ES cell potency fluctuates with endogenous retrovirus activity

Todd S. Macfarlan[1,2], **Wesley D. Gifford**[1], **Shawn Driscoll**[1], **Karen Lettieri**[1], **Helen M. Rowe**[3], **Dario Bonanomi**[1], **Amy Firth**[4], **Oded Singer**[4], **Didier Trono**[3], and **Samuel L. Pfaff**[1,*]

[1]Howard Hughes Medical Institute, Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines, La Jolla, CA, 92037, USA [2]Current Address, Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD 20892 USA [3]School of Life Sciences, Ecole Polytechnique Federale de Lausanne (EPFL), 1015 Lausanne, Switzerland [4]Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines, La Jolla, CA, 92037, USA

Summary

Embryonic stem (ES) cells are derived from blastocyst stage embryos and are believed to be functionally equivalent to the inner cell mass, which lacks the ability to produce all extraembryonic tissues. Here we report the identification of a rare transient cell population within mouse ES and induced pluripotent stem (iPS) cell cultures that express high levels of transcripts found in two-cell (2C) embryos in which the blastomeres are totipotent. We genetically tagged these 2C-like ES cells and show that they lack the ICM pluripotency proteins Oct4, Sox2, and Nanog and have acquired the ability to contribute to both embryonic and extraembryonic tissues. We show that nearly all ES cells cycle in and out of this privileged state, which we find is partially controlled by histone modifying enzymes. Transcriptome sequencing and bioinformatic analyses revealed that a significant number of 2C-transcripts are initiated from long terminal repeats derived from murine endogenous retroviruses, suggesting this foreign sequence has helped to drive cell fate regulation in placental mammals.

The zygote and its daughter cells are totipotent because they are able to develop into all embryonic and extraembryonic cell types[1,2]. The progeny of these first two daughter cells become progressively more fate restricted as they activate distinct patterns of gene expression that first direct them toward one of three broad lineages: Oct4/Sox2/Nanog[+] epiblast cells that give rise to the embryo, Gata4/6[+] primitive endoderm cells that contribute to extraembryonic membranes that encase the embryo, and Cdx2[+] trophectoderm cells that form a large part of the placenta[3]. These early cell fate decisions represent a major and relatively recent advance in mammalian evolution in which the placenta and extraembryonic tissues that support the intrauterine nourishment of the fetus allow development to progress further before birth. The epigenetic landscape of the zygote changes dramatically during the first cell divisions. Shortly after fertilization the oocyte maternal transcripts are replaced

with newly synthesized RNAs generated by activating transcription of the zygotic genome[4–6]. The unique transcriptional profile of the zygote and its daughter cells defines a brief period when the cells are totipotent.

Murine ES cells are isolated from the inner cell mass (ICM) of blastocysts that have already become a separate lineage from the trophectoderm[7,8]. ICM-derived ES cells are regarded as pluripotent because they have the capacity to generate tissues of the fetus but are extremely inefficient at colonizing the extraembryonic tissues[9]. The rare contribution of ES cells to extraembryonic tissues could be explained by contamination of ES cultures with trophectoderm or primitive endoderm-committed cells, or occur because rare ES cells have acquired the ability to produce extra-embryonic tissues in addition to embryonic tissues. This latter possibility is intriguing because recent evidence shows that ES cultures are a heterogeneous mixture of metastable cells with fluctuating expression of genes such as *Zscan4*, *Stella*, *Nanog*, *Sox17* and *Gata6*, which could account for special attributes of individual cells[10–14].

A large number of retrotransposons are expressed when the zygotic genome is first transcribed, including the endogenous retroviruses (ERVs), LINE-1 elements, and the non-autonomous SINE elements[15]. At the 2C stage, MuERV-L/MERVL retrovirus-like elements are transiently de-repressed and produce 3% of the transcribed mRNAs[15–17]. Following the 2C stage, MERVL-retroelement expression is silenced[18,19]. We discovered that this regulated pattern of MERVL expression overlapped with greater than one hundred 2C-specific genes that have co-opted regulatory elements from these foreign retroviruses to initiate their transcription. We exploited the regulated activity of these 2C virus-derived promoters to label cells and found that both ES and iPS cultures contain a small but relatively constant fraction of cells that have entered into the 2C-transcriptional state. Purification of these 2C-like cells reveals that they have unique developmental characteristics and efficiently produce progeny for extraembryonic and embryonic lineages.

Identification of a 2C-like state within ES cultures

To identify zygotically activated genes we performed deep RNA-sequencing on mouse oocytes and 2C stage embryos. A comparison of the transcripts in these cells identified a large number of genes and retrotransposons that became expressed in the 2C embryo as well as numerous transcripts that were downregulated (Fig. 1a, Supplementary Table 1). The most highly activated repeat was the MERVL-family of retroviruses and their corresponding long terminal repeat (LTR) promoters (Mt2_mm) which were activated greater than 300-fold (Supplementary Table 1). Sequence alignments revealed that greater than 25% of the nearly 700 copies of MERVL elements were activated, and that 307 genes generated chimeric transcripts with junctions to MERVL elements (Fig.1a, Supplementary Table 2), including 10 that were previously described[15]. Of the 626 chimeric transcripts generated, >90% were 5'LTR- exon fusions that generated open reading frames (ORFs), suggesting that these LTRs had become functional promoters for protein coding genes (Fig. 1b, Supplementary Fig. 1a). The most significantly enriched Gene Ontology categories representing these chimeric proteins were regulation of transcription, ion binding, translation, nucleotide binding, and mRNA transport (Fig. 1c). Two notable transcription factors that utilized alternate MERVL-LTR promoters were *Gata4* and *Tead4*, factors important for the specification of primitive endoderm and trophectoderm, respectively[20–22].

Since greater than 300 out of nearly 700 copies of the MERVL endogenous retroviruses still encode Gag viral protein, we stained 2C and early blastocyst embryos to confirm that viral Gag was expressed and developmentally regulated. We found that 2C embryos express Gag but lack the pluripotency marker Oct4, whereas blastula cells lack Gag but express Oct4

(Fig. 1d–e). Thus, MERVL activity is developmentally regulated and these retroviral promoters have been co-opted by many cellular genes to impose tight control over their expression.

Next we asked whether it was possible to use the regulatory sequences from MERVL elements to label 2C cells. We cloned the MERVL 5'LTR, primer binding site, and a portion of the Gag gene upstream of a tandem-tomato fluorescent reporter (*2C::tomato*). We injected fertilized eggs with the *2C::tomato* construct and monitored the expression of tomato during culture *in vitro*. Tomato expression was highest in arrested zygotes and 2C embryos and became downregulated at the morula stage (Fig. 1f, Supplementary Movie 1). Interestingly, when we introduced the *2C::tomato* construct into ES cells and selected for clonal stable integrants, we found several colonies that contained 1–5 cells that were strongly labeled with tomato amongst cells lacking expression of the reporter (Fig. 1g). Importantly, we also found that rare ES cells expressed MERVL mRNA and Gag protein, and that these overlapped with *2C::tomato*+ cells (Fig. 1g, Supplementary Fig. 1b–c). The correspondence between the *2C::tomato* reporter and the expression of MERVL was further confirmed by immunoblotting, and EM imaging of viral epsilon particles encoded by MERVL within the endoplasmic reticulum of tomato+ cells but not tomato− cells (Supplementary Fig. 1d, e). Thus MERVL expression is restricted *in vivo* to 1-4C embryos and is reactivated within a small subpopulation of ES cells derived from blastocysts.

