Cite as: Y. J. Choi *et al.*, *Science* 10.1126/science.aag1927 (2017).

Deficiency of microRNA *miR-34a* expands cell fate potential in pluripotent stem cells

Yong Jin Choi,[1]* Chao-Po Lin,[1]*† Davide Risso,[2]* Sean Chen,[1] Thomas Aquinas Kim,[1] Meng How Tan,[3] Jin B. Li,[3] Yalei Wu,[4] Caifu Chen,[5] Zhenyu Xuan,[6] Todd Macfarlan,[7] Weiqun Peng,[8] K. C. Kent Lloyd,[9] Sang Yong Kim,[10] Terence P. Speed,[11,12,13] Lin He[1]†

[1]Division of Cellular and Developmental Biology, Department of Molecular and Cell Biology, University of California, Berkeley, CA 94705, USA. [2]Division of Biostatistics, School of Public Health, University of California, Berkeley, CA 94720, USA. [3]Department of Genetics, Stanford University, Stanford, CA 94305, USA. [4]Thermo Fisher Scientific, 180 Oyster Point Boulevard, South San Francisco, CA 94080, USA. [5]Integrated DNA Technologies, 200 Chesapeake Drive, Redwood City, CA 94063, USA. [6]Department of Molecular and Cell Biology, University of Texas at Dallas, Richardson, TX 75080, USA. [7]Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD 20892, USA. [8]Department of Physics, George Washington University, Washington, DC 20052, USA. [9]Mouse Biology Program, University of California, Davis, CA 95616, USA. [10]Department of Pathology, New York University School of Medicine, New York, NY 10016, USA. [11]Department of Statistics, University of California, Berkeley, CA 94720, USA. [12]Department of Mathematics and Statistics, University of Melbourne, Parkville, VIC 3010, Australia. [13]Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville, VIC 3052, Australia.

*These authors contributed equally to this work.

†Corresponding author. Email: lhe@berkeley.edu (L.H.); newcplin@gmail.com (C.-P.L.)

Embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs) efficiently generate all embryonic cell lineages, but rarely generate extra-embryonic cell types. We show that microRNA *miR-34a* deficiency expands the developmental potential of mouse pluripotent stem cells to yield both embryonic and extra-embryonic lineages and strongly induce MuERV-L (MERVL) endogenous retroviruses, similar to what is seen with totipotent 2-cell blastomeres. *miR-34a* restricts the acquisition of expanded cell fate potential in pluripotent stem cells, and represses MERVL expression through transcriptional regulation, at least in part, by targeting the transcription factor Gata2. Altogether, our studies reveal a complex molecular network that defines and restricts pluripotent developmental potential, raising the tantalizing possibility of culturing bi-potential ESCs to explore the molecular basis for totipotency.

Mouse embryonic stem cells (ESCs), derived from the inner cell mass (ICM) of blastocysts, and induced pluripotent stem cells (iPSCs), generated by somatic reprogramming, are classically defined as pluripotent stem cells (*1*–*3*). As a population, ESCs and iPSCs contribute to all embryonic cell types in vitro and in vivo, but rarely to extra-embryonic cell lineages in the placenta and yolk sac (*4*). This restricted pluripotent potential contrasts with that of totipotent mouse zygotes and 2-cell stage (2C) blastomeres, which give rise to both embryonic and extra-embryonic cell lineages during normal development (*5*, *6*). Rare pluripotent stem cell populations with expanded cell fate potential have been identified in culture as a result of genetic alterations, derivation methods, culture conditions, or enrichment with specific molecular markers (*7*–*10*). Such pluripotent stem cells exhibit bi-directional cell fate potential, contributing to both embryonic and extra-embryonic lineages in vitro and in vivo (*7*–*10*). A key molecular signature shared by ESCs/iPSCs with bi-directional potential is the strong induction of the MuERV-L (MERVL) family of murine endogenous retroviruses (ERVs), which only occurs in totipotent 2C blastomeres during normal mouse preimplantation development (*7*, *9*). These studies suggest that rare, cultured ESCs/iPSCs could acquire features of cell fate plasticity reminiscent of early blastomeres; however, little is known about the molecular barriers restricting ESC/iPSC developmental potential to a pluripotent cell state. In this study, we identified the *miR-34a* microRNA (miRNA) as a non-coding regulator that restricts pluripotent cell fate potential in cultured ESCs/iPSCs, showing that deficiency of this miRNA in pluripotent stem cells yields bi-directional cell fate potential and MERVL induction.

Expanded cell fate potential of *miR-34a*$^{-/-}$ pluripotent stem cells

miRNAs are a class of small, regulatory non-coding RNAs that regulate gene expression post-transcriptionally through a combined mechanism of mRNA degradation and translational repression (*11*, *12*). These small non-coding RNAs have been implicated in numerous cellular processes

in development and disease, and they have become increasingly recognized as key regulators of cell fate specification in diverse developmental systems, including pluripotent stem cells (*13–16*).

Initially identified as *bona fide* p53 transcriptional targets in tumor suppression (*17–22*), the *miR-34* miRNAs (*miR-34a*, *miR-34b*, and *miR-34c*) were previously characterized as key barriers to somatic reprogramming (*17*). *miR-34a* deficiency significantly enhances the efficiency of iPSC generation, producing iPSCs with normal self-renewal and pluripotency (fig. S1, A to C) (*17*). However, teratomas generated from *miR-34a*$^{-/-}$ iPSCs, but not those from wild-type iPSCs, contained cellular features reminiscent of trophoblast giant cells in the placenta, characterized by PL-1 (placental lactogen 1) expression, large cell volume, enlarged nuclei, and close proximity to internal hemorrhages (Fig. 1A). In ESCs, *miR-34a* accounts for the majority of *miR-34* family miRNA expression (fig. S1D). Similar to *miR-34a*$^{-/-}$ iPSCs, *miR-34a*$^{-/-}$ ESCs also generated teratomas that contained not only derivatives of the three germ layers (fig. S1B), but also features of extra-embryonic placental lineages (Fig. 1A). Consistently, *miR-34a*$^{-/-}$ teratomas exhibited induction of multiple trophectoderm (TE) markers (Fig. 1B and fig. S1E), including *cdx2*, *elf5*, *psx1*, *fgfr2*, *egfr* and *mdfi* (*23*, *24*). Although we did not identify areas that morphologically resembled the visceral endoderm of the yolk sac in *miR-34a*$^{-/-}$ teratomas, we detected a strong induction of primitive endoderm (PE) markers (fig. S1E), including *gata4*, *gata6* and *sox17* (*25*). These findings suggest that *miR-34a*$^{-/-}$ pluripotent stem cells likely differentiate toward both embryonic and extra-embryonic cell lineages during teratoma formation.

The expanded potential of *miR-34a*$^{-/-}$ ESCs/iPSCs was also evident upon embryoid body (EB) differentiation (Fig. 1, C and D, fig. S1, F, G and H). Although markers of all three germ layers were similarly induced in wild-type and *miR-34a*$^{-/-}$ EBs (Fig. 1D and fig. S1F), TE markers, extra-embryonic endoderm markers and trophoblast cell markers (*23*, *26–28*) were strongly induced only in *miR-34a*$^{-/-}$ EBs (Fig. 1, C and D, and fig. S1, F and G). Using immunofluorescence (IF) staining, we confirmed that a significant percentage of *miR-34a*$^{-/-}$ EBs were Cdx2 positive, and that these Cdx2-positive cells preferentially localized to the EB periphery (Fig. 1C and fig. S1H). Thus, upon EB differentiation, *miR-34a*$^{-/-}$ pluripotent stem cells exhibited expanded cell fate potential, generating cells with molecular features of both embryonic and extra-embryonic lineages.

To define the cell fate potential of *miR-34a*$^{-/-}$ pluripotent stem cells in normal development, we traced the cell fate of their progenies in chimeric blastocysts. Initially, four GFP-labeled wild-type or *miR-34a*$^{-/-}$ mouse ESCs were microinjected into each C57BL/6N recipient mouse morula to generate chimeric blastocysts (Fig. 2A). Although wild-type ESCs exclusively gave rise to cells localized to the ICM (Fig. 2A and table S1), *miR-34a*$^{-/-}$ ESC progenies localized to both ICM and TE in ~60% of chimeric blastocysts (Fig. 2A and table S1). This expanded cell fate potential is unlikely to be due to extra-embryonic contamination during *miR-34a*$^{-/-}$ ESC derivation since *miR-34a*$^{-/-}$ iPSCs derived from mouse embryonic fibroblasts (MEFs) phenocopied *miR-34a*$^{-/-}$ ESCs in this expanded developmental potential (fig. S1I). When aggregated with recipient C57BL/6J morulae, *miR-34a*$^{-/-}$ ESC and iPSC progenies colonized both ICM and TE in ~25% of chimeric blastocysts, whereas passage- and littermate-controlled wild-type ESCs and iPSCs exclusively contributed to the ICM (Fig. 2A and fig. S1I).

Given the expanded cell fate potential in both *miR-34a*$^{-/-}$ ESCs and *miR-34a*$^{-/-}$ iPSCs, we compared the miRNA profiles between *miR-34a*$^{-/-}$ pluripotent stem cells and their wild-type controls. We demonstrated that *miR-34a* was the only miRNA strongly depleted in both *miR-34a*$^{-/-}$ ESCs and iPSCs, suggesting that this phenotype is mediated solely by *miR-34a* deficiency (fig. S1J).

The expanded cell fate potential of *miR-34a*$^{-/-}$ ESCs could result from cells with bi-directional potential of individual cells; alternatively, the expanded fate potential could result from a heterogeneous population containing cells that preferentially differentiate into embryonic or extra-embryonic lineages. To distinguish between these two possibilities, we microinjected single, GFP-labeled *miR-34a*$^{-/-}$ ESCs into each C57BL/6N recipient morula to generate chimeric blastocysts (Fig. 2B). In the two independent *miR-34a*$^{-/-}$ ESC lines tested, single *miR-34a*$^{-/-}$ ESCs colonized both ICM and TE in 33% and 38% of chimeric blastocysts (n = 13/40 and 8/21) (Fig. 2B and table S1), respectively, suggesting that a significant portion of *miR-34a*$^{-/-}$ ESCs have bi-directional developmental potential at the single-cell level.

We then generated post-gastrulation chimeric embryos by microinjecting 10-15 GFP-labeled wild-type or *miR-34a*$^{-/-}$ ESCs into each C57BL/6N recipient blastocyst. Wild-type ESCs contributed exclusively to the three embryonic germ layers (ectoderm, mesoderm, and endoderm); in comparison, *miR-34a*$^{-/-}$ ESCs contributed to both embryonic and extra-embryonic cell lineages in E9.5, E12.5, and E14.5 chimeric embryos (Fig. 2, C and D, fig. S1K, and table S1). Although *miR-34a*$^{-/-}$ ESCs were more efficient in generating embryonic lineages than extra-embryonic lineages, we observed GFP-positive, *miR-34a*$^{-/-}$ ESC progenies in the visceral endoderm of the yolk sac, and to a lesser degree, in extra-embryonic trophoblast lineages of the placenta (trophoblast giant cells, spongiotrophoblasts, syncytiotrophoblasts (STBs) or sinusoidal trophoblast giant cells (s-TGCs) (Fig. 2, C and D, and table S1). In some chimeric embryos,

GFP-positive, *miR-34a*$^{-/-}$ ESC progenies were observed in both yolk sac visceral endoderm and placental extra-embryonic lineages. (table S1). Notably, the number of GFP-positive cells in extra-embryonic cell lineages greatly exceeds the number of *miR-34a*$^{-/-}$ ESCs injected (Fig. 2, C and D), suggesting that injected *miR-34a*$^{-/-}$ ESCs had undergone substantial proliferation before committing to terminally differentiated extra-embryonic lineages.

