miR-203 controls timing of developmental transitions during early preimplantation embryogenesis

José González-Martínez,[1] Estefanía Ayala,[1,2,*] Agustín Sanchez-Belmonte,[1,3,*] Alejandro García,[1] Enrique Nogueira,[1,2] Anna Melati,[1] Daniel Gimenez,[4] Ana Losada,[4] Sagrario Ortega[3] and Marcos Malumbres[1,2,5]

[1] Cell Division and Cancer group, Spanish National Cancer Research Centre (CNIO) Madrid.

[2] Cancer Cell Cycle group, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain.

3 Mouse Genome Unit, Spanish National Cancer Research Centre (CNIO) Madrid.

4 Chromosome Dynamics group, Spanish National Cancer Research Centre (CNIO) Madrid.

5 ICREA, Barcelona.

Keywords: Embryo development, Histone acetylation, microRNA, Pluripotency, Totipotency

*, EA and ASB contributed equally to this work

Abstract

Commonly expressed at developmental transitions, microRNAs operate as fine tuners of gene expression. However, how they regulate the earliest developmental transitions in early mammalian embryogenesis remains obscure. Here, we identify the microRNA miR-203 as a critical regulator of timing within the totipotency to pluripotency transition in mouse embryos. Genetically engineered mouse models show that loss of miR-203 delays developmental timing during preimplantation leading to the accumulation of delayed embryos with high expression of totipotency-associated markers. By introducing miR-203-null alleles in a new embryonic reporter (eE-Reporter) transgenic mouse carrying MERVL-Tomato and Sox2-GFP transgenes, we observed that lack of miR-203 leads to sustained expression of MERVL retroviral elements and reduced Sox2 expression in later preimplantation developmental stages. A combination of single-cell transcriptional studies and epigenetic analyses identified the central coactivator and histone acetyltransferase P300 as a major miR-203 target at the totipotency to pluripotency transition in vivo. By fine tuning P300 levels, miR-203 carves the epigenetic rewiring process needed for this developmental transition, allowing a timely and correctly paced development.

Introduction

Preimplantation embryonic development is a meticulously orchestrated process characterized by tightly regulated cell transitions, each marking distinct stages of cellular specialization. The transition from a zygote to a 2-cell embryo involves the first round of cell division and establishes the totipotent nature of these cells, i.e. the ability to differentiate to any cell type of an organism including extraembryonic tissues crucial for supporting embryonic development. The transitions to the 8-cell stage and the morula are characterized by compaction and cell adhesion processes, leading to the formation of a tightly-packed cellular mass. The morula progresses to the blastocyst stage, which occurs at embryonic (E)3.5 day in the mouse, a critical juncture in embryonic development where cellular differentiation becomes more pronounced. At this stage, the outer epithelial layer forms the trophectoderm (TE), which will give rise to extraembryonic structures, whereas the inner cell mass (ICM) at the cavity will be separated into the extraembryonic primitive endoderm (PrE) or hypoblast, which continues to form the yolk sac; and the pluripotent epiblast (Epi), which gives rise to the embryo proper. The more external layer of TE cells exits the cell cycle and differentiate into trophoblast giant cells that invade the endometrium to promote implantation (Menchero *et al*, 2017; Wilkinson *et al*, 2023).

Epigenetic modifications, including histone modifications, DNA methylation, and non-coding RNAs, dynamically regulate the chromatin landscape and gene expression patterns during early embryonic development, contributing to the establishment of cellular identity. A common finding among these processes is that chromatin accessibility increases between the 2-cell-stage and blastocyst, accompanied by decreasing marks at chromatin and reduced DNA methylation until the generation of the three blastocyst layers (Wilkinson *et al.*, 2023). A critical step in this process is the 4C-to-8C transition, in which embryonic genome activation occurs accompanied by increased chromatin accessibility after repressive H3K27 tri-methylation (H3K27me3) marks are progressively replaced by H3K27 acetylation (H3K27ac) in the enhancers and promoters of genes transcriptionally induced during this transition (Gao *et al*, 2018; Lu *et al*, 2021; Millan-Zambrano *et al*, 2022). How this switch occurs in different species and cell lineages is an outstanding question with critical implications in early embryonic development (Wilkinson *et al.*, 2023). However, the scarcity of in vivo studies addressing these issues makes the identification of these core developmental regulators still elusive.

Reprogramming of DNA methylation during early embryonic development typically involves several rounds of de-methylation in the zygote, 2-cell stage and from the 8C to morula and blastocyst stages (Zhu *et al*, 2018). How demethylation occurs is unclear, although several studies have implicated TET enzymes, which oxidize 5-methylcytosines (5mCs) and promote locus-specific reversal of DNA methylation (Yang *et al*, 2020b). In addition, we recently reported that a specific microRNA, miR-203, can directly target the transcripts of the DNA methyltransferases DNMT3A and DNMT3B, leading to strong hypomethylation in pluripotent cells (Salazar-Roa *et al*, 2020). miR-203 was initially identified as a microRNA with critical roles in skin differentiation (Yi *et al*, 2008) and soon recognized as a target of multiple alterations in cancer (Bueno *et al*, 2008; Michel & Malumbres, 2013). Ectopic expression of this microRNA leads general hypomethylation of the genome in embryonic stem (ES) cells and induced pluripotent stem (iPS) cells, in a DNMT3A/B-dependent manner, leading to a "tabula rasa" pluripotent cellular state in which cell differentiation towards multiple lineages is enhanced (Salazar-Roa *et al.*, 2020). However, whether miR-203 plays a role during early embryonic development remains unexplored.

Results

miR-203 is required for timely transitions during early embryonic development

In mouse embryos, miR-203 is significantly induced at the 8-cell and morula stages, in parallel to compaction and polarization of blastomeres, and its expression levels decrease before the first cell-fate decision to generate outer-residing polar cells of the trophectoderm (TE) and apolar cells in the inner cell mass (ICM; Figure 1a). miR-203 is also undetectable in cultured ESCs derived from the ICM (Supplementary Figure 1a).

To study the functional relevance of miR-203 in early developmental stages, we generated mice with a conditional deletion of the entire *mmu-mir-203* locus using the Cre-loxP system. The miR-203-encoding sequence is located at Chr. 12 in the mouse, in an intergenic region ~3.5-kb telomeric to *Aspg* and ~15-kb centromeric to the *Kif26a* gene. We placed two *loxP* sites flanking *mmu-mir-203* generating the *miR-203*(lox) allele after recombination of these sequences in ES cells (Figure 1b and Supplementary Figure 1b). Cre-mediated recombination resulted in a null allele [*miR-203*(–)], which was used to generate germline miR-203-deficient mice. Expression of miR-203 was reduced in *miR-203*(+/–) 8-cell-stage embryos and was undetectable in *miR-203*(–/–) homozygous mutant embryos (Figure 1b). Lack of miR-203 did not result in reduced viability during

embryogenesis as we observed no difference compared to the expected ratio of miR-203-null pups (Figure 1c). On the other hand, adult *miR-203*(–/–) mice were viable and did not display obvious alterations, apart from reduced thickness and complexity of the stratified epithelium in the skin (Supplementary Figure 1c), in agreement with a previously described role of miR-203 in skin differentiation (Yi *et al.*, 2008). However, lack of miR-203 resulted in delayed kinetics of transitions in the early embryo. When analyzing the structure of E3.5 embryos directly isolated from the uterus, we observed that ~95% of E3.5 wild-type embryos had already reached the blastocyst stage. However, miR-203-deficient embryos displayed a reduced representation of these structures (21%) in favor of more incipient blastocysts, abundant morulas at different stages as well as a few earlier (4-, 8-cell stages) embryos and only a few aberrant structures (Figure 1d). Intriguingly, despite these defects, the majority of *miR-203*(–/–) embryos were able to progress to the blastocyst stage by E4.5 (Supplementary Figure 1d), suggesting a specific role for miR-203 in ensuring a smooth and timely developmental transition through early developmental stages.

We also made use of an inducible knock-in allele, *miR-203*(tet), previously generated in our laboratory (Salazar-Roa *et al.*, 2020), in which miR-203 can be induced by doxycycline (Dox) in the presence of the reverse transactivator rtTA (Rosa26-rtTA; Figure 1e). Treatment of *miR-203*(+/tet); Rosa26-rtTA [from now on, *miR-203*(+/tet); with Dox resulted in a significant upregulation of miR-203 at different developmental stages (Figure 1f), and resulted in a significant reduction in the number of viable embryos in vitro (Figure 1g and Supplementary Figure 1e).

To generate detailed in vivo data about early transitions in the mouse embryo, we monitored expression of the murine class III ERV with leucine transfer RNA primer binding site (MERVL), an endogenous retrovirus specially expressed during the zygotic genome activation at the 2C stage (Macfarlan *et al*, 2012). We generated a new transgenic moouse in which a MERVL::tdTomato reporter (Macfarlan *et al.*, 2012) was injected in mouse zygotes. After selection of mice with stable germline transmission of the allele, we observed strong expression of the Tomato reporter at the 2-cell stage and decreasing expression in the subsequent developmental stages with residual or patched expression in some 8-cell embryos (Figure 2a-c and Supplementary Figure 2a). By E2.75, most *miR-203*(–/–); MERVL-Tomato 8-cell embryos displayed higher levels of expression of the Tomato reporter when compared to control embryos (Figure 2d,e and Supplementary

Figure 2b). We also combined this allele with a previously generated Sox2-GFP transgenic (Suh *et al*, 2007), in which the green fluorescence reporter is strongly expressed later at the morula and blastocyst stages, thus generating a double MERVL-Tomato; Sox2-GFP transgenic line (early Embryo reporter or eE-Reporter mouse; Figure 2a-c). Introducing these two alleles into the miR-203 deficient mouse allowed the identification of compacted morulas with either low or high expression of MERVL-Tomato that were not present in E3.5 control embryos (Figure 2f,g), as well as morulas with separated green and red clusters of cells, suggesting not only developmental delay but also an asynchronous cell fate transition within the blastomeres of some embryos.

miR-203 modulates totipotency and pluripotency programs in preimplantation embryos

To understand the molecular basis behind these defects, we collected 289 E3.5 and 540 E4.5 *miR-203*(–/–) (KO) embryonic cells, and compared them to 306 E3.5 and 1,314 E4.5 wild-type embryonic cells using single-cell RNA sequencing (scRNAseq; Supplementary Figure 3a). We also collected 251 E3.5 and 951 E4.5 *miR-203*(+/tet) cells after treatment with Dox when plug was detected (from now on, *miR-203*(+/tet)^{Dox}). Dimensional reduction of expression data clearly separated miR-203-deficient cells from control of *miR-203*(+/tet)^{Dox} embryonic cells (Figure 3a and Supplementary Figure 3a). Initial mapping with markers of trophoectoderm (TE; *Cdx2*, *Gata3* and *Krt8*), inner cell mass (ICM; *Nanog*, *Klf4* and *Pouf5f1*) and primitive endoderm (PrE, also known as hypoblast; *Gata6, Gata4, Sox17*) suggested that a clear separation of these three lineages in E4.5 cells as well as control and *miR-203*(+/tet)^{Dox} E3.5 embryos (Figure 3b). E3.5 miR-203-deficient cells, however, were not well defined with these markers.

