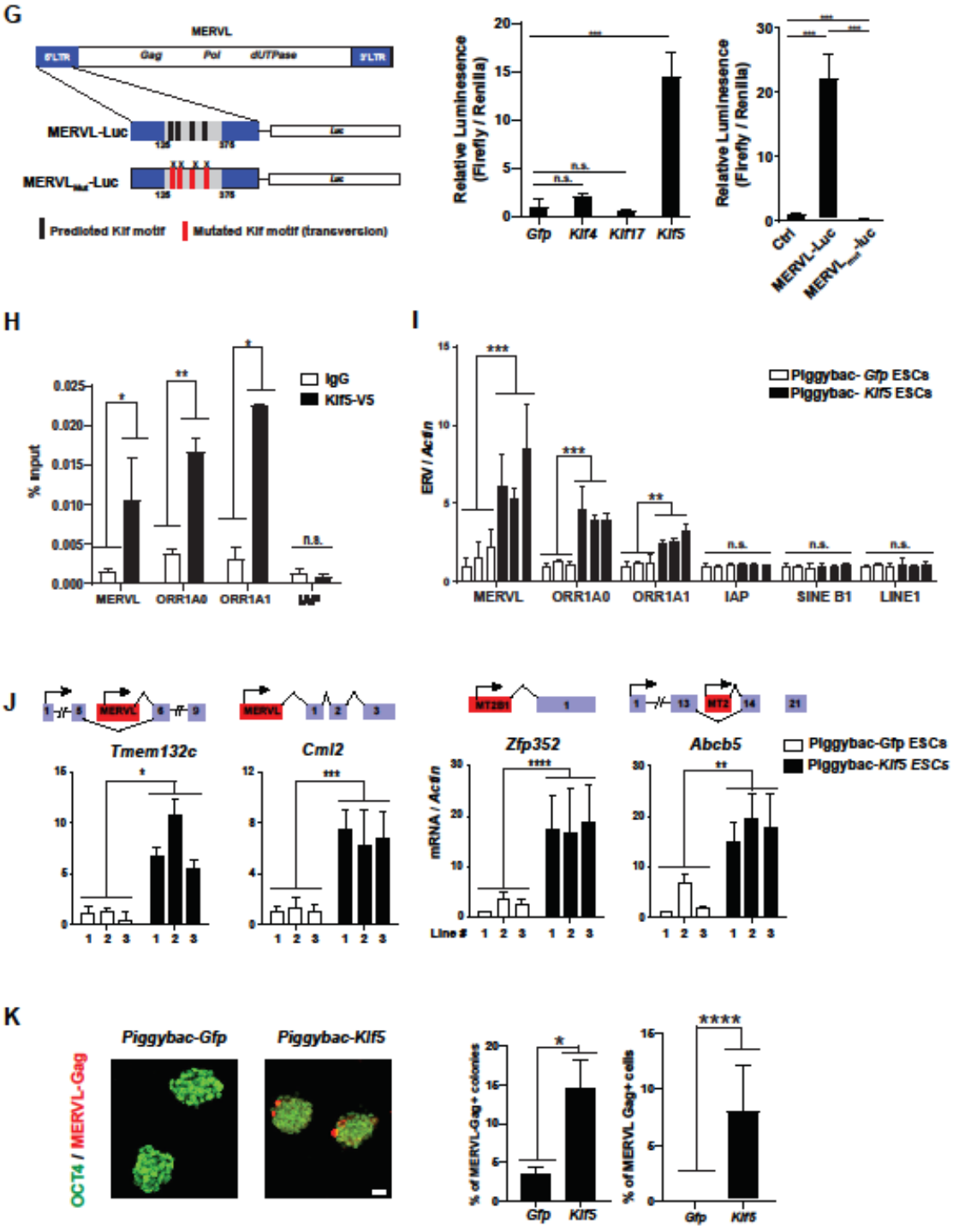
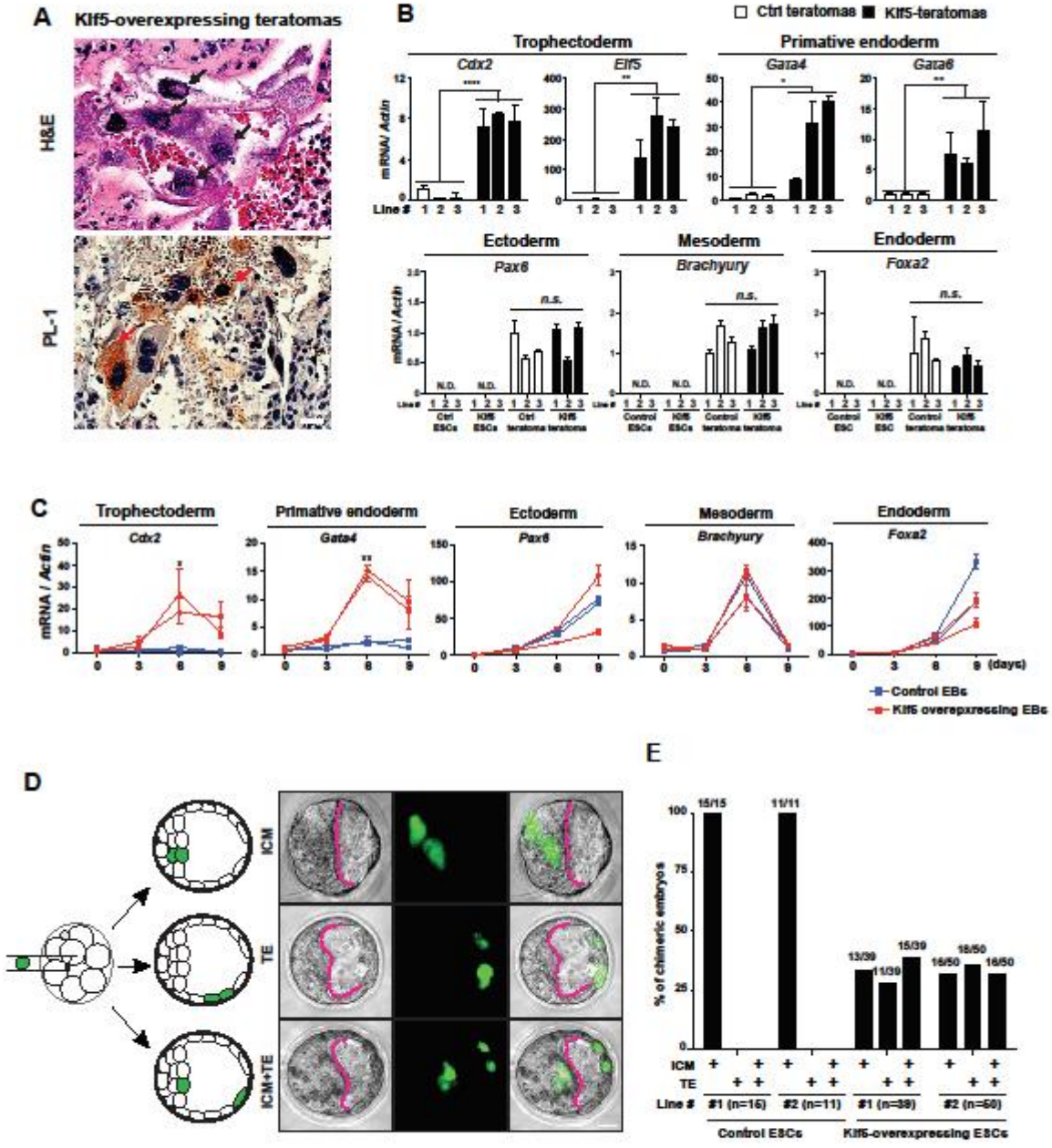

G
MERVL
5'LTR
Gag
Pol
dUTPase
3'LTR
MERVL-Luc
Luc
Predicted Klf motif
Mutated Klf motif (transversion)
Relative Luminescence (Firefly / Renilla)
Gfp
Klf4
Klf17
Klf5
Ctrl
MERVL-Luc
H
IgG
Klf5-V5
% input
MERVL
ORR1A0
ORR1A1
I
Piggybac- Gfp ESCs
Piggybac- Klf5 ESCs
ERV / Actin
IAP
SINE B1
LINE1
J
Tmem132c
Cml2
Zfp352
Abcb5
Piggybac-Gfp ESCs
Piggybac-Klf5 ESCs
mRNA / Actin
Line #
K
Piggybac-Gfp
Piggybac-Klf5
OCT4 / MERVL-Gag
% of MERVL-Gag+ colonies
% of MERVL Gag+ cells

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, 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 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, sixteen-cell stage; BL, blastocyst. (Bottom) Gene structures were 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 blastocyst embryos (n= 5) experimentally validated the 2C-specfic induction of MERVL (Pro vs 2C, **P* = 0.0116, t = 3.024, df = 11; Pro vs 8C *****P* < 0.0001, t = 13.57, df = 5), ORR1A0 (Pro vs 2C **P* = 0.0295, t = 2.843, df = 6; Pro vs 8C ****P* = 0.0002, t = 9.278, df = 5) and ORR1A1 (Pro vs 2C ***P* = 0.0063, t = 3.447, df = 10; Pro vs 8C **P* = 0.0103, t = 4.567, df = 4). Error bars = s.d. **D**. MERVL, ORR1A0 and ORR1A1 are coordinately induced in multiple lines of 2C-like ESCs. (Left) Upon bioinformatic analyses of published RNA-seq data (Ishiuchi et al., 2015b), *Caf-1* deficient ESCs, including *P60* knockdown and *P150* knockdown ESCs, exhibited a coordinated induction of MERVL, ORR1A0 and ORR1A1 (MERVL, control vs *P60* knockdown, adjusted *P* = 1.99e-224, control vs *P150* knockdown, adjusted *P* =2.14e-212; ORR1A0 , control vs *P60* knockdown adjusted *P* = 9.79e-64, control vs *P150* knockdown, adjusted *P* = 2.67e-86; ORR1A1, control vs *P60* knockdown, adjusted *P* = 1.92e-81, control vs *P150* knockdown, adjusted *P* = 2.68e-86). Error bars = s.d. P value was computed with the DESeq2 package in R. (Right) Real time PCR analyses confirmed a 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 to 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. Pseudotime expression plot showing the dynamic expression profiles of *Klf17, Klf4* and *Klf5* across preimplantation development. Klf5 exhibited an early induction that peaked at 4C stage, 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 (125bp – 375bp) 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, 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 = s.d.; miniP-luc vs MERVL-luc, *P* = 0.0005, df = 4, t = 10.39; MERVL-luc vs 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 a robust activation of the MERVL-associated transcriptome. **H.** Klf5 specifically occupies to 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, s.d.; 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, s.d.; 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 a 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 positive ESC colonies (ESC colonies with 1 or more MERVL-Gag positive cells) and the percentage of MERVL-Gag positive cells in the total population were each quantified using fluorescent staining (Right). Scale bar, 20μm. error bars = s.d.. % of MERVL Gag positive colonies: **P* = 0.0405, df = 4, t = 2.987. % of MERVL Gag positive cells: *****P* < 0.0001, df = 32, t = 6.712. * *P* < 0.05, ** *P* < 0.01, *** *P* < 0.001, **** *P* < 0.0001, *n.s*., not significant. All *P* values were computed using unpaired, two-tailed student's t-test unless otherwise indicated.

Figure 2


A
Klf5-overexpressing teratomas
H&E
PL-1
B
Ctrl teratomas
Klf5-teratomas
Trophectoderm
Cdx2
Elf5
Primative endoderm
Gata4
Gata6
Ectoderm
Pax6
Mesoderm
Brachyury
Endoderm
Foxa2
mRNA/ Actin
n.s.
N.D.
Line #
Ctrl ESCs
Klf5 ESCs
Ctrl teratoma
Klf5 teratoma
Control ESC
Klf5 ESC
Control teratoma
Klf5 teratoma
C
Trophectoderm
Cdx2
Primative endoderm
Gata4
Ectoderm
Pax6
Mesoderm
Brachyury
Endoderm
Foxa2
mRNA / Actin
(days)
Control EBs
Klf5 overepxressing EBs
D
ICM
TE
ICM+TE
E
% of chimeric embryos
15/15
11/11
13/39
11/39
15/39
16/50
18/50
16/50
ICM
TE
Line #
#1 (n=16)
#2 (n=11)
#1 (n=39)
#2 (n=50)
Control ESCs
Klf5-overexpressing ESCs

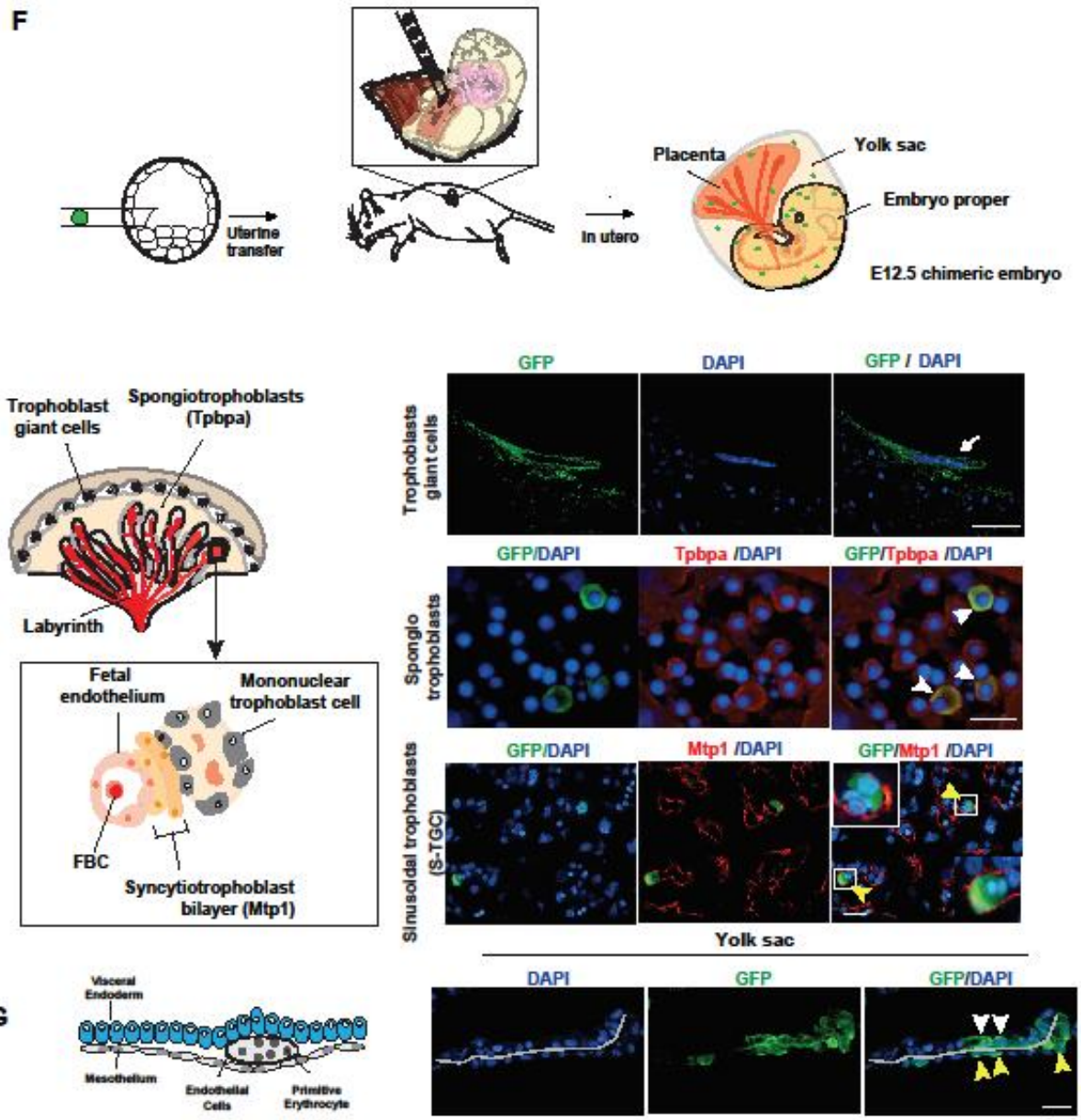
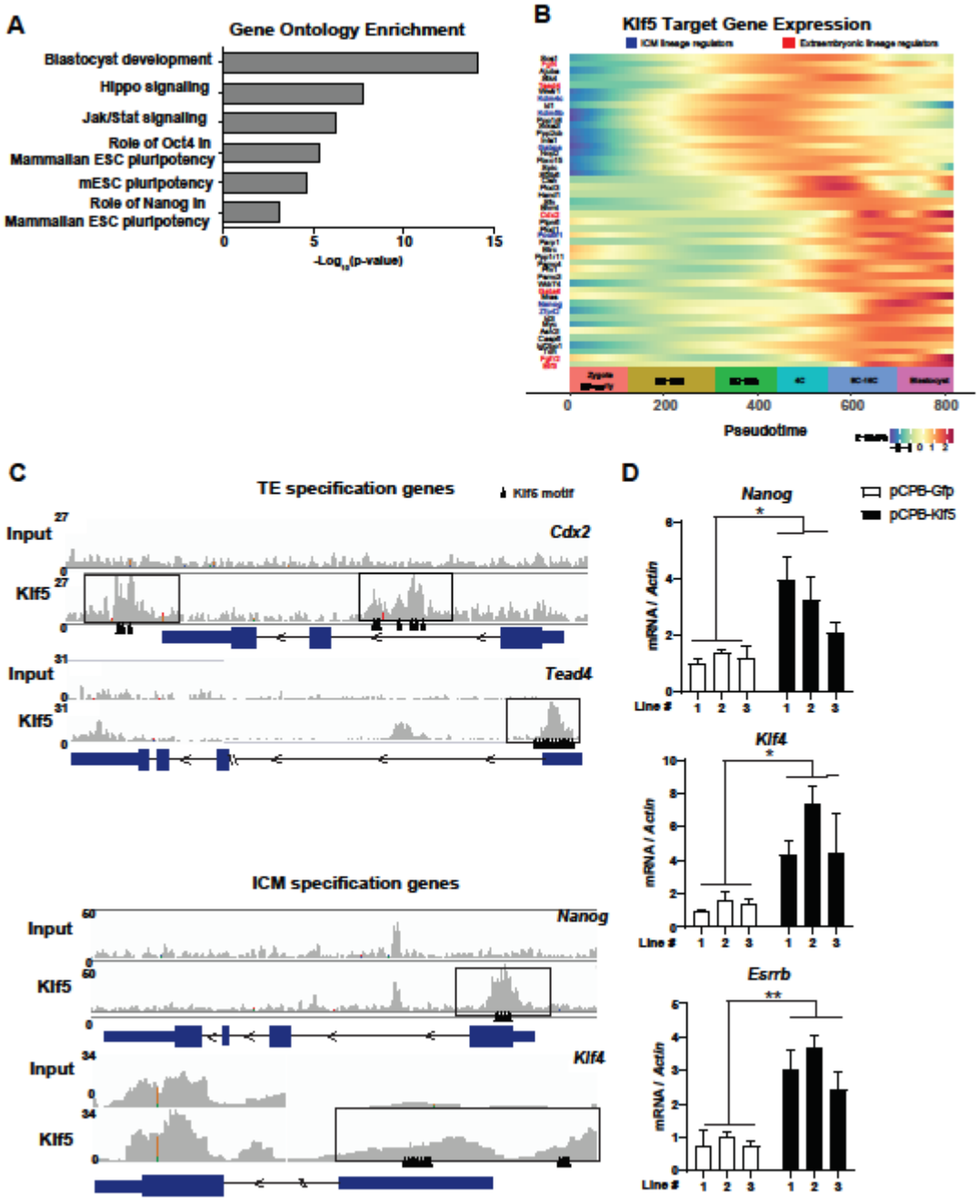