MERVL induction in *miR-34a*$^{-/-}$ ESCs/iPSCs

To investigate the molecular basis for the bi-directional potential of *miR-34a*$^{-/-}$ pluripotent stem cells, we compared the transcriptomes of wild-type and *miR-34a*$^{-/-}$ iPSCs using RNA-sequencing (RNA-seq). We compared the abundance of all annotated transcripts between wild-type and *miR-34a*$^{-/-}$ iPSCs, including protein-coding genes, long non-coding RNAs (lncRNAs), pseudogenes, antisense transcripts, and retrotransposons using 100-bp paired-end RNA-seq data (Fig. 3A). Given the repetitive nature of retrotransposons, we quantified retrotransposon expression at the family level, using both uniquely and non-uniquely mapped reads: we assigned each multi-mapped read to the best possible alignment and subsequently summed the reads from all the retrotransposon loci to obtain family-level expression estimates (see supplementary materials). Some of the most highly expressed and differentially regulated transcripts in *miR-34a*$^{-/-}$ iPSCs were transcribed from the MERVL family of endogenous retroviruses (ERVs) (Fig. 3A and fig. S2A), which is also highly induced in totipotent 2C blastomeres and previously characterized bi-potential ESCs (*7*, *9*, *29*, *30*). In fact, transcripts from MERVL ERVs constitute nearly 3% of the transcriptome in 2C blastomeres (*29–31*). Retrotransposon induction in *miR-34a*$^{-/-}$ ESCs/iPSCs was largely specific to the MERVL family, as this induction was not observed in other families, such as IAP, LINEs and SINEs (Fig. 3, A and B, fig. S2A, and table S2). The majority of differentially expressed retrotransposons in *miR-34a*$^{-/-}$ iPSCs belongs to the canonical MERVL family of ERVs (a class-III ERV) (fig. S2A and table S2); a small fraction of differentially expressed loci belongs to the MT2A, MT2B, MT2B1, or MT2B2 ERV families that are highly related to the canonical MERVL solo LTR, MT2_Mm (fig. S2A and table S3).

Consistent with our RNA-seq results, we confirmed the increased MERVL expression in *miR-34a*$^{-/-}$ iPSCs and ESCs using real-time PCR primer pairs designed from multiple highly conserved MERVL regions (Fig. 3, B and C, and table S6). Whereas MERVL induction in *miR-34a*$^{-/-}$ iPSCs persisted for more than 27 passages (fig. S2B), MERVL was only induced in early-passage *miR-34a*$^{-/-}$ ESCs, and became completely silenced around passage 12 (fig. S2B). It is conceivable that MERVL expression in *miR-34a*$^{-/-}$ ESCs triggers additional mechanisms to re-establish their silencing. A compensatory induction of *miR-34a* homologs (*miR-34b/34c* and *miR-449*) occurred in late-passage *miR-34a*$^{-/-}$ ESCs (fig. S2B), which could help to re-establish MERVL silencing. The expanded cell fate potential of *miR-34a*$^{-/-}$ ESCs correlated with MERVL induction, as late-passage *miR-34a*$^{-/-}$ ESCs lost both MERVL induction and the bi-directional cell fate potential (fig. S2, B and C).

The MERVL ERVs have been retained throughout mammalian evolution, with independent expansion in the murine and primate genomes (*32*). There are 2,502 MERVL loci in the C57B6/J mouse genome (*32*), 26% of which encode elements with an intact retroviral structure, comprising *5′*- and *3′*-LTRs flanking the coding sequences for *gag*, *pol*, and *dUTPase*, but lacking *env*-like open reading frames (ORFs) (Fig. 3C). Another 32% MERVL loci exhibit truncated retroviral structure, missing one or both LTRs (Fig. 3C). The remaining 41% MERVL loci have undergone homologous recombination, yielding solo LTRs (MT2_Mm) with varying degrees of sequence degeneration (Fig. 3C). We obtained bioinformatic estimates of locus-specific MERVL expression in wild-type and *miR-34a*$^{-/-}$ iPSCs using our RNA-seq data (table S3 and supplementary materials). Notably, definitive evidence for MERVL induction in *miR-34a*$^{-/-}$ iPSCs was observed predominantly for loci harboring MERVLs with a complete retroviral structure, but not for those with truncated structure (Fig. 3D, fig. S2D, and table S3). A fraction of MT2_Mm solo LTRs, along with a few elements from the highly related MERVL solo LTRs (MT2B, MT2B1, MT2B2 and MT2A), also exhibited a similar induction (fig. S2A and table S3). These findings suggest that an intact LTR, which likely contains the retrotransposon promoter sequences, is likely required for the induction of a MERVL locus.

Approximately 300 MERVL loci still encode intact Gag viral protein (*7*). We observed a significant increase in MERVL-Gag expression and in the percentage of MERVL-Gag-positive cells in *miR-34a*$^{-/-}$ pluripotent stem cells (Fig. 3, E and F). Intriguingly, *miR-34a*$^{-/-}$ ESCs and iPSCs were both heterogeneous populations, containing ~12% and ~20% MERVL-Gag-positive cells, respectively, in otherwise Oct4-positive colonies (Fig. 3, E and F, and fig. S2, E and F). A fraction of individual *miR-34a*$^{-/-}$ iPSC colonies exhibited a significantly greater MERVL induction than the bulk population (fig. S2G), suggesting that the extent of MERVL induction in individual cells or colonies was largely underestimated by data from the bulk population. Notably, the expression of MERVL-Gag and Oct4 were mutually exclusive in *miR-34a*$^{-/-}$ ESCs and iPSCs (Fig. 3E and fig. S2, E and F). Hence, the MERVL-positive, Oct4-negative *miR-34a*$^{-/-}$ cells could possess a unique transcriptional state that correlates with a developmental potency distinct from that of classic, Oct4-positive pluripotent stem cells (*7*, *9*, *33*).

The global gene expression profiles of *miR-34a*$^{-/-}$ iPSCs

resemble those of previously reported bi-potential ESCs (designated as 2C-like ESCs in some studies), which are distinct from those of pluripotent ESCs (fig. S3, A and B). Hierarchical clustering and principal component analysis suggest that these bi-potential cells are closer in transcriptional space to 2C blastomeres, compared to pluripotent ESCs (fig. S3, A and B, and supplementary materials). Intriguingly, among the most differentially expressed genes between *miR-34a*$^{-/-}$ and wild-type iPSCs were those proximal to MERVL loci (Fig. 4A). Indeed, differential expression analysis between previously reported bi-potential ESCs (*lsd1*$^{-/-}$ (*34*), MERVL-expressing wild-type ESCs (*7*), and *p60* knockdown and *p150* knockdown ESCs (*9*)) and their pluripotent ESC controls revealed the induction of MERVL-proximal genes as a key signature (fig. S3C). The MERVL derepression in *miR-34a*$^{-/-}$ iPSCs and ESCs also correlates with the induction of many genes that harbor either a proximal upstream MERVL or an intronic MERVL on the same strand (Fig. 4A and fig. S3D). In some cases, MERVL or related loci, particularly solo LTRs, act as alternative promoters, generating chimeric transcripts of proximal genes that differ in 5′-UTRs and/or ORFs (Fig. 4B, fig. S3, E and F, and tables S4 and S5). These MERVL-gene chimeric transcripts can be unambiguously identified by their corresponding splice junctions from the RNA-seq data (table S5). Using real-time PCR analyses, we validated the induction of multiple MERVL-proximal genes in *miR-34a*$^{-/-}$ ESCs and iPSCs, including *tcstv1*, *tcstv3*, *zfp352*, *cml2* and *p4ha2* (harboring a proximal upstream MERVL element), as well as *abcb5*, *tmem132c* and *chit1* (harboring an intronic MERVL element) (Fig. 4B and fig. S3, E and F). The induction level of MERVL-gene isoforms varied among individual *miR-34a*$^{-/-}$ iPSC colonies, but appeared to correlate with the extent of MERVL induction (fig. S3G). Consistent with these observations, *miR-34a* overexpression in *miR-34a*$^{-/-}$ iPSCs not only decreased MERVL expression, but also reduced the level of the MERVL-driven chimeric transcript, such as *zfp352* (Fig. 4C).

Given the technical difficulty of mapping repetitive RNA-seq reads, we repeated our analysis using 150-bp pair-ended RNA-seq data to gain confidence in the accuracy of our retrotransposon mapping. Our RNA-seq analysis using longer sequence reads confirmed the induction of MERVL ERVs and MERVL proximal genes in *miR-34a*$^{-/-}$ iPSCs (fig. S4, A and B).

Transcriptional activation of MERVL in *miR-34a*$^{-/-}$ ESCs/iPSCs

Given the correlation between MERVL induction and bi-potential pluripotent stem cells (*7*, *9*, *34*) (Fig. 3, A and B), the molecular pathway that mediates *miR-34a*-dependent MERVL repression could also regulate *miR-34a*-dependent restriction of pluripotent potential. To determine the key MERVL sequences required for its induction in *miR-34a*$^{-/-}$ pluripotent stem cells, we transfected wild-type and *miR-34a*$^{-/-}$ ESCs with a MERVL$_{1\text{-}1000}$-Luc (luciferase) reporter containing the full-length LTR (MT2_Mm) and a portion of the *gag* sequence as the promoter (*34*). This reporter showed elevated luciferase activity in *miR-34a*$^{-/-}$ ESCs compared to wild-type ESCs, faithfully recapitulating the endogenous MERVL induction (Fig. 5A). The LTR sequence was both necessary and sufficient for the MERVL$_{1\text{-}1000}$-Luc reporter activity in *miR-34a*$^{-/-}$ ESCs (Fig. 5A); furthermore, a minimal 250-bp fragment, MERVL$_{125\text{-}375}$, containing a predicted TATA box, was sufficient to drive strong luciferase activity specifically in *miR-34a*$^{-/-}$ ESCs (Fig. 5A). The MERVL$_{125\text{-}375}$ fragment is conserved among the most highly induced MERVL loci in *miR-34a*$^{-/-}$ iPSCs (fig. S5A), suggesting a sequence-dependent transcriptional mechanism for MERVL induction. Consistent with this inference, *miR-34a*$^{-/-}$ ESCs also exhibited H3K4Me3 enrichment near the LTR of MERVL ERVs, as well as the MERVL LTR proximal to *tcstv1*, but not near other ERVs such as IAP or MMERK10C (Fig. 5B). Thus, MERVL loci are specifically enriched for active transcription machinery in *miR-34a*$^{-/-}$ pluripotent stem cells.

Gata2 mediates MERVL induction in *miR-34a*$^{-/-}$ ESCs/iPSCs

The MERVL$_{125\text{-}375}$ fragment likely contains *cis*-regulatory elements necessary and sufficient to enable MERVL induction in *miR-34a*$^{-/-}$ ESCs/iPSCs. We failed to detect any significant sequence complementarity between *miR-34a* (pri-, pre-, or mature miRNA sequences) and the MERVL$_{125\text{-}375}$ sequence, thus precluding a direct, RNA base-pairing-dependent repression mechanism. We bioinformatically predicted 70 candidate transcription factors that could potentially bind within MERVL$_{125\text{-}375}$, among which, only GATA-binding protein 2 (Gata2) exhibits an expression pattern similar to that of MERVL during early pre-implantation development, as confirmed by both our real time PCR analyses and previously reported RNA-seq data (fig. S5B, based on data from previous studies (*35*, *36*)).

Aligning the LTR sequences from 18 MERVL loci that were strongly induced in *miR-34a*$^{-/-}$ iPSCs, we found that all of the loci harbored at least two putative gata2 binding sites (BS), with most loci harboring three binding sites (fig. S5A). Mutating the two best-conserved, putative Gata2 binding sites (BS1 and BS3) within the MERVL$_{125\text{-}375}$-Luc reporter significantly reduced its activity in *miR-34a*$^{-/-}$ pluripotent stem cells (Fig. 5C). Similarly, *gata2* knockdown in *miR-34a*$^{-/-}$ pluripotent stem cells effectively abolished the induction of MERVL and MERVL-proximal genes (*zfp352*, *tmem132c* and *chit1*) (Fig. 5D and fig. S5C), and significantly

decreased MERVL and *cdx2* induction in the subsequent teratoma analyses (fig. S5D). *gata2* knockdown in *miR-34a*$^{-/-}$ iPSCs also reduced H3K4Me3 deposition on MERVL and the MERVL-proximal gene *tcstv1*, suggesting a decrease in active transcriptional machinery on MERVL loci (Fig. 5E). Finally, using chromatin immunoprecipitation (ChIP), we demonstrated that specific Gata2 binding to the MERVL LTR, but not to the MERVL internal regions or other ERVs (such as IAP or MMERK10C), in Gata2-overexpressing *miR-34a*$^{-/-}$ pluripotent stem cells (Fig. 5F). Additionally, Gata2 binding at the MERVL LTR was greater in *miR-34a*$^{-/-}$ pluripotent stem cells than that in wild-type controls (fig. S5E). Altogether, our data suggest that Gata2 plays an essential role in directly promoting the induction of MERVL ERVs and MERVL-proximal genes in *miR-34a*$^{-/-}$ pluripotent stem cells.