We then used gene signatures typical of the 2-cell, 8-cell, ICM/epiblast, TE and PrE lineages (Supplementary Table 1 and Supplementary Figure 3c) to classify all cells in this dataset (Figure 3c,d). This analysis suggested that the E3.5 miR-203-deficient cell cluster was composed of 8-cell-like and ICM/epiblast cells, as well as a small number of 2-cell-like cells that were not present in other groups. Pseudotime inference in this group of cells confirmed the expected transition from 2c- to 8c-like cells before giving rise to the three major embryonic lineages (Figure 3e-g and Supplementary Figure 3d,e). By E4.5, all control, *miR-203*(+/tet)^{Dox} and miR-203-deficient embryos had separated populations representing TE, as well as the two lineages generated in the second cell fate decision in the embryo: the epiblast and PrE. However, the ratio in the number of cells

belonging to these populations was altered, with miR-203-deficient and *miR-203*(+/tet)^{Dox} embryos having a lower proportion of TE cells (Figure 3h). Interestingly, *miR-203*(–/–) E3.5 embryos showed a significant overexpression of MERVL transcripts when compared to wild-type or *miR-203*(+/tet)^{Dox} embryos, and these differences were still evident in E4.5 embryos (Figure 3i and Supplementary Figure 3f). MERVL high-cells were mostly found in the 2-cell- and 8-cell-like clusters described above in the *miR-203*(–/–) E3.5 embryos (Figure 3j and Supplementary 3g,h), but were also present in some *miR-203*(–/–) and *miR-203*(+/tet)^{Dox} Epiblast-like clusters at E4.5 (Figure 3j). In agreement with the increased levels of 2-cell-like markers, *miR-203*(–/–) embryos displayed decreased expression of pluripotency markers by E3.5 (*Nanog, Oct4, Sox2;* Supplementary Figure 3i). However, by E4.5 pluripotent transcripts recovered and were more abundant in the epiblast lineage of miR-203-null embryos. These mutant cells displayed decreased abundance of TE markers and higher expression of hypoblast markers, in agreement with the differences in the representation of those cells in mutant embryos (Figure 1h).

miR-203 modulates the P300 histone acetyltransferase during early developmental stages

To look for possible miR-203 targets during early development, we first analyzed pathways differentially expressed in E3.5 embryos. Overexpression of miR-203 at this stage resulted in downregulation of chromatic modifying enzymes (Figure 4a) as the most downregulated pathway suggesting a possible effect of miR-203 in the transcripts encoding these enzymes. Downregulated transcripts encoding chromatin regulators in *miR-203*(tet)^{ON} embryos included *Ep300, Pbrm1, and Smarcd1*, among others (*Atf7ip, Kmt2a, Kat6a, and Tada2b*; Figure 4b). On the other hand, these three transcripts were upregulated in *miR-203*(–/–) embryos, together with *Kat6b, Arid1a, Kdm6a and Dr1*; Figure 4c). Expression of the corresponding genes remained low in all samples overexpressing miR-203 (Figure 4d). We tested the ability of miR-203 in directly repressing these transcripts by generating reporter constructs with the 3'-UTR of each of them downstream of a luciferase reporter. Expression of miR-203 was able to repress most of these transcripts (Figure 4e) with the possible exception of *Arid2* and *Atf2*, two chromatin regulators whose transcript levels were also reduced although did not reach statistical significance in the differential expression analysis. Mutation of the predicted

miR-203 binding sites of these luciferase-3'-UTR reporters prevented miR-203-driven repression (Figure 4e), suggesting a direct control of these transcripts by miR-203.

Among the validated candidates, we observed that P300, but not the related acetyltransferase CREB binding protein (CBP), displayed an expression pattern inverse to that of miR-203, with increasing levels during the 2-4-cell stages and low levels at the 8-cell and morula stages when miR-203 peaks (Figure 4f), in agreement with previous studies (Wang *et al*, 2022). Expression of P300 increased after miR-203 was silenced in normal blastocysts, but upregulation did not happen when miR-203 expression was maintained high in the miR-203(tet)^{ON} model (Figure 4g).

miR-203 regulates developmental timing by controlling acetyltransferase levels in preimplantation embryos

To test the relevance of P300 as a miR-target we made use of Trichostatin A (TSA), a molecule that induces the acetylation and stability of P300, in E2.5 and E3.5 embryos isolated from the eE-Reporter mouse. As depicted in Figure 5a and Supplementary Figure 4a, the P300 stabilizer TSA induced a phenotype similar to that observed in *miR-203*(–/–) embryos, including developmental delay and strong expression of MERVL at later developmental stages. The effect of TSA was reproduced with suberoylanilide hydroxamic acid (SAHA), a pan-histone deacetylase (HDAC) inhibitor (Supplementary Figure 4b). On the other hand, A485, a catalytic inhibitor of the P300 and CREB binding protein (CBP) acetyltransferases, also induced a delay in embryo development (Figure 5a and Supplementary Figure 4a). Importantly, this P300/CBP inhibitor prevented the expression of MERVL in *miR-203*(–/–) embryos (Figure 5a and Supplementary Figure 4a), suggesting that these acetyltransferases were required for MERVL expression in the absence of miR-203. Similarly, short interfering RNAs (siRNAs) against *Ep300* also displayed a significant effect in delaying preimplantation embryonic development, and this effect was reproduced to a lesser extent after knock down of the lysine acetyltransferases *Kat6a* or *Kat6b* (Figure 5b).

To analyze more in detail the molecular effects of these treatments, we exposed E2.5 8-cell stage embryos with the above-mentioned small molecule inhibitors for 24h in vitro, and samples were processed for low input RNA-seq. Stabilization of P300/CBP or HDAC inhibition resulted in significant increase in cells displaying a 2-cell-like transcriptional program (Figure 5c and Supplementary Figure 4c), including the expression of MERVL transcripts (Figure 5d). The effect of specifically modulating P300

levels was validated by microinjecting zygotes with siRNAs against P300, 72 h before harvesting samples for RNAseq analysis. Knock down of *Ep300* resulted in overrepresensation of 2-cell-like transcripts (Figure 5e and Supplementary Figure 4d) as well as overexpression of MERVL sequences (Figure 5f). Interestingly, these effects we are also reproduced after known down of *Kat6b*, suggesting that miR2-30 may modulate preimplantation developmental timing by regulating P300 and KAT6B levels in preimplantation embryos.

We finally tested the effect of miR-203 levels in lysine acetylation at H3K9 and H3K27, two sites involved in the activation of transcripts, by testing the genome-wide profile of acetylation of these sites using low input bulk CUT&RUN on E3.5 dissagregated embryos (Meers *et al*, 2019; Wang *et al.*, 2022). Lack of miR-203 resulted in increased acetylation of genes belonging to the 2-cell-like signature whereas overexpression of miR-203 had a small effect on reducing acetylation of these gene promoters (Figure 6a). Statistical analysis using gene set enrichment indicated increased H3K9ac and H3K27ac marks in 2-cell genes as well as MERVL sequences in miR-203-null embryos (Figure 6b,c). All these data suggest that lack of miR-203 induces increased levels of acetyltransferases such as EP300 and possibly KAT6B, thereby resulting in increased acetylation and expression of 2-cell stage transcripts during the 8-cell-to blastocyst stages of preimplantation development.

Discussion

The epigenetic control of cell fate transitions during preimplantation development in mammals is well established, although the exact mechanisms controlling changes in time and space remain unclear. Totipotency is first established around zygotic genome activation at the 2C stage, and MERVL is considered one of the first transcripts expressed during the early zygotic activation at the 2C stage (Macfarlan *et al.*, 2012). Its activation in pluripotent cells is sufficient to induce a 2C-like cell state (De Iaco *et al*, 2017; Hendrickson *et al*, 2017; Yang *et al*, 2020a). MERVL sequences function as chromatin organizers in early embryos (Asimi *et al*, 2022) and full-length MERVL transcripts, but not encoded retroviral proteins, are essential for preimplantation development (Sakashita *et al*, 2023).

Several microRNAs have been involved in the epigenetic control of MERVL expression in preimplantation embryos. miR344 is activated by DUX at the 2C stage and directly silences ZMYM2 and its partner LSD1, preventing the recruitment of

LSD1/HDAC corepressor complexes to MERVL sequences (Yang *et al.*, 2020a). Interestingly, miR-344 expression increases from pronucleus to 2C-8C stages, peaks at the 8C stage, and it is then downregulated in later stages, showing a complementary pattern to that of miR-203 (Figure 1a). miR-34a, on the other hand, represses GATA2, a transcriptional activator of MERVL, and lack of this microRNA results in upregulation of MERVL and 2C genes leading to expanded developmental potency in mouse embryos (Choi *et al*, 2017). These data suggest a model in which miR-344 is permissive for the expression of MERVL sequences at the 2C stage by targeting ZMYM2/LSD1 and thereby preventing the recruitment of HDAC repressor complexes. miR-34a is later induced at the 4C to 8C transition targeting GATA2 thereby preventing efficient MERVL expression. The parallel DUX silencing during the exit of the 2C state would lead to reduction in miR-344 levels and upregulation of ZMYM2, which can also directly bind and repress GATA2 during this transition. miR-203 would later be induced during the 8C-blastocyst transition leading to repression of acetyltransferases such as P300, thereby potentiating the HDAC repressive activity on MERVL/2C promoters. miR-34a could also participate in repressing MERVL sequences later in the ICM and epiblast lineage after miR-203 silencing.

These multiple mechanisms of control also suggest high redundancy. It is well established that microRNAs fine-tune gene expression, rather than being the primary determinant of expression levels, by using an integrative and combinatorial approach, such that multiple miRNAs may target an individual gene and an individual miRNA may target multiple genes (DeVeale *et al*, 2021). For instance, no differences in the levels of MERVL and 2C transcripts were observed in miR-34a-deficient preimplantation embryos, suggesting highly redundant molecular networks mediating MERVL silencing (Choi *et al.*, 2017). In addition to the presence of multiple miR-34 family members, it is also conceivable that multiple microRNAs could act redundantly to repress MERVL expression to ensure rapid and efficient MERVL silencing in vivo. However, microRNA activity is dispensable for pre-implantation development (Bernstein *et al*, 2003; Wang *et al*, 2007). The fact that homozygous mutations in genes encoding many epigenetic regulators cause early developmental failure (Grosswendt *et al*, 2020) suggests additional layers of regulation beyond these non-coding RNAs.