F
Uterine transfer
In utero
Placenta
Yolk sac
Embryo proper
E12.5 chimeric embryo
Trophoblast giant cells
Spongiotrophoblasts (Tpbpa)
Labyrinth
Fetal endothelium
Mononuclear trophoblast cell
FBC
Syncytiotrophoblast bilayer (Mtp1)
GFP
DAPI
GFP / DAPI
Trophoblasts giant cells
GFP/DAPI
Tpbpa /DAPI
GFP/Tpbpa /DAPI
Sponglo trophoblasts
GFP/DAPI
Mtp1 /DAPI
GFP/Mtp1 /DAPI
Sinusoidal trophoblasts (S-TGC)
Yolk sac
G
Visceral Endoderm
Mesothelium
Endothelial Cells
Primitive Erythrocyte
DAPI
GFP
GFP/DAPI

Figure 2. Klf5 overexpression confers a bi-potential cell fate in ESCs both *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 trophoblast giant cell 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 (PE) markers (*Gata4* and *Gata6*) were highly induced in *Klf5-overexpressing* teratomas in real-time PCR analyses, but not in control teratomas. In contrast, the expression of *Pax6 (*an ectoderm marker), *Brachyury* (a mesoderm marker) and *Foxa2* (an endoderm marker) are similarly induced in control and *Klf5-overexpressing* teratomas. Teratomas were generated from three independent pairs of passage-controlled control and *Klf5-overexpressing* ESC lines. Error bars: *s.d*.; *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 a significant induction in 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, *s.d.*; *Cdx2* (day 6), * $P = 0.0372$, df = 2, t = 5.036; *Gata4* (day 6), ** $P = 0.0019$, df = 2, t = 23.20. **D, E.** Single *Klf5-overexpressing* ESCs confer bi-potential cell fate in chimeric blastocyst embryos**.** Representative images (**D**) and quantitation (**E**) of chimeric blastocyst embryos with ESC contribution to the ICM, the TE, or both. **D.** A diagram illustrates the experimental strategy to generate chimeric blastocysts embryos by microinjecting a single GFP-labeled, *Klf5-overexpressing* ESC into each C57BL/6N recipient morula. ESC contribution to the ICM and/or the TE was determined by the localization of GFP-positive ESC progenies in fluorescence imaging. Scale bar, 20 μm. **D, E.** Chimeric blastocysts were generated from two independent pairs of passage-controlled control and *Klf5-overexpressing* ESCs lines. **F, 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 blastocysts, followed by embryo transfer into 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* overexpressed ESC generated trophoblast giant cells with a characteristic cellular morphology (white arrows), spongiotrophoblasts with Tpbpa expression (white arrowheads), and syncytio trophoblasts with Mtp1 expression (yellow arrowheads). Scale bars, 100μm. **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 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; * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, *n.s*., not significant. All *P* values were computed using unpaired, two-tailed student's t-test.

Figure 3


A
Gene Ontology Enrichment
Blastocyst development
Hippo signaling
Jak/Stat signaling
Role of Oct4 in Mammalian ESC pluripotency
mESC pluripotency
Role of Nanog in Mammalian ESC pluripotency
0
5
10
15
-Log10(p-value)
B
Klf5 Target Gene Expression
ICM lineage regulators
Extraembryonic lineage regulators
0
200
400
600
800
Pseudotime
C
TE specification genes
Klf5 motif
Input
Klf5
Cdx2
Tead4
ICM specification genes
Nanog
Klf4
D
pCPB-Gfp
pCPB-Klf5
Nanog
Klf4
Esrrb
mRNA / Actin
Line #
1
2
3

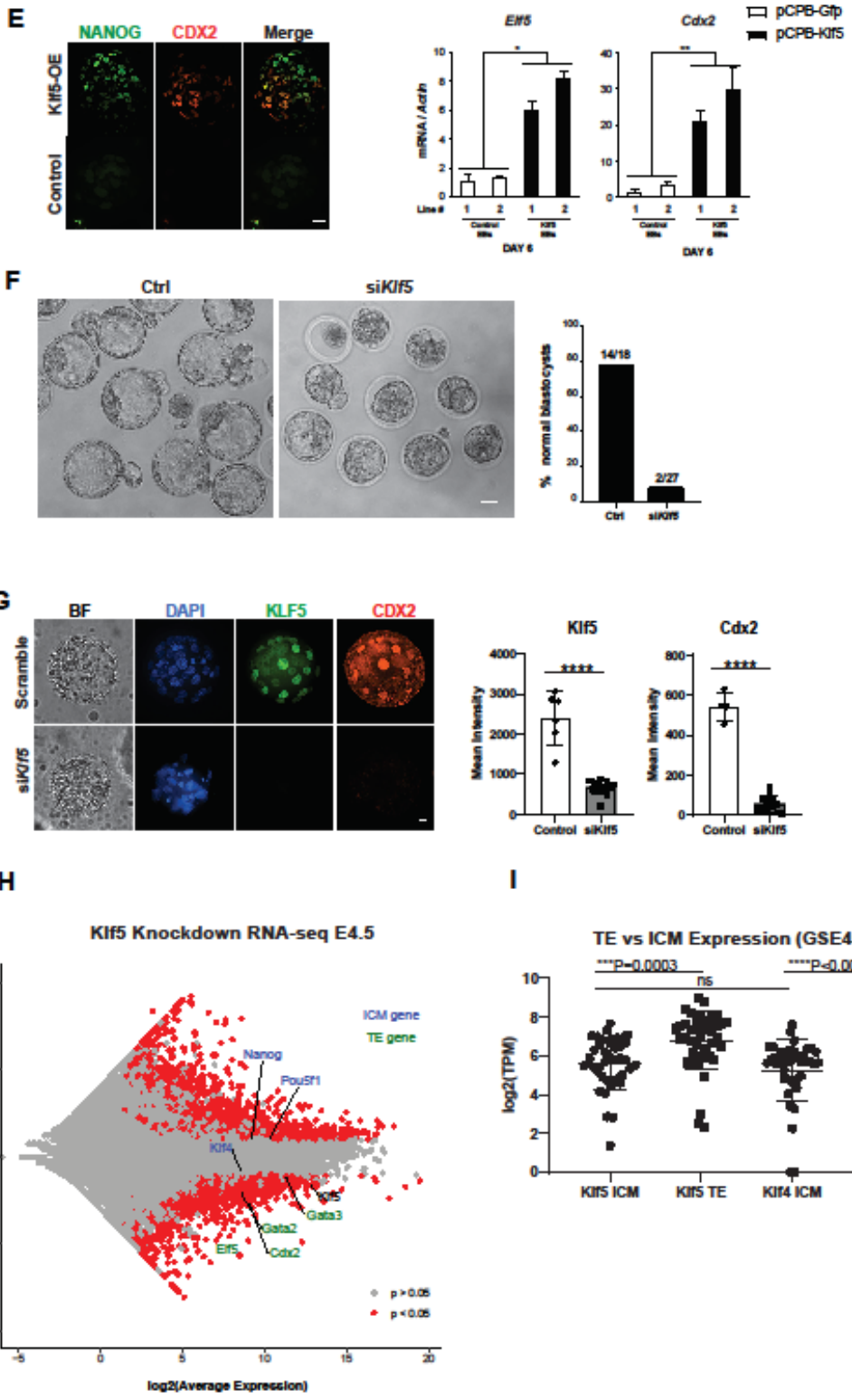
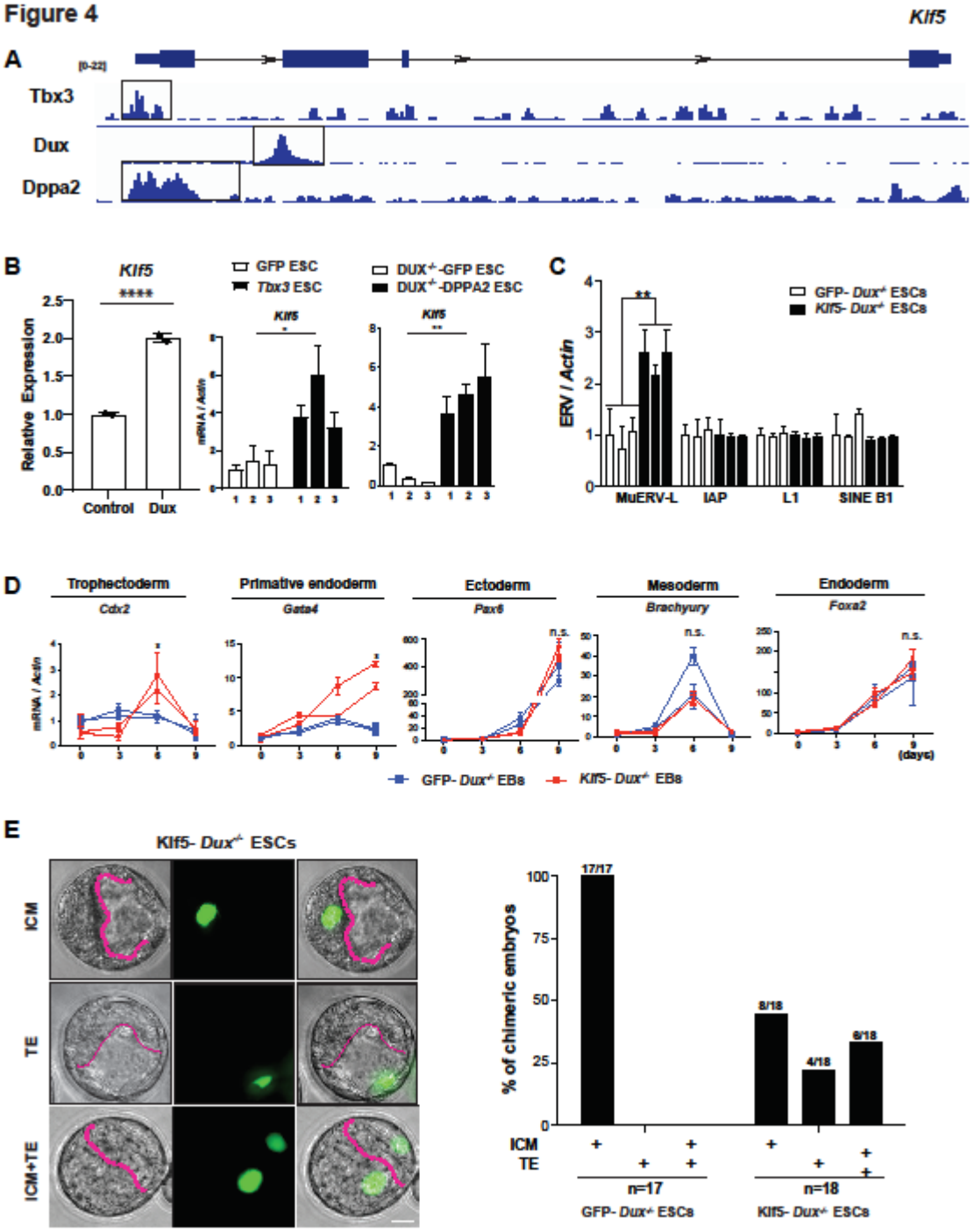
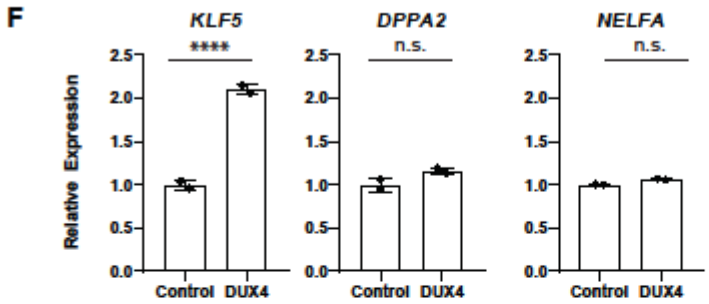
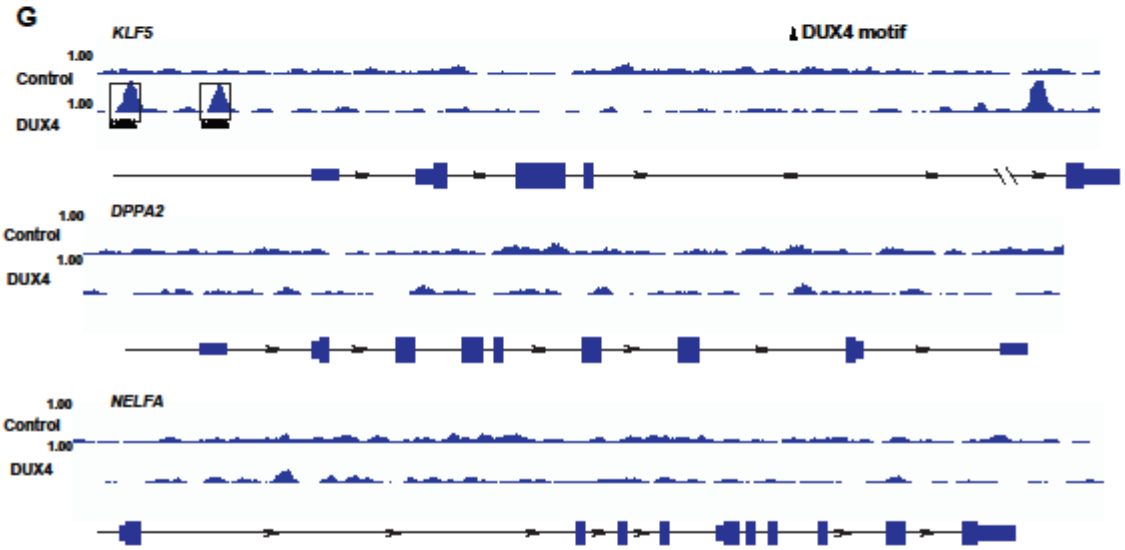

E
NANOG
CDX2
Merge
Klf5-OE
Control
Elf5
Cdx2
pCPB-Gfp
pCPB-Klf5
mRNA / Actin
Line #
DAY 6
F
Ctrl
siKlf5
% normal blastocysts
14/18
2/27
G
BF
DAPI
KLF5
CDX2
Scramble
siKlf5
Klf5
Cdx2
Mean Intensity
Control
siKlf5
H
Klf5 Knockdown RNA-seq E4.5
ICM gene
TE gene
Nanog
Pou5f1
Klf4
Klf5
Gata3
Gata2
Elf5
Cdx2
p > 0.05
p < 0.05
log2 FoldChange [Klf5-KD / Control]
log2(Average Expression)
I
TE vs ICM Expression (GSE45719)
***P=0.0003
ns
****P<0.0001
log2(TPM)
Klf5 ICM
Klf5 TE
Klf4 ICM
Klf4 TE

Figure 3. Klf5 promotes dual induction of cell fate specification genes for both embryonic and extraembryonic commitment. A. Ingenuity pathway analysis 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 = s.d. **E.** EBs derived from *Klf5-overexpressing* ESCs contain Cdx2-positive cells as well as Nanog-positive cells, whereas control EBs failed to activate Cdx2 and completely lost Nanog expression. Real time PCR analyses confirmed the induction of TE markers (Cdx2 and Elf5) in *Klf5-overexpressing* EBs. Scale bars, 100 μm (left). Error bars, s.d..; *Cdx2*, ***P* = 0.0372, df = 2, t = 5.036; *Elf5*, **P* = 0.0339, df = 2, t = 5.289. **F, G.** *Klf5* knockdown stalls preimplantation development before the blastocyst stage and abolishes Cdx2 expression in TE. **F.** Representative bright field images (left) and quantitation (right) are shown for E4.0 *Klf5* knockdown (siKlf5) or control (scramble siRNA) blastocyst embryos. **G**. Representative Cdx2 immunofluorescence images (left) and relative fluorescence quantitation (right) are shown for control and *Klf5* knockdown blastocyst embryos. 3 independent *Klf5* knockdown experiments were performed using a total of 18 control embryos and 27 *Klf5* knockdown embryos. Scale bar = 20uM; 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 TE-marker genes: *Cdx2, Elf5, Gata2* and *Gata3* whereas expression of ICM genes *Klf4, Nanog* and *Pou5f1* is not significantly impacted. **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 comparable levels to *Klf5* in the ICM and is expressed 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 = 20uM.

