**Highlights**

- Allele-specific reporters revealed dynamic DNA methylation of Sox2 and miR290 SEs
- DNMTs and transcription factor binding regulate methylation dynamics
- SE DNA methylation directly regulates transcription \textit{in cis}
- Dynamic DNA methylation is co-regulated with MED1 recruitment and H3K27ac level

**Authors**

Yuelin Song, Patrick R. van den Berg, Styliani Markoulaki, ..., Stefan Semrau, Yonatan Stelzer, Rudolf Jaenisch

**Correspondence**

young@wi.mit.edu (R.A.Y.), semrau@physics.leidenuniv.nl (S.S.), yonatan.stelzer@weizmann.ac.il (Y.S.), jaenisch@wi.mit.edu (R.J.)

**In Brief**

Song et al. used an allelic reporter approach to show super-enhancers allelic DNA methylation dynamics underlies locus-specific heterogeneity, which functionally impacts transcription and cellular states of mouse embryonic stem cells.
Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs

Yuelin Song, 1,2 Patrick R. van den Berg, 4 Styliani Markoulaki, 1 Frank Soldner, 1 Alessandra Dall’Agnese, 1 Jonathan E. Henninger, 1 Jesse Drotar, 1 Nicholas Rosenau, 1 Malkiel A. Cohen, 1 Richard A. Young, 1,2,* Stefan Semrau, 4,* Yonatan Stelzer, 3,* and Rudolf Jaenisch 1,2,5,*

1Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA
2Biology Department, Massachusetts Institute of Technology, Cambridge, MA 02142, USA
3Department of Molecular Cell Biology, Weizmann Institute of Science, 76100 Rehovot, Israel
4Leiden Institute of Physics, Leiden University, 2300 RA Leiden, the Netherlands
5Lead Contact
*Correspondence: young@wi.mit.edu (R.A.Y.), semrau@physics.leidenuniv.nl (S.S.), yonatan.stelzer@weizmann.ac.il (Y.S.), jaenisch@wi.mit.edu (R.J.)
https://doi.org/10.1016/j.molcel.2019.06.045

SUMMARY
Variable levels of DNA methylation have been reported at tissue-specific differential methylation regions (DMRs) overlapping enhancers, including super-enhancers (SEs) associated with key cell identity genes, but the mechanisms responsible for this intriguing behavior are not well understood. We used allele-specific reporters at the endogenous Sox2 and Mir290 SEs in embryonic stem cells and found that the allelic DNA methylation state is dynamically switching, resulting in cell-to-cell heterogeneity. Dynamic DNA methylation is driven by the balance between DNA methyltransferases and transcription factor binding on one side and co-regulated with the Mediator complex recruitment and H3K27ac level changes at regulatory elements on the other side. DNA methylation at the Sox2 and the Mir290 SEs is independently regulated and has distinct consequences on the cellular differentiation state. Dynamic allele-specific DNA methylation at the two SEs was also seen at different stages in pre-implantation embryos, revealing that methylation heterogeneity occurs in vivo.

INTRODUCTION
Tissue-specific differential methylation regions (T-DMRs) have been found to strongly associate with low CpG density and inter-genic enhancers (Ehrlich et al., 2016; Fleischer et al., 2017; Izzi et al., 2016; Jones, 2012; Rinaldi et al., 2016), and the vast majority of cell-type specific DNA methylation changes occur at distal regulatory elements (Luo et al., 2018; Stadler et al., 2011). Whole-genome bisulfite sequencing (WGBS) data indicate a low but detectable level of DNA methylation at T-DMRs overlapping active enhancers (Elliott et al., 2015; Heyn et al., 2016; Hon et al., 2013; Jiang et al., 2015; King et al., 2016; Shull et al., 2016; Stadler et al., 2011). Recent single-cell WGBS (scWGBS) data from mouse embryonic stem cells (ESCs) and the early mouse embryo suggest that the variable low-to-intermediate DNA methylation levels found at enhancer regions in bulk-cell measurements are largely due to averaging signals across cells with heterogeneous methylation states (Cheow et al., 2015; Guo et al., 2013, 2015, 2017; Rulands et al., 2018; Smallwood et al., 2014). However, due to the static snapshot view of sequencing-based methods, it has been difficult to define the basis, regulation, and functional impact of DNA methylation heterogeneity on gene expression and cellular states.

The hierarchy and casual relationship between the regulation of enhancer DNA methylation, active enhancer histone marks, transcription factor (TF) binding, and cis-regulated transcription has been challenging to define due to the epigenetic heterogeneity among cells (Jin et al., 2011; King et al., 2016; Zhu et al., 2016). While genome-wide epigenetic profiling provided insights into the relationship between DNA methylation, histone marks, and TFs and coactivators binding (King et al., 2016; Kundaje et al., 2015; Wilson and Filipp, 2018), these approaches, even at the single-cell level, did not allow resolving fast dynamics of individual epigenetic processes in heterogeneous tissues and cell populations. Thus, currently there is no clear understanding of the basis, regulation, and functional consequences of DNA methylation heterogeneity.

