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

Author
Choi, Yong Jin

Publication Date
2016

Peer reviewed|Thesis/dissertation

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

by

YONG JIN CHOI

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Comparative Biochemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Lin He, Chair

Fenyong Liu

Sangwei Lu

Dirk Hockemeyer

Summer 2016

Abstract

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

by

YONG JIN CHOI

Doctor of Philosophy in Comparative Biochemistry

University of California, Berkeley

Professor Lin He, Chair

Embryonic stem cells and induced pluripotent stem cells have pluripotent developmental potential, efficiently giving rise to all embryonic cell types, but rarely extra-embryonic lineages. Here, we identify a microRNA *miR-34a*, whose deficiency in mouse pluripotent stem cells expands their developmental potential to generate both embryonic and extra-embryonic lineages *in vitro* and *in vivo*. *miR-34a*$^{-/-}$ pluripotent stem cells with this bidirectional cell fate potential resemble totipotent 2-cell (2C) blastomeres not only in their cell fate potential, but also in the key molecular signature, namely a strong induction of the MuERV-L (MERVL) family of murine endogenous retroviruses (ERVs). *miR-34a* represses MERVL expression through transcriptional regulation, at least in part, by repressing the transcription factor GATA-binding protein 2 (Gata2). Consistently, the *miR-34a*/Gata2 pathway restricts the acquisition of bidirectional cell fate potential in pluripotent stem cells. Altogether, our findings provide vital insights into the complex molecular network that defines and restrict the developmental potential of pluripotent stem cells.

Table of Contents

Table of Contents i

List of Figures and Tables ii

Acknowledgements iii

Chapter 1: Introduction 1

Pluripotent stem cell 2
Transposable element 3
microRNA 4

Chapter 2: Materials and Methods 6

Chapter 3: Results 14

miR-34a$^{-/-}$ pluripotent stem cells exhibit expanded cell fate potential 15
miR-34a$^{-/-}$ pluripotent stem cells exhibit an induction of MERVL ERVs 16
MERVL induction in *miR-34a*$^{-/-}$ pluripotent stem cells is regulated transcriptionally 17
Gata2 mediates elevated MERVL expression in *miR-34a*$^{-/-}$ pluripotent stem cells 18
miR-34a restricts cell fate potential of pluripotent stem cells by directly repressing Gata2 19

Chapter 4: Conclusions 21

References 23

List of Figures and Tables

Figure 1: *miR-34a*$^{-/-}$ pluripotent stem cells exhibit expanded cell fate potential.

Figure 2: *miR-34a*$^{-/-}$ pluripotent stem cells exhibit specific induction of the MERVL ERVs.

Figure 3: Gata2 is essential for the MERVL induction in *miR-34a*$^{-/-}$ pluripotent stem cells.

Figure 4: *miR-34a* restricts cell fate potential of pluripotent stem cells by targeting *gata2*.

Figure S1: *miR-34a*$^{-/-}$ pluripotent stem cells exhibit an expanded cell fate potential *in vitro* and *in vivo*

Figure S2: MERVL ERVs are specifically induced in *mir-34a*$^{-/-}$ pluripotent stem cells.

Figure S3: The MERVL induction in *miR-34a*$^{-/-}$ pluripotent stem cells alters the expression and structure of a subset of MERVL proximal genes.

Figure S4: The effect of the length and depth of RNA-seq data on the transcriptional profile characterization of *miR-34a*$^{-/-}$ iPSCs.

Figure S5: Gata2 mediates the MERVL induction in *miR-34a*$^{-/-}$ pluripotent stem cells.

Figure S6: Gata2 is a key target of *miR-34a* in pluripotent stem cells.

Table S1: *miR-34a*$^{-/-}$ ESCs contribute to both embryonic and extra-embryonic cell lineages in chimeric analyses *in vivo*

Table S2: Expression quantification of all retrotransposon families in wild-type and *miR-34a*$^{-/-}$ iPSCs using RNA-seq data

Table S3: Expression quantitation of individual MERVL loci and MERVL-related ERV loci in wild-type and *miR-34a*$^{-/-}$ iPSCs using RNA-seq data

Table S4: A summary of genes differentially expressed between wild-type and *miR-34a*$^{-/-}$ iPSCs using RNA-seq data

Table S5: Quantitation of chimeric junction reads between MERVL or MERVL-related loci and proximal protein-coding genes in wild-type and *miR-34a*$^{-/-}$ iPSCs using RNA-seq data

Table S6: The quantitative PCR primers used in this study

Acknowledgements

First and foremost, I would like to express my sincere gratitude to my advisor Professor Lin He for the continuous support of my Ph.D study and related research, for her patience, motivation and immense knowledge. Her aupport and inspiring suggestions have been precious for the development of this thesis.

Besides my advisors, I would also like to thank my committee members: Professor Fenyong Liu, Professor Sangwei Liu and Professor Dirk Hockemeyer for their insightful comments and encouragement, but also for the hard question which incented me to widen research from various perspectives.

I thank my fellow lab members for the stimulating discussions. I would never forget all the chats and beautiful moments I shared with them. They were fundamental in supporting me during these stressful and difficult moments. A special thanks goes to Paul (Chao-Po Lin) because without him I would have never completed this wonderful experience in Berkeley.

My deepest gratitude goes to my family for all their love and encouragement. For my parents who raised me with a love of science and supported me in all my pursuits. Currently my father fighting cancer after he was diagnosed with cancer of ampulla of vater and my mother helping his fight. What I want him to know is that he isn't fighting cancer. We are all fighting cancer. I have no doubt that he will fight cancer with the same ferocious determination that he have with other challenges in his life. I know that the great strength he have will prevail. I also would like to thank my sister and brother in law for their unflagging love and unconditional support throughout my studies. And most of all for my loving, supportive, encouraging, and patient my wife and my children whose faithful support during the final stages of this Ph.D is so appreciated.

Chapter 1
Introduction

Introduction

Mouse embryonic stem cells (ESCs) derived from the inner cell mass (ICM) of peri-implantation blastocyst embryos, as well as induced pluripotent stem cells (iPSCs) generated by somatic reprogramming, are classically defined as pluripotent stem cells [1–3]. As a population, ESCs and iPSCs efficiently contribute to all embryonic cell types in vitro and in vivo, yet rarely to extra-embryonic cell lineages in placenta and yolk sac [4,5]. This restricted developmental potential of mouse pluripotent stem cells contrasts with that of totipotent zygotes and 2C blastomeres, which give rise to both embryonic and extra-embryonic cell lineages during normal development and ultimately generate the entire organism[6,7].

Differentiated somatic cells can be induced to generate pluripotent stem cells that functionally resemble ESCs, which means iPSCs can differentiate into tissues of all three germ layers, give rise to chimaeric mice, and transmit to the germline[8,9]. This reprogramming process, rooted in the remarkable cellular plasticity retained during differentiation, can be triggered by exogenous expression of a set of ES-cell specific genes[3]. Among the best-characterized reprogramming factors are a defined set of transcriptional regulators, Oct4 and Sox2, Klf4 and c-Myc [3,10–14], which constitute the core gene regulatory circuits that coordinately control pluripotency and self-renewal. With the enforced expression of these reprogramming transcription factors, iPSC generation occurs with low efficiency and slow kinetics, reflecting the existence of cellular and molecular barriers as impediments for this process [15].
Several studies have revealed considerable mechanistic overlap between somatic cell reprogramming and malignant transformation [16]. Many cellular mechanisms that enhance reprogramming, such as increased cell proliferation and cell survival, evasion of DNA damage response and cell immortalization, have long been demonstrated to promote tumorigenesis [17–20]; and several potent oncogenes and tumor suppressors are essential regulators for somatic reprogramming [3,15,16]. In particular, the inactivation of p53, one of the most important tumor suppressors, significantly enhances iPSC generation [17–19,21–23]. As a tumor suppressor, p53's transcriptional regulation converges onto multiple target genes that collectively mediate its downstream effects [24]. Similarly, p53's role in repressing somatic reprogramming is likely to be mediated through multiple targets as well. To date, the cell-cycle regulator p21 is the only p53 target with a demonstrated role in repressing reprogramming, yet its effect only partially phenocopies that of p53 [17,21,25]. Thus, additional p53 targets may exist to mediate the effects of p53 on iPSC generation through a yet unidentified mechanism(s).

Transposable elements (TE) can be colonized the genomes of all eukaryotes and it can be divided into two main classes accordingly to their mechanism of transposition, retrotransposons and DNA transposons[26–29]. Retrotransposons, also known as class I transposable elements, replicates through reverse transposition and amplify via a "copy-and-paste" process [30–32]. This process involves the transcription of an RNA intermediate by the enzyme machinery of the host cell, the subsequent reversetranscription to cDNA, and the integration into the host genome by the enzymes encoded by the retrotransposon. The class I transposable element is composed of two sub-types, the long terminal repeat (LTR) and the non-LTR retrotransposons[30,33]. Non-LTR retrotransposons, including the long interspersed nuclear elements (LINE) and short interspersed nuclear elements (SINE), are the most abundant elements in mammalian genomes.

Approximately 33% of human genome can be recognized as a non-LTR retrotranspons. LTR retrotransposons, also known as endogenous retroviruses (ERVs), constitute approximately 8 and 10% of the human and mouse genome, respectively and are divided into classe I, class II and class III elements [34,35]. Retrotransposition of a subset of ERVs is responsible for up to 10% of all spontaneous mutations in mice and therefore can have detrimental effects on the host fitness[36]. ERVs are endogenous viral elements that have infected by exogenous retroviruses and then integrated into the host cell. Retroviruses usually infect somatic cells and retroviral genes integrated into genomic DNA are not passed on to host progeny[37]. However, some types of retrovirus can occasionally infect germ cells, which means these viral elements are maintained as heritable genetic elements in the host species[38]. When the exogenous retrovirus infect host cell, viral reverse transcriptase make DNA copy of viral genome from RNA viral genome and then the retroviral sequences were integrated into host genome by integrase. The integrated retroviral element is called as a provirus and is recognized as a part of the host genome. The provirus consists of four retroviral genes, gag, pro, pol and env, and two LTRs flanked the 5' and 3' region. The gag gene encodes for the core structural proteins of a retrovirus, which include the viral matrix (MA), capsid (CA) and nucleoproteins (NC). The pro gene encodes the protease, which cleaves the gag polyprotein precursor. The pol gene encodes the reverse transcriptase and integrase proteins, which are required for amplification and integration. The env gene is a viral protein that serves to form the viral envelope. However, most of ERVs are defective of retroviral genes due to the self-defense mechanism of hosts.

In mice, ERVs are divided into three classes based on the sequence of their pol genes and range greatly in copy number from several to tens of thousands of copies per genome[39]. They also vary depends on different mouse strains[40].

First of all, class I ERVs/gammaretroviruses are classified as type C based on virion morphology and constitute up to 0.7% of the mouse genome[41]. This ERV class members are grouped together based on their similarity to the Moloney Murine Leukemia Virus (MoMLV or MLV), which was identified in leukemic cells of the AKR mutant mouse strain[42]. Depending on the mouse strain, there are 25 to 70 copies of MLV elements in the genome[43]. A few members of this class, such as GLN, have the capacity to produce functional virions. Heidmann et al demonstrated that GLN family of highly reiterated ERVs, one copy that encodes retroviral particles prone to infection of mouse cells[44].

Class II ERVs/betaretroviruses are classified as Type B and D based on their viral particle morphology and comprise of around 3% of the mouse genome[45]. This class consists of many more members capable of producing functional retroviral particles as compared to class I[46,47]. Class II ERVs show some sequence similarity to the Mouse Mammary Tumor Virus (MMTV) and Susan et al firstly identified as a factor capable of causing mammary cancers in mice, showing in vertical transmission from mother to offspring via viral particles released into the milk[48]. Another member of this class includes the well-studied non-infectious family of Intracisternal A Particles (IAPs), which is present in approximately 2000 copies in the mouse genome[49,50]. Other class II ERVs includes MusD and the closely related Early Transposon (ETn) elements. The LTR sequences of these two families are virtually identical and are present at a combined ~400 copies in the mouse genome[40].

Class III ERVs, also known as spumavirus-like elements, are the most numerous of all three ERV classes, comprising of about 5.4% of the mouse genome[51]. These elements consist largely of Murine ERV-L (MERV-L)[52]. Class III ERVs are amongst the most transcriptionally active and some elements have even been co-opted by the host for use as promoters during specific

stages of early embryogenesis[53].
As retrotransposition is deleterious, numerous pathways have evolved to repress these retroelements[54]. For example, DNA methylation is required for transcriptional silencing of ERVs in differentiated cells. However, this epigenetic mark is dispensable for silencing of ERV during early stages of mouse embryogenesis and in mouse embryonic stem cells (mESCs). Hutnick et al reported that in demethylated ESCs cultures carrying mutations of DNA methyltransferase I (Dnmt1) show no increased expression of IAPs, which is murine endogenous retroviral repetitive elements, relative to the wild type ESCs and a dramatic increase of IAP mRNA and protein expression was observed upon induction of differentiation through the withdrawal of leukemia-inhibitory factor for 6 or more days suggesting mESCs, in contrast to somatic cells, are a unique stem cell type possessing a DNA methylation-independent IAP repression mechanism[55].
Interestingly, recent studies reported that histone modification and DNA methylation to function together to repress retrotransposons. Tamaru et al suggested that the H3K9 lysine methyltransferase (KMTase) DIM5 is required for CpG methylation in the filamentous fungus Neurospora crassa[56]. Lehnertz et al also demonstrated that the related H3K9 KMTase SUV39H1 and SUV39H2 are required for DNA methylation of pericentromeric repeats, a heterochromatic region in mice[57]. By contrast, DNA methylation is not required for H3K9 methylation of the same region. Thus, H3K9 methylation generally acts upstream of DNA methylation, but the role of this histone modification in silencing of ERVs in mESCs had not been addressed until recently.
Interestingly, unlike most differentiated cell types, particular families of ERVs are expressed in cells of the early embryo and placenta[58,59]. As such elements have a greater chance to amplify and integrate into the germ line, they have the highest copy numbers. Among the most active ERVs are the IAP and MusD/ETn elements[60]. IAP expression has been observed in several mouse tumors and cell lines and ETn expression in undifferentiated embryonic carcinoma (EC) and embryonic stem cells (ESCs)[61]. Both families are highly expressed during early embryogenesis, but are silenced as development progresses. Pfaff etl al demonstrated that at the two-cell (2C) embryo stage, murine endogenous retrovirus (MuERVL, also known as ERV4) elements are transiently derepressed and produced 3% of the transcribed messenger RNAs and after 2C stage, MuERVL retroelement expression is silenced, suggesting this foreign sequence has helped to drive cell-fate regulation in early embryogenesis[62].

Totipotency is a unique feature of mammalian zygotes and early blastomeres, which becomes gradually restricted during preimplantation development [63]. Totipotency can be induced by somatic nuclear transfer, suggesting that this transient developmental state can be re-established experimentally [64]. Rare populations with expanded cell fate potential have been identified in cultured mouse pluripotent stem cells as a result of genetic alterations, specific culture and derivation conditions, or enrichment with specific molecular markers [65–68]. Such mouse pluripotent stem cells resemble 2C-blastomeres, not only differentiating into embryonic and extra-embryonic lineages in functional assays, but also carrying a key molecular signature, namely the strong induction of the MuERV-L (MERVL) family of murine endogenous retroviruses (ERVs). Thus, cultured ESCs/iPSCs retain the cell fate plasticity to acquire features of early blastomeres.

microRNAs (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 [69–75]. These small non-coding RNAs are increasingly recognized as key regulators of cell fate specification in normal development [76] and in pluripotent stem cells [77–80]. Here, we have identified a miRNA, miR-34a, whose deficiency in ESCs and iPSCs expands their cell fate potential, giving rise to both embryonic and extra-embryonic lineages in vitro and in vivo. miR-34a$^{-/-}$ ESCs and iPSCs also exhibit a strong and specific induction of MERVL ERVs, a unique molecular hallmark shared by totipotent 2C-blastomeres and reported 2C-like ESCs [66,67,81]. Our mechanistic studies demonstrate that miR-34a restricts cell fate potential of pluripotent stem cells in the embryonic linages, and silences MERVL expression in ESCs/iPSCs, at least in part, by directly repressing a transcription factor, Gata2. Taken together, we reveal the functional importance of the miR-34a/Gata2 pathway in regulating cell fate plasticity in pluripotent stem cells.

Chapter 2

Materials and Methods

Materials and Methods

Mouse breeding and genotyping

The generation of miR-34a$^{-/-}$ mice was described previously (16). Both wild-type and miR-34a$^{-/-}$ mice were maintained on an isogenic C57BL/6J background, and housed in a non-barrier animal facility at UC-Berkeley. The following primers were used for genotyping, with parenthetical values indicating the size of the diagnostic PCR product: mir-34a-Common-R, ACTGCTGTACCCTGCTGCTT, with mir-34a-WT-F, GTACCCCGACATGCAAACTT (wild-type band, 400 bp), or mir-34a-KO-F, GCAGGACCACTGGATCATTT (knockout band, 263 bp) (16). NCr-nu/nu female athymic mice used for teratoma generation were purchased from Taconic (Taconic, Cat. # NCRNU). All the mouse work was done with approval of University of California, Berkeley's Animal Care and Use Committee. University of California, Berkeley's assurance number is A3084-01, and is on file at the National Institutes of Health Office of Laboratory Animal Welfare.

Derivation of embryonic stem cells (ESCs)

Mouse ESCs were isolated based on published protocols with slight modifications. Uteri containing E3.5 wild-type or miR-34a$^{-/-}$ embryos were isolated from timed pregnant females, and put in Knockout DMEM (Life Technologies, Cat. # 10829-018) supplemented with 10mM HEPES (Life Technologies, Cat. # 15630-080). E3.5 blastocysts were flushed with 1ml syringes with 18G needles and individually transferred to a 12-well plate with irradiated MEF (mouse embryonic fibroblasts) feeders in 1 ml N2B27 medium containing 100 U/ml LIF (EMD Millipore, Cat. # ESG1107), 1 μM PD0325901 (Sigma, Cat. # PZ0162) and 3 μM CHIR99021 (EMD Millipore, Cat. # 361559). After 5 days of incubation, embryo outgrowth was separated from the trophectoderm (TE) and picked up by a 10 μl pipette and transferred to 20 μl Accutase (Life Technologies, Cat. # A11105-01) and incubated at 37^{o}C for 20 min to dissociate cells. Dissociated cells were then cultured on irradiated MEF feeder cells with N2B27 medium containing LIF and two inhibitors for one passage. Subsequently, ESCs were passaged with 0.25% Trypsin-EDTA and maintained in regular mouse ES medium. ESCs were also derived in regular ES medium (see below) to test for variation among derivation protocols.

Generation of induced pluripotent stem cells (iPSCs)

Wild-type and miR-34a$^{-/-}$ iPSCs were generated from primary mouse embryonic fibroblasts (MEFs) by somatic reprogramming (3). Primary MEFs were isolated from littermate-controlled E13.5 wild-type and miR-34a$^{-/-}$ embryos, infected with pMX retroviral vectors that encode mouse Oct4, Sox2 and Klf4 (Addgene, Cat. # 13366, 13367 and 13370), and cultured on irradiated MEF feeder in ES medium containing Knockout DMEM (Life Technologies, Cat. # 10829-018), 15% ES-grade FBS (Omega scientific, Cat. # FB-01), 2mM L-glutamine (Life Technologies, Cat. # 25030-164), 1x10^{-4}M MEM non-essential amino acids (Life Technologies, Cat. # 11140-076), 1x10^{-4} M 2-mercaptoethanol (Sigma, Cat. # M3148) and 1X Penicillin-Streptomycin (Life Technologies, Cat. # 15140-122). Subsequently, single iPSC-like colonies were individually picked and expanded on irradiated MEF feeders to establish a stable line. At least three independent iPSC lines were generated for each genotype.

RNA-seq data analysis

RNA-seq reads were mapped to the GRCm38 (mm10) reference genome with TopHat to quantify gene and retrotransposon expression levels (36). 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) (Table S4). We used EdgeR to test for differential expression between miR-34a$^{-/-}$ and wild-type iPSCs (37). 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 (Supplementary Text; Table S2-S5).

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; Table S2-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. 3C) and yielded similar results (data now shown). 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 blastocyst embryos and chimeric mice from ESCs/iPSCs

To generate chimeric blastocysts by morula aggregation, we followed the method described by Nagy et. al. with minor modifications. One-cell stage, C57B6/J wild-type zygotes were collected at 0.5 day postcoitum (dpc), cultured in EmbryoMax KSOM Medium (Millipore, Cat. # MR-121-D) for 48h and only well-developed morula embryos were selected for aggregation. The ESCs or iPSCs were combined with morula embryos by sandwich method after removal of zona pellucida by acid Tyrode's solution (Sigma, Cat. # T1788) and then cultured overnight.

To generate chimeric blastocysts by microinjection, four ESCs or one ESC of the desired genotype were injected into E2.5 C57Bl/6N wild-type recipient morula embryos (16-32 cell

stage) and then cultured in vitro overnight to obtain the chimeric blastocysts. To generate chimeric mice by microinjection, 10-15 ESCs of the desired genotype were injected into E3.5 recipient blastocyst embryos before implanted into pseudopregnant females. Chimeric embryos were collected at E9.5 and E12.5 and subjected to IF staining. In both experiments, the recipient embryos for microinjection were generated from 4 week old, super-ovulated C57Bl/6N female mice. These female mice were first treated with intraperitoneal injection with 5 IU of PMSG (Sigma, pregnant mare serum gonadotropin, Cat. # G4877-1000U) and 5 IU of HCG (Sigma, human chorionic gonadotropin, Cat. # CG5-1VL) and then mated with male mice from the same strain. Subsequently, one-cell embryos were collected from those carrying vaginal plugs 24 hour after intraperitoneal injection of HCG. Collected embryos were cultured for 2 or 3 days in vitro in EmbryoMax KSOM Medium (Millipore, Cat. # MR-121-D); and properly developed morulae or blastocysts were selected for microinjection.

Preimplantation embryo expression analysis

Superovulated wild-type and miR-34a$^{-/-}$ females were mated with males of matching genotype to generate 2C, 8C and blastocyst embryos at E1.5, E2.5, and E3.5, respectively. Oocytes were collected at E0.5 from unmated, superovulated females. 2C and 8C embryos were recovered by oviduct flushing with DMEM (Thermo Fisher, Cat. #11995-040), while blastocysts were recovered by uterine flushing. Embryos were washed in PBS and subject to real-time PCR analyses using a Single Cell-to-Ct qRT-PCR kit (Life Technologies, Cat. # 4458236). All primers used are listed in Table S6.

Generation of teratomas from pluripotent stem cells and histological analyses

$1x10^6$ of WT or *miR-34a*$^{-/-}$ iPSCs or ESCs were injected into the dorsal flanks of 6-7 week old immune-deficient NCr-*nu/nu* female mice (Taconic, Cat# NCRNU). After 4-5 weeks, resulting teratomas were collected by surgical removal, fixed overnight in 10% buffered formalin (Fisher Scientific, Cat. # SF100-4), dehydrated in a graded series of ethanol solutions, embedded in Paraplast X-TRA paraffin (Fisher Scientific, Cat. # 23-021-401), sectioned at 6 μm thickness, mounted on glass slides, and stained with hematoxylin and eosin (H&E) using standard procedures (*16*). Additionally, the paraffin sections were subjected to IF and immunohistochemistry.

Immunofluorescence (IF) and Immunohistochemistry (IHC)

For IF staining of differentiated EBs or ESCs/iPSCs, samples were fixed with 4% paraformaldehyde for 10 min at room temperature and incubated with blocking solution (0.1%Triton X-100 and 5% normal goat serum in PBS) for 1 hour at room temperature. To detect the expression of pluripotent or lineage markers, EBs/cells were incubated overnight at 4°C with antibodies against MERVL-Gag (1:2000, a gift from T. Heidmann laboratory), Oct4 (1:100, Santa Cruz Biotechnology, Cat. # sc-5279), Gata4 (1:100, Santa Cruz Biotechnology, Cat. # sc-9053) or Cdx2 (1:100, Abcam, Cat. # ab76541 or #157524), followed by staining with goat anti-rabbit IgG (H+L) secondary antibody, Alexa Fluor 594-conjugated secondary antibody (1:500, Life Technologies, Cat. # A11037) for 1 hour at room temperature. EBs/cells were then stained with DAPI (300 nM, Sigma, Cat. # D9564) and subjected to imaging analyses using Laser scanning confocal microscopy (Zeiss LSM710) and Zeiss Observer.A1 microscope.

For IF staining of chimeric blastocysts, samples were fixed in 4% paraformaldehyde (PFA) for 20 min at room temperature and then transferred to phosphate-buffered saline (PBS) containing 0.1% bovine serum albumin (BSA). Embryos were permeabilized using 0.1% Triton X-100 in PBS containing 0.1% BSA for 5 min, and then blocked for 1 hour at room temperature in blocking solution (10% goat serum diluted in PBS/0.1% BSA). Blastocysts were then incubated with anti-Cdx2 primary antibody (1:100, Abcam, Cat. # ab76541) in blocking solution at 4°C overnight and stained with goat anti-rabbit, Alexa Fluor 594-conjugated secondary antibody (1:300, Life Technologies, Cat. # A11037) in blocking solution for 1 hour at room temperature. Blastocysts were then placed individually into chamber slides (Lab-Tek, Cat. # 155411) in 400 ml PBS/0.1%BSA solution. Images were taken using an Olympus Revolution XD spinning disk confocal microscope.

For IF staining of chimeric mouse embryos, including placentas and yolk sacs were fixed with 4% PFA for 2 hours, incubated in 30% sucrose for overnight in 4°C, embedded in Tissue-Tek O.C.T. compound (VWR, Cat. #25608-930), and cryo-sectioned at 8 mm. These sections were either directly visualized for GFP expression or subjected to IF using mouse anti-GFP (1:100, Abcam, Cat. # ab38689), rabbit anti-Tpbpa (1:200, Abcam, Cat. # ab104401), or rabbit anti-MTP1 (1:150, Alpha Diagnostic, Cat. # MTP11-A) primary antibodies. Trophoblast giant cells were identified based on their location in the placenta and the morphology of enlarged nuclei. Spongiotrophoblasts were identified based on the staining of the molecular marker, Tpbpa. The bilaminar structure of the yolk sac is identified by DAPI staining, and the visceral endoderm part is identified by its columnar, epithelial morphology. The GFP signals in three embryonic germ layers of chimeric mouse embryos (E12.5) and yolk sacs were directly visualized without staining.

For immunohistochemistry (IHC) analyses on teratomas or placentas of chimeric embryos , 5 mm paraffin sections were deparaffinized, dehydrated, and subjected to heat-induced antigen retrieval in a pressure cooker using Target Retrieval solution (DAKO, Cat. # S1699). Slides were incubated for 10 minutes with 3% H_2O_2, blocked for 3 hours with PBS containing 5% BSA and 0.3% Triton X-100, and incubated with primary antibodies against PL-1 (1:75, Santa Cruz Biotechnology, Cat. # sc-34713) or GFP (1: 100, Abcam, Cat. # ab38689) overnight in PBS buffer containing 1% BSA and 0.3% Triton X-100. Slides were then incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies for 2 hours at room temperature, and then subjected to 3,3'-Diaminobenzidine (DAB) staining (Life Technologies, Cat. # 00-2014) followed by a counterstain with Mayer's hematoxylin (Electron Microscopy Sciences, Cat. # 26503-04). The sinusoidal trophoblast giant cells (s-TGCs) were identified by their enlarged nuclei and adjacent location in the maternal blood sinusoid space.