Figure 4


A
Klf5
Tbx3
Dux
Dppa2
B
Klf5
****
Relative Expression
Control
Dux
GFP ESC
Tbx3 ESC
DUX+-GFP ESC
DUX+-DPPA2 ESC
mRNA / Actin
C
ERV / Actin
GFP- Dux+ ESCs
Klf5- Dux+ ESCs
MuERV-L
IAP
L1
SINE B1
D
Trophectoderm
Cdx2
Primative endoderm
Gata4
Ectoderm
Pax6
Mesoderm
Brachyury
Endoderm
Foxa2
n.s.
(days)
GFP- Dux+ EBs
Klf5- Dux+ EBs
E
Klf5- Dux+ ESCs
ICM
TE
ICM+TE
% of chimeric embryos
17/17
8/18
4/18
6/18
n=17
n=18
GFP- Dux+ ESCs
Klf5- Dux+ ESCs


F
KLF5
DPPA2
NELFA
****
n.s.
n.s.
Relative Expression
2.5
2.0
1.5
1.0
0.5
0.0
Control
DUX4


G
KLF5
DUX4 motif
Control
DUX4
1.00
DPPA2
NELFA

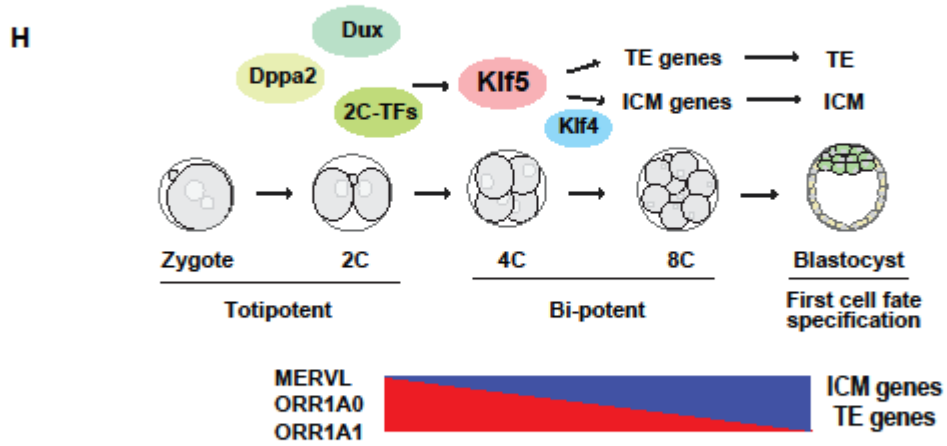
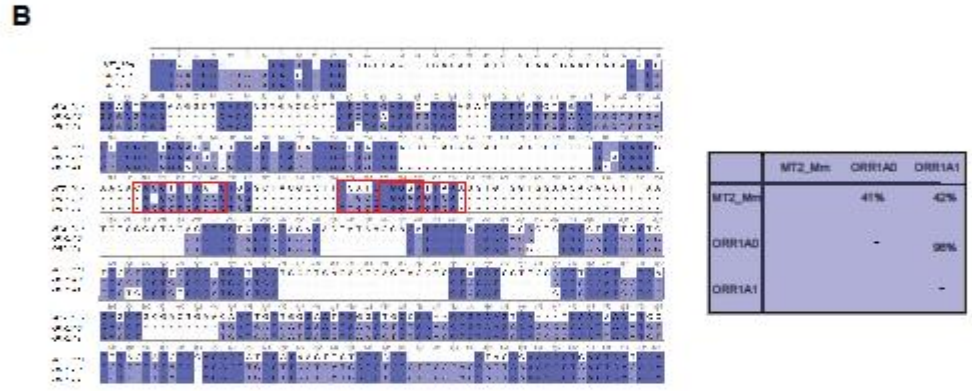
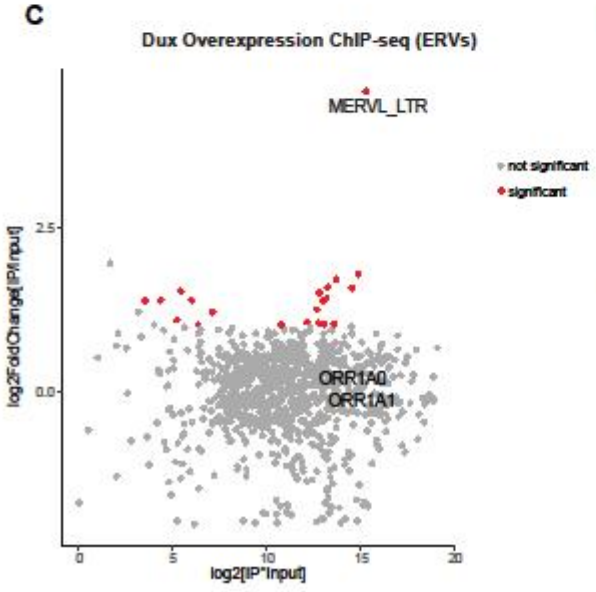

H
Dux
Dppa2
2C-TFs
Klf5
Klf4
TE genes
TE
ICM genes
ICM
Zygote
2C
4C
8C
Blastocyst
Totipotent
Bi-potent
First cell fate specification
MERVL
ORR1A0
ORR1A1
ICM genes
TE genes

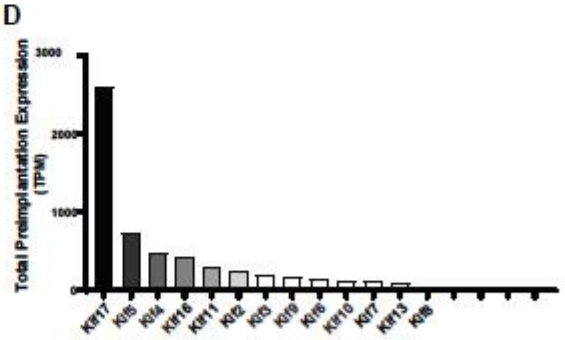
Figure 4 Klf5 induction is regulated by the 2C-specific transcription factor Dux. A. Multiple 2C-specific transcription factors show enrichment of occupancies proximal to *Klf5*. We mined all ESC ChIP-seq experiments in Cistrome DB and identified specific binding of Klf5 to a number of published 2C-specific 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 a robust induction of *Klf5* by *Dux*. Additionally, *Tbx3* and *Dppa2* overexpression also upregulates *Klf5*. Error bars = s.d.; *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 . *P* values for Dux overexpression were calculated with the DEseq2 package in R, otherwise *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 a *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 extra-embryonic TE marker *Cdx2* and primitive endoderm marker *Gata4*. Markers for endoderm, mesoderm and ectoderm (*Foxa2, Brachyury* and *Pax6*, respectively) were similarly induced in *Klf5-overexpressing* and control EBs tested. Data were generated from 2 independent, passage-controlled ESC lines. Error bars = s.d.; *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 blastocysts. ESC contribution to the ICM and/or the TE was determined by the localization of GFP-positive ESC progenies (left). The percentage of chimeric blastocysts with ESC contribution to the ICM, the TE, or both were quantified (right). Scale bar, 20 μm. **F, 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 the DUX/KLF5 regulation, but not the DUX/DPPA2 or DUX/NELFA regulation, 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 the early 2C stage, 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.

Sup Fig. 1

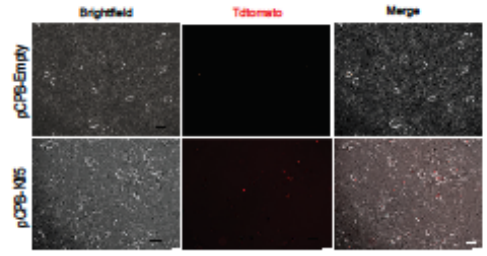

A
Zygote
2cell-early
2cell-mid
2cell-late
4cell
8cell
16cell
Blastocyst


B
MT2_Mm
ORR1A0
ORR1A1
41%
42%
98%

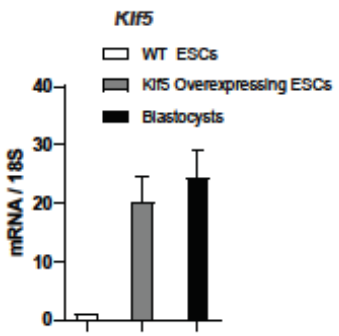

C
Dux Overexpression ChIP-seq (ERVs)
MERVL_LTR
not significant
significant
ORR1A0
ORR1A1
log2FoldChange[IP/Input]
log2[IP*Input]


D
Total Preimplantation Expression (TPM)
3000
2000
1000
0
Kff17
Kff5
Kff4
Kff16
Kff11
Kff2
Kff3
Kff9
Kff6
Kff10
Kff7
Kff13
Kff8

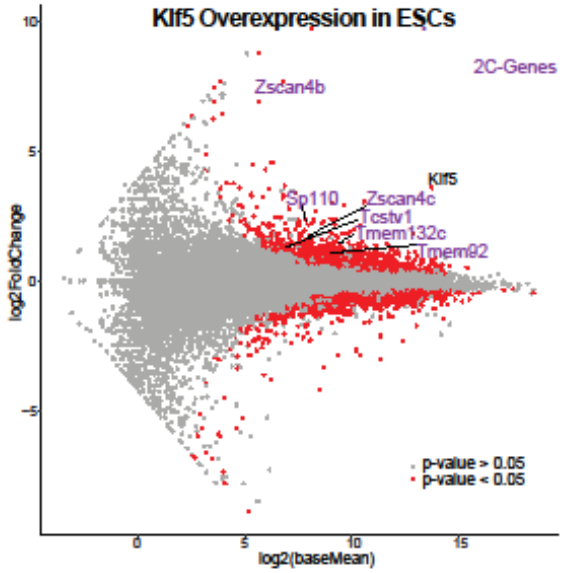
F


Brightfield
Tdtomato
Merge
pCPG-Empty
pCPG-Klf5

G

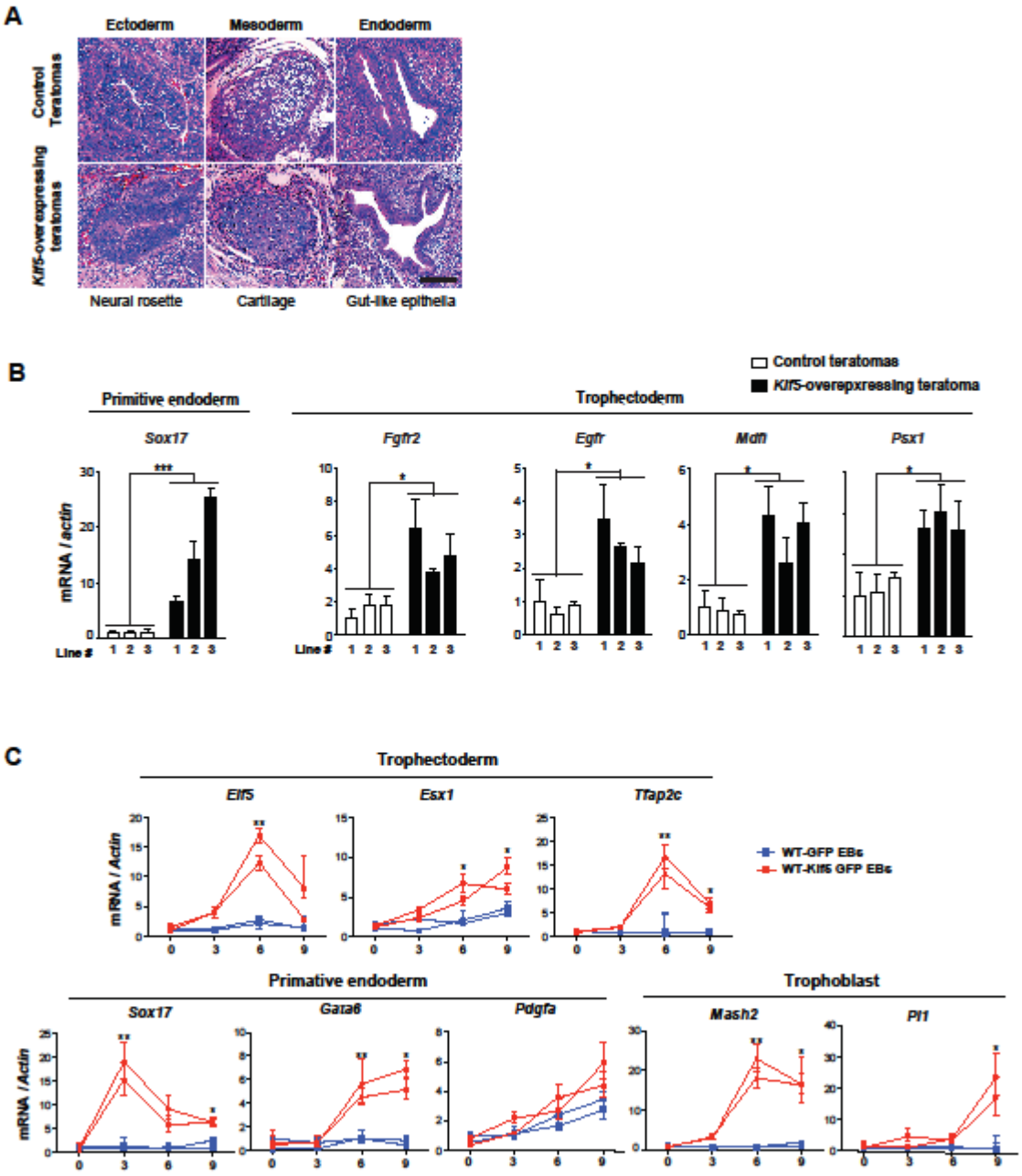

Klf5
WT ESCs
Klf5 Overexpressing ESCs
Blastocysts
mRNA / 18S
40
30
20
10
0

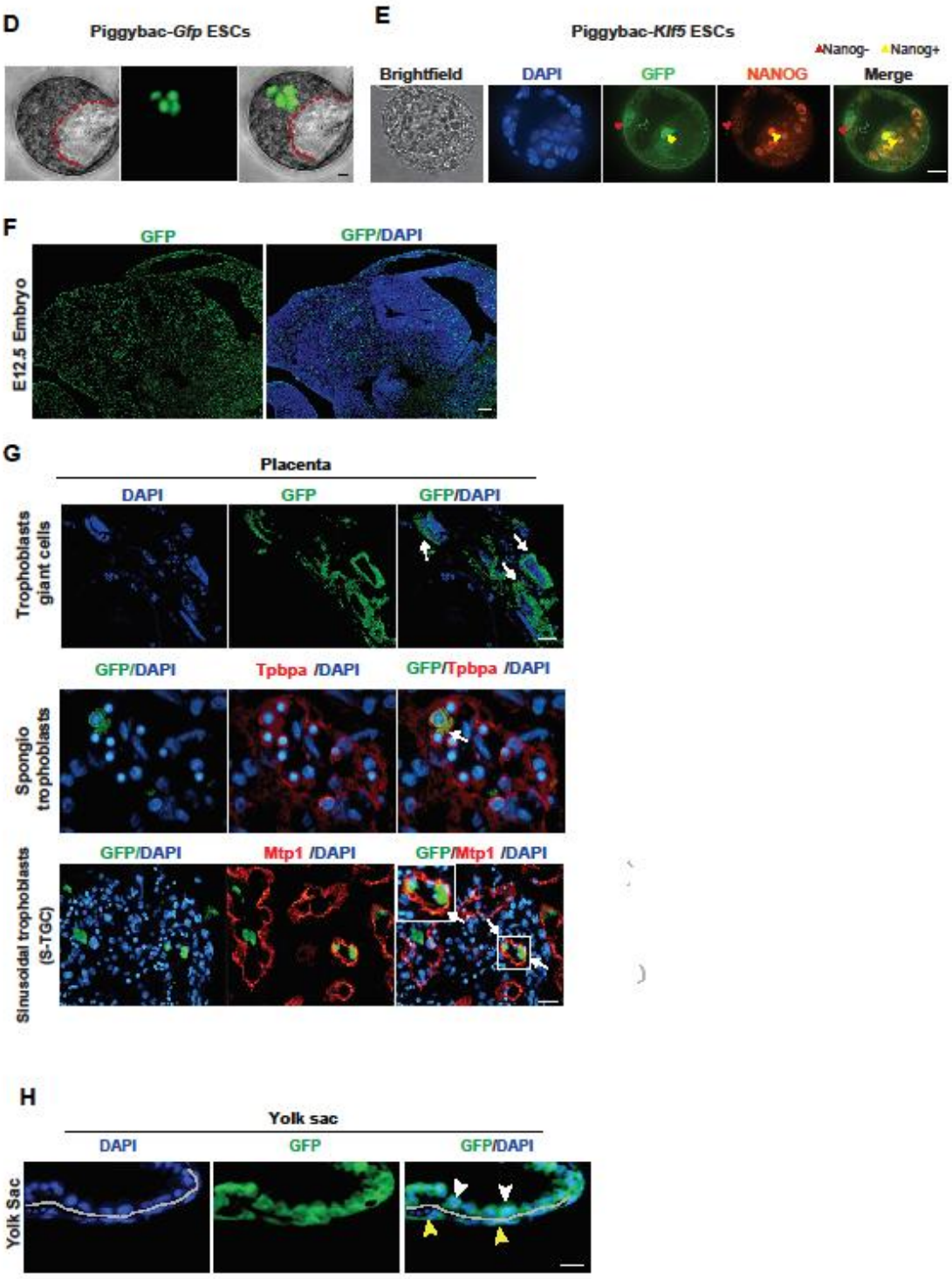
H


Klf5 Overexpression in ESCs
2C-Genes
Zscan4b
Klf5
Sp110
Zscan4c
Tcstv1
Tmem132c
Tmem92
p-value > 0.05
p-value < 0.05
log2FoldChange
log2(baseMean)
10
5
0
-5
0
5
10
15

Supplementary Figure S1. Klf5 directly regulates three major 2C-specific ERV families A. QPCR comparison of *Klf5* expression in WT ESCs, *Klf5* overexpressing ESCs and WT blastocyst embryos demonstrating that the level of *Klf5* overexpression is within a relevant physiological range. **B**. A Pseudotime plot showing the percentage of all reads that mapped to each of MERVL, ORR1A0 and ORR1A1 in preimplantation development. RNA-seq data from Deng et al. are used for this analysis. **C.** Multiple sequence alignment of MERVL, ORR1A0 and ORR1A1 consensus LTR sequences showing the sequence similarity among the three ERV families. ORR1A0 and ORR1A1 are highly homologous in sequences, while MERVL contains stretches of shared sequences with ORR1A0 and ORR1A1. **D.** Diagrams illustrating Klf5 ChIP occupancy at 2C-ERVs that act as alternative promoters for 2C-gene expression. **E.** Dux specifically binds to MERVL, but not ORR1A0 or ORR1A1. An MA-plot was shown to indicate the retrotransposon families enriched for Dux binding using ChIP-seq data of Dux overexpression ESCs from Hendrickson et al. **F.** A bar plot ranking Klf family members by their total expression level (TPM) across all preimplantation developmental stages using published RNA-seq dataset (Deng et al., 2014) . *Klf17*, *Klf4*, and *Klf5* are the three most highly and dynamically expressed Klf transcription factors in mouse preimplantation embryos. **G**. *Klf5* overexpression induces MERVL expression in a subset of 2C::Tdtomato wildtype ESCs. Representative images were shown for cultured 2C::Tdtomato ESCs transfected with control and *Klf5* expressing piggyback vectors. **H.** MA-plot of RNA-seq data in control versus *Klf5* overexpressing ESCs showing the upregulation of several well-recognized 2C genes.