Our recently developed Reporter of Genome Methylation (RGM) allows tracing of locus-specific DNA methylation based on the on-and-off of a fluorescent signal in single cells in real time, and has been shown to faithfully reflect the endogenous DNA methylation states at multiple genomic loci (Stelzer et al., 2015, 2016). This system allows for robustly tracking locus-specific DNA methylation at enhancer regions and for functionally dissecting the hierarchy of epigenetic events that regulate enhancer activity and cellular states, overcoming the challenges faced by bulk measurements or sequence-based methods. We
Please cite this article in press as: Song et al., Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs, Molecular Cell (2019), https://doi.org/10.1016/j.molcel.2019.06.045
utilized this system at two pluripotency super-enhancers (SEs), Sox2 and Mir290 SEs, in ESCs. Both SEs overlap with ESC-specific DMRs, which display consistently low levels of methylation, indicating potential heterogeneity (Kobayashi et al., 2012; Leung et al., 2014; Rulands et al., 2018; Seisenberger et al., 2012; Stadler et al., 2011). We targeted RGMs to both alleles of the two SEs in F1 129xCastaneous (129xCAST) hybrid ESCs allowing to visualize allele-specific DNA methylation changes. We observed highly dynamic switching between different methylation states on individual alleles resulting in cell-to-cell heterogeneity and were able to distinguish the DNA methylation pathways driving these changes. The RGM system enables isolation of rare and transient populations exclusively based on their locus-specific methylation states, which allowed defining the relationship between dynamic SE DNA methylation changes, the Mediator complex condensation, histone H3K27 acetylation, TF binding, cis-regulated target gene expression, and changes in cellular states. Finally, transgenic methylation reporter mice for both SEs revealed the previously underappreciated epigenetic heterogeneity and dynamics of the pluripotent cells in cleavage embryos, recapitulating and extending the observations in ESCs.

RESULTS

DNA Methylation at the Sox2 and Mir290 SEs Is Heterogeneous at the Allelic Level

Sox2 and Mir290 SEs reside on chromosome 3 and 7, respectively. Both SEs overlap with T-DMRs, which are hypo-methylated in ESCs but become de novo methylated upon differentiation (Stelzer et al., 2015). The T-DMR of the Sox2 SE is located about 100 kb upstream of the Sox2 gene, whereas the Mir290 SE, consisting of hypo-methylated DMR constituents interspersed by small hyper-methylated regions, is proximal to the Mir290-295 cluster (Figure S1A). WGBS of ESCs indicates that the Sox2 and Mir290 SE DMRs have overall DNA methylation levels higher than that of hypo-methylated promoters of highly expressed genes in ESCs, such as Gapdh and Oct4, but lower than that of imprinting control regions or retroelements, which are monoallelically and hyper-methylated, respectively (Figure S1B) (Kobayashi et al., 2012; Leung et al., 2014; Seisenberger et al., 2012; Stadler et al., 2011). This low-to-intermediate level of methylation at both SEs in bulk cell WGBS suggests that they are hypermethylated in a small population of cells. Re-analysis of published scWGBS data (Smallwood et al., 2014) revealed that the T-DMRs of both SEs belong to the 5% regions with the most variable DNA methylation level compared to other regions of chromosome 7 or chromosome 3 (Figure S1C), further supporting the presence of rare cells with hypermethylated SE DMRs.

Consistent with published scWGBS studies reporting heterogeneity in the wild-type genome (Guo et al., 2013; 2015; Guo et al., 2017; Hu et al., 2016; Singer et al., 2014; Smallwood et al., 2014), we previously observed methylation heterogeneity in ESCs with the endogenous Nanog tagged with eGFP and RGM-tdTomato reporter inserted mono-allelically into the Sox2 or Mir290 SE DMRs (Stelzer et al., 2015). The heterogeneity at these two specific loci was manifested by the bi-modal distribution of RGM activity in Nanog positive (Nanog+) pluripotent cells as seen in fluorescence-activated cell sorting (FACS) (Figure 1A). Sorting cells based on florescence intensity, followed by bisulfite PCR (BS-PCR) and sequencing, validated that RGM methylation strictly correlates with the endogenous methylation in both regions (Figure 1A). Analyzing the Sox2 SE revealed that hypermethylation occurred on both the targeted and the untargeted alleles in the pluripotent ESC population (Nanog+), indicating that rare allelic methylation exists among cells (Figure S1D). The rare methylated alleles were also detected at the Mir290 SE by high-throughput sequencing of BS-PCR amplicons from the wild-type allele. Figure 1B shows that, comparing to Dnmt3a/b double-knockout cells (described later in Figure S3A), we found methylation at the Mir290 SE in non-manipulated wild-type ESCs as well as on the untargeted allele in the Nanog+RGM+ ESCs. These results indicate that SE DNA methylation heterogeneity is created by allele-specific hypermethylation in rare ESC populations independent of RGM targeting. To track DNA methylation heterogeneity on each allele, we targeted the Mir290 and the Sox2 SE independently in 129xCastaneous F1 hybrid ESCs with allele-specific RGM reporters and generated two cell lines, Sox2-129SE-RGM-tdTomato/Sox2-CASTSE-RGM-eGFP (abbreviated below as SOX2-SE-TG) and Mir290-129SE-RGM-tdTomato/Mir290-CASTSE-RGM-eGFP (abbreviated below as Mir290-SE-TG) (Figures 1C and S1E) allowing to visualize the SE locus-specific DNA methylation state at allelic and single-cell resolution. These cell lines also enabled dissection of allelic functional output of SE methylation states by distinguishing the two alleles based on the abundance of 129 or CAST allele-specific SNPs at both the DNA and the mRNA level.

The initial FACS analysis detected a small fraction of single-positive (T’G+ T’G+) as well as double-negative (T’G-) cells in both cell lines, though the majority of cells were double-positive (T’G+) (Figure 1D), consistent with the heterogeneity reported in scWGBS data by others (Figure S1C) and in our BS-PCR