Previous studies also reported the importance of epigenetic regulation in MERVL induction (*34*, *37*, *38*). However, no obvious alterations were detected between wild-type and *miR-34a*$^{-/-}$ iPSCs in the extent of DNA methylation (fig. S5, F and G), or the deposition of H3K27Ac and H3K9Me2 modifications (fig. S5, H and I), on MERVL loci. Although *miR-34a*$^{-/-}$ ESCs and iPSCs exhibited a modest H3K4Me1 enrichment on MERVL loci (fig. S5J), *miR-34a* overexpression efficiently silenced MERVL in *lsd1*$^{-/-}$ ESCs (fig. S5K), suggesting that a global increase of H3K4Me1 had no effect on the direct MERVL silencing by *miR-34a*. Hence, none of the tested epigenetic mechanisms appeared to be essential for *miR-34a*-mediated MERVL repression.

miR-34a directly represses Gata2

Gata2 plays an essential role in mediating MERVL activation in *miR-34a*$^{-/-}$ pluripotent stem cells (Fig. 5D), and also emerges as a strong candidate for a direct *miR-34a* target (Fig. 6A). *gata2* harbors three predicted *miR-34a* binding sites (*39*), two of which (site 1 and site 3) were strong predictions, with a perfect seed match and/or a strong 3′ complementary base-pairing (Fig. 6A). Consistent with this hypothesis, Gata2 exhibited *miR-34a*-dependent regulation in pluripotent stem cells—*gata2* protein and mRNA levels are increased in *miR-34a*$^{-/-}$ pluripotent stem cells and reduced upon ectopic *miR-34a* overexpression (Fig. 6, B and C). Additionally, a luciferase reporter containing a fragment of the *gata2* gene with all three predicted *miR-34a* binding sites exhibited *miR-34a*-dependent repression, and mutating all three predicted *miR-34a* binding sites in this reporter abolished this regulation (Fig. 6D). Consistent with *gata2* derepression in *miR-34a*$^{-/-}$ ESCs/iPSCs, a number of previously validated Gata2 targets, including *dab2*, *cd34*, and *gata1* (*40*–*42*), exhibited Gata2-dependent upregulation in *miR-34a*$^{-/-}$ iPSCs (fig. S6, A and B). Taken together, *gata2* is likely a direct *miR-34a* target in ESCs/iPSCs, thereby mediating MERVL expression transcriptionally.

Knockdown of *gata2* in *miR-34a*$^{-/-}$ iPSCs phenocopies *miR-34a* overexpression, not only downregulating the expression of MERVL and MERVL-proximal genes (Fig. 4C, Fig. 5D, and fig. S5C), but also abolishing their bi-directional developmental potential (Fig. 6, E to G). In an EB differentiation assay, *gata2* knockdown or *miR-34a* overexpression in *miR-34a*$^{-/-}$ pluripotent cells impaired the induction of the TE marker *cdx2*, without affecting the induction of the markers for all three germ layers (Fig. 6, E and F). More importantly, although the control infected *miR-34a*$^{-/-}$ iPSCs contributed to both ICM and TE in 50% of the chimeric blastocysts (n = 7/14, Fig. 6G and table S1), *miR-34a*$^{-/-}$ iPSCs with *gata2* knockdown lost this expanded cell fate potential and failed to differentiate into both embryonic and extra-embryonic lineages (n = 0/13, Fig. 6G and table S1). Consistent with these findings, late-passage *miR-34a*$^{-/-}$ ESCs that have lost MERVL induction and bi-directional potential also exhibit a decrease in *gata2* level through an unknown mechanism (fig. S2, B and C, and fig. S6C). Hence, *miR-34a* is a non-coding RNA that restricts cell fate potential of ESCs/iPSCs to a pluripotent state. Whereas multiple *miR-34a* targets are likely to act collectively to restrict cell fate potential and repress MERVL expression in pluripotent stem cells, the *miR-34a*/Gata2 axis clearly plays an essential role in this process (Fig. 7).

Discussion

Mouse zygotes and early blastomeres possess a totipotent cell fate potential, generating both embryonic and extra-embryonic cell types during normal development (*5*, *6*). As preimplantation development progresses, this totipotent or totipotent-like cell fate potential is gradually restricted (*43*, *44*). By the late blastocyst stage, the separation of TE (which develops into the placenta), epiblast (which forms the embryo proper) and PE (which develops into the yolk sac) signals the completion of the first cell fate specification event in mammalian development (*45*). This cell fate specification commits the developmental potential of cells to either embryonic or extra-embryonic lineages (*45*). ESCs and iPSCs in culture faithfully recapitulate the developmentally restricted, pluripotent cell fate potential of the epiblast, as they efficiently contribute to all embryonic cell lineages in vivo, but rarely to extra-embryonic lineages (*4*). Experimentally, stem cells with bi-directional developmental potential can be obtained, albeit with low efficiency, using somatic nuclear transfer, genetic modifications or specific enrichment/culture/derivation procedures in ESC/iPSC culture (*7*–*10*, *23*, *46*). This low efficiency in generating bi-potential ESCs reflects the existence of cellular and molecular impediments that restrict the pluripotent cell fate potential.

miR-34a plays an important role in maintaining cell fate

identity in multiple contexts, as it restricts pluripotent cell fate potential of ESCs/iPSCs and acts as a barrier to somatic reprogramming (*17*). *miR-34a* is a non-coding RNA whose deficiency in ESCs or iPSCs generates a significant fraction of cells with bi-directional developmental potential, and strongly induces the expression of MERVL ERVs. Unlike previously reported bi-potential pluripotent stem cells, *miR-34a*$^{-/-}$ ESCs/iPSCs are more stable in maintaining their cell fate potential, and contain a high percentage of MERVL-positive cells in culture (*7*, *9*). MERVL induction in *miR-34a*$^{-/-}$ ESCs/iPSCs could simply be an indicator of a unique 2C-like transcriptional and epigenetic state. An interesting parallel can be drawn to human ESCs, wherein the levels of specific retrotransposon families marks a dynamic population of naive-state pluripotency (*47*, *48*). Alternatively, MERVL could functionally contribute to the establishment and/or maintenance of the 2C-like cell fate potential in part by rewiring gene regulatory networks to induce many 2C-specific, MERVL proximal genes (Fig. 4, A and B) (*7*, *9*). Although most MERVL proximal genes are poorly characterized so far, we found a few genes that mediate important cellular events in preimplantation development, including Hnf4a that regulates visceral endoderm differentiation (*49*) and Usp17lc (Dub-2) that impacts the development of trophectoderm (*50*). Taken together, these findings suggest the possibility that ERVs, while traditionally viewed as evolutionary remnants of invading foreign DNA sequences, could have important yet unrecognized developmental functions.

Although *miR-34a* deficiency expands cell fate potential in pluripotent stem cells, *miR-34a*$^{-/-}$ mice undergo normal preimplantation development under laboratory conditions (*17*, *51*). Interestingly, *miR-34a* is induced at the 4C to 8C transition during preimplantation development (fig. S5B), precisely coinciding with the initiation of MERVL silencing (*29–31*, *52*). We compared the levels of MERVL-Gag protein, MERVL transcripts, and MERVL proximal gene expression in wild-type and *miR-34a*$^{-/-}$ preimplantation embryos, but observed no obvious differences (fig. S7, A and B). It is conceivable that the molecular network mediating MERVL silencing in preimplantation embryos is highly redundant. The six *miR-34* family miRNAs could act redundantly to repress MERVL expression; and *miR-34*-independent pathways could also act in parallel to ensure a rapid and efficient MERVL silencing in vivo. In addition, it is important to recognize the incomplete penetrance of the expanded cell fate potential in *miR-34a*$^{-/-}$ ESCs/iPSCs (Fig. 2, A and B). Pluripotent stem cell culture provides a highly sensitive experimental system that enables detection of altered cell fate potential at the single cell level, whereas a partially altered cell fate potential in early blastomeres could be tolerated in mouse preimplantation embryos due to their unusually robust cell fate plasticity (*6*, *53*).

Our studies identified Gata2 as an important *miR-34a* target that plays an essential role in mediating the MERVL induction and the bi-directional cell fate potential of pluripotent stem cells. In our studies, Gata2 is necessary but not sufficient in promoting MERVL induction in pluripotent stem cells. Gata2 knockdown in *miR-34a*$^{-/-}$ ESCs abolished MERVL induction (Fig. 5D and fig. S5C), yet Gata2 overexpression in wild-type ESCs failed to induce MERVL expression (fig. S8A) or enhance Gata2 binding to MERVL loci (fig. S8B). Since each miRNA often regulates many targets, additional *miR-34a* targets could act cooperatively with elevated Gata2 to confer its binding to MERVL loci and to induce MERVL expression.

In summary, pluripotent stem cell culture contains mutually exclusive populations of pluripotent, MERVL^{low}/Oct4^{high} cells and bi-potent, MERVL^{high}/Oct4^{low} cells (*7*). In wild-type ESC/iPSC culture, this equilibrium strongly favors the MERVL^{low}/Oct4^{high} population at the expense of the MERVL^{high}/Oct4^{low} population (Fig. 7). *miR-34a* deficiency increases Gata2-dependent transcription of MERVL endogenous retroviruses, shifting the equilibrium to enable more cells to acquire a bi-potential cell fate (Fig. 7). Our findings indicate that an intricate network of protein-coding genes, miRNAs, and retrotransposons act cooperatively to restrict cell fate plasticity and to define the pluripotent state of ESCs/iPSCs. Furthermore, these observations raise the tantalizing possibility of establishing in vitro conditions for culturing bi-potential pluripotent stem cells to explore the molecular basis for totipotency.

Materials and methods

Derivation of ESCs and generation of iPSCs

For ESC derivation, uteri containing E3.5 wild-type or *miR-34a*$^{-/-}$ embryos were isolated from timed pregnancies, and transferred individually to a 12-well plate with irradiated MEF (mouse embryonic fibroblasts) feeders. After 5 days of incubation, embryo outgrowth was separated from TE, individually picked, and expanded in mouse ES medium. For iPSC generation, wild-type and *miR-34a*$^{-/-}$ primary MEFs were isolated from littermate-controlled E13.5 wild-type and *miR-34a*$^{-/-}$ embryos, infected with retroviruses generated by pMX retroviral vectors that encode mouse Oct4, Sox2 and Klf4, and cultured on irradiated MEF feeder in ES medium. Subsequently, single iPSC-like colonies were individually picked and expanded on irradiated MEF feeders to establish stable lines. Cell lines used for in vivo experiments were validated free from microplasma contamination by PCR.

RNA-seq data analysis

RNA-seq reads were mapped to the GRCm38 (mm10)

reference genome with TopHat to quantify gene and retrotransposon expression levels. MERVL-gene junctions were defined as those junctions, identified by TopHat, which overlap on one side with an annotated Ensembl gene (including protein coding genes, long ncRNAs, pseudogenes and antisense transcripts) and on the other side with an annotated element of MERVL (including both complete, truncated and solo LTR copies). We used EdgeR to test for differential expression between *miR-34a*$^{-/-}$ and wild-type iPSCs. We defined a gene as differentially expressed (DE) if it had an absolute $\log_2$-fold-change greater or equal than 2 and a False Discover Rate of 0.05.

We performed the analysis using three datasets: HiSeq2000 100bp paired-end data (100PE); NextSeq500 150bp paired-end data (150PE); and a combined dataset obtained by pooling the reads of the two (combined). This allowed us to quantify the effect of read length and sequencing depth on our analyses: the results are highly reproducible across the three datasets (fig. S4 and tables S2 to S5).

Real-time PCR analysis for gene expression

RNA was isolated by Trizol extraction following manufacturer's instruction (Life Technologies, Cat. # 15596). cDNA was reverse-transcribed using iScript Advanced Reverse-Transcriptase (Bio-Rad, Cat. # 1725037). For single colony analysis, cDNA was prepared using a Single Cell-to-Ct qRT-PCR kit (Life Technologies, Cat. # 4458236). All real-time qPCR analyses were performed using SYBR FAST qPCR Master Mix (Kapa Biosystems, Cat. # KK4604). All primers used are listed in table S6. To detect MERVL expression, four pairs of primers were designed to amplify specific regions of MERVL (Fig. 2C) and yielded similar results. One pair of primers detecting the MERVL *pol* region was used for all other MERVL real-time PCR analyses.

Embryoid body (EB) differentiation

For EB differentiation, ESCs or iPSCs were plated in 10cm petri dish (150,000 cells/ml) in ES cell medium without LIF and gently cultured on a rotator after removal of feeder cells. Samples were collected at day 0, 3, 6 and 9 post differentiation for real-time PCR analyses and for immunofluorescence staining.