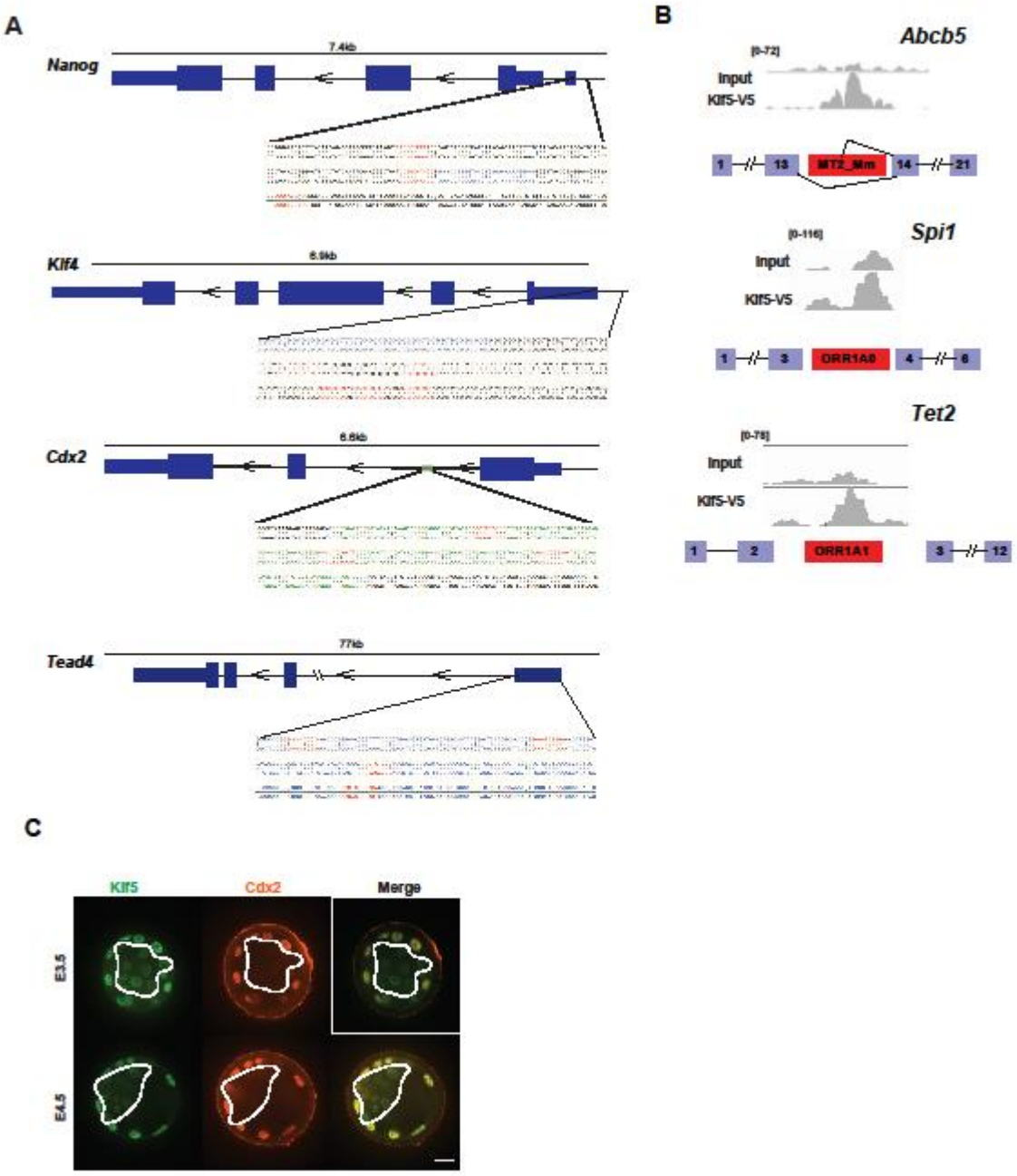
Sup. Fig. 2

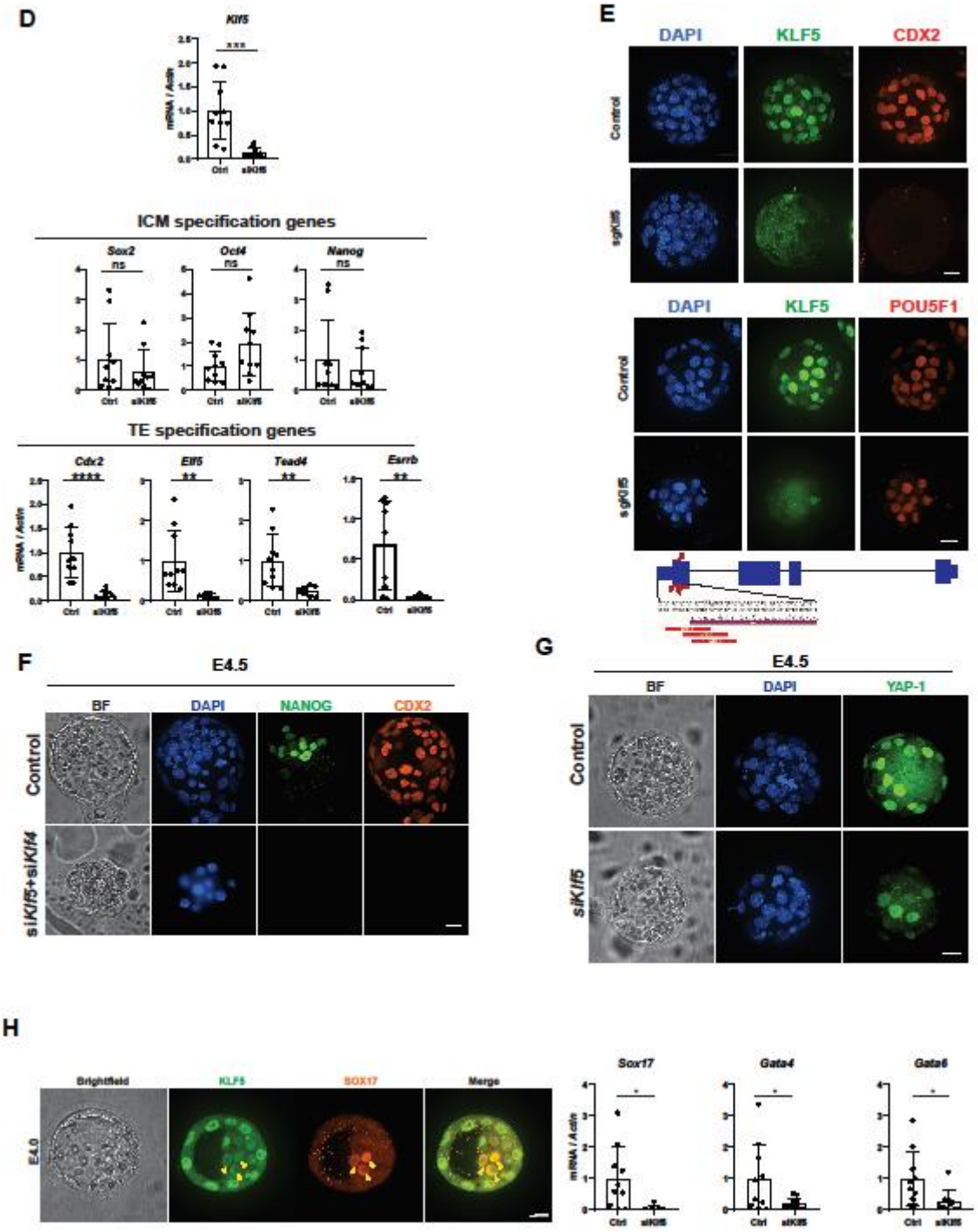

A
Ectoderm
Mesoderm
Endoderm
Control Teratomas
Klf5-overexpressing teratomas
Neural rosette
Cartilage
Gut-like epithelia
B
Control teratomas
Klf5-overepxressing teratoma
Primitive endoderm
Trophectoderm
Sox17
Fgfr2
Egfr
Mdfi
Psx1
mRNA / actin
Line #
C
Trophectoderm
Eif5
Esx1
Tfap2c
WT-GFP EBs
WT-Klf5 GFP EBs
mRNA / Actin
Primative endoderm
Trophoblast
Sox17
Gata6
Pdgfa
Mash2
Pl1


D
Piggybac-Gfp ESCs
E
Piggybac-Klf5 ESCs
Nanog-
Nanog+
Brightfield
DAPI
GFP
NANOG
Merge
F
GFP
GFP/DAPI
E12.5 Embryo
G
Placenta
DAPI
GFP
GFP/DAPI
Trophoblasts giant cells
GFP/DAPI
Tpbpa /DAPI
GFP/Tpbpa /DAPI
Spongio trophoblasts
GFP/DAPI
Mtp1 /DAPI
GFP/Mtp1 /DAPI
Sinusoidal trophoblasts (S-TGC)
H
Yolk sac
DAPI
GFP
GFP/DAPI
Yolk Sac

Supplementary Figure S2. Klf5 confers a bi-potential cell fate in ESCs. A. Control and *Klf5*-overexpressing ESCs generate differentiated teratomas containing tissues representative of the three embryonic germ layers (ectoderm, mesoderm and endoderm) as shown by H&E staining. Scale bars, 50 μm. **B.** A real time PCR panel of primitive endoderm (PE) marker (*Sox17*) and TE markers (*Fgfr2, Egfr, Mdfi* and *Psx1*) confirms that Klf5-overexpressing ESCs, but not control ESCs, are capable of robustly activating extra-embryonic transcriptional trajectories in differentiated teratomas. Error bars, *s.d.*, * $P < 0.05$, *** $P < 0.001$. **A**, **B**. A total of three independent pairs of passage-controlled ESCs were compared in the teratoma assay. **C.** A real time PCR panel performed during EB differentiation showed that *Klf5*-overexpressing EBs, but not control EBs, differentiate towards PE (*Sox17),* TE (*Elf5, Esx1* and *Tfap2c*) and more mature trophoblast-like lineages (*Mash2* and *Pl1)*. * $P < 0.05$, *** $P < 0.001$. All *P*-values were calculated using unpaired, two-tailed Student's *t*-test. **D.** Representative images of a chimeric blastocyst generated by microinjecting a single Piggybac-*Gfp* transduced wildtype ESC into a recipient morula. The progenies of the GFP-labeled ESC only contributed to the ICM. Scale bar = 20μm, red dashed line indicates blastocoel cavity. **E.** The progenies of GFP-labeled *Klf5* overexpressing ESCs can localize to both the ICM and TE compartments. Progeny that localize within the TE compartment lose NANOG protein expression, consistent with a shift from ICM to TE specification whereas progeny in the ICM regain NANOG expression. Red arrows indicate NANOG- progeny in the TE, yellow arrows indicate NANOG+ progeny in the ICM. Scale bar = 20um **F, G.** Klf5 overexpression in ESCs confers a bi-potential cell fate in E12.5 chimeric embryos. **F.** Progenies from Klf5-overexpressing ESCs contribute to embryonic cell lineages in E12.5 chimeric embryos. E12.5 chimeric embryos were generated by injecting either a single ESC or 10-15 ESCs into the recipient blastocysts. Scale bar = 500 μm. **G.** 10-15 GFP-labeled Klf5-overexpressing ESCs were microinjected into each 8 cell morula, followed by embryo transfer into pseudo-pregnant mothers to generate E12.5 chimeric embryos. GFP positive progenies from Klf5-overexpressing ESCs generated trophoblast giant cells (white arrows, identified by their unique cell morphology), spongiotrophoblasts (yellow arrows, identified based on their co-expression of Tpbpa), and syncytiotrophoblasts (orange arrows, identified based on their co-expression of Mtp1). Scale bar = 100μm. **H.** GFP-labeled Klf5-overexpressing ESCs yield both embryonic and extra-embryonic yolk sac lineages. White arrowheads indicate visceral endoderm cells identified based on the bilaminar structure of the yolk sac and their characteristic columnar epithelial morphology. Yellow arrowheads indicate mesothelium derived from embryonic mesoderm. Scale bar, 20 μm.

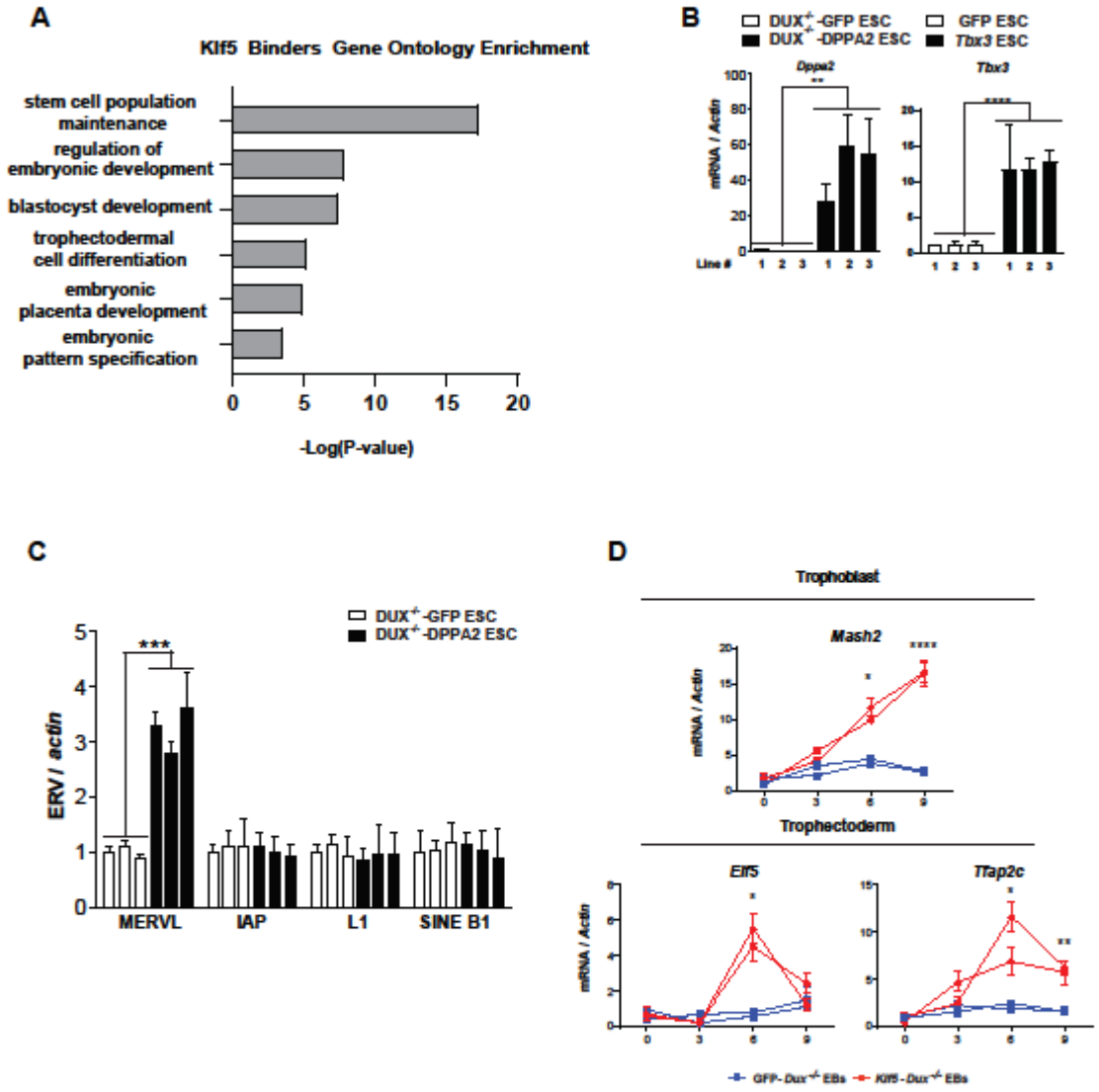
Supplementary Figure S3




D
Klf5
mRNA / Actin
Ctrl
siKlf5
ICM specification genes
Sox2
Oct4
Nanog
ns
TE specification genes
Cdx2
Elf5
Tead4
Esrrb
E
DAPI
KLF5
CDX2
POU5F1
Control
siKlf5
F
E4.5
BF
DAPI
NANOG
CDX2
Control
siKlf5+siKlf4
G
E4.5
BF
DAPI
YAP-1
Control
siKlf5
H
Brightfield
KLF5
SOX17
Merge
E4.0
Sox17
Gata4
Gata6

Supplementary Figure S3. Klf5 acts at the branch point between embryonic and extra-embryonic commitment. A. Cis-regulatory elements of *Oct4, Nanog*, *Cdx2* and *Tead4* contain multiple predicted Klf5 binding sites within ChIP-seq peaks identified from *Klf5*-overexpressing ESCs. Blue, a cis-regulatory region containing the TSS; green, known enhancer region of *Cdx2* (Wang and Shashikant, 2007), *red*, predicted Klf binding sites. **B.** Real time PCR analyses from *siKlf5* or scramble siRNAs treated E4.0 embryos demonstrated the efficiency of *Klf5* knockdown. Error bar = s.d.. *** $P = 0.0003$, df = 18, t = 4.510. The *P*-value is calculated based on unpaired, two-tailed student's unpaired t-test. **C, D.** *Klf5* knockout cells in blastocysts exhibit a decreased expression of Cdx2 yet retain intact Oct4 expression. Using CRISPR-EZ, we generated E4.0 blastocysts that contained a large subset of *Klf5* deficient cells. Immunofluorescence staining of Klf5 confirming the efficient ablation of *Klf5* by CRISPR-EZ in blastocysts. A marked decrease of Cdx2 in TE was observed in *Klf5* deficient cells in CRISPR edited blastocysts (**C**) yet Oct4 expression remained largely intact (**D**). Scale bars, 20 μm. **E.** Immunofluorescence panel of *Klf5+Klf4* double knockdown embryos showing that downregulation of both factors leads to loss of both NANOG and CDX2 protein in E4.5 embryos prior to apoptosis. **F.** Immunofluorescence panel of YAP-1 in control and *Klf5* knockdown embryos at E4.5 showing both nuclear and cytoplasmic localization of YAP-1 even in si*Klf5* treated embryos, indicated in-tact Hippo signaling despite the loss of *Klf5.* Scale bar = 20um. **G.** Real time PCR analyses confirmed the downregulation of TE specification genes *in Klf5* knockdown blastocyst embryos. Error bars = s.d.; *Cdx2*, ****$P < 0.0001$, df = 18, t = 5.297; *Elf5*, **$P = 0.0071$, df = 15, t = 3.114; *Tead4*, **$P = 0.0050$, df = 16, t = 3.248; *Esrrb*, **$P = 0.0016$, df = 17, t = 3.737. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, *n.s*., not significant. All P values were calculated using unpaired, two-tailed student's t-test. **H.** Comparison of *Klf5* and *Klf4* expression from published mouse preimplantation data (GSE45719) demonstrating that *Klf5* is enriched in the TE and expressed at comparable levels to *Klf4* in the ICM. **I.** Immunofluorescence panel of KLF5 and CDX2 protein confirming transcriptional results that KLF5 is enriched in the TE. White lines show the boundary between the TE and the ICM. Scale bar = 20um. **J.** MA-plot from microarray data published in Azami et al Development 2017 illustrating that, in this system, PrE marker genes are not significantly downregulated following loss of *Klf5.* Additionally, multiple probes for *Klf5* itself are not significantly differentially expressed.

Supplementary Figure S4



Supplementary Figure S4. Klf5 induction is regulated by the 2C-specific transcription factor Dux. A. A gene ontology enrichment plot summarizing key enriched functional terms shared by *Klf5*-binding transcription factors. Specific processes in embryonic development and stem cell biology converge on the regulation of *Klf5*. **B.** Real time PCR analyses for *Dppa2* and *Klf5* in *Dppa2*-overexpressing and control $Dux^{-/-}$ ESCs showing that *Klf5* is upregulated by *Dppa2* in ESCs, independently of *Dux*. Three independent $Dux^{-/-}$ ESC lines were employed for this analysis. Error bars = s.d., *Dppa2*, $^{**}P = 0.0090$, df = 4, t = 4.743; *Klf5*, $^{**}P = 0.0027$, df = 4, t = 6.622. **C.** EBs derived from *Klf5*-overexpressing $Dux^{-/-}$ ESCs induce markers of TE and trophoblasts. Real time PCR analyses confirmed the additional induction of TE lineage markers and trophoblast markers differentiating EBs derived from *Klf5*-overexpressing $Dux^{-/-}$ ESCs, but not from control $Dux^{-/-}$ ESCs. Two independent $Dux^{-/-}$ ESC lines were employed for these analyses. Error bars = s.d. *Elf5* (day 6), $^{*}P = 0.0147$, df = 2, t = 8.158; *Tfap2c* (day 6), $^{*}P = 0.0403$, df = 2, t = 4.830; *Tfap2C* (day 9), $^{**}P = 0.0014$, df = 2, t = 26.55; *Mash2* (day 6), $^{*}P = 0.0132$, df = 2, t = 8.602; *Mash2* (day 9), $^{****}P < 0.0001$, df = 2, t = 138.8. All P values were calculated on the basis of unpaired, two-tailed Student's t-tests. $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$

Chapter 3: MERVL-*Gag* is required for one cell to two cell transition in mouse preimplantation embryos

Abstract

Recent reports on the functionality of ERVs has precipitated a paradigm shift away from these elements being viewed merely as parasitic genomic entities, to a view that considers a subset of ERVs as essential to their host's biological processes. This work describes the essential function of the Gag protein encoded by the ERV, MERVL. Depletion of MERVL by RNAi causes developmental arrest just before nuclear envelope breakdown (NEBD) after the parental pronuclei have met in the central region of the nascent embryo. Interestingly, MERVL-*Gag* localizes to both the perinuclear region and the midbody bridge in murine preimplantation embryos. Further, MERVL-*Gag* directly binds to, and colocalizes, with Tsg101, an ESCRT subunit that has been shown to be required for cytokinesis. Lastly, depletion of Tsg101 phenocopies defects observed in the absence of MERVL-*Gag*, although at incomplete penetrance. Altogether, this work details the domestication of MERVL-*Gag* for the fundamental process of cell division during the one cell to two cell transition and suggests that MERVL-*Gag* functionally cooperates with Tsg101.