The fact that heterozygous or hemizygous mutations in the genes encoding the acetyltransferases p300 and CBP cause the Rubinstein-Taybi syndrome, characterized by distinctive facial features and varying degrees of intellectual disability (Roelfsema *et al*,

2005; Van Gils *et al*, 2021), suggests that p300 is happloinsufficient for cell fate decisions in development. In fact, even a relatively small imbalance in these proteins has significant developmental effects (Roelfsema *et al.*, 2005). p300 and CBP act as integrators of the signals from various molecular pathways that compete for these factors to determine chromatin accessibility (Goodman & Smolik, 2000). For instance, P300 is known to participate as a SOX2 co-activator in pluripotent cells (Yoo *et al*, 2023), as well as a mediator of trophoblast stem cell pluripotency (Dou *et al*, 2024).Our data suggest that miR-203 may participate in regulating p300 levels during preimplantation development. In the absence of miR-203, p300 levels would rise, allowing its binding to additional enhancers such as those involved in the 2C stage (including MERVL), allowing the expression of both totipotency and pluripotency genes, and increasing lineages other than TE, as observed in *miR-203*(–/–) embryos. On the other hand, overexpression of miR-203 would lead to a general deficiency in the availability of p300 for totipotency and pluripotency enhancers, leading to specific defects in later lineage determination. Whereas these hypotheses remain speculative, our data along with previous observations reporting the control of DNMT3A/B by miR-203 in pluripotent cells (Salazar-Roa *et al.*, 2020), suggest new functions for this microRNA in early developmental transitions. Given the increasing relevance of assisted reproductive technologies, these data may have implications for controlling the regenerative potential of pluripotent cells for medical needs (Gonzalez-Martinez & Malumbres, 2022).

Methods

Mouse models

The miR-203 conditional knockout model was generated by flanking the *mmu-mir203* locus with loxP sites (Genebridges) using homologous recombination into ES cells (Supplementary Figure 1b). A frt-flanked neomycin-resistant cassette was used for selection of recombinant clones. Elimination of the neomycin-resistant cassette was achieved using a CAG-Flpe transgenic line (Rodriguez *et al*, 2000) and loxP sites were recombined using a EIIa-Cre transgene (Schwenk *et al*, 1995) resulting in the *miR-203*(–/–) allele (Figure 1b and Suppl. Figure 1b). The miR-203 inducible model (Salazar-Roa *et al.*, 2020) and Sox2-GFP transgenic allele (Suh *et al.*, 2007) were described previously. For generation of the MERVL-Tomato transgenic allele, the expression vector reported previously (Macfarlan *et al.*, 2012) was isolated from a bacterial extract using

the EndoFree Plasmid Maxi Kit (Qiagen #12362), digested with SacII for linearization, phenol/chloroform treated and resuspended in microinjection buffer 10 mM Tris-HCl pH 7.5, 0.1 mM EDTA at 1 ng/µl. Transgenic mice were produced by pronuclear injection of the linearized vector into zygotes obtained from superovulated B6.CBA females crossed with males containing the *Col1a1_tetO-miR203* and *Rosa26_rtTA* alleles (Salazar-Roa *et al.*, 2020) using standard protocols. Out of 66 mice born, 10 were confirmed positive for transgene integration. Founders were crossed with C57BL6/J females to test for germ line transmission.

These animals were maintained in a B6.CBA genetic background and mice were bred under specific-pathogen-free conditions at the Spanish National Cancer Research Center (CNIO) animal facility, following the animal care standards of the institution. The animals were observed daily, and diseased mice were humanely euthanized in accordance with the Guidelines for Humane Endpoints for Animals Used in Biomedical Research (Directive 2010/63/EU of the European Parliament and Council and the Recommendation 2007/526/CE of the European Commission). All animal protocols were approved by the committee for animal care and research of the Instituto de Salud Carlos III-Comunidad de Madrid (Madrid, Spain).

Mice superovulation, embryos isolation, and microinjection

For female mice superovulation, 4 to 5 weeks old females were superovulated by intraperitoneal injection of 5 IU of pregnant mare serum gonadotropin per female (PMSG; Sigma G4877-1000IU or ProSpec #HOR-272) on Day 1 at 1:00 p.m. On day 3 at 12:00 p.m, females were injected intraperitoneally 5IU of human chorionic gonadotropin (hCG; sigma C1063-10VL) and mated with the desired males (8 to 20 weeks old). The next day vaginal plugs were assessed at 7-8 am (0.5dpc or E0.5). Zygotes were collected at E0.5, 2-cell stage embryos at E1.5, 8-cell stage embryos at E2.5, morulae at E2.75, and blastocysts at E3.5 (mid blastocyst stage) or E4.5 (late blastocyst stage). Isolation of zygotes was achieved by dissecting out the cumulus zygotes complexes and briefly treating them with type IV hyaluronidase (Merck #H4272), whereas 2-cell, 8-cell and morula stage embryos were obtained by flushing the oviduct with M2 medium (Sigma-Aldrich # M7167) by inserting a fine needle through the infundibulum and collecting the embryos. Blastocysts were obtained by flushing the uterus at E3.5/4.5. After isolation, embryos were processed for analysis or cultured in 20uL drops of KSOM EmbryoMax medium (Sigma-Aldrich #MR-106-D) carefully covered by mineral oil suitable for

embryo culture (lifeguard oil #LGUA-5 00) in P35 culture dishes. For embryo microinjection, E0.5 zygotes or E1.5 2-cell stage embryos were isolated, and siRNAs or DNA constructs diluted in microinjection buffer (10 mM Tris pH 7.4, 0.1 mM EDTA) in the indicated concentrations were microinjected in the cytoplasm or the pronucleus, respectively. When needed, the resulted microinjected embryos were transferred to Crl:CD1(ICR) pseudo-pregnant females.

For analysis of postimplantation embryos, E3.5 blastocysts were obtained as previously described, zona pellucida was removed, and 5 embryos were seeded in each well of an 8 µ-Slide 8 Well high Glass Bottom IBIDI chamber (IBIDI # 80807) filled with IVC1 culture medium (Bedzhov *et al*, 2014). On day 3, culture medium was changed to IVC2 and 48 h later embryos were fixed with 4% PFA and immunostained. The composition of the culture mediums is stated in (Bedzhov *et al.*, 2014).

Analysis of microRNA expression levels in preimplantation embryos

For real time PCR, embryos were harvested and pipetted into 8uL of cell lysis buffer from the TaqMan MicroRNA Cells-to-CT Kit (Invitrogen #4391848) in a 200 µL low binding tube, pipetted up and down 10 times with low binding pipette tips, and incubated for 8min at room temperature. An average of 10-20 embryos per experimental condition and replicate were used. After incubation, 1 µL of stop buffer was added to the mix, which was pipetted up and down 5 times and kept on ice. After miRNA isolation, miRNA samples were shortly subjected to cDNA synthesis and real time PCR amplification with two-step miRNA assays from TaqMan probes following manufacturer instructions, using the TaqMan Universal PCR Master Mix, no AmpErase UNG (ThermoFisher #4364343) in a QuantStudio 5 real-time PCR system. The Taqman two-step miRNA expression assays used in this study were the miR-203 pri-miRNA assay (Assay ID: Mm03306530_pri); the miR-16 pri-miRNA assay (Assay ID: Mm03306798_pri); the miR-203 miRNA assay (Assay ID: 000507), and the miR-16 miRNA assay (Assay ID: 000391).

Single-cell RNA sequencing

For single-cell RNA sequencing of E3.5 blastocysts, zona pellucida was removed passing the embryos by mouth pipetting through three consecutive drops of Tyrode's acidic solution pH 2.5 (Sigma T1788). After careful examination under the microscope for disintegration of the pellucida, tyrode's solution was immediately neutralized with a double volume of PBS pH 7.4 (Gibco #10010023), and embryos were placed into 20 µL

of dissociation buffer composed of TrypLE™ Express Enzyme (Gibco # 12604013) / Accutase (Gibco # A1110501) in a 1:1 ratio, in a low binding 200 µL microcentrifuge tube. To avoid attachment of the embryos to the pulled crystal capillaries, capillaries were coated with 1% BSA in PBS beforehand. Blastocysts were incubated for 20 minutes in dissociation buffer at 37ºC, pipetted up and down gently 60 times with 20 µL low binding pipette tips, incubated 10 additional minutes at 37ºC and pipetted up and down again gently 40 times to disaggregate the last ICM cell clumps. After dissociation, the resulting cell suspension was neutralized with 40uL of 1% cell culture grade BSA (Sigma # A8806). In the case of 8-cell and morula stage embryos, pellucida was removed as abovementioned and embryos were placed in 20 µL of Accutase (Gibco # A1110501) for 20 min at 37ºC. After incubation, embryos were pipetted up and down 50 times until complete disaggregation and neutralized by the addition of 40 µL of 1% cell culture grade BSA (Sigma # A8806) to avoid the formation of clumps after dissociation. Then, cell suspensions were immediately placed on ice until processing for scRNA-seq. For scRNA-seq analysis, cell samples were loaded onto a 10x Chromium Single Cell controller chip B (10x Genomics) following manufacturer's intructions (Chromium Single Cell 3'GEM, Library & Gel Bead Kit v3, ref. PN-1000075). Intended targeted cell recovery of ~1000 - 10000 cells depending on the embryonic stage. Generation of gel beads in emulsion (GEMs), barcoding, GEM-RT clean-up, cDNA amplification and library construction were all performed as recommended by the manufacturer. scRNA-seq libraries were sequenced with an Illumina NextSeq 550 (using v2.5 reagent kits). Raw images generated by the sequencer are submitted to analysis, per-cycle base calling and quality score assignment with Illumina's Real Time Analysis integrated primary analysis software (RTA v2). Conversion of BCL (base calls) binary files to FASTQ format is subsequently performed with bcl2fastq2 (Illumina).

Ultra-low input bulk RNA-seq

For ultra-low input RNA-seq, 10-15 embryos were collected per sample and resuspended in 8 mL of low input lysis buffer of the NEBNext Single Cell/Low input cDNA Synthesis & Amplification module (#E6421S). Whole cells were processed with the "NEBNext Single Cell/Low Input RNA Library Prep" kit (NEB #E6420) by following manufacturer instructions. Double stranded cDNA production was performed by limited-cycle PCR. Sequencing libraries were completed with the "NEBNext Ultra II FS DNA Library Prep Kit for Illumina" (NEB #E7805) and subsequently analysed on an Illumina instrument by

following manufacturer's protocols. The resulting purified cDNA libraries were applied to an Illumina flow cell for cluster generation and sequenced on Illumina NovaSeq X (with NovaSeq X Series Reagent Kits) by following manufacturer's protocols using PairEnd (2x151 bases) in KO experiment and Illumina NextSeq 550 (with 75 cycle High Output v2.5 reagent kit) using SingleRead (86 bases) in pharmacological modulation experiments. Raw images generated by the sequencer are submitted to analysis, per-cycle basecalling and quality score assignment with Illumina's Real Time Analysis integrated primary analysis software (RTA4). Conversion of BCL (base calls) binary files to FASTQ format is subsequently performed with BCL Convert (Illumina).

