Embryonic stem cell potency fluctuates with endogenous retrovirus activity

Todd S. Macfarlan1,†, Wesley D. Gifford1, Shawn Driscoll1, Karen Lettieri1, Helen M. Rowe2, Dario Bonanomi1, Amy Firth3, Oded Singer1, Didier Trono2 & Samuel L. Pfaff1

Embryonic stem (ES) cells are derived from blastocyst-stage embryos and are thought to be functionally equivalent to the inner cell mass, which lacks the ability to produce all extraembryonic tissues. Here we identify a rare transient cell population within mouse ES and induced pluripotent stem (iPS) cell cultures that expresses 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 inner cell mass pluripotency proteins Oct4 (also known as Pou5f1), 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 is partially controlled by histone-modifying enzymes. Transcriptome sequencing and bioinformatic analyses showed that many 2C transcripts are initiated from long terminal repeats derived from endogenous retroviruses, suggesting that 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 types1–2. The progeny of these two daughter cells become progressively more fate-restricted as they activate distinct patterns of gene expression that first direct them towards 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+ trophoderm cells that form a large part of the placenta3. 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 markedly 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 genome4–6. The unique transcriptional profile of the zygote and its daughter cells defines a brief period when the cells are totipotent.

Mouse ES cells are isolated from the inner cell mass (ICM) of blastocysts that have already become a separate lineage from the trophoderm7–8. ICM-derived ES cells are regarded as pluripotent because they have the capacity to generate tissues of the fetus, but they are extremely inefficient at colonizing the extraembryonic tissues9. The rare contribution of ES cells to extraembryonic tissues could be explained by contamination of ES cultures with trophoderm or primitive endoderm-committed cells, or may occur because rare ES cells have acquired the ability to produce extraembryonic tissues in addition to embryonic tissues. This latter possibility is intriguing, because recent evidence shows that ES cell cultures are a heterogeneous mixture of metastable cells with fluctuating expression of genes such as Zscan4, stella (also known as Dppa3), Nanog, Sox17 and Gata6, which could account for special attributes of individual cells10–14.

A large number of retrotransposons are expressed when the zygotic genome is first transcribed, including the endogenous retroviruses (ERVs), long interspersed nuclear element-1 (LINE-1), and the non-autonomous short interspersed elements (SINEs)15. At the 2C stage, murine endogenous retrovirus with leucine tRNA primer (MuERV-L, also known as MERV-L and Erv4) elements are transiently derepressed and produce 3% of the transcribed messenger RNAs15–17. After the 2C stage, MuERV-L retroelement expression is silenced18,19. We discovered that this regulated pattern of MuERV-L expression overlapped with more than 100 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 cell cultures contain a small but relatively constant fraction of cells that has entered into the 2C-transcriptional state. Purification of these 2C-like cells shows 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 (RNA-seq) 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 and Supplementary Table 1). The most highly activated repeat was the MuERV-L family of retroviruses and their corresponding long terminal repeat (LTR) promoters (Mt2_mm), which were activated more than 300-fold (Supplementary Table 1). Sequence alignments showed that more than 25% of the nearly 700 copies of MuERV-L elements were activated, and that 307 genes generated chimaeric transcripts with junctions to MuERV-L elements (Fig. 1a and Supplementary Table 2), including 10 that were previously described15. Of the 626 chimaeric transcripts generated, >90% were 5’ LTR–exon fusions that generated open reading frames (ORFs), suggesting that these LTRs had become functional promoters for
Next we asked whether it was possible to use the regulatory sequences from MuERV-L elements to label 2C cells. We cloned the MuERV-L 5′ LTR, the primer-binding site, and a portion of the gag gene upstream of the red fluorescent protein tandem dimeric Tomato (tdTomato). We injected fertilized eggs with the 2C::tdTomato construct and monitored the expression of tdTomato during culture in vitro. tdTomato expression was highest in arrested zygotes and 2C cells but lacked Oct4 protein, as determined by immunofluorescence. Scale bars, 20 µm.

Because more than 300 of the nearly 700 copies of the MuERV-L 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, MuERV-L activity is developmentally regulated and these retroviral promoters have been co-opted by many cellular genes to impose tight control over their expression.