Background

Endogenous retroviruses (ERVs) comprise 8% and 10% of mouse and human genomes, respectively (Gifford and Tristem, 2003). However, given their substantial contribution to their host genomes, their functions, and mechanisms of action in their hosts, are just beginning to be described. Anecdotal reports of ERV derived enhancers (Todd et al., 2019), promoters (Modzelewski et al., 2021) and co-opted proteins (Dupressoir et al., 2009, 2011) have driven a paradigm shift away from the notion of these elements simply being junk DNA, towards the theory that they, in fact, function in fundamental host processes from telomere length maintenance (Levis et al., 1993) to the evolution of the placenta (Dupressoir et al., 2009, 2011). A particular ERV, MERVL, has been described as among the earliest transcribed genes during murine embryogenesis. Intriguingly, MERVL has been described as a marker for cells that have totipotent-like attributes and enforced expression of MERVL in mouse embryonic stem cells (mESCs) correlates with the expansion of developmental potential of mESCs. Further, multiple reports have demonstrated an ambiguous, but essential, function for MERVL during preimplantation development (Huang et al., 2017; Kigami et al., 2003). However, because MERVL encodes both *Gag* and *Pol* open reading frames, and since MERVL loci themselves have been reported to regulate chromatin architecture (Kruse et al., 2019), the functional modality of MERVL in preimplantation development has remained elusive.
In the mouse, following fertilization, the paternal pronucleus migrates towards the central region of the egg where it meets the maternal pronucleus. After the congress of the parental pronuclei, the embryo then initiates nuclear envelope breakdown (NEBD). In vertebrates NEBD is regulated, in part, by the phosphorylation of nuclear Lamin A by Cdk1/CyclinB (De Castro et al., 2018). This phosphorylation of nuclear lamina triggers their disassembly and the initiation of mitosis.

Mouse embryos are unusual in that they undergo acentriolar mitosis for the first few cleavages as both male and female games degrade their centrosomes (Courtois et al., 2012; Manandhar et al., 1998, 2005; Schuh and Ellenberg, 2007). It is not until the blastocyst stage, once centrosomes have assembled *de novo*, that they gradually transition to more typical cell division mechanisms (Courtois et al., 2012). At the onset of mitosis, microtubule organizing centers (MTOCs) serve as spindle nucleation points. Just prior to metaphase, these individual MTOC clusters coalesce into a broad bipolar spindle. Once mitosis has completed and the

daughter cells have re-entered interphase, a persistent interphase bridge can be observed between nascently formed daughter blastomeres. Intriguingly, a recent report has shown that this interphase bridge can serve as an MTOC, and ablation of the interphase bridge triggers entry into the next mitosis.

It has been shown that there is significant overlap in the proteins involved in retroviral budding and those required to the completion of cytokinesis (Carlton and Martin-Serrano, 2007). Particularly, the endosomal sorting complex required for protein transport (ESCRT) are a series of complexes involved in membrane scission. Tsg101 has been shown to be essential for cytokinesis and is also an important host factor that mediates HIV viral vesicle budding (Usami et al., 2009; Votteler and Sundquist, 2013; Wagner et al., 2003). During the late stages of cytokinesis, Tsg101 is recruited to the midbody via its interaction with Cep55. At the midbody, Tsg101 works to recruit downstream ESCRT complexes, thus mediating abscission between the nascent daughter cells.

Through a series of bioinformatic, molecular biology, biochemical and loss of function studies, I discovered that the MERVL-*Gag* protein localizes to perinuclear spaces and to the midbody ring, with the latter being cell cycle dependent. Further, I show that MERVL-*Gag* binds to ESCRT-I subunit, Tsg101 and that depletion of MERVL-*Gag*, and to a lesser extend Tsg101, causes arrest at the one cell stage. Together, these data suggest that MERVL-*Gag* plays an essential role in mouse preimplantation development and may act upstream of NEBD following fertilization.

Results

Expression and subcellular localization of MERVL-*Gag*

To establish when MERVL is expressed, and hence the point in development where it would function, we re-analyzed publicly available single cell RNA-seq (scRNA-seq) datasets that span preimplantation development in the mouse (Zhao et al., 2018). These data revealed that MERVL expression initiates at the one cell stage, is sharply upregulated in two cell and four cell embryos, and rapidly declines by the eight cell stage before becoming transcriptionally silent in blastocysts (Figure 1A). These data reflect MERVL as a family, and not independent loci, as MERVL elements are highly similar in sequence and are not readily distinguishable with available short read data (Teissandier et al., 2019). Therefore, in order to focus in on a functional modality of MERVL, we reasoned that conserved regions in the MERVL gene body across MERVL loci suggest a beneficial role in mouse preimplantation embryos. Classification of MERVL, using the REannotate tool (Pereira, 2008), revealed that there are 659, 811 and 1032 loci that are complete in structure, truncated or solo LTRs, respectively (Figure 1B). Alignment of the internal stretches of the full length MERVL elements revealed that while the *Pol* gene contained interspersed deletions, the *Gag* open reading frame remained largely intact in a majority of full length MERVL copies (Figure 1B). The consensus sequence of MERVL contains the sequence for the full length ancestral *Gag* open reading frame (Bénit et al., 1997; Kigami et al., 2003). Utilizing NCBI blast revealed that roughly 600 insertions retain a *Gag* with at least 95% sequence identity to the consensus at the nucleotide level (Figure 1C). Altogether, these data establish that the *Gag* open reading frame of MERVL is conserved across many full length loci and are suggestive of a function for MERVL *Gag* during mouse preimplantation development.

Typical examples of retroviruses such as HIV and SFV can be expressed as Gag-*Pol* fusions that are then processed into various functional truncated sub-complexes (Bell and

Lever, 2013; Müllers, 2013). To ascertain whether MERVL-*Gag* is itself processed, we measured MERVL-*Gag* expression in oocytes and two cell stage embryos by single cell western. Expectedly, MERVL-*Gag* expression was very low in oocytes and became sharply upregulated by the two cell stage (Figure 1D). Intriguingly, the molecular size for MERVL-*Gag* was a single sharp peak at the expected size for MERVL-*Gag* and not a *Gag-Pol* fusion (Figure 1D). These results suggest that MERVL-*Gag* is not processed. However, it is possible that a MERVL-*Gag-Pol* fusion is expressed *in vivo*, but at a level below the limit of detection of the single cell western assay.

To see how well the expression of MERVL-*Gag* correlates to the transcriptional profile of MERVL, I stained embryos at different stages across preimplantation development. MERVL-*Gag* demonstrated prominent perinuclear staining in the blastomeres of two cell and four cell stage embryos. Unexpectedly, in these stages MERVL-*Gag* appeared to form ring structures between blastomeres (Figure 1E). These ring structures were highly reminiscent of the midbody ring that forms at the late stages of cytokinesis, just prior to cytokinetic bridge abscission (Hu et al., 2012). To confirm whether this could be the midbody ring, I performed a time course where I stained embryos undergoing the one cell to two cell transition and used immunofluorescence to track the spatiotemporal localization of MERVL-*Gag.* During interphase, MERVL-*Gag* could be seen as a bridge connecting the one cell embryo with the polar body. The expression of MERVL-*Gag* remained broadly cytoplasmic until telophase where it could be seen in, presumably, the cytokinetic bridge and the midbody ring. Finally, by the late two cell stage MERVL-*Gag* appeared to be highly concentrated in the perinuclear region, with only faint staining in the remnants of the midbody ring (Figure 1F). To gain a better understanding of the structure of the ring feature MERVL-*Gag* is recruited to, we performed STORM on two cell stage embryos that had been stained for MERVL-*Gag*. STROM confirmed that MERVL-*Gag* does form a ring structure, approximately 4um in diameter, surrounding the cytokinetic bridge (Figure 1G). Interestingly, by the morula stage it was evident that MERVL-*Gag* expression had quickly diminished. However, a subset of blastomeres, seemingly those situated on the exterior of the embryo, retained a low level of MERVL-*Gag* expression (Figure 1H). Altogether, these data show that MERVL-*Gag* protein expression is very concordant with MERVL expression at the transcriptional level and that MERVL-*Gag* localizes to the midbody ring and perinuclear region up until the morula stage when its low expression appears biased to outer blastomeres.

MERVL is required for the one cell to two cell transition

Informed by MERVL-*Gag*'s biased expression in morula stage embryos (Figure 1H) and the established literature correlating MERVL expression with totipotent-like cell fate potential (Choi et al., 2017a; Hu et al., 2020; Ishiuchi et al., 2015; Macfarlan et al., 2011, 2012; Yan et al., 2019), I sought to determine whether depletion of MERVL *in vivo* would alter cell fate specification. Blastomeres of two cell stage embryos retain totipotency as well as high MERVL expression (Figure 1A). Therefore, I performed single blastomere injections co-delivering a validated siRNA that targets MERVL (Huang et al., 2017) as well as mRNA encoding H2B-Gfp in order to track injected cells at subsequent developmental stages. Strikingly, while single blastomeres injected with control scramble siRNA and H2B-Gfp mRNA were able to contribute to approximately half of the resulting blastocyst, blastomeres knocked down for MERVL showed a dramatic arrest in proliferation (Figure 2A). Further, uninjected blastomeres also appeared to be negatively impacted by the knockdown of MERVL in the sister blastomere as shown by the markedly reduced cell count in the resulting blastocysts, despite activation of TE marker *Cdx2* (Figure 2A). This observation suggests that MERVL may have cell nonautonomous function. In order to confirm the requirement of MERVL for proliferation during development, I repeated the knockdown of MERVL, but in one cell stage embryos rather than single blastomeres of two

cell embryos. Consistent with the single blastomere knockdown results, depletion of MERVL as early as the one cell stage also inhibited developmental progression to the two cell stage. Approximately half of all MERVL knockdown embryos arrested at the late one cell stage (Figure 1B). Interestingly, they arrested at a stage where the maternal and paternal pronuclei were in very close proximity and in the central region of the embryo (Figure 1C). Consistent with the implied functionality of MERVL-*Gag* due to its extensive conservation among full length MERVL loci (Figure 1B), co-delivery of the MERVL targeting siRNA with siRNA resistant mRNA, encoding MERVL-*Gag,* significantly improved faithful progression to the two cell stage (Figure 1C). This suggests that it is a lack of MERVL-*Gag* that leads to arrested development of MERVL knockdown embryos and blastomeres. The proximity of the parental pronuclei in the central region of the embryo indicated the arrest following siMERVL treatment was towards the late one cell stage, prior to the onset of the first mitotic cleavage. However, it was ambiguous from DAPI staining alone whether the arrest was prior to, or before, nuclear envelope breakdown (NEBD) (Figure 2C). In order to ascertain when, during the one cell to two cell transition, siMERVL treated embryos arrest, I performed additional knockdown and scramble control microinjections and co-delivered mOrange-LaminA to visualize the nuclear envelope and Tubulin-Gfp mRNA to visualize the microtubules. Intriguingly, whereas scramble control two cell embryos had both nuclear envelope and nucleoplasmic staining for mOrange-LaminA, siMERVL treated embryos' nuclear envelope appeared far less patterned. It has been observed that at the onset of M-phase the kinase activity of Cdk1 phosphorylates LaminA at Ser22, and this phosphorylation event leads to nucleoplasmic, rather than nuclear envelope associated LaminA localization. Further, phosphorylation of LaminA at these sites precedes the onset of nuclear envelope breakdown {}. Finally, embryos depleted for MERVL also demonstrated elevated pH2AX signal, indicating increased double stranded breaks in the genome. The elevated pH2AX signal could not be rescued by co-delivery of siRNA resistant MERVL-*Gag*, indicating that it may be the lack of another aspect of the MERVL genome that is responsible this phenotype. Altogether, these data establish that MERVL is required for the one cell to two cell transition, likely in part due to MERVL-*Gag* function and suggest that MERVL depleted embryos arrest prior to NEBD at the late one cell stage.

MERVL-*Gag* harbors a PSAP domain and interacts directly with Tsg101

It has been reported that the topology of retroviral budding from the cell surface is comparable to membrane dynamics during the late stages of cytokinesis (Carlton and Martin-Serrano, 2007). Additionally, many of the of the host factors that mediate cytokinesis are also involved in retroviral budding (Votteler and Sundquist, 2013). Particularly, the endosomal complex required for protein transport (ESCRT) has been shown to be required for cytokinesis as well as HIV budding (Carlton and Martin-Serrano, 2007; Elia et al., 2011; Morita et al., 2007; Usami et al., 2009). In order to determine whether MERVL-*Gag* interacts with ESCRT components, I scanned the primary sequence for late domain motifs, usually recognized by ESCRTs during retroviral budding (Martin-Serrano et al., 2004; Usami et al., 2009; Votteler and Sundquist, 2013). Intriguingly, I identified a PSAP motif within the MERVL-*Gag* open reading frame (Figure 3A). It has been shown that the ESCRT-I component, Tsg101, recognizes and binds to PT/SAP motifs via its UEV domain (Pornillos et al., 2003). Further, utilizing fluorescently tagged versions of MERVL-*Gag* and Tsg101, I showed that these two proteins co-localize in the perinuclear space of one cell embryos and that both MERVL-*Gag* and Tsg101-Halo can be seen in the midbody ring at the two cell stage (Figure 3A). A limitation of this data is that MERVL-*Gag* and Tsg101 were not observed in the same midbody ring in the same embryo at the same time. Further work to colocalize MERVL-*Gag* and Tsg101 in the same midbody ring is required. Consistently, cointroduction of native MERVL-*Gag* fused to DsRed (PSAP-DsRED) and Tsg101-Halo into HEK293Ts also revealed virtually perfect colocalization of PSAP-DsRED and Tsg101-Halo (Figure 3B). However, MERVL-*Gag* with a mutated PSAP domain (LSAP-

DsRED) failed to colocalize with Tsg101 (Figure 3B). In order to determine whether the observed colocalization between MERVL-*Gag* and Tsg101 was due to direct association of the two, I performed coimmunoprecipitation (Co-IP) experiments. Consistent with the colocalization observed in embryos and in cultured cells, MERVL-*Gag* directly bound Tsg101 in a PSAP motif dependent manner, as mutation of this motif entirely abrogated MERVL-*Gag* and Tsg101 binding (Figure 3C). Strikingly, Cep55 could also be pulled down with MERVL-*Gag*, again in a PSAP dependent fashion as mutation of the PSAP motif greatly reduced Cep55 binding to MERVL-*Gag* (Figure 3C). This suggests that MERVL-*Gag*, Tsg101 and Cep55 can form a complex.

Depletion of Tsg101 phenocopies MERVL knockdown albeit with incomplete penetrance

The interaction observed between MERVL-*Gag*, Tsg101 and Cep55 raised the possibility that MERVL-*Gag* may impact the subcellular localization of these interaction partners. While depletion of MERVL had no observable impact on Tsg101's localization to the midbody ring in two cell stage embryos (Figure 3D), there was an obvious loss of Tsg101 localization in the perinuclear space of embryos that were able to make it to the two cell stage (Figure 3D). This altered localization led me to speculate that depletion of Tsg101 in preimplantation development may phenocopy observed MERVL knockdown defects. A previous study on the requirement of Tsg101 during development concluded that Tsg101 appears to be dispensable for the early stages of preimplantation development, as Tsg101 knockout blastocysts were indistinguishable from controls (Wagner et al., 2003). However, this study utilized heterozygous intercrosses and can be confounded by maternal contributions of Tsg101 mRNA. In order to overcome this ambiguity, I depleted Tsg101 from one cell preimplantation embryos via microinjection of an siRNA targeting Tsg101 and co-delivered Tubulin-Gfp and mOrange-LaminA to visualize microtubules and the nuclear envelope, respectively. Consistent with published results, most of the embryos treated with an siRNA against Tsg101 were able to make it to the two cell stage without any obvious morphological defects (Figure 3D). However, in 10% of cases, depletion of Tsg101 led to one cell arrest. Intriguingly, the arrested embryos lacked nuclear mOrange-LaminA signal, resembling the MERVL depleted embryos (Figure 3E). While this data certainly requires additional validation, it suggests that both MERVL (presumably MERVL-*Gag*) and Tsg101 are both involved in regulating events upstream of nuclear envelope breakdown and entry into mitosis.

Discussion

In this study I show that MERVL-*Gag* localizes to perinuclear spaces and to the midbody ring in preimplantation embryos. Further, this localization mirrors that of Tsg101, an ESCRT-I subunit that has been demonstrated as essential for cytokinesis (Carlton and Martin-Serrano, 2007; Wagner et al., 2003) and that MERVL-*Gag* directly associates with Tsg101, and through this may form a complex with Cep55. Strikingly, RNAi mediated depletion of MERVL-*Gag* led to one cell arrest, characterized by an increase in pH2AX staining. Additionally, MERVL knockdown embryos also lacked phosphorylated laminA on their nuclear envelope, suggestive of a failure to initiate NEBD. Depletion of Tsg101 also led to a similar defect, though with incomplete penetrance. Altogether, these data suggest that MERVL-*Gag* has an essential function in the early stages of preimplantation development, and my cooperate with Tsg101 and downstream interacting partners to regulate the onset of NEBD.

Previous studies have attempted loss of function analyses to determine what, if any, role MERVL may have *in vivo* (Huang et al., 2017; Kigami et al., 2003). Consistent with my analysis, they show that depletion of MERVL leads to developmental arrest, though timepoints between studies differ. One possible explanation for this apparent heterogeneity in MERVL knockdown associated phenotypes is the efficiency of depletion. Further, the timing of depletion appears to

be a critical determinant as roughly half of MERVL knockdown embryos in this study were able to successfully progress to the two cell stage. As embryos were harvested following natural mating, it is likely the spread in fertilization time impacts the observed phenotype upon MERVL knockdown. As this remains a limitation of the current study, it would be beneficial to systematically assess the impacts of MERVL depletion at different times following in vitro fertilization, which can aide in developmental synchrony.

MERVL transcribes all of its open reading frames as a single mRNA (Kigami et al., 2003; Ribet et al., 2008), therefore any attempt, by RNAi, at depleting MERVL-*Gag* will also have an impact on the other open reading frames in the MERVL genome. In order to better attribute the defects observed when MERVL is depleted to MERVL-*Gag*, an orthogonal approach, such as Trim Away (Clift et al., 2018), should be attempted. Finally, previous reports suggest that Tsg101 is not essential for the first few cleavages of preimplantation development, as Tsg101 null blastocysts are morphologically normal (Wagner et al., 2003). However, this study's analyses are confounded by maternally deposited Tsg101 mRNA. In conjunction with MERVL-*Gag*, Trim away on Tsg101 would also be deeply informative, especially if this depletion of Tsg101 continues to phenocopy the Trim Away mediated depletion of MERVL-*Gag*.