Chromatin immunoprecipitation (ChIP)

For each ChIP experiment, 10^6 ESCs or iPSCs were fixed with 1% formaldehyde (VWR, Cat. # 5016-02) to extract chromatin for immunoprecipitation. Nuclei were isolated by Farnham lysis buffer (5 mM PIPES pH 8.0, 85 mM KCl, 0.5% NP-40 with Protease Inhibitor Cocktail Tablet (Roche, Cat. # 11836153001) and lysed in nuclear lysis buffer (50 mM Tris pH 8.0, 10 mM EDTA, and 1% SDS with the protease inhibitor cocktail). Chromatin was fragmented by a Covaris S220 Focused ultrasonicator (peak power 175, duty factor 20, cycles/burst 200, duration

30s, with 35 treatments) and diluted in RIPA buffer (10 mM Tris pH 7.6, 1 mM EDTA, 0.1% sodium deoxycholate, and 1% Triton X-100 with protease inhibitor cocktail) in a 1:9 ratio. The pull-down was performed at 4°C overnight using 40 ml Dynabeads protein A (Life Technologies, Cat. # 10001D) and 2 mg antibodies against Gata2 (Santa Cruz, Cat. # sc-9008), H3K4Me (Abcam, Cat. # ab8895), H3K4Me3 (EMD Millipore, Cat. # 17-614), H3K9Me2 (Abcam, Cat. # ab1220), and H3 (Abcam, Cat. #1791). Washes were performed twice with the low-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 150 mM NaCl), three times with the high-salt wash buffer (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl pH 8.0, 500 mM NaCl), and four times with the LiCl wash buffer (0.25 M LiCl, 1% IGEPAL CA630, 1% sodium deoxycholate 1 mM EDTA, 10 mM Tris pH 8.0). Chromatin-immunoprecipitated DNA was analyzed by qPCR using SYBR FAST qPCR Master mix (KAPA biosystems, Cat. # KK4602) and a 7900HT Fast Real-Time PCR machine (Applied Biosystems). Percent input of immunoprecipitated samples was calculated according to the real-time PCR results of serially-diluted lysates. The enrichment of each individual locus is calculated as (the percentage of input of modified histone H3)/(the percentage of input of histone H3). The following primers were used for qPCR analysis: mervl forward, 5'-CTTCCATTCACAGCTGCGAC-3'; mervl reverse 5'-CTAGAACCACTCCTGGTACC-3'; iap forward; 5'-GCTCCTGAAGATGTAAGCAATAAAG-3'; iap reverse, 5'-CTTCCTTGCGCCAGTCCCGAG-3'; mmervk10c forward, 5'-TTCGCCTCTGCAATCAAGCTCTC; mmervk10c reverse, 5'-TCGCTCRTGCCTGAAGATGTTTC-3'; tcstv1 forward, 5'-ATCTACTTGGGGTGCCTGGT-3'; tcstv1 reverse, 5'-GAAGACCAGCTGAACCATCC-3'.

Luciferase Assays

For MERVL-luciferase reporter assays, we used pGL3 luciferase reporter vectors (Promega, Cat. # E1751) harboring $MERVL_{1\text{-}1000}$, $MERVL_{1\text{-}493}$, and $MERVL_{500\text{-}1000}$ fragments, as described by Macfarlan et. al. (25). The $MERVL_{125\text{-}375}$-Luc reporter was constructed by truncating the $MERVL_{1\text{-}493}$ fragment using a QuikChange Site-Directed Mutagenesis Kit (Strategene, Cat. # 200518). The two fully conserved Gata2 binding sites (BS1 and BS3) were ablated in the $MERVL_{125\text{-}375}$-Luc reporter construct, either individually or in combination, using a QuikChange Site-Directed Mutagenesis Kit. The following primers were designed for mutagenesis: 1-375 Forward, TGGACTTCCATTCACCTCGAGATCTGCGATCTAAGTAAGC; 1-375 Reverse, ATCGCAGATCTCGAGGTGAATGGAAGTCCAAGGATCTAGC; 125-375 Forward, CTTACGCGTGCTAGCGATCTTGAGCCATAGTGGCTATGGA; 125-375 Reverse, CTATGGCTCAAGATCGCTAGCACGCGTAAGAGCTCGGTAC; Gata2ΔBS1 Forward, TCTCCGAGTTTAAGGAACACACCTTTGGGCTACGCCTTTC; Gata2ΔBS1 Reverse, AATCCCAGATGAAAGGCGTAGCCCAAAGGTGTGTTCCTTA; Gata2ΔBS3 Forward, TTAAAGGTGTGGTGGAACACACCTTTGGGCTACACCTTCT; and Gata2ΔBS3 Reverse, TGTCTCCAGCAGAAGGTGTAGCCCAAAGGTGTGTTCCACC. MERVL-Luc reporters and control Renilla luciferase reporter pRL-TK (Promega, Cat. # E2241) were co-transfected (600 ng and 150 ng per well of a 12-well plate, respectively) using Lipofectamine 2000 reagent (Life Technologies, Cat. # 11668027) into ESCs. Transfection complexes containing the reporter constructs were prepared in Opti-MEM Reduced-Serum Medium (Life Technologies, Cat. # 31985062) according to manufacturer's instructions. After trypsinization with 0.25% Trypsin + EDTA (Life Technologies, Cat. # 25200-056), 100,000 cells were resuspended in ES media lacking Pen Strep, incubated with transfection complexes for 10 minutes at 37° C, and then

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

For gata2 3'UTR luciferase assays, a fragment that includes the 3' portion of the ORF and the entire Gata2 3'UTR was amplified by PCR. The fragment was then cloned into a psiCheck-2 vector (Promega, Cat. # C8021) to generate the gata2 3'UTR-Luc reporter.miR-34a binding site mutants were generated using a QuikChange Site-Directed Mutagenesis Kit. The following primers were used: Gata2 3'UTR F XhoI, CTCGAGAGTCTCTCTTTTGGCCACCC; Gata2 3'UTR R NotI, GCGGCCGCCAAGGCCACCTGACAGCTTA; Gata2 3'UTRΔ34aBS F, CCGTCCAGCATGGTGATGGGCTAGGCAAGCCTCCCACTGG; Gata2 3'UTRΔ34aBS R, GCTTGCCTAGCCCATCACCATGCTGGACGGGTGGGGGTGG; Gata2 3'UTRΔ34aBS2 F AGAGACCCACTTCCTGCCTAGCCTGGCCGAAGCCACCTCT; Gata2 3'UTRΔ34aBS2 R, TCGGCCAGGCTAGGCAGGAAGTGGGTCTCTTGGGATGGGC; Gata2 3'UTRΔ34aBS3 F, CTTCTTTGGGACCTCCCAGTCAGGGCTCTCGGGGGCAGAC; Gata2 3'UTRΔ34aBS3 R, GAGAGCCCTGACTGGGAGGTCCCAAAGAAGGACCCCAAGA. The gata2 3'UTR-Luc reporters (2 ng per well of 12-well plate) were co-transfected with 100 nM siGFP or mature miR-34a RNA mimics using Lipofectamine 2000 reagent (Life Technologies, Cat. #11668027) into a feeder-free mouse ESC line (39). After 48 hours, cells were lysed and assayed for luciferase activity by Dual-Luciferase® Reporter Assay System (Promega, Cat. # E1910) using a Glomax 20/20 Luminometer (Promega).

Transfection and retrovirus/lentivirus transduction

To overexpress miR-34a in wild-type and miR-34a$^{-/-}$ iPSCs, cells were infected with MSCV (murine stem cell virus) retrovirus that encoded a LTR-miR-34a and a PGK-puromycin-IRES-GFP cassette (40). MSCV and MSCV-miR-34a transduced iPSCs were selected with 3 mg/ml puromycin for two days before collected for real-time PCR analyses and western blotting.

The ESCs and iPSCs used for microinjection were labeled by green fluorescence protein (GFP) using the PiggyBac-GFP plasmid. The PiggyBac vector contains an EF1a-driven GFP expression cassette and an ubiquitin-puromycin selection marker. The PiggyBac-GFP plasmid was mixed with the PiggyBac transposase plasmid in a 1:1 ratio (41), and subsequently transfected into ESCs or iPSCs using Lipofectamine 2000 (Life Technologies, Cat. # 12566014). Cells were selected with 3 mg/ml puromycin for two days and cultured in puromycin-free ES medium for following analyses.

To knock down gata2 by RNAi, two Gata2 shRNAs were cloned into pLKO.1 lentiviral vector (Addgene, #10878) using the following oligos (shgata2#1 sense: 5'-CCGGGAGGTGGATGTCTTCTTCAACCACTCGAGTGGTTGAAGAAGACATCCACCTCTTTTTG-3'; shGata2#1 antisense: 5'-AATTCAAAAAGAGGTGGATGTCTTCTTCAACCACTCGAGTGGTTGAAGAAGACATCCACCTC-3'; shGata2#2 sense: 5'-CCGGGGACGAGGTGGATGTCTTCTTCAACTCGAGTTGAAGAAGACATCCACCTCGTCCTTTTTG-3'; shGata2#2 antisense: 5'-AATTCAAAAAGGACGAGGTGGATGTCTTCTTCAACTCGAGTTGAAGAAGACATCCA

CCTCGTCC-3') (42); and the corresponding lentiviruses were produced by co-transfecting pLKO.1 shRNA vectors with pMD2.G and psPAX2 to HEK293T cells. After infection, iPSCs were selected in 3 mg/ml puromycin for two days and expanded for in vitro and in vivo analyses.

Western blotting

For ESC or iPSC collection, trypsinized cells were plated on a gelatin-coated plate for 1 hour to remove feeders. Cells separated from the feeders were then lysed in Laemmli sample buffer (60 mM Tris-Cl pH 6.8, 2% SDS, 100 mM DTT, 10% glycerol, 0.02% bromophenol blue) and subjected to western analyses. Antibodies against mouse Gata2 (Santa Cruz Biotechnology, Cat # CG2-96) was used at 1:500 dilution, and α-tubulin (Sigma, clone B-5-1-2) was used at a 1:4,000 dilution as a loading control. The quantitation of all western analyses was carried out with ImageJ (NIH).

Bisulfite sequencing

Genomic DNAs were purified by the standard phenol/chloroform method. 2 mg DNA was subjected to bisulfite conversion and subsequent purification using the EZ DNA Methylation-Gold Kit (Zymo Research, Cat. # D5005). Bisulfite-treated DNAs were amplified using Jumpstart REDTaq Readymix (Sigma, Cat. # P1107) with the following primers: MERVL forward, ATATGAATAAAGTGGTTATGGTGGT; MERVL reverse, AATTCCTAAACCCATAAATCCTAAC; IAP forward, TTGATAGTTGTGTTTTAAGTGG; IAP reverse, AAAACACCACAAACCAAAATC. The amplified DNA fragments were cloned to pGEM-T Easy vector (Promega, Cat. # A1360) for sequencing. The methylation patterns were analyzed by QUMA (Quantification tool for methylation analysis, http://quma.cdb.riken.jp).

Chapter 3

Results

miR-34a$^{-/-}$ pluripotent stem cells exhibit expanded cell fate potential

microRNAs (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 [69,71,75]. These small non-coding RNAs are increasingly recognized as key regulators of cell fate specification in normal development and in pluripotent stem cells [77,78].

Initially identified as bona fide p53 transcriptional targets in tumor suppression, the miR-34 miRNAs (miR-34a, miR-34b and miR-34c), particularly miR-34a, have been previously characterized as a key barrier for somatic reprogramming [82]. miR-34a deficiency significantly enhances the efficiency of iPSC generation [82], producing iPSCs with normal self-renewal and pluripotency (Fig. S1A, S1B and S1C). Surprisingly however, teratomas generated from miR-34a$^{-/-}$ iPSCs, but not 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 constitutes the majority of expressed miR-34 miRNAs (Fig. S1D). Similarly, miR-34a$^{-/-}$ ESC derived teratomas, but not the wild-type controls, also contained areas reminiscent of extra-embryonic placental cell lineages (Fig.1A) and exhibited an induction of trophectoderm (TE) markers (Fig. S1E), including cdx2, elf5, psx1, fgfr2, egfr and mdfi [83,84]. While we did not identify any areas morphologically resembling the visceral endoderm of the yolk sac, we detected a strong induction of primitive endoderm (PE) markers (gata4, gata6 and sox17) in miR-34a$^{-/-}$ teratomas, but not in wild-type controls (Fig. S1E). These findings suggest that miR-34a$^{-/-}$ pluripotent stem cells likely differentiate towards both embryonic and extra-embryonic cell lineages during teratoma formation.

The expanded potential of miR-34a$^{-/-}$ ESCs/iPSCs is also evident upon embryoid body (EB) differentiation (Fig. 1B and 1C). While markers from all three germ layers were similarly induced in wild-type and miR-34a$^{-/-}$ EBs, significant upregulation of TE markers (cdx2, elf5, esx1, tfap2c and gata3) [83,85,86] was observed only in miR-34a$^{-/-}$ EBs (Fig. 1B, 1C and S1F). Immunofluorescence (IF) staining confirmed that a significant percentage of miR-34a$^{-/-}$ EBs was Cdx2 positive (Fig. 1B), and these Cdx2-positive cells preferably localized to the periphery (Fig. 1B, S1G and S1H). Additionally, the extra-embryonic endoderm marker gata4 and pdgfra, as well as the trophoblast lineage marker mash2 (ascl2) and pl1 (prl3d1), were also induced in miR-34a$^{-/-}$ EBs (Fig. S1F). Thus, upon EB differentiation, miR-34a$^{-/-}$ ESCs exhibited expanded cell fate potential, generating cells with molecular features characteristic of both embryonic and extra-embryonic lineages.

To define the cell fate potential of miR-34a$^{-/-}$ pluripotent stem cells in normal development, we traced their lineage in chimeric blastocysts following microinjection or aggregation with recipient morulae. Initially, four GFP-labeled wild-type or miR-34a$^{-/-}$ ESCs were injected into each C57BL/6N recipient morula to generate chimeric blastocysts (Fig. 1D). While wild-type ESCs exclusively gave rise to cells localized to the ICM (Fig. 1D; Table S1), miR-34a$^{-/-}$ ESC

progenies localized to both ICM and TE in ~60% of chimeric blastocysts (Fig. 1D; Table S1). This expanded cell fate potential is unlikely due to extra-embryonic contamination during miR-34a$^{-/-}$ ESC derivation, as miR-34a$^{-/-}$ iPSCs derived from mouse embryonic fibroblasts (MEFs) phenocopied miR-34a$^{-/-}$ ESCs in their developmental potential. When aggregated with recipient C57BL/6J morulae, miR-34a$^{-/-}$ ESCs and miR-34a$^{-/-}$ iPSCs colonize both ICM and TE of chimeric blastocysts, while passage- and littermate-controlled wild-type ESCs and iPSCs exclusively colonized the ICM (Fig. S1I).

The expanded cell fate potential of miR-34a$^{-/-}$ ESCs in chimeric blastocysts could be due to the presence of cells with bidirectional potential; alternatively, miR-34a$^{-/-}$ ESCs could contain a heterogeneous population of cells that preferentially differentiate into embryonic or extra-embryonic cell lineages. To distinguish between these two possibilities, we injected single, GFP-labeled miR-34a$^{-/-}$ ESCs into each recipient morula to generate chimeric blastocysts (Fig. 1E). In 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. 1E; Table S1) respectively, suggesting that a significant portion of miR-34a$^{-/-}$ ESCs exhibit a bidirectional developmental potential at the single-cell level.

We then generated chimeric embryos by microinjecting 10-15 GFP-labeled wild-type or miR-34a$^{-/-}$ ESCs into C57BL/6N recipient blastocysts. While wild-type ESCs contributed exclusively to lineages of the three embryonic germ layers, miR-34a$^{-/-}$ ESCs contributed to both embryonic and extra-embryonic cell lineages in E9.5, E12.5 and E14.5 chimeric embryos (Fig. 1F, 1G and S1J; Table S1). In particular, we observed clusters of GFP-positive, miR-34a$^{-/-}$ ESC progenies in the visceral endoderm of the yolk sac, as well as in multiple extra-embryonic trophoblast lineages of the placenta (trophoblast giant cells, spongiotrophoblasts, syncytiotrophoblasts (STBs) and sinusoidal trophoblast giant cells (s-TGCs), Fig. 1F, 1G; Table S1). In these chimeric embryos, the number of GFP-positive cells in extra-embryonic cell lineages greatly surpasses the number of miR-34a$^{-/-}$ ESCs injected (Fig. 1F and 1G), suggesting that injected miR-34a$^{-/-}$ ESCs had undergone substantial proliferation before committing to multiple terminally differentiated extra-embryonic lineages.

miR-34a$^{-/-}$ pluripotent stem cells exhibit an induction of MERVL ERVs.

To investigate the molecular basis for the bidirectional 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 (ncRNAs), pseudogenes, antisense transcripts, and retrotransposons using 100 bp paired end RNA-seq data (Fig. 2A). Given the repetitive nature of retrotransposons, we quantified retrotransposon expression at the family level using both uniquely and non-uniquely mapped reads (Supplemental Information S1). Surprisingly, the most highly expressed and differentially regulated transcript in miR-34a$^{-/-}$ iPSCs was transcribed from the MERVL family of ERVs (Fig. 2A and S2A), which were also highly induced in totipotent 2C blastomeres and reported ESCs with expanded potential [58,66,67,87]. ERV induction in miR-34a$^{-/-}$ ESCs/iPSCs was largely specific

to the MERVL family (Fig. 2A, 2B and S2A; Table S2). The majority of differentially expressed retrotransposons in miR-34a$^{-/-}$ iPSCs belonged to the canonical MERVL family of ERVs (a class-III ERV) (Fig. S2A; Table S2); a small fraction of differentially expressed loci belonged to the MT2A, MT2B, MT2B1, and MT2B2 ERV families that are highly related to the canonical MERVL solo LTR, MT2_Mm (Fig. S2A; Table S3).

Consistent with our RNA-seq results, we invariably detected a significant increase of MERVL expression in miR-34a$^{-/-}$ iPSCs and ESCs, using real-time PCR primer pairs designed from multiple highly conserved MERVL regions (Fig. 2B and 2C; data not shown). Interestingly, while MERVL induction in miR-34a$^{-/-}$ iPSCs persisted for more than 27 passages (Fig. S2B), MERVL was only induced in early passages of 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. The expanded cell fate potential of miR-34a$^{-/-}$ ESCs was highly correlated with the strong MERVL induction, as the late passage (passage 17) miR-34a$^{-/-}$ ESCs lost both MERVL induction and the bidirectional potential (Fig. S2B and S2C).

The MERVL ERVs have been retained throughout mammalian evolution, with independent expansion in the murine and primate genomes [52]. There are 2502 loci in the C57B6/J mouse genome, ~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. 2C) [52]. Another 32% of MERVL loci exhibit truncated retroviral structure, missing one or both LTRs (Fig. 2C). The remaining 41% of MERVL loci have undergone homologous recombination, yielding solo LTRs (MT2_Mm) with varying degrees of sequence degeneration (Fig. 2C). We obtained bioinformatic estimates of locus-specific MERVL expression in wild-type and miR-34a$^{-/-}$ iPSCs using our RNA-seq data (Table S3; Supplemental Information). Notably, definitive evidence for MERVL reactivation in miR-34a$^{-/-}$ iPSCs was observed predominantly for loci harboring MERVLs with a complete retroviral structure, but not for those with truncated structure (Fig. 2D and S2D; 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; Table S3).

Approximately 300 MERVL loci still encode intact Gag viral protein [67]. 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. 2E and 2F). Interestingly, miR-34a$^{-/-}$ ESCs and iPSCs were heterogeneous populations, containing ~12% and ~20% MERVL-Gag-positive cells, respectively, in otherwise Oct4-positive colonies (Fig. 2E, 2F, S2E and S2G). Consistent with this observation, a fraction of individual miR-34a$^{-/-}$ iPSC colonies exhibited a significantly greater MERVL induction than the bulk population (Fig. S2F), suggesting that the extent of MERVL induction in individual cells was largely underestimated using the bulk population. In miR-34a$^{-/-}$ ESCs and iPSCs, the expression of MERVL-Gag and Oct4 were mutually exclusive (Fig. 2E, S2E and S2G), suggesting that the MERVL-positive, Oct4 negative miR-34a$^{-/-}$ cells possess a unique state of developmental potency, distinct from that of classic pluripotent stem cells characterized by Oct4 expression [66,67,88].

The global protein-coding gene expression profiles of miR-34a$^{-/-}$ iPSCs resemble those of reported bi-potential ESCs in hierarchical clustering (Fig. S3A). Intriguingly, among the most differentially expressed protein-coding genes in miR-34a$^{-/-}$ iPSCs were those proximal to MERVL loci (Fig. 2G). Indeed, differential expression analysis between reported ESCs with bidirectional cell fate potential (lsd1$^{-/-}$, 2C+, p60 knockdown and p150 knockdown ESCs) and their pluripotent controls revealed the induction of MERVL proximal genes as a key feature (Fig. S3B) [66,67,81]. The MERVL derepression in miR-34a$^{-/-}$ iPSCs and ESCs also correlates with the induction of many protein-coding genes that harbor either a proximal upstream MERVL or an intronic MERVL on the same strand (Fig. 2G; Fig. S3C). In many cases, MERVL or related loci, particularly solo LTRs, act as alternative promoters, generating chimeric transcripts of proximal genes that differ in 5'-UTRs (Fig. 2H, S3D and S3E; Table S4 and S5). These MERVL-gene chimeric transcripts can be unambiguously identified by the 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. 2H; Fig. S3D and S3E). The induction level of MERVL-gene isoforms varied among individual miR-34a$^{-/-}$ iPSC colonies, and largely correlated with the extent of MERVL induction (Fig. S3F). Consistently, miR-34a overexpression in miR-34a$^{-/-}$ iPSCs not only decreased MERVL expression, but also reduced the level of the MERVL-driven chimeric transcript (Fig. 2I).

We repeated our analysis using 150 bp pair-end RNA-seq data to gain confidence in the accuracy of our retrotransposon mapping (Fig. S4A). The analysis using longer sequence reads confirmed all our observations (Fig. S4A and S4B).

MERVL induction in miR-34a$^{-/-}$ pluripotent stem cells is regulated transcriptionally

Given the correlation between MERVL induction and bi-potential pluripotent stem cells (Fig. S3A and S3B)[66,67], the molecular pathway that mediates miR-34a-dependent MERVL repression could also regulate miR-34a-dependent restriction on pluripotent potential. To determine the key sequences required for MERVL induction, 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 [81]. This reporter showed an elevated luciferase activity in miR-34a$^{-/-}$ ESCs compared to wild-type ESCs, faithfully recapitulating the endogenous MERVL induction (Fig. 3A). The LTR sequence was both necessary and sufficient for the MERVL$_{1\text{-}1000}$-Luc reporter activity in miR-34a$^{-/-}$ ESCs (Fig. 3A); furthermore, a minimal fragment, MERVL$_{125\text{-}375}$, containing a direct repeat and a TATA box, was sufficient to drive strong luciferase activity specifically in miR-34a$^{-/-}$ ESCs (Fig. 3A). The MERVL$_{125\text{-}375}$ fragment is highly conserved among all highly induced MERVL loci in miR-34a$^{-/-}$ iPSCs (Fig. S5A), suggesting a sequence-dependent transcriptional mechanism for MERVL induction. Consistently, miR-34a$^{-/-}$ ESCs and iPSCs also exhibited H3K4Me3 enrichment near the LTR of MERVL retrotransposons and the MERVL LTR proximal to tcstv1, but not near other ERVs such as IAP or MMERK10C (Fig.

3B). Thus, MERVL loci are specially enriched for active transcription machinery in miR-34a$^{-/-}$ pluripotent stem cells.

Gata2 mediates elevated MERVL expression in miR-34a$^{-/-}$ pluripotent stem cells.

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 predicted 70 candidate transcription factors that 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 (Fig. S5A; ref. 52, 53). Gata2 is also reported to play an important role in cell fate potency of ESCs [89].

We aligned the LTR sequences from 18 MERVL loci that were strongly induced in miR-34a$^{-/-}$ iPSCs, most of which contain three predicted Gata2 binding sites (Fig. S5A). Mutating the two most conserved sites within the MERVL$_{125\text{-}375}$-Luc reporter significantly reduced its activity in miR-34a$^{-/-}$ pluripotent stem cells (Fig. 3C). Similarly, gata2 knockdown in miR-34a$^{-/-}$ pluripotent stem cells effectively abolished the induction of MERVL and MERVL proximal genes (zfp352, tmem132c and chit1) in cell culture (Fig. 3D; Fig. S5C), and significantly decreased MERVL and cdx2 induction during teratoma formation (Fig. S5D). Gata2 knockdown also reduced H3K4Me3 deposition on MERVL elements and on the MERVL proximal gene tcstv1, suggesting a specific decrease in active transcriptional machinery on MERVL loci (Fig. 3E). Finally, using chromatin immunoprecipitation (ChIP), we demonstrated specific binding of Gata2 to the LTR region of MERVL, and not to the MERVL internal region or to other ERVs such as IAP or MMERK10C (Fig. 3F). Taken together, Gata2 plays an essential role in directly promoting the induction of MERVL ERVs and MERVL proximal genes in miR-34a$^{-/-}$ pluripotent stem cells.

Epigenetic modifications constitute another possible mechanism for MERVL induction in miR-34a$^{-/-}$ pluripotent stem cells. We investigated the role of DNA methylation on MERVL induction, but no difference was detected between wild-type and miR-34a$^{-/-}$ iPSCs (Fig. S5E), consistent with intact MERVL silencing in dnmt3a$^{-/-}$; dnmet3b$^{-/-}$ ESCs/iPSCs (Fig. S5F). MERVL was previously shown to be induced by a global decrease in H3K9Me2 or a global increase in H3K4Me1 in g9a/glp$^{-/-}$ ESCs or lsd1$^{-/-}$ ESCs, respectively [81,90]. Using ChIP, we found that H3K27Ac and H3K9Me2 deposition on MERVL was unaltered in miR-34a$^{-/-}$ pluripotent stem cells (Fig. S5G and S5H). Additionally, while miR-34a$^{-/-}$ ESCs and iPSCs exhibited a ~2-3 fold H3K4Me1 enrichment on MERVL loci (Fig. S5I), miR-34a overexpression silenced MERVL expression in wildtype and lsd1$^{-/-}$ ESCs with a similar efficiency (Fig. S5J), suggesting that the direct mechanism through which miR-34a silenced MERVL was likely independent of a global alteration of H3K4Me1. Hence, none of the tested epigenetic mechanisms appeared to be essential mechanisms for miR-34a-mediated MERVL repression in ESCs/iPSCs.

miR-34a restricts cell fate potential of pluripotent stem cells by directly repressing Gata2.

Gata2 not only plays an essential role in mediating MERVL activation in miR-34a$^{-/-}$ ESCs/iPSCs (Fig. 3D), it also emerges as a strong candidate as a direct miR-34a target. gata2 harbors three predicted miR-34a binding sites [91,92] (Fig. 4A), and exhibited miR-34a-dependent regulation in pluripotent stem cells—gata2 mRNA and protein are increased in miR-34a$^{-/-}$ pluripotent stem cells and reduced upon ectopic miR-34a overexpression (Fig. 4B and 4C). Similarly, a luciferase reporter containing a fragment of the gata2 gene with all three predicted miR-34a binding sites exhibited miR-34a-dependent repression in luciferase assays (Fig. 4D). Mutating all three predicted miR-34a binding sites in this reporter abolished this miR-34a-dependent regulation (Fig. 4D). These findings suggest that gata2 is a direct miR-34a target in ESCs/iPSCs that mediates MERVL regulation. Consistent with Gata2 derepression in miR-34a$^{-/-}$ ESCs/iPSCs, a number of previously characterized Gata2 targets, including dab2, cd34 and gata1 [89,93,94], exhibited a Gata2-dependent upregulation in miR-34a$^{-/-}$ iPSCs (Fig. S6A and S6B).

Knockdown of gata2 in miR-34a$^{-/-}$ iPSCs phenocopies miR-34a overexpression, not only downregulating the expression of MERVL and MERVL-proximal genes (Fig. 3D and S5C), but also abolishing their bidirectional developmental potential to differentiate into both embryonic and extra-embryonic lineages (Fig. 4E and 4F). In 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. 4E and 4F). More importantly, while the control infected miR-34a$^{-/-}$ iPSCs contributed to both ICM and TE in 53% of the chimeric blastocysts (n=8/15, Fig. 4G; Table S1), miR-34a$^{-/-}$ iPSCs with gata2 knockdown lost this expanded cell fate potential (n=0/13, Fig. 4G; Table S1). To our knowledge, miR-34a is the first non-coding RNA known to restrict pluripotent cell fate potential in ESCs/iPSCs. While multiple miR-34a targets could 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. 4H).