Figure 1. DNA Methylation at the Sox2 and Mir290 SEs Is Heterogeneous at the Allelic Level
(A) Left, DNA methylation heterogeneity at both the Sox2 and the Mir290 SE in v6.5-Nanog-eGFP ESC where the RGM-tdTomato reporter was mono-allelically targeted. Right, BS-PCR followed by sequencing of the Sox2 SE in different populations of the bimodal distribution.
(B) Average methylation percentage and standard errors were quantified from high-throughput sequencing of BS-PCR amplicons of the Mir290 SE wild-type alleles in Dnmt3a/b double-knockout ESCs, in Nanog+ RGM+ ESCs and in untargeted wild-type ESCs. BS-PCRs were amplified allele-specifically as illustrated from potential epigenetic states indicated above. Standard error was estimated assuming number of methylated counts as a binomial random variable.
(C) Targeting strategy for generating SOX2-SE-TG and MIR290-SE-TG ESCs using CRISPR/Cas9 and targeting vectors. Methylation tracks from (Stadler et al., 2011) were used as the genome reference with blue bars highlighting the DMRs of the two SEs. Red tracks, 129 allele; green tracks, CAST allele.
(D) FACS analysis of CASTx129 F1 ESC clones targeted with allele-specific RGMs at either the Mir290 or the Sox2 SE.
(E) Allele-specific BS-PCR of the SEs with RGM (Snprn-tdTomato or Snprn-eGFP) in single PCR amplicons followed by Sanger sequencing in sorted cells from both SOX2-SE-TG and MIR290-SE-TG. See also Figure S1.
analysis on both targeted and wild-type alleles (Figures 1A, 1B, and S1D). To confirm that the RGM reporter activity faithfully reflected the allele-specific endogenous DNA methylation state, we sorted the four populations and performed allele-specific BS-PCR followed by Sanger sequencing of the DMRs upstream of the reporters. Figure 1E shows that the reporter activities on both alleles were consistent with the DNA methylation levels of the genomic SE regions and the inserted RGMs in all sorted populations. Quantitative pyro-sequencing further confirmed that T°G° and T°G-targeted SEs undergo the predicted methylation changes when exiting pluripotency (Figure S1G).

**Dynamic Allele-Specific SE DNA Methylation Is Regulated by De Novo Methylation and Passive Demethylation during Cell Proliferation**

To gain insights into the origin of DNA methylation heterogeneity, we FACS sorted equal numbers of the four populations from both reporter cell lines and monitored the RGM activity upon passing in serum + LIF medium (Figure 2A). Figure 2B (serum + LIF) and Figure S2A show that the SE DNA methylation states in the four sorted populations were not stable but highly dynamic with each allele independently switching the RGM on-and-off over the course of only a few days. This indicates that the observed SE DNA methylation heterogeneity is a result of fast dynamic and reversible switching of allelic DNA methylation states. When sorted cells were passaged and cultured in “2i” (GSKi and MAPKi) medium, the kinetics of the transitions between different methylation states was significantly altered with slowed de novo methylation for both SEs and an initial acceleration of demethylation states was significantly altered with slowed de novo methylation (Figure S1F). As expected, both unmethylated alleles in sorted T°G° cells from both cell lines gained methylation synchronously upon retinoic acid (RA)-induced differentiation. This confirms that the RGM-targeted SEs undergo the predicted methylation changes after disruption of these genes. When Dnmt1 or Uhrf1 were disrupted, the 129SE-RGM-tdTomato allele became demethylated in a substantial fraction of cells (Figure 3E). In contrast, transduction of sgRNAs against enzymes implicated in mediating active demethylation (Tets/Tdg/Aid), Figure 3D shows the predicted outcomes of 129SE-RGM-tdTomato allele demethylation (changes of the fraction of T°G° cells) after disruption of these genes. When Dnmt1 or Uhrf1 were disrupted, the 129SE-RGM-tdTomato allele became demethylated in a substantial fraction of cells (Figure 3E).

Demethylation in “2i” suggests that changes in DNA methyltransferase (Dnmt) activities modulate the observed dynamics. To determine the main de novo methyltransferase driver for SE methylation, we compared RGM activities in Dnmt3a or Dnmt3b single-knockout and Dnmt3a/3b double-knockout (DKO) cells (Figure S3A). Although the number of RGM negative cells was reduced in Dnmt3a or Dnmt3b single-knockout cells, cells with methylated SEs were eliminated only in the absence of both de novo methyltransferases in DKO cells preventing any de novo methylation (Figures 3A and S3B). The hypomethylation of both SEs was further confirmed by pyro-sequencing in DKO ESCs as well as in cells induced to differentiate by RA (Figure S3C). These results suggest that both DNMT3A and DNMT3B have redundant functions and independently contribute to de novo methylation of SE DMRs.

DNA demethylation can occur either passively in rapidly dividing cells, caused by inhibition of DNMT1 or by active removal of the methyl group mediated by Tet enzymes and base excision repair (BER) pathways (Wu and Zhang, 2017). To assess whether demethylation of the SEs involved active or passive mechanisms, we analyzed whether DNA demethylation would be affected in cells upon delaying cell-cycle progression using thymidine block. In all three populations carrying at least one methylated allele, the kinetics of demethylation upon thymidine block was significantly decreased upon 3 days in culture (Figures 3B and 3C). This suggests that cell proliferation-driven passive demethylation is responsible for SE demethylation. To confirm this observation genetically, we transfected 129SE-RGM-tdTomato T°G° cells with Cas9 and single-guide RNAs (sgRNAs) against genes encoding the maintenance enzymes DNMT1/UHRF1, which upon downregulation would lead to genome-wide passive dilution of methylation. In addition, we used sgRNAs against enzymes implicated in mediating active demethylation (Tets/Tdg/Aid). Figure 3D shows the predicted outcomes of 129SE-RGM-tdTomato allele demethylation (changes of the fraction of T°G° cells) after disruption of these genes. When Dnmt1 or Uhrf1 were disrupted, the 129SE-RGM-tdTomato allele became demethylated in a substantial fraction of cells (Figure 3E).