Materials and Methods

Superovulation and Embryo Collection

Wild-type B6 or F1 female mice were injected with 5 IU of PMSG (Prospec), followed by 5 IU of HCG (Prospec) 46-48 hours later. Once injected with HCG, females were paired with stud males. The next morning, females were checked for the presence of copulation plugs and pronuclear stage embryos were harvested from plugged females. Embryos were dissociated from cumulus cells using hyaluronidase and washed in M2 buffer.

Immunofluorescence Staining

Embryos were fixed with 4% paraformaldehyde (PFA) for 15 min at room temperature and incubated in permeabilization buffer for 5 min. Embryos were blocked in 10% goat serum for 1 hour at room temperature and incubated in primary antibody overnight at 4C. Primary antibodies used were MERVL Gag (1:100, Epigentek), Tsg101 (1:200, Proteintech), and pH2AX (1:100, Cell Signaling). Embryos were stained with secondary antibody for 1 hour at 4C and 1x DAPI for 10 min at room temperature. Secondary antibodies used were goat anti-rabbit IgG (H+L) conjugated to Alexa Fluor 488 and 594 (1:200, Life Technologies). After each step except blocking, embryos were washed 2-3 times with wash buffer (PBS with 0.2% BSA). Embryos were imaged using confocal microscopy (Nikon Eclipse TE200-E) at 60x water.

Co-Immunoprecipitation

HEK293 cells were grown on 15 cm cell culture dishes with DMEM (Life Technologies). At around 60-70% confluency, cells were transfected with 500 µL DMEM, 50 µL polyethylenimine (PEI), and a total of 10 µg DNA (5 µg for each construct if transfecting with two constructs) for each condition. 48 hours after transfection, media was aspirated, and cells were washed with 2 mL ice-cold PBS + PMSF (1:100 dilution). Then 1 mL PBS + PMSF was added, and cells were scraped off dishes using cell scrapers, then transferred to 15 mL conical tubes. Cells were centrifuged at 800 rpm for 5 min and supernatant was removed.

After harvesting the cells, all steps were completed on ice. 200 µL lysis buffer containing 10 mM Tris, 0.5 mM EDTA, 150 mM NaCl, 2% Triton X-100, and PMSF was added to the cells. The

resulting cell lysates were pipetted up and down 20 times and put through a 1 mL syringe and 26G x 1/2 needle into a new 15 mL conical. The lysates were incubated on ice for 5 min, then placed in a Taitec MicroMixer E-36 on high speed for 5 min. Following the shaking step, cell membranes came out of solution and were placed on the side of the tube. Two more times, the remaining supernatant was pipetted up and down 20 times and put through a syringe into the same conical, incubated on ice for 5 min, and placed on a shaker for 5 min. The lysates were transferred to a 1.5 mL tube and centrifuged for 10 min at 4C, 20000 RCF.

While the lysates were centrifuging, anti-HA beads were equilibrated. Three times, 750 µL lysis buffer was added to 25 µL beads for each condition. The beads were inverted and centrifuged for 30 sec at 10 RPM, and the lysis buffer was removed.

Once the lysates were finished centrifuging, the supernatant was transferred to a new 1.5 mL tube. 300 µL dilution buffer containing 10 mM Tris-HCl, 150 mM NaCl, 0.5 mM EDTA, and PMSF. was added to the lysates and the resulting mixture was pipetted up and down. 50 µL lysate for each condition was set aside in a separate 1.5 mL tube for Western blot analysis. 50 µL Laemmli buffer was added and the input samples were vortexed briefly, boiled at 95C for 5 min, and stored at -80C. Meanwhile, 25 µL anti-HA beads were added to the lysates, and the resulting IP samples were then incubated overnight in a rotating mixer at 4C.

The next day, the IP samples were centrifuged for 30 sec at 4C, 8200 RCF and the supernatant was removed. Three times, 750 µL lysis buffer was added to the samples and the samples were vortexed and inverted for 20 sec. They were then centrifuged for 30 sec at 4C, 8200 RCF and the supernatant was removed. 50 µL Laemmli buffer was added to the samples, which were vortexed briefly, boiled at 95C for 5 min, and stored at -80C.

Western Blot Analysis

One day after co-immunoprecipitation, the lysates and IP samples were separated using 10% SDS-PAGE. 5 µL Dual Color ladder (Bio-Rad) and 12 µL of sample were loaded into the gel. After electrophoresis, the proteins were transferred onto 0.45 µm-pore sized nitrocellulose membranes overnight at 4C. The next day, the membranes were washed with TBS-T containing 0.1% Tween-20 and blocked with 5% milk for 1 hour at room temperature, then incubated with primary antibody overnight at 4C. Primary antibodies used were MERVL Gag (Epigentek), Tsg101 (Proteintech), Cep55 (Proteintech), and β-tubulin (Cell Signaling Technology), each at 1:1000 dilution. After washing with TBS-T, the membranes were incubated in HRP-conjugated secondary antibody for 1 hour at room temperature. Secondary antibody used was mouse anti-rabbit IgG-HRP (1:5000, Santa Cruz Biotechnology). The membranes were washed with TBS-T and visualized using chemiluminescence on the ChemiDoc XRS+ Gel Imaging System (Bio-Rad).

Cell Imaging

HEK293 cells were grown in 6-well plates with DMEM. At around 50% confluency, cells were transfected with 500 µL Opti-MEM (Life Technologies), 10 µL polyethylenimine (PEI), and a total of 2 µg DNA (1 µg for each construct if transfecting with two constructs) for each condition. 24 hours after transfection, cells were passaged 1:3 onto 35 mm glass bottom dishes. 24 hours after passaging, 2 µL 100x Hoechst and 0.1 µL 100x Janelia Fluor 646 HaloTag Ligand were added to dishes and cells were incubated for 1 hour at 37C. Media was aspirated, and cells were washed with 0.5 mL PBS. PBS was aspirated and cells were fixed with 1.5 mL 4% PFA for 15 min at room temperature. After incubation, PFA was aspirated, and 1.5 mL PBS was added to cells. Cells were imaged using confocal microscopy at 60x water.

Embryo Microinjection

All mRNA was made using the HiScribe T7 ARCA mRNA kit with tailing (New England Biolabs) according to manufacturer's instructions. Mice were first superovulated by administration of 5 IU of PMSG (Prospec) and HCG (Prospec) spaced 46-48 hours apart and then paired with a stud male. On E0.5, embryos were harvested and cleared of cumulus cells with hyaluronidase in M2 (Sigma). Embryos were then microinjected with mRNA at a concentration of 300-500 ng/μL using an Eppendorf FemtoJet.

For the Gag knockdown experiment, siMERVL embryos were microinjected with an siRNA specific to the Gag region of MERVL (ordered custom from ThermoFisher). Control embryos were treated with a scramble control siRNA. Both siRNAs were used at a concentration of 100 μM. For the rescue embryos, siRNA resistant Gag mRNA was included with the siRNA.

Golden Gate Assembly

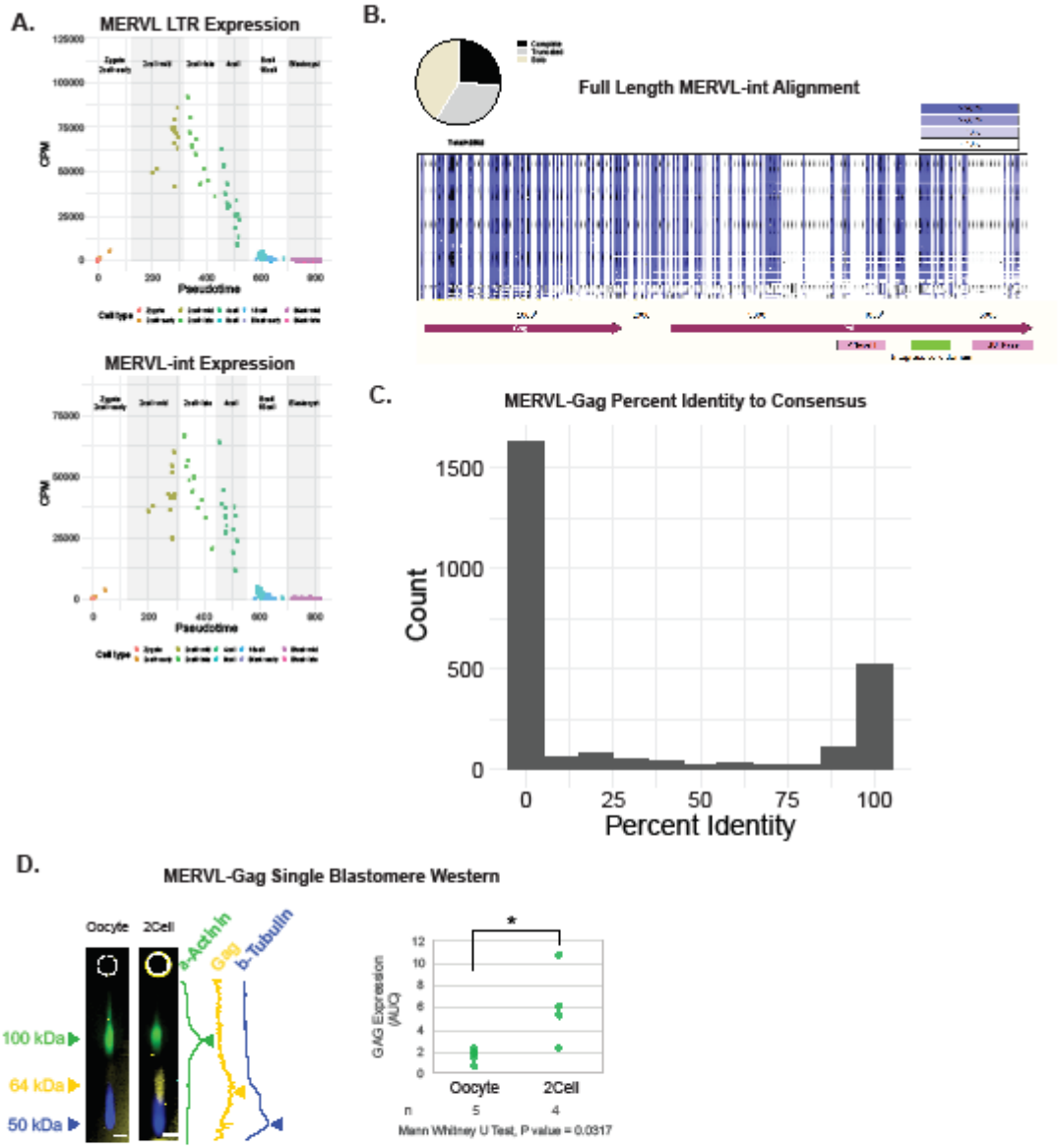
Full-length and Split Halo constructs for Gag, PSAP, Tsg101, Cep55, and Alix were cloned using Golden Gate assembly. All fusion proteins were designed using SnapGene and cloned into a PiggyBac vector containing an EF1α promoter, BsmB1 cutting site, and Poly (A) signal sequence. Primers were ordered as custom DNA oligos from ThermoFisher.

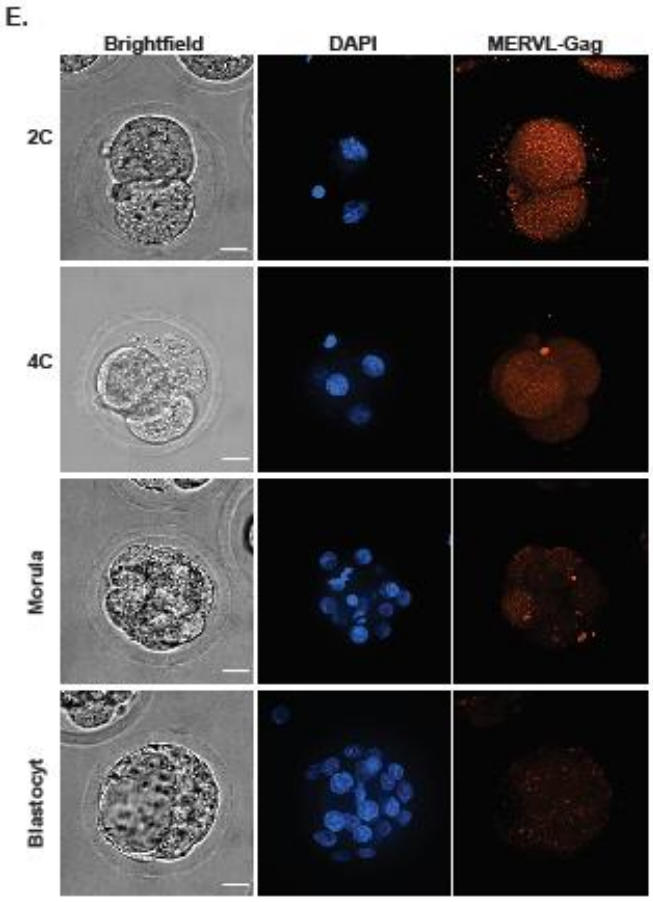
For each construct, primers were designed to amplify the two fragments of interest with BsmB1 cutting sites added to the ends. Each primer began with the sequence GCCGTCTCT, with the first 2 bases being spacer nucleotides and CGTCTC (N) being the BsmB1 site. For the first fragment in the fusion protein, an additional 4 nucleotides (CACC) for ligation into the vector were added to the forward primer and the first 4 nucleotides of the second fragment were added to the reverse primer. For the second fragment, an additional 4 nucleotides (AGGG) for ligation into the vector were added to the reverse primer. All full-length Halo tags were added to the C-terminal ends of the proteins of interest. For the Split Halo constructs used in this study, N-terminal Halo was added to the N-terminal end of Cep55, Alix, Gag, and PSAP, and C-terminal Halo was added to the C-terminal end of Tsg101.

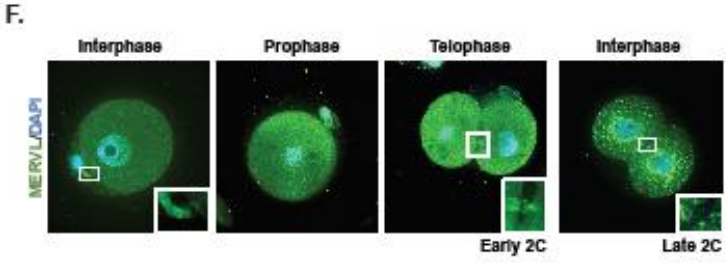
DNA templates were amplified using the following PCR settings: 3 min at 98C, followed by 34 cycles of 98C denaturation for 10 sec, 58C annealing for 30 sec, 72C extension for 30 sec, then 10 min at 72C. The amplified fragments were extracted using a DNA gel extraction kit (Bio Basic) after gel electrophoresis and then quantified using a NanoDrop 2000c (ThermoFisher). For each fusion protein, 20 fmol of both fragments were incubated with vector, Esp3I (an isoschizomer of BsmB1), T4 ligase, and T4 ligase buffer at the following settings: 30 min at 37C, followed by 40 cycles of 5 min at 37C and 5 min at 16C, then 30 min at 37C, and finally 5 min at 80C.

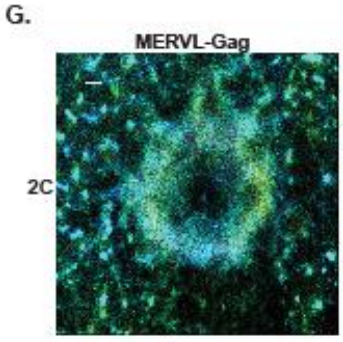
Figures

Figure 1


A.
MERVL LTR Expression
CPM
Pseudotime
MERVL-int Expression
CPM
Pseudotime
B.
Full Length MERVL-int Alignment
C.
MERVL-Gag Percent Identity to Consensus
Count
Percent Identity
D.
MERVL-Gag Single Blastomere Western
Oocyte
2Cell
a-Actinin
Gag
b-Tubulin
100 kDa
64 kDa
50 kDa
GAG Expression (AUC)
Oocyte
2Cell
n
5
4
Mann Whitney U Test, P value = 0.0317


E.
Brightfield
DAPI
MERVL-Gag
2C
4C
Morula
Blastocyt


F.
Interphase
Prophase
Telophase
Interphase
MERVL/DAPI
Early 2C
Late 2C


G.
MERVL-Gag
2C

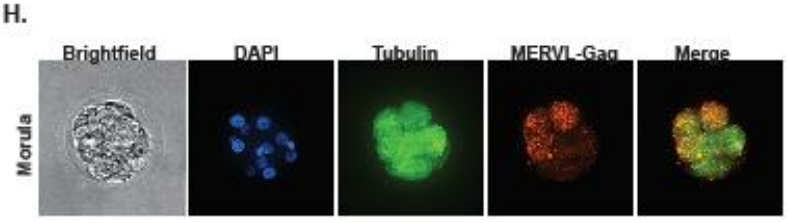
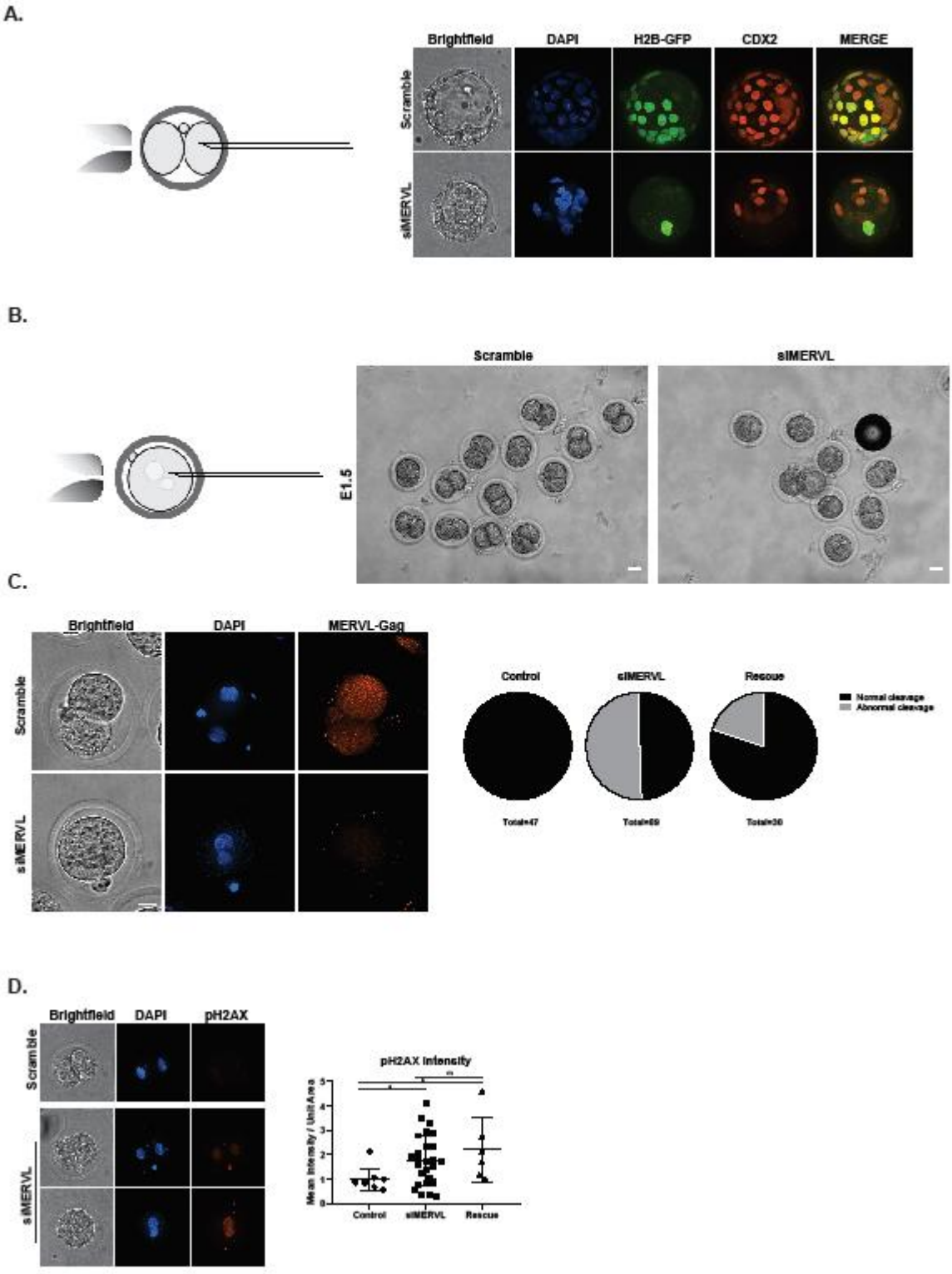
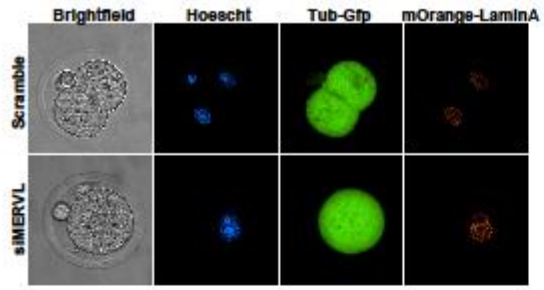