Computational analysis of transcriptomic data

For ultra-low input RNAseq, we used the cluster_rnaseq pipeline (https://github.com/cnio-bu/cluster_rnaseq/tree/master). Sequencing quality control was performed using multiqc preprocessing (v1.18; https://multiqc.info), and adapters were removed with BBDuk (v39.01, https://github.com/BioInfoTools/BBMap/blob/master/sh/bbduk.sh). For each sample, we performed read alignment to the mm10 version of GRCh38 mouse reference genome using STAR (v2.7.11a) and quantification using featureCounts from Subread package (v2.0.6; (Liao *et al*, 2019)) and count matrices were used for downstream analysis in scanpy v. 1.8.2. We used TEtranscripts (v2.2.3; https://github.com/mhammell-laboratory/TEtranscripts) in order to quantify TE elements in our low input RNA samples using bam files extracted with STAR in aligment step, annotation file for mm10 version of GRCh38 mouse reference genome and for transcript elements of GRCh38 mouse reference pre-generated by RepeatMasker 3.0 (http://www.repeatmasker.org).

For single-cell RNAseq data, quality control was performed using fastqc (version 0.11.9). For each sequenced single-cell library, we performed read alignment to the mm10 version of GRCh38 mouse reference genome. Quantification and initial quality control were performed using STARsolo of STAR (version 2.7.10a) and STARsolo filtered count matrices were used for downstream analysis. We integrated the filtered count matrices from STARsolo and analysed them with scanpy v. 1.8.2 following their recommended standard practices (Wolf *et al*, 2018). In brief, we excluded genes expressed by less than three cells, excluded cells expressing fewer than 150 genes, and cells with more than 15% mitochondrial and 25% ribosomal content. Expression data was normalized and converted to log(X + 1). Next, for dimensionality reduction we identified highly variable

genes, scaled the data and calculated PCA to observe the variance ratio plot and decide on an elbow point which defined n_pcs. This was used for calculating neighbourhood graph, UMAP and further Louvain clustering. RNA velocity was used to investigate directed dynamic information by leveraging splicing kinetics using scVelo v. 0.2.4 (La Manno *et al*, 2018). For the identification of transposable elements we used scTE v. 1.0 (He *et al*, 2021) following the same parameters used with STARsolo.

The origin of gene signatures used for signature scores in shown in Supplementary Table 1. All differential gene expression analysis were performed with scanpy using t-test method for comparisons between subpopulations and conditions. Only significant differentially expressed genes were considered for downstream analysis (FDR (bonferroni) < 0.05). GSEA Preranked was used to perform gene set enrichment analysis for the selected gene signatures on a pre-ranked gene list, setting 1,000 gene set permutations (Subramanian *et al*, 2005) using the Python wrap (Fang *et al*, 2023). Only those gene sets with significant enrichment levels (FDR q-value < 0.25) were considered.

Luciferase assays

Luciferase assays were performed in HEK293T cells. Briefly, 2.10^5 cells per well were seeded on 6-well plates, and in 24h cells were transfected using Lipofectamine 2000 (Invitrogen), following the manufacturer's instructions. The 3′-UTR regions from the murine genes *Kat6a, Kat6b, Ep300, Arid1a, Arid2, Dr1, Smarcd2, Kmt2c* and *Atf2* were amplified by PCR with specific primers (Supplementary Table 2) using mouse cDNA from ES cell cultures. PCR products were verified by sequencing cloned into the pGL3-Control vector (Promega) by Gibson Assembly, downstream of the luciferase reporter gene. Mutations in the miR-203 seed sequence binding sites were generated by site-directed mutagenesis (Supplementary Table 3) and subsequently verified by sequencing. Overexpression of miR-203 or control GFP was achieved by transfection with either pMCSV-GFP or pMCSV-miR-203-GFP vectors (Salazar-Roa *et al.*, 2020), in combination with the pGL3-derived vectors, and *Renilla* as a control. Luciferase measurement was achieved 48h post-transfection using a luminescence microplate reader (Bio-Tek).

Ultra-low input CUT&RUN

CUT&RUN samples were generated following the original protocol of (Meers *et al.*, 2019) and a modified protocol (Wang *et al.*, 2022) for early embryos. For CUT&RUN, 100 disaggregated blastocysts were used per sample. For each sample, 10 µl of BioMag

Plus Concanavalin A beads (Polysciences # 86057-3) were washed twice in 500 µL of Binding Buffer (20 mM HEPES-KOH pH 7.9, 10 mM KCl, 1 mM $CaCl_2$, 1 mM $MnCl_2$), and after buffer removal were resuspended in 10 µl Binding Buffer. Then, 8-cell stage embryos or disaggregated blastocysts (as stated in the scRNA-seq section) were added to 45µl Wash Buffer (150 mM NaCl, 0.5 mM Spermidine, 20 mM HEPES-NaOH pH 7.5, 1X Roche Protease Inhibitor Cocktail), mixed with the 10µl Binding Buffer containing magnetic beads, gently tapped twice, and incubated for 10 min at room temperature in a 200 µL low binding tube. Buffer was removed by placing the tube on a magnet stand, the beads were resuspended in 50 µl antibody Incubation Buffer (1X Wash Buffer with 0.02% Digitonin and 2 mM EDTA) with the indicated antibody and kept at 4°C overnight. Then, samples were washed twice with 100 µl Wash Buffer with 0.02% Digitonin. The beads were resuspended in 100 µl Wash Buffer with 0.02% Digitonin and 500 ng/ml CNIO home-made pA-MNase, and rotated for 2h at 4°C. Beads were then washed twice with 100 µl of Wash Buffer with 0.02% Digitonin, resuspended in 100 µl ice-cold Wash Buffer with 0.02% Digitonin and 2 µl 100 mM $CaCl_2$, and incubated in ice in a metallic rack for 30 min. The reaction was stopped by adding 12 µl of 10X Stop Buffer (1.700 mM NaCl, 100 mM EDTA, 20 mM EGTA, 0.02% Digitonin, 250 µg/ml RNaseA, 250 µg/ml Glycogen). The tube was then incubated at 37°C for 15 min to release the CUT&RUN fragments into the supernatant. After spinning the sample for 1min and putting it in a magnetic stand, the supernatant was transferred to a new 1.5 ml ml low binding tube while discarding the beads. DNA was purified by fist adding 2.5 µl 10% SDS, and 2.5µl Proteinase K (20 mg/ml) to the supernatant and incubating it at 55°C for 1 h. GlycoBlue (Invitrogen #AM9516) was added to the supernatant before the supernatant was mixed with 200 µl Phenol:Chloroform:Isoamyl Alcohol (Invitrogen) by vortexing. Samples were then centrifuged at 16000 g for 5min at room temperature. After carefully transferring the supernatant to a new 1.5 ml LoBind tube with low binding pipette tips, 26 µl 3M Sodium Acetate, 1 µl Glycogen (20 mg/ml), and 750 µl cold 100% Ethanol were added, mixed by vortexing, and incubated at -20°C for 1h. Then, the tube was centrifuged at 16,000 g for 20 min at 4°C and the supernatant was discarded. The precipitated DNA pellet was washed with 500uL room temperature 70% Ethanol, and then dried at 55ºC for 15min. Finally, the DNA pellet was dissolved overnight in 20 µl ultrapure H_2O. DNA samples were processed through subsequent enzymatic treatments for end-repair, dA-tailing, and ligation to adapters with "NEBNext Ultra II DNA Library Prep Kit for Illumina" (NEB #E7645). Adapter-ligated libraries were completed by a PCR of 21 cycles

of amplification, with 63º T(ann/ext) for 10s, and extracted with a (single) double-sided SPRI size selection. Libraries were applied to an Illumina flow cell for cluster generation and sequenced on an Illumina NextSeq550 in paired-end 43-bp mode by following manufacturer's recommendations.

Quality control was performed using fastqc. Paired-end reads were aligned against mm10 genome using Bowtie2 (version 2.4.2) with the following parameters "-local --very-sensitive-local --no-unal --no-mixed --no-discordant --phred33 -I 10 -X 700". Then we excluded reads that fail in the alignment using samtools (version 1.11) and duplicated reads with GATK (version 4.1.9). Finally, the bam files were converted to bigwig using bamCoverage from deeptools (version 3.5.1). To extract the signal of the promoters, we quantified the amount of signal in a 2.5kb window centered in the TSS of the genes and normalized the values using the median value of the 5000 more expressed genes for each condition (Dahl *et al*, 2016). Additionally, we quantified the signal in the MERVL sequences and normalized them using the value previously calculated. GSEA preranked (version 4.3.2) was used to evaluate the enrichment in the different conditions compared to controls.

Statistics

Statistics was performed using Prism 9 software (GraphPad Software) or Python 3.8. All statistical tests of comparative data were done using an unpaired, 2-tailed Student's t test with Welch's correction, or 1-way ANOVA with Tukey's multiple-comparison test when appropriate. Data are expressed as the mean of at least 3 independent experiments ± SEM; with a *P* value of less than 0.05 considered statistically significant.

Acknowledgements

J.G.M. and A.S.B. were supported by FPI contracts from the Spanish Ministry of Science and Innovation (MICINN). A.G. received a predoctoral contract from Foundation laCaixa. M.M. lab was supported by research grants from MICINN/AEI/FEDER (PID2021-128726 and PDC2022-133408-I00), and Comunidad de Madrid (Y2020/BIO-6519 and S2022/BMD-7437). VHIO would like to acknowledge the Cellex Foundation for providing research facilities and equipment and the CERCA Program from the Generalitat de Catalunya for their support on this research. CNIO (CEX2019-000891-S) and VHIO (CEX2020-001024-S/AEI/10.13039/501100011033) are Centers of Excellence Severo Ochoa (Agencia Estatal de Investigación, MICINN).