Figure 1 | The MuERV-L 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 MuERV-L are shown in red and green, with those in red denoting significant change in expression. b, ORF status of predicted MuERV-L-linked chimaeric transcript. c, Gene Ontology (GO) analysis of MuERV-L-linked protein-coding transcripts. The number of genes from the ten most enriched GO categories are shown. d, e, 2C (d) and blastocyst (e) embryos were mixed and immunostained with MuERV-L-Gag and Oct4 antibodies. Scale bars, 20 µm. f, Zygotes were injected with the 2C::tdTomato transgene, and allowed to develop in vitro for 48 h before imaging. DIC, differential interference contrast. Scale bar, 50 µm. g, 2C::tdTomato+ ES cells express MuERV-L-Gag protein, as detected by immunofluorescence. DAPI, 4′,6-diamidino-2-phenylindole. Scale bars, 50 µm. h, Microarray analysis of 2C::tdTomato+ and 2C::tdTomato− cells. Red indicates genes with a greater than fourfold change in expression. i, 2C::tdTomato+ MuERV-L-Gag+ ES and iPSCs cells lack Oct4 protein, as determined by immunofluorescence. Scale bars, 20 µm.
correspondence between the 2C::tdTomato reporter and MuERV-L expression was further confirmed by immunoblotting, and electron microscopy imaging of viral episomal particles encoded by MuERV-L within the endoplasmic reticulum of tdTomato+ cells but not tdTomato− cells (Supplementary Fig. 1d, e). Thus, MuERV-L expression is restricted in vivo to 1–4-cell-stage embryos and is reactivated within a small subpopulation of ES cells derived from blastocysts.

To characterize the unexpected 2C::tdTomato-labelled cells within ES cultures, we sorted tdTomato+ and tdTomato− cells and performed microarray and mRNA sequencing analyses (Fig. 1h and Supplementary Tables 3–5). tdTomato+ cells expressed 55-fold higher levels of MuERV-L transcripts than tdTomato− cells, but the vast majority of other retrotransposons were unaffected (Supplementary Table 3). Notably, tdTomato− cells had 165 transcripts activated more than fourfold, and no genes repressed more than fourfold compared with tdTomato− cells (Fig. 1h, Supplementary Table 4 and Supplementary Fig. 2a–f). Among the genes that were highly enriched in tdTomato− cells, several were previously shown to be restricted to the 2–4-cell stage of development, including Zscan4, Tsttl1/3, Efia, Gm3430 (also known as Thoc4), Tdup125 and Zfp352 (refs 23–25). In total, 525 genes that were enriched in 2C::tdTomato− cells were also activated at the 2C stage, including 52 genes that generated chimaeric transcripts linked to MuERV-L elements (Supplementary Tables 6 and 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::tdTomato+ cells lacked Oct4, Sox2 and Nanog (Fig. 1i and Supplementary Fig. 1f). (Fig. 1d, e). We generated a transgenic mouse line using the MuERV-L regulatory elements driving expression of a tamoxifen-inducible Cre recombinase (2C::ERT2-Cre-ERT2; Supplementary Fig. 3a). These mice were then mated with Cre-responsive reporter lines (ROSA::Lsl-tdTomato 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 (4HT) to the ES cultures we detected nuclear Cre expression in MuERV-L–Gag− cells (Supplementary Fig. 3d). When ES cultures were grown for 2–6 days with 4HT 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::tdTomato+ and 2C::tdTomato− cells we performed flow cytometry to collect tdTomato+ and tdTomato− cells. When these purified subpopulations were cultured we found that tdTomato− cells produced tdTomato+ cells and vice versa (Fig. 2c). Within 24 h nearly 50% of the tdTomato+ cells convert to tdTomato− independently of the starting percentages of the two different cell populations (Fig. 2c and data not shown). Under hypoxic conditions (5% O2), the percentage of cells expressing the 2C::tdTomato reporter was decreased, which could be reversed by shifting the cultures back to 20% O2 (Fig. 2d). We also found that growing cells for 48 h in ‘ground-state’ media conditions (2i media) reduced but did not eliminate the presence of tdTomato+ cells relative to media containing knockout serum replacement, suggesting that growth in 2i media might arise from contamination with trophectoderm or primitive endoderm. To exclude this possibility, we examined iPS cells derived from mouse fibroblasts because they should not be contaminated with cells from blastocyst embryos. Similar to ES cells, we found that sporadic iPS cells express the MuERV-L–Gag protein and lack Oct4 (Fig. 1i). Thus, the heterogeneity within ES cultures is a property that is shared with iPSC cell cultures and is unlikely to arise from a cell contaminant.

Next we examined whether the 2C::tdTomato+ cells represent a stable cell population or whether ES cells transition in and out of this 2C-like state. We used 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 mating with Cre-responsive reporter lines (ROSA::Lsl-tdTomato 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 (4HT) to the ES cultures we detected nuclear Cre expression in MuERV-L–Gag− cells (Supplementary Fig. 3d). When ES cultures were grown for 2–6 days with 4HT 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.