Generation of chimeric blastocysts or chimeric mice from ESCs/iPSCs

To generate chimeric blastocysts by microinjection, four ESCs or one ESC of the desired genotype were injected into each E2.5 C57Bl/6N wild-type, recipient morulae, which was then cultured overnight to obtain chimeric blastocysts. To generate chimeric blastocysts by morula aggregation, one-cell embryos were cultured for 48h into morulae. After the removal of zona pellucida, morulae were aggregated with ESCs or iPSCs and then cultured overnight. To generate chimeric mice by microinjection, 10-15 ESCs of the desired genotype were injected into E3.5 recipient blastocyst embryos before implanted into the uterus of pseudo-pregnant mothers. Chimeric embryos were collected at E9.5, E12.5 and E14.5 for immunofluorescence (IF) and immunohistochemistry analyses.

Immunofluorescence (IF) and immunohistochemistry (IHC)

For IF staining, cells, EBs or embryos paraffin sections were fixed and blocked before incubated with primary antibodies (1:100, Cdx2, Abcam, Cat. # ab76541; 1:100), followed by secondary antibody incubation and DAPI staining. For IHC of teratomas and chimeric mouse embryos, tissue paraffin sections were incubated with primary antibodies against PL-1 (1:75, Santa Cruz Biotechnology, Cat. # sc-34713) or GFP (1: 100, Abcam, Cat. # ab38689).

REFERENCES AND NOTES

1. M. J. Evans, M. H. Kaufman, Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154–156 (1981). doi:10.1038/292154a0 Medline
2. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006). doi:10.1016/j.cell.2006.07.024 Medline
3. G. R. Martin, 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.* **78**, 7634–7638 (1981). doi:10.1073/pnas.78.12.7634 Medline
4. R. S. Beddington, E. J. Robertson, An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733–737 (1989). Medline
5. V. E. Papaioannou, J. Mkandawire, J. D. Biggers, Development and phenotypic variability of genetically identical half mouse embryos. *Development* **106**, 817–827 (1989). Medline
6. A. K. Tarkowski, Experiments on the development of isolated blastomers of mouse eggs. *Nature* **184**, 1286–1287 (1959). doi:10.1038/1841286a0 Medline
7. T. S. Macfarlan, W. D. Gifford, S. Driscoll, K. Lettieri, H. M. Rowe, D. Bonanomi, A. Firth, O. Singer, D. Trono, S. L. Pfaff, Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63 (2012). doi:10.1038/nature11244 Medline
8. S. M. Morgani, M. A. Canham, J. Nichols, A. A. Sharov, R. P. Migueles, M. S. H. Ko, J. M. Brickman, Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep.* **3**, 1945–1957 (2013). doi:10.1016/j.celrep.2013.04.034 Medline
9. T. Ishiuchi, R. Enriquez-Gasca, E. Mizutani, A. Bošković, C. Ziegler-Birling, D. Rodriguez-Terrones, T. Wakayama, J. M. Vaquerizas, M.-E. Torres-Padilla, Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671 (2015). doi:10.1038/nsmb.3066 Medline
10. M. Abad, L. Mosteiro, C. Pantoja, M. Cañamero, T. Rayon, I. Ors, O. Graña, D. Megías, O. Domínguez, D. Martínez, M. Manzanares, S. Ortega, M. Serrano, Reprogramming in vivo produces teratomas and iPS cells with totipotency features. *Nature* **502**, 340–345 (2013). doi:10.1038/nature12586 Medline
11. M. Chekulaeva, W. Filipowicz, Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr. Opin. Cell Biol.* **21**, 452–460 (2009). doi:10.1016/j.ceb.2009.04.009 Medline
12. S. L. Ameres, P. D. Zamore, Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14**, 475–488 (2013). doi:10.1038/nrm3611 Medline
13. C. Kanellopoulou, S. A. Muljo, A. L. Kung, S. Ganesan, R. Drapkin, T. Jenuwein, D. M. Livingston, K. Rajewsky, Dicer-deficient mouse embryonic stem cells are

defective in differentiation and centromeric silencing. *Genes Dev.* **19**, 489–501 (2005). doi:10.1101/gad.1248505 Medline
14. R. L. Judson, J. E. Babiarz, M. Venere, R. Blelloch, Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat. Biotechnol.* **27**, 459–461 (2009). doi:10.1038/nbt.1535 Medline
15. S. R. Viswanathan, G. Q. Daley, Lin28: A microRNA regulator with a macro role. *Cell* **140**, 445–449 (2010). doi:10.1016/j.cell.2010.02.007 Medline
16. S. R. Viswanathan, C. H. Mermel, J. Lu, C.-W. Lu, T. R. Golub, G. Q. Daley, microRNA expression during trophectoderm specification. *PLOS ONE* **4**, e6143 (2009). doi:10.1371/journal.pone.0006143 Medline
17. Y. J. Choi, C.-P. Lin, J. J. Ho, X. He, N. Okada, P. Bu, Y. Zhong, S. Y. Kim, M. J. Bennett, C. Chen, A. Ozturk, G. G. Hicks, G. J. Hannon, L. He, miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nat. Cell Biol.* **13**, 1353–1360 (2011). doi:10.1038/ncb2366 Medline
18. N. Okada, C.-P. Lin, M. C. Ribeiro, A. Biton, G. Lai, X. He, P. Bu, H. Vogel, D. M. Jablons, A. C. Keller, J. E. Wilkinson, B. He, T. P. Speed, L. He, A positive feedback between p53 and miR-34 miRNAs mediates tumor suppression. *Genes Dev.* **28**, 438–450 (2014). doi:10.1101/gad.233585.113 Medline
19. T.-C. Chang, E. A. Wentzel, O. A. Kent, K. Ramachandran, M. Mullendore, K. H. Lee, G. Feldmann, M. Yamakuchi, M. Ferlito, C. J. Lowenstein, D. E. Arking, M. A. Beer, A. Maitra, J. T. Mendell, Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. *Mol. Cell* **26**, 745–752 (2007). doi:10.1016/j.molcel.2007.05.010 Medline
20. N. Raver-Shapira, E. Marciano, E. Meiri, Y. Spector, N. Rosenfeld, N. Moskovits, Z. Bentwich, M. Oren, Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. *Mol. Cell* **26**, 731–743 (2007). doi:10.1016/j.molcel.2007.05.017 Medline
21. L. He, X. He, L. P. Lim, E. de Stanchina, Z. Xuan, Y. Liang, W. Xue, L. Zender, J. Magnus, D. Ridzon, A. L. Jackson, P. S. Linsley, C. Chen, S. W. Lowe, M. A. Cleary, G. J. Hannon, A microRNA component of the p53 tumour suppressor network. *Nature* **447**, 1130–1134 (2007). doi:10.1038/nature05939 Medline
22. H. Tazawa, N. Tsuchiya, M. Izumiya, H. Nakagama, Tumor-suppressive miR-34a induces senescence-like growth arrest through modulation of the E2F pathway in human colon cancer cells. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 15472–15477 (2007). doi:10.1073/pnas.0707351104 Medline
23. H. Niwa, Y. Toyooka, D. Shimosato, D. Strumpf, K. Takahashi, R. Yagi, J. Rossant, Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell* **123**, 917–929 (2005). doi:10.1016/j.cell.2005.08.040 Medline
24. M. Giakoumopoulos, T. G. Golos, Embryonic stem cell-derived trophoblast differentiation: A comparative review of the biology, function, and signaling mechanisms. *J. Endocrinol.* **216**, R33–R45 (2013). doi:10.1530/JOE-12-0433 Medline
25. J. Artus, A. Piliszek, A.-K. Hadjantonakis, The primitive endoderm lineage of the mouse blastocyst: Sequential transcription factor activation and regulation of differentiation by Sox17. *Dev. Biol.* **350**, 393–404 (2011). doi:10.1016/j.ydbio.2010.12.007 Medline
26. D. G. Simmons, A. L. Fortier, J. C. Cross, Diverse subtypes and developmental origins of trophoblast giant cells in the mouse placenta. *Dev. Biol.* **304**, 567–578 (2007). doi:10.1016/j.ydbio.2007.01.009 Medline
27. C. Kubaczka, C. E. Senner, M. Cierlitza, M. J. Araúzo-Bravo, P. Kuckenberg, M. Peitz, M. Hemberger, H. Schorle, Direct induction of trophoblast stem cells from murine fibroblasts. *Cell Stem Cell* **17**, 557–568 (2015). doi:10.1016/j.stem.2015.08.005 Medline
28. H. Benchetrit, S. Herman, N. van Wietmarschen, T. Wu, K. Makedonski, N. Maoz, N. Yom Tov, D. Stave, R. Lasry, V. Zayat, A. Xiao, P. M. Lansdorp, S. Sebban, Y. Buganim, Extensive nuclear reprogramming underlies lineage conversion into functional trophoblast stem-like cells. *Cell Stem Cell* **17**, 543–556 (2015). doi:10.1016/j.stem.2015.08.006 Medline
29. D. Kigami, N. Minami, H. Takayama, H. Imai, MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **68**, 651–654 (2003). doi:10.1095/biolreprod.102.007906 Medline
30. A. E. Peaston, A. V. Evsikov, J. H. Graber, W. N. de Vries, A. E. Holbrook, D. Solter, B. B. Knowles, Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004). doi:10.1016/j.devcel.2004.09.004 Medline
31. A. V. Evsikov, W. N. de Vries, A. E. Peaston, E. E. Radford, K. S. Fancher, F. H. Chen, J. A. Blake, C. J. Bult, K. E. Latham, D. Solter, B. B. Knowles, Systems biology of the 2-cell mouse embryo. *Cytogenet. Genome Res.* **105**, 240–250 (2004). doi:10.1159/000078195 Medline
32. L. Bénit, J. B. Lallemand, J. F. Casella, H. Philippe, T. Heidmann, ERV-L elements: A family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J. Virol.* **73**, 3301–3308 (1999). Medline
33. J. Nichols, B. Zevnik, K. Anastassiadis, H. Niwa, D. Klewe-Nebenius, I. Chambers, H. Schöler, A. Smith, Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95**, 379–391 (1998). doi:10.1016/S0092-8674(00)81769-9 Medline
34. T. S. Macfarlan, W. D. Gifford, S. Agarwal, S. Driscoll, K. Lettieri, J. Wang, S. E. Andrews, L. Franco, M. G. Rosenfeld, B. Ren, S. L. Pfaff, Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25**, 594–607 (2011). doi:10.1101/gad.2008511 Medline
35. Z. Xue, K. Huang, C. Cai, L. Cai, C. Y. Jiang, Y. Feng, Z. Liu, Q. Zeng, L. Cheng, Y. E. Sun, J. Y. Liu, S. Horvath, G. Fan, Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* **500**, 593–597 (2013). doi:10.1038/nature12364 Medline
36. F. Tang, C. Barbacioru, E. Nordman, S. Bao, C. Lee, X. Wang, B. B. Tuch, E. Heard, K. Lao, M. A. Surani, Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLOS ONE* **6**, e21208 (2011). doi:10.1371/journal.pone.0021208 Medline
37. I. A. Maksakova, P. J. Thompson, P. Goyal, S. J. M. Jones, P. B. Singh, M. M. Karimi, M. C. Lorincz, Distinct roles of KAP1, HP1 and G9a/GLP in silencing of the two-cell-specific retrotransposon MERVL in mouse ES cells. *Epigenetics Chromatin* **6**, 15 (2013). doi:10.1186/1756-8935-6-15 Medline
38. M. Hayashi, K. Maehara, A. Harada, Y. Semba, K. Kudo, H. Takahashi, S. Oki, C. Meno, K. Ichiyanagi, K. Akashi, Y. Ohkawa, Chd5 regulates MuERV-L/MERVL expression in mouse embryonic stem cells via H3K27me3 modification and histone H3.1/H3.2. *J. Cell. Biochem.* **117**, 780–792 (2016). doi:10.1002/jcb.25368 Medline
39. P. Loher, I. Rigoutsos, Interactive exploration of RNA22 microRNA target predictions. *Bioinformatics* **28**, 3322–3323 (2012). doi:10.1093/bioinformatics/bts615 Medline
40. C. Zhang, X. Ye, H. Zhang, M. Ding, H. Deng, GATA factors induce mouse embryonic stem cell differentiation toward extraembryonic endoderm. *Stem Cells Dev.* **16**, 605–613 (2007). doi:10.1089/scd.2006.0077 Medline
41. S. H. Orkin, GATA-binding transcription factors in hematopoietic cells. *Blood* **80**, 575–581 (1992). Medline
42. M. Suzuki, M. Kobayashi-Osaki, S. Tsutsumi, X. Pan, S. Ohmori, J. Takai, T. Moriguchi, O. Ohneda, K. Ohneda, R. Shimizu, Y. Kanki, T. Kodama, H. Aburatani, M. Yamamoto, GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. *Genes Cells* **18**, 921–933 (2013). doi:10.1111/gtc.12086 Medline
43. K. Piotrowska-Nitsche, A. Perea-Gomez, S. Haraguchi, M. Zernicka-Goetz, Four-cell stage mouse blastomeres have different developmental properties. *Development* **132**, 479–490 (2005). doi:10.1242/dev.01602 Medline
44. I. Tabansky, A. Lenarcic, R. W. Draft, K. Loulier, D. B. Keskin, J. Rosains, J. Rivera-Feliciano, J. W. Lichtman, J. Livet, J. N. H. Stern, J. R. Sanes, K. Eggan, Developmental bias in cleavage-stage mouse blastomeres. *Curr. Biol.* **23**, 21–31 (2013). doi:10.1016/j.cub.2012.10.054 Medline
45. K. Cockburn, J. Rossant, Making the blastocyst: Lessons from the mouse. *J. Clin. Invest.* **120**, 995–1003 (2010). doi:10.1172/JCI41229 Medline
46. I. Wilmut, A. E. Schnieke, J. McWhir, A. J. Kind, K. H. Campbell, Viable offspring derived from fetal and adult mammalian cells. *Nature* **385**, 810–813 (1997). doi:10.1038/385810a0 Medline
47. J. Wang, G. Xie, M. Singh, A. T. Ghanbarian, T. Raskó, A. Szvetnik, H. Cai, D. Besser, A. Prigione, N. V. Fuchs, G. G. Schumann, W. Chen, M. C. Lorincz, Z. Ivics, L. D. Hurst, Z. Izsvák, Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516**, 405–409 (2014). doi:10.1038/nature13804 Medline
48. T. W. Theunissen, M. Friedli, Y. He, E. Planet, R. C. O'Neil, S. Markoulaki, J. Pontis, H. Wang, A. Iouranova, M. Imbeault, J. Duc, M. A. Cohen, K. J. Wert, R. Castanon,