H.
Brightfield
DAPI
Tubulin
MERVL-Gag
Merge
Morula

Figure 1 MERVL-*Gag* expression and dynamics during preimplantation development A. Expression plots for the MERVL long terminal repeat (MERVL LTR) and the internal region of MERVL (MERVL-int) across preimplantation development showing its transient expression from the two cell to the eight cell stage. **B.** Pie chart showing the distribution of complete, truncated, and solo LTR MERVL loci in the mm10 genome (Top left) as well as a nucleotide level alignment of all full length MERVL insertions revealing the striking conservation of MERVL-*Gag* compared to the *Pol* open reading frame. **C.** Histogram of the percent identity each MERVL loci's *Gag* open reading frame shares with the consensus *Gag* at the nucleotide level revealing ~600 insertions with an intact (>90% identity) *Gag*. **D.** Single blastomere western performed for MERVL-*Gag* (yellow), Actin (green) and b-tubulin (Blue) at the oocyte and two cell stages demonstrating signal consistent with the molecular weight of MERVL-*Gag* that significantly increases in two cell stage embryos relative to oocytes. P value = 0.0317 calculated on the basis of a Mann Whitney U Test. **E.** Immunofluorescent staining for MERVL-*Gag* from the two cell to blastocyst stages demonstrating the consistent monotonically decreasing expression profile of MERVL-*Gag* across preimplantation development. **F.** Immunofluorescent staining for MERVL-*Gag* during the one cell to two cell transition demonstrating MERVL-*Gag* localizes to the interphase bridge and the midbody ring. **G.** STORM imaging of MERVL-*Gag* in at the midbody ring in two cell stage embryos. **H.** Immunofluorescent staining for MERVL-*Gag* (red) and Tubulin (green) at the morula stage that shows the asymmetrical expression of MERVL-*Gag* at this stage.

Figure 2


A.
Brightfield
DAPI
H2B-GFP
CDX2
MERGE
Scramble
siMERVL
B.
Scramble
siMERVL
E1.5
C.
Brightfield
DAPI
MERVL-Gag
Scramble
siMERVL
Control
siMERVL
Rescue
Normal cleavage
Abnormal cleavage
Totaln=47
Totaln=69
Totaln=30
D.
Brightfield
DAPI
pH2AX
Scramble
siMERVL
pH2AX Intensity
Mean Intensity / Unit Area
Control
siMERVL
Rescue

E.


Brightfield
Hoescht
Tub-Gfp
mOrange-LaminA
Scramble
siMERVL

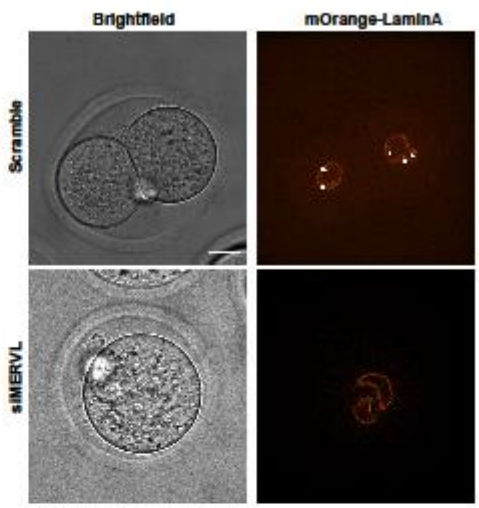

Brightfield
mOrange-LaminA
Scramble
siMERVL

Figure 2 Phenotypic impacts observed upon MERVL depletion A. Schematic of single blastomere knockdown of MERVL (left) and immunofluorescent staining for Cdx2 at the blastocyst stage following delivery of either scramble or control siRNAs, co-delivered with H2B-Gfp mRNA. **B.** Schematic of one cell stage knockdown of MERVL (left) as well as brightfield images of scramble or siMERVL treated embryos. **C.** Immunofluorescent staining for pH2AX in scramble or siMERVL treated embryos as well as quantification demonstrating the significant increase in pH2AX after MERVL depletion. **E.** Tubulin-Gfp and mOrange-LaminA staining in scramble or siMERVL treated embryos with white arrows highlighting nucleoplasmic LaminA present in scramble, but no siMERVL, treated embryos.

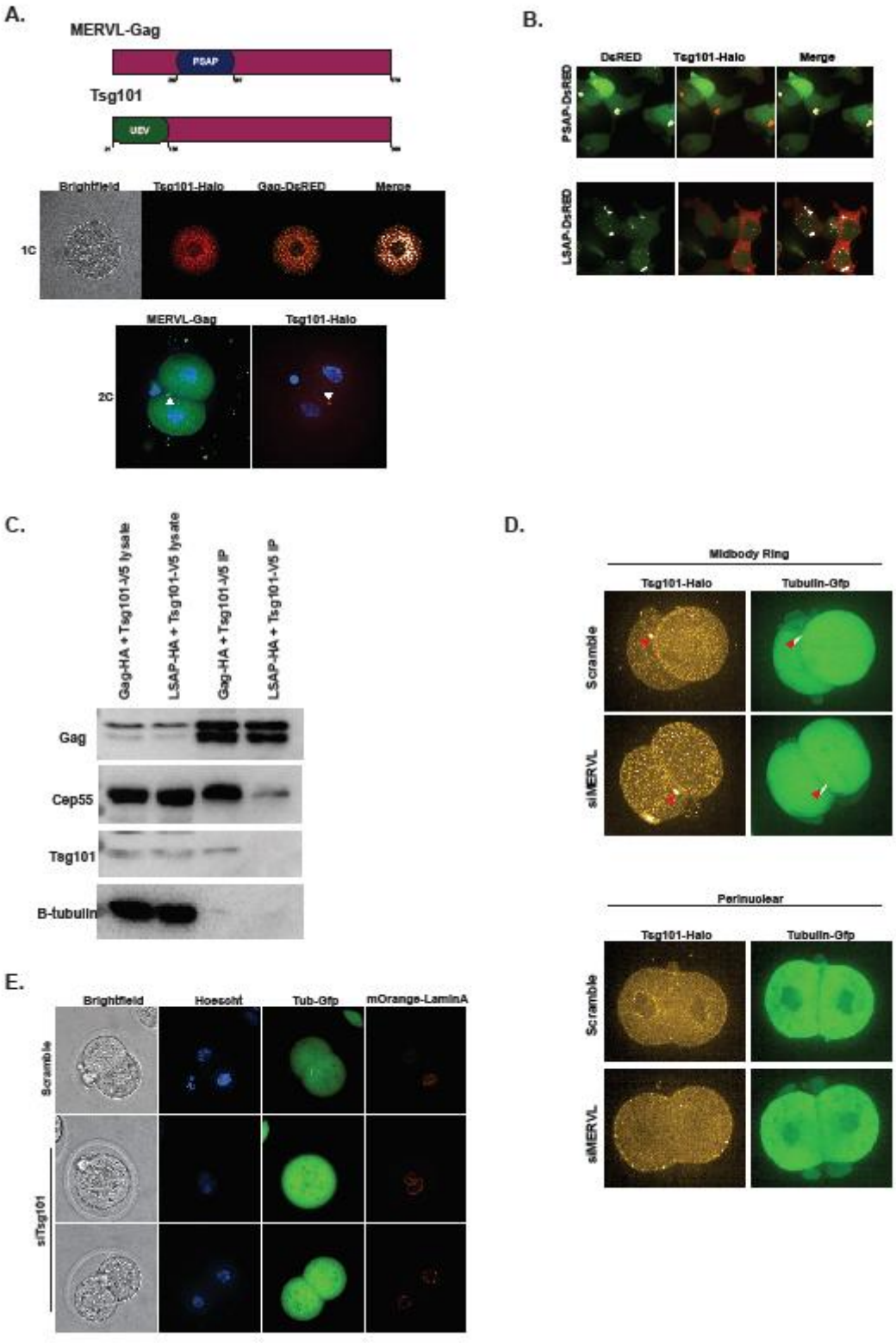

A.
MERVL-Gag
PSAP
Tsg101
UEV
Brightfield
Tsg101-Halo
Gag-DsRED
Merge
1C
MERVL-Gag
Tsg101-Halo
2C
B.
DsRED
Tsg101-Halo
Merge
PSAP-DsRED
LSAP-DsRED
C.
Gag-HA + Tsg101-V5 lysate
LSAP-HA + Tsg101-V5 lysate
Gag-HA + Tsg101-V5 IP
LSAP-HA + Tsg101-V5 IP
Gag
Cep55
Tsg101
B-tubulin
D.
Midbody Ring
Tsg101-Halo
Tubulin-Gfp
Scramble
siMERVL
Perinuolear
Tsg101-Halo
Tubulin-Gfp
Scramble
siMERVL
E.
Brightfield
Hoescht
Tub-Gfp
mOrange-LaminA
Scramble
siTsg101

Figure 3 MERVL-*Gag* directly binds Tsg101 and can alter its subcellular localization A. Schematic of MERVL-*Gag* and Tsg101 proteins (top) as well as co-localization of Tsg101-Halo and Gag-Dsred at the one cell stage. Additionally, immunofluorescent staining for MERVL-*Gag* at the midbody ring and Tsg101-Halo localization to the midbody ring in two cell stage embryos. **B.** Staining for native MERVL-*Gag* (PSAP-DsRED), Tsg101-Halo and MERVL-*Gag* with a mutated PSAP domain (LSAP-DsRED) demonstrating PSAP dependent colocalization between Tsg101 and MERVL-*Gag*. **C.** Western blot summarizing co-immunoprecipitation experiments probing the interactions between MERVL-*Gag*, Tsg101 and Cep55 suggesting that they may form a complex in a manner dependent of MERVL-*Gag's* PSAP domain. **D.** Staining for Tsg101-Halo and Tubulin-Gfp demonstrating that Tsg101 can still localize to the midbody in MERVL depleted embryos but loses its perinuclear localization.

References

Al-Sady, B., Madhani, H.D., and Narlikar, G.J. (2013). Division of labor between the chromodomains of HP1 and Suv39 methylase enables coordination of heterochromatin spread. Mol. Cell.
Alda-Catalinas, C., Bredikhin, D., Hernando-Herraez, I., Santos, F., Kubinyecz, O., Eckersley-Maslin, M.A., Stegle, O., and Reik, W. (2020). A Single-Cell Transcriptomics CRISPR-Activation Screen Identifies Epigenetic Regulators of the Zygotic Genome Activation Program. Cell Syst.
Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol.
Ashoti, A., Alemany, A., Sage, F., and Geijsen, N. (2021). DUX4 induces a homogeneous sequence of molecular changes, culminating in the activation of a stem-cell-like transcriptional network and induction of apoptosis in somatic cells. BioRxiv.
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. Dev.
Bell, N.M., and Lever, A.M.L. (2013). HIV Gag polyprotein: Processing and early viral particle assembly. Trends Microbiol.
Ben-Shahar, T.R., Castillo, A.G., Osborne, M.J., Borden, K.L.B., Kornblatt, J., and Verreault, A. (2009). Two Fundamentally Distinct PCNA Interaction Peptides Contribute to Chromatin Assembly Factor 1 Function. Mol. Cell. Biol.
Bénit, L., De Parseval, N., Casella, J.F., Callebaut, I., Cordonnier, A., and Heidmann, T. (1997). Cloning of a new murine endogenous retrovirus, MuERV-L, with strong similarity to the human HERV-L element and with a gag coding sequence closely related to the Fv1 restriction gene. J. Virol.
Benner, C., Heinz, S., and Glass, C.K. (2017). HOMER - Software for motif discovery and next generation sequencing analysis. Http://Homer.Ucsd.Edu/.
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.
Bessonnard, S., Mot, L. De, Gonze, D., Barriol, M., Dennis, C., Goldbeter, A., Dupont, G., and Chazaud, C. (2014). Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. Dev.
Bessonnard, S., Coqueran, S., Vandormael-Pournin, S., Dufour, A., Artus, J., and Cohen-Tannoudji, M. (2017). ICM conversion to epiblast by FGF/ERK inhibition is limited in time and requires transcription and protein degradation. Sci. Rep.
Boskovic, A., Bing, X.Y., Kaymak, E., and Rando, O.J. (2020). Control of noncoding RNA production and histone levels by a 5' tRNA fragment. Genes Dev.
Bosnakovski, D., Shams, A.S., Yuan, C., da Silva, M.T., Ener, E.T., Baumann, C.W., Lindsay, A.J., Verma, M., Asakura, A., Lowe, D.A., et al. (2020). Transcriptional and cytopathological hallmarks of FSHD in chronic DUX4-expressing mice. J. Clin. Invest.
Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.
Bryja, V., Bonilla, S., and Arenas, E. (2006). Derivation of mouse embryonic stem cells. Nat. Protoc.
Canzio, D., Liao, M., Naber, N., Pate, E., Larson, A., Wu, S., Marina, D.B., Garcia, J.F., Madhani, H.D., Cooke, R., et al. (2013). A conformational switch in HP1 releases auto-inhibition to drive heterochromatin assembly. Nature.
Carlton, J.G., and Martin-Serrano, J. (2007). Parallels between cytokinesis and retroviral

budding: A role for the ESCRT machinery. Science (80-.).
Casa, V., and Gabellini, D. (2012). A repetitive elements perspective in Polycomb epigenetics. Front. Genet.
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*.
De Castro, I.J., Gil, R.S., Ligammari, L., Laura, M., Giacinto, D., and Vagnarelli, P. (2018). Oncotarget 7763 www.impactjournals.com/oncotarget CDK1 and PLK1 coordinate the disassembly and reassembly of the nuclear envelope in vertebrate mitosis. Oncotarget.
Chen, C., Liu, W., Guo, J., Liu, Y., Liu, X., Liu, J., Dou, X., Le, R., Huang, Y., Li, C., et al. (2021). Nuclear m6A reader YTHDC1 regulates the scaffold function of LINE1 RNA in mouse ESCs and early embryos. Protein Cell.
Choi, Y.J., Lin, C.P., Ho, J.J., He, X., Okada, N., Bu, P., Zhong, Y., Kim, S.Y., Bennett, M.J., Chen, C., et al. (2011a). MiR-34 miRNAs provide a barrier for somatic cell reprogramming. Nat. Cell Biol.
Choi, Y.J., Lin, C.P., Ho, J.J., He, X., Okada, N., Bu, P., Zhong, Y., Kim, S.Y., Bennett, M.J., Chen, C., et al. (2011b). MiR-34 miRNAs provide a barrier for somatic cell reprogramming. Nat. Cell Biol.
Choi, Y.J., Lin, C.P., Risso, D., Chen, S., Kim, T.A., Tan, M.H., Li, J.B., Wu, Y., Chen, C., Xuan, Z., et al. (2017a). Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. Science (80-.).
Choi, Y.J., Lin, C.P., Risso, D., Chen, S., Kim, T.A., Tan, M.H., Li, J.B., Wu, Y., Chen, C., Xuan, Z., et al. (2017b). Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. Science (80-.). *355*.
Clift, D., So, C., McEwan, W.A., James, L.C., and Schuh, M. (2018). Acute and rapid degradation of endogenous proteins by Trim-Away. Nat. Protoc.
Courtois, A., Schuh, M., Ellenberg, J., and Hiiragi, T. (2012). The transition from meiotic to mitotic spindle assembly is gradual during early mammalian development. J. Cell Biol.
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.
Deng, M., and Li, R. (2009). Sperm chromatin-induced ectopic polar body extrusion in mouse eggs after ICSI and delayed egg activation. PLoS One.
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 (80-.). *343*, 193–196.
Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: Ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.
Dupressoir, A., Vernochet, C., Bawa, O., Harper, F., Pierron, G., Opolon, P., and Heidmann, T. (2009). Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. Proc. Natl. Acad. Sci. U. S. A.
Dupressoir, A., Vernochet, C., Harper, F., Guégan, J., Dessen, P., Pierron, G., and Heidmann, T. (2011). A pair of co-opted retroviral envelope syncytin genes is required for formation of the two-layered murine placental syncytiotrophoblast. Proc. Natl. Acad. Sci. U. S. A.
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.
Eckersley-Maslin, M.A., Svensson, V., Krueger, C., Stubbs, T.M., Giehr, P., Krueger, F., Miragaia, R.J., Kyriakopoulos, C., Berrens, R. V., Milagre, I., et al. (2016). MERVL/Zscan4 Network Activation Results in Transient Genome-wide DNA Demethylation of mESCs. Cell Rep.
Egan, C.M., Nyman, U., Skotte, J., Streubel, G., Turner, S., O'Connell, D.J., Rraklli, V., Dolan,