To explore additional regulators of SE DNA methylation dynamics besides DNMTs activities and cell division, we investigated the impact of TF binding on the transition between DNA methylation states. Some TFs can serve as readers of DNA methylation or inducing changes to DNA methylation states upon binding to target sequences (Feldmann et al., 2013; Maurano et al., 2015; Yin et al., 2017; Zhu et al., 2016). The Sox2 SE harbors multiple enrichment sites for the master TFs OCT4 and NANOG in ESCs (Hnisz et al., 2013) (Figure 4A, top). We deleted enrichment sites for the two TFs (peak 1 for NANOG and 2 for both OCT4 and
OCT4) at the Sox2 SE DMR on either the 129SE-RGM-tdTomato or the CASTSE-RGM-eGFP allele using sgRNAs against allele-specific SNPs (Figure 4A, bottom) and generated ESC clones harboring allele-specific peak deletions (ΔPeak 1-CAST, ΔPeak 2-CAST, and ΔPeak 2-129 clones; Figure S4A). We sorted the T’G’ and T’G+ populations from these clones and monitored the

Figure 2. SE DNA Methylation Heterogeneity Is Created by Dynamic Switching of Methylation States
(B) FACS analyses on the dynamics of T’G’, T’G+, T’G+, and T’G– populations 4 days post-sorting for both MIR290-SE-TG in serum + LIF or “2i” medium.
(C) Quantifications of the dynamics of 4 sorted populations from MIR290-SE-TG in percentages change over time when cultured in the serum + LIF or the “2i” medium after sorting.
See also Figure S2.

Please cite this article in press as: Song et al., Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs, Molecular Cell (2019), https://doi.org/10.1016/j.molcel.2019.06.045
Please cite this article in press as: Song et al., Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs, Molecular Cell (2019), https://doi.org/10.1016/j.molcel.2019.06.045
re-establishment of allelic heterogeneity across deletion genotypes (Figure 4B). The fraction of T+G− or T G+ cells transitioning from T+G− or T G− cells were quantified as allelic de novo methylation or demethylation rates, respectively (Figure 4C). We found that both the 129SE-RGM-tomato and the CAST1-SE-RGM-eGFP allele exhibited a faster de novo methylation rate after deletion of its TF enrichment sites as compared to the intact wild-type allele (Figure 4D, top), indicating higher susceptibility to de novo methylation upon loss of TF binding. Similarly, the allele that had its TF enrichment site deleted showed a slower demethylation rate than the wild-type allele, indicating less resistance to maintenance methylation upon loss of TF binding (Figure 4D, bottom). To confirm that the observed RGM activity changes correspond to changes in DNA methylation, we performed BS-PCR followed by Sanger sequencing on sorted cells from ΔPeak 1-129 clones. This analysis confirmed that the methylation status of the endogenous SE region was consistent with that of the Snrpn promoter as well as RGM activities at allelic resolution after genetic manipulation (Figure S4B). The TF binding effect on methylation dynamics was seen not only in cloned cells but also in sorted T+G− cell population transfected with allele-specific sgRNAs against TF enrichment sites (Figure 4E). Consistent with the single-cell clone analyses, the allele with TF enrichment site deletion showed a faster de novo methylation rates than the wild-type allele that was not targeted by the sgRNAs (Figure 4F).

DNA Methylation Decreases MED1 Association with SEs, Enhancer-Promoter H3K27ac, and in cis Transcription of the Target Genes

We investigated whether the rapid changes in SE DNA methylation would dynamically affect target gene transcription. Promoter DNA methylation has long been associated with stable silencing of gene expression (Deaton and Bird, 2011; Dor and Cedar, 2018; Schübeler, 2015; Smith and Meissner, 2013); in comparison, enhancer methylation’s role in transcription is less well characterized. The Mediator complex has been shown to be dynamically involved in phase-separated condensates concentrating at SEs for transcription of key cell-identity genes (Sabari et al., 2018). Since SE DNA methylation is dynamically changing, we investigated whether different allelic methylation states affect association of MED1 condensates with the Mir290 SE. We performed DNA FISH at the Mir290 SE locus and MED1 immunostaining on sorted cell populations. Figure 5A and Figure S5A show that MED1 was not enriched at the methylated Mir290 SE as T+G− cell populations did not have DNA FISH foci that overlapped with MED1 enrichment as compared to cells in which at least one Mir290 SE was unmethylated. Since the Mediator complex interacts with both the SE and the promoter (Whyte et al., 2013), a loss of MED1 enrichment upon SE DNA methylation may affect promoter activity as well. We therefore performed H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) as a proxy epigenetic mark defining active enhancers and promoters on four sorted populations from both reporter cell lines. H3K27ac was significantly reduced at both methylated SE regions, as measured by total (Figures 5B and 5C, Sox2 SE and Mir290 SE boxes; Figure S5B, enhancer panels) as well as allele-specific H3K27ac enrichment (Figure S5C, enhancer panels). As expected, a decrease in H3K27ac was also observed at promoters residing on the same chromosome with the methylated SE (Figures 5B and 5C, Sox2 and Mir290 boxes, and Figures S5B and S5C, promoter panels) but not at adjacent regions (Figure S5B, adjacent regions panels). This demonstrates that SE methylation affects the promoter H3K27ac level, likely through a loss of enhancer-promoter communication.

To test whether synchronized H3K27ac changes upon transient DNA methylation at enhancers and promoters affects in cis target gene expression, we performed allele-specific qRT-PCR on the four sorted cell populations from both reporter cell lines. As shown in Figure 5D, methylation of either allele of the SEs resulted in decreased target gene expression on the same chromosome. However, the Sox2 SE and the Mir290 SE have different effects on the total expression level of their respective target genes. The suppressive effect of transient DNA methylation was independent and additive when either Mir290 SE allele was methylated (Figure 5E, left). In contrast, total Sox2 expression only significantly decreased when both Sox2 SE alleles were methylated (Figure 5E, right), and in single-positive cells only single-molecule RNA FISH (smFISH) could detect a slight decrease of Sox2 transcripts (Figures S5D and S5E), indicating a compensating mechanism on total Sox2 transcripts when one SE allele is methylated. Notably, DNA methylation at two SEs exclusively anti-correlated with their respective in cis target genes, and little difference is seen in Mir290-295 expression if cells were sorted based on the methylation state at the Sox2 SE locus and vice versa (Figure 5F). This indicates that the DNA methylation state of the two SEs switches independently of each other.