Z. Zhang, Y. Huang, J. R. Nery, J. Drotar, T. Lungjangwa, D. Trono, J. R. Ecker, R. Jaenisch, Molecular criteria for defining the naive human pluripotent state. *Cell Stem Cell* **19**, 502–515 (2016). doi:10.1016/j.stem.2016.06.011 Medline
49. S. A. Duncan, A. Nagy, W. Chan, Murine gastrulation requires HNF-4 regulated gene expression in the visceral endoderm: Tetraploid rescue of Hnf-4$^{-/-}$ embryos. *Development* **124**, 279–287 (1997). Medline
50. K.-H. Baek, H. Lee, S. Yang, S.-B. Lim, W. Lee, J. E. Lee, J.-J. Lim, K. Jun, D.-R. Lee, Y. Chung, Embryonic demise caused by targeted disruption of a cysteine protease Dub-2. *PLOS ONE* **7**, e44223 (2012). doi:10.1371/journal.pone.0044223 Medline
51. C. P. Concepcion, Y.-C. Han, P. Mu, C. Bonetti, E. Yao, A. D'Andrea, J. A. Vidigal, W. P. Maughan, P. Ogrodowski, A. Ventura, Intact p53-dependent responses in miR-34-deficient mice. *PLOS Genet.* **8**, e1002797 (2012). doi:10.1371/journal.pgen.1002797 Medline
52. P. Svoboda, P. Stein, M. Anger, E. Bernstein, G. J. Hannon, R. M. Schultz, RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269**, 276–285 (2004). doi:10.1016/j.ydbio.2004.01.028 Medline
53. J. Rossant, Postimplantation development of blastomeres isolated from 4- and 8-cell mouse eggs. *J. Embryol. Exp. Morphol.* **36**, 283–290 (1976). Medline
54. M. H. Tan, K. F. Au, A. L. Yablonovitch, A. E. Wills, J. Chuang, J. C. Baker, W. H. Wong, J. B. Li, RNA sequencing reveals a diverse and dynamic repertoire of the *Xenopus tropicalis* transcriptome over development. *Genome Res.* **23**, 201–216 (2013). doi:10.1101/gr.141424.112 Medline
55. B. A. Westerman, A. K. Braat, N. Taub, M. Potman, J. H. A. Vissers, M. Blom, E. Verhoeven, H. Stoop, A. Gillis, A. Velds, W. Nijkamp, R. Beijersbergen, L. A. Huber, L. H. J. Looijenga, M. van Lohuizen, A genome-wide RNAi screen in mouse embryonic stem cells identifies Mp1 as a key mediator of differentiation. *J. Exp. Med.* **208**, 2675–2689 (2011). doi:10.1084/jem.20102037 Medline
56. R. T. Willett, L. A. Greene, Gata2 is required for migration and differentiation of retinorecipient neurons in the superior colliculus. *J. Neurosci.* **31**, 4444–4455 (2011). doi:10.1523/JNEUROSCI.4616-10.2011 Medline
57. C. Trapnell, L. Pachter, S. L. Salzberg, TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009). doi:10.1093/bioinformatics/btp120 Medline
58. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014). doi:10.1093/bioinformatics/btt656 Medline
59. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. H. Yang, J. Zhang, Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004). doi:10.1186/gb-2004-5-10-r80 Medline
60. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010). doi:10.1093/bioinformatics/btp616 Medline
61. B. Li, C. N. Dewey, RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323 (2011). doi:10.1186/1471-2105-12-323 Medline
62. J. H. Bullard, E. Purdom, K. D. Hansen, S. Dudoit, Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. *BMC Bioinformatics* **11**, 94 (2010). doi:10.1186/1471-2105-11-94 Medline
63. W. E. Johnson, C. Li, A. Rabinovic, Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007). doi:10.1093/biostatistics/kxj037 Medline
64. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012). doi:10.1093/bioinformatics/bts034 Medline
65. K. V. Mardia, J. T. Kent, J. M. Bibby, *Multivariate Analysis* (Academic Press, 1979).
66. V. Pereira, Automated paleontology of repetitive DNA with REANNOTATE. *BMC Genomics* **9**, 614 (2008). doi:10.1186/1471-2164-9-614 Medline
67. H. R. Chiang, L. W. Schoenfeld, J. G. Ruby, V. C. Auyeung, N. Spies, D. Baek, W. K. Johnston, C. Russ, S. Luo, J. E. Babiarz, R. Blelloch, G. P. Schroth, C. Nusbaum, D. P. Bartel, Mammalian microRNAs: Experimental evaluation of novel and previously annotated genes. *Genes Dev.* **24**, 992–1009 (2010). doi:10.1101/gad.1884710 Medline
68. W.-M. Liu, R. T. K. Pang, P. C. N. Chiu, B. P. C. Wong, K. Lao, K.-F. Lee, W. S. B. Yeung, Sperm-borne microRNA-34c is required for the first cleavage division in mouse. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 490–494 (2012). doi:10.1073/pnas.1110368109 Medline

ACKNOWLEDGMENTS

We thank V. Prideaux, W. Wang, R. Huang, K. N. Li, H. Aaron, J-Y. Lee, W. Xu, J. Ong, P. Cheung, B. Zaghi, M. Chung, J. Choi, A. Li, A. Perez, W. Bao, S. Tindall, K. Zhao, K. Cui, B. Xue, O. Tam, K. Heydari, A. Valeros, MJ Bennett, D. Young, N. Anchell, JA Wood and H. Noller for technical assistance. We thank L. Xie, V. A. Modzelewski and R. Song for discussion and input. We also thank T. Heidmann, J. Rossant, A. Li, V. Krizhanovsky, M. Stadtfeld, M.C. Lorincz, Y. Shinkai, D. Trono, T. Chen and R. Jaenisch for sharing valuable reagents. P. Margolis for carefully reading the manuscript and M. Rape and N. Patel for sharing Olympus Revolution XD spinning disk confocal microscope and Zeiss LSM 700 confocal microscope with us. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley (supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303) and the computing resource provided by the Center for systems Biology in UT Dallas. L.H. is supported by a new faculty award by CIRM (RN2-00923-1), an R01 (R01 CA139067) from the National Cancer Institute (NCI), an R01 (R01GM114414) from NIGMS, and a HHMI faculty scholar award. S.C. is supported by a CIRM predoctoral fellowship and a CRCC predoctoral fellowship. C-P.L. is supported by a Siebel postdoctoral fellowship and a CIRM postdoctoral fellowship. D.R. is supported by an NIH grant U01MH105979. Z.X. is partially supported by the UT-Dallas faculty startup fund. The RNA sequencing data are publicly available at the NCBI Gene Expression Omnibus with the accession number GSE69484. The annotation of retrotransposons in GFF format is available as supplementary data.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/cgi/content/full/science.aag1927/DC1
Materials and Methods
Figs. S1 to S8
Tables S1 to S6
References (*54*–*68*)
Data S1