Chapter 4
Conclusions

Mouse zygotes and early blastomeres possess a totipotent cell fate potential, generating both embryonic and extra-embryonic cell types during normal development. This totipotent cell fate potential is gradually restricted during preimplantation development. By the late blastocyst stage, the separation of TE (which develops into the placenta to support embryonic development), epiblast (which forms the embryo proper) and PE (which develops into the yolk sac) signals the completion of the first cell fate specification event, which commits the developmental potential of cells to either embryonic or extra-embryonic lineages. ESCs and iPSCs in culture faithfully recapitulate the developmentally restricted, pluripotent cell fate potential of the epiblast, efficiently contributing to all embryonic cell lineages in vivo, but rarely to extra-embryonic lineages [4]. Experimentally, ESCs can be induced into cells with bidirectional developmental potential, using somatic nuclear transfer, genetic modifications, and specific ESC/iPSC enrichment/culture/derivation procedures [64–68,83]. The low efficiency of such experimental manipulations reflects the existence of multiple cellular and molecular impediments for the acquisition of a bidirectional 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 for somatic reprogramming [82]. To our knowledge, miR-34a is the first non-coding RNA whose deficiency in ESCs or iPSCs expands their developmental potential, generating a significant fraction of cells with bidirectional developmental potential. It is intriguing that miR-34a deficient pluripotent stem cells also induce the expression of MERVL ERVs; and this MERVL induction in ESCs/iPSCs appears correlated with their expanded cell fate potential in a number of experimental systems (Fig. 4H). MERVL induction could simply be an indicator of an unique 2C-like transcriptional and epigenetic state; alternatively, MERVL could functionally contribute to the establishment and/or maintenance of the 2C-like cell fate potential by rewiring gene regulatory networks to induce many 2C-specific, MERVL-driven genes [66,67]. An interesting parallel can be drawn to human ESCs, wherein fluctuating levels HERV-H family ERVs marks a dynamic population of naive-state pluripotency [95]. 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.

While miR-34a deficiency expands cell fate potential in pluripotent stem cells, miR-34a$^{-/-}$ mice undergo normal preimplantation development in laboratory conditions [82,96]. It is conceivable that the miR-34 family miRNAs act redundantly with other mechanisms to repress MERVL expression in preimplantation development. It is also possible that an impaired totipotency to pluripotency transition in miR-34a$^{-/-}$ embryos is tolerated in mouse preimplantation development due to its considerable cell fate plasticity [7,97]. Nevertheless, miR-34a$^{-/-}$ pluripotent stem cells constitute a powerful experimental system to investigate the molecular basis underlying the developmental potential to both embryonic and extra-embryonic linages. Our studies provides vital insights into an intricate network of protein-coding genes, ncRNAs, and retrotransposons that act cooperatively to define cell fate plasticity and cell fate potential in pluripotent stem cells.

References

1. Evans, M.J & Kaufman,M.H Establishment in culture of pluripotential cells from mouse embryos. Nature. 292, 154-156.pdf.
2. Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells Developmental Biology : **78,** 7634–7638 (1981).
3. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–76 (2006).
4. 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).
5. Tam, P. P. L. & Rossant, J. Mouse embryonic chimeras: tools for studying mammalian development. *Development* **130,** 6155–63 (2003).
6. Papaioannou, V. E., Mkandawire, J. & Biggers, J. D. Development and phenotypic variability of genetically identical half mouse embryos. *Development* **106,** 817–27 (1989).
7. Karlson, P. & Luscher, M. 1959 Nature Publishing Group. *Nature* **183,** 55–56 (1959).
8. Wernig, M. *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448,** 318–24 (2007).
9. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313–7 (2007).
10. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861–72 (2007).
11. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917–20 (2007).
12. Park, I.-H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451,** 141–6 (2008).
13. Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nat. Biotechnol.* **25,** 1177–81 (2007).
14. Huangfu, D. *et al.* Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nat. Biotechnol.* **26,** 1269–75 (2008).
15. Stadtfeld, M. & Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. *Genes Dev.* **24,** 2239–2263 (2010).
16. Krizhanovsky, V. & Lowe, S. W. NEWS & VIEWS The promises and perils of p53. *Nature* **460,** 4–5 (2009).
17. Li, H. *et al.* The Ink4/Arf locus is a barrier for iPS cell reprogramming. *Nature* **460,** 1136–9 (2009).
18. Marión, R. M. *et al.* A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity. *Nature* **460,** 1149–53 (2009).
19. Utikal, J. *et al.* Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. *Nature* **460,** 1145–8 (2009).
20. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* (2009). doi:10.1038/nature08592
21. Hong, H. *et al.* Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. *Nature* **460,** 1132–5 (2009).
22. Kawamura, T. *et al.* Linking the p53 tumour suppressor pathway to somatic cell reprogramming. *Nature* **460,** 1140–4 (2009).
23. Zhao, Y. *et al.* Two Supporting Factors Greatly Improve the Efficiency of Human iPSC

Generation. *Cell Stem Cell* **3,** 475–479 (2008).
24. Riley, T., Sontag, E., Chen, P. & Levine, A. Transcriptional control of human p53-regulated genes. *Nat. Rev. Mol. Cell Biol.* **9,** 402–412 (2008).
25. Banito, A. *et al.* Senescence impairs successful reprogramming to pluripotent stem cells. *Genes Dev.* **23,** 2134–9 (2009).
26. McClintock, B. The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci.* **36,** 344–355 (1950).
27. Ravindran, S. Barbara McClintock and the discovery of jumping genes. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 20198–9 (2012).
28. Fedoroff, N. V. McClintock's challenge in the 21st century. *Proc. Natl. Acad. Sci. U. S. A.* **109,** 20200–3 (2012).
29. Fedoroff, N. V. Transposable Elements , Epigenetics , and Genome Evolution. **338,** (2012).
30. Koito, A. & Ikeda, T. Intrinsic immunity against retrotransposons by APOBEC cytidine deaminases. *Front. Microbiol.* **4,** 28 (2013).
31. Elbarbary, R. A., Lucas, B. A. & Maquat, L. E. Retrotransposons as regulators of gene expression. *Science* **351,** aac7247 (2016).
32. Boeke, J. D. & Chapman, K. B. Retrotransposition mechanisms. *Curr. Opin. Cell Biol.* **3,** 502–507 (1991).
33. Blomberg, J., Benachenhou, F., Blikstad, V., Sperber, G. & Mayer, J. Classification and nomenclature of endogenous retroviral sequences (ERVs): problems and recommendations. *Gene* **448,** 115–23 (2009).
34. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
35. Stocking, C. & Kozak, C. a. Murine endogenous retroviruses. *Cell. Mol. Life Sci.* **65,** 3383–98 (2008).
36. Maksakova, I. a *et al.* Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.* **2,** e2 (2006).
37. Alamgir, A. S. M., Owens, N., Lavignon, M., Malik, F. & Evans, L. H. Precise Identification of Endogenous Proviruses of NFS / N Mice Participating in Recombination with Moloney Ecotropic Murine Leukemia Virus (MuLV) To Generate Polytropic MuLVs Precise Identification of Endogenous Proviruses of NFS / N Mice Participating. (2005). doi:10.1128/JVI.79.8.4664
38. Belshaw, R. *et al.* Long-term reinfection of the human genome by endogenous retroviruses. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 4894–9 (2004).
39. Stoye, J. P. Studies of endogenous retroviruses reveal a continuing evolutionary saga. *Nat. Rev. Microbiol.* **10,** 395–406 (2012).
40. Zhang, Y., Maksakova, I. a, Gagnier, L., van de Lagemaat, L. N. & Mager, D. L. Genome-wide assessments reveal extremely high levels of polymorphism of two active families of mouse endogenous retroviral elements. *PLoS Genet.* **4,** e1000007 (2008).
41. D'Souza, V., Dey, A., Habib, D. & Summers, M. F. NMR structure of the 101-nucleotide core encapsidation signal of the Moloney murine leukemia virus. *J. Mol. Biol.* **337,** 427–42 (2004).
42. Steffen, D., Bird, S., Rowe, W. P. & Weinberg, R. a. Identification of DNA fragments carrying ecotropic proviruses of AKR mice. *Proc. Natl. Acad. Sci. U. S. A.* **76,** 4554–8 (1979).

43. Stoye, J. P. & Coffin, J. M. proviruses revealed by using virus Polymorphism of Murine Endogenous Proviruses Revealed by Using Virus Class-Specific Oligonucleotide Probes. **62,** (1988).
44. Ribet, D., Harper, F., Esnault, C., Pierron, G. & Heidmann, T. The GLN family of murine endogenous retroviruses contains an element competent for infectious viral particle formation. *J. Virol.* **82,** 4413–9 (2008).
45. Retroviruses, I. E. *et al.* Evolution and Distribution of Class Evolution and Distribution of Class II-Related Endogenous Retroviruses †. (2005). doi:10.1128/JVI.79.10.6478
46. Baillie, G. J., Lagemaat, L. N. Van De, Baust, C. & Mager, D. L. Multiple Groups of Endogenous Betaretroviruses in Mice , Rats , and Other Mammals Multiple Groups of Endogenous Betaretroviruses in Mice , Rats , and Other Mammals. (2004). doi:10.1128/JVI.78.11.5784
47. McCarthy, E. M. & McDonald, J. F. Long terminal repeat retrotransposons of Mus musculus. *Genome Biol.* **5,** R14 (2004).
48. Czarneski, J., Rassa, J. C. & Ross, S. R. Mouse Mammary Tumor Virus. 469–479 (2003).
49. Chase, D. G. & Pikó, L. Expression of A- and C-type particles in early mouse embryos. *J. Natl. Cancer Inst.* **51,** 1971–5 (1973).
50. Qin, C. *et al.* Intracisternal A particle genes: Distribution in the mouse genome, active subtypes, and potential roles as species-specific mediators of susceptibility to cancer. *Mol. Carcinog.* **49,** 54–67 (2010).
51. Waterston, R. H. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–62 (2002).
52. Bénit, L., Lallemand, J., Philippe, H. & Heidmann, T. ERV-L Elements : a Family of Endogenous Retrovirus-Like Elements Active throughout the Evolution of Mammals ERV-L Elements : a Family of Endogenous Retrovirus-Like Elements Active throughout the Evolution of Mammals. (1999).
53. Peaston, a E., Knowles, B. B. & Hutchison, K. W. Genome plasticity in the mouse oocyte and early embryo. *Biochem. Soc. Trans.* **35,** 618–22 (2007).
54. Leung, D. C. & Lorincz, M. C. Silencing of endogenous retroviruses: when and why do histone marks predominate? *Trends Biochem. Sci.* **37,** 127–133 (2011).
55. Hutnick, L. K., Huang, X., Loo, T.-C., Ma, Z. & Fan, G. Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. *J. Biol. Chem.* **285,** 21082–91 (2010).
56. Tamaru, H. & Selker, E. U. A histone H3 methyltransferase controls DNA methylation in Neurospora crassa. *Nature* **414,** 277–83 (2001).
57. Lehnertz, B. *et al.* Suv39h-Mediated Histone H3 Lysine 9 Methylation Directs DNA Methylation to Major Satellite Repeats at Pericentric Heterochromatin. **13,** 1192–1200 (2003).
58. Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7,** 597–606 (2004).
59. Biosci, C., Lib, N. R. & Dupressoir, A. Germ line-specific expression of intracisternal A-particle retrotransposons in transgenic mice . These include : Germ Line-Specific Expression of Intracisternal A-Particle Retrotransposons in Transgenic Mice. **16,** (1996).
60. Brûlet, P., Condamine, H. & Jacob, F. Spatial distribution of transcripts of the long repeated ETn sequence during early mouse embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **82,** 2054–8 (1985).

61. Baust, C. *et al.* Structure and Expression of Mobile ETnII Retroelements and Their Coding-Competent MusD Relatives in the Mouse Structure and Expression of Mobile ETnII Retroelements and Their Coding-Competent MusD Relatives in the Mouse. (2003). doi:10.1128/JVI.77.21.11448
62. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).
63. Rossant, J. & Tam, P. P. L. Blastocyst lineage formation, early embryonic asymmetries and axis patterning in the mouse. *Development* **136,** 701–713 (2009).
64. I. Wilmut et al. Viable offspring derived from fetal and adult mammalian cells Nature 385, 810-813 (1997).
65. Mosteiro, L. *et al.* Reprogramming in vivo produces teratomas and iPS cells with totipotency features. doi:10.1038/nature12586
66. Ishiuchi, T. *et al.* Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22,** 662–671 (2015).
67. Macfarlan, T. S. *et al.* Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487,** 57–63 (2012).
68. Morgani, S. M. *et al.* Totipotent embryonic stem cells arise in ground-state culture conditions. *Cell Rep.* **3,** 1945–57 (2013).
69. Ambros, V. The functions of animal microRNAs. *Nature* **431,** 350–5 (2004).
70. Ameres, S. L. & Zamore, P. D. Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14,** 475–88 (2013).
71. Bartel, D. P. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. *Cell* **116,** 281–297 (2004).
72. Chekulaeva, M. & Filipowicz, W. Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. *Curr. Opin. Cell Biol.* **21,** 452–460 (2009).
73. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet.* **5,** 522–531 (2004).
74. Kim, V. N., Han, J. & Siomi, M. C. Biogenesis of small RNAs in animals. *Nat. Rev. Mol. Cell Biol.* **10,** 126–139 (2009).
75. Zamore, P. D. & Haley, B. Ribo-gnome: the big world of small RNAs. *Science (80-.).* **309,** 1519–24 (2005).
76. Chen, L., Wang, D., Wu, Z., Ma, L. & Daley, G. Q. Molecular basis of the first cell fate determination in mouse embryogenesis. *Cell Res.* **20,** 982–93 (2010).
77. Judson, R. L., Babiarz, J. E., Venere, M. & Blelloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. *Nat. Biotechnol.* **27,** 459–61 (2009).
78. Kanellopoulou, C. *et al.* Dicer-deficient mouse embryonic stem cells are defective in differentiation and centromeric silencing. *Genes Dev.* **19,** 489–501 (2005).
79. Viswanathan, S. R. & Daley, G. Q. Lin28: A MicroRNA Regulator with a Macro Role. *Cell* **140,** 445–449 (2010).
80. Viswanathan, S. R. *et al.* MicroRNA expression during trophectoderm specification. *PLoS One* **4,** 17–19 (2009).
81. Macfarlan, T. S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25,** 594–607 (2011).
82. Choi, Y. J. *et al.* miR-34 miRNAs provide a barrier for somatic cell reprogramming. *Nat Cell Biol* **13,** 1353–1360 (2011).
83. Niwa, H. *et al.* Interaction between Oct3/4 and Cdx2 determines trophectoderm

differentiation. *Cell* **123,** 917–929 (2005).
84. Giakoumopoulos, M. & Golos, T. G. Embryonic stem cell-derived trophoblast differentiation: A comparative review of the biology, function, and signaling mechanisms. *J. Endocrinol.* **216,** (2013).
85. Kubaczka, C. *et al.* Direct Induction of Trophoblast Stem Cells from Murine Fibroblasts. *Cell Stem Cell* **17,** 557–568 (2015).
86. Benchetrit, H. *et al.* Extensive Nuclear Reprogramming Underlies Lineage Conversion into Functional Trophoblast Stem-like Cells. *Cell Stem Cell* **17,** 543–556 (2015).
87. Kigami, D. MuERV-L Is One of the Earliest Transcribed Genes in Mouse One-Cell Embryos. *Biol. Reprod.* **68,** 651–654 (2002).
88. Nichols, J. *et al.* Formation of pluripotent stem cells in the mammalian embryo dependes on the POU transcription factor Oct4. *Cell* **95,** 379–391 (1998).
89. Zhang, C., Ye, X., Zhang, H., Ding, M. & Deng, H. GATA factors induce mouse embryonic stem cell differentiation toward extraembryonic endoderm. *Stem Cells Dev.* **16,** 605–613 (2007).
90. Maksakova, I. A. *et al.* 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).
91. Lewis, B. P., Shih, I., Jones-Rhoades, M. W., Bartel, D. P. & Burge, C. B. 33-Prediction of Mammalian MicroRNA Targets. *Cell* **115,** 787–798 (2003).
92. Miranda, K. C. *et al.* A Pattern-Based Method for the Identification of MicroRNA Binding Sites and Their Corresponding Heteroduplexes. *Cell* **126,** 1203–1217 (2006).
93. Orkin, H. GATA- Binding Transcript ion Factors in Hema to poie t ic Cells. *Cancer* 575–581 (1992).
94. Suzuki, M. *et al.* GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. *Genes to Cells* **18,** 921–933 (2013).
95. Wang, J. *et al.* Primate-specific endogenous retrovirus-driven transcription defines naive-like stem cells. *Nature* **516,** 405–9 (2014).
96. Concepcion, C. P. *et al.* Intact p53-dependent responses in miR-34-deficient mice. *PLoS Genet.* **8,** (2012).
97. Rossant, J. Postimplantation development of blastomeres isolated from 4- and 8-cell mouse eggs. *J. Embryol. Exp. Morphol.* **36,** 283–90 (1976).

Figures and Tables

Fig. 1

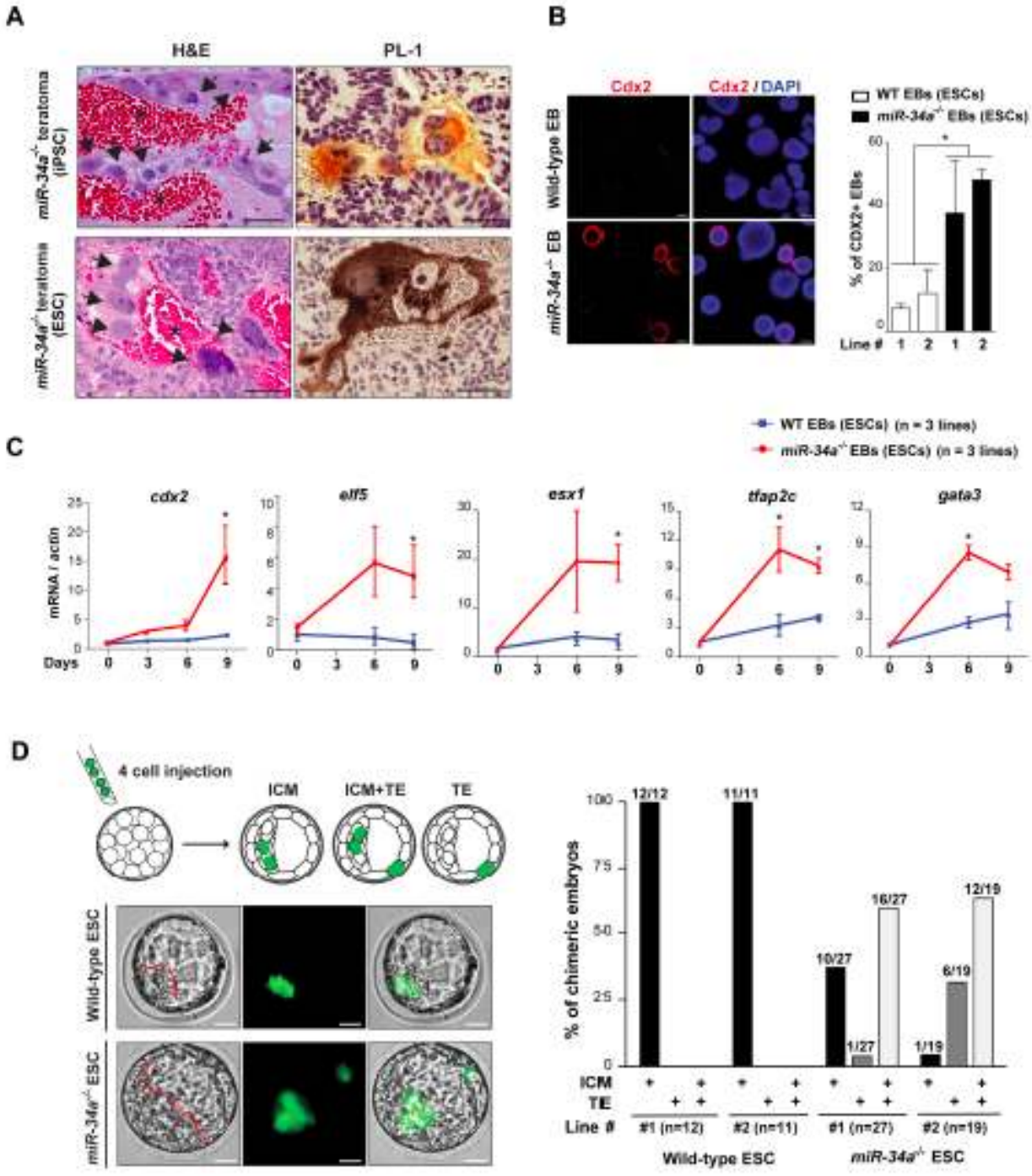

A
H&E
PL-1
miR-34a-/- teratoma (iPSC)
miR-34a-/- teratoma (ESC)
B
Cdx2
Cdx2 / DAPI
Wild-type EB
miR-34a-/- EB
WT EBs (ESCs)
miR-34a-/- EBs (ESCs)
% of CDX2+ EBs
Line # 1 2 1 2
C
WT EBs (ESCs) (n = 3 lines)
miR-34a-/- EBs (ESCs) (n = 3 lines)
cdx2
eif5
esx1
tfap2c
gata3
mRNA / actin
Days
D
4 cell injection
ICM
ICM+TE
TE
Wild-type ESC
miR-34a-/- ESC
% of chimeric embryos
12/12
11/11
10/27
1/27
16/27
1/19
6/19
12/19
ICM
TE
Line #
#1 (n=12)
#2 (n=11)
#1 (n=27)
#2 (n=19)
Wild-type ESC
miR-34a-/- ESC

Fig. 1 (Cont'd)

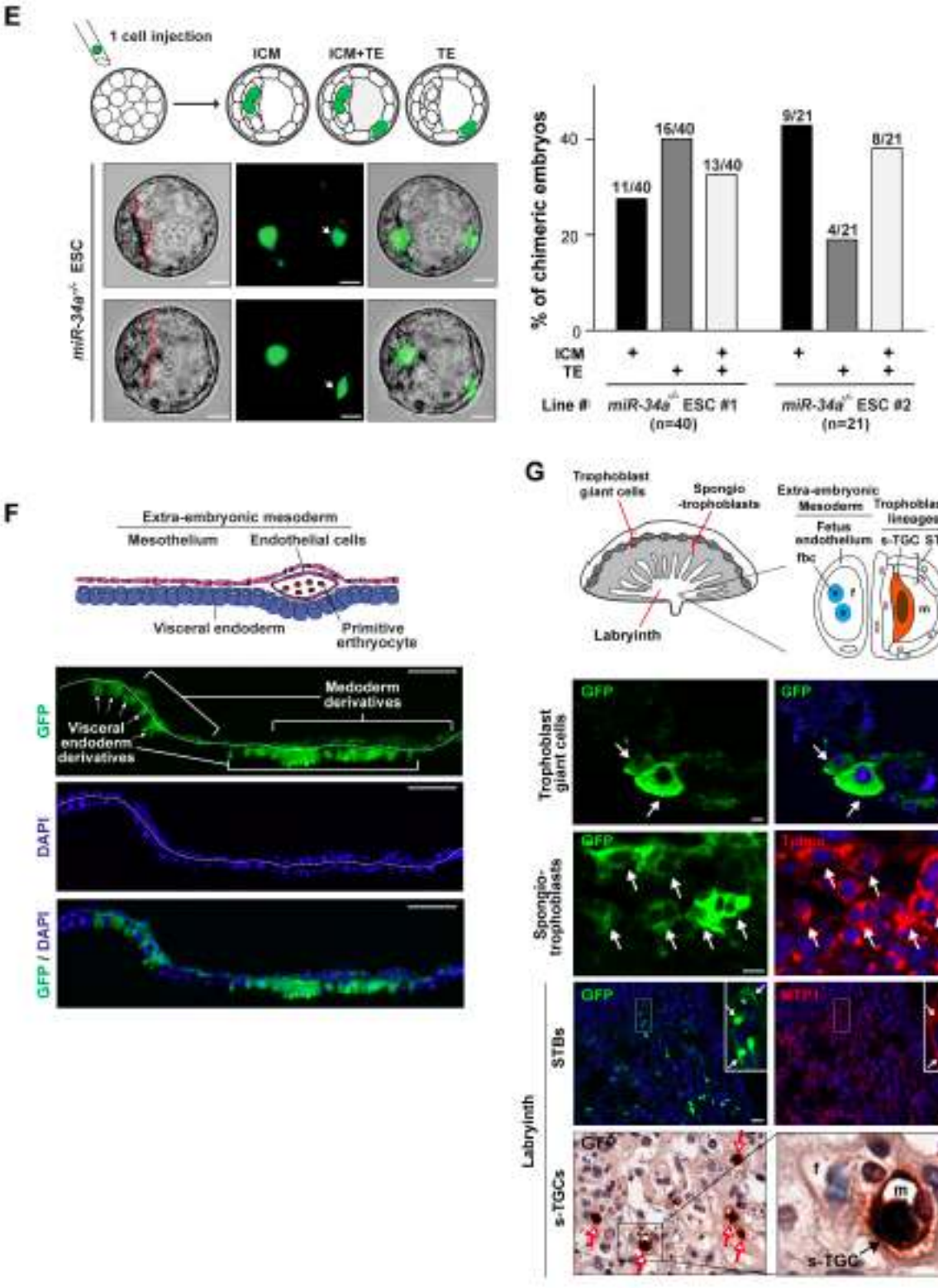

E
1 cell injection
ICM
ICM+TE
TE
miR-34a-/- ESC
% of chimeric embryos
11/40
16/40
13/40
9/21
4/21
8/21
ICM
TE
Line #
miR-34a-/- ESC #1 (n=40)
miR-34a-/- ESC #2 (n=21)
F
Extra-embryonic mesoderm
Mesothelium
Endothelial cells
Visceral endoderm
Primitive erthyrocyte
GFP
Medoderm derivatives
Visceral endoderm derivatives
DAPI
GFP / DAPI
G
Trophoblast giant cells
Spongio -trophoblasts
Extra-embryonic Mesoderm
Trophoblast lineages
Fetus endothelium
s-TGC
STB
fbc
Labryinth
Trophoblast giant cells
Spongio-trophoblasts
STBs
s-TGCs
GFP
s-TGC

Figure Legends

Fig 1. miR-34a$^{-/-}$ pluripotent stem cells exhibit expanded cell fate potential. (A). miR-34a$^{-/-}$ teratomas contain extra-embryonic cell lineages and extra-embryonic cell markers. Teratomas generated from miR-34a$^{-/-}$ iPSCs and miR-34a$^{-/-}$ ESCs contain cells with the typical placental trophoblast giant cell morphology (black arrows) and placental lactogen 1 (PL-1) expression. Asterisks: the blood-filled lacunae associated with placenta giant cell-like cells. Scale bars, 50 mm. B, C. miR-34a$^{-/-}$ embryoid bodies (EBs) exhibit an induction of both embryonic and extra-embryonic cell markers in immunofluorescence (IF) staining (B) and real-time PCR analyses (C). B. miR-34a$^{-/-}$ EBs yield a greater percentage of Cdx2-positive EBs in IF staining and exhibit an induction of the TE marker Cdx2 primarily in cells at the periphery. Scale bars, 100 mm. Error bars: s.d., n=5-7 (randomly selected 10x fields). C. miR-34a$^{-/-}$ EBs showed an increase in TE markers cdx2, elf5, esx1, tfap2c and gata3. Three independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ ESC lines were compared. Error bars, s.d., n=3. D-G. miR-34a$^{-/-}$ ESCs contribute to both embryonic and extra-embryonic cell lineages in chimeric assays in vivo. D. Four GFP-labeled wild-type or miR-34a$^{-/-}$ ESCs were microinjected into each C57BL/6N recipient morula, and the contribution of their progenies to the inner cell mass (ICM) and the trophectoderm (TE) were determined by the localization of GFP-positive cells (left). Scale bar, 20 mm. The percentage of chimeric blastocyst embryos with ESC contribution to the ICM, the TE, and ICM+TE were measured for both wild-type and miR-34a$^{-/-}$ ESCs (right). Two independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ ESCs were compared. n, the number of chimeric embryos obtained for each ESC line from three independent injections (Table S1). Two independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ ESC lines were compared. E. Single GFP-labeled miR-34a$^{-/-}$ ESCs are able to contribute to both ICM and TE (white arrows) of chimeric blastocysts. Representative images were shown for two chimeric blastocysts (top). Scale bar, 20 mm. The percentage of chimeric embryos with ESC contribution to the ICM, the TE, and ICM+TE were quantified (bottom). Two independent miR-34a$^{-/-}$ ESC lines were examined. n, the number of chimeric blastocyst embryos for each ESC line from three independent injections for each line. All P-values were calculated on a basis of a two-tailed student's t-test. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$. Two independent miR-34a$^{-/-}$ ESC lines were examined. F, G. miR-34a$^{-/-}$ ESCs contribute to multiple differentiated cell lineages in embryo, yolk sac and placenta in chimeric analyses in vivo. 10-15 GFP-labeled wild-type or miR-34a$^{-/-}$ ESCs were microinjected into C57BL/6N blastocysts or aggregated with CD1 recipient morulae to generate chimeric embryos. GFP-labeled miR-34a$^{-/-}$ ESCs contributed to the visceral endoderm derivatives of the yolk sac (F, scale bar, 50 mm), and multiple trophoblast lineages of the placenta (trophoblast giant cells, spongiotrophoblasts, syncytiotrophoblasts (STBs) and sinusoidal trophoblast giant cells (s-TGCs) (G, scale bar, 50 mm) of the chimeric embryos at E12.5 and E14.5. F. A diagram illustrates the yolk sac tissue architecture and major cell types (top). GFP-positive visceral endoderm cells are identified based on the bilaminar structure of the yolk sac and their characteristic columnar epithelial morphology (bottom). G. A diagram illustrates the placenta tissue architecture and major cell types (top). (Bottom) GFP-positive trophoblast giant cells (white arrows) and s-TGCs (red arrows) 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 protein alpha (Tpbpa) or ferroportin (MTP1). fbc, nucleated fetal blood cells; f, fetal blood; m, maternal blood. The

statistic summary of multiple passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ ESC lines were summarized in Table S1.