To characterize the unexpected *2C::tomato*-labeled cells within ES cultures we sorted tomato+ and tomato− cells and performed microarray and mRNA-sequencing analyses (Fig. 1h, Supplementary Tables 3–5). Tomato+ cells expressed 55 fold higher levels of MERVL transcripts than tomato− cells but the vast majority of other retrotransposons were unaffected (Supplementary Table 3). Strikingly, tomato+ cells had 165 transcripts activated more than 4-fold and zero genes repressed more than 4-fold compared with tomato− cells (Fig. 1h, Supplementary Table 4, Supplementary Fig. 2a–f). Amongst the genes that were highly enriched in tomato+ cells, several were previously shown to be restricted to the 2-4C stage of development, including *Zscan4*, *Tcstv1/3*, *Eif1a*, *Tho4* (*Gm4340*) *Tdpoz*, and *Zfp352*[23–25]. In total, 525 genes that were enriched in *2C::tomato*+ cells were also genes activated at the 2C stage, including 52 genes that generated chimeric transcripts linked to MERVL elements (Supplementary Tables 6–7).

A hallmark of the ICM and ES cells is their expression of Oct4, Sox2, and Nanog, whereas totipotent 2C embryos do not express Oct4 (Fig. 1d–e). We found that *2C::tomato*+ cells within ES cultures also lacked Oct4, Sox2, and Nanog (Fig. 1i, Supplementary Fig. 1f). The reduction in Oct4, Sox2, and Nanog protein-labeling occurred despite changes in their mRNA levels, suggesting the regulation is occurring post-transcriptionally (Fig. 1h, Supplementary Fig. 2g). In summary, *2C::tomato* labels a subset of ES cells that share transcriptional and proteomic features of 2C embryos and display strikingly different patterns of pluripotency markers from the majority of ES cells in culture.

ES cells cycle in and out of the 2C state

We considered the possibility that the expression of the *2C::tomato* reporter and MERVL-Gag protein in sporadic cells within ES cultures might arise from contamination with trophectoderm or primitive endoderm. To exclude this possibility we examined induced pluripotent cells (iPS) derived from mouse fibroblasts since they should not be contaminated with cells from blastocyst embryos. Similar to ES cells, we found that sporadic iPS cells express the MERVL-Gag protein and lack Oct4 (Fig. 1i). Thus, the heterogeneity within ES cultures is a property that is shared with iPS cultures and is unlikely to arise from a cell contaminant.

Next we examined whether the *2C::tomato*[+] cells represent a stable cell population or whether ES cells transition into and out of this 2C-like state. We utilized a Cre-LoxP fate mapping strategy to indelibly mark cells that had expressed 2C genes (Supplementary Fig. 3a–c). We generated a transgenic mouse line using the MERVL regulatory elements driving expression of a tamoxifen inducible Cre recombinase (*2C::ERT2CreERT2*, Supplementary Fig. 3a). These mice were then mated with Cre responsive reporter lines (*ROSA::LSL-tomato* and *ROSA::LSL-LacZ*, Supplementary Fig. 3b). ES cell lines were derived from double positive transgenic blastocysts (Supplementary Fig. 3c). After addition of 4 hydroxytamoxifen (4OHT) to the ES cultures we detected nuclear Cre expression in MERVL-Gag[+] cells (Supplementary Fig. 3d). When ES cultures were grown for 2–6 days with 4OHT we found a steady increase in the percentage of reporter positive cells (Fig. 2a). Remarkably, over extended passages nearly every ES cell activated the reporter (Fig. 2b), demonstrating this transient state is regularly entered by ES cells.

To monitor the kinetics of the interconversion between *2C::tomato*[+] and[−] cells we performed flow cytometry to collect tomato[+] and tomato[−] cells. When these purified subpopulations were cultured we found that tomato[+] cells produced tomato[−] cells and vice versa (Fig. 2c). Within 24 hours nearly 50% of tomato[+] cells convert to tomato[−], independent of the starting percentages of the two cell populations (Fig. 2c and data not shown). Under hypoxic conditions (5% O_2), the percentage of cells expressing the *2C::tomato* reporter was decreased, which could be reversed by shifting the cultures back to 20% O_2 (Fig. 2d). We also found that growing cells for 48 hours "ground state" media conditions (2i-media[26]) reduced but did not eliminate the presence of tomato[+] cells relative to media containing knockout serum replacement (KOSR), suggesting extrinsic and intrinsic mechanisms regulate the MERVL/2C gene network (Fig. 2e).

The 2C-ES switch is regulated by histone modification

After activation of the zygotic genome in mouse development, histone deacetylation and histone H1 synthesis leads to the formation of repressive chromatin which is thought to limit the broad pattern of transcription present in 2C embryos[27,28]. Using indirect immunofluorescence, we found that tomato[+] cells had significantly higher levels of active histone marks including methylation of histone 3 lysine 4 (H3 K4) and acetylation of H3 and H4 (Supplementary Fig. 4a), a finding confirmed using immunoblot analysis of sorted cell populations (Fig. 3a). This type of chromatin mirrors that found in two cell embryos[28]. Next we tested whether tomato[+] cells had different levels of DNA methylation compared to non-labeled ES cells. We found that the MERVL sequences were hypomethylated in tomato[+] cells compared to tomato[−] cells. In contrast to the MERVL sequences, IAP retroviruses were highly methylated in both tomato[+] and tomato[−] cells suggesting the altered pattern of methylation was not uniform across the genome (Supplementary Fig. 4b). In sum these data suggest that as ES cells (re)enter into the 2C state, their chromatin and DNA is altered to favor transcription in a way that mirrors the two-cell embryo.

We previously demonstrated that MERVL and 2C-specific genes were increased in the absence of the histone lysine specific demethylase *Lsd1/Kdm1a*[29]. To test whether other proviral co-repressors and histone-modifying enzymes also influence these genes we profiled the transcriptome of ES cells with homozygous mutations in the KRAB box associated transcriptional repressor *Kap1* and the histone H3 K9 methyltranferase *G9a*[29–31]. We found that MERVL and a number of 2C genes were significantly upregulated in *Kdm1a*, *Kap1*, and *G9a* mutant ES cells (Fig. 3b, Supplementary Fig. 5b–c, Supplementary Table 7). These findings were confirmed using *in situ* hybridization and immunofluorescence microscopy (Fig. 3c, Supplementary Fig. 5d). Treatment of *2C::tomato* ES lines with the histone deacetylase inhibitor TSA also increased the number of tomato[+] cells four-fold (Fig.