27 May 2016; accepted 14 December 2016
Published online 12 January 2017
10.1126/science.aag1927

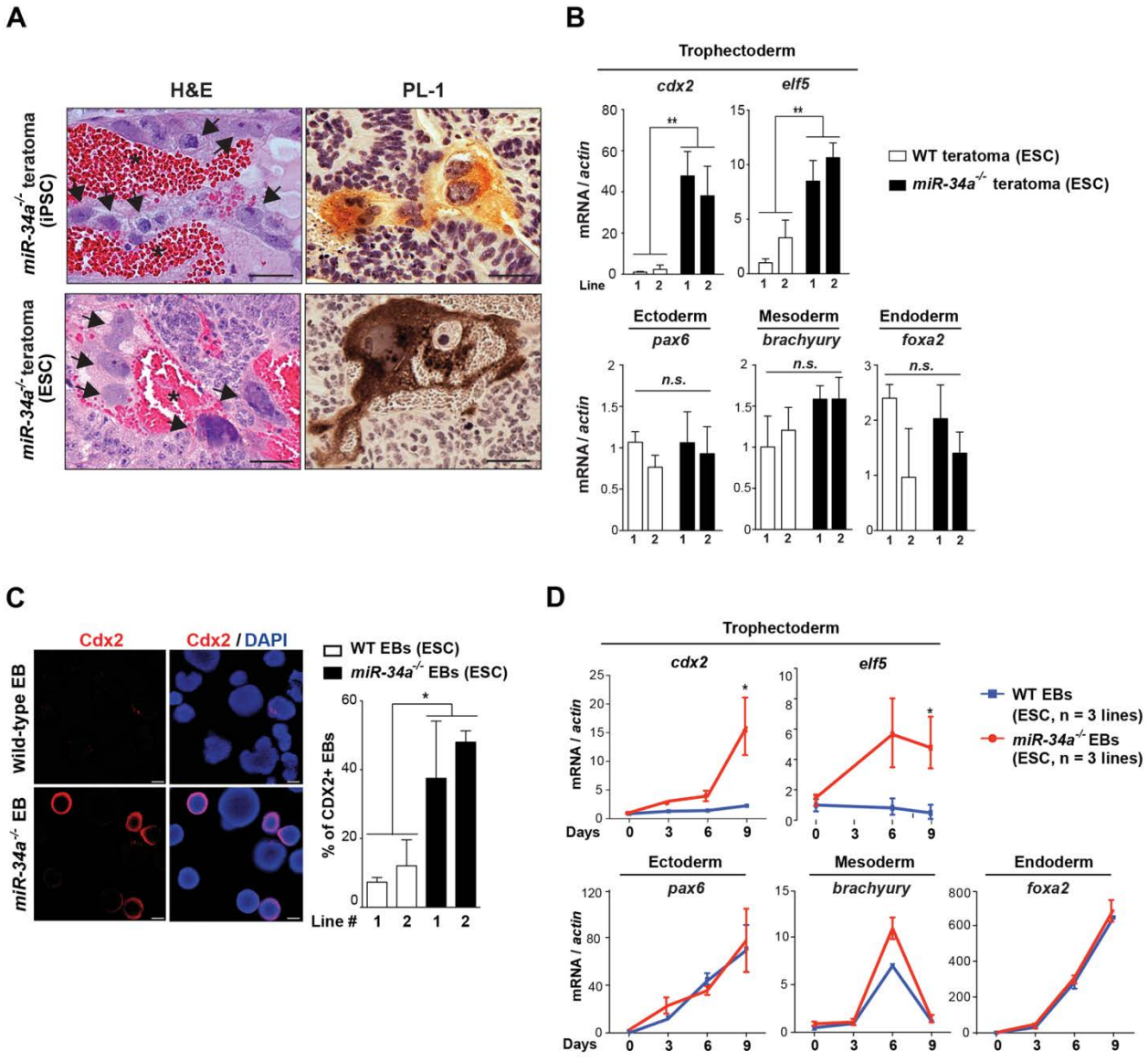

Fig. 1. *miR-34a*$^{-/-}$ pluripotent stem cells generate both embryonic and extra-embryonic lineages in teratomas and EBs. (**A**) *miR-34a*$^{-/-}$ teratomas contain embryonic and extra-embryonic cell lineages. Teratomas generated from *miR-34a*$^{-/-}$ iPSCs and ESCs contain cells characterized by placental trophoblast giant cell like morphology (black arrows) and strong placental lactogen 1 (PL-1) expression. Asterisks: the blood-filled lacunae associated with placental trophoblast giant cell-like cells. Scale bars, 50 μm. (**B**) The trophectoderm marker *cdx2* and *elf5* were highly induced in *miR-34a*$^{-/-}$ teratomas, but not wild-type controls, as determined by real-time PCR. In contrast, wild-type and *miR-34a*$^{-/-}$ teratomas similarly induced the expression of *pax6* (an ectoderm marker), *brachyury* (a mesoderm marker) and *foxa2* (an endoderm marker). Teratomas were generated from two independent pairs of passage- and littermate- controlled wild-type and *miR-34a*$^{-/-}$ ESC lines. Error bars: *s.d.*, n = 3. (**C**) *miR-34a*$^{-/-}$ EBs yield a greater percentage of Cdx2-positive EBs and a stronger Cdx2 expression, with the Cdx2 induction primarily in cells at the periphery. Scale bars, 100 μm. Error bars: *s.d.*, n, 5-7 randomly selected 10x fields. (**D**) *miR-34a*$^{-/-}$ EBs showed an increase in the expression of TE markers *cdx2* and *elf5*. In contrast, wild-type and *miR-34a*$^{-/-}$ EBs similarly induced the expression of *pax6*, *brachyury* and *foxa2.* EBs were generated from three independent pairs of passage- and littermate-controlled wild-type and *miR-34a*$^{-/-}$ ESC lines. Error bars, *s.e.m.*, n = 3. * $P < 0.05$, ** $P < 0.01$, *n.s.*, not significant. All *P*-values were calculated on a basis of a two-tailed Student's *t*-test.

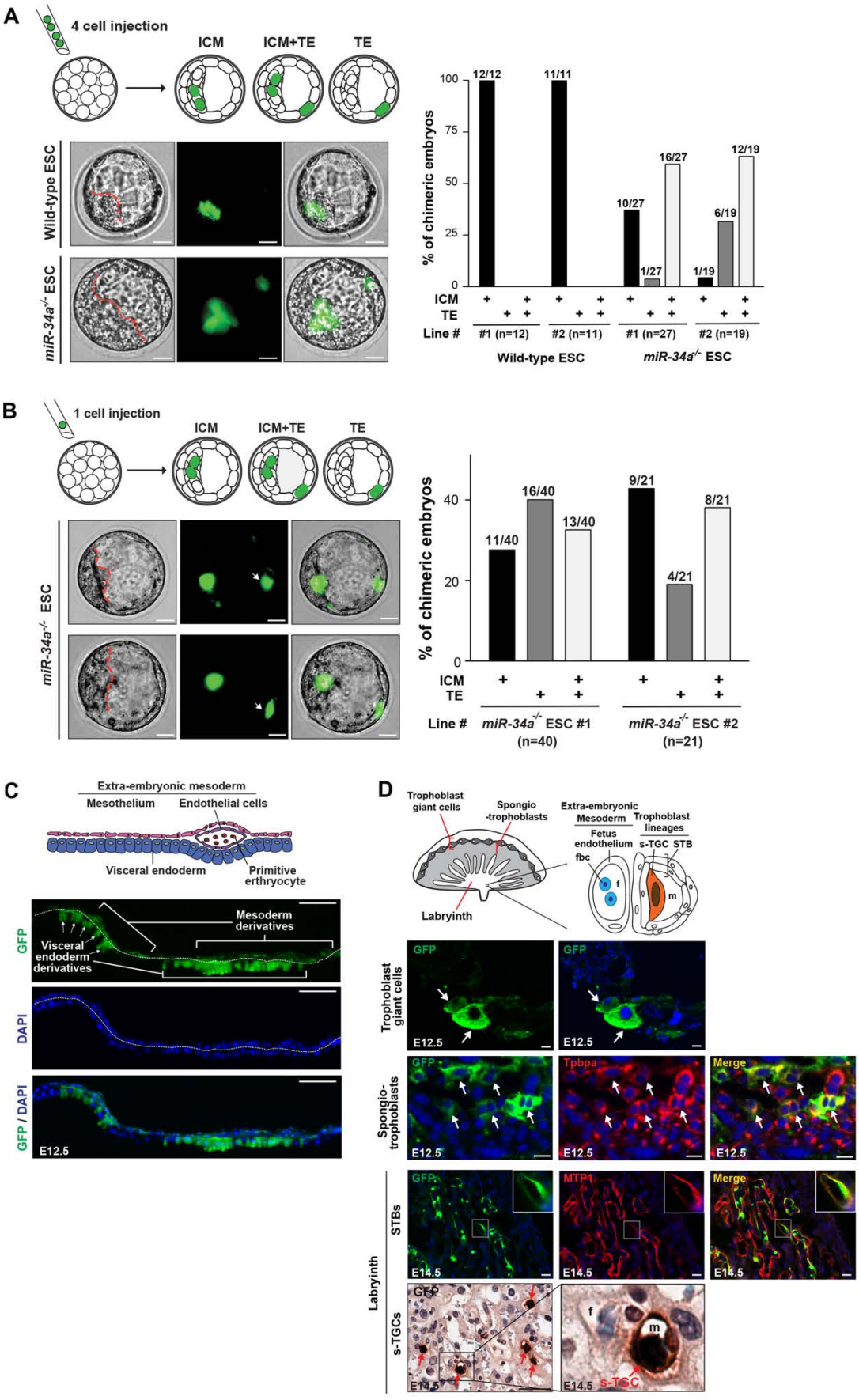

Fig. 2 (preceding page). *miR-34a*$^{-/-}$ ESCs exhibit an expanded cell fate potential in chimeric embryos in vivo. (**A**) Four GFP-labeled wild-type or *miR-34a*$^{-/-}$ ESCs were microinjected into each C57BL/6N recipient morula; and their contribution to the ICM and the TE were determined by the localization of GFP-positive ESC progenies (left). Scale bar, 20 μm. The percentage of chimeric blastocyst embryos with ESC contribution to the ICM, the TE, and the ICM+TE were measured for both wild-type and *miR-34a*$^{-/-}$ ESCs (right). (**B**) A single GFP-labeled *miR-34a*$^{-/-}$ ESC is able to contribute to both the ICM and the TE (white arrows) of a chimeric blastocyst. Representative images were shown for two chimeric blastocysts (left). Scale bar, 20 μm. The percentage of chimeric embryos with single ESC contribution to the ICM, the TE, and the ICM+TE were quantified (right). Two independent pairs of passage- and littermate-controlled wild-type and *miR-34a*$^{-/-}$ ESCs were compared in (A). Two independent *miR-34a*$^{-/-}$ ESC lines were examined in (B). *n* indicates the number of chimeric blastocyst embryos for each ESC line from three independent injections. (**C** and **D**) *miR-34a*$^{-/-}$ ESCs contribute to differentiated cell lineages in embryo, yolk sac, and placenta in vivo. 10-15 GFP-labeled wild-type or *miR-34a*$^{-/-}$ ESCs were microinjected into each C57BL/6N blastocyst to generate chimeric embryos. (C) A diagram illustrates the yolk sac tissue architecture and its major cell types (top). GFP-labeled *miR-34a*$^{-/-}$ ESCs yield visceral endoderm cells that are identified based on the bilaminar structure of the yolk sac and their characteristic columnar epithelial morphology (bottom). Scale bar, 50 μm. (D) A diagram illustrates the placenta tissue architecture and its major cell types (top). GFP-labeled *miR-34a*$^{-/-}$ ESCs generate trophoblast giant cells (white arrows) and s-TGCs (red arrows) that are identified based on their specific distribution in placenta and their unique cell morphology; GFP-positive spongiotrophoblasts or STBs (white arrows) are identified based on the IF co-staining with trophoblast-specific proteins alpha (Tpbpa) or ferroportin (MTP1) (bottom). Scale bar, 50 μm. The specific extra-embryonic cell lineages contributed by GFP labeled *miR-34a*$^{-/-}$ ESCs are not identical among chimeric embryos (table S1), yet ~30% chimeric embryos have at least two different extra-embryonic cell types derived from the *miR-34a*$^{-/-}$ ESCs.