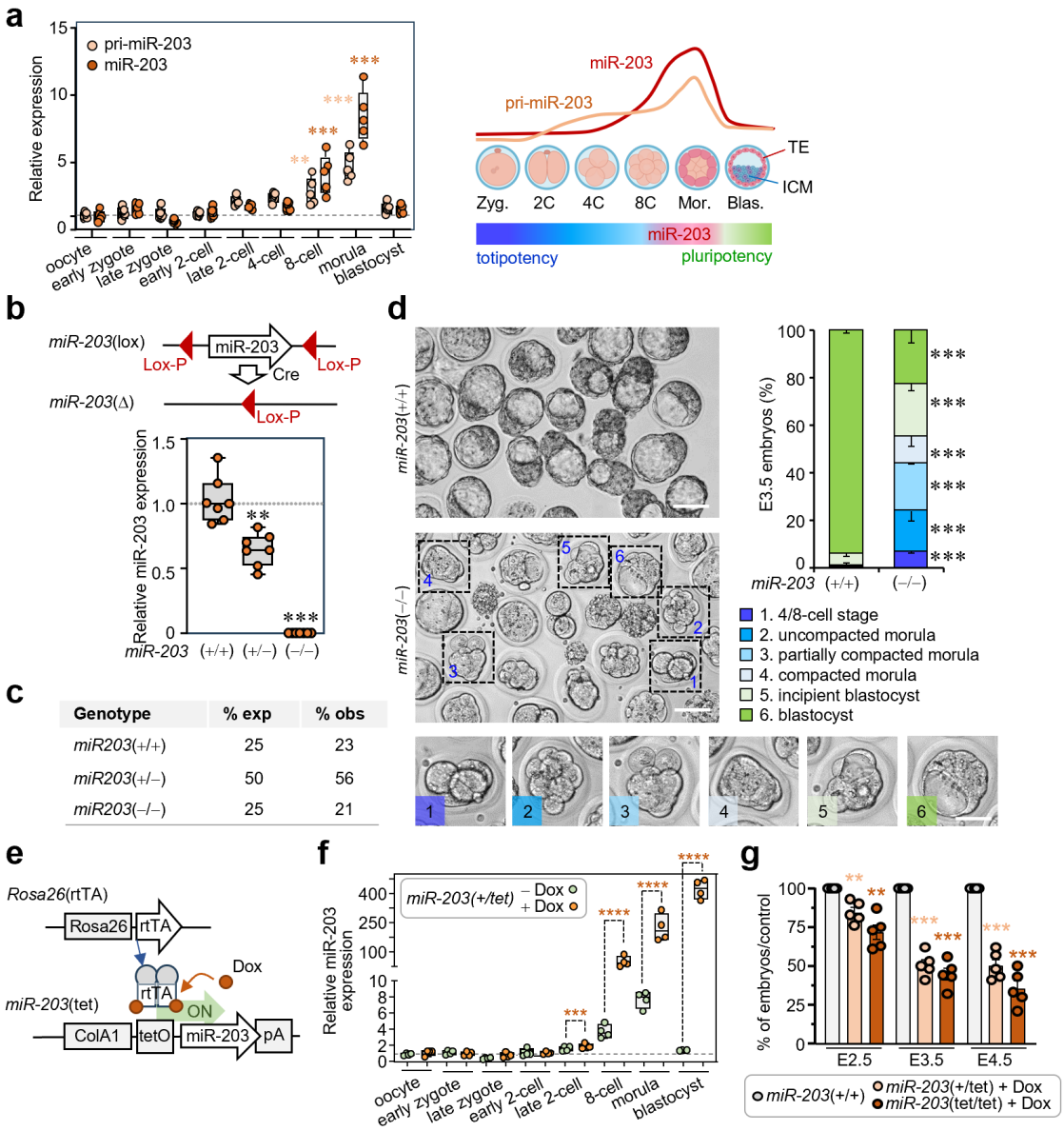
Author contributions

J.G.M.: Conceptualization, Methodology, Formal analysis, Investigation, Supervision, Writing; E.A.: Methodology, Formal analysis, Investigation; A.S.B.: Methodology, Formal analysis, Investigation, Writing – Review & Editing; A.G.: Methodology, Formal analysis, Investigation; D.G.: Methodology, Formal analysis, Investigation; A.L.: Methodology, Investigation; S.O.: Methodology, Formal analysis, Investigation, Writing – Review & Editing; M.M.: Conceptualization, Methodology, Formal analysis, Resources, Supervision, Writing, Project administration and Funding acquisition.

Conflicts of interest

The authors declare no conflict of interest.

Figures



Genotype	% exp	% obs
miR203(+/+)	25	23
miR203(+/–)	50	56
miR203(–/–)	25	21

Figure 1. **miR-203 regulates developmental timing during early embryogenesis**. **a**, Expression of *pri-mmu-mir-203* (pale brown) and *mmu-mir-203* (brown) at the indicated stages during early embryonic development in the mouse. Each data point represents 30 embryos of the stated developmental stage from the same female donor (n=150 total embryos/experimental group). RNA expression is normalized by a housekeeping miRNA or pri-miRNA (miR-16 or pri-miR-16) that is maintained invariable during early embryogenesis. The panel to the right shows a summary of pri-miR-203 and miR-203 expression during the indicated developmental stages. **b,** Strategy used to generate miR-203-deficient mice for this work (top; see Suppl. Figure 1b for details), and relative levels of miR-203 respect miR-16 as housekeeping control miRNA in wild-type, heterozygous

and homozygous null 8-cell stage embryos (20 embryos from the same female per data point). **c**, Quantification of the ratio of adult mice with the indicated genotypes in intercrosses between *miR-203*(+/–) mice. **d**. Representative images and quantification of developmental stages of *miR-203*(+/+) or *miR-203*(–/–) embryos flushed out from superovulated females at E3.5. n= 84 embryos miR-203 *miR-203*(+/+) and n= 169 embryos *miR-203*(–/–). Scale bars, 100 μm (50 μm in the bottom insets). **e**, Schematic representation of the miR-203-inducible knock-in model, [*ColA1*(miR-203/miR-203); *Rosa26*(rtTA/rtTA)], in which the reverse tetracycline transactivator is expressed from the *Rosa26* locus, whereas miR-203 is driven by the tetracycline operator downstream of the *ColA1* locus in the presence of doxycycline (Dox). **f**, Expression of *miR-203* in *miR-203*(+/tet) in the absence (green) or presence (orange) of Dox, compared to miR-16 levels in the indicated preimplantation stages, after treatment the mother with Dox when pug was detected. Each data point represents 25 embryos of the stated developmental stage from the same female donor (n=100 total embryos/experimental group). **g**, Percentage of embryos recovered after flushing oviducts (E2.5) or uteri (E3.5) of females of the indicated genotypes and treatments respect control *miR-203*(+/+) females. n=185 embryos (E2.5), 126 embryos (E3.5) and 86 embryos (E4.5). Data in a, b, g are analyzed by a 1-way ANOVA with Tukey multiple comparisons post-hot test, and data in d and f are analyzed by a student t-test with Welsh correction. **, $P<0.01$; ***, $P<0.001$ of each experimental group respect early zygote in a, *miR-203*(+/+) in b, d and g, and respect the – dox group in f.

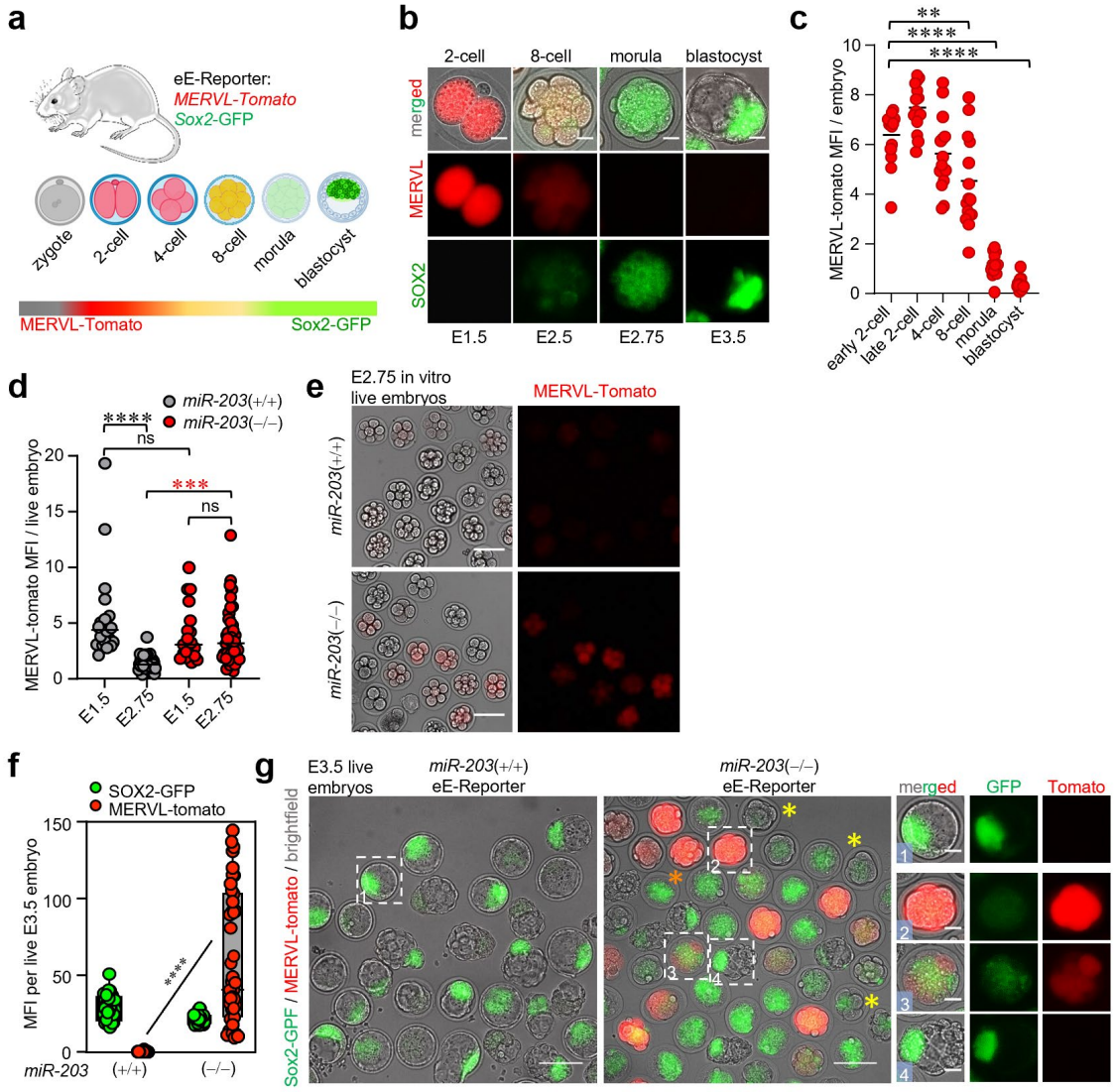


Figure 2. **Loss of *miR-203* slows down developmental pace and extends the activity of totipotency-associated endogenous retroviral elements in vivo. a**, Schematic representation of the eE-reporter mouse. **b**, Totipotent embryos with active MERVL retroviral elements display a peak of red fluorescence (td-tomato) in the 2-cell stage concomitant with the expression of totipotency genes, which decreases upon after the exit of totipotency in the 4-cell and 8-cell stages. Entry in pluripotency is reported by the expression of GFP under the control of the *Sox2* gene promoter at the morula and blastocyst stages. Scale bars: 20µm. **c**, Quantification of MERVL-tomato mean fluorescence intensity (MFI) in arbitrary units, per embryo, in the indicated embryonic preimplantation stages. Each data point corresponds to one embryo. Horizontal bars represent the mean of each biological group. **d**, Quantification of MERVL-tomato MFI in embryos of the indicated genotypes during E1.5 and E2.75 stages. Horizontal bars represent the mean of each biological group. Note the sustained MERVL-tomato MFI