ES cells cycle in and out of the 2C state

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

Next we examined whether the 2C::tdTomato+ cells represent a stable cell population or whether ES cells transition in and out of this 2C-like state. We used 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 mating with Cre-responsive reporter lines (ROSA::Lsl-tdTomato 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 (4HT) to the ES cultures we detected nuclear Cre expression in MuERV-L–Gag− cells (Supplementary Fig. 3d). When ES cultures were grown for 2–6 days with 4HT 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::tdTomato+ and 2C::tdTomato− cells we performed flow cytometry to collect tdTomato+ and tdTomato− cells. When these purified subpopulations were cultured we found that tdTomato− cells produced tdTomato+ cells and vice versa (Fig. 2c). Within 24 h nearly 50% of the tdTomato+ cells convert to tdTomato− independently of the starting percentages of the two different cell populations (Fig. 2c and data not shown). Under hypoxic conditions (5% O2), the percentage of cells expressing the 2C::tdTomato reporter was decreased, which could be reversed by shifting the cultures back to 20% O2 (Fig. 2d). We also found that growing cells for 48 h in ‘ground-state’ media conditions (2i media) reduced but did not eliminate the presence of tdTomato+ cells relative to media containing knockout serum replacement, suggesting that growth in 2i media might arise from contamination with trophectoderm or primitive endoderm. To exclude this possibility, we examined iPS cells derived from mouse fibroblasts because they should not be contaminated with cells from blastocyst embryos. Similar to ES cells, we found that sporadic iPS cells express the MuERV-L–Gag protein and lack Oct4 (Fig. 1i). Thus, the heterogeneity within ES cultures is a property that is shared with iPSC cell cultures and is unlikely to arise from a cell contaminant.

Next we examined whether the 2C::tdTomato+ cells represent a stable cell population or whether ES cells transition in and out of this 2C-like state. We used a Cre/loxP fate-mapping strategy to indelibly mark cells that had expressed 2C genes (Supplementary Fig. 3a–c).
extrinsic and intrinsic mechanisms regulate the MuERV-L and 2C gene network (Fig. 2c).

The 2C–ES switch is regulated by histone modification

After activation of the zygotic genome in mouse development, histone deacetylation and histone H1 synthesis lead to the formation of repressive chromatin that is thought to limit the broad pattern of transcription present in 2C embryos. Using indirect immunofluorescence, we found that tdTomato+ cells had markedly higher levels of active histone marks, including methylation of histone 3 lysine 4 demethylase gene were increased in mutant ES cells lacking the histone lysine-specific demethylase gene (H3K4) 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 2C embryos. Next we tested whether tdTomato+ cells had different levels of DNA methylation compared with non-labelled ES cells. We found that the MuERV-L sequences were hypomethylated in tdTomato+ cells compared with tdTomato− cells. In contrast to the MuERV-L sequences, intracisternal A-type particle retroviruses were highly methylated in both tdTomato+ and tdTomato− cells, suggesting the altered pattern of methylation was not uniform across the genome (Supplementary Fig. 4b). In summary, these data suggest that as ES cells (re)enter into the 2C state, their chromatin and DNA is altered to favour transcription in a way that mirrors the 2C embryo.

We previously demonstrated that MuERV-L and 2C-specific genes were increased in mutant ES cells lacking the histone lysine-specific demethylase gene Kdm1a (also known as LSD1). To test whether other proviral co-repressors and histone-modifying enzymes also influence 2C-specific gene expression we profiled the transcriptome of ES cells with homozygous mutations in the KRAB (Kruppel-associated box)-associated transcriptional repressor Kap1 and the H3K9 histone methyltransferase G9a. We found that MuERV-L and several 2C genes were significantly upregulated in Kdm1a, Kap1 and G9a mutant ES cells (Fig. 3b, Supplementary Fig. 5a and Supplementary Table 7). These findings were confirmed using in situ hybridization and immunofluorescence microscopy (Fig. 3c and Supplementary Fig. 5d). Treatment of 2C::tdTomato ES lines with the histone deacetylase inhibitor trichostatin A also increased the number of tdTomato+ cells fourfold (Fig. 3d). To understand better how 2C gene regulation is controlled when chromatin repressors are acutely eliminated we used a stably integrated 2C::tdTomato ES line that is homozygous for a floxed allele of Kdm1a and contains a Cre-ERT transgene that can be activated with 4HT. Within 24 h of deleting Kdm1a we found a tenfold increase in tdTomato+ cells that was steadily maintained (Fig. 3e). In addition, fluorescence-activated cell sorting (FACS)-purified tdTomato+ cells more rapidly became tdTomato− cells in the absence of Kdm1a (Fig. 3f), and stayed in this state longer (Supplementary Movies 2 and 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::tdTomato+ and 2C::tdTomato− states.