To determine whether SE methylation has a causal role in suppressing enhancer-promoter H3K27ac and transcription, we transfected Cas9-sgRNAs targeting Dnmt1 and Uhrf1 and removed DNA methylation in sorted T+G− MIR290-SE-TG cells to induce rapid passive demethylation (Figure 6A). Figure 6B shows that cells deficient for Dnmt1 or Uhrf1 displayed significantly faster demethylation resulting in a higher proportion of T+G− cells as compared to the control. Both acetylation of H3K27 at the SE (Figure 6C) and Mir290-295 expression (Figure 6D) were significantly increased upon Dnmt1/Uhrf1
Figure 4. TF Binding at SEs Promotes Demethylation and Inhibits De Novo Methylation

(A) Top, schematic representation of TF enrichment sites (based on the ChIP-seq data of NANOG [pink track, peak 1 and 2] and OCT4 [blue track, peak 2]. ENCODE: ENCSR779CZG and ENCSR392DGA) relative to the RGM targeted site (orange). Bottom, allele-specific deletions of individual peaks after overlapping NANOG (N) and OCT4 (O) ChIP tracks. Red: 129SE-RGM-tdTomato allele, green: CASTSE-RGM-eGFP allele. Scissors illustrate sgRNA targeting sites. ChIP-seq value is presented as fold-change-over-control.

(B) Experimental setup using cells with different allelic TF enrichment site deletions in assessing the impacts of TF binding on SE methylation dynamics.

(C) Top, T+G+ cells were sorted from the genotyped single-cell clones with allelic TF enrichment site deletions. Bottom, T–G– cells were sorted from the same clones.

(D) Quantification of allele-specific de novo methylation rates (top panels, T+G– or T–G+ cells derived from T+G+ cells) and demethylation rates (bottom panels, T+G– or T–G+ cells derived from T–G– cells) of the respective ESC clones compared to that of an unmodified parental wild-type clone (dotted line level).

(E) Bulk T+G+ cells were sorted from the SOX2-SE-TG cell line and transfected with allele-specific sgRNA pairs to delete TF enrichment sites or with empty vectors.

(F) Quantification of allele-specific de novo methylation rates of the bulk cells transfected with different sgRNAs.

See also Figure S4.
Figure 5. DNA Methylation Decreases MED1 Association at SE, Enhancer-Promoter H3K27ac and in cis Transcription of the Target Genes.

(A) Averaged DNA FISH (Magenta, Mir290 SE) and co-immunofluorescence staining (Green, MED1) signal in the nuclei of MIR290-SE-TG cells sorted based on allelic methylation states. Random spots were selected in the same image away from the DNA FISH spots.

(B) ChIP-seq data showing the binding of Sox2 and Sox2 SE at the Mir290 and Mir290 SE regions.

(C) Gene expression analysis showing the fold change of total Mir290 and Sox2 relative to T+G-.

(D) Bar graph illustrating the ratio of total Mir290-295 to total Mir290-SE-TG cell and total Sox2 to total Sox2-SE-TG cell.

(E) Bar graph showing the fold change of total Mir290 and Sox2 relative to T+G-.

(F) Fold change of total Mir290 and Sox2 relative to T+G-.

(legend continued on next page)
disruptions, as measured by ChIP-qPCR and qRT-PCR from the same cultures, respectively. This suggests that change in DNA methylation directly regulates SE function and transcription in cis.

Since correlating abundance in RNA allele-specific SNPs with allele-specific RGM activities allows distinguishing direct targets regulated in cis by the SE methylation status versus expression changes caused by secondary effects, we searched additional genomic targets on the same chromosomes that are directly regulated by SE methylation by allele-specific RNA sequencing (RNA-seq) analysis on sorted populations. We quantified allele-specific expression of genes on chromosome 3 (for MIR290-SE-TG) and chromosome 7 (for SOX2-SE-TG) in single positive cells and calculated the ratio between expressions from the allele with an unmethylated SE over that of the other allele with a methylated SE. We plotted this ratio of each gene calculated in T+G+ cells as the x-axis value and the ratio calculated in T+G– cells as the y-axis value (Figure S6A). As expected, Mir290-295 and Sox2 both appeared in the upper right corner as they were in cis directly suppressed by allelic SE methylation. Surprisingly, two antisense transcripts relative to Sox2 and Mir290-295, Ecm1 and AU018091, respectively, were oppositely regulated by allele-specific Sox2 or Mir290 SE methylation: SE hyper-methylation strongly correlated with upregulations of both anti-sense transcripts, whereas SE hypo-methylation correlated with inhibition (Figures S6B and S6C). This result shows that direct transcriptional targets of SE methylation are highly specific with possibly opposite effects on some cis-regulated genes. Though the detailed mechanism of such regulation remains to be elucidated, Ecm1 was upregulated in Sox2 SE deletion cells (Gagnon et al., 2014).

Our results suggest that DNA methylation at both SEs fluctuates independently and dynamically, altering Mediator complex condensates at the SE and allelic H3K27ac at enhancers and promoters in cis and ultimately leading to heterogeneous allelic transcription of the target genes (Figure 6E).