Fig. 2

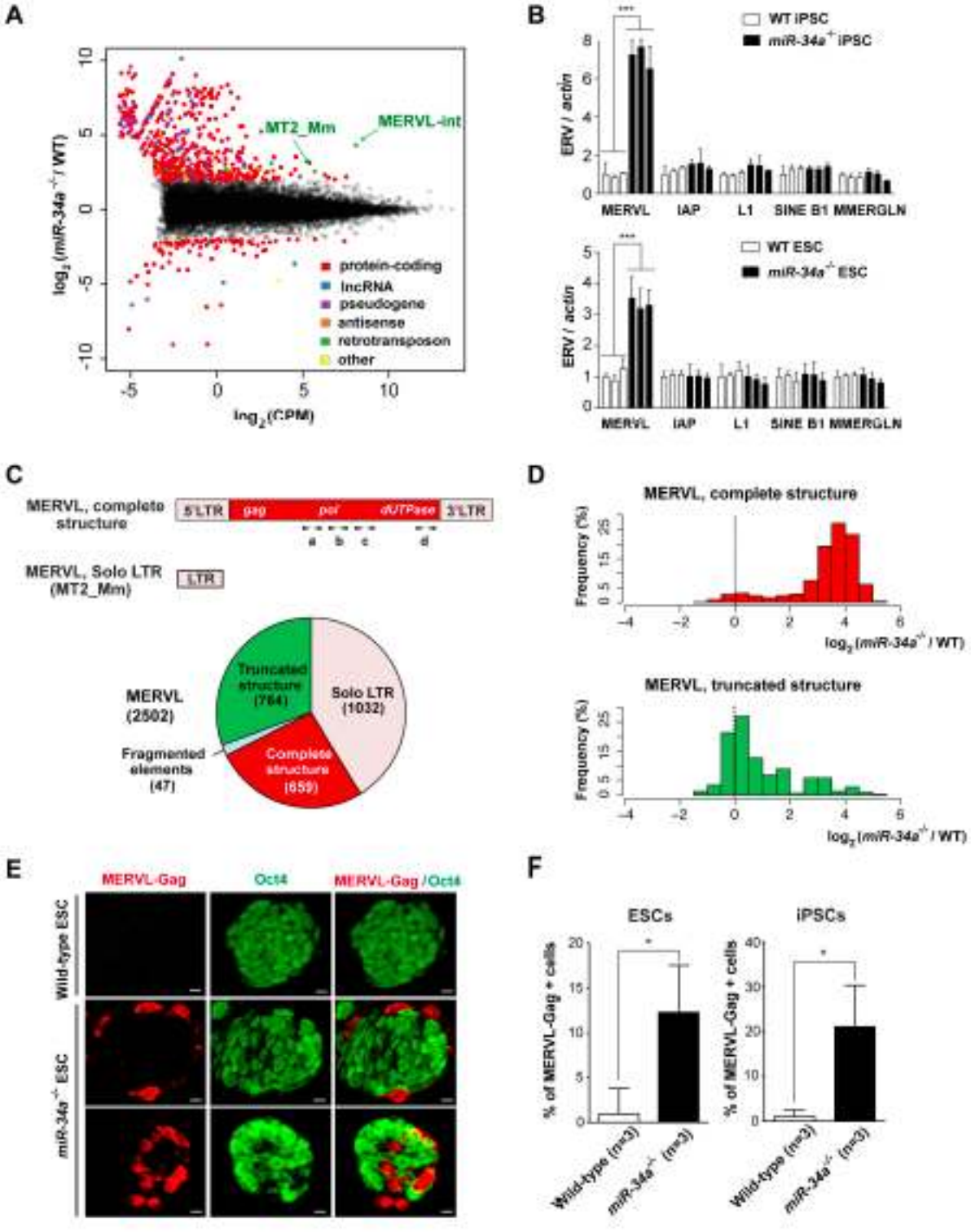

A
MT2_Mm
MERVL-int
protein-coding
lncRNA
pseudogene
antisense
retrotransposon
other
log2(CPM)
B
WT iPSC
miR-34a-/- iPSC
WT ESC
miR-34a-/- ESC
ERV / actin
MERVL
IAP
L1
SINE B1
MMERGLN
C
MERVL, complete structure
5'LTR
gag
pol
dUTPase
3'LTR
MERVL, Solo LTR (MT2_Mm)
LTR
MERVL (2502)
Truncated structure (764)
Solo LTR (1032)
Fragmented elements (47)
Complete structure (659)
D
MERVL, complete structure
MERVL, truncated structure
Frequency (%)
E
MERVL-Gag
Oct4
MERVL-Gag / Oct4
Wild-type ESC
miR-34a-/- ESC
F
ESCs
iPSCs
% of MERVL-Gag + cells
Wild-type (n=3)
miR-34a-/- (n=3)

Fig. 2 (Cont'd)

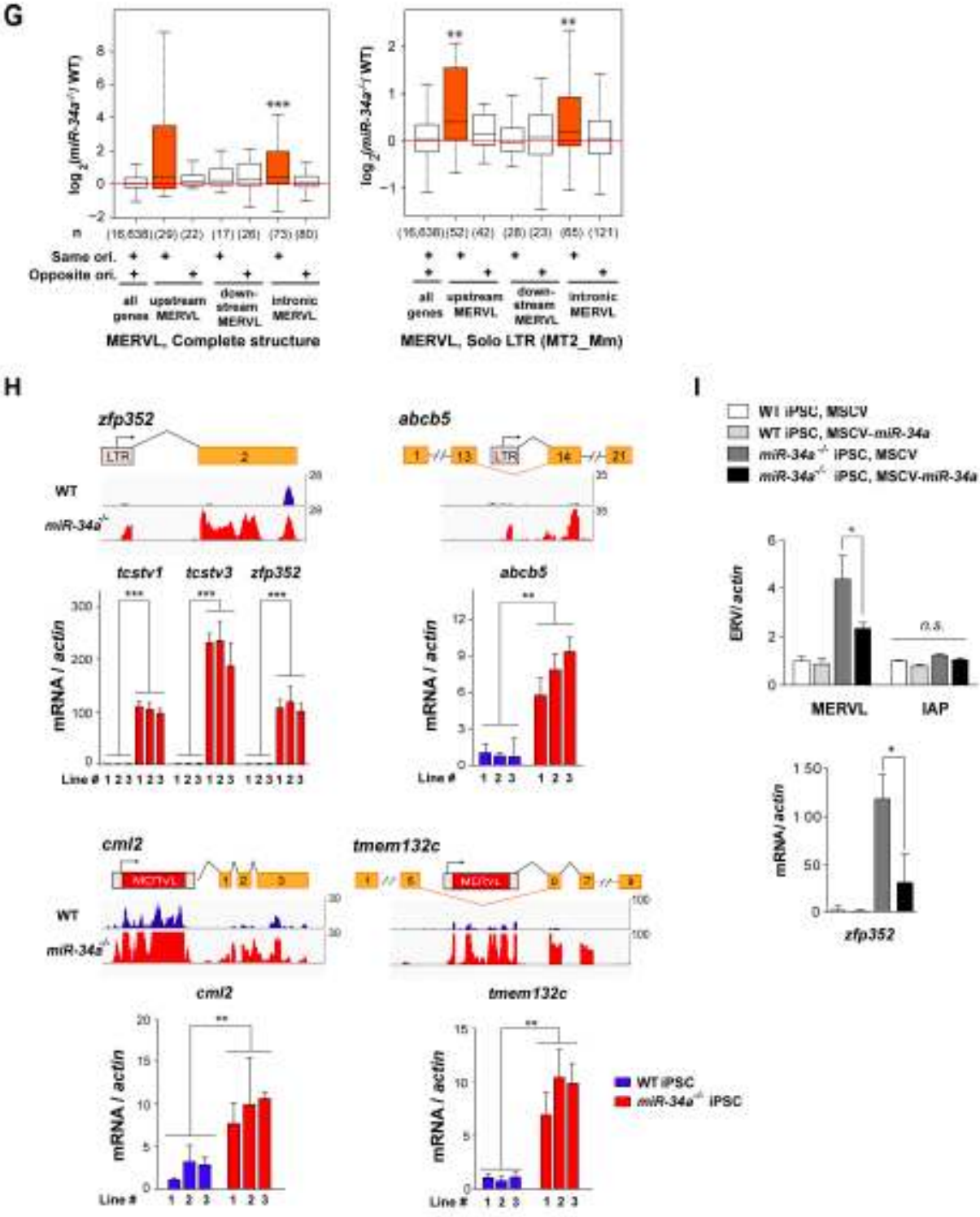

G
MERVL, Complete structure
MERVL, Solo LTR (MT2_Mm)
all genes
upstream MERVL
down-stream MERVL
intronic MERVL
Same ori.
Opposite ori.
H
zfp352
abcb5
WT
miR-34a-/-
tcstv1
tcstv3
zfp352
mRNA / actin
Line # 1 2 3 1 2 3 1 2 3 1 2 3
cml2
tmem132c
WT iPSC
miR-34a-/- iPSC
I
WT iPSC, MSCV
WT iPSC, MSCV-miR-34a
miR-34a-/- iPSC, MSCV
miR-34a-/- iPSC, MSCV-miR-34a
ERV/ actin
MERVL
IAP
n.s.
mRNA/ actin

Fig. 2. miR-34a$^{-/-}$ pluripotent stem cells exhibit specific induction of the MERVL ERVs. A. The MERVL ERVs are highly induced and differentially expressed (DE) transcriptional unit in miR-34a$^{-/-}$ iPSCs compared to wild-type iPSCs. An MA-plot compares the transcription profiles of miR-34a$^{-/-}$ and wild-type (WT) iPSCs using RNA-seq data. DE transcriptional units, including protein-coding genes, long non-coding RNAs (lncRNAs), pseudogenes, antisense transcripts, and retrotransposons, are color-coded by class. The 441 DE transcription units between miR-34a$^{-/-}$ and wild-type iPSCs (False Discovery Rate 5%, absolute $\log_2$ fold-change ≥ 2) include 352 protein-coding genes, 54 pseudogenes, 13 lncRNAs, 6 antisense transcripts, 4 retrotransposon families, and 12 others. Among these DE genes, MERVL-int (the MERVL internal sequence that encodes gag, pol and dUTPase) and MT2_Mm (the solo-LTR of the canonical MERVL) are strongly induced and highly expressed in miR-34a$^{-/-}$ iPSCs. RNA-seq data were generated from three pairs of independently derived, passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSCs. CPM: counts per million. B. MERVL ERVs are specifically induced in miR-34a$^{-/-}$ iPSCs and ESCs, as measured 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. Three independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC and ESC lines were compared. Error bars: s.d., n=3. C. A schematic diagram illustrates the transcript structure of the MERVL family of ERVs, with the position of four pairs of validated real-time PCR primers indicated (a, b, c and d) (top). (Bottom) A pie chart shows the relative abundance for the different MERVL subclasses, categorized according to their structural features. MERVL loci with a complete structure (red) carry both 5' and 3' LTRs, along with an internal sequence (annotated by Repeat Masker as MERVL-int); truncated MERVL elements (green) lack one or both LTRs; and solo-LTRs of canonical MERVL (pink), also designated as MT2_Mm, are 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, either with a complete structure (top) or with a truncated structure (bottom). E. The MERVL-Gag and Oct4 expression is mutually exclusive in a subset of miR-34a$^{-/-}$ ESCs, while MERVL-Gag staining is absent in wild-type ESCs. Scale bars, 20 mm. 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., n=5-6 (randomly selected 10x fields). E-F. Three independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC and ESC lines were compared. G-I. The MERVL activation in miR-34a$^{-/-}$ pluripotent stem cells induces many MERVL proximal gene isoforms. G. 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 between miR-34a$^{-/-}$ and wild-type iPSCs 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 for reference. Same ori. (same orientation): MERVL and its proximal gene are on the same strand; opposite ori. (opposite orientation): MERVL 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, have significantly larger fold-changes than the rest of the genes. **, $P < 0.01$; *** $P < 0.001$; Wilcoxon-Mann-Whitney test. H. Examples are shown for induced expression and altered transcript structure of MERVL proximal genes in miR-34a$^{-/-}$

iPSCs. The MT2B1 solo LTR elements upstream of zfp352, tcstv1 or tcstv3 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. Intronicly localized MERVLs, either a solo LTR (as that in intron 13 of abcb5) or a complete ERV element (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. Error bars: s.d., n=3, *** $P < 0.001$. I. miR-34a overexpression in miR-34a$^{-/-}$ iPSCs using a MSCV retroviral vector significantly suppresses the level of MERVL and the MERVL-zfp352 chimeric transcript in real-time PCR analysis, but causes no alteration in the level of IAP. One pair of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSCs were examined. Error bars: s.d., n=3. All P-values were calculated on a basis of a two-tailed Student's t-test unless stated otherwise. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Fig. 3

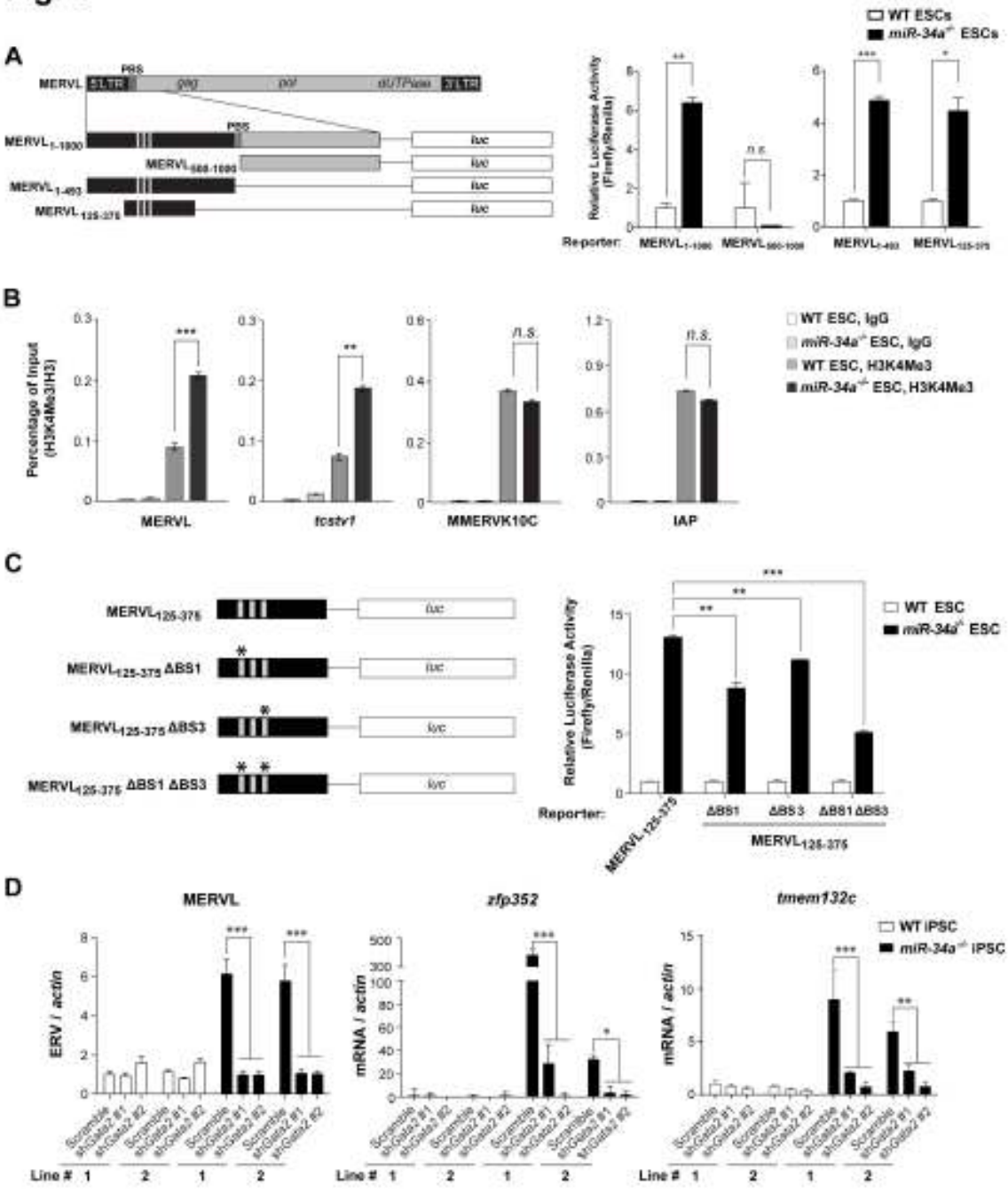

A
MERVL
PBS
gag
pol
dUTPase
MERVL1-1000
MERVL500-1000
MERVL1-493
MERVL125-375
luc
Relative Luciferase Activity (Firefly/Renilla)
WT ESCs
miR-34a-/- ESCs
n.s.
Reporter:
B
Percentage of Input (H3K4Me3/H3)
WT ESC, IgG
miR-34a-/- ESC, IgG
WT ESC, H3K4Me3
miR-34a-/- ESC, H3K4Me3
MERVL
fcstv1
MMERVK10C
IAP
C
MERVL125-375
MERVL125-375 ΔBS1
MERVL125-375 ΔBS3
MERVL125-375 ΔBS1 ΔBS3
WT ESC
miR-34a-/- ESC
ΔBS1
ΔBS3
ΔBS1 ΔBS3
D
MERVL
zfp352
tmem132c
ERV / actin
mRNA / actin
WT iPSC
miR-34a-/- iPSC
Scramble
shGata2 #1
shGata2 #2
Line # 1 2 1 2

Fig. 3 (Cont'd)

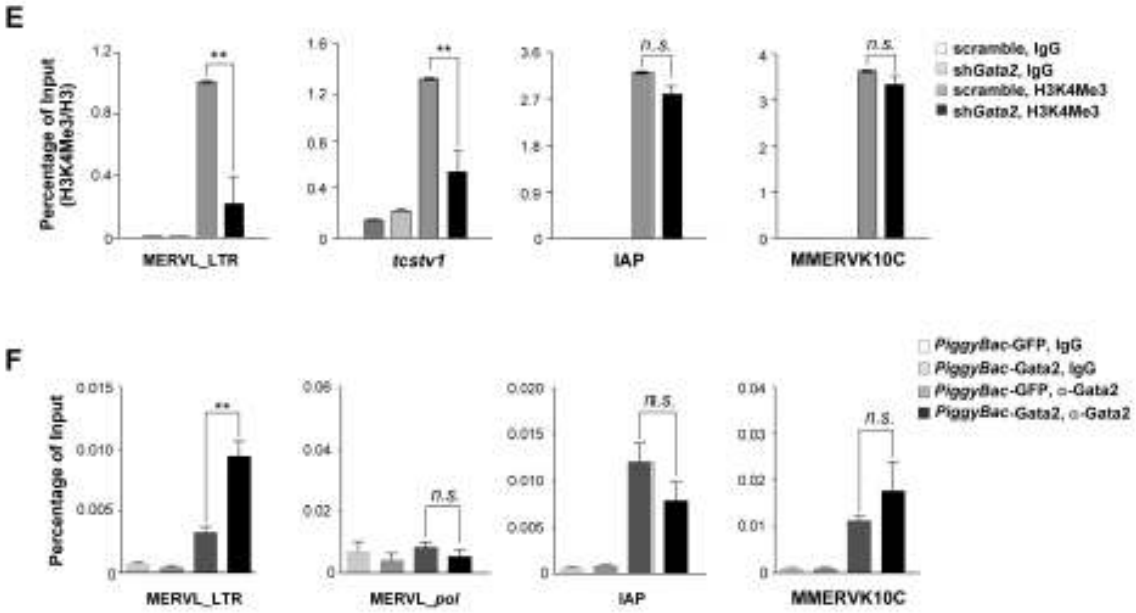

E
Percentage of Input (H3K4Me3/H3)
MERVL_LTR
tcstv1
IAP
MMERVK10C
**
n.s.
scramble, IgG
shGata2, IgG
scramble, H3K4Me3
shGata2, H3K4Me3
F
Percentage of Input
MERVL_LTR
MERVL_pol
IAP
MMERVK10C
PiggyBac-GFP, IgG
PiggyBac-Gata2, IgG
PiggyBac-GFP, α-Gata2
PiggyBac-Gata2, α-Gata2

Fig.3. Gata2 is essential for the MERVL induction in miR-34a$^{-/-}$ pluripotent stem cells. A. A full length or a fragment of MERVL LTR can be activated specifically in miR-34a$^{-/-}$ESCs. (Left) A schematic diagram shows the MERVL LTR and the truncated 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 Luc reporter containing the truncated MERVL$_{125\text{-}375}$ fragment completely 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 an increased H3K4Me3 modification on the MERVL LTR and the MERVL-tcstv1 chimeric gene in miR-34a$^{-/-}$ ESCs. In comparison, the H3K4Me3 level is unaltered on IAP LTR and MMERVK10C LTR. C. Mutations of two predicted Gata2 binding sites (BS1 and BS3) in the MERVL$_{125\text{-}375}$-Luc reporter synergistically impair the reporter activity in miR-34a$^{-/-}$ ESCs. (Left) A schematic diagram shows the MERVL$_{125\text{-}375}$-Luc reporter and the mutated derivatives. (Right) While the mutation of BS1 or BS3 alone in the MERVL$_{125\text{-}375}$-Luc reporter modestly impairs its activity in miR-34a$^{-/-}$ ESCs, mutations of both BS1 and BS3 significantly reduces the reporter activity. Error bars: s.d., n=2. D. gata2 knockdown significantly decreases the expression of MERVL and MERVL proximal genes. Using two independent shRNAs targeting gata2, we effectively knocked down gata2 in miR-34a$^{-/-}$ iPSCs using RNA interference (RNAi) (also see Fig. S5C), and observed a decreased expression of MERVL and MERVL proximal genes (zfp352 and tmem132c). E. gata2 knockdown in miR-34a$^{-/-}$ iPSCs significantly decreases the H3K4Me3 deposition on MERVL elements and on the specific MERVL element upstream of 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, IAP LTR, or MMERVK10C LTR is unaltered upon Gata2 overexpression. C, E-F. Two independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC lines were compared. Error bars: s.d., n=3. All P-values were calculated on a basis of a two-tailed Student's t-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, n.s., not significant

Fig. 4

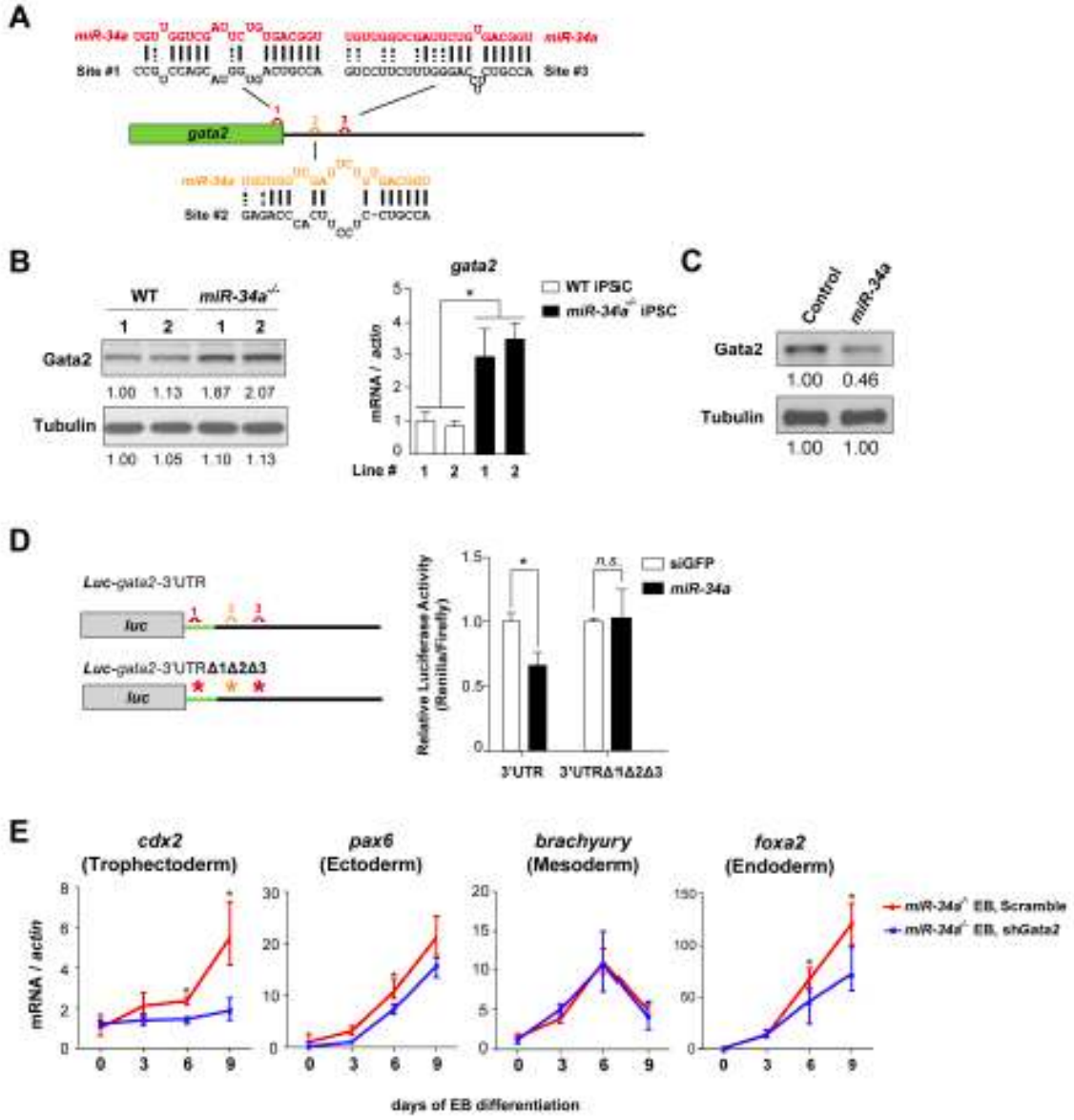
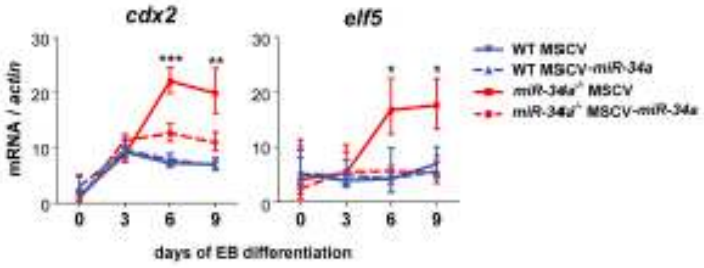
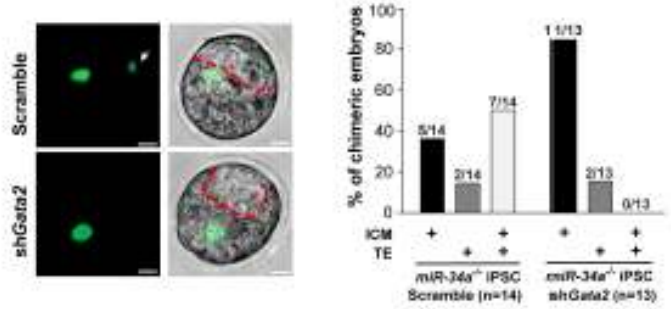