2C-like ES cells have expanded fate potential

Because 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::tdTomato+ cells have acquired the ability to produce extraembryonic tissues, a characteristic that ES cells lack. We used FACS to collect tdTomato+ and tdTomato− cells from a 2C::tdTomato ES line, and injected four cells into morula-stage embryos. The tdTomato− cells contributed exclusively to the ICM of all five chimaeric blastocysts analysed (Fig. 4a). By contrast, the tdTomato+ cells contributed to the trophectoderm (in four out of five chimaeric embryos) in addition to the ICM (in three out of five chimaeric embryos) (Fig. 4a). To track the fate of the 2C::tdTomato ES cells later in development, we injected blastocysts with tdTomato− or tdTomato+ cells into blastocysts later in development, we injected blastocysts with tdTomato− or tdTomato+ cells into blastocysts.
subject to semiquantitative PCR with extraembryonic (X) tissues were separated from chimaeric embryos, and that were implanted into pseudopregnant females. (GT)

[two-color figure]

Figure 4 | Activation of the 2C state is associated with expanded potency in chimaeric embryos towards extraembryonic lineages. a, 2C::tdTomato− or 2C::tdTomato+ cytomegalovirus (CMV)–GFP ES cells were injected into morula-stage embryos, which were then grown in vitro. The resulting blastocysts were imaged to visualize the position of injected cells in either the trophectoderm (TE) or ICM. Scale bars, 20 μm. b, 2C::tdTomato− or 2C::tdTomato+ Ef1a::GFP+ cells were injected into blastocysts that were then implanted into pseudopregnant females to generate chimaeric embryos. Arrows indicate 2C::tdTomato+, GFP+ cells contributing to the yolk sac and placenta. Bright denotes bright-field microscopy. c, 2C::tdTomato−, Ef1a::GFP+ cells contribute to embryonic endoderm, mesoderm, ectoderm, yolk sac, placental tissues (including giant trophoblast cells, white arrows) and primordial germ cells (PGCs, colabelled with anti-Ddx4 antibody in red, blue arrows). Scale bars, 500 μM (endoderm, mesoderm and yolk sac) and 50 μm (placenta and PGCs). d, Heterozygous (+/GT) or homozygous (GT/GT) Kdm1a−geo gene-trap ES cells were injected into wild-type blastocysts that were then implanted into pseudopregnant females. (β-geo is a fusion of β-galactosidase and neomycin-resistance genes.) Embryonic (E) and extraembryonic (X) tissues were separated from chimaeric embryos, and subject to semiquantitative PCR with β-geo primers to determine the relative contribution of the injected cells to these lineages. Error bars represent s.e.m. A 1:1 mixture of Kdm1a+/− control and Kdm1a knockout (KO) ES cells were co-injected into wild-type blastocysts. At embryonic day 12.5, chimaeric embryonic (E) tissue was separated from placenta (P), yolk sac (Y) and amnion (A) and subject to PCR to detect the floxed (f) and knockout alleles of the injected cells relative to wild-type alleles of the resident injected embryo. M denotes PCR marker. f, Kdm1a+/−, Ef1a::GFP+ cells contribute to embryonic endoderm, mesoderm, ectoderm, yolk sac, placental tissues (including giant trophoblast cells, white arrow) and primordial germ cells (PGCs, colabelled with anti-Ddx4 antibody in red, blue arrow). Scale bars, 500 μM (endoderm, mesoderm, ectoderm and yolk sac) and 50 μm (placenta and PGCs).

tdTomato− cells that were pre-injected with a lentivirus encoding green fluorescent protein (GFP) from a constitutively active Ef1α promoter (Ef1α::GFP). tdTomato− GFP+ cells contributed exclusively to embryonic tissues, whereas tdTomato+ GFP+ cells contributed to embryonic endoderm, ectoderm, mesoderm, the germ lineage as well as the yolk sac and placenta (Fig. 4b, c and Supplementary Fig. 6a, b). The extraembryonic contribution of the tdTomato+ GFP+ cells included giant trophoblast cells of the placenta (Fig. 4c). Thus, the developmental potential of 2C::tdTomato+ cells includes embryonic plus extraembryonic tissues in contrast to most ES cells in culture, which are restricted to generating only embryonic cell types.