3d). To better understand how 2C gene regulation is controlled when chromatin repressors are acutely eliminated we utilized a stably integrated *2C::tomato* ES line that is homozygous for a floxed allele of *Kdm1a* and contains a Cre-ERT transgene that can be activated with 4OHT. Within 24 hours of deleting *Kdm1a* we found a ten-fold increase in tomato+ cells that was steadily maintained (Fig. 3e). In addition, FACS purified tomato− cells more rapidly became tomato+ in the absence of *Kdm1a* (Fig. 3f), and stayed in this state longer (Supplementary Movies 2–3). These findings suggest that Kdm1a, Kap1, G9a, and histone deacetylases all contribute to the repression of 2C genes in ES cells and that they function by altering the equilibrium between the *2C::tomato*+ and− states.

2C-like ES cells have expanded fate potential

Since 2C-like cells within ES cultures express high levels of 2C-restricted genes found in totipotent blastomeres and reduced levels of pluripotency associated proteins, we reasoned that this subpopulation of ES cells might have distinct functional characteristics. We tested whether *2C::tomato*+ cells have acquired the ability to produce extraembryonic tissues, a characteristic that ES cells lack. We used FACS to collect tomato+ and tomato− cells from a *2C::tomato* ES line and injected four cells into morula stage embryos. The tomato− cells contributed exclusively to the ICM of all 5 chimeric blastocysts analyzed (Fig. 4a). In contrast, the tomato+ cells contributed to the trophectoderm (in 4 of 5 chimeric embryos) in addition to the ICM (in 3 of 5 chimeric embryos) (Fig. 4a). To track the fate of the 2C::tomato+ ES cells later in development, we injected blastocysts with tomato+ or tomato− cells that were pre-infected with a lentivirus encoding GFP from a constitutively active Ef1a promoter (*Ef1a::GFP*). Tomato−/GFP+ cells contributed exclusively to embryonic tissues, whereas tomato+/GFP+ cells contributed to embryonic endoderm, ectoderm, mesoderm, the germ lineage as well as the yolk sac, and placenta (Fig. 4b–c, Supplementary Fig. 6a–b). The extraembryonic contribution of the tomato+/GFP+ cells included trophoblast giant cells of the placenta (Fig. 4c). Thus, the developmental potential of *2C::tomato*+ cells includes embryonic plus extraembryonic tissues in contrast to the majority of ES cells in culture, which are restricted to generating only embryonic cell types.

We next examined whether *Kdm1a* mutant ES lines, that contain higher frequencies of *2C::tomato*+ cells also had increased potency in mouse chimera assays. As expected, *Kdm1a* heterozygous ES cells contributed exclusively to embryonic tissues (in 5/5 chimeric embryos) but never to extraembryonic tissues (Fig. 4d). In contrast, *Kdm1a* homozygous mutant ES cells generated both embryonic (in 4/6 chimeric embryos) and extraembryonic tissues (in 5/6 chimeric embryos) (Fig. 4d). To confirm the increased potential of *Kdm1a* mutant ES cells, we used a competition chimera assay. We co-injected a 1:1 mixture of *Kdm1a* Fl/Fl and KO/KO ES cells into five wild type blastocysts. PCR was then used to detect the appearance of *Kdm1a* Fl/Fl or KO/KO cells in dissected tissues. We detected *Kdm1a* Fl/Fl ES cells in the embryonic tissues and amnion, but not the yolk sac or placenta (Fig. 4e). In contrast, *Kdm1a* mutant ES cells contributed to embryonic tissues, the amnion, yolk sac, and placental tissues, including trophoblast giant cells and PGCs (Fig. 4e–f). Thus, the artificial activation of 2C genes achieved by removing *Kdm1a* is associated with expanded fate potential.

We have shown that *2C:tomato*+ cells within ES cultures have increased potency, but it is unclear whether entrance into this state is essential for their long-term pluripotency. To test this possibility, we performed serial depletion of 2C-like ES cells by genetic ablation with diptheria toxin (DTA). We generated ES lines by crossing *2C::ERT2CreERT2* mice (Supplementary Fig. 3a) with a Cre responsive DTA allele (*ROSA::LSL-DTA*) and treated the ES line with 4OHT for 20 passages (Supplementary Fig. 7a). We found that these 2C-depleted ES cultures were still capable of generating high contribution chimeras

(Supplementary Fig. 7b), although their differentiation was biased toward mesoderm and ectoderm lineages *in vitro* (Supplementary Fig. 7c). These data suggest that occasional entry into the 2C-like state might help to preserve the broad embryonic fate potential of ES cells.

Discussion

In mammalian development, the zygote and its daughter cells progress from totipotent cells capable of generating an entire mouse to more lineage restricted inner and outer cells of the morula capable of generating embryonic or extraembryonic lineages, respectively. A key transcriptional feature of the totipotent cells is the onset of zygote genome activation in which the embryo switches from a maternal to a zygotic transcriptome. To mark cells at this early stage of embryonic development, we generated a reporter with the regulatory elements from the endogenous retrovirus MERVL, which is highly restricted to the zygote/2C stage. Surprisingly, we found that rare ES and iPS cells expressed the reporter. When we characterized these cells, we found that they lacked expression of the pluripotency proteins Oct4, Sox2, and Nanog. Instead, these rare cells expressed a large number of genes restricted to the 2C stage, and most importantly, were capable of forming both embryonic and extraembryonic lineages (Fig. 5a–b). Our studies identify a rare 2C-like cell in ES cultures that has expanded fate potential.

Although it is unclear how MERVL and other 2C genes regulate potency, several lines of evidence indicate that 2C-like cells are required for the health and maintenance of ES cultures. First, we found that nearly all cells enter into the 2C-like state over increasing passage. Second, when we depleted 2C-like ES cells from cell cultures we found that their differentiation characteristics were altered to generate more ectoderm and mesoderm derivatives. Third, functional studies of the *Zscan4* gene, found adjacent to a full-length MERVL element and highly enriched within *2C::tomato*+ cells, have shown that it is required for the maintenance of telomeres within ES cultures[1414]. Another important question that remains is whether the selection of these special ES cells can be utilized for practical purposes, such as reprogramming somatic nuclei. This idea is supported by the finding that 2C genes are not properly activated in cloned embryos, and that reprogramming efficiency is enhanced by inhibition of histone deacetylases and Kdm1a, which repress the 2C-state[32,33,34]. Thus overexpression of one or multiple MERVL-linked 2C genes or inhibition of other 2C-gene repressors may be useful strategies to facilitate reprogramming. This possibility is supported by the recent finding that forced Zscan4 expression in fibroblasts enhances their iPS reprogramming efficiency[35].

Transposable elements are a major driving force of evolution. Our findings support the notion that the co-option of retrotransposable elements by cellular genes can serve as an evolutionary mechanism for coordinately linking the expression of many genes[15,29,36]. Transposon sequences have recently been shown to play a critical role in rewiring gene regulatory networks in ES cells and in the endometrium that contributed to the evolution of pregnancy in mammals[37,38]. It has also been speculated that ERVs played a role in the evolution of the placenta by providing fusogenic envelope genes adapted for formation of the syncytiotrophoblasts[39]. We suggest that ERVL elements, that are found in all placental mammals[40], may have played an equally important gene regulatory role in early mammalian development, by contributing to the specification of cell types and leading to formation of placental tissues.