**Sox2 and Mir290 SE Methylation Heterogeneities Have Different Biological Impacts on ESC State**

Culture in “2i” medium has been shown to only allow naïve pluripotent cells to proliferate (Nichols and Smith, 2009). Long-term culture of MIR290-SE-TG and SOX2-SE-TG cells in “2i” after passingage from serum + LIF media, though favoring T+G+ population decreased but did not abolish heterogeneity completely (Figure S6D). The persistence of all four populations in both reporter lines indicates that DNA methylation at both SEs have different degrees of heterogeneity in different culture conditions. Both Sox2 and Mir290-295 are highly expressed in ESCs (Calabrese et al., 2007; Hnisz et al., 2013; Jaenisch and Young, 2008; Thomson et al., 2011), raising the possibility that allelic transcriptional heterogeneity caused by SE methylation heterogeneity may lead to co-existing heterogeneous cellular states of ESCs. In “2i” media, SOX2-SE-TG T+G+ cells exhibited significantly impaired colony-forming ability and proliferation (Figures 6F and 6G). However, under the same condition, the heterogeneous DNA methylation at the Mir290 SE did not lead to any obvious changes of ESCs, despite the slight colony formation disadvantage of MIR290-SE-TG T+G+ cells (Figures 6F and 6G). We further explored the functional differences among populations in vivo by injecting sorted cells to form teratomas. Surprisingly, despite the significant growth disadvantage of SOX2-SE-TG T+G+ population, they were able to contribute to all three germ layers in teratoma formation assays with no obvious contribution bias towards any germ layer compared to SOX2-SE-TG T+G+, MIR290-SE-TG T+G+, and T–G– cells (Figure S6E). This indicates that ESCs with biallelic methylation at the Sox2 SE are still pluripotent. However, when examined at the molecular level, these cells were distinct from other populations in principal-component analysis on difference in the 5% most highly variably expressed genes (Figure 6H) and 17,000 uniquely distinct H3K27ac enrichment peaks in ChIP-seq (Figure S6F). GO analysis on RNA-seq revealed that the SOX2-SE-TG T+G+ population preferentially expressed genes in differentiation-related pathways as compared to the MIR290-SE-TG T+G+ population (Figure S7A). The epigenetic and transcriptional differences of SOX2-SE-TG T+G+ cells indicate that these cells downregulate Sox2 expression and are prone to differentiate but not as yet committed to a certain fate. Our results are consistent with the notion that pluripotent ESC are heterogeneous as reflected by the dynamic allelic DNA methylation of key pluripotency SEs.

**DNA Methylation Is Dynamic at Both SEs in Blastocysts while Exhibiting Spatial-Temporal Differences in Pre-implantation Embryos**

In vivo, both Sox2 and Mir290-295 are expressed in preimplantation embryos. As reported previously Sox2 expression increases between the morula and the blastocyst stage (Mistrì et al., 2018) and Mir290-295 expression significantly upregulates at the 4-cell stage (Medeiros et al., 2011). To investigate changes in DNA methylation of the two SEs at single-cell and allelic resolution, we generated transgenic mice homozygous for the 129SE-RGM-tdTomato allele or the CASTSE-RGM-eGFP allele and obtained 2–4 cell embryos carrying one 129SE-RGM-tdTomato allele and one CASTSE-RGM-eGFP allele by mating animals homozygous for RGM-eGFP or RGM-tdTomato (Figure 7A). The two
Figure 6. DNA Methylation Directly Suppresses SE Activity and Affects ESC State

(A) Experimental setup for assessing the causal role of SE DNA methylation suppresses H3K27ac. FACS (DNA methylation), RT-qPCR (Mir290–295), and ChIP-qPCR (H3K27ac) were co-assessed from the same pool of cells from each sample.

(B) Loss of DNA methylation in MIR290-SE-TG T–G– cells 8 days post-Dnmt1 and Uhrf1 sgRNA transfection as compared to controls.

(C) H3K27ac ChIP-qPCR at the Mir290 SE from the experimental groups in (B), respectively. Data are represented as mean ± SD.

(D) Mir290-295 pri-miRNA level from the experimental groups in (B). Data are represented as mean ± SD.

(E) Summary of the dynamic regulation and functional impact of allelic SE methylation.

(F) Colony formation assays in "2i" starting from 100 sorted cells. Data are represented as mean ± SD.

(G) Growth curves measured by AlamarBlue Cell Viability Reagent. Data are represented as mean ± SD.

(H) Principal-component analysis of the top 5% highly variable genes from different populations of SOX2-SE-TG (Labeled as S. red: T+G–, green, T–G+, black: T–G–, yellow: T+G+) and MIR290-SE-TG (labeled as M; color code same as S).

See also Figures S5–S7.
Figure 7. DNA Methylation Is Dynamic at Both SEs in Blastocysts while Exhibiting Spatial-Temporal Differences in Pre-implantation Embryos
(A) Mating scheme for generating SOX2-SE-TG and MIR290-SE-TG mice and heterozygous pre-implantation embryos genetically carry 129SE-RGM-tdTomato and CASTSE-RGM-eGFP at the Sox2 or the Mir290 SE for imaging analyses.
(B) Live 4–8 cell (MIR290-SE-TG) and morula stage (SOX2-SE-TG) embryos.
(C) Live E3.5-E4.5 blastocysts of SOX2-SE-TG and MIR290-SE-TG in 10× low magnification, 40× high magnification, and 3D projections (left to right in each group). Red: tdTomato, green: eGFP, blue: Hoechst 33342.
(D) Tracking Sox2 SE DNA methylation dynamics in vivo. Columns are sorted and injected populations and rows are different imaging channels. Red: RGM-tdTomato; Green: RGM-eGFP; Cy5: Qtracker 705 was used to label and track injected cells. White arrows indicate demethylation, and yellow arrows indicate de novo methylation, at 2 days post-injection compared to 5 h post-injection. Channels were adjusted for brightness and contrast for optimal visibility. See also Figure S7.
SEs gained allelic DNA methylation heterogeneity at different times: reporter activity became apparent as early as the 4-cell stage for the Mir290 SE but only at the morula stage for the Sox2 SE (Figure 7B). At the blastocyst stage, Sox expression was restricted to the inner cell mass (ICM), whereas the Mir290-295 displayed broad expression in both ICM and trophoderm (TE) (Nichols and Smith, 2009; Paikari et al., 2017; Wicklow et al., 2014). Heterogeneous SE DNA methylation was consistent with the established spatial expression pattern of the two genes in blastocysts (Figure 7C). We further investigated whether the observed methylation heterogeneity was due to dynamic allelic methylation state switching in vivo. We sorted the four populations from SOX2-SE-TG and MIR290-SE-TG ESCs, injected each population into 8-cell stage wild-type CD1-IGS host embryos, and cultured embryos for 2 days to monitor de novo methylation or demethylation pattern at single-cell resolution (Figure S7B). A long-term membrane bound dye (Cy5) was used to track the injected cells (Figures 7D; Figure S7C). Figure 7D (SOX2-SE-TG cells) and Figure S7C (MIR290-SE-TG cells) show that, at the blastocyst stage, injected T+G- cells demethylated the SE as they turned on the RGMs on either or both alleles and became single positive or T+G+ cells (T+G- columns, white arrows). Demethylation also was observed in injected single-positive cells as the originally methylated allele at injection became unmethylated and cells became T+G+ (T+G- and T- G+ columns, white arrows). Similarly, dynamic de novo methylation was observed in vivo, as injected T+ G- cells shut down RGM activities on either or both alleles (T+G- columns, yellow arrows) and single-positive cells became T+ G+ cells (T+ G- and T- G+ columns, yellow arrows).