We next examined whether Kdm1a mutant ES lines, which contain higher frequencies of 2C::tdTomato+ cells, also had increased potency in mouse chimaera assays. As expected, Kdm1a heterozygous ES cells contributed exclusively to embryonic tissues (in five out of five chimaeric embryos) but never to extraembryonic tissues (Fig. 4d). By contrast, Kdm1a homozygous mutant ES cells generated both embryonic (in four out of six chimaeric embryos) and extraembryonic (in five out of six chimaeric embryos) tissues (Fig. 4d). To confirm the increased potential of Kdm1a mutant ES cells, we used a competition chimaera assay. We co-injected a 1:1 mixture of control loxP-flanked (floxed) Kdm1a+α or homozygous Kdm1a knockout ES cells into five wild-type blastocysts. PCR was then used to detect the appearance of Kdm1a+α or knockout cells in dissected tissues. We detected Kdm1a+α ES cells in the embryonic tissues and amnion, but not the yolk sac or placenta (Fig. 4e). By contrast, Kdm1a mutant ES cells contributed to extraembryonic tissues, the amnion, yolk sac and placental tissues, including giant trophoblast cells and primordial germ cells (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::tdTomato+ 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::ERT2-Cre-ERT2 mice (Supplementary Fig. 3a) with a Cre-responsive DTA allele (ROSA::LSL-DTA) and treated the ES line with 4HT for 20 passages (Supplementary Fig. 7a). We found that these 2C-depleted ES cultures were still capable of generating high contribution chimaeras (Supplementary Fig. 7b), although their differentiation was biased towards 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 MuERV-L, 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 MuERV-L 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 MuERV-L element and highly enriched within 2C::tdTomato+ cells, have shown that it is required for the maintenance of telomeres within ES cultures14. Another important question that remains is whether the selection of
METHODS SUMMARY

2C::tdTomato was created by digesting the MuERV-L–LTR-Gag clone 9 (ref. 29) with MluI and HindIII, resulting in MuERV-L–LTR 1–730, and was ligated into pcDNA3 hygro tdTomato with the cytomegalovirus (CMV) promoter removed. To generate 2C::tdTomato ES cells, Kdm1a<sup>fl/+</sup>, Cre-ERT ES cells were transfected with 2C::tdTomato using Lipofectamine 2000 (Invitrogen) and selected with 150 μg ml<sup>−1</sup> hygromycin for 7 days. Colonies containing tdTomato<sup>+</sup> cells were then picked and expanded. 2C::ERT2-Cre-ERT2 was generated by replacing tdTomato with an ERT2-Cre-ERT2 insert using EcoR1 and NotI sites. DNA was linearized with MluI and AvrII sites before injection into embryos to generate transgenic mice. The resulting mice were mated with ROSA::Lsl-tdTomato mice (JAX 007905), ROSA::Lsl-DTA mice (JAX 010527) or ROSA::Lsl-LacZ mice (gift from D. Anderson laboratory), and ES lines were derived using standard procedures. Kdm1a<sup>Cre/Ganc</sup>, Kdm1a<sup>fl/+</sup>, Kap1 and G9a mutant ES cells were described previously<sup>32–34</sup>. 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, quantitative PCR with reverse transcription (qRT–PCR), immunostaining and chimeric mouse injections were performed as described<sup>35</sup>. All animal experiments were performed in accordance with the Salk Institute Institutional Animal Care and Use Committee guidelines.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 November 2011; accepted 21 May 2012.
Published online 13 June 2012.

1. Tarkowski, A. K. Experiments on the development of isolated blastomeres of mouse eggs. Nature 184, 1286–1287 (1959).
2. Papaioannou, V. E., Mkiandawire, J. & Biggers, J. D. Development and phenotypic variability of genetically identical half mouse embryos. Development 106, 817–827 (1989).
3. Cockburn, K. & Rossant, J. Making the blastocyst: lessons from the mouse. J. Clin. Invest. 120, 995–1003 (2010).
4. Latham, K. E. & Schultz, R. M. Embryonic genome activation. Front. Biosci. 6, d748–d759 (2001).
5. Schultz, R. M. The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. Theriogenology 59, 3–19 (2003).
6. Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154–156 (1981).
7. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc. Natl Acad. Sci. USA 78, 7634–7638 (1981).
8. Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. Development 105, 733–737 (1989).
9. Niakan, K. K. et al. Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extrarotransposon gene expression and indirectly antagonizing self-renewal. Genes Dev. 24, 312–326 (2010).
10. Hayashi, K., Lopes, S. M., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. Cell Stem Cell 3, 391–401 (2008).
11. Singh, A. M., Hamazaki, T., Hankins, K. E. & Terada, N. A heterogeneous expression pattern for Nanog in embryonic stem cells. Stem Cells 25, 2534–2542 (2007).
12. Chambers, I. et al. Nanog safeguards pluripotency and mediates germline development. Nature 450, 1230–1234 (2007).
13. Zalzman, M. et al. Zscan4 regulates telomere elongation and genomic stability in ES cells. Nature 464, 858–863 (2010).
14. Peaston, A. E. et al. Retrotansposons regulate host genes in mouse oocytes and preimplantation embryos. Dev. Cell 7, 597–606 (2004).
15. Eyskuiok, A. V. et al. Systems biology of the 2-cell mouse embryo. Cytogenet. Genome Res. 105, 240–250 (2004).
16. Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. Biol. Reprod. 68, 651–654 (2003).
17. Svoboda, P. et al. RNAi and expression of retrotansposons MuERV-L and IAP in preimplantation mouse embryos. Dev. Biol. 269, 276–286 (2004).
18. Riet, B. et al. Murine endogenous retrovirus MuERV-L is the progenitor of the “orphan” epsilon virus-like particles of the early mouse embryo. J. Virol. 82, 1622–1625 (2008).
19. 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 121, 3877–3888 (1995).
20. Yagi, R. et al. Transcription factor TEF4 specifies the trophectoderm lineage at the beginning of mammalian development. Development 134, 3827–3836 (2007).
21. Nishihoka, N. et al. Tef4 is required for specification of trophectoderm in pre-implantation mouse embryos. Mech. Dev. 125, 270–283 (2008).
23. Choo, K. B., Chen, H. H., Cheng, W. T., Chang, H. S. & Wang, M. In silico mining of EST databases for novel pre-implantation embryo-specific zinc finger protein genes. Mol. Reprod. Dev. 59, 249–255 (2001).