Methods Summary

2C::tomato was created by digesting the MERVL LTR-Gag clone #9[29] with MluI and HindIII, resulting in MERVL LTR 1-730, and was ligated into pcDNA3 hygro tdtomato

with CMV promoter removed. To generate 2C::tomato ES cells, *Kdm1a* Fl/Fl; Cre:ERT ES cells were transfected with *2C::tomato* using Lipofectamine 2000 (Invitrogen) and selected with 150ug/ml hygromycin for 7 days. Colonies containing tomato[+] cells were then picked and expanded. *2C::ERT2-Cre-ERT2* was generated by replacing tdtomato with an ERT2-Cre-ERT2 insert using EcoR1 and Not1 sites. DNA was linearized with Mlu1 and AvrII sites before injection into embryos to generate transgenic mice. The resulting mice were mated with *ROSA::LSL*-tomato mice (JAX 007905), *ROSA::LSL-DTA* mice (JAX 010527), or *ROSA::LSL-LacZ* mice (Gift of Anderson lab), and ES lines were derived using standard procedures. *Kdm1a* GT/GT, *Kdm1a* Fl/Fl, KAP1 ES3 Cre, and G9A TT2 ES cells were described previously[29–31]. RNA-Seq from oocytes and 2C embryos was performed by lysing litters of embryos (5–10 embryos) in Prelude Direct lysis buffer (Nugen) and amplifying RNA using the Ovation RNA-Seq system (Nugen) before library construction using the Tru-Seq RNA sample prep kit (Illumina). Microarray, QRT-PCR, immunostaining, and chimeric mouse injections were performed as described[29]. All animal experiments we performed in accordance with the Salk Institute's IACUC guidelines.

Full Methods

RNA-Seq

For RNA-Seq analysis of early stage embryos, three independent litters of superovulated oocytes or naturally fertilized superovulated oocytes were collected and lysed directly in 2 ul of Prelude Direct Lysis buffer (Nugen). RNA was then subject to amplification using the Ovation RNA-Seq System (Nugen). Amplified cDNA was fragmented using Covaris, and single end (oocytes) or paired end (2C embryos) libraries were then constructed using the mRNA-Seq sample prep kit (Illumina) or Tru-Seq RNA library construction kit (Illumina) starting with end repair. Sequencing was performed on either Illumina Genome Analyzer (oocytes) or Hi-Seq (2C embryos). 72 bp single-end reads (oocytes) or 100 bp paired-end reads (2C embryos) were aligned to the mouse genome using Bowtie, allowing up to three mismatches per alignment and up to 20 alignments per read, filtering out any read aligning in more than 20 locations. To compare the Oocyte data to the 2C data, read lengths were cut down to 72 bp (from the 3'end). The Oocyte data had an average of 33 million alignments per sample while the 2C data had an average of 49 million alignments per sample. Read counts were quantified using a custom gene reference based on UCSC's Knowngene reference. At each gene locus, all isoforms belonging to a single gene were fused into one transcript containing all exons from each isoform. Counts aligning in multiple locations were counted as a fraction of their total number of alignments at each location. Differential expression testing was performed with DESeq[41]. Genes with adjusted p-values less than 0.05 were marked as significant. Chimeric transcripts were identified utilizing the spliced alignment data produced by Tophat. Tophat identifies exons based on alignment pileup and it follows by aligning previously unaligned reads across potential splice junctions. We split the junction information into two lists, left and right side of each junction, and compared to both the UCSC known gene database for mm9 and to the RepeatMasker database, also from UCSC's database. Only junctions that hit an exon of a known model on one end and a repeat element on the other were retained. GO analysis was performed using the David Bioinformatic Resource (http://david.abcc.ncifcrf.gov/)[42].

For RNA-Seq of *2C::tomato*[+] and[−] cells, *Kdm1a* KO ES cells, *Kap1* KO ES cells, and *G9a* KO ES cells, sample libraries were prepared from 500-5ug of total RNA using the mRNA-Seq sample prep kit (Illumina) or Tru-Seq RNA library construction kit. Library samples were amplified on flow cells using cluster generation kit (Illumina) and then sequenced using consecutive 36 cycle sequencing kit on the Genome Analyzer (Illumina) or 100bp paired end reads on the Hi-Seq (Illumina). Raw sequence data was then aligned to the mouse genome using the short read aligner Bowtie and the default setting (2 mismatches per 25 bp

and up to 40 genomic alignments) (http://bowtie-bio.sourceforge.net/index.shtml). RPKM values were also determined by Bowtie. For repetitive sequences, we aligned sequencing reads to the Repbase database using Bowtie (http://www.girinst.org/repbase/index.html). Differential expression was determined using DESeq as described above. To compare gene expression in *G9a, Kdm1a*, and *Kap1* mutant ES cells, we identified upregulated genes by combining previously identified upregulated genes[29–31] and our own DESeq analysis.

ES culture and generation of *2C∷tomato* and *2C∷ERT2-Cre-ERT2* ES lines

The derivation and culture of *Kdm1a* GT/GT, Fl/Fl, and Fl/Fl, Cre-ERT ES cells were described previously[29]. The *2C∷tomato* construct was created by digesting the MERVL LTR-Gag clone #9 in pGL3 basic with MluI and HindIII, resulting in MERVL LTR 1-730, and was ligated into pcDNA3 hygro tdtomato digested with MluI and HindIII (to remove CMV promoter). To generate *2C∷tomato* ES cells, *Kdm1a* Fl/Fl; Cre:ERT2 ES cells were transfected with *2C∷tomato* using Lipofectamine 2000 (Invitrogen) and selected with 150ug/ml hygromycin for 7 days. Colonies containing tomato positive cells were then picked and expanded. 2C∷tomato ES cells were also derived from a transgenic mouse generated by pronuclear injection of the *2C∷tomato* ES line. *2C∷ERT2-Cre-ERT2* was generated by replacing tdtomato with an ERT2-Cre-ERT2 insert using EcoR1 and Not1 sites. DNA was linearized with Mlu1 and AvrII sites before injection into embryos to generate transgenic mice. The resulting mice were mated with *ROSA∷LSL*-tomato mice (JAX 007905), *ROSA∷LSL-DTA* mice (JAX 010527), or *ROSA∷LSL-LacZ* mice (Gift of Anderson lab) and ES lines were derived using standard procedures.

KAP1 ES3Cre and G9A TT2 ES cells were described previously[30,31]. To recombine and delete the *Kdm1a, KAP1*, and *G9A* floxed alleles, cells were treated with 1uM 4OHT for 24 hours. Cells were then harvested at a minimum of 48 hours later to allow for loss of residual protein. To activate the *2C∷ERT2-Cre-ERT2* transgene, cells were maintained in 1uM 4OHT and fed daily. 2C∷tomato positive and negative cells were counted using a FACScan and sorted using a FACSDiVA. For differentiation assays, ES cells were grown in suspension in the absence of Lif as described[29].