during E2.75 stage in *miR-203* deficient morulae respect wildtype controls (red asterisks). **e**, Fluorescence micrographs of live E3.5 embryos harboring the eE-Reporter and *miR-203* alelles. Insets to the right represent a *miR-203*(+/+) blastocyst bearing GFP fluorescence in the ICM (1), and *miR-203*(–/–) embryos with intense td-tomato fluorescence and GFP abscense (2), mixed td-tomato and GFP expression (3), and GFP expression only (4). Yellow asterisks indicate embryos in the morula stage which already lost MERVL-tomato expression. All embryos harbor the *MERVL-tomato* and Sox2-GFP alleles in the same alellic dose. Scale bars: 100µm (micrographs) and 25µm (insets). **f**, Quantification of green (*Sox-GFP*) and red (*MERVL-tomato*) mean fluorescence intensity in arbitrary units (MFI) of the embryos displayed in **d**. Each data point corresponds to one embryo. **g**, Live fluorescence micrographs of the E2.75 morulas of the indicated genotypes. Scale bars: 100µm. Data in **c**, **d**, **e** are analyzed by a 1-way ANOVA with Tukey multiple comparisons post-hot test. **, $P<0.01$; ***, $P<0.001$, ****, $P<0.0001$, ns: non-significant within the indicated comparisons.

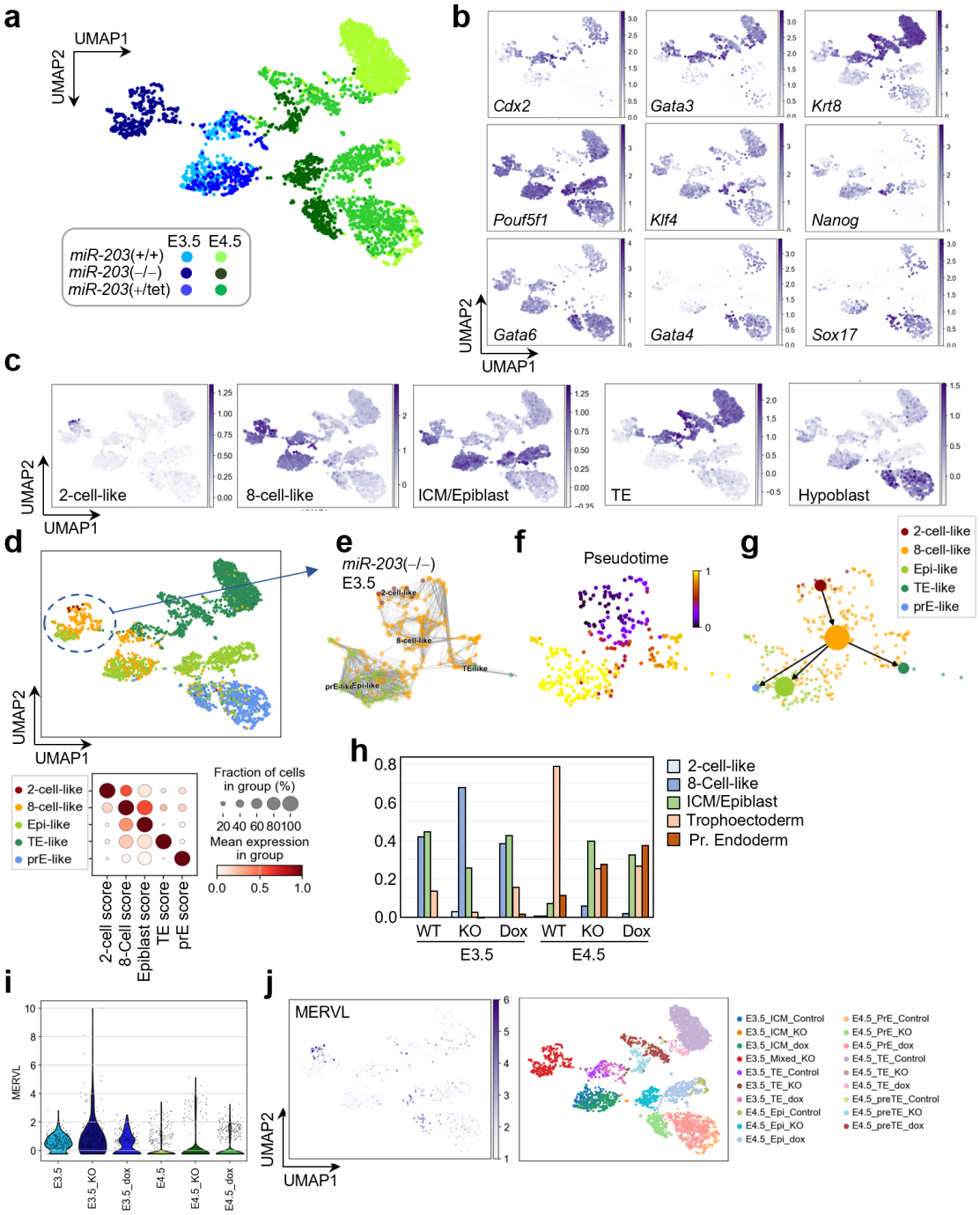


Figure 3. Single cell transcriptomics reveals a heterogeneous cluster of cells with delayed developmental timing and totipotency-associated transcriptional features upon genetic ablation of *miR-203 in vivo*. a, scRNA-seq analysis of *miR-203* wildtype, knockout and knockin embryos at E3.5 and E4.5. The UMAP plot displays six main clusters displayed below in different colors. **b**, UMAP plots displaying the expression of prototypical trophectoderm identity genes (*Cdx2*, *Gata3*, *Krt8*), inner cell mass / epiblast (*Pou5f1*, *Klf4*, *Nanog*), and primitive endoderm identity genes (*Gata6*, *Gata4*, *Sox17*). Scale bars indicate gene expression. **c**, UMAP plots showing the expression of distinctive

gene signatures for 2-cell-like, 8-cell-like, inner cell mass (ICM) / epiblast, throphectoderm (TE) and hypoblast identity transcriptional programs. Scale bars indicate the gene expression. **d**, UMAP plot representing the distinct cell populations presented in **a** classified according to the lineage identity gene signatures stated in **c**. Note the presence of a hybrid, heterogeneous cluster constituted by *miR-203*(–/–) cells exhibiting 2-cell, 8-cell and ICM transcriptional features which cluster together (dashed circle). The dot plot below represents the numeric score for each one of the signatures in each cell cluster. **e**, Detailed UMAP plot of the *miR-203*(–/–) E3.5 cell cluster showing the distribution of each cell type according to the expression of the indicated lineage identity gene signatures. **f**, Pseudotime analysis projected on the *miR-203*(–/–) E3.5 cell cluster. **g**, PAGA velocity analysis delineating the main developmental trajectories within the *miR-203*(–/–) E3.5 cell cluster. **h**, Bar plot representing gene expression for each lineage identity gene signature stated in the different cell clusters. **i**, Violin plot depicting MERVL expression across the different cell clusters displayed in **a**. Note the increased expression of MERVL in *miR-203*(–/–) E3.5 embryos. **j**, UMAP plots showing the expression MERVL retroelements projected to the UMAP plot (left) and the separation of cells by stage, genotype and cell lineage (right).

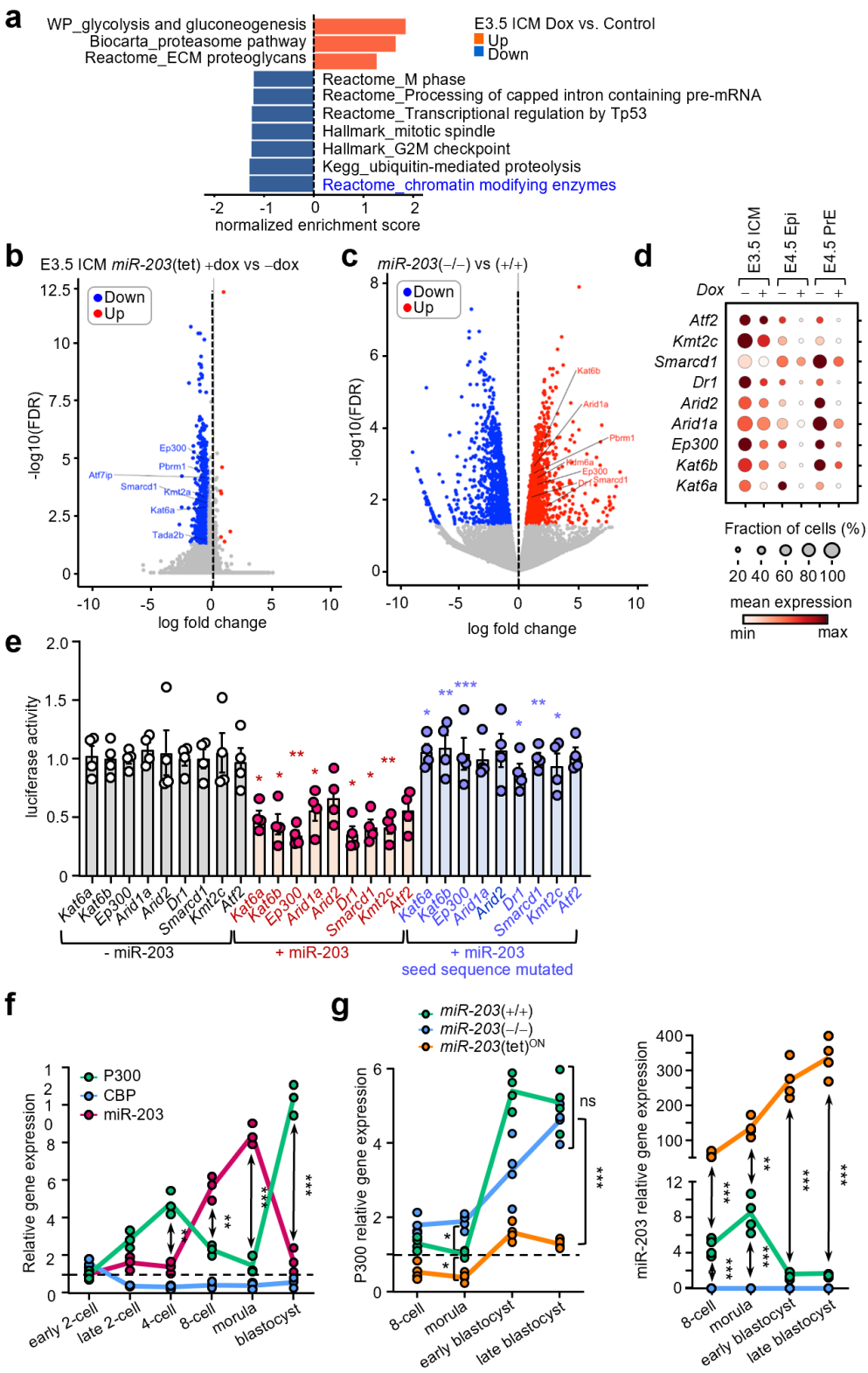


Figure 4. Histone acetyltransferase enzymes are direct biochemical targets of miR-203. **a**, Top categories in the Gene Ontology analysis of the genes significantly up- and downregulated in the ICM of E3.5 blastocysts developed in utero in the presence or absence of doxycycline to conditionally induce *miR-203* expression. Note that the go-terms for chromatin modifying enzymes are amongst the most downregulated in these