24. Huang, C. J., Chen, G. Y., Chen, H. H., Tsai, S. F. & Choo, K. B. TDPOZ, a family of bipartite animal and plant proteins that contain the TRAF (TD) and POZ/BTB domains. Gene 324, 117–127 (2004).

25. Zhang, W. et al. Zip206 regulates ES cell gene expression and differentiation. Nucleic Acids Res. 34, 4781–4790 (2006).

26. Yokochi, T. The ground state of embryonic stem cell self-renewal. Nature 453, 519–523 (2008).

27. Ma, J., Svoboda, P., Schultz, R. M. & Stein, P. Regulation of zygotic gene activation in the preimplantation mouse embryo: global activation and repression of gene expression. Biol. Reprod. 64, 1713–1721 (2001).

28. Wieckowski, M., Miranda, M., Nothias, J. Y. & DePamphilis, M. L. 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. 110, 1147–1158 (1997).

29. Macfarlan, T. S. et al. Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. Genes Dev. 25, 594–607 (2011).

30. Rowe, H. M. et al. KAP1 controls endogenous retroviruses in embryonic stem cells. Nature 463, 237–240 (2010).

31. Yokochi, T. et al. G9a selectively represses a class of late-replicating genes at the nuclear periphery. Proc. Natl Acad. Sci. USA 106, 19363–19368 (2009).

32. Suzuki, T., Minami, N., Kono, T. & Imai, H. Zygotically activated genes are suppressed in mouse nuclear transferred embryos. Cloning Stem Cells 8, 295–304 (2006).

33. Shao, G. B. et al. Effect of trichostatin A treatment on gene expression in cloned mouse embryos. Theriogenology 71, 1245–1252 (2009).

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

35. Hirata, T. et al. Zscan4 transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells. Sci. Rep. 2, 208 (2012).

36. Feschotte, C. Transposable elements and the evolution of regulatory networks. Nature Rev. Genet. 9, 397–405 (2008).

37. Kunarso, G. et al. Transposable elements have rewired the core regulatory network of human embryonic stem cells. Nature Genet. 42, 631–634 (2010).

38. Lynch, V. J., Leclerc, R. D., May, G. & Wagner, G. P. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nature Genet. 43, 1154–1159 (2011).

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. USA 106, 12127–12132 (2009).

40. Bénit, L., Lallemand, J. B., Casella, J. F., Philippe, H. & Heidmann, T. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. J. Virol. 73, 3301–3308 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank members of the Pfaff laboratory for discussion: S. Andrews, D. Chambers, Y. Dayn, T.-C. Sung, J. Fitzpatrick, M. Joens, Y. Sigal, D. Gibbs and L. Ouyang for technical assistance, Y. Shinkai and D. Gilbert for the G9a mutant ES cells, and T. Heidmann for MuERV-L-Gag antibodies. This research was supported by the National Institute of Neurological Disorders and Stroke (R37NS037116) and the Marshall Heritage Foundation. T.S.M. and W.D.G. were supported by the California Institute for Regenerative Medicine and S.L.P. is an investigator of the Howard Hughes Medical Institute and Benjamin H. Lewis Chair in Neurobiology.

Author Contributions T.S.M. designed and performed all experiments with assistance from W.D.G., S.D., D.B. and K.L. under the supervision of S.L.P. D.T. generated KAP1 mutant ES cells and H.M.R. and D.T. provided mRNA-seq data from these cells. A.F. and M.R. generated and provided iPS cell lines and lentivirus constructs. T.S.M., W.D.G. and H.M.R. provided materials from W.D.G., S.D., D.B. and K.L. under the supervision of S.L.P. D.T. generated mutant ES cells, and T. Heidmann for MuERV-L-Gag antibodies. This research was supported by the National Institute of Neurological Disorders and Stroke (R37NS037116) and the Marshall Heritage Foundation. T.S.M. and W.D.G. were supported by the National Institute of Neurological Disorders and Stroke (R37NS037116) and the Marshall Heritage Foundation. T.S.M. and W.D.G. were supported by the California Institute for Regenerative Medicine and S.L.P. is an investigator of the Howard Hughes Medical Institute and Benjamin H. Lewis Chair in Neurobiology.