Immunofluorescence Staining and Microscopy

ES cells and iPS cells were plated on gelatinized glass coverslips on PMEFs. Cells were fixed with 4% PFA for 10 minutes, followed by washing with PBS-T (0.05% tween). Cells were then blocked in PBS-T containing 3% BSA for ten minutes and stained with primary antibody for 1 hour at room temperature. Antibodies used: mouse anti KAP1 (Abcam), 1:1000; mouse anti OCT3/4, Santa Cruz sc-5279, rabbit anti MERVL-GAG, gift of Heidmann lab, 1:2000; rat anti E-Cadherin, Abcam ab11512, 1:500, rabbit anti Pan Acetylated histone H3, Upstate #06-599, 1:1000; rabbit anti Pan acetyl H4, Upstate #06-598, 1:1000; and rabbit anti H3 DiMeK4, clone AW30, Abcam, 1:1000. After washing 3 times for 10 minutes with PBST, cells were stained with secondary antibody (1:1000 anti mouse, rat, or rabbit IgG Alexa fluor 488, 555, or 647) for one hour at room temperature and washed again 3 times with PBS-T. Coverslips were stained with DAPI in PBS for 5 minutes before inverting onto slides in mounting medium. Cells were then imaged using either an Olympus FV1000 confocal microscope and 60× oil objective, or a Zeiss Axioskop 2 epifluorescence microscope and 40× objective. Quantification of histone stains were performed with Fluoview. Preimplantation embryos were stained as described with minor modifications. Embryos were fixed in 4% PFA for 30 minutes and permeabilized in 0.25% Triton for 20 minutes, prior to blocking in 10% FBS for 1 hour in 0.1% Triton-PBS. Primary antibodies were incubated overnight at 4 degrees C in blocking buffer. Subsequent washes and secondary antibody incubations were at room temperature in 0.1% Triton-PBS.

In situ hybridization

A MERVL probe was generated by PCR from mouse ES cDNA using the forward primer 5' ccatccctgtcattgctca 3' and reverse primer 5' cctttccaccccttgatt 3' and cloned into the PCR2.1 TOPO vector. A DIG labeled probe was prepared using in vitro transcription with the T7 polymerase. ES samples were fixed in 4% PFA, digested for two minutes with proteinase K, washed with PBS, acetylated, and hybridized with denatured probe overnight at 68 degrees. After washing with $5\times$ SSC and $0.2\times$ SSC, DIG labeled probe was visualized using an anti DIG antibody coupled to alkaline phosphatase.

Immunoblotting

Whole cell extracts were prepared by pelleting ES cells at $200\times g$ and resuspending in 1:5 volume of 1% NP40 lysis buffer containing 10mM Tris, 150mM NaCl, and $1\times$ protease inhibitors. To solubilize histones, extracts were also sonicated using a bioruptor on the high setting for 10 minutes. 10–50ug of total protein in LDS sample buffer (Invitrogen) was then loaded onto a 4–12% NuPage gel (Invitrogen), electrophoresed at 200V for 60 minutes, and transferred to nitrocellulose membranes at 30V for 90 minutes. Membranes were blocked in PBS-T containing 5% nonfat dry milk. Primary antibodies were incubated overnight at 4 degrees. Antibodies utilized: rabbit anti GAPDH, Santa Cruz sc25-778, 1:1000, rabbit and MERVL-GAG, gift of Heidmann lab, 1:1000, anti Pan AcH3, Upstate #06-599, 1:1000; anti Pan AcH4, Upstate #06-598, 1:1000; anti H3 DiMeK4 clone AW30, Abcam, 1:500; anti H4, Novus ab10158, 1:1000; anti H3, Novus, NB 500–171, 1:500. After washing extensively with PBS-T, secondary antibodies (anti rabbit or mouse HRP conjugate, 1:10,000 dilution) were incubated for one hour at room temperature. After washing extensively with PBS-T and water, blots were developed using ECL plus detection system (Amersham).

Bisulfite Sequencing

ES cells were lysed in tail lysis buffer (0.1M Tris pH8.5, 5mM EDTA, 0.2% SDS, 0.2M NaCl,) containing proteinase K (Roche) for 1 hour at 55 degrees C, followed by treatment with DNase free RNase for 30 minutes at 37 degrees C. DNA was then sonicated briefly and purified using Qiagen PCR purification columns. Bisulfite conversion of genomic DNA was carried out using the Epitect Bisulfite Kit (Qiagen). Bisulfite converted DNA was then PCR amplified using Accuprime Taq polymerase (Invitrogen) followed by TOPO TA cloning (Invitrogen). At least 10 individual clones per primer pair were sequenced (Eton Bio). Primer sequences were described previously[29].

QRT-PCR

For QRT-PCR Analysis, first strand cDNA was generated from up to 5ug total RNA using Superscript III (Invitrogen) and polydT or random hexamer priming. QPCR was performed using SYBR green master mix (Applied Biosystems) in 96 well dishes in triplicate and repeated with at least two biological replicates with similar results. Standard curves were generated for each primer pair (described previously[29]) and expression levels were plotted relative to *Gapdh* (in arbitrary units).

Microarray

Total RNA was prepared from *2C::tomato*[+] and[−] cells using RNEasy kits (Qiagen). Labeling of 100ng of total RNA was performed using the Whole Transcript (WT) Sense Target Labeling Assay kit (Affymetrix) before hybridization to Genechip Mouse Gene 1.0 ST Arrays. Probeset normalization and summarization were prepared using Robust Multichip Analysis (RMA) in Expression Console (Affymetrix).

Mouse Chimera Assay

ES cells were injected into either E2.5 or E3.5 C57Bl/6J embryos and cultured in vitro or implanted into pseudopregnant females. For PCR assays dissected tissues were placed in lysis buffer (1% SDS, 150mM NaCl, 10mM TriS pH8.0, 1mM EDTA pH 8.0) containing proteinase K overnight at 55 degrees C. DNA was then isolated by phenol chloroform extraction and ethanol precipitation, followed by PCR analysis with primers designed to amplify the *Betageo* cassette or the wild type *Kdm1a* foxed allele. For embryo imaging, chimeric mice were harvested between E9.5 and E12.5 and fixed with 4% PFA for two hours, washed extensively in PBS overnight, incubated in 30% sucrose for 4 hours, and frozen on dry ice in OCT. Cryosections were then taken and stained with DAPI before imaging.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Tarkowski AK. Experiments on the development of isolated blastomers of mouse eggs. Nature. 1959; 184:1286–1287. [PubMed: 13836947]

2. Papaioannou VE, Mkandawire J, Biggers JD. Development and phenotypic variability of genetically identical half mouse embryos. Development. 1989; 106(4):817–827. [PubMed: 2562672]

3. Cockburn K, Rossant J. Making the blastocyst: lessons from the mouse. J Clin Invest. 2010; 120(4): 995–1003. [PubMed: 20364097]