A
miR-34a
Site #1
Site #3
Site #2
gata2
B
WT
miR-34a-/-
Gata2
Tubulin
1.00 1.13 1.87 2.07
1.00 1.05 1.10 1.13
gata2
WT iPSC
miR-34a-/- iPSC
mRNA / actin
Line #
C
Control
miR-34a
Gata2
Tubulin
1.00 0.46
1.00 1.00
D
Luc-gata2-3'UTR
Luc-gata2-3'UTRΔ1Δ2Δ3
luc
siGFP
miR-34a
Relative Luciferase Activity (Renilla/Firefly)
n.s.
3'UTR
3'UTRΔ1Δ2Δ3
E
cdx2 (Trophectoderm)
pax6 (Ectoderm)
brachyury (Mesoderm)
foxa2 (Endoderm)
mRNA / actin
days of EB differentiation
miR-34a-/- EB, Scramble
miR-34a-/- EB, shGata2

Fig. 4 (Cont'd)

F



G



H

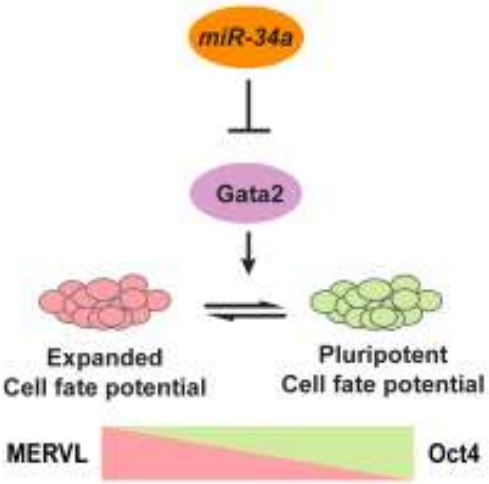


Fig. 4. miR-34a restricts cell fate potential of pluripotent stem cells by targeting gata2. A. A schematic representation of three predicted miR-34a binding sites in the gata2 mRNA, with one site (1) located at the 3' end of the open reading frame (ORF) and two sites (2 and 3) located within the 3'UTR. Site 1 (red) is predicted as a strong miR-34a binding site by both duplex folding energy and the 7mer-A1 seed-match rule (28, 29). While site 3 (red) does not have a perfect seed sequence, it contains a compensatory 3' base-pairing and exhibits a strong folding energy. In contrast, site 2 (orange) represents a weaker prediction (28). B. Gata2 exhibits miR-34a-dependent repression in miR-34a$^{-/-}$ pluripotent stem cells. Gata2 protein (left) and gata2 mRNA (right) are elevated in miR-34a$^{-/-}$ iPSCs as compared with 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, * $P < 0.05$. C. Overexpression of miR-34a in miR-34a$^{-/-}$ iPSCs represses the Gata2 protein level. The quantification of western blot analyses was performed by ImageJ. D. Mutating all three predicted miR-34a binding sites within the Luc-gata2-3UTR luciferase reporter completely abolishes its miR-34a-dependent repression. Error bars: s.d., n=2. E. gata2 knockdown in miR-34a$^{-/-}$ iPSCs by RNAi 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 significantly suppresses the induction of TE markers cdx2 and elf5 upon EB differentiation. Error bars: s.d., n=3. E-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 analyses. We injected four GFP-labeled cells into each recipient morula. While 50% of chimera blastocysts generated from control infected miR-34a$^{-/-}$ iPSCs contain iPSC contribution to both ICM and TE (n=7/14), miR-34a$^{-/-}$ iPSCs with gata2 knockdown lost this expanded cell fate potential (n=0/13), and primarily contributed to the ICM (n=11/13). All P-values were calculated on a basis of a two-tailed Student's t-test. * $P < 0.05$, ** $P < 0.01$. H. A diagram shows our proposed model in which the miR-34a/gata2 pathway restricts the pluripotent cell fate potential of pluripotent stem cells.

Fig. S1

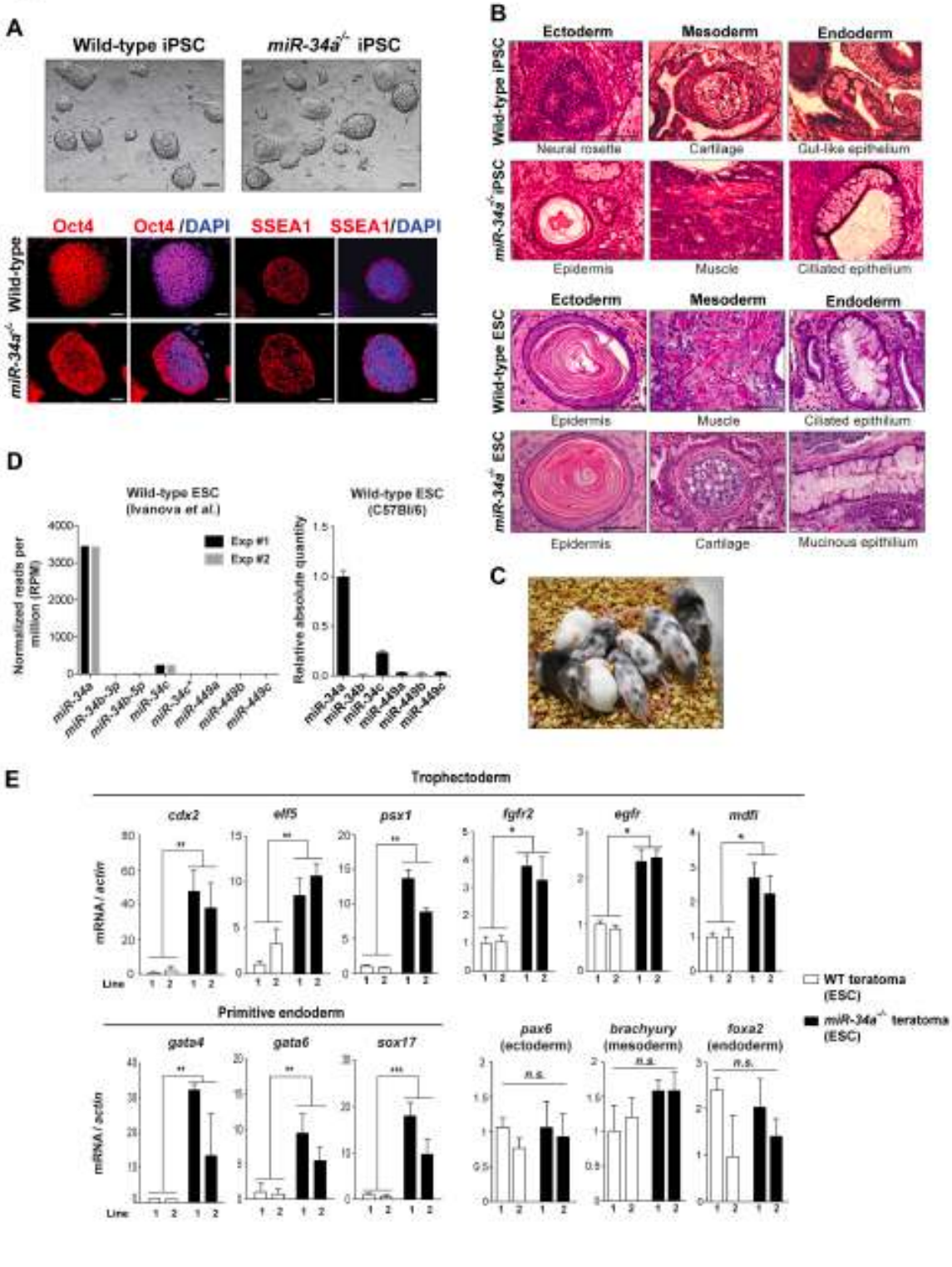

A
Wild-type iPSC
miR-34a-/- iPSC
Oct4
Oct4 /DAPI
SSEA1
SSEA1/DAPI
Wild-type
miR-34a-/-
B
Ectoderm
Mesoderm
Endoderm
Wild-type iPSC
Neural rosette
Cartilage
Gut-like epithelium
miR-34a-/- iPSC
Epidermis
Muscle
Ciliated epithelium
Wild-type ESC
Epidermis
Muscle
Ciliated epithelium
miR-34a-/- ESC
Epidermis
Cartilage
Mucinous epithelium
C
D
Wild-type ESC (Ivanova et al.)
Wild-type ESC (C57Bl/6)
Normalized reads per million (RPM)
Relative absolute quantity
Exp #1
Exp #2
miR-34a
miR-34b-3p
miR-34b-5p
miR-34c
miR-34c*
miR-449a
miR-449b
miR-449c
E
Trophectoderm
cdx2
elf5
psx1
fgfr2
egfr
mdfi
mRNA / actin
Line
Primitive endoderm
gata4
gata6
sox17
pax6 (ectoderm)
brachyury (mesoderm)
foxa2 (endoderm)
n.s.
WT teratoma (ESC)
miR-34a-/- teratoma (ESC)

Fig. S1 (Cont'd)

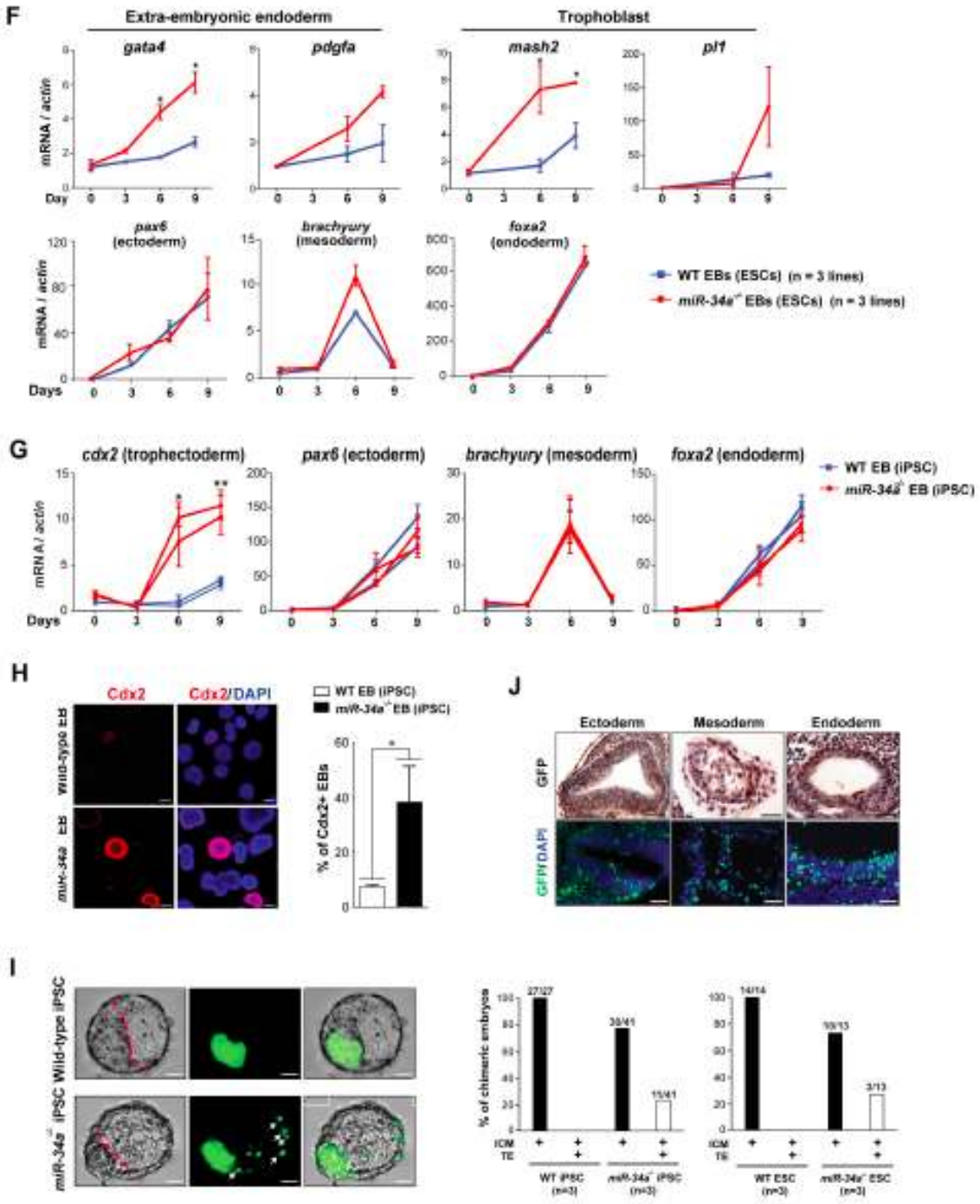

F
Extra-embryonic endoderm
Trophoblast
gata4
pdgfa
mash2
plf
mRNA / actin
Day
pax6 (ectoderm)
brachyury (mesoderm)
foxa2 (endoderm)
Days
WT EBs (ESCs) (n = 3 lines)
miR-34a-/- EBs (ESCs) (n = 3 lines)
G
cdx2 (trophectoderm)
pax6 (ectoderm)
brachyury (mesoderm)
foxa2 (endoderm)
WT EB (iPSC)
miR-34a-/- EB (iPSC)
H
Cdx2
Cdx2/DAPI
Wild-type EB
miR-34a-/- EB
% of Cdx2+ EBs
J
Ectoderm
Mesoderm
Endoderm
GFP
GFP/DAPI
I
Wild-type iPSC
miR-34a-/- iPSC
% of chimeric embryos
ICM
TE
WT iPSC (n=3)
miR-34a-/- iPSC (n=3)
WT ESC (n=3)
miR-34a-/- ESC (n=3)

Fig. S1. miR-34a$^{-/-}$ pluripotent stem cells exhibit an expanded cell fate potential in vitro and in vivo.

A. Wild-type and miR-34a$^{-/-}$ iPSCs exhibit ESC-like morphology and express pluripotency markers Oct4 and SSEA1. Scale bars, 50 mm for differential interference contrast (DIC) images and 20 mm for immunofluorescence (IF) images. B. Wild-type and miR-34a$^{-/-}$ iPSCs (top) and ESCs (bottom) generate differentiated teratomas containing tissues derived from three germ layers (ectoderm, mesoderm and endoderm) as shown by hematoxylin and eosin (H&E) staining. Scale bars, 25 mm. Two independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSCs and ESCs are compared. C. miR-34a$^{-/-}$ iPSCs efficiently contribute to adult chimeric mice as determined by coat color pigmentation. Shown here is a representative image of three independent experiments, in which three independent lines of Oct4-Gfp/+; a/a; miR-34a$^{-/-}$ iPSCs were microinjected into albino-C57BL/6/cBrd/cBrd/cr blastocysts. D. The expression level of miR-34/449 family miRNAs in wild-type ESCs measured by small RNA sequencing (left) and real-time PCR analyses (right). miR-34a is the most highly expressed miR-34/449 miRNA in two independent wild-type ESC lines measured. E. The trophectoderm (TE) marker cdx2, elf5, psx1, fgfr2, egfr and mdfi, as well as the primitive endoderm marker gata4, gata6, and sox17, were highly induced in miR-34a$^{-/-}$ teratomas, 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: standard deviation (s.d.), n=3. F. Compared to wild-type EBs, miR-34a$^{-/-}$ EBs derived from ESCs yield a greater expression of extra-embryonic endoderm marker gata4 and pdgfra, and trophoblast lineage marker mash2 (ascl2) and pl1 (prl3d1). The real-time PCR analyses were performed using data collected from three independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ ESC lines. Error bars: s.e.m., n=3. G. During EB differentiation, miR-34a$^{-/-}$ iPSCs, but not wild-type iPSCs, exhibited strong induction of the TE marker cdx2. In contrast, wild-type and miR-34a$^{-/-}$ EBs similarly induced the expression of pax6 (an ectoderm marker), brachyury (a mesoderm marker) and foxa2 (an endoderm marker). Two independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC lines are compared. Error bars: s.d., n=3. H. Compared to wild-type EBs, miR-34a$^{-/-}$ EBs derived from iPSCs yield a greater percentage of Cdx2-positive EBs, as well as an increased percentage of Cdx2+ cells in each EB. Scale bars, 100 mm. Shown here are representative images and quantitation from one pair of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC and ESC lines. Error bars, s.d., n=6 random fields measured. I. miR-34a$^{-/-}$ iPSCs contribute to both ICM and TE when aggregated with recipient wild-type morulae. The ICM versus the TE contributions from the GFP-labeled wild-type or miR-34a$^{-/-}$ iPSCs were determined by the localization of GFP positive cells in the chimeric blastocysts. While wild-type iPSCs exclusively contribute to the ICM, a fraction of miR-34a$^{-/-}$ iPSCs contributes to both ICM and TE (white arrows) (top). Scale bar, 20 mm. The percentage of chimeric embryos with iPSC (left) or ESC (right) contribution to the ICM, the TE and ICM+TE in the chimeric blastocysts was quantified for both wild-type and miR-34a$^{-/-}$ iPSCs from six independent aggregation experiments (bottom). Three independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC and ESC lines were compared. J. miR-34a$^{-/-}$ESCs contributed to all three lineage germ layers. Scale bars, 50 mm for IHC images and 100 mm for IF images. All P-values were calculated on a basis of a two-tailed Student's t-test. * $P < 0.05$, ** $P < 0.01$.

Fig. S2

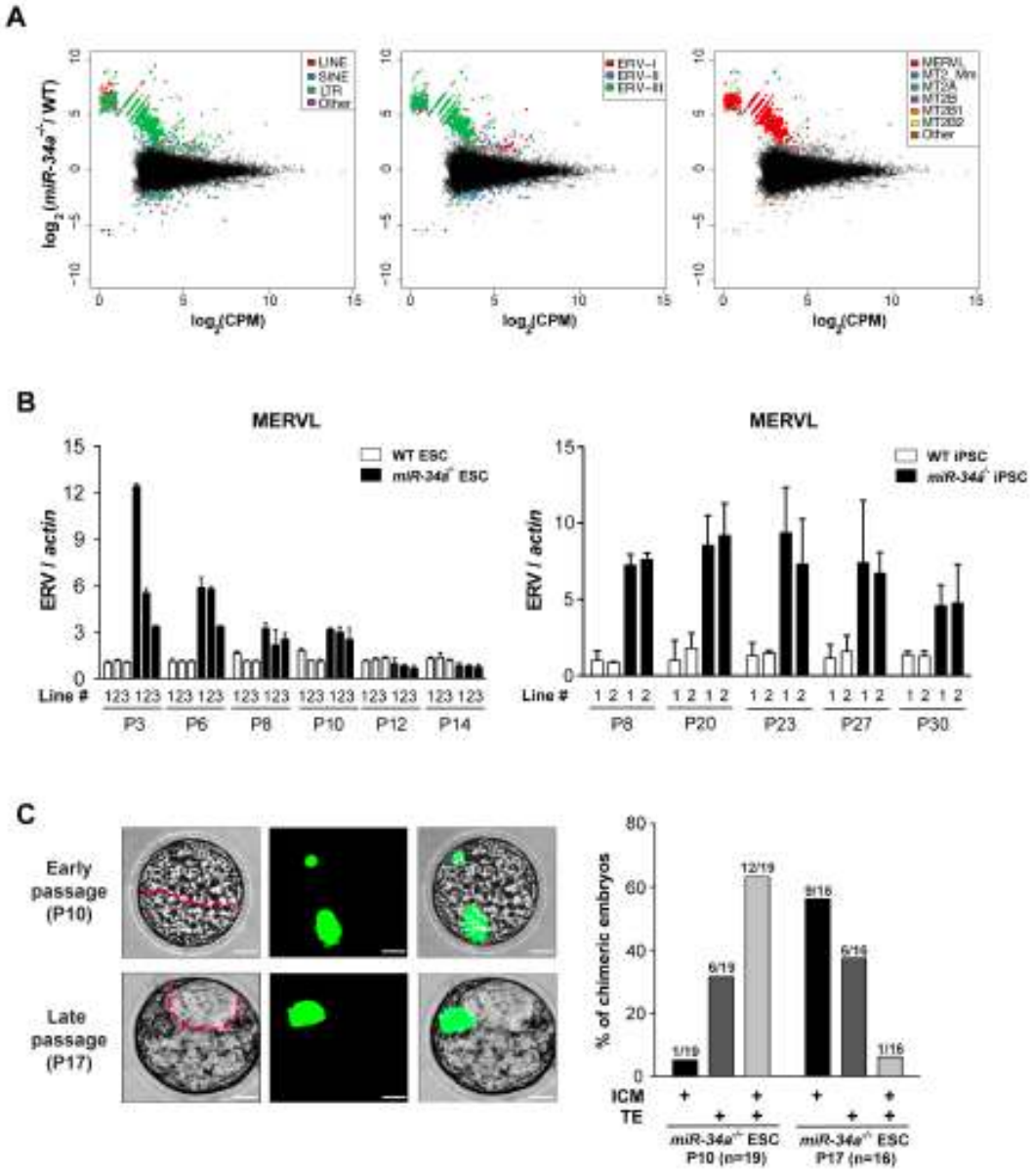

A
LINE
SINE
LTR
Other
ERV-I
ERV-II
ERV-III
MERVL
MT2_Mm
MT2A
MT2B
MT2B1
MT2B2
Other
log2 (miR-34a-/- / WT)
log2(CPM)
B
MERVL
WT ESC
miR-34a-/- ESC
ERV / actin
Line #
P3
P6
P8
P10
P12
P14
MERVL
WT iPSC
miR-34a-/- iPSC
P8
P20
P23
P27
P30
C
Early passage (P10)
Late passage (P17)
% of chimeric embryos
1/19
6/19
12/19
9/16
6/16
1/16
ICM
TE
miR-34a-/- ESC P10 (n=19)
miR-34a-/- ESC P17 (n=16)

Fig. S2 (Cont'd)

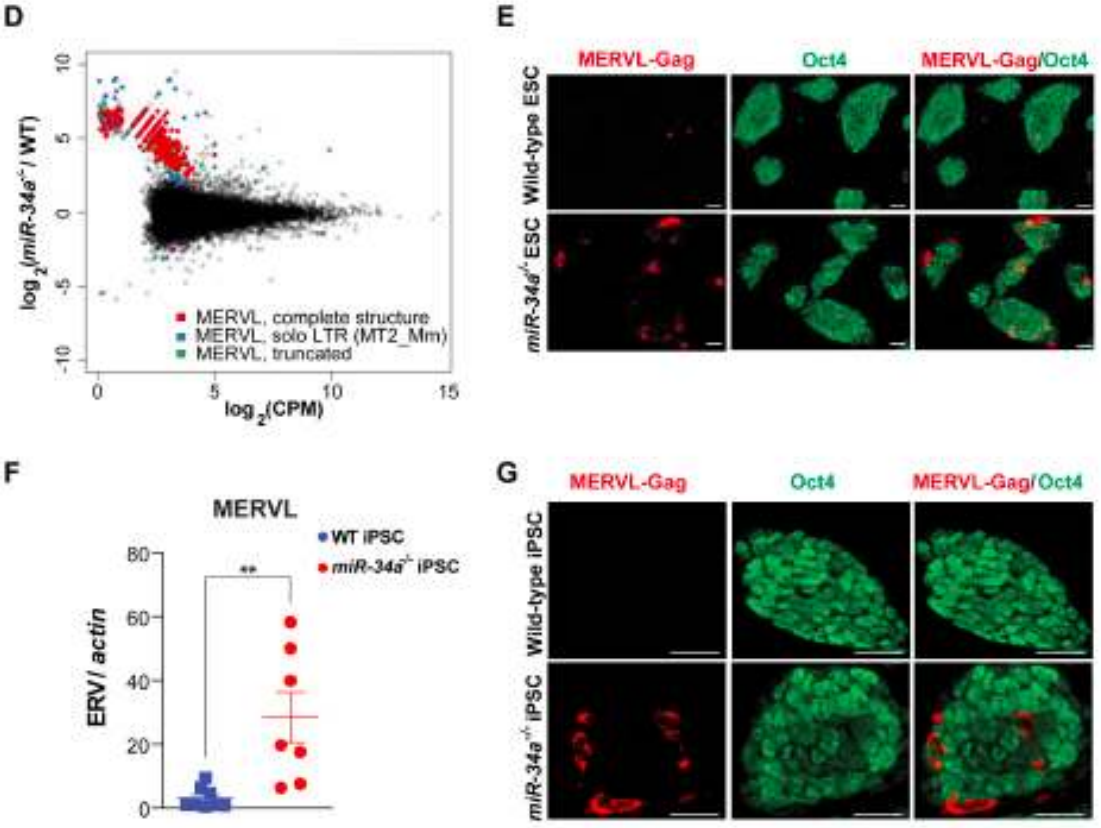

D
log2(miR-34a-/- / WT)
log2(CPM)
MERVL, complete structure
MERVL, solo LTR (MT2_Mm)
MERVL, truncated
E
MERVL-Gag
Oct4
MERVL-Gag/Oct4
Wild-type ESC
miR-34a-/- ESC
F
MERVL
WT iPSC
miR-34a-/- iPSC
ERV/ actin
**
G
MERVL-Gag
Oct4
MERVL-Gag/Oct4
Wild-type iPSC
miR-34a-/- iPSC

Fig. S2. MERVL ERVs are specifically induced in mir-34a$^{-/-}$ pluripotent stem cells.