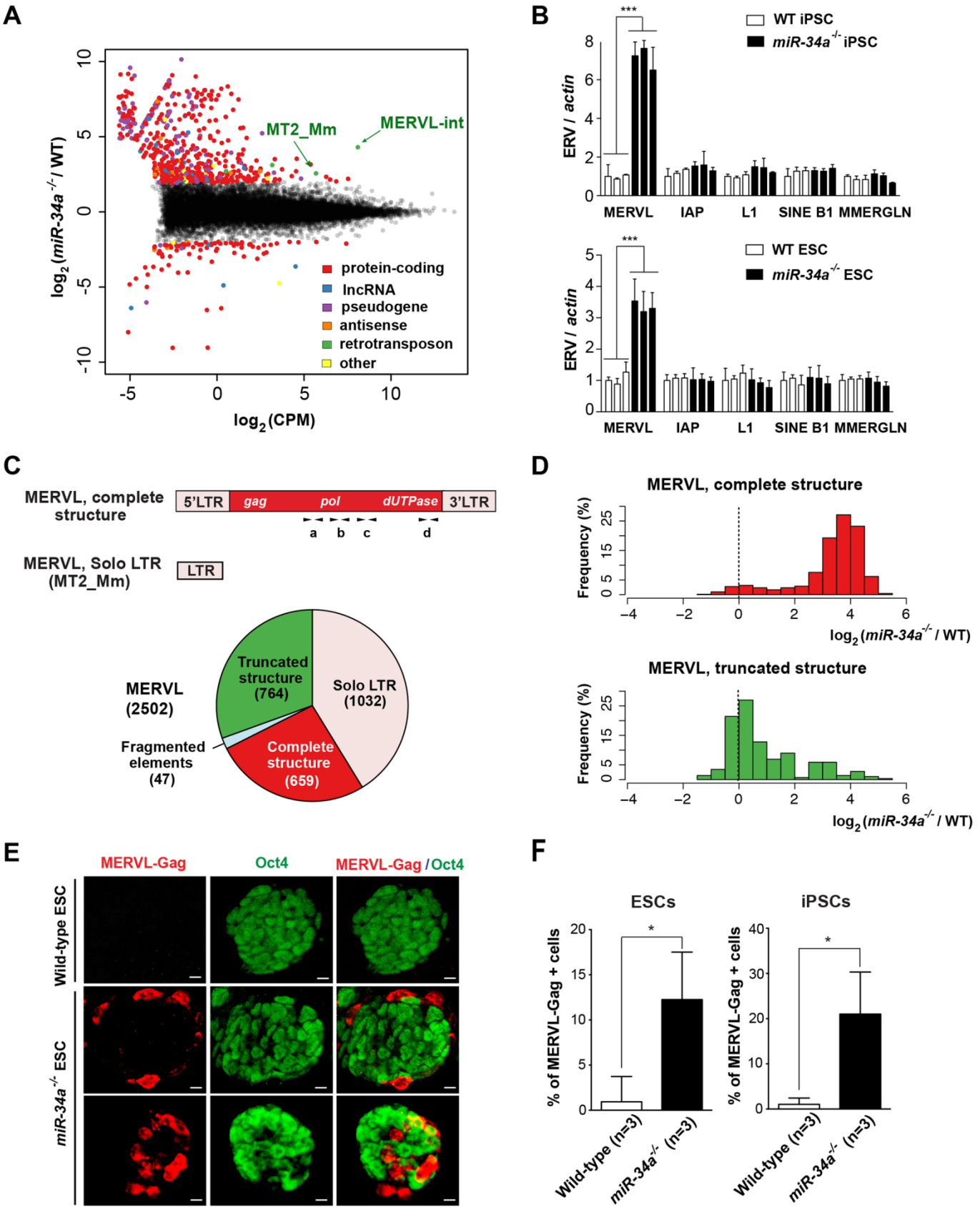

Fig. 3 (preceding page). *miR-34a*$^{-/-}$ pluripotent stem cells exhibit specific induction of the MERVL ERVs. The MERVL ERVs are a highly induced and differentially expressed (DE) transcriptional unit in *miR-34a*$^{-/-}$ iPSCs compared to wild-type iPSCs. (**A**) An MA-plot compares the transcription profiles of *miR-34a*$^{-/-}$ and wild-type (WT) iPSCs using RNA-seq data. DE transcriptional units (False Discovery Rate 5%, absolute $\log_2$ fold-change ≥ 2), including protein-coding genes (n = 352), long non-coding RNAs (lncRNAs) (n = 13), pseudogenes (n = 54), antisense transcripts (n = 6), and retrotransposon families (n = 4), are color-coded by class. Among these DE transcripts, MERVL-int (the MERVL internal sequence that encodes Gag, Pol and dUTPase) and MT2_Mm (the canonical MERVL solo-LTR) are strongly induced and highly expressed in *miR-34a*$^{-/-}$ iPSCs. CPM: counts per million. (**B**) MERVL induction in *miR-34a*$^{-/-}$ iPSCs and ESCs is confirmed by real-time PCR analyses. All other ERVs tested, including intracisternal A-particle (IAP), LINE L1, SINE B1, and MMERGLN, exhibited no significant difference between wild-type and *miR-34a*$^{-/-}$ pluripotent stem cells. In (A) and (B), error bars: *s.d.*, n = 3; three independent pairs of passage- and littermate-controlled wild-type and *miR-34a*$^{-/-}$ iPSC and ESC lines were compared. (**C**) (Top) A schematic diagram illustrates the transcript structure of a canonical MERVL element. The positions of four pairs of validated real-time PCR primers (a, b, c, and d) are indicated. (Bottom) A pie chart shows the relative abundance for the different MERVL subclasses in the C57B6/J genome, categorized according to their structural features. Complete MERVL loci (red) carry both 5′- and 3′-LTRs, flanking the internal sequence MERVL-int; truncated MERVL elements (green) lack one or both LTRs; solo-LTRs (pink), also designated as MT2_Mm, were generated through homologous recombination during evolution. (**D**) The MERVL induction in *miR-34a*$^{-/-}$ iPSCs occurs primarily in loci with a complete structure. Histograms are shown for the $\log_2$ fold-change of the expressed MERVL loci between *miR-34a*$^{-/-}$ and wild-type iPSCs derived from the RNA-seq data, either with a complete structure (top) or with a truncated structure (bottom). The expression of solo-LTRs cannot be accurately measured by the current RNA-seq data, due to their short sequences and repetitive nature. (**E**) The MERVL-Gag and Oct4 expression is mutually exclusive in *miR-34a*$^{-/-}$ ESCs. In comparison, MERVL-Gag staining is mostly absent in wild-type ESCs. Scale bars, 20 μm. (**F**) The percentage of MERVL-Gag-positive cells were quantified in early passage of passage- and littermate-controlled wild-type and *miR-34a*$^{-/-}$ ESCs (left) and iPSCs (right). Error bars: *s.d.*, 5-6 randomly selected 10x fields were quantified for each ESC line tested. * $P < 0.05$, *** $P < 0.001$. All *P*-values were calculated on a basis of a two-tailed Student's *t*-test.

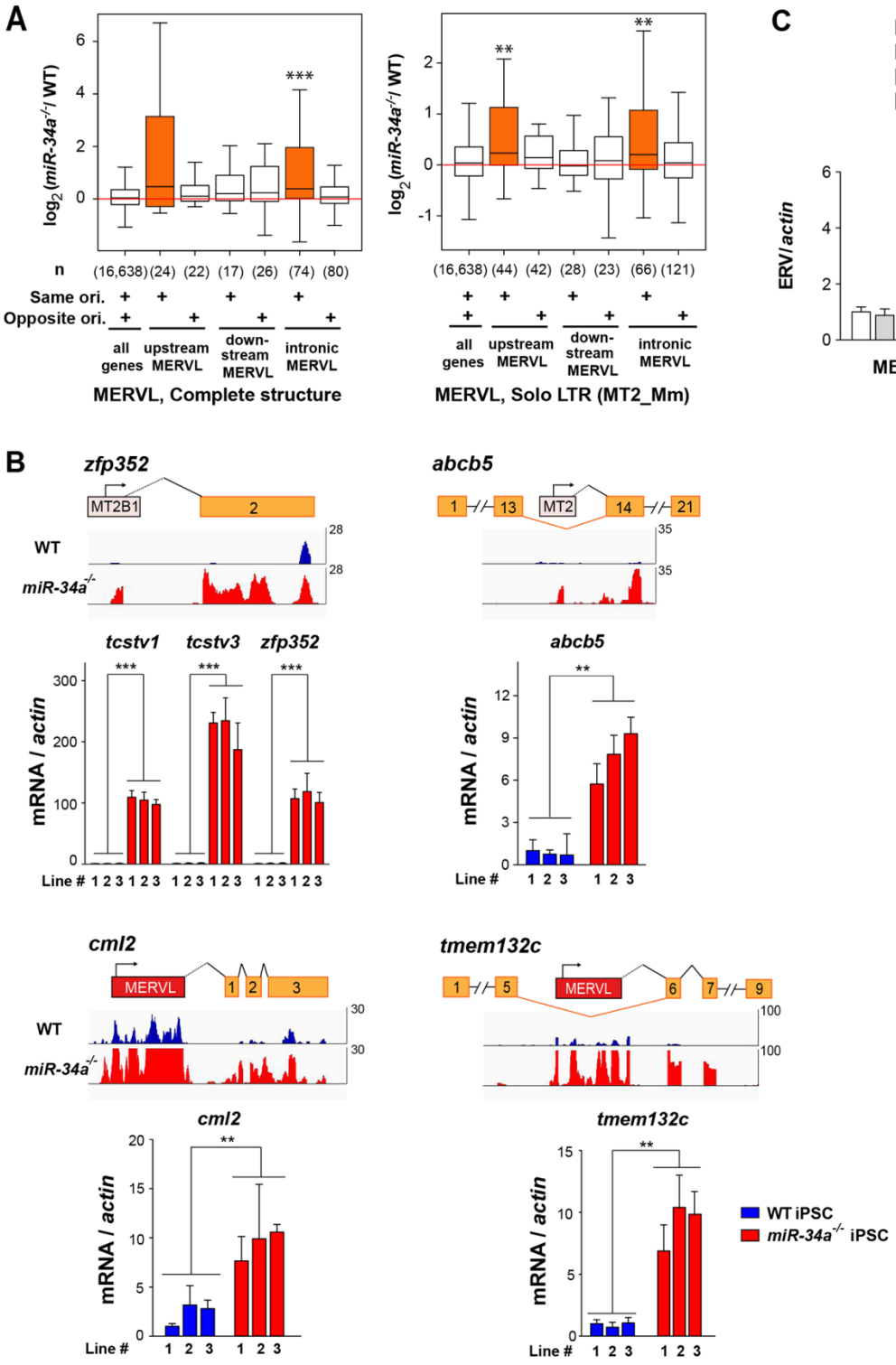

Fig. 4 (preceding page). MERVL reactivation in *miR-34a*$^{-/-}$ pluripotent stem cells induces proximal host gene isoforms. (**A**) MERVL-proximal genes with an upstream or intronic MERVL element on the same strand are preferentially up-regulated in *miR-34a*$^{-/-}$ iPSCs. Box-plots of $\log_2$ fold-change f ESCs between *miR-34a*$^{-/-}$ and wild-type iPSCs derived from the RNA-seq data are shown for genes proximal to annotated, complete MERVL loci (left) or MERVL solo LTR (MT2_Mm) loci (right). The numbers in parentheses indicate the number of protein-coding genes in each category; the box plot of the $\log_2$ fold-change of all protein-coding genes is included as a reference. Same ori, (same orientation): the MERVL element and its proximal gene are on the same strand; opposite ori, (opposite orientation): the MERVL element and its proximal gene are on opposite strands. Genes harboring a complete MERVL copy in their introns on the same strand, as well as genes with MT2_Mm on the same strand (either upstream or in their introns), exhibited significantly larger fold-change compared to other genes. **, $P < 0.01$; *** $P < 0.001$; Wilcoxon-Mann-Whitney test. (**B**) Examples are shown for induced expression and altered transcript structure of MERVL-proximal genes in *miR-34a*$^{-/-}$ iPSCs. The MT2B1 solo LTR elements, a subfamily of MERVL with highly related LTR sequences, reside upstream of *zfp352*, *tcstv1*, and *tcstv3*, and act as promoters to strongly induce their expression in *miR-34a*$^{-/-}$ iPSCs. Similarly, a complete MERVL element upstream of *cml2* acts as an alternative promoter to induce the expression of an MERVL-*cml2* isoform in *miR-34a*$^{-/-}$ iPSCs. MERVL elements located within introns, either as solo LTRs (as the MT2_Mm element in intron 13 of *abcb5*) or as complete ERV elements (as that in intron 5 of *tmem132c*), also act as alternative promoters to drive the expression of truncated gene isoforms that contain only the downstream exons. Given the repetitive nature of MERVL LTRs, we can not definitively determine which LTR of the full length MERVL element acts as the alternative promoters for the proximal genes using the RNA-seq data. (**C**) *miR-34a* overexpression in *miR-34a*$^{-/-}$ iPSCs using a MSCV (murine stem cell virus) retroviral vector significantly suppressed the level of MERVL and the MERVL-*zfp352* chimeric transcript in real-time PCR analysis. In comparison, *miR-34a* overexpression caused no alteration in the level of IAP retrotransposon. One pair of passage- and littermate-controlled wild-type and *miR-34a*$^{-/-}$ iPSCs were examined. In (B) and (C), error bars: *s.d.*, n = 3. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, n.s., not significant. All *P*-values were calculated on a basis of a two-tailed Student's *t*-test.