samples. **b**, RNA-seq analysis of the inner cell mass (ICM) of E3.5 *miR-203*(–/–) blastocysts. The volcano plot displays the gene expression change, red and blue dots indicate upregulated and downregulated genes, respectively. Grey dots are non-significant (NS). Note the upregulation of key genes encoding histone modification enzymes, including the H3K27 acetyltransferase P300. **c**, Volcano plot displaying the gene expression changes in E3.5 blastocysts overexpressing miR-203 respect wiltype controls. Note that several genes encoding members of the histone acetyltransferase enzymes family are downregulated, especially *Ep300*. **d**, Dot plot depicting the expression of individual genes contained in the Go-terms for histone modifying enzymes in embryos from the stated genotypes and treatments (+ dox: +; – dox: –). Scale bar represents the mean gene expression. **e**, Bar plot displaying luciferase activity (fold increase in arbitrary units) in HEK293T cells transfected with the corresponding luciferase vectors containing a wildtype or mutated seed sequence within the 3'UTR region of the predicted miR-203 target transcripts. Each data point represents a biological replicate of the assay, and error bars indicate SEM. **f**, Gene expression graph showing the expression levels of the genes *Cbp* and *Ep300* (respect *Hprt* as a housekeeping control) and miR-203 (respect miR-16 as housekeeping control miRNA). Note the antagonistic pattern of gene expression displayed between *Ep300* and *miR-203,* and that the expression of *Cbp* remains stably low during preimplantation development. **g**, Relative expression of *Ep300* respect *Hprt* as a housekeeping control in early embryos of the indicated genotypes. Note the different levels of *Ep300* transcripts upon genetic manipulation of *miR-203* within the morula stage and the clearance of these differences with the course of development. **h**, Relative expression of *miR-203* within embryos isolated in the different preimplantation stages stated. In **f**, **g**, **h** each data point represents one biological replicate containing a minimum of 20 embryos of the indicated stage. Data in **e** - **h** are analyzed by a 1-way ANOVA with Tukey multiple comparisons post-hot test. *, $P<0.05$, **, $P<0.01$, ***, $P<0.001$, ****, $P<0.0001$, ns: non-significant. In e, magenta asterisks represent the comparisons between the – miR-203 and + miR-203 groups; and blue asterisks the comparisons between the + miR-203 and + miR-203/seed sequence mutated groups.

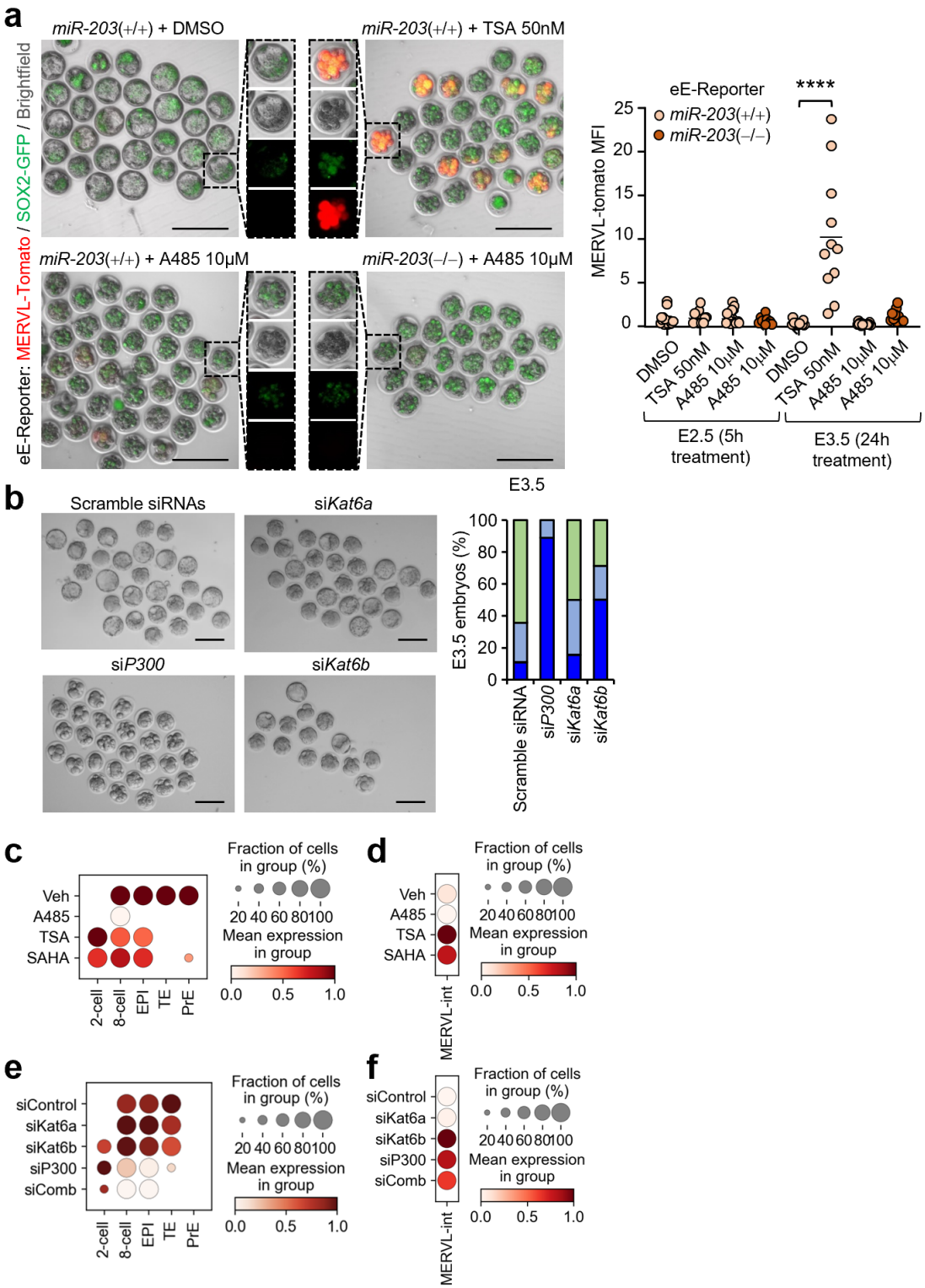


Figure 5. Direct manipulation of H3K27 histone acetyltransferase and deacetylase activity mimics miR-203 phenotypes in early embryos. **a**, Brightfield and fluorescence micrographs of live embryos of the stated genotypes isolated at E2.75 and treated with the indicated small molecule inhibitors for 24h. Plot to the right depicts the quantification of the MERVL-tomato mean fluorescence intensity per live embryo (MFI) after 5 and 24h

of treatment. **b**, Brightfield micrographs of preimplantation embryos injected with the indicated siRNAs at the zygote stage (E0.5) and developed in vitro for 3 days. Scale bar: 100µm. The bar plot to the right shows the quantification of the percentage of embryos in each of the embryonic stages listed after microinjection with siRNAs at the zygote stage (E0.5) and their development in vitro for 72h. **c-f**, Dot plots depicting the expression of the stated gene signatures (c, e) or MERVL sequences (d, f) for the indicated experimental conditions. Scale circles and bars represent the fraction of samples and the the mean gene expression per group, respectively. Data in **a** are analyzed by a 1-way ANOVA with Tukey multiple comparisons post-hot test. ****, $P<0.0001$.