A. MA plots illustrate the comparison of the transcription profiles of all retrotransposon loci between wild-type and miR-34a$^{-/-}$ iPSCs using the RNA-seq data. We found 949 differentially expressed (DE) retrotransposon loci (FDR of 5%, absolute $\log_2$ fold-change ≥ 2). (Left) The DE loci are color-coded by retrotransposon class, with 69 LINEs (red, 7%), 101 SINEs (blue, 11%), 777 ERVs (green, 82%) and 2 others (purple, <1%). (Middle) Loci from the ERV-III class are preferentially derepressed in miR-34a$^{-/-}$ iPSCs. When only the ERV loci are highlighted for DE retrotransposons in the MA plot, 46 belong to the ERV-I class (red, 6%), 97 belong to the ERV-II class (blue, 12%) and 634 belong to the ERV-III glass (green, 82%). (Right) Loci from the MERVL family of ERVs are preferentially derepressed in miR-34a$^{-/-}$ iPSCs. When only ERV-III loci are highlighted for DE retrotransposons in the MA plot, we identified 552 MERVL (red, 87%), 27 MT2_Mm (blue, 4%), 1 MT2A (green, <1%), 3 MT2B (purple, <1%), 6 MT2B1 (orange, 1%), 11 MT2B2 (yellow, 2%) and 34 other ERV-III loci (brown, 5%). MT2_Mm is designated as the solo LTR of the canonical MERVL ERV; MT2A, MT2B, MT2B1, and MT2B2 refer to solo-LTRs that are related to MT2_Mm. CPM, counts per million. B, C. The extent of MERVL derepression (B) and the expanded cell fate potential (C) is decreased in late passages of miR-34a$^{-/-}$ ESCs. In contrast, MERVL derepression is more stable in miR-34a$^{-/-}$ iPSCs. The level of MERVL expression was measured using real-time PCR analyses for three (ESC) or two (iPSC) independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ cells. Error bars: s.d., n=3. C. When four GFP-labeled ESCs were microinjected into each recipient morula, 63% of early passage (P10) miR-34a$^{-/-}$ ESCs contribute to both ICM and TE (n=12/19), while only 6% of late passage (P17) miR-34a$^{-/-}$ ESCs exhibit the same phenotype (n=1/16). Scale bar, 20 mm. D. An MA-plot compares the transcription profiles of all retrotransposon loci between wild-type and miR-34a$^{-/-}$ iPSCs using the RNA-seq data. Among the 949 DE retrotransposon loci (FDR of 5%, absolute $\log_2$ fold-change of 2 or more); 600 belong to the MERVL family (highlighted in color, 63%), revealing a specific derepression of this ERV family in miR-34a$^{-/-}$ iPSCs. Interestingly, 86.5% (519) are MERVL elements with a complete structure (red); 7.5% (45) are solo LTRs (blue); and 6% (36) are truncated copies (green). CPM, counts per million. E. miR-34a$^{-/-}$ ESC cultures exhibit an increase of cells expressing MERVL-Gag but lacking Oct4, as determined by IF staining. Scale bars, 20 mm. Images shown are representative from two pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSCs. F. The MERVL expression level is heterogeneous among miR-34a$^{-/-}$ iPSC colonies. Using single-cell real-time PCR technology, we measured the MERVL expression level in individual colonies of wild-type and miR-34a$^{-/-}$ iPSCs, demonstrating the existence of colonial heterogeneity. Error bars: s.e.m., n=7-8. All P-values were calculated on a basis of a two-tailed Student's t-test. ** $P < 0.01$. G. A representative miR-34a$^{-/-}$ iPSC colony contains cells with strong MERVL-Gag expression, yet no Oct4 expression. Scale bars, 50 mm.

Fig. S3

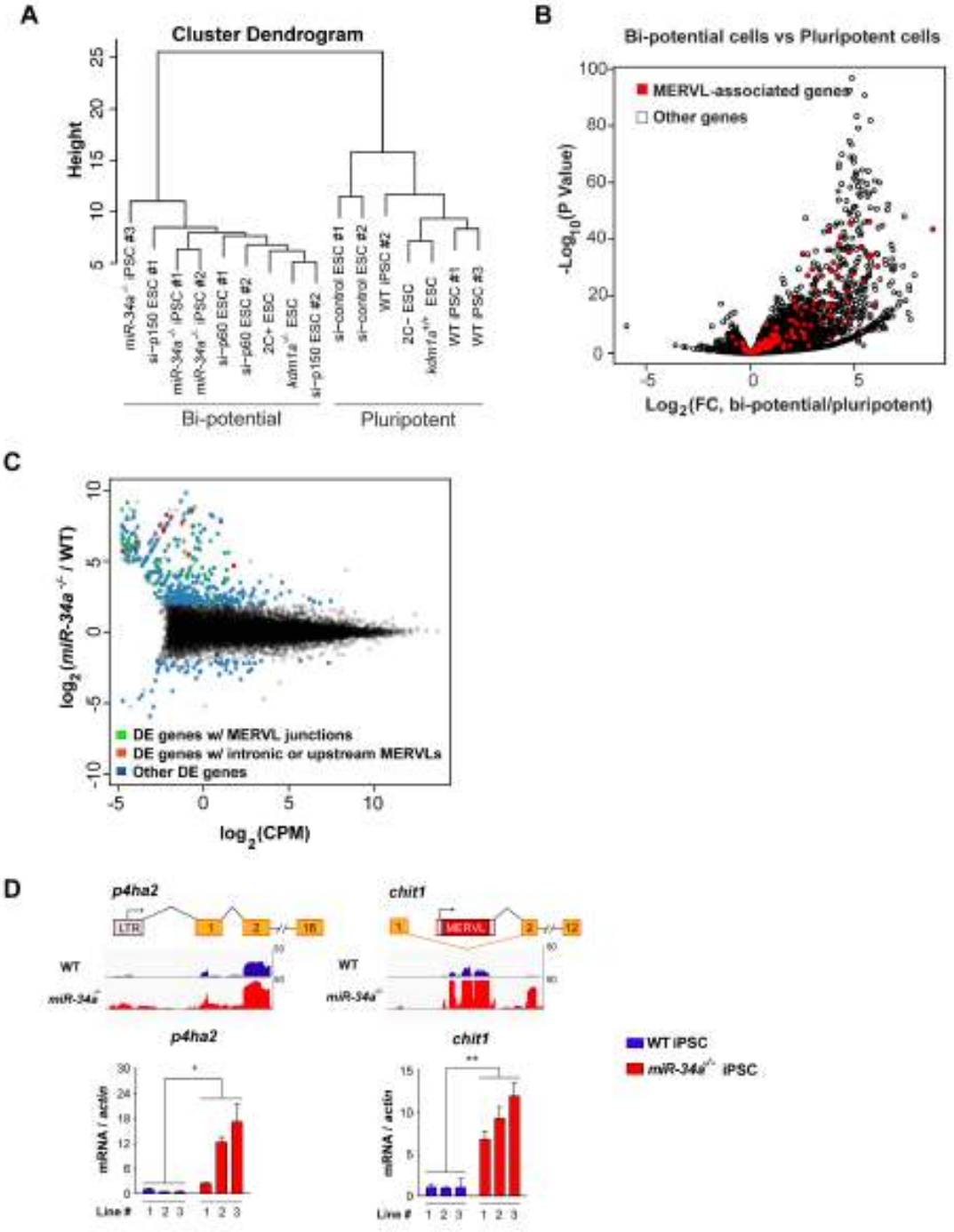

A
Cluster Dendrogram
Height
Bi-potential
Pluripotent
B
Bi-potential cells vs Pluripotent cells
MERVL-associated genes
Other genes
-Log10(P Value)
Log2(FC, bi-potential/pluripotent)
C
log2 (miR-34a-/- / WT)
DE genes w/ MERVL junctions
DE genes w/ intronic or upstream MERVLs
Other DE genes
log2(CPM)
D
p4ha2
chit1
LTR
MERVL
WT
miR-34a-/-
miRNA / actin
Line #
WT iPSC
miR-34a-/- iPSC

Fig. S3 (Cont'd)

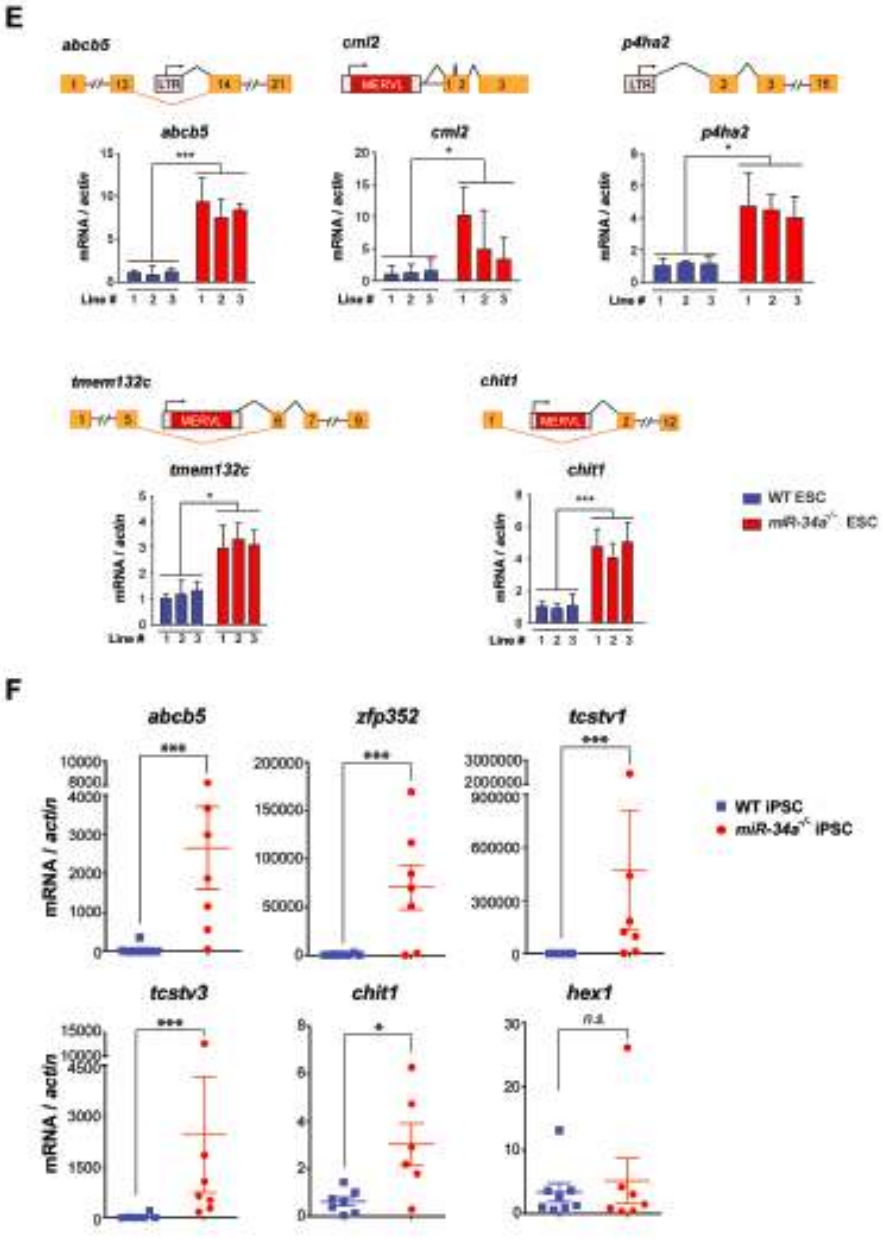

E
abcb5
cml2
p4ha2
tmem132c
chit1
mRNA / actin
Line #
WT ESC
miR-34a-/- ESC
F
abcb5
zfp352
tcstv1
tcstv3
chit1
hex1
n.s.
WT iPSC
miR-34a-/- iPSC

Fig. S3. The MERVL induction in miR-34a$^{-/-}$ pluripotent stem cells alters the expression and structure of a subset of MERVL proximal genes.

A-B. miR-34a$^{-/-}$ pluripotent stem cells and other reported totipotent-like ESCs share a similar expression profile. We reanalyzed several publicly available datasets (see Supplementary Text), and revealed a similar transcriptional profile between miR-34a$^{-/-}$ iPSC and other published totipotent-like ESCs, such as 2C+ ESCs (7), kdm1a$^{-/-}$ ESCs (25), and chromatin assembly factor-1 (CAF-1)-deficient (si-p60 and si-150) ESCs (9). A. Hierarchical clustering of published datasets (7, 9, 25) effectively clusters the expression data in two main branches: bi-potential vs. pluripotent. B. Differential expression analysis of bi-potential cells versus pluripotent stem cells using published datasets (7, 9, 25). We excluded our miR-34a$^{-/-}$ and wild-type iPSC data from this analysis to avoid the whole dataset driven by the miR-34a$^{-/-}$ iPSC data. The volcano plot shows that a large fraction of MERVL-associated genes (58/224) is differential upregulated in totipotent-like cells. FC: fold-change (bi-potential/pluripotent); CPM: counts per million. C. The differentially upregulated protein-coding genes in miR-34a$^{-/-}$ iPSCs are enriched for genes with an observed MERVL-gene junction or genes proximal to MERVL (n=50/242, Fisher's exact test, $P < 10^{-15}$). An MA-plot is shown to compare the transcription profiles of protein-coding genes between miR-34a$^{-/-}$ and wild-type iPSCs using the RNA-seq data. The DE protein-coding genes are color-coded by their relation to MERVL. Green: DE genes with an observed junction read with an adjacent MERVL element (see Supplementary Text); orange: DE genes with an upstream or intronic MERVL element; blue: all other DE genes. CPM: counts per million. D, E. Examples are shown for induced expression and altered transcript structure of MERVL proximal genes in miR-34a$^{-/-}$ iPSCs and ESCs. MERVL associated genes are induced in miR-34a$^{-/-}$ iPSCs (D) and miR-34a$^{-/-}$ ESCs (E). D. In miR-34a$^{-/-}$ iPSCs, the p4ha2 transcription is induced by an upstream MT2B element, a solo LTR related to MERVL LTR; the chit1 gene contains an intronic MERVL element in intron 1, which acts as an alternative promoter to drive a chit1 isoform containing all downstream exons. The expression level of chit1 and p4ha2 was measured using three independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC lines. Error bars: s.d., n=3. E. Real-time PCR analyses validate the induction of MERVL-gene isoforms in miR-34a$^{-/-}$ ESCs, including cml2 and p4ha2 (with an upstream MERVL element) as well as abcb5, tmem132c and chit1 (with an intronic MERVL element). Intronicly localized MERVLs, either a solo LTR (as that in intron 13 of abcb5) or a complete ERV (as that in intron 5 of tmem132c or intron 1 of chit1), act as alternative promoters to drive the expression of truncated gene isoforms that only contain the downstream exons. Three independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ ESC lines were compared. Error bars: s.d., n=3. F. Induction of MERVL proximal genes is heterogeneous among different miR-34a$^{-/-}$ iPSC colonies. Using single cell real-time PCR analyses, we measured the expression of MERVL proximal genes (abcb5, zfp352, tcstv1, tcstv3, and chit1) in individual colonies of wild-type and miR-34a$^{-/-}$ iPSCs, demonstrating the colonial heterogeneity in their expression level. Interestingly, one of a previously reported totipotency marker hex1 (8) is expressed at a similar level between wild-type and miR-34a$^{-/-}$ ESCs, suggesting that the Hex1-positive ESCs are likely a different cell population from the totipotent-like miR-34a$^{-/-}$ ESC population. Error bars: s.d., n=3. All P-values were calculated on a basis of a two-tailed Student's t-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, n.s., not significant.

Fig. S4

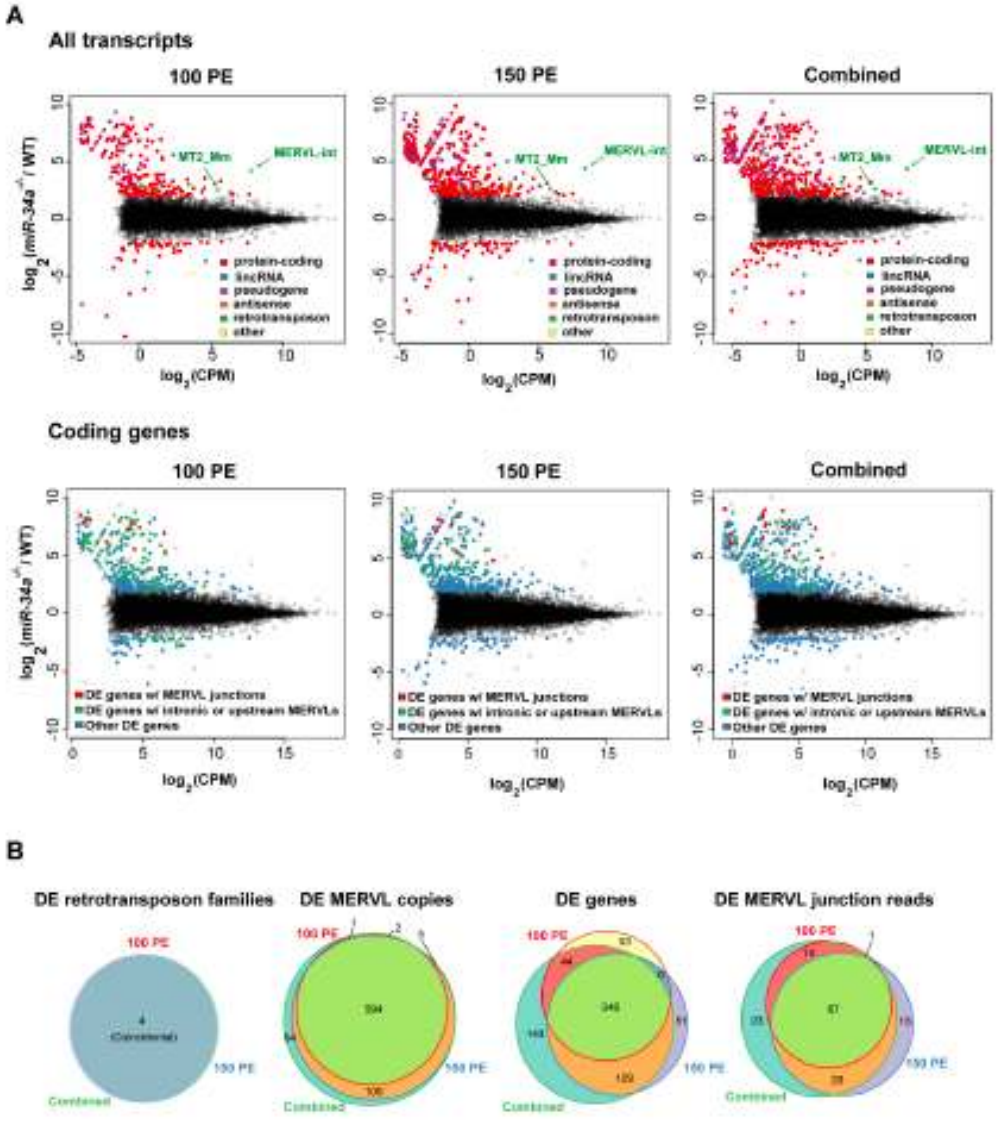

A
All transcripts
100 PE
150 PE
Combined
MT2_Mm
MERVL-int
protein-coding
lincRNA
pseudogene
antisense
retrotransposon
other
log2(miR-34a-/- / WT)
log2(CPM)
Coding genes
DE genes w/ MERVL junctions
DE genes w/ intronic or upstream MERVLs
Other DE genes
B
DE retrotransposon families
DE MERVL copies
DE genes
DE MERVL junction reads
100 PE
150 PE
Combined

Fig. S4. The effect of the length and depth of RNA-seq data on the transcriptional profile characterization of *miR-34a*$^{-/-}$ iPSCs.

We re-sequenced the wild-type and *miR-34a*$^{-/-}$ iPSC RNA-seq libraries to a greater length (150 bp paired-end (PE)) and sequencing depth (additional 192 million reads), in order to explore to what extent our previous RNA-seq analysis is limited by the technical difficulty to precisely map the repetitive RNA-seq reads. A. (Top) The MERVL ERVs are the most highly induced and differentially expressed (DE) transcriptional unit in *miR-34a*$^{-/-}$ iPSCs. MA-plots compare the transcription profiles of *miR-34a*$^{-/-}$ and wild-type (WT) iPSCs. DE transcriptional units, including protein-coding genes, long non-coding RNAs (lncRNAs), pseudogenes, antisense transcripts, and retrotransposons, are color-coded by class, as in Fig. 2A. The results of 100 bp PE RNA-seq (left, same as Fig. 2A), the results of 150 bp PE RNA-seq (middle), and the results of the combined dataset by pooling together the two sequencing datasets (right) are shown as MA plots. (Bottom) The upregulated protein-coding genes in *miR-34a*$^{-/-}$ iPSCs are enriched for genes with an observed MERVL-gene junction or genes proximal to MERVL. MA-plots compare the transcription profiles of protein-coding genes between *miR-34a*$^{-/-}$ and WT iPSC using RNA-seq data. The DE protein-coding genes are color-coded by their relation to MERVL, as in Fig. S3C. The results of 100 bp PE RNA-seq (left, same as Fig. S3C), the results of 150 bp PE RNA-seq (middle), and the results of the combined datasets (right) are shown as MA plots. B. Venn diagrams showing the concordance between the 100 bp PE and 150 bp PE RNA-seq datasets. For each of the datasets (100 bp PE, 150 bp PE, and combined), we identified DE retrotransposons families, MERVL loci, DE genes, and DE MERVL-gene junctions (see Supplementary Text). The datasets are largely in agreement, suggesting that 100 bp PE reads are sufficient to detect most DE MERVL-gene junctions and DE MERVL loci. As expected, the combined dataset has more power to detect differential expression, since we are effectively doubling the sequencing depth.

Fig. S5

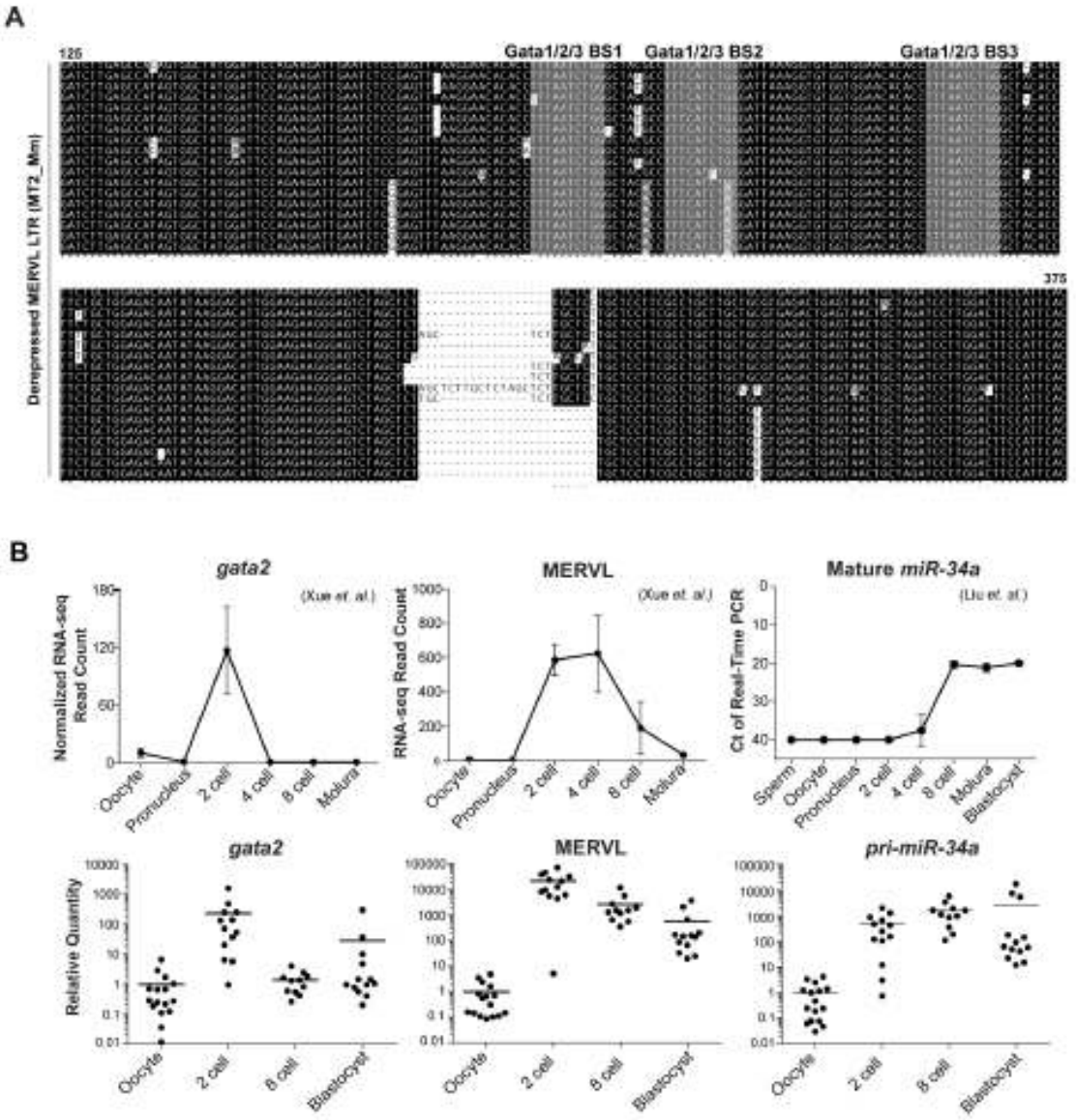

A
126
Gata1/2/3 BS1
Gata1/2/3 BS2
Gata1/2/3 BS3
Derepressed MERVL LTR (MT2_Mm)
375
B
gata2
(Xue et. al.)
Normalized RNA-seq Read Count
MERVL
(Xue et. al.)
RNA-seq Read Count
Mature miR-34a
(Liu et. al.)
Ct of Real-Time PCR
Oocyte
Pronucleus
2 cell
4 cell
8 cell
Morula
Sperm
Blastocyst
gata2
MERVL
pri-miR-34a
Relative Quantity

Fig. S5 (Cont'd)

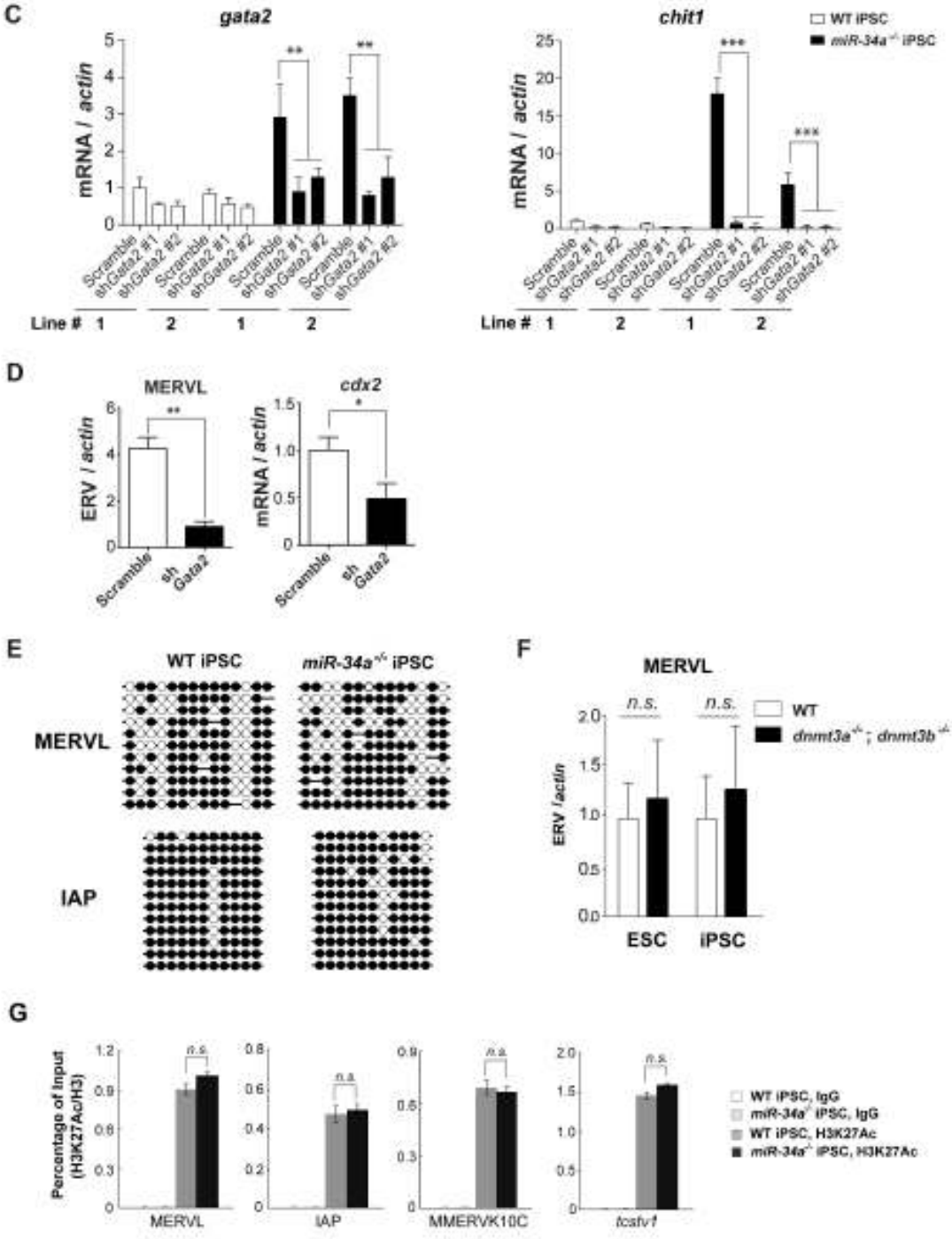

C
gata2
chit1
WT iPSC
miR-34a-/- iPSC
mRNA / actin
Scramble
shGata2 #1
shGata2 #2
Line # 1 2 1 2
D
MERVL
cdx2
ERV / actin
mRNA / actin
Scramble
sh Gata2
E
WT iPSC
miR-34a-/- iPSC
MERVL
IAP
F
MERVL
n.s.
WT
dnmt3a-/-; dnmt3b-/-
ERV / actin
ESC
iPSC
G
Percentage of Input (H3K27Ac/H3)
MERVL
IAP
MMERVK10C
tcstv1
WT iPSC, IgG
miR-34a-/- iPSC, IgG
WT iPSC, H3K27Ac
miR-34a-/- iPSC, H3K27Ac