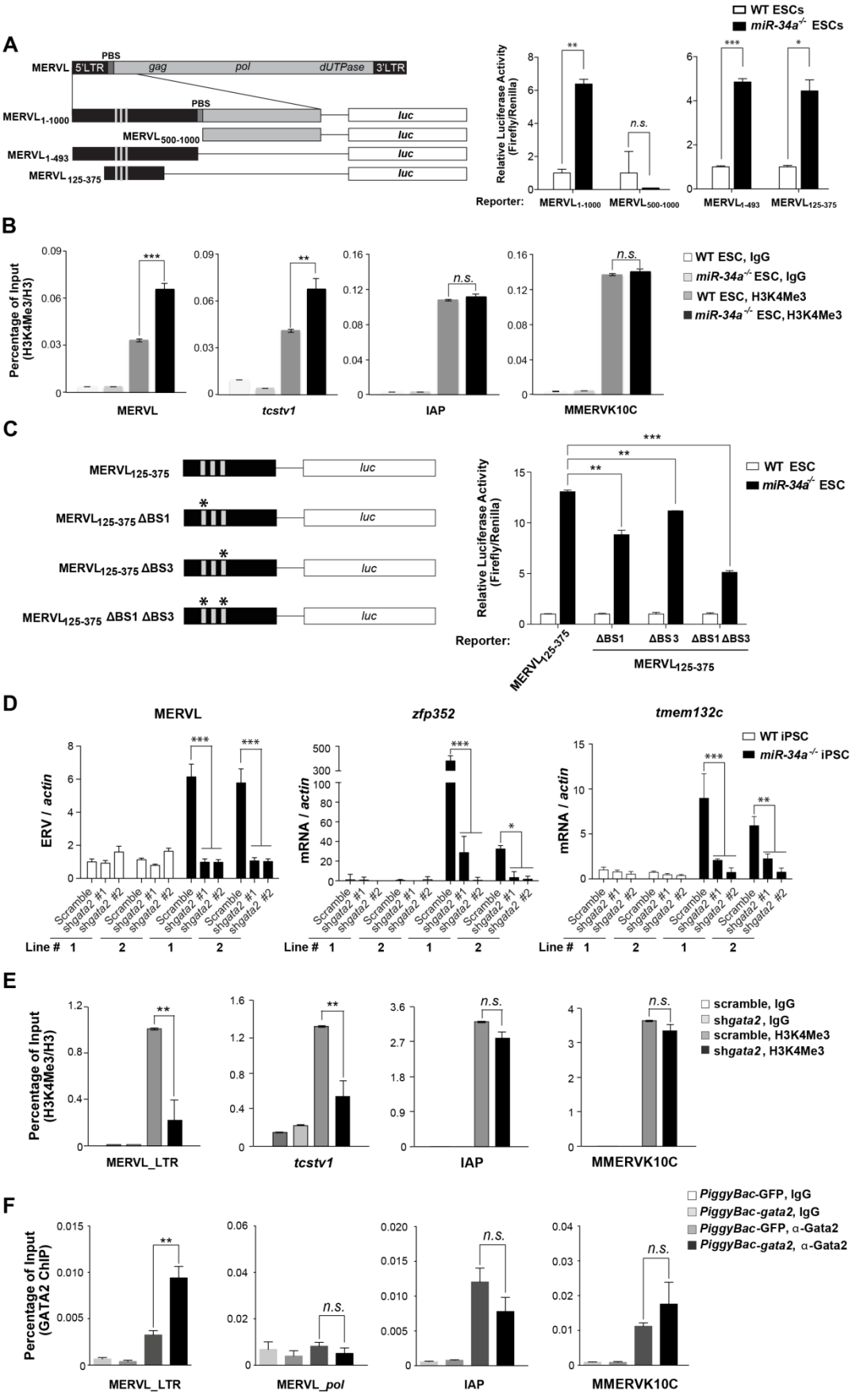

Fig. 5 (preceding page). Gata2 is essential for the MERVL induction in *miR-34a*$^{-/-}$ pluripotent stem cells. (**A**) A full-length or partial MERVL LTR can be activated specifically in *miR-34a*$^{-/-}$ESCs in a luciferase reporter assay. (Left) A schematic diagram shows the MERVL LTR and the truncated MERVL LTR fragments that were tested for promoter activity using luciferase (Luc) assays. (Right) The Luc reporters driven by MERVL fragments containing the full length LTR (MERVL$_{1\text{-}1000}$-Luc and MERVL$_{1\text{-}493}$-Luc) exhibit a strong activation in *miR-34a*$^{-/-}$ ESCs, but not in wild-type ESCs. A 250-bp truncated MERVL$_{125\text{-}375}$ fragment recapitulates this differential reporter activity in wild-type and *miR-34a*$^{-/-}$ ESCs. Error bars: *s.d.*, n = 2. PBS: primer binding site. (**B**) Chromatin Immunoprecipitation (ChIP) reveals increased H3K4Me3 modification on the LTR region of MERVL loci and the MERVL-*tcstv1* chimeric gene in *miR-34a*$^{-/-}$ ESCs. The H3K4Me3 level is unaltered on IAP LTR and MMERVK10C LTR. (**C**) Gata2 is important for the MERVL$_{125\text{-}375}$-Luc reporter activity in *miR-34a*$^{-/-}$ ESCs. (Left) A schematic diagram shows the MERVL$_{125\text{-}375}$-Luc reporter and the mutated derivatives that harbor mutations of predicted Gata2 binding sites. (Right) Mutation of both BS1 and BS3 in the MERVL$_{125\text{-}375}$-Luc reporter synergistically impairs the promoter reporter activity in *miR-34a*$^{-/-}$ ESCs. Error bars: *s.d.*, n = 2. (**D**) *gata2* knockdown in *miR-34a*$^{-/-}$ iPSCs significantly decreases the expression of MERVL elements and MERVL-proximal genes. Using two independent shRNAs targeting *gata2*, we effectively knocked down *gata2* in *miR-34a*$^{-/-}$ iPSCs (fig. S5C), and observed a decreased expression of MERVL elements and MERVL-proximal genes (*zfp352* and *tmem132c*) by real time PCR. (**E**) *gata2* knockdown in *miR-34a*$^{-/-}$ iPSCs significantly decreases the H3K4Me3 deposition on MERVL elements and MERVL_ *tcstv1*, but has no effects on H3K4Me3 deposition on IAP or MMERVK10C. (**F**) ChIP reveals an increase of Gata2 binding to the MERVL LTR region upon Gata2 overexpression in *miR-34a*$^{-/-}$ iPSCs. Gata2 binding to MERVL *pol* region, IAP LTR, or MMERVK10C LTR was unaltered. Error bars: *s.d.*, n = 3. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, *n.s.*, not significant. All *P*-values were calculated on a basis of a two-tailed Student's *t*-test.

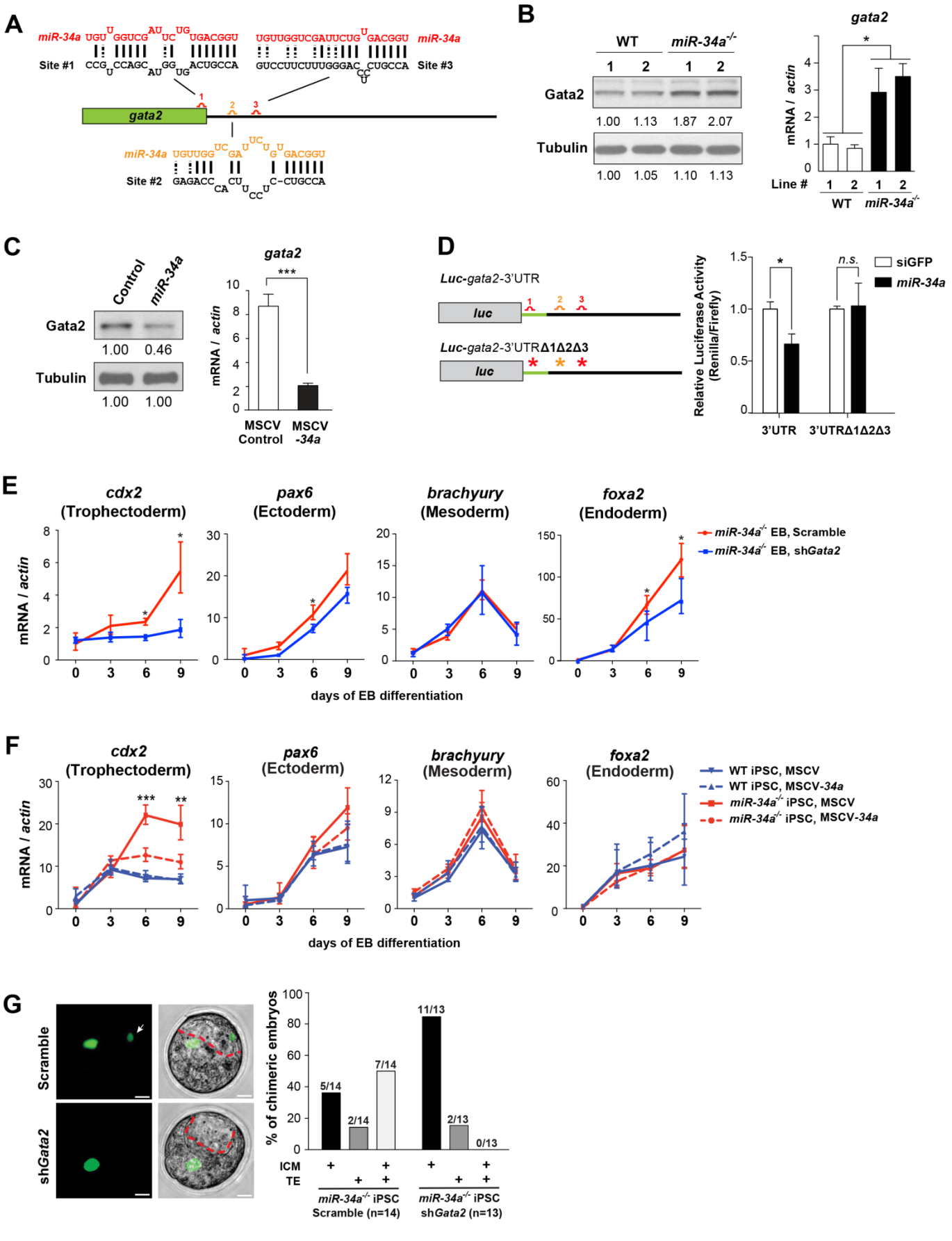

Fig. 6 (preceding page). *miR-34a* restricts cell fate potential of pluripotent stem cells by targeting *gata2.* (**A**) A schematic representation of three predicted *miR-34a* binding sites in the *gata2* mRNA, with one site (1) located at the 3′ end of the open reading frame (ORF) and two sites (2 and 3) located within the 3′-UTR. Site 1 (red) is predicted as a strong *miR-34a* binding site by both the 7mer-A1 seed-match rule and duplex folding energy (*39*). Site 3 (red) does not have a perfect seed match, but contains compensatory 3′ base-pairing and exhibits a strong folding energy. In comparison, site 2 (orange) represents a weaker prediction yet with a reasonable folding energy (*39*). (**B** and **C**) *gata2* exhibits *miR-34a*-dependent repression in *miR-34a*$^{-/-}$ pluripotent stem cells. (B) Gata2 protein (left) and *gata2* mRNA (right) levels are elevated in *miR-34a*$^{-/-}$ iPSCs compared to wild-type iPSCs. Two independent pairs of passage- and littermate-controlled wild-type and *miR-34a*$^{-/-}$ iPSC lines were measured by Western blotting and real-time PCR analyses. Error bars: *s.d.*, n = 3. (C) Overexpression of *miR-34a* in *miR-34a*$^{-/-}$ iPSCs represses Gata2 protein (left) and mRNA levels (right). (**D**) Mutating all three predicted *miR-34a* binding sites within the Luc-gata2-3UTR luciferase reporter completely abolishes *miR-34a*-dependent repression. Error bars: *s.d.*, n = 2. (**E**) RNAi-mediated *gata2* knockdown in *miR-34a*$^{-/-}$ iPSCs abolishes the induction of the TE marker *cdx2* during EB differentiation, but has no effects on the induction of ectoderm (*pax6*), mesoderm (*brachyury*), or endoderm (*foxa2*) markers. Error bars: *s.d.*, n = 3. (**F**) *miR-34a* overexpression in *miR-34a*$^{-/-}$ iPSCs using MSCV retroviral vectors significantly suppresses the induction of TE markers *cdx2* upon EB differentiation. The induction of ectoderm (*pax6*), mesoderm (*brachyury*), and endoderm (*foxa2*) markers are not affected. Error bars: *s.d.*, n = 3. In (E) and (F), results shown are representative of two independent experiments using the same *miR-34a*$^{-/-}$ iPSC line. (**G**) The expanded cell fate potential of *miR-34a*$^{-/-}$ iPSCs is restricted by *gata2* knockdown in chimeric assays. We injected four GFP-labeled iPSCs into each recipient morula to generate chimeric blastocysts. While 50% of chimeric blastocysts generated from control infected *miR-34a*$^{-/-}$ iPSCs contain iPSC contribution to both the ICM and the TE (n = 7/14), *miR-34a*$^{-/-}$ iPSCs with *gata2* knockdown lacked this expanded cell fate potential (n = 0/13), and primarily contributed to the ICM (n = 11/13). * $P < 0.05$, *** $P < 0.001$, *n.s.*, not significant. All *P*-values were calculated on a basis of a two-tailed Student's *t*-test.

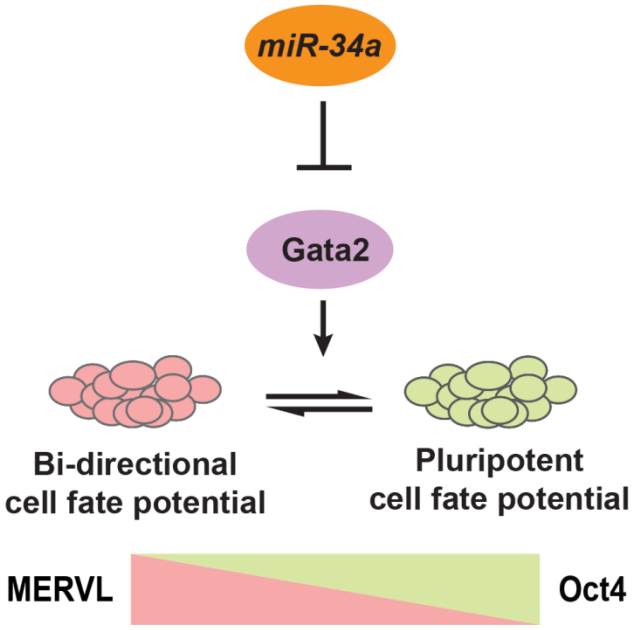

Fig. 7. A model on the role of the *miR-34a*/Gata2 pathway in restricting cell fate potential in ESCs/iPSCs to a pluripotent state.

Deficiency of microRNA *miR-34a* expands cell fate potential in pluripotent stem cells

Yong Jin Choi, Chao-Po Lin, Davide Risso, Sean Chen, Thomas Aquinas Kim, Meng How Tan, Jin B. Li, Yalei Wu, Caifu Chen, Zhenyu Xuan, Todd Macfarlan, Weiqun Peng, K. C. Kent Lloyd, Sang Yong Kim, Terence P. Speed and Lin He (January 12, 2017)
published online January 12, 2017

Editor's Summary

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005.