In summary, our data indicate that dynamic DNA methylation exists at active SEs in early preimplantation embryos creating locus-specific epigenetic heterogeneity, recapitulating and extending our observations in ESCs in vitro.

**DISCUSSION**

The importance of DNA methylation regulation at cis-regulatory elements is increasingly recognized as many developmental- and disease-associated DMRs overlap with these regions (Schultz et al., 2015; Weigel et al., 2016; Ziller et al., 2013). Locus-specific DNA methylation heterogeneity across cells has been shown by recent scWGBS as a potential explanation for the variable low-to-intermediate levels of methylation at active enhancers in bulk measurements. The present work was based on an experimental paradigm that overcomes some of the limitations of single-cell sequencing approaches using an allele-specific reporter system. This allowed us to address questions that were not resolved by previously used sequencing-based methods. (1) Our study shows that in ESCs the methylation state of the two alleles of the Sox2 and Mir290 SEs change dynamically and independently of each other. (2) We demonstrate that the dynamic change of SE DNA methylation is driven by the balance between three DNMTs and cell proliferation, with TF binding promoting the hypomethylated state. (3) We show that DNA methylation dynamically regulates target genes in cis and inhibits formation of Mediator complex condensates at the SE as well as enhancer-promoter H3K27 acetylation. (4) Allelic variation of SE DNA methylation, reflecting the epigenetic heterogeneity of ESCs, can originate from cells of different transcriptional landscapes and proliferative potentials as for the Sox2 SE or of developmentally identical states as for the Mir290 SE. (5) Finally, we show that dynamic DNA methylation is not only seen in cultured ESCs but also in preimplantation embryos.

Allele-specific RGM reporters targeted to the endogenous Sox2 and the Mir290 SEs allowed us to trace DNA methylation both in vitro and in vivo. Detailed analyses showed that the low levels of DNA methylation of the Sox2 and the Mir290 SE are due to the presence of a small fraction of cells with hypermethylated SE alleles. The methylation heterogeneity in these cells results from highly dynamic and reversible switching between allelic DNA methylation states. Because the RGM reporter allowed isolation of cells with defined allele-specific SE DNA methylation states, we were able to demonstrate that dynamic changes in SE DNA methylation are tightly anti-correlated in cis with enhancer-promoter H3K27ac levels. This is likely due to disruption of enhancer-promoter interactions consistent with the Mediator complex condensates showing decreased association at the methylated Mir290 SE. The Mediator complex and its unit MED1 have been shown previously to form condensates with liquid-like properties, which allows dynamic interactions with TFs and the transcription apparatus (Cho et al., 2018; Sabari et al., 2018). Our study shows that DNA methylation can affect these transcriptional condensates. Given the dynamic state switching of allelic SE DNA methylation as well as the dynamic nature of MED1 condensate formation, it is highly likely that one process mediates the other. We also show that SE DNA methylation can have opposing effects on transcription of different genes located on the same chromosome: the direct target genes Sox2 and Mir290-295 were repressed, while the antisense genes Ecm1 and AU018091 were activated by SE methylation. By removing DNA methylation at the Mir290 SE through Dnmt1/Uhrf1 deletion, we showed that changes in SE DNA methylation is a dynamic process actively regulating its transcriptional activity. By enabling sorting for a particular epigenetic state and combined with allelic expression analyses, we demonstrate that dynamic DNA methylation serves as an epigenetic basis for allelic heterogeneity in gene expression and that dynamic DNA methylation at SEs is a likely mechanism for dynamic random monoallelic transcription seen in mammalian cells (Deng et al., 2014; Reinius and Sandberg, 2015). However, it warrants further exploration to establish the causal link between allelic epigenetic and transcriptional heterogeneity in vivo.

While Sox2 and Mir290 SE methylation affect target gene expression similarly, we detected some differences on cellular growth and differentiation. Cells with biallelically methylated Sox2 SE revealed impaired growth and upregulation of differentiation-related pathways (Figures 6F and 6G; Figure S7A). In contrast, Mir290 SE methylation had little effects on cell state. We identified additional differences of how DNA methylation suppresses activity of the two SEs. Mir290-295 expression was independently suppressed by methylation at either Mir290 SE DMR allele consistent with the observation that individual DMR constituents have independent activities (Suzuki et al., 2017). In contrast, monoallelic Sox2 SE methylation did not...
significantly affect the overall Sox2 expression, suggesting additional regulatory mechanisms.