Fig. S5 (Cont'd)

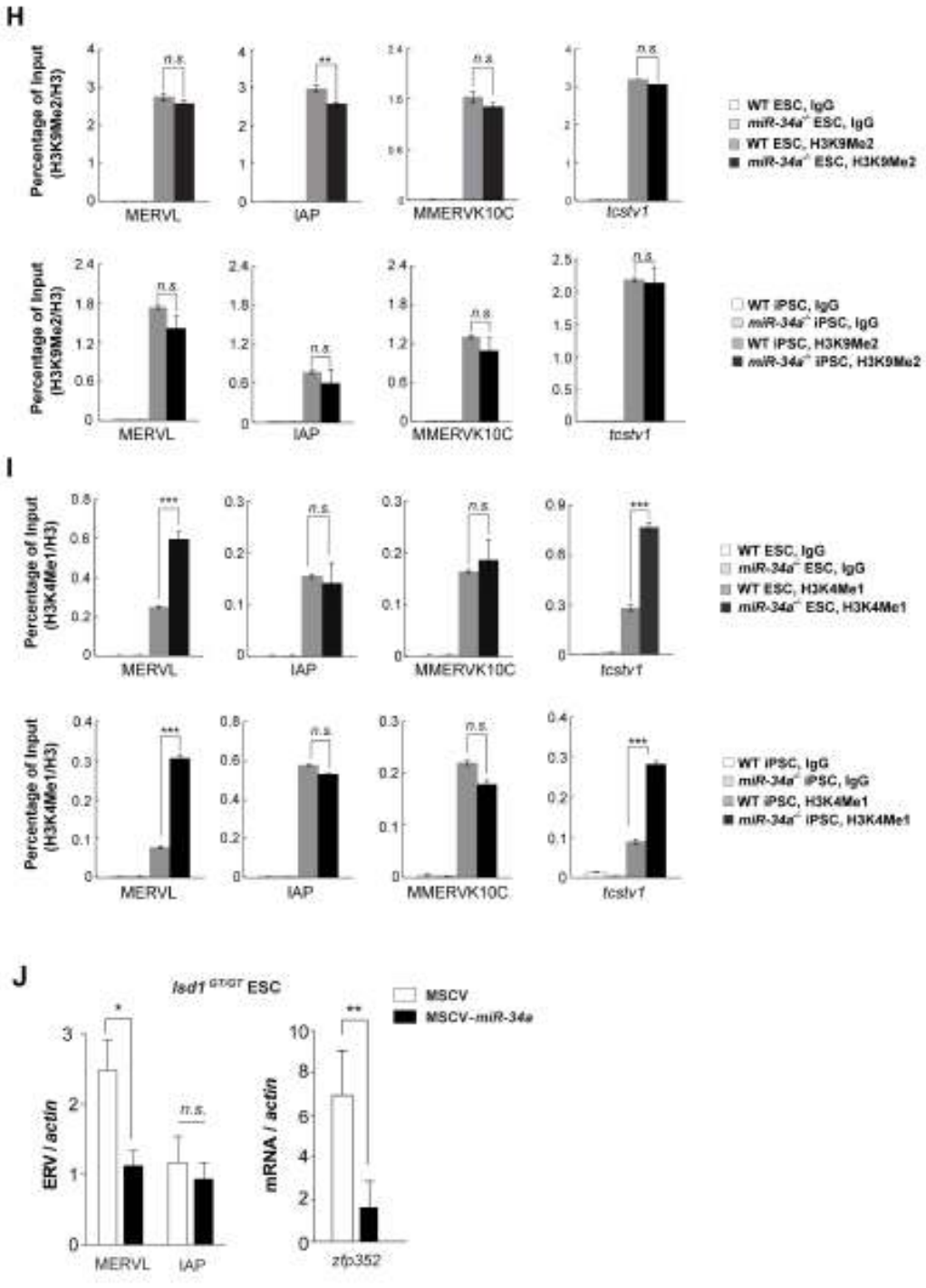

H
Percentage of Input (H3K9Me2/H3)
MERVL
IAP
MMERVK10C
tcstv1
n.s.
**
WT ESC, IgG
miR-34a-/- ESC, IgG
WT ESC, H3K9Me2
miR-34a-/- ESC, H3K9Me2
WT iPSC, IgG
miR-34a-/- iPSC, IgG
WT iPSC, H3K9Me2
miR-34a-/- iPSC, H3K9Me2
I
Percentage of Input (H3K4Me1/H3)
***
WT ESC, IgG
miR-34a-/- ESC, IgG
WT ESC, H3K4Me1
miR-34a-/- ESC, H3K4Me1
WT iPSC, IgG
miR-34a-/- iPSC, IgG
WT iPSC, H3K4Me1
miR-34a-/- iPSC, H3K4Me1
J
lsd1 GT/GT ESC
MSCV
MSCV-miR-34a
ERV / actin
mRNA / actin
zfp352
*
**

Fig. S5. Gata2 mediates the MERVL induction in miR-34a$^{-/-}$ pluripotent stem cells.

A. Clustal-W LTR sequence alignment of 18 differentially expressed MERVL loci in miR-34a$^{-/-}$ iPSCs reveals three conserved predicted Gata1/2/3 binding sites within the minimal region of MERVL LTR, MERVL$_{125\text{-}375}$. Among the three predicted GATA binding sites (designated as BS1, BS2, and BS3 and highlighted in yellow), BS1 and BS3 are fully conserved across all 18 MERVL elements, while BS2 is partially conserved. B. Expression patterns of MERVL, gata2 and miR-34a during mouse preimplantation development. In published datasets (52, 53), the levels of MERVL and gata2 both peak in 2C embryos; and the mature miR-34a is highly expressed from the 8C to the blastocyst stage (top). Using single-embryo real-time PCR analyses, we validated the expression patterns of MERVL, gata2, and pri-miR-34a in mouse preimplantation embryos (bottom). C. The expression level of the MERVL proximal genes is dependent on gata2. Two independent shRNAs against gata2 are able to effectively knock down gata2 and the MERVL proximal gene chit1 in miR-34a$^{-/-}$ iPSCs. Two independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC lines were compared. Error bars: s.d., n=3. D. gata2 is necessary for MERVL and cdx2 induction during teratoma formation. In teratomas generated from miR-34a$^{-/-}$ iPSCs, knockdown of gata2 in miR-34a$^{-/-}$ iPSCs reduces the MERVL and cdx2 levels during teratoma formation. E, F. DNA methylation is not essential for the MERVL induction in miR-34a$^{-/-}$ pluripotent stem cells. E. Wild-type and miR-34a$^{-/-}$ iPSCs have similar level of modest DNA methylation on MERVL elements, as determined by bisulfite sequencing. In contrast, iPSCs of both genotypes exhibit a high level of DNA methylation on the IAP elements. Black circle, methylated CpG; open circle, unmethylated CpG. F. No MERVL induction is detected in dnmt3a$^{-/-}$; dnmt3b$^{-/-}$ ESCs and iPSCs that are deficient for de novo DNA methylation. Error bars: s.d., n=3. n.s., not significant. G-I Characterization of epigenetic modifications on MERVL in wild-type and miR-34a$^{-/-}$ pluripotent stem cells. Wild-type and miR-34a$^{-/-}$ pluripotent stem cells have similar deposition of H3K27Ac (G) and H3K9Me2 (H) on the MERVL LTR and the MERVL-tcstv1 chimeric gene, yet the deposition of H3K4Me1 (I) on MERVL is increased in miR-34a$^{-/-}$ pluripotent stem cells. As a control, H3K27Ac (G), H3K9Me2 (H) and H3K4Me1 (I) deposition on IAP LTR or MMERVK10C LTR is similar between wild-type and miR-34a$^{-/-}$ pluripotent stem cells. Error bars: s.d., n=3. Two independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ ESCs and iPSCs are compared. J. miR-34a overexpression in lsd1 deficient (lsd1$^{GT/GT}$) ESCs using a murine stem cell virus (MSCV) retroviral vector effectively suppresses the level of MERVL and the MERVL proximal gene zfp352, but causes no alteration in IAP. Error bars: s.d., n=3. All P-values were calculated on a basis of a two-tailed Student's t-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, n.s., not significant.

Fig. S6

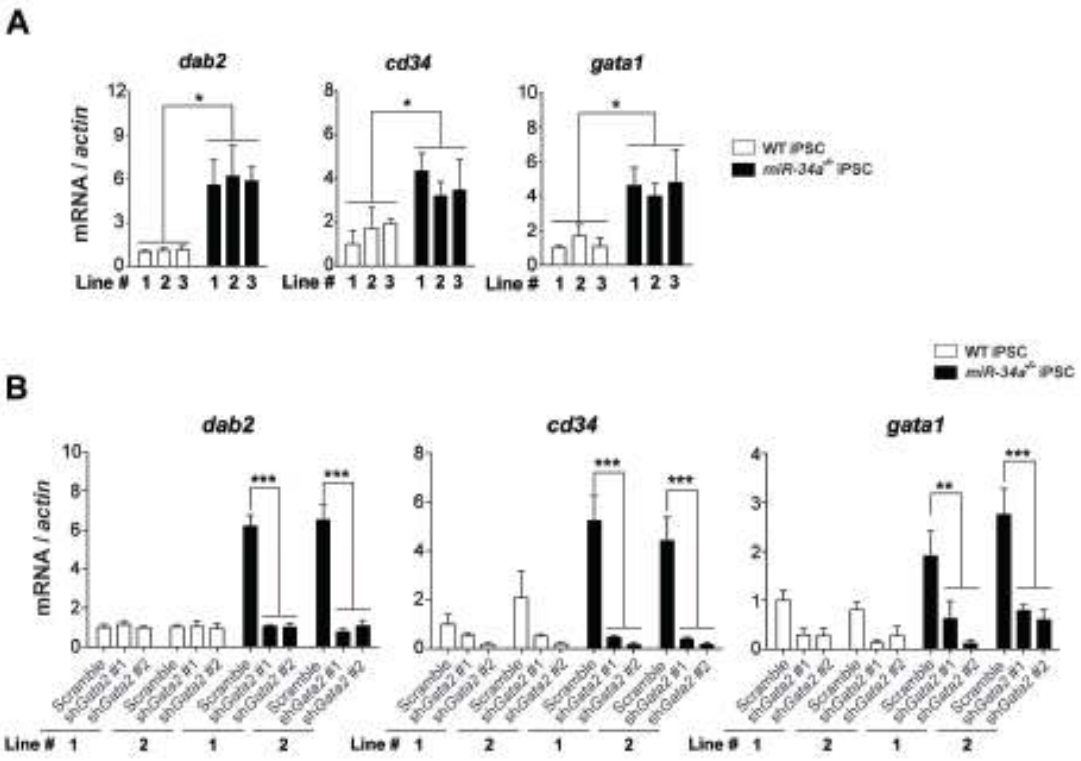

A
dab2
cd34
gata1
mRNA / actin
Line # 1 2 3 1 2 3
WT iPSC
miR-34a-/- iPSC
B
dab2
cd34
gata1
mRNA / actin
Scramble
shGata2 #1
shGata2 #2
Line # 1 2 1 2

Fig. S6. Gata2 is a key target of miR-34a in pluripotent stem cells.

A. Consistent with Gata2 being a key target for miR-34a, miR-34a$^{-/-}$ iPSCs exhibit an increase of multiple well-characterized gata2 targets (dab2, cd34 and gata1) in our real-time PCR analyses. Three independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC lines were compared. Error bars, s.d., n=3. B. The expression level of characterized gata2 targets is dependent on gata2 in miR-34a$^{-/-}$ iPSCs. Two independent gata2 shRNAs are able to effectively suppress dab2, cd34 and gata1 in miR-34a$^{-/-}$ iPSCs. Two independent pairs of passage- and littermate-controlled wild-type and miR-34a$^{-/-}$ iPSC lines were compared by real-time PCR analyses. Error bars: s.d., n=3. All P-values were calculated on a basis of a two-tailed Student's t-test. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table. S1 A summary of all chimeric analyses performed using wild-type and *miR-34a*$^{-/-}$ pluripotent stem cells [a]

A. Chimeric analyses by morula microinjection

Cell line	# of cells injected	# of injected embryos	# of blastoycsts	# of blastocysts with GFP+ cells in the ICM	# of blastocysts with GFP+ cells in the TE	# of blastocysts with GFP+ cells in ICM+TE
Wild-type ESC #1	4	12	12	12	0	0
Wild-type ESC #2	4	15	11	11	0	0
miR-34a$^{-/-}$ ESC #1	4	31	27	10	1	16
miR-34a$^{-/-}$ ESC #2	4	25	19	1	6	12
miR-34a$^{-/-}$ ESC #1	1	45	40	11	16	13
miR-34a$^{-/-}$ ESC #2	1	25	21	9	4	8
miR-34a$^{-/-}$ iPSCs w/ Scramble control	4	20	14	5	2	7
miR-34a$^{-/-}$ iPSCs w/ sh*Gata2*	4	20	13	11	2	0

B. Chimeric analyses by blastocyst microinjection

Embryo stage	ESC line	# of live embryos	# of GFP+ embryos	# of embryos w/ GFP+ embryonic tissues	# of embryos w/ GFP+ extra-embryonic tissues	# of embryos w/ GFP+ extra-embryonic placental tissues[b]	# of embryos w/ GFP+ extra-embryonic yolk sac tissues[c]
E9.5	Wild-type #1	7	4	4	0	0	0
	miR-34a$^{-/-}$ #1	11	10	10	4	3	4
E12.5	Wild-type #1	1	1	1	0	0	0
	miR-34a$^{-/-}$ #1	12	8	8	5	5	2
E14.5	Wild-type #2	10	7	7	0	0	0
	miR-34a$^{-/-}$ #2	16	15	15	12	7	9

[a] All chimera analyses are performed using passage controlled and littermate controlled wildtype and *miR-34a*$^{-/-}$ pluripotent stem cells

[b] GFP contribution to the placenta tissue is determined by the presence of GFP positive TGCs, spongiotrophoblasts or s-TGCs.

[c] GFP contribution to the yolk sac is determined by the presence of GFP positive visceral endoderm cells

Table S2: Expression quantification of all retrotransposon families in wild-type and *miR-34a*$^{-/-}$ iPSCs in RNA-seq data

CPM: counts per million; FC: fold change; WT: wild-type.

* Test for differential expression between *miR-34a*$^{-/-}$ and wild-type iPSCs

RT_FAMILY	RT_CLASS	WT CPM	*miR-34a*$^{-/-}$ CPM	$\log_2$FC (*miR-34a*$^{-/-}$/ WT)	*P* Value*
MERVL-int	LTR/ERVL	61.5344743	1213.427274	4.301681028	0
MT2_Mm	LTR/ERVL	21.22242793	125.3473841	2.564816204	1.57E-242
ERVB4_2-LTR_MM	LTR/ERVK	3.067982223	26.56770717	3.10365701	3.52E-147
MamGyp-int	LTR/Gypsy	39.72775214	94.79477821	1.25430832	2.14E-74
RLTR45-int	LTR/ERVK	63.31236563	175.6871821	1.472616655	1.92E-73
RLTR4_MM-int	LTR/ERV1	112.75341	226.8487448	1.008682793	2.13E-47
X7C_LINE	LINE/CR1	0.561927503	3.655149136	2.683031644	1.75E-39
MuRRS-int	LTR/ERV1	6.076885233	13.18256957	1.116592206	1.18E-27
MER92B	LTR/ERV1	13.32337721	6.319528673	-1.069532014	8.87E-26
MMETn-int	LTR/ERVK	1135.880309	482.6561107	-1.234914319	4.19E-24
ERVB4_2-I_MM	LTR/ERVK	8.729885059	16.59976977	0.925146461	7.81E-24
ETnERV-int	LTR/ERVK	440.6674441	220.6603116	-0.998305411	4.97E-21
RLTR45	LTR/ERVK	13.09988465	23.08570641	0.816223554	1.74E-19
L1MC2	LINE/L1	19.92879912	11.61525448	-0.784197591	2.60E-17
MYSERV6-int	LTR/ERVK	33.75342132	50.64821161	0.584294922	1.32E-16
ERVB4_1C-LTR_Mm	LTR/ERVK	63.59288423	21.45389144	-1.571060935	1.07E-15
RLTR6B_Mm	LTR/ERV1	3.685362658	9.561620702	1.369434014	1.42E-15
RLTR44C	LTR/ERVK	7.145278689	13.06741983	0.867730309	1.61E-14
RLTR6-int	LTR/ERV1	38.58567691	54.33021032	0.495279572	2.01E-13
ERVB3_1-I_MM	LTR/ERVK	1.707993958	3.94510182	1.200249506	2.36E-13
IAPLTR4	LTR/ERVK	3.001091254	6.313611227	1.049645193	4.24E-13
MamGypsy2-LTR	LTR/Gypsy	9.622304277	5.554597332	-0.79483822	1.45E-12
MLTR25A	LTR/ERVK	39.74564499	58.30395229	0.553212623	3.87E-12
Lx3B	LINE/L1	52.63205315	74.41015327	0.499125777	1.32E-11
MT2B1	LTR/ERVL	48.30916487	70.1700798	0.5373857	1.82E-11
IAPEz-int	LTR/ERVK	262.4295406	180.0654888	-0.543536565	2.36E-11
RLTR44-int	LTR/ERVK	4.571645615	7.973594603	0.787074408	8.19E-11
IAP-d-int	LTR/ERVK	13.15502434	19.68614617	0.577926748	3.16E-10
MLT1E1A	LTR/ERVL-MaLR	17.4154184	11.72116933	-0.574839919	4.48E-10
L1Md_F3	LINE/L1	56.37107839	76.80861257	0.445870643	1.02E-09
RLTR13D5	LTR/ERVK	2.974615165	5.304441396	0.835859184	2.15E-09
MLT2E	LTR/ERVL	0.36466038	1.169973827	1.688327401	4.42E-09

RLTR13G	LTR/ERVK	15.90105174	10.21864601	-0.633775555	6.74E-09
L1MEd	LINE/L1	33.39386169	44.61902729	0.41850849	6.86E-09
MamSINE1	SINE/tRNA-RTE	93.64805998	70.75162486	-0.404162454	1.05E-08
RLTRETN_Mm	LTR/ERVK	52.60316333	34.06998488	-0.62952437	2.62E-08
ETnERV2-int	LTR/ERVK	8.872732687	14.02958415	0.659454327	2.74E-08
Lx2A1	LINE/L1	3.68321392	6.28392368	0.772227493	3.47E-08
MER89	LTR/ERV1	3.657598753	6.028137548	0.709659825	3.88E-08
RLTR10D	LTR/ERVK	4.877443406	7.632608254	0.648563086	3.90E-08
RLTR12BD_Mm	LTR/ERVK	24.20578737	17.00619763	-0.50562603	7.00E-08
MERV1_I-int	LTR/ERV1	22.35119124	31.3139192	0.48356735	1.96E-07
RLTR49	LTR/ERVK	13.62320161	8.801454438	-0.626857468	4.73E-07
L1ME4a	LINE/L1	15.9891478	10.44246576	-0.607704176	5.33E-07
RLTR9E	LTR/ERVK	68.67563417	90.65588023	0.4001975	8.34E-07
RLTR23	LTR/ERV1	37.12396706	27.25902075	-0.441306396	9.11E-07
HERVL40-int	LTR/ERVL	4.731676161	2.458570754	-0.956854896	9.41E-07
Lx3_Mus	LINE/L1	45.73903624	60.9459896	0.413769444	1.03E-06
RLTR10D2	LTR/ERVK	1.601319309	2.824646521	0.821325882	1.80E-06
MER57D	LTR/ERV1	3.55090093	1.930937504	-0.876743772	1.88E-06
RMER16C	LTR/ERVK	14.72414527	9.505458563	-0.637790494	2.15E-06
RLTR31D_MM	LTR/ERVK	68.35976393	55.15479494	-0.309126298	2.38E-06
ERVB4_1B-LTR_MM	LTR/ERVK	61.13965226	77.77208995	0.348199979	3.45E-06
L1ME3G	LINE/L1	7.00445261	4.579776735	-0.603586702	4.04E-06
LTR53B	LTR/ERVL	4.208238582	2.669239916	-0.652390759	4.48E-06
MMERGLN_LTR	LTR/ERV1	33.94322746	24.2057027	-0.490979267	4.64E-06
X7A_LINE	LINE/CR1	20.7578552	13.42906142	-0.622125639	5.07E-06
MMERGLN-int	LTR/ERV1	1767.068178	1357.46491	-0.38050749	6.50E-06
MER66-int	LTR/ERV1	0.863813257	1.802941383	1.071130726	7.58E-06
LTR78B	LTR/ERV1	59.98722739	74.67613202	0.317055358	1.55E-05
LTR16A1	LTR/ERVL	0.847065849	1.772109511	1.082162116	1.75E-05
LTR103b_Mam	LTR/ERV1?	0.625774343	1.320919211	1.124305069	2.16E-05
RMER17D	LTR/ERVK	20.66314247	27.11998139	0.391701814	3.42E-05
B1_Mur3	SINE/Alu	762.5629308	624.187827	-0.288878644	3.76E-05
Lx7	LINE/L1	153.964323	124.9166959	-0.301198595	4.11E-05
RMER13A2	LTR/ERVK	31.35812743	38.17637009	0.281823401	4.31E-05
RLTR27	LTR/ERVK	3.530223643	5.18371288	0.54776716	4.53E-05
L1Md_Gf	LINE/L1	5.640095083	8.319666995	0.553311853	5.34E-05
B2_Mm2	SINE/B2	1154.621811	905.238155	-0.351018666	6.80E-05
RLTR9A3A	LTR/ERVK	1.268961326	2.114716236	0.740905133	8.10E-05

Table S3: Expression quantitation of individual MERVL loci and related loci in wild-type and *miR-34a*$^{-/-}$ iPSCs in RNA-seq data

% Divergence, % deletion and % insertion include data from Repbase when comparing each MERVL or related locus and the corresponding Repbase consensus sequence.

CPM: counts per million; WT: wild-type; FC: fold change.

* Test for differential expression between *miR-34a*$^{-/-}$ and wild-type iPSCs

Coordinates of MERVL loci	Family	Structure	Length	% Divergence	% Deletion	% Insertion	Log_2FC (*miR-34a*$^{-/-}$ / WT)
chr12:83624988-83631434:+	MT2_Mm	complete	6446	4.2	1.12	1.18	3.514301014
chr7:11169946-11176384:+	MT2_Mm	complete	6438	4.56	1.04	0.78	5.647252993
chr5:62772576-62775853:-	MT2_Mm	complete	3277	1.78	0.56	0.2	6.160274387
chr7:13186591-13192986:-	MT2_Mm	complete	6395	4.14	1.28	0.7	5.437707033
chr3:96200245-96206681:-	MT2_Mm	complete	6436	4.2	1.12	0.78	6.654981299
chr11:83071002-83077434:-	MT2_Mm	complete	6432	0.85	0.175	0	6.610572105
chr9:67198624-67205005:+	MT2_Mm	complete	6381	4.12	3.54	0.7	5.543811836
chr13:76077211-76083662:-	MT2_Mm	complete	6451	4.32	1	0.7	4.631055098
chr14:45748430-45754862:+	MT2_Mm	complete	6432	4.16	1.08	0.72	7.31264545
chr1:82396584-82403045:-	MT2_Mm	complete	6461	3.32	0.58	0.8	6.079522897
chr9:122328594-122335061:-	MT2_Mm	complete	6467	3.18	0.38	0.8	3.929366301
chr5:67037609-67044038:+	MT2_Mm	complete	6429	3.3	0.52	1	5.582157817
chr19:44218506-44224942:+	MT2_Mm	complete	6436	3.28	0.38	0.8	5.285869781
chr12:23033920-23040440:+	MT2_Mm	complete	6520	4.06	0.96	2.98	6.219576956
chr11:72326538-72332964:+	MT2_Mm	complete	6426	1.35	0.3	0	6.222553402
chr5:70542295-70548729:-	MT2_Mm	complete	6434	4.26	1.12	0.7	5.647625232
chr7:21244735-21251204:-	MT2_Mm	complete	6469	5.08	0.92	0.7	7.069215618
chr3:62356899-62363368:-	MT2_Mm	complete	6469	4.1	0.92	0.7	4.35858159
chr4:45289987-45305097:-	MT2_Mm	complete	15110	5.566666667	1.7	1.266666667	5.608340382
chr4:143470783-143481089:-	MT2_Mm	complete	10306	7.75	3.483333333	1.2	4.651725584
chr13:114039988-114046421:+	MT2_Mm	complete	6433	4.76	1.2	0.7	4.323091538
chr11:33062306-33068740:+	MT2_Mm	complete	6434	4.58	1.12	0.7	5.37612561
chr5:94371096-94377526:+	MT2_Mm	complete	6430	3.78	0.42	0.8	6.132921687
chr6:73737238-73743703:-	MT2_Mm	complete	6465	3.4	0.54	1.06	5.184816059
chr5:113383853-113390286:-	MT2_Mm	complete	6433	4.26	1.12	0.7	4.139328147
chr13:17932086-17938482:+	MT2_Mm	complete	6396	4.22	1.24	0.7	4.867596782
chr5:146741734-146748164:+	MT2_Mm	complete	6430	4.04	1.28	0.7	5.799094929
chr19:61168774-61175244:-	MT2_Mm	complete	6470	4.3	0.92	0.7	3.754596061
chrX:13317610-13324040:-	MT2_Mm	complete	6430	4.18	0.4	0.74	5.541256972
chr5:122383278-122389751:-	MT2_Mm	complete	6473	3.94	0.34	2.2	4.700750008
chr15:81145468-81151865:+	MT2_Mm	complete	6397	3.38	0.54	0.8	4.466450151
chr9:8048218-8054662:+	MT2_Mm	complete	6444	4.74	1.2	1.18	5.513482681
chr8:124846028-124852464:-	MT2_Mm	complete	6436	4.28	1.04	0.7	2.64812077
chr5:127537960-127541824:+	MT2_Mm	complete	3864	4.2	0.833333333	1.216666667	6.938647393
chr17:33867142-33873611:+	MT2_Mm	complete	6469	4.2	0.92	0.7	4.777705615