4. Latham KE, Schultz RM. Embryonic genome activation. Front Biosci. 2001; 6:D748–759. [PubMed: 11401780]

5. Schultz RM. The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. Hum Reprod Update. 2002; 8(4):323–331. [PubMed: 12206467]

6. Kanka J. Gene expression and chromatin structure in the pre-implantation embryo. Theriogenology. 2003; 59(1):3–19. [PubMed: 12499014]

7. Evans MJ, Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature. 1981; 292(5819):154–156. [PubMed: 7242681]

8. Martin GR. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A. 1981; 78(12):7634–7638. [PubMed: 6950406]

9. Beddington RS, Robertson EJ. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. Development. 1989; 105(4):733–737. [PubMed: 2598811]

10. Niakan KK, et al. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. Genes Dev. 2010; 24(3):312–326. [PubMed: 20123909]

11. Hayashi K, Lopes SM, Tang F, Surani MA. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell. 2008; 3(4): 391–401. [PubMed: 18940731]

12. Singh AM, Hamazaki T, Hankowski KE, Terada N. A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem Cells. 2007; 25(10):2534–2542. [PubMed: 17615266]

13. Chambers I, et al. Nanog safeguards pluripotency and mediates germline development. Nature. 2007; 450(7173):1230–1234. [PubMed: 18097409]

14. Zalzman M, et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature. 2010; 464(7290):858–863. [PubMed: 20336070]

15. Peaston AE, et al. Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. Dev Cell. 2004; 7(4):597–606. [PubMed: 15469847]

16. Evsikov AV, et al. Systems biology of the 2-cell mouse embryo. Cytogenet Genome Res. 2004; 105(2–4):240–250. [PubMed: 15237213]

17. Kigami D, Minami N, Takayama H, Imai H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. Biol Reprod. 2003; 68(2):651–654. [PubMed: 12533431]

18. Svoboda P, et al. RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. Dev Biol. 2004; 269(1):276–285. [PubMed: 15081373]

19. Ribet D, et al. Murine endogenous retrovirus MuERV-L is the progenitor of the "orphan" epsilon viruslike particles of the early mouse embryo. J Virol. 2008; 82(3):1622–1625. [PubMed: 18045933]

20. Soudais C, et al. Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation in vitro. Development. 1995; 121(11):3877–3888. [PubMed: 8582296]

21. Yagi R, et al. Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. Development. 2007; 134(21):3827–3836. [PubMed: 17913785]

22. Nishioka N, et al. Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mech Dev. 2008; 125(3–4):270–283. [PubMed: 18083014]

23. Choo KB, Chen HH, Cheng WT, Chang HS, Wang M. In silico mining of EST databases for novel pre-implantation embryo-specific zinc finger protein genes. Mol Reprod Dev. 2001; 59(3):249–255. [PubMed: 11424210]

24. Huang CJ, Chen CY, Chen HH, Tsai SF, Choo KB. TDPOZ, a family of bipartite animal and plant proteins that contain the TRAF (TD) and POZ/BTB domains. Gene. 2004; 324:117–127. [PubMed: 14693377]

25. Zhang W, et al. Zfp206 regulates ES cell gene expression and differentiation. Nucleic Acids Res. 2006; 34(17):4780–4790. [PubMed: 16971461]

26. Ying QL, et al. The ground state of embryonic stem cell self-renewal. Nature. 2008; 453(7194): 519–523. [PubMed: 18497825]

27. Ma J, Svoboda P, Schultz RM, Stein P. Regulation of zygotic gene activation in the preimplantation mouse embryo: global activation and repression of gene expression. Biol Reprod. 2001; 64(6):1713–1721. [PubMed: 11369600]

28. Wiekowski M, Miranda M, Nothias JY, DePamphilis ML. Changes in histone synthesis and modification at the beginning of mouse development correlate with the establishment of chromatin mediated repression of transcription. J Cell Sci. 1997; 110(Pt 10):1147–1158. [PubMed: 9191039]

29. Macfarlan TS, et al. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev. 2011

30. Rowe HM, et al. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature. 2010; 463(7278):237–240. [PubMed: 20075919]

31. Yokochi T, et al. G9a selectively represses a class of late-replicating genes at the nuclear periphery. Proc Natl Acad Sci U S A. 2009; 106(46):19363–19368. [PubMed: 19889976]

32. Suzuki T, Minami N, Kono T, Imai H. Zygotically activated genes are suppressed in mouse nuclear transferred embryos. Cloning Stem Cells. 2006; 8(4):295–304. [PubMed: 17196094]

33. Shao GB, et al. Effect of trychostatin A treatment on gene expression in cloned mouse embryos. Theriogenology. 2009; 71(8):1245–1252. [PubMed: 19246084]

34. Li W, et al. Generation of human-induced pluripotent stem cells in the absence of exogenous Sox2. Stem Cells. 2009; 27(12):2992–3000. [PubMed: 19839055]

35. Hirata T, et al. Zscan4 transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells. Sci Rep. 2012; 2:208. [PubMed: 22355722]

36. Feschotte C. Transposable elements and the evolution of regulatory networks. Nat Rev Genet. 2008; 9(5):397–405. [PubMed: 18368054]

37. Kunarso G, et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nat Genet. 2010; 42(7):631–634. [PubMed: 20526341]

38. Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat Genet. 2011; 43(11):1154–1159. [PubMed: 21946353]

39. Dupressoir A, et al. 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. 2009; 106(29):12127–12132. [PubMed: 19564597]

40. Benit L, Lallemand JB, Casella JF, Philippe H, Heidmann T. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. J Virol. 1999; 73(4):3301–3308. [PubMed: 10074184]

41. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11(10):R106. [PubMed: 20979621]

42. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4(1):44–57. [PubMed: 19131956]
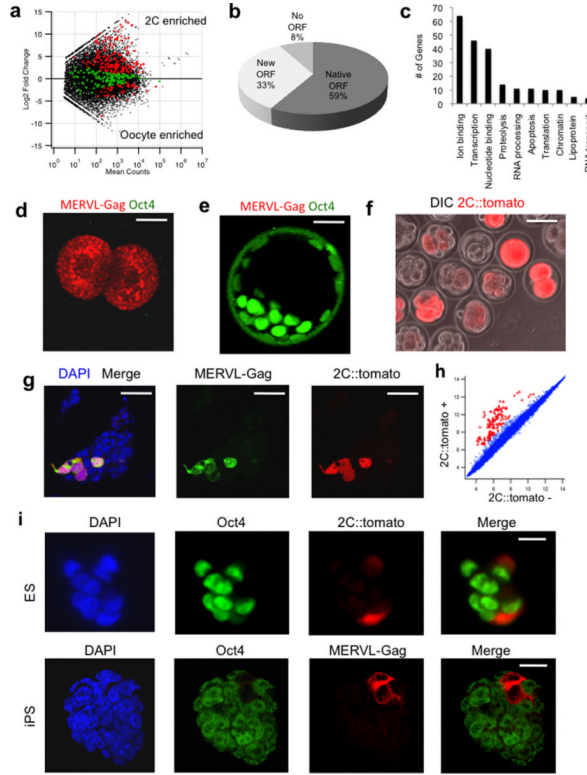
Figure 1. The MERVL retrovirus and a reporter driven by its LTR marks the 2C state
a, Comparison of gene expression between oocytes and 2C embryos. Genes generating junctions to MERVL are shown in red/green, with those in red showing significant change in expression. **b,** ORF status of predicted MERVL-linked chimeric transcripts. **c,** GO analysis of MERVL-linked protein coding transcripts. The number of genes from the 10 most enriched GO categories are shown. **d-e,** 2C (d) and blastocyst embryos (e) were mixed and immunostained with MERVL-Gag and Oct4 antibodies. Scale bar 20 µm. **f,** Zygotes were injected with the *2C::tomato* transgene, and allowed to develop *in vitro* for 48 hours before imaging. Scale bar 50µm. **g,** *2C::tomato* [+] ES cells express MERVL-Gag protein detected by immunofluorescence. Scale bar 50µm. **h,** Microarray analysis of *2C::tomato*[+] vs. [−] cells. Red indicates genes with greater than 4-fold change in expression. **i,** *2C::tomato*[+]/MERVL-Gag[+] ES cells and iPS cells lack Oct4 protein as determined by immuno-fluorescence. Scale bar 20µm
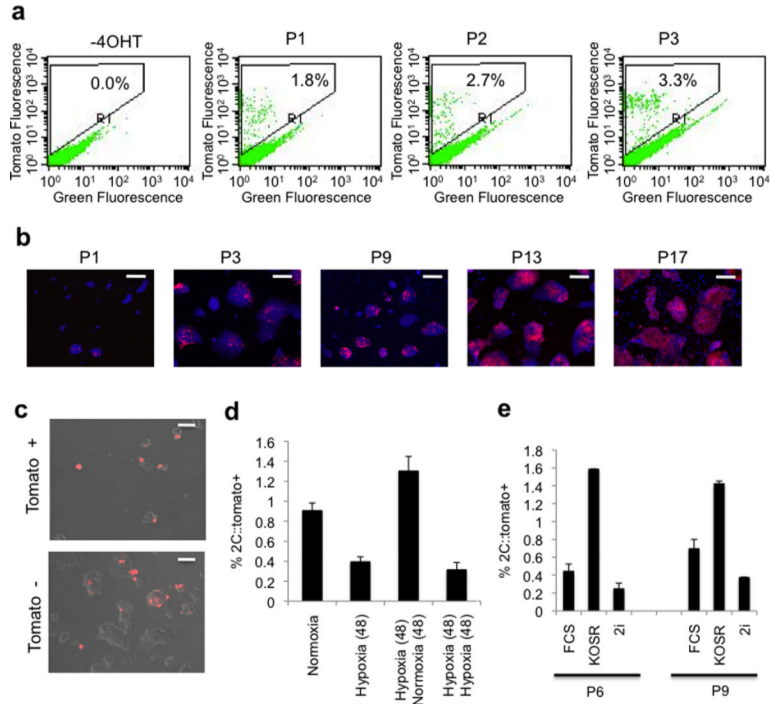
Figure 2. The 2C epigenetic state is transient, is activated within the majority of cells in ES cultures at increasing passages, and is controlled by cell intrinsic and extrinsic factors
a,, FACS analysis of *2C::ERT2-Cre-ERT2, ROSA LSL::Tomato* ES cells at increasing passage in the presence of 4OHT. The percentage of tomato[+] cells is indicated. **b,** *2C::ERT2-Cre-ERT2, LSL::LacZ* ES cells were cultured in the presence of 4OHT, and at increasing passage, cells were fixed and immunostained with anti Beta-galactosidase antibodies and counterstained with DAPI. Scale bar 50μm. **c,** *2C::tomato[+]* and [−] cells were collected by FACS and plated before imaging 48 hours later. Scale bar 50μm. **d,** *2C::tomato* ES cells were cultured in 20% O_2 (normoxia) or 5% O_2 (hypoxia) for the indicated number of hours, and the percentage of tomato[+] cells was determined by FACS. Error bars represent s.d., n=3 **e,** *2C::tomato* ES cells at the indicated passage were cultured in media containing 15% fetal calf serum (FCS), 20% knockout serum replacement (KOSR), or N2B27 media containing 3mM GSK3β and MEK inhibitors (2i) for 48 hours before counting the percentage of *2C::tomato[+]* cells by FACS. Error bars represent s.d., n=3.
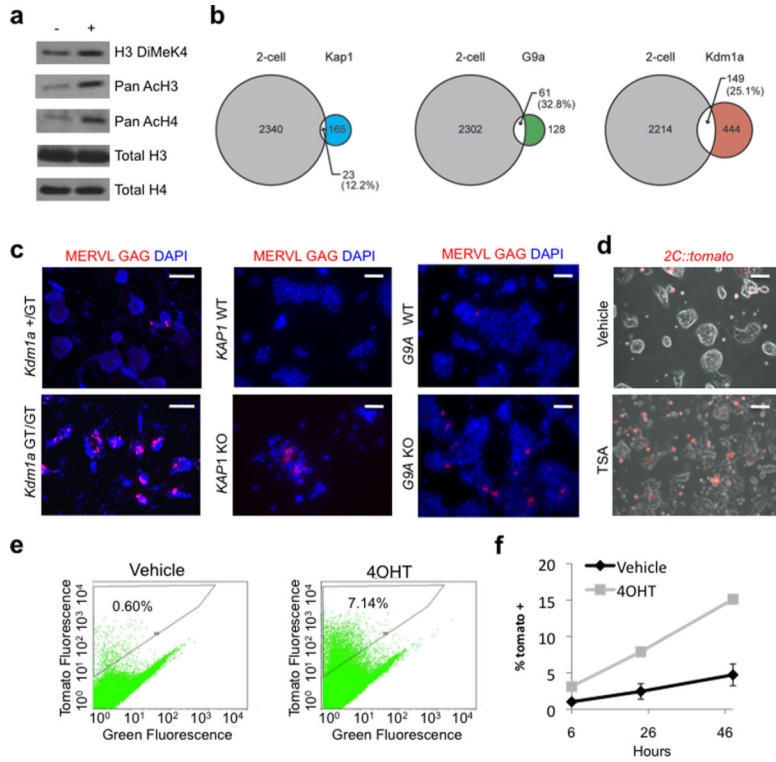
Figure 3. The 2C state is associated with an active epigenetic signature and is antagonized by repressive chromatin modifying enzymes