M.J., Chadderton, N., Hansen, K., et al. (2013). CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression. Dev. Cell.
Elia, N., Sougrat, R., Spurlin, T.A., Hurley, J.H., and Lippincott-Schwartz, J. (2011). Dynamics of endosomal sorting complex required for transport (ESCRT) machinery during cytokinesis and its role in abscission. Proc. Natl. Acad. Sci. U. S. A.
Ema, M., Mori, D., Niwa, H., Hasegawa, Y., Yamanaka, Y., Hitoshi, S., Mimura, J., Kawabe, Y. ichi, 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.
Gifford, R., and Tristem, M. (2003). The evolution, distribution and diversity of endogenous retroviruses. Virus Genes.
Göke, J., Lu, X., Chan, Y.S., Ng, H.H., Ly, L.H., Sachs, F., and Szczerbinska, I. (2015). Dynamic transcription of distinct classes of endogenous retroviral elements marks specific populations of early human embryonic cells. Cell Stem Cell.
Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: The reference human genome annotation for the ENCODE project. Genome Res. *22*, 1760–1774.
Hatanaka, Y., Inoue, K., Oikawa, M., Kamimura, S., Ogonuki, N., Kodama, E.N., Ohkawa, Y., Tsukada, Y.I., and Ogura, A. (2015). Histone chaperone CAF-1 mediates repressive histone modifications to protect preimplantation mouse embryos from endogenous retrotransposons. Proc. Natl. Acad. Sci. U. S. A.
Haussecker, D., Huang, Y., Lau, A., Parameswaran, P., Fire, A.Z., and Kay, M.A. (2010). Human tRNA-derived small RNAs in the global regulation of RNA silencing. RNA.
Hayashi, M., Maehara, K., Harada, A., Semba, Y., Kudo, K., Takahashi, H., Oki, S., Meno, C., Ichiyanagi, K., Akashi, K., et al. (2016). Chd5 Regulates MuERV-L/MERVL Expression in Mouse Embryonic Stem Cells Via H3K27me3 Modification and Histone H3.1/H3.2. J. Cell. Biochem.
van der Heijden, G.W., Derijck, A.A.H.A., Ramos, L., Giele, M., van der Vlag, J., and de Boer, P. (2006). Transmission of modified nucleosomes from the mouse male germline to the zygote and subsequent remodeling of paternal chromatin. Dev. Biol.
Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017a). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet.
Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.-W.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017b). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. Nat. Genet. *49*, 925–934.
Hirate, Y., Hirahara, S., Inoue, K. ichi, Kiyonari, H., Niwa, H., and Sasaki, H. (2015). Par-aPKC-dependent and -independent mechanisms cooperatively control cell polarity, Hippo signaling, and cell positioning in 16-cell stage mouse embryos. Dev. Growth Differ.
Home, P., Ray, S., Dutta, D., Bronshteyn, I., Larson, M., and Paul, S. (2009). GATA3 is selectively expressed in the trophectoderm of peri-implantation embryo and directly regulates Cdx2 gene expression. J. Biol. Chem.
Hu, C.K., Coughlin, M., and Mitchison, T.J. (2012). Midbody assembly and its regulation during cytokinesis. Mol. Biol. Cell.
Hu, Z., Tan, D.E.K., Chia, G., Tan, H., Leong, H.F., Chen, B.J., Lau, M.S., Tan, K.Y.S., Bi, X., Yang, D., et al. (2020). Maternal factor NELFA drives a 2C-like state in mouse embryonic stem cells. Nat. Cell Biol.
Huang, Y., Kim, J.K., Do, D.V., Lee, C., Penfold, C.A., Zylicz, J.J., Marioni, J.C., Hackett, J.A., and Surani, M.A. (2017). Stella modulates transcriptional and endogenous retrovirus programs during maternal-to-zygotic transition. Elife.

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., and Wheeler, T.J. (2016). The Dfam database of repetitive DNA families. Nucleic Acids Res.
Iaco, A. De, Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). A family of double-homeodomain transcription factors regulates zygotic genome activation in placental mammals. Nat. Genet.
De Iaco, A., Coudray, A., Duc, J., and Trono, D. (2019). DPPA2 and DPPA4 are necessary to establish a 2C-like state in mouse embryonic stem cells. EMBO Rep.
Ishiuchi, T., Enriquez-Gasca, R., Mizutani, E., Boškoviä, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.E. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. Nat. Struct. Mol. Biol.
Johnson, M.H., and Ziomek, C.A. (1981). Induction of polarity in mouse 8-cell blastomeres: Specificity, geometry, and stability. J. Cell Biol.
Judd, J., Sanderson, H., and Feschotte, C. (2020). Evolution of mouse circadian enhancers from transposable elements. BioRxiv.
Kadokawa, Y., Kato, Y., and Eguchi, G. (1987). Cell lineage analysis of the primitive and visceral endoderm of mouse embryos cultured in vitro. Cell Differ.
Kahn, T.G., Dorafshan, E., Schultheis, D., Zare, A., Stenberg, P., Reim, I., Pirrotta, V., and Schwartz, Y.B. (2016). Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. Nucleic Acids Res.
Kang, M., Piliszek, A., Artus, J., and Hadjantonakis, A.K. (2013). FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse. Dev.
Kang, M., Garg, V., and Hadjantonakis, A.K. (2017). Lineage Establishment and Progression within the Inner Cell Mass of the Mouse Blastocyst Requires FGFR1 and FGFR2. Dev. Cell.
Katayama, M., Ellersieck, M.R., and Roberts, R.M. (2010). Development of monozygotic twin mouse embryos from the time of blastomere separation at the two-cell stage to blastocyst. Biol. Reprod.
Kelly, S.J. (1977). Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. J. Exp. Zool.
Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. Biol. Reprod.
Kim, J. Do, Faulk, C., and Kim, J. (2007). Retroposition and evolution of the DNA-binding motifs of YY1, YY2 and REX1. Nucleic Acids Res.
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.
Kovacs, G., Szabo, V., and Pirity, M.K. (2016). Absence of rybp compromises neural differentiation of embryonic stem cells. Stem Cells Int.
Kruse, K., Díaz, N., Enriquez-Gasca, R., Gaume, X., Torres-Padilla, M.-E., and Vaquerizas, J. (2019). Transposable elements drive reorganisation of 3D chromatin during early embryogenesis. BioRxiv.
Kumar, R.P., Ray, S., Home, P., Saha, B., Bhattacharya, B., Wilkins, H.M., Chavan, H., Ganguly, A., Milano-Foster, J., Paul, A., et al. (2018). Regulation of energy metabolism during early mammalian development: Tead4 controls mitochondrial transcription. Dev.
Lawson, K.A., Meneses, J.J., and Pedersen, R.A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. Development.
Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., and Sheen, F. miin (1993). Transposons in place of telomeric repeats at a Drosophila telomere. Cell.
Li, H., Lai, P., Jia, J., Song, Y., Xia, Q., Huang, K., He, N., Ping, W., Chen, J., Yang, Z., et al. (2017). RNA Helicase DDX5 Inhibits Reprogramming to Pluripotency by miRNA-Based

Repression of RYBP and its PRC1-Dependent and -Independent Functions. Cell Stem Cell.
Liao, Y., Smyth, G.K., and Shi, W. (2014). FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
Lin, S.-C.J., Wani, M. a, Whitsett, J. a, and Wells, J.M. (2010). Klf5 regulates lineage formation in the pre-implantation mouse embryo. Development *137*, 3953–3963.
Lorthongpanich, C., Messerschmidt, D.M., Chan, S.W., Hong, W., Knowles, B.B., and Solter, D. (2013). Temporal reduction of LATS kinases in the early preimplantation embryo prevents ICM lineage differentiation. Genes Dev.
Lu, X., Sachs, F., Ramsay, L.A., Jacques, P.É., Göke, J., Bourque, G., and Ng, H.H. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol.
Macfarlan, T.S., Gifford, W.D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., Andrews, S.E., Franco, L., Rosenfeld, M.G., Ren, B., et al. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev.
Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature.
Maeda, Y., Yanagimachi, H., Tateno, H., Usui, N., and Yanagimachi, R. (1998). Decondensation of the mouse sperm nucleus within the interphase nucleus. Zygote.
Maemura, M., Taketsuru, H., Nakajima, Y., Shao, R., Kakihara, A., Nogami, J., Ohkawa, Y., and Tsukada, Y. ichi (2021). Totipotency of mouse zygotes extends to single blastomeres of embryos at the four-cell stage. Sci. Rep.
Maksakova, I.A., Thompson, P.J., Goyal, P., Jones, S.J.M., Singh, P.B., Karimi, M.M., and Lorincz, M.C. (2013). Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. Epigenetics and Chromatin.
Manandhar, G., Sutovsky, P., Joshi, H.C., Stearns, T., and Schatten, G. (1998). Centrosome reduction during mouse spermiogenesis. Dev. Biol.
Manandhar, G., Schatten, H., and Sutovsky, P. (2005). Centrosome reduction during gametogenesis and its significance. Biol. Reprod.
Martin-Serrano, J., Perez-Caballero, D., and Bieniasz, P.D. (2004). Context-Dependent Effects of L Domains and Ubiquitination on Viral Budding. J. Virol.
Mathelier, A., Fornes, O., Arenillas, D.J., Chen, C.Y., 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.
McLaren, A., Molland, P., and Signer, E. (1995). Does monozygotic twinning occur in mice? Genet. Res.
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.
Mihajlovic, A.I., Thamodaran, V., and Bruce, A.W. (2015). The first two cell-fate decisions of preimplantation mouse embryo development are not functionally independent. Sci. Rep.
Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell.
Modzelewski, A.A., Shao, W., Chen, J., Lee, A., Xin, F., Noon, M., Tjokro, K., Sales, G., Biton, A., Speed, T., et al. (2021). Title : A species-specific retrotransposon drives a conserved Cdk2ap1 isoform essential for preimplantation development. BioRxiv.
Modzelewski, A.J., Chen, S., Willis, B.J., Lloyd, K.C.K., Wood, J.A., and He, L. (2018). Efficient mouse genome engineering by CRISPR-EZ technology. Nat. Protoc.
Morita, E., Sandrin, V., Chung, H.Y., Morham, S.G., Gygi, S.P., Rodesch, C.K., and Sundquist, W.I. (2007). Human ESCRT and ALIX proteins interact with proteins of the midbody and

function in cytokinesis. EMBO J.
Müllers, E. (2013). The foamy virus gag proteins: What makes them different? Viruses.
Nishioka, N., Yamamoto, S., Kiyonari, H., Sato, H., Sawada, A., Ota, M., Nakao, K., and Sasaki, H. (2008). Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mech. Dev.
Nishioka, N., Inoue, K. ichi, Adachi, K., Kiyonari, H., Ota, M., Ralston, A., Yabuta, N., Hirahara, S., Stephenson, R.O., Ogonuki, N., et al. (2009). The Hippo Signaling Pathway Components Lats and Yap Pattern Tead4 Activity to Distinguish Mouse Trophectoderm from Inner Cell Mass. Dev. Cell.
Ono, R., Nakamura, K., Inoue, K., Naruse, M., Usami, T., Wakisaka-Saito, N., Hino, T., Suzuki-Migishima, R., Ogonuki, N., Miki, H., et al. (2006). Deletion of Peg10, an imprinted gene acquired from a retrotransposon, causes early embryonic lethality. Nat. Genet.
Pastuzyn, E.D., Day, C.E., Kearns, R.B., Kyrke-Smith, M., Taibi, A. V., McCormick, J., Yoder, N., Belnap, D.M., Erlendsson, S., Morado, D.R., et al. (2018). The Neuronal Gene Arc Encodes a Repurposed Retrotransposon Gag Protein that Mediates Intercellular RNA Transfer. Cell.
Percharde, M., Lin, C.J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. Cell.
Pereira, V. (2008). Automated paleontology of repetitive DNA with REANNOTATE. BMC Genomics.
Pérez-Palacios, R., Macías-Redondo, S., Climent, M., Contreras-Moreira, B., Muniesa, P., and Schoorlemmer, J. (2016). In vivo chromatin targets of the transcription factor Yin Yang 2 in trophoblast stem cells. PLoS One.
Plusa, B., Piliszek, A., Frankenberg, S., Artus, J., and Hadjantonakis, A.K. (2008). Distinct sequential cell behaviours direct primitive endoderm formation in the mouse blastocyst. Development.
Pornillos, O., Higginson, D.S., Stray, K.M., Fisher, R.D., Garrus, J.E., Payne, M., He, G.P., Wang, H.E., Morham, S.G., and Sundquist, W.I. (2003). HIV Gag mimics the Tsg101-recruiting activity of the human Hrs protein. J. Cell Biol.
Posfai, E., Petropoulos, S., de Barros, F.R.O., Schell, J.P., Jurisica, I., Sandberg, R., Lanner, F., and Rossant, J. (2017). Position- and hippo signaling-dependent plasticity during lineage segregation in the early mouse embryo. Elife.
Presnell, J.S., Schnitzler, C.E., and Browne, W.E. (2015). KLF/SP transcription factor family evolution: Expansion, diversification, and innovation in eukaryotes. Genome Biol. Evol.
Quivy, J.P., Gérard, A., Cook, A.J.L., Roche, D., and Almouzni, G. (2008). The HP1-p150/CAF-1 interaction is required for pericentric heterochromatin replication and S-phase progression in mouse cells. Nat. Struct. Mol. Biol.
Ribet, D., Louvet-Vallée, S., Harper, F., de Parseval, N., Dewannieux, M., Heidmann, O., Pierron, G., Maro, B., and Heidmann, T. (2008). Murine Endogenous Retrovirus MuERV-L Is the Progenitor of the “Orphan” Epsilon Viruslike Particles of the Early Mouse Embryo. J. Virol.
Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47–e47.
Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler, B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between polycomb repressive complexes. Elife.
Rowe, H.M., Jakobsson, J., Mesnard, D., Rougemont, J., Reynard, S., Aktas, T., Maillard, P. V., Layard-Liesching, H., Verp, S., Marquis, J., et al. (2010). KAP1 controls endogenous retroviruses in embryonic stem cells. Nature.
Santoni, F.A., Guerra, J., and Luban, J. (2012). HERV-H RNA is abundant in human embryonic stem cells and a precise marker for pluripotency. Retrovirology.

Santos, F., Peters, A.H., Otte, A.P., Reik, W., and Dean, W. (2005). Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. Dev. Biol.
Sasaki, H. (2015). Position- and polarity-dependent Hippo signaling regulates cell fates in preimplantation mouse embryos. Semin. Cell Dev. Biol.
Scheffler, K., Uraji, J., Jentoft, I., Cavazza, T., Mönnich, E., Mogessie, B., and Schuh, M. (2021). Two mechanisms drive pronuclear migration in mouse zygotes. Nat. Commun.
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.
Schrode, N., Saiz, N., Di Talia, S., and Hadjantonakis, A.K. (2014). GATA6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. Dev. Cell.
Schuh, M., and Ellenberg, J. (2007). Self-Organization of MTOCs Replaces Centrosome Function during Acentrosomal Spindle Assembly in Live Mouse Oocytes. Cell.
Sekita, Y., Wagatsuma, H., Nakamura, K., Ono, R., Kagami, M., Wakisaka, N., Hino, T., Suzuki-Migishima, R., Kohda, T., Ogura, A., et al. (2008). Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. Nat. Genet.
Sharma, U., Conine, C.C., Shea, J.M., Boskovic, A., Derr, A.G., Bing, X.Y., Belleannee, C., Kucukural, A., Serra, R.W., Sun, F., et al. (2016). Biogenesis and function of tRNA fragments during sperm maturation and fertilization in mammals. Science (80-.).
Sotomaru, Y., Kato, Y., and Tsunoda, Y. (1998). Production of Monozygotic Twins after Freezing and Thawing of Bisected Mouse Embryos. Cryobiology.
Street, K., Risso, D., Fletcher, R.B., 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*.
Strumpf, D., Mao, C.A., 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.
Teissandier, A., Servant, N., Barillot, E., and Bourc'His, D. (2019). Tools and best practices for retrotransposon analysis using high-throughput sequencing data. Mob. DNA.
Todd, C.D., Deniz, Ö., Taylor, D., and Branco, M.R. (2019). Functional evaluation of transposable elements as enhancers in mouse embryonic and trophoblast stem cells. Elife.
TOGASHI, M., SUZUKI, H., MIYAI, T., and OKAMOTO, M.T. (1987). Production of monozygotic twins by splitting of 2-cell stage embryos in mice. Japanese J. Anim. Reprod.
Tsunoda, Y., and McLaren, A. (1983). Effect of various procedures on the viability of mouse embryos containing half the normal number of blastomeres. J. Reprod. Fertil.
Ujhelly, O., Szabo, V., Kovacs, G., Vajda, F., Mallok, S., Prorok, J., Acsai, K., Hegedus, Z., Krebs, S., Dinnyes, A., et al. (2015). Lack of Rybp in Mouse Embryonic Stem Cells Impairs Cardiac Differentiation. Stem Cells Dev.
Usami, Y., Popov, S., Popova, E., Inoue, M., Weissenhorn, W., and Göttlinger, H.G. (2009). The ESCRT pathway and HIV-1 budding. Biochem. Soc. Trans.
Verreault, A., Kaufman, P.D., Kobayashi, R., and Stillman, B. (1996). Nucleosome assembly by a complex of CAF-1 and acetylated histones H3/H4. Cell.
Votteler, J., and Sundquist, W.I. (2013). Virus budding and the ESCRT pathway. Cell Host Microbe.
Wagner, K.-U., Krempler, A., Qi, Y., Park, K., Henry, M.D., Triplett, A.A., Riedlinger, G., Rucker, E.B., and Hennighausen, L. (2003). Tsg101 Is Essential for Cell Growth, Proliferation, and Cell Survival of Embryonic and Adult Tissues. Mol. Cell. Biol.
Wang, W.C.H., and Shashikant, C.S. (2007). Evidence for positive and negative regulation of the mouse Cdx2 gene. J. Exp. Zool. Part B Mol. Dev. Evol.
Wang, J., Singh, M., Sun, C., Besser, D., Prigione, A., Ivics, Z., Hurst, L.D., and Izsvák, Z. (2016). Isolation and cultivation of naive-like human pluripotent stem cells based on HERVH

expression. Nat. Protoc.
Wang, M., Kato, Y., and Tsunoda, Y. (1997). Effects of several factors on the monozygotic twin production in the mouse. J. Reprod. Dev.
Wang, Q., Racowsky, C., and Deng, M. (2011). Mechanism of the chromosome-induced polar body extrusion in mouse eggs. Cell Div.
Wang, Y., Long, H., Yu, J., Dong, L., Wassef, M., Zhuo, B., Li, X., Zhao, J., Wang, M., Liu, C., et al. (2018). Histone variants H2A.Z and H3.3 coordinately regulate PRC2-dependent H3K27me3 deposition and gene expression regulation in mES cells. BMC Biol.
Wang, Y., Na, Q., Li, X., Tee, W.W., Wu, B., and Bao, S. (2021). Retinoic acid induces NELFA-mediated 2C-like state of mouse embryonic stem cells associates with epigenetic modifications and metabolic processes in chemically defined media. Cell Prolif.
Xenopoulos, P., Kang, M., Puliafito, A., DiTalia, S., and Hadjantonakis, A.K. (2015). Heterogeneities in nanog expression drive stable commitment to pluripotency in the mouse blastocyst. Cell Rep.
Yan, Y.L., Zhang, C., Hao, J., Wang, X.L., 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.
Zalzman, M., Falco, G., Sharova, L. V., Nishiyama, A., Thomas, M., Lee, S.L., Stagg, C.A., Hoang, H.G., Yang, H. Te, Indig, F.E., et al. (2010). Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature.
Zhang, W., Chen, F., Chen, R., Xie, D., Yang, J., Zhao, X., Guo, R., Zhang, Y., Shen, Y., Göke, J., et al. (2019). Zscan4c activates endogenous retrovirus MERVL and cleavage embryo genes. Nucleic Acids Res.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.
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.