Author Information Microarray and RNA-seq files have been submitted to the NCBI Gene Expression Omnibus database under accession GSE33923. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.L.P. (pfaff@salk.edu).
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 μl of prelude direct lysis buffer (Nugen). RNA was then subjected to amplification using the ovation RNA-seq system (Nugen). Amplified complementary DNA 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 a Illumina genome analyzer (oocytes) or a Hi-Seq (2C embryos). A total of 72-base pair (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 with 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, whereas the 2C data had an average of 49 million alignments per sample. Read counts were quantified using a custom gene reference based on the University of California, Santa Cruz (UCSC) database. 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 several locations were counted as a fraction of their total number of alignments at each location. Differential expression testing was performed with DESeq41. Genes with adjusted P values less than 0.05 were marked as significant. Chimaeric transcripts were identified using 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, the left and right side of each junction, and compared with both the UCSC knowgene database for mm9 and the RepeatMasker database, also from the UCSC 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/). For RNA-seq of 2C::tdTomato+ and 2C::tdTomato− cells, and Kdm1a, Kap1 and G9a knockout ES cells, sample libraries were prepared from 500 ng–5 μg 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 100-bp paired-end reads on the HiSeq (Illumina). Raw sequence data was then aligned to the mouse genome using the short read aligner Bowtie and the default settings (two mismatches per 25 bp and up to 40 genomic alignments) (http://bowtie.bio.sourceforge.net/index.shtml). Reads per kilobase of exon model per million mapped reads (RPKM) values were also determined by Bowtie. For repetitive sequences, we aligned sequencing reads to the Repbase database using Bowtie (http://www.girinst.org/rebase/index.html). Differential expression was determined using DESeq as described earlier. To compare gene expression in G9a, Kdm1a and Kap1 mutant ES cells, we identified upregulated genes by combining previously identified upregulated genes29 and our own DESeq analysis.

ES culture and generation of 2C::tdTomato+ and 2C::ERT2-Cre-ERT2 ES lines. The derivation and culture of Kdm1a+/−29 and Kdm1a−/−29, Cre-ERT2 ES cells were described previously29. The 2C::tdTomato+ construct was created by digesting the MuERV-L-LTR-Gag clone 9 in pGL3 basic with MluI and HindIII (to remove the CMV promoter). To generate 2C::tdTomato ES cells, Kdm1a−/−; Cre-ERT2 ES cells were transfected with 2C::tdTomato using Lipofectamine 2000 (Invitrogen) and selected with 150 μg/ml of prelude direct lysis buffer (Nugen). Cells were then picked and expanded. 2C::tdTomato ES cells were also derived from a transgenic mouse generated by pronuclear injection of the 2C::tdTomato 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 MluI and AvrII sites before infection into embryonic stem (ES) cells, and cultured with MluI and HindIII (to remove the CMV promoter), with MluI and HindIII (to remove the CMV promoter). DNA was linearized with Mlu1 and AvrII sites before infection into embryonic stem (ES) cells, and cultured

Immunofluorescence staining and microscopy. ES and iPS cells were plated on gelatinized glass coverslips on primary mouse embryonic fibroblasts. Cells were fixed with 4% paraformaldehyde (PFA) for 10 min, followed by washing with PBS-T (0.05% tween). Cells were blocked in PBS-T containing 3% BSA for 10 min and stained with primary antibody for 1 h at room temperature. Antibodies used: mouse anti-Kap1 (Abcam, 1:1,000); mouse anti-Oct3/4 (Santa Cruz, 1:500); rabbit anti-MuERV-L-Gag (gift from T. Heidmann laboratory, I:2,000); rat anti-E-cadherin (Abcam, 1:500); rabbit anti-Pan-acetylated histone H3 (Upstate, 1:1,000); rabbit anti-Pan-acetyl H4 (Upstate, 1:1,000); and rabbit anti-H3K4me2 (Abcam, 1:1,000). After washing three times for 10 min with PBS-T, cells were stained with secondary antibody (1:1,000 anti-mouse, -rat or -rabbit IgG Alexa fluor 488, 55 or 647) for 1 h at room temperature and washed again three times with PBS-T. Coverslips were stained with DAPI in PBS for 5 min before inverting onto slides in mounting medium. Cells were then imaged using either an Olympus FX1000 confocal microscope and a ×60 oil objective, or a Zeiss Axioskop 2 epifluorescence microscope and ×40 objective. Quantification of histone stains was performed with Fluoview. Preimplantation embryos were stained as described with minor modifications. Embryos were fixed in 4% PFA for 30 min and permeabilized in 0.25% Triton for 20 min, before blocking in 10% PBS for 1 h in 0.1% Triton-PBS. Primary antibodies were incubated overnight at 4 °C in blocking buffer. Subsequent washes and secondary antibody incubations were at room temperature in 0.1% Triton-PBS.