a, *2C::tomato*[+] and [−] cells were collected by FACS and subjected to immunoblot analysis with indicated antibodies. **b,** Pairwise comparisons of the number of genes activated in *Kap1*, *G9a*, and *Kdm1a* mutant ES cells compared with genes activated in 2C embryos. **c,** ES cell lines homozygous for mutant alleles of *Kdm1a*, *Kap1*, and *G9a*, and corresponding wild type ES lines were immunostained with MERVL Gag antibodies and counterstained with DAPI. Scale bar 50 μm. **d,** *2C::tomato* ES cells were treated with 40nM TSA for 24 hours before imaging. Scale bar 50μm. **e,** *Kdm1a* Fl/Fl, Cre-ERT ES cells containing a stably integrated *2C::tomato* transgene were treated with vehicle or 4OHT and subject to FACS to count the percentage of tomato[+] cells. **f,** *2C::tomato*; *Kdm1a* Fl/Fl, Cre:ERT ES cells were treated with 4OHT or vehicle for 24 hours, then passaged for 72 hours before collecting tomato negative cells by FACS. The percentage of tomato[+] cells was plotted after increasing hours in culture. Error bars represent s.d., n=3.
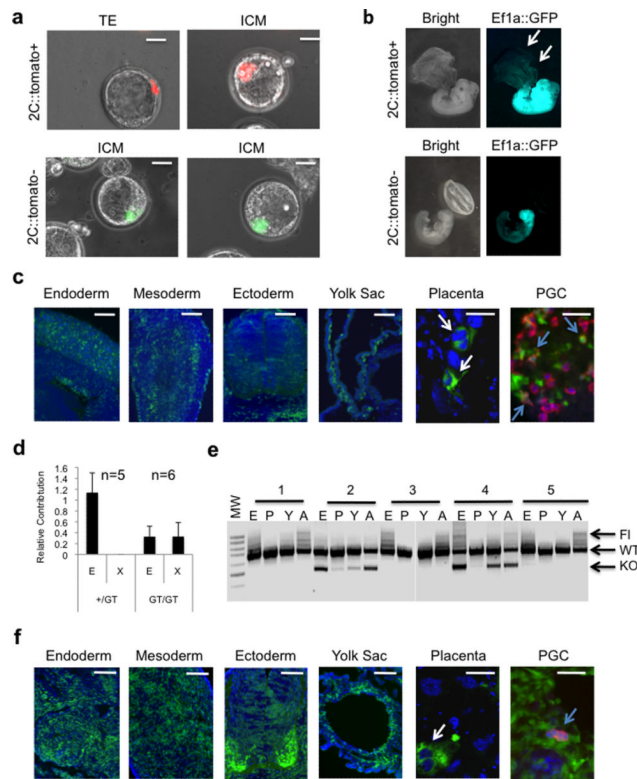
Figure 4. Activation of the 2C state is associated with expanded potency in chimeric embryos towards extraembryonic lineages

a, *2C::tomato$^+$* or *2C::tomato$^-$* ,CMV::GFP ES cells were injected into morula-stage embryos, which were then grown *in vitro*. The resulting blastocysts were imaged to visualize position of injected cells in either the trophectoderm (TE) or ICM. Scale bar 20μm. **b,** *2C::tomato$^+$* or *2C::tomato$^-$* , Ef1a::GFP$^+$ cells were injected into blastocysts which were then implanted into pseudopregnant females to generate chimeric embryos. Arrows indicate 2C::tomato$^+$, GFP$^+$ cells contributing to the yolk sac and placenta. **c,** *2C::tomato$^+$*, Ef1a::GFP$^+$ cells contribute to embryonic endoderm, mesoderm, ectoderm, yolk sac, placental tissues (including trophoblast giant cells-white arrows) and primordial germ cells (PGCs, colabeled with anti-Ddx4 antibody in red, blue arrows). Scale bars represent 500uM (Endoderm, mesoderm, ectoderm, yolk sac) or 50um (placenta and PGCs). **d,** Heterozygous (+/GT) or homozygous (GT/GT)*Kdm1a*-ßgeo genetrap ES cells were injected into wild type blastocysts which were implanted into pseudopregnant females. Embryonic (E) and extraembryonic (X) tissues were separated from chimeric embryos, and subject to semi-quantitative PCR with ßgeo primers to determine the relative contribution of the injected cells to these lineages. Error bars represent s.e.m. **e,** A 1:1 mixture of *Kdm1a* Fl/Fl and KO/KO ES cells were co-injected into wild type blastocysts. At E12.5, chimeric embryonic tissue (E) was separated from placenta (P), yolk sac (Y) and amnion (A) and subject to PCR to detect the FL and KO alleles of the injected cells relative to WT alleles of the resident injected embryo. **F,** *Kdm1a* GT/GT, Ef1a::GFP$^+$ cells contribute to embryonic endoderm, mesoderm, ectoderm, yolk sac, placental tissues (including trophoblast giant cells-white arrows) and primordial germ cells (PGCs, colabeled with anti Ddx4 antibody in red, blue arrows). Scale bars represent 500uM (Endoderm, mesoderm, ectoderm, yolk sac) or 50um (placenta and PGCs).
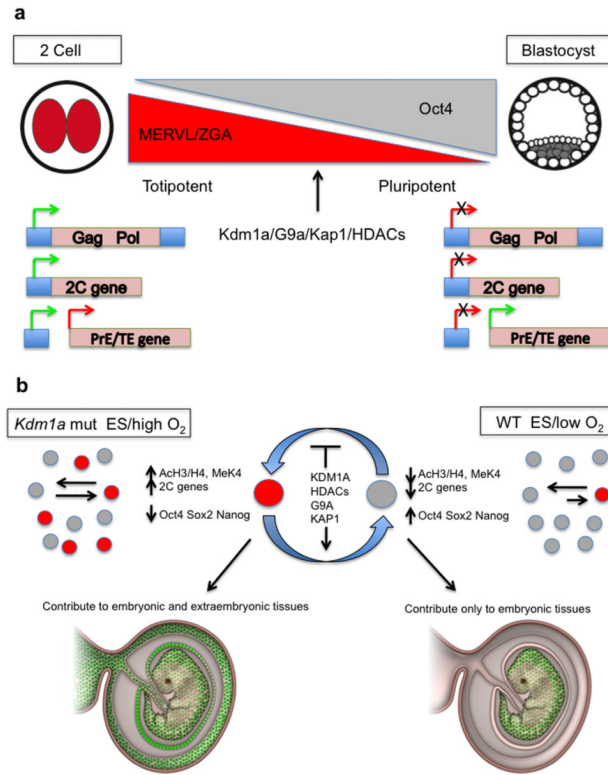
Figure 5. Model of the role of the MERVL-LTR linked 2C gene network in regulating embryonic potency

a, During zygote genome activation, a network of genes that utilize MERVL-LTRs as promoters are activated. This stage correlates with a period where blastomeres are totipotent. As development progresses, the MERVL-LTR linked 2C gene network is silenced by chromatin repressors as the ICM segregates from the TE and PrE. **b,** During the derivation of ES cells from blastocysts, a rare transient population of cells marked by the *2C::tomato* reporter express high levels of 2C genes and low levels of pluripotency markers. In mouse chimera assays, these cells contribute to embryonic and extraembryonic tissues (shown in green). Increasing the oxidative tension of ES cultures or deletion/inhibition of repressive histone modifying enzymes alters the equilibrium between the 2C and ES states.