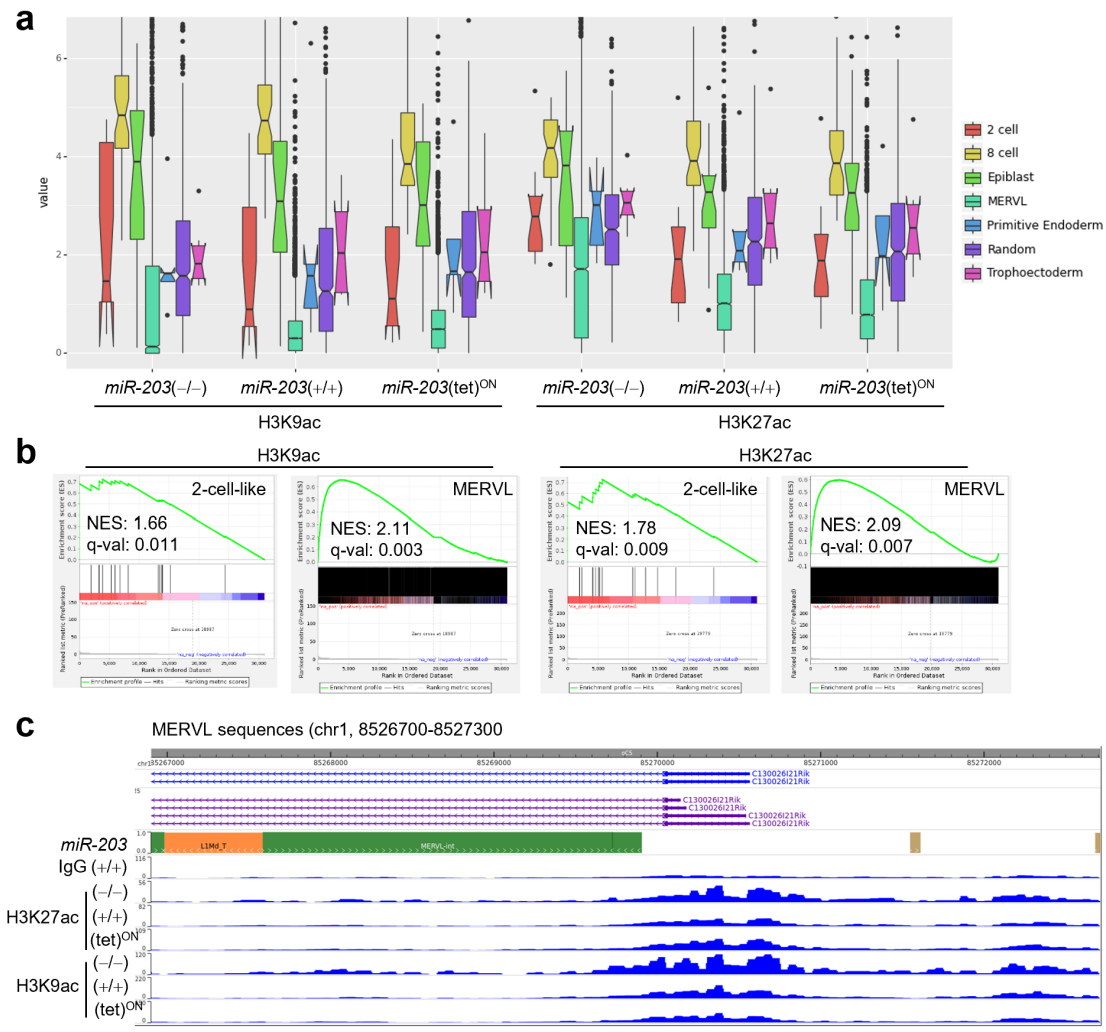


Figure 6. **Loss of miR-203 promotes increased histone acetylation in totipotency-associated genetic loci. a**, Box plot depicting the levels of histone H3 lysine 9 acetylation (H3K9ac) and histone H3 lysine 27 acetylation (H3K27ac) in genomic loci corresponding to genes contained in the signatures listed to the right within E3.5 embryos of the indicated genotypes and treatments. **b**, Gene enrichment profiles corresponding to H3K9ac and H3K27ac histone marks in 2-cell-like and MERVL gene signatures from the embryos analyzed in a. **c**, Genome view analysis of the indicated histone acetylation marks within the chromosomal region corresponding to one specific MERVL sequence. Note the increase in H3K9 and H3K27 acetylation upon miR-203 loss.

References

Asimi V, Sampath Kumar A, Niskanen H, Riemenschneider C, Hetzel S, Naderi J, Fasching N, Popitsch N, Du M, Kretzmer H *et al* (2022) Hijacking of transcriptional condensates by endogenous retroviruses. *Nat Genet* 54: 1238-1247

Bedzhov I, Leung CY, Bialecka M, Zernicka-Goetz M (2014) In vitro culture of mouse blastocysts beyond the implantation stages. *Nat Protoc* 9: 2732-2739

Bernstein E, Kim SY, Carmell MA, Murchison EP, Alcorn H, Li MZ, Mills AA, Elledge SJ, Anderson KV, Hannon GJ (2003) Dicer is essential for mouse development. *Nat Genet* 35: 215-217

Bueno MJ, Perez de Castro I, Gomez de Cedron M, Santos J, Calin GA, Cigudosa JC, Croce CM, Fernandez-Piqueras J, Malumbres M (2008) Genetic and epigenetic silencing of microRNA-203 enhances ABL1 and BCR-ABL1 oncogene expression. *Cancer Cell* 13: 496-506

Choi YJ, Lin CP, Risso D, Chen S, Kim TA, Tan MH, Li JB, Wu Y, Chen C, Xuan Z *et al* (2017) Deficiency of microRNA miR-34a expands cell fate potential in pluripotent stem cells. *Science* 355

Dahl JA, Jung I, Aanes H, Greggains GD, Manaf A, Lerdrup M, Li G, Kuan S, Li B, Lee AY *et al* (2016) Broad histone H3K4me3 domains in mouse oocytes modulate maternal-to-zygotic transition. *Nature* 537: 548-552

De Iaco A, Planet E, Coluccio A, Verp S, Duc J, Trono D (2017) DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat Genet* 49: 941-945

DeVeale B, Swindlehurst-Chan J, Blelloch R (2021) The roles of microRNAs in mouse development. *Nat Rev Genet* 22: 307-323

Dou C, Wu L, Zhang J, He H, Xu T, Yu Z, Su P, Zhang X, Wang J, Miao YL *et al* (2024) The transcriptional activator Klf5 recruits p300-mediated H3K27ac for maintaining trophoblast stem cell pluripotency. *J Mol Cell Biol* 15

Fang Z, Liu X, Peltz G (2023) GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 39

Gao L, Wu K, Liu Z, Yao X, Yuan S, Tao W, Yi L, Yu G, Hou Z, Fan D *et al* (2018) Chromatin Accessibility Landscape in Human Early Embryos and Its Association with Evolution. *Cell* 173: 248-259 e215

Gonzalez-Martinez J, Malumbres M (2022) Expanding the Differentiation Potential of Already-Established Pluripotent Stem Cells. *Methods Mol Biol* 2454: 95-107

Goodman RH, Smolik S (2000) CBP/p300 in cell growth, transformation, and development. *Genes Dev* 14: 1553-1577

Grosswendt S, Kretzmer H, Smith ZD, Kumar AS, Hetzel S, Wittler L, Klages S, Timmermann B, Mukherji S, Meissner A (2020) Epigenetic regulator function through mouse gastrulation. *Nature* 584: 102-108

He J, Babarinde IA, Sun L, Xu S, Chen R, Shi J, Wei Y, Li Y, Ma G, Zhuang Q *et al* (2021) Identifying transposable element expression dynamics and heterogeneity during development at the single-cell level with a processing pipeline scTE. *Nat Commun* 12: 1456

Hendrickson PG, Dorais JA, Grow EJ, Whiddon JL, Lim JW, Wike CL, Weaver BD, Pflueger C, Emery BR, Wilcox AL *et al* (2017) Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat Genet* 49: 925-934

La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lonnerberg P, Furlan A *et al* (2018) RNA velocity of single cells. *Nature* 560: 494-498

Liao Y, Smyth GK, Shi W (2019) The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. *Nucleic Acids Res* 47: e47

Lu X, Zhang Y, Wang L, Wang L, Wang H, Xu Q, Xiang Y, Chen C, Kong F, Xia W *et al* (2021) Evolutionary epigenomic analyses in mammalian early embryos reveal species-specific innovations and conserved principles of imprinting. *Sci Adv* 7: eabi6178

Macfarlan TS, Gifford WD, Driscoll S, Lettieri K, Rowe HM, Bonanomi D, Firth A, Singer O, Trono D, Pfaff SL (2012) Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487: 57-63

Meers MP, Bryson TD, Henikoff JG, Henikoff S (2019) Improved CUT&RUN chromatin profiling tools. *Elife* 8

Menchero S, Rayon T, Andreu MJ, Manzanares M (2017) Signaling pathways in mammalian preimplantation development: Linking cellular phenotypes to lineage decisions. *Dev Dyn* 246: 245-261

Michel CI, Malumbres M (2013) microRNA-203: Tumor Suppression and Beyond. *Microrna* 2: 118-126

Millan-Zambrano G, Burton A, Bannister AJ, Schneider R (2022) Histone post-translational modifications - cause and consequence of genome function. *Nat Rev Genet* 23: 563-580

Rodriguez CI, Buchholz F, Galloway J, Sequerra R, Kasper J, Ayala R, Stewart AF, Dymecki SM (2000) High-efficiency deleter mice show that FLPe is an alternative to Cre-loxP. *Nature genetics* 25: 139-140

Roelfsema JH, White SJ, Ariyurek Y, Bartholdi D, Niedrist D, Papadia F, Bacino CA, den Dunnen JT, van Ommen GJ, Breuning MH *et al* (2005) Genetic heterogeneity in Rubinstein-Taybi syndrome: mutations in both the CBP and EP300 genes cause disease. *Am J Hum Genet* 76: 572-580

Sakashita A, Kitano T, Ishizu H, Guo Y, Masuda H, Ariura M, Murano K, Siomi H (2023) Transcription of MERVL retrotransposons is required for preimplantation embryo development. *Nat Genet* 55: 484-495

Salazar-Roa M, Trakala M, Alvarez-Fernandez M, Valdes-Mora F, Zhong C, Munoz J, Yu Y, Peters TJ, Grana-Castro O, Serrano R *et al* (2020) Transient exposure to miR-203 enhances the differentiation capacity of established pluripotent stem cells. *EMBO J* 39: e104324

Schwenk F, Baron U, Rajewsky K (1995) A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic acids research* 23: 5080-5081

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES *et al* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102: 15545-15550

Suh H, Consiglio A, Ray J, Sawai T, D'Amour KA, Gage FH (2007) In vivo fate analysis reveals the multipotent and self-renewal capacities of Sox2+ neural stem cells in the adult hippocampus. *Cell Stem Cell* 1: 515-528

Van Gils J, Magdinier F, Fergelot P, Lacombe D (2021) Rubinstein-Taybi Syndrome: A Model of Epigenetic Disorder. *Genes (Basel)* 12

Wang M, Chen Z, Zhang Y (2022) CBP/p300 and HDAC activities regulate H3K27 acetylation dynamics and zygotic genome activation in mouse preimplantation embryos. *EMBO J* 41: e112012

Wang Y, Medvid R, Melton C, Jaenisch R, Blelloch R (2007) DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* 39: 380-385

Wilkinson AL, Zorzan I, Rugg-Gunn PJ (2023) Epigenetic regulation of early human embryo development. *Cell Stem Cell* 30: 1569-1584

Wolf FA, Angerer P, Theis FJ (2018) SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* 19: 15

Yang F, Huang X, Zang R, Chen J, Fidalgo M, Sanchez-Priego C, Yang J, Caichen A, Ma F, Macfarlan T *et al* (2020a) DUX-miR-344-ZMYM2-Mediated Activation of MERVL LTRs Induces a Totipotent 2C-like State. *Cell Stem Cell* 26: 234-250 e237

Yang J, Bashkenova N, Zang R, Huang X, Wang J (2020b) The roles of TET family proteins in development and stem cells. *Development* 147

Yi R, Poy MN, Stoffel M, Fuchs E (2008) A skin microRNA promotes differentiation by repressing 'stemness'. *Nature* 452: 225-229

Yoo W, Song YW, Kim J, Ahn J, Kim J, Shin Y, Ryu JK, Kim KK (2023) Molecular basis for SOX2-dependent regulation of super-enhancer activity. *Nucleic Acids Res* 51: 11999-12019

Zhu P, Guo H, Ren Y, Hou Y, Dong J, Li R, Lian Y, Fan X, Hu B, Gao Y *et al* (2018) Single-cell DNA methylome sequencing of human preimplantation embryos. *Nat Genet* 50: 12-19