In situ hybridization. A MuERV-L-LTR-Gag probe was generated by PCR from mouse ES cDNA using the forward primer 5′-CCATCCCTGTCAAGGTC-3′ and reverse primer 5′-CTTTTCCACCCCTGATT-3′, and cloned into the PCR2.1 TOPO vector. A digoxigenin (DIG)-labelled probe was prepared using in vitro transcription with the T7 polymerase. ES cell samples were fixed in 4% PFA, digested for 2 min with proteinase K, washed with PBS, acetylated and hybridized with denatured probe overnight at 68 °C. After washing with 5× saline sodium citrate (SSC) and 0.2× SSC, DIG-labelled probe was visualized using an anti-DIG antibody coupled to alkaline phosphatase.

Immunoblotting. Whole cell extracts were prepared by pelleting ES cells at 200g and resuspending in 1:5 volume of 1% NP40 lysis buffer containing 10 mM Tris, 150 mM NaCl and 1% protease inhibitors. To solubilize histones, extracts were also sonicated using a bioruptor on the high setting for 10 min. Ten to fifty micrograms of total protein in LDS sample buffer (Invitrogen) was then loaded onto a 4–12% NuPage gel (Invitrogen), electrophoresed at 200 V for 60 min, and transferred to nitrocellulose membranes at 30 V for 90 min. Membranes were blocked in PBS-T containing 5% non-fat dry milk. Primary antibodies were incubated overnight at 4 °C. Antibodies used: rabbit anti-GAPDH (Santa Cruz, 1:1,000); rabbit anti-MuERV-L-Gag (gift from T. Heidmann laboratory, I:1,000); anti-Pan-Ach3 (Upstate, 1:1,000); anti-Pan-Ach4 (Upstate, 1:1,000); anti-H3K4me2 (Abcam, 1:500); anti-H4 (Novus, 1:1,000); and anti-H3 (Novus, 1:500). After washing extensively with PBS-T, secondary antibodies (anti-rabbit or mouse horseradish peroxidase (HRP) conjugate, 1:10,000 dilution) were incubated for 1 h at room temperature. After washing extensively with PBS-T and water, blots were developed using ECL plus detection system (Amersham).

Bisulphite sequencing. ES cells were lysed in tail lysis buffer (0.1 M Tris, pH 8.5, 5 mM EDTA, 0.2% SDS and 0.2 M NaCl) containing proteinase K (Roche) for 1 h at 55 °C, followed by treatment with Dnase-free RNase for 30 min at 37 °C. DNA was then sonicated briefly and purified using Qiagen PCR purification columns. Bisulphite conversion of genomic DNA was carried out using the epitect bisulphite kit (Qiagen). Bisulphite-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 Biosciences). PCR sequences were separated previously through qRT–PCR. For quantitative PCR with reverse transcription (qRT–PCR) analysis, first-strand cDNA was generated from up to 5 μg total DNA 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) and expression levels were plotted relative to Gapdh (in arbitrary units).

Microarray. Total RNA was prepared from 2C::tdTomato+ and 2C::tdTomato− cells using RNAeasy kits (Qiagen). Labelling of 100 ng of total RNA was performed using the whole transcript sense target labelling assay kit (Affymetrix) before hybridization to mouse gene 1.0 ST arrays. Probe set normalization and summarization were prepared using robust multichip analysis (RMA) in expression console (Affymetrix).

Mouse chimera assay. ES cells were injected into either embryonic day (E) 2.5 or E3.5 C57Bl/6j embryos, and cultured in vitro or implanted into pseudopregnant

©2012 Macmillan Publishers Limited. All rights reserved
females. For PCR assays, dissected tissues were placed in lysis buffer (1% SDS, 150 mM NaCl, 10 mM Tris, pH 8.0, 1 mM EDTA, pH 8.0) containing proteinase K overnight at 55°C. DNA was then isolated by phenol–chloroform extraction and ethanol precipitation, followed by PCR analysis with primers designed to amplify the β-geo cassette or the wild-type Kdm1a foxed allele. For embryo imaging, chimaeric mice were collected between E9.5 and E12.5 and fixed with 4% PFA for 2 h, washed extensively in PBS overnight, incubated in 30% sucrose for 4 h, and frozen on dry ice in OCT. Cryosections were then taken and stained with DAPI before imaging.

41. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
42. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).