chr13:7068416-7074796:-	MT2_Mm	complete	6380	4.54	3.56	0.7	6.400111522
chr2:34929316-34935764:+	MT2_Mm	complete	6448	4.46	1.04	1.18	5.089851303
chrX:20033944-20040353:+	MT2_Mm	complete	6409	4.16	1.66	0.7	6.400498085
chr5:123547625-123554025:+	MT2_Mm	complete	6400	4.18	1.2	0.72	4.028989236
chr9:94821689-94828083:+	MT2_Mm	complete	6394	3.24	0.62	0.8	5.696152442
chr8:95114678-95121111:-	MT2_Mm	complete	6433	4.22	1.14	0.78	3.474817543
chr15:64591754-64598211:-	MT2_Mm	complete	6457	0.95	0.175	1.2	2.962503294
chr13:62726497-62732966:-	MT2_Mm	complete	6469	4.08	0.92	0.7	4.737338167
chr13:97726504-97732964:-	MT2_Mm	complete	6460	4.42	1.04	1.66	4.378503352
chr12:19696552-19701686:-	MT2_Mm	complete	5134	2.0375	0.05	1.125	6.852541136
chr7:73672203-73678713:-	MT2_Mm	complete	6510	3.04	0.52	1.2	9.480339806
chr15:102596789-102603186:+	MT2_Mm	complete	6397	4.22	1.2	0.7	5.974552422
chr1:144674371-144680805:+	MT2_Mm	complete	6434	3.62	0.46	0.8	6.850015853
chr9:90808146-90814576:-	MT2_Mm	complete	6430	4.14	1.08	0.7	5.034498636
chr4:107506471-107512917:-	MT2_Mm	complete	6446	3.94	1.12	1.18	4.842387769
chr3:24399196-24405593:+	MT2_Mm	complete	6397	3.96	0.54	0.8	4.695490778
chr14:59301928-59308330:-	MT2_Mm	complete	6402	4.22	1.2	0.72	4.143044968
chr4:86892645-86896219:-	MT2_Mm	complete	3574	4.116666667	1.516666667	0.6	6.8083275
chr12:19931733-19938183:-	MT2_Mm	complete	6450	3.6	0.68	1.86	5.935944337
chr18:74749909-74756343:-	MT2_Mm	complete	6434	4.52	1.12	0.7	6.300679946
chr13:12653183-12659616:+	MT2_Mm	complete	6433	4.2	1.14	0.7	5.9222435
chr2:113095789-113102258:+	MT2_Mm	complete	6469	4.7	0.92	0.7	5.165792718
chr8:107732118-107738587:+	MT2_Mm	complete	6469	4.18	0.92	0.7	5.595372807
chr18:81299741-81306177:-	MT2_Mm	complete	6436	4.14	1.04	0.7	4.282376323
chr17:23686020-23692350:-	MT2_Mm	complete	6330	4.2	5.88	0.7	3.593243403
chr4:144006213-144012649:-	MT2_Mm	complete	6436	3.98	1.04	0.7	6.266543458
chr12:72557482-72563914:-	MT2_Mm	complete	6432	4.22	1.14	0.7	5.138318948
chr5:22566117-22572514:+	MT2_Mm	complete	6397	4.12	1.28	0.7	3.893788609
chr2:140345880-140352349:-	MT2_Mm	complete	6469	4.1	0.92	0.7	6.767489173
chr12:76755591-76762027:-	MT2_Mm	complete	6436	4.42	1.04	0.7	5.306698297
chr14:64504787-64511194:+	MT2_Mm	complete	6407	4.14	1.28	1.18	5.33127887
chr4:74434630-74441057:+	MT2_Mm	complete	6427	4.7	1.08	0.7	5.330581463
chr18:58992838-58999274:-	MT2_Mm	complete	6436	4.52	1.04	0.7	6.74749798
chr17:41657796-41664230:-	MT2_Mm	complete	6434	3.94	1.12	0.7	4.942679724
chr8:13860724-13867158:-	MT2_Mm	complete	6434	4.44	1.12	0.7	5.580156693
chr12:37064708-37071132:+	MT2_Mm	complete	6424	4.36	1.46	0.7	4.784666616
chr1:129394080-129400395:-	MT2_Mm	complete	6315	1.725	0.325	0.125	4.367338094
chr5:27985195-27991273:-	MT2_Mm	complete	6078	3.883333333	0.783333333	0.666666667	6.247394429
chr5:69771812-69778246:+	MT2_Mm	complete	6434	4.58	1.12	0.7	6.239480417
chr5:116147050-116153484:-	MT2_Mm	complete	6434	4.3	1.12	0.7	3.640806692
chr17:83140245-83146681:+	MT2_Mm	complete	6436	4.56	1.04	0.7	6.237973898
chr1:87741515-87747759:-	MT2_Mm	complete	6244	3.916666667	0.916666667	0.583333333	4.920402297
chr5:94847059-94853528:+	MT2_Mm	complete	6469	5.04	0.92	0.7	7.51918194
chr5:95170190-95176342:+	MT2_Mm	complete	6152	5.12	0.92	0.7	6.230045818

Table S4: Genes differentially expressed between wild-type and *miR-34a*$^{-/-}$ iPSCs in RNA-seq data (combined)

CPM: counts per million; WT: wild-type; FC: fold change.

* Test for differential expression between *miR-34a*$^{-/-}$ and wild-type iPSCs

Gene Symbol	Ensembl ID	Gene Type	WT CPM	*miR-34a*$^{-/-}$ CPM	Log_2FC (*miR-34a*$^{-/-}$ / WT)	*p* Value*
Gm21542	ENSMUSG00000098010	pseudogene	0	5.533581642	10.74492173	1.72E-96
AF067061	ENSMUSG00000095071	protein_coding	0	4.164728715	10.33942014	4.15E-92
Gm8994	ENSMUSG00000094973	protein_coding	0	4.065571232	10.29525326	4.83E-94
AA623943	ENSMUSG00000081478	pseudogene	0	3.842826331	10.22128408	1.57E-90
Gm13083	ENSMUSG00000066688	protein_coding	0	3.315054402	10.00016241	1.10E-84
Gm5698	ENSMUSG00000086151	pseudogene	0.0072015	8.128811858	9.433317937	1.99E-163
Gm10424	ENSMUSG00000096066	protein_coding	0	1.868043748	9.174302322	8.26E-49
Gm11238	ENSMUSG00000095935	protein_coding	0	1.756342442	9.103962547	1.31E-39
Gm6502	ENSMUSG00000095718	protein_coding	0	1.666090447	9.001238982	7.46E-47
Gm21818	ENSMUSG00000095653	protein_coding	0.0188789	9.699973163	8.903744247	1.80E-149
Gm11487	ENSMUSG00000066137	protein_coding	0	1.495003758	8.848559132	6.86E-41
Gm10696	ENSMUSG00000074424	protein_coding	0	1.461642387	8.816703798	6.48E-37
Gm3176	ENSMUSG00000094043	protein_coding	0	1.439048752	8.788807831	4.49E-41
Gm11756	ENSMUSG00000093962	protein_coding	0	1.382929318	8.738865971	9.12E-41
Gm4302	ENSMUSG00000091101	protein_coding	0	1.093045837	8.428744206	2.76E-25
BC147527	ENSMUSG00000094796	protein_coding	0.0072015	3.92936949	8.382965816	5.29E-90
Gm3987	ENSMUSG00000094412	pseudogene	0	1.060299747	8.364104123	8.72E-26
Tmem92	ENSMUSG00000075610	protein_coding	0.0850616	27.64655647	8.287345604	0
Gm14742	ENSMUSG00000085973	lincRNA	0	0.974560944	8.241599068	2.22E-30
Gm5117	ENSMUSG00000093862	protein_coding	0	0.935386344	8.160228846	6.90E-27
Zscan4c	ENSMUSG00000054272	protein_coding	0.0692936	20.28365203	8.143363742	1.68E-201
AF067063	ENSMUSG00000094237	protein_coding	0	0.871444237	8.095401258	8.34E-25
Olfr376	ENSMUSG00000063881	protein_coding	0	0.87633908	8.081428893	2.93E-27
Gm4340	ENSMUSG00000090854	protein_coding	0.0072015	3.092959244	8.04871043	5.58E-73
C86695	ENSMUSG00000079112	protein_coding	0	0.835752535	8.040104608	3.17E-25
Gm2056	ENSMUSG00000095717	protein_coding	0	0.847170288	8.02485789	5.56E-26
Gm4023	ENSMUSG00000093420	pseudogene	0	0.839512043	8.021318127	3.10E-24
Gm11546	ENSMUSG00000083227	pseudogene	0.0072015	2.964356085	7.987635594	8.91E-70
Tcstv3	ENSMUSG00000095821	protein_coding	0.007884	2.881181794	7.942275226	5.64E-73
Gm16522	ENSMUSG00000094560	protein_coding	0.0504147	13.20456496	7.917680847	1.04E-183
Usp17ld	ENSMUSG00000057321	protein_coding	0.0188789	4.885193105	7.913943012	2.02E-107
Gm6763	ENSMUSG00000091779	protein_coding	0.0116774	2.750915279	7.879196007	4.74E-60
Zscan4f	ENSMUSG00000070828	protein_coding	0.1163004	26.87330119	7.868673399	1.47E-127
Gm13078	ENSMUSG00000046435	protein_coding	0.0501177	10.67102471	7.858391467	7.38E-199

Gm11232	ENSMUSG00000066141	protein_coding	0	0.751331002	7.853348438	5.43E-22
Zscan4e	ENSMUSG00000095936	protein_coding	0.0236519	6.674653748	7.852423156	3.66E-141
Gm13040	ENSMUSG00000070616	protein_coding	0	0.739774294	7.8170672	4.15E-14
Gm11543	ENSMUSG00000081007	pseudogene	0.0312388	6.479571018	7.814446629	9.57E-119
Gm7942	ENSMUSG00000092166	protein_coding	0.0373724	10.26028173	7.802450762	6.05E-189
Clec10a	ENSMUSG00000000318	protein_coding	0	0.716886542	7.794900783	2.31E-23
Usp17lc	ENSMUSG00000058976	protein_coding	0.0717222	15.55836669	7.76722619	4.62E-173
Gm4778	ENSMUSG00000089696	protein_coding	0.007884	2.539085583	7.761369935	1.15E-64
Gm16429	ENSMUSG00000096230	protein_coding	0.0366899	9.954847851	7.76085456	1.33E-194
Gm5989	ENSMUSG00000094300	pseudogene	0.0072015	2.562986673	7.759359432	6.48E-61
Gm21761	ENSMUSG00000096175	protein_coding	0.0535257	11.6435412	7.741842776	5.39E-208
Tcstv1	ENSMUSG00000096284	protein_coding	0.0312388	6.166019604	7.735079018	3.71E-118
Gm2022	ENSMUSG00000071217	pseudogene	0	0.691061419	7.729380159	3.86E-15
Gm8332	ENSMUSG00000095799	protein_coding	0.0072015	2.496826967	7.718510252	1.25E-44
Gm16026	ENSMUSG00000079455	pseudogene	0.0116774	2.43785817	7.704252886	5.02E-40
Gm4858	ENSMUSG00000096879	protein_coding	0.0116774	2.421536584	7.688292222	1.98E-61
Gm11236	ENSMUSG00000096700	protein_coding	0	0.66224853	7.684901437	6.62E-22
Tdpoz4	ENSMUSG00000060256	protein_coding	0	0.647400123	7.675966687	3.33E-20
Gm8038	ENSMUSG00000091981	pseudogene	0.0116774	2.406905126	7.671087325	2.33E-46
Gm12625	ENSMUSG00000094005	pseudogene	0	0.646967053	7.649328857	5.82E-19
BC080695	ENSMUSG00000070618	protein_coding	0.0490498	10.7120006	7.61681381	9.28E-159
Gm5939	ENSMUSG00000083258	pseudogene	0.0072015	2.285309918	7.61388108	6.45E-56
AA792892	ENSMUSG00000073497	protein_coding	0.0294884	7.169218644	7.586131479	5.42E-115
BB287469	ENSMUSG00000079031	protein_coding	0	0.625369523	7.564652009	2.90E-14
Gm6803	ENSMUSG00000096803	protein_coding	0.0072015	2.184185423	7.529893406	1.25E-56
Gm11239	ENSMUSG00000095048	protein_coding	0	0.591462624	7.524985632	2.20E-19
B020031M17Rik	ENSMUSG00000078537	protein_coding	0.1763451	31.4527565	7.498399249	1.397612898
Gm13119	ENSMUSG00000070619	protein_coding	0.0360074	8.219358245	7.486893132	2.54E-158
Ly6g	ENSMUSG00000022582	protein_coding	0	0.569337073	7.461601	7.63E-19
Gm13043	ENSMUSG00000095409	protein_coding	0	0.569969585	7.444191265	9.26E-15
Gm11425	ENSMUSG00000082419	pseudogene	0.0222869	5.015597001	7.437653281	3.76E-93
Gm11757	ENSMUSG00000096750	protein_coding	0	0.554436421	7.434745624	6.58E-17
Gm6902	ENSMUSG00000094682	protein_coding	0	0.531514205	7.369861246	4.19E-17
Vmn2r12	ENSMUSG00000090688	protein_coding	0	0.516215221	7.342274043	1.20E-16
Usp17lb	ENSMUSG00000062369	protein_coding	0.0727017	12.8628946	7.332550625	3.38E-206
Pramef6	ENSMUSG00000078512	protein_coding	0.014403	3.277595128	7.331254183	1.68E-80
Gm21731	ENSMUSG00000063846	protein_coding	0	0.506027478	7.330872709	1.46E-15
Gm27205	ENSMUSG00000098832	pseudogene	0.0229694	4.586638316	7.321685414	6.53E-80

Table S5: Quantitation of chimeric junction reads between MERVL or MERVL-related loci and proximal genes in wild-type and *miR-34a*$^{-/-}$ iPSCs

RT: retrotransposon; CPM: counts per million; WT: wild-type; FC: fold change.

* Test for differential expression between *miR-34a*$^{-/-}$ and wild-type iPSCs

Gene Symbol	Gene Type	Gene Strand	RT Family	RT class	RT Structure	RT Coordinates	log_2FC (*miR-34a*$^{-/-}$/ WT)
Tmem132c	protein_coding	+	MT2_Mm	LTR/ERVL	complete	chr5:127537960-127541824:+	11.46609761
Limch1	protein_coding	+	MT2_Mm	LTR/ERVL	complete	chr5:67037609-67044038:+	9.983892081
Aqr	protein_coding	-	MT2_Mm	LTR/ERVL	complete	chr2:114180872-114187339:-	9.800226786
A530040E14Rik	antisense	+	MT2_Mm	LTR/ERVL	truncated	chr1:85106228-85109031:+	9.527660259
C130026I21Rik	protein_coding	-	MT2_Mm	LTR/ERVL	truncated	chr1:85106228-85109031:+	9.527659504
A530032D15Rik	protein_coding	-	MT2_Mm	LTR/ERVL	truncated	chr1:85106228-85109031:+	9.527642537
B020004J07Rik	protein_coding	-	MT2_Mm	LTR/ERVL	solo	chr4:101843830-101844302:-	9.285655204
Gm12114	lincRNA	+	MT2_Mm	LTR/ERVL	complete	chr11:33062306-33068740:+	8.538015608
Arhgap8	protein_coding	+	MT2_Mm	LTR/ERVL	complete	chr15:84746573-84752968:+	8.473487756
Kifc3	protein_coding	-	MT2_Mm	LTR/ERVL	complete	chr8:95114678-95121111:-	8.131282972
Bola1	protein_coding	-	MT2_Mm	LTR/ERVL	complete	chr3:96200245-96206681:-	8.087166203
Gm20765	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr10:104185230-104185716:+	8.015759915
Gm6763	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr10:104142684-104143170:+	7.923881383
Gm4340	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr10:104193739-104194225:+	7.909265075
Gm6763	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr10:104151194-104151680:+	7.870201415
Gm6763	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr10:104151194-104151680:+	7.870158166
Gm21312	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr10:104176721-104177207:+	7.844302475
Gm8764	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr10:104159703-104160189:+	7.82786612
Gm14742	lincRNA	-	MT2_Mm	LTR/ERVL	solo	chrX:77803090-77803582:-	7.751140244
Gm16026	pseudogene	+	MT2_Mm	LTR/ERVL	truncated	chr1:85266398-85269727:+	7.706472424
C130026I21Rik	protein_coding	-	MT2_Mm	LTR/ERVL	truncated	chr1:85266398-85269727:+	7.706468753
Gm21304	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr10:104168212-104168698:+	7.39416661
AU019990	lincRNA	+	MT2_Mm	LTR/ERVL	solo	chr2:132645643-132646135:+	7.265326658
Arsk	protein_coding	-	MT2_Mm	LTR/ERVL	complete	chr13:76077211-76083662:-	7.068406993
Prex2	protein_coding	+	MT2_Mm	LTR/ERVL	complete	chr1:11257740-11264173:+	7.039844934
Cln6	protein_coding	+	MT2_Mm	LTR/ERVL	complete	chr9:62861832-62868299:-	6.927696195
Calml4	protein_coding	+	MT2_Mm	LTR/ERVL	complete	chr9:62861832-62868299:-	6.927694728
Gm10653	pseudogene	-	MT2_Mm	LTR/ERVL	complete	chr9:62861832-62868299:-	6.92769278
Ppig	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr2:69728872-69729370:+	6.776649315
Pemt	protein_coding	-	MT2_Mm	LTR/ERVL	solo	chr11:60040027-60040494:-	6.46367756
Ddit4l	protein_coding	+	MT2_Mm	LTR/ERVL	solo	chr3:137621221-137621696:+	6.247839539
Abcb5	protein_coding	-	MT2_Mm	LTR/ERVL	solo	chr12:118924403-118924960:-	5.241922708
Dcaf17	protein_coding	+	MT2_Mm	LTR/ERVL	complete	chr2:71064812-71071246:-	4.626997411
Dhtkd1	protein_coding	-	MT2_Mm	LTR/ERVL	solo	chr2:5924600-5925150:-	3.815720697
Rbm25	protein_coding	+	MT2_Mm	LTR/ERVL	complete	chr12:83624988-83631434:+	3.593480454

Arih2	protein_coding	-	MT2_Mm	LTR/ERVL	solo	chr9:108653495-108653928:-	3.03632234
Slc16a6	protein_coding	-	MT2_Mm	LTR/ERVL	solo	chr11:109466198-109466685:-	-0.281347296
Gm25540	miRNA	-	MT2_Mm	LTR/ERVL	solo	chr11:109466198-109466685:-	-0.428445988
Dhx33	protein_coding	-	MT2_Mm	LTR/ERVL	solo	chr11:71000803-71001242:+	-2.803464926
Abcc5	protein_coding	-	MT2A	LTR/ERVL	solo	chr16:20379035-20379434:+	2.290609778
Zfp868	protein_coding	-	MT2A	LTR/ERVL	solo	chr8:69710660-69711051:-	0.380940563
Zfp868	protein_coding	-	MT2A	LTR/ERVL	solo	chr8:69619231-69619738:-	0.212033171
Zfp869	protein_coding	-	MT2A	LTR/ERVL	solo	chr8:69710660-69711051:-	0.147382801
Gm26758	lincRNA	+	MT2A	LTR/ERVL	solo	chr13:65825591-65825827:-	-0.21246167
1110018N20Rik	processed_transc	+	MT2A	LTR/ERVL	solo	chr2:167191735-167192261:+	-0.384170625
Ptgis	protein_coding	-	MT2A	LTR/ERVL	solo	chr2:167191735-167192261:+	-0.384176187
Dpep3	protein_coding	-	MT2A	LTR/ERVL	solo	chr8:105973920-105974668:+	-0.408725248
R74862	processed_transc	-	MT2A	LTR/ERVL	solo	chr7:143044421-143044941:-	-0.411852998
R74862	processed_transc	-	MT2A	LTR/ERVL	solo	chr7:143033544-143033843:-	-0.414704588
Cd81	protein_coding	+	MT2A	LTR/ERVL	solo	chr7:143044421-143044941:-	-0.499390557
Gsr	protein_coding	+	MT2A	LTR/ERVL	solo	chr8:33675652-33675905:-	-0.929286406
Hat1	protein_coding	+	MT2A	LTR/ERVL	solo	chr2:71387445-71388329:+	-0.930093018
Cmc1	protein_coding	-	MT2A	LTR/ERVL	solo	chr9:118069014-118069462:+	-0.941615254
Azi2	protein_coding	+	MT2A	LTR/ERVL	solo	chr9:118069014-118069462:+	-0.941623837
Mtmr2	protein_coding	+	MT2A	LTR/ERVL	solo	chr9:13783156-13783676:-	-0.949461802
Ifitm1	protein_coding	+	MT2A	LTR/ERVL	solo	chr7:140967128-140967588:+	-0.959896675
Pot1a	protein_coding	-	MT2A	LTR/ERVL	solo	chr6:25800255-25800523:+	-0.990259623
Gm12827	lincRNA	+	MT2A	LTR/ERVL	solo	chr4:117502714-117503219:+	-2.116880023
Pla1a	protein_coding	-	MT2B	LTR/ERVL	solo	chr16:38417057-38417131:-	2.612142332
RP23-400A19.2	lincRNA	-	MT2B	LTR/ERVL	solo	chr1:119319245-119319518:-	1.505671904
Lmbr1l	protein_coding	-	MT2B	LTR/ERVL	solo	chr15:98905937-98906036:+	1.121734273
S100a1	protein_coding	-	MT2B	LTR/ERVL	solo	chr3:90511959-90512063:+	0.961315949
Ctcf	protein_coding	+	MT2B	LTR/ERVL	solo	chr8:105653401-105653527:-	0.872175214
Nusap1	protein_coding	+	MT2B	LTR/ERVL	solo	chr2:119631787-119631876:+	0.721630014
Acot9	protein_coding	+	MT2B	LTR/ERVL	solo	chrX:155263394-155263728:-	0.576484506
Spag5	protein_coding	+	MT2B	LTR/ERVL	solo	chr11:78313909-78314051:-	0.495990958
Slc5a11	protein_coding	+	MT2B	LTR/ERVL	solo	chr7:123228507-123229071:+	0.388316964
Slc5a11	protein_coding	+	MT2B	LTR/ERVL	solo	chr7:123232599-123233196:+	0.381356797
Cbx1	protein_coding	+	MT2B	LTR/ERVL	solo	chr11:96797938-96798525:-	-0.000407603
Poc5	protein_coding	+	MT2B	LTR/ERVL	solo	chr13:96391667-96391796:+	-0.064449069
Ggta1	protein_coding	-	MT2B	LTR/ERVL	truncated	chr2:35423159-35427212:+	-0.115044479
Nedd8	protein_coding	-	MT2B	LTR/ERVL	solo	chr14:55631571-55663636:+	-0.159932734
Cdk4	protein_coding	+	MT2B	LTR/ERVL	solo	chr10:127065322-127065398:-	-0.18401604
Vdac1	protein_coding	+	MT2B	LTR/ERVL	solo	chr11:52386042-52386389:-	-0.264479648
Gm20748	lincRNA	+	MT2B	LTR/ERVL	solo	chr18:61644203-61644354:+	-0.294536437
Mum1	protein_coding	+	MT2B	LTR/ERVL	solo	chr10:80234451-80234630:-	-0.301699071
Zscan10	protein_coding	+	MT2B	LTR/ERVL	solo	chr17:23607887-23608077:-	-0.404751101
Ppm1a	protein_coding	+	MT2B	LTR/ERVL	solo	chr12:72735540-72736079:+	-0.410792826
Mib2	protein_coding	-	MT2B	LTR/ERVL	solo	chr4:155666827-155666927:+	-0.608729739

Table S6: Sequences of real-time PCR primers

Primer name	Primer sequence
MervI F	GGTGGTCGAGATGGAGGTTA
MervI R	CGGATTGCGGGTTTGATCTC
Iap F	GCTCCTGAAGATGTAAGCAATAAAG
Iap R	CTTCCTTGCGCCAGTCCCGAG
Line1 F	AAACCCCTTCCACTCCACTC
Line1 R	AGGGTAGCACTCTCCTTAGT
Sine B1 F	GTGGCGCACGCCTTTAATC
Sine B1 R	GACAGGGTTTCTCTGTGTAG
Gln F	CGTAAGGACCCTAGTGGCTG
Gln R	GCACTCACTCTTCTTCACTCTG
Zfp352 F	AAGGTCCCACATCTGAAGAA
Zfp352 R	GGGTATGAGGATTCACCCA
Tcstv1 F	GACCACCTGAACCATCCATC
Tcstv1 R	CACCTCAGGCTGCAGTTGTA
Tcstv3 F	ACCAGCTGAAACATCCATCC
Tcstv3 R	CCATGGATCCCTGAAGGTAA
Chit1 F	AAAACGTGGATGCTGCTGTG
Chit1 R	CCAGGACCCCTTTGTCCTTC
Cml2 F	AGGTTTTACTGGATGTCATCGGA
Cml2 R	CCCTGAGCCCTTTGGGAAC
Tmem132c F	GGCTGTCTGTCACCCTACAC
Tmem132c R	AGCCTCGTCCAAAGAGATGG
Tmem92 F	GGCACACTCACCTTGACCTT
Tmem92 R	GCCAGGATGACCAAGAACCTAA
P4ha2 F	CCCCAAGACAGGTGTCCTCA
P4ha2 R	ATCTTGCTCATCGCTCCTTGA
Abcb5 F	TACATCTGCCCTGGACACAG
Abcb5 R	TAGTACAGCCCCTGCTTTGC
Gata2 F	CACCCCGCCGTATTGAATG
Gata2 R	CCTGCGAGTCGAGATGGTTG
Cdx2 F	AGGCTGAGCCATGAGGAGTA
Cdx2 R	CGAGGTCCATAATTCCACTCA
Elf5 F	ATTCGCTCGCAAGGTTACTCC
Elf5 R	GGATGCCACAGTTCTCTTCAG
Gata3 F	CGGGTTCGGATGTAAGTCGA
Gata3 R	GTAGAGGTTGCCCCGCAGT
Tfap2c F	AAGCGGTGGCTGACTATTTAA
Tfap2c R	CAGGCTGAAATGAGACAAACAG
Psx1 F	GAATTGGTTTCGGATGAGGA
Psx1 R	GTGGCTCAGAAGAAGCCATC
Hand1 F	CCCCTCTTCCGTCCTCTTAC
Hand1 R	CTGCGAGTGGTCACACTGAT
Mash2 (Ascl2) F	GGTGACTCCTGGTGGACCTA
Mash2 (Ascl2) R	TCCGGAAGATGGAAGATGTC
Gata4 F	CGAGATGGGACGGGACACT
Gata4 R	CTCACCCTCGGCCATTACGA

Gata6 F	GAGCTGGTGCTACCAAGAGG
Gata6 R	TGCAAAAGCCCATCTCTTCT
Sox17 F	GAGGGCCAGAAGCAGTGTTA
Sox17 R	AGTGATTGTGGGGAGCAAGT
Foxa2 F	GTTAAAGTATGCTGGGAGCCG
Foxa2 R	CGCCCACATAGGATGACATG
Brachuary F	GCTCTAAGGAACCACCGGTCATC
Brachuary R	ATGGGACTGCAGCATGGACAG
Pax6 F	ACAGAGTTCTTCGCAACCTG
Pax6 R	CATCTGAGCTTCATCCGAGT
CD34 F	GGTAGCTCTCTGCCTGATGAG
CD34 R	TGGTAGGAACTGATGGGGATATT
Gata1 F	TGGGGACCTCAGAACCCTTG
Gata1 R	GGCTGCATTTGGGGAAGTG
Dab2 F	TCTCAGCCTGCATCTTCTGA
Dab2 R	GAGCGAGGACAGAGGTCAAC
β-actin F	GATCTGGCACCACACCTTCT
β-actin R	GGGGTGTTGAAGGTCTCAAA
Pdgfra F	CCTCAGCGAGATAGTGGAGAAC
Pdgfra R	ACCGATGTACGCATTATCAGAGT
Fgfr2 F	ACCAAATACCAAATCTCCCAAC
Fgfr2 R	ATTCATTCTCCACCAGGCA
Pl1 (Prl3d1) F	TGGAGCCTACATTGTGGTGG
Pl1 (Prl3d1) R	TGGCAGTTGGTTTGGAGGA
Egfr F	GCCATCTGGGCCAAAGATACC
Egfr R	GTCTTCGCATGAATAGGCCAAT
Mdfi (I-sma) F	GTAGCAAGATCCACTCACCCT
Mdfi (I-sma) R	CTGGAGCCATTTGGCAGTAGT