The experimental platform described here allows rapid tracing and isolating rare cell populations based on their transient methylation signatures at specific loci and thus can provide mechanistic insights into the nature of enhancer DNA methylation in heterogeneous cell populations both in vivo and in vitro in real time, which is difficult in sequencing-based approaches. Furthermore, this system enables manipulation of different molecular components to define interactions and hierarchies between layers of epigenetic regulation in dynamic systems with rapid changes. Our study provides a path towards the mechanistic understanding of dynamic T-DMR regulation in heterogeneous tissues and complex biological processes, such as development and diseases (Heyn and Esteller, 2012; Robertson, 2005).

STAR METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **LEAD CONTACT AND MATERIALS AVAILABILITY**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - ESC cell lines
  - Animals
- **METHOD DETAILS**
  - Southern blots
  - Flow cytometry
  - Bisulfite conversion-PCR (BS-PCR) and pyrosequencing
  - Retinoic acid differentiation
  - Double thymidine block
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - qRT-PCR and TaqMan assays
  - ChIP-qPCR
  - H3K27ac ChIP-Seq and analysis
  - RNA-Seq and analysis
  - RNA smFISH and image analyses
  - DNA FISH, Med1 IF and average image analyses
  - High-throughput sequencing of bisulfite PCR
- **DATA AND CODE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.molcel.2019.06.045.

ACKNOWLEDGMENTS

We thank George Bell, Prathapan Thiru, and Bingbing Yuan for their help in ChIP-seq analysis and BS-PCR sequencing analysis; Ruth Flannery and Dina Rooney for their help with animal husbandry, injections of the ESCs, and harvesting pre-implantation embryos; and Dongdong Fu for sectioning and processing of teratoma samples. We would like to thank Tom Volkert, Sumeet Gupta, Kevin Truong, Amanda Chiakla, and Jennifer Love of the Whitehead Genome Technology Core for their help in ChIP-seq; Wendy Salmon of the W.M. Keck Microscopy Facility for help with confocal microscopy; Glenn Paradis, Patti Wisniewski, Patrick Autissier, Michael Jennings, Michele Griffin, Mervelina Satumo-Condon, Hanna Aharonov, and Eleanor Kincaid of the Whitehead Institute and MIT flow cytometry facilities for their help with cell sorting. We thank Dr. Roderick Bronson and Kathleen Cormier at the KI Swanson Biotechnology Center Histology Core for teratoma sample consultation. We thank Raaji Alagappan, Tenzin Lungjangwa, and Carrie Garrett-Engele for their technical support. We thank Alicia V. Zamudio from Young Lab, Jian Shu, Shawn Liu, Haiteq Ma, Emile Wogram, and all of the members of the Jae- nisch lab for helpful discussions. Y.S. was supported by HFSP long-term fellowship, ISF grant no. 1610/18 and is the incumbent of the Louis and Ida Rich Career Development Chair. R.J. was supported by NIH grants HD 045022, 1U19AI131155, SR01MH104610, and 1R01GM114964.

AUTHOR CONTRIBUTIONS

Y. Song, Y. Stelzer, and R.J. conceived the project. Y. Stelzer and R.J. designed and supervised the experiments, S.S., R.A.Y., and R.J. acquired funding for this study. Y. Song conducted experiments, interpreted results, and wrote the manuscript with input from all authors. S.M., J.D., and N.R. contributed instrumentally to the writing of the manuscript. M.A.C. assisted with teratoma injection.

DECLARATION OF INTERESTS

R.J. is a cofounder of Fate Therapeutics, Fulcrum Therapeutics, and Omega Therapeutics and an advisor to Dewpoint Therapeutics. R.A.Y. is a founder and shareholder of Syros Pharmaceuticals, Camp4 Therapeutics, Omega Therapeutics, and Dewpoint Therapeutics.

Received: February 25, 2019
Revised: May 6, 2019
Accepted: June 26, 2019
Published: August 15, 2019

REFERENCES


### STAR METHODS

#### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit poly-clonal anti-H3K27ac</td>
<td>Abcam</td>
<td>Cat#ab4729; RRID:AB_2118291</td>
</tr>
<tr>
<td>Rabbit poly-clonal anti-MED1</td>
<td>Abcam</td>
<td>Cat#ab64965; RRID:AB_1142301</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Puromycin</td>
<td>Sigma Aldrich</td>
<td>Cat#P7255</td>
</tr>
<tr>
<td>Thymidine</td>
<td>Sigma Aldrich</td>
<td>Cat#T1895</td>
</tr>
<tr>
<td>Retinoic acid</td>
<td>Sigma Aldrich</td>
<td>Cat#R2625</td>
</tr>
<tr>
<td>LIF recombinant protein</td>
<td>House-made</td>
<td>N/A</td>
</tr>
<tr>
<td>MAPK inhibitor PD0325901</td>
<td>Stemgent</td>
<td>Cat#04-0006-10</td>
</tr>
<tr>
<td>GSK-3β inhibitor CHIR99021</td>
<td>Stemgent</td>
<td>Cat#04-0004-10</td>
</tr>
<tr>
<td><strong>Critical Commercial Assays</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zymoclean Gel DNA Recovery Kit</td>
<td>Zymo Research</td>
<td>Cat#D4002</td>
</tr>
<tr>
<td>X-tremeGENE 9 DNA Transfection Reagent</td>
<td>Sigma Aldrich</td>
<td>Cat#6365809001</td>
</tr>
<tr>
<td>Xfect ESC Transfection Reagent</td>
<td>Clontech</td>
<td>Cat#631320</td>
</tr>
<tr>
<td>AlamarBlue Cell Viability Reagent</td>
<td>Bio-Rad</td>
<td>BUF012A</td>
</tr>
<tr>
<td>NEBNext Ultra DNA Library Prep Kit for Illumina</td>
<td>NEB</td>
<td>Cat#E7370S</td>
</tr>
<tr>
<td>NEBNext Multiplex Oligos for Illumina®</td>
<td>NEB</td>
<td>Cat#E7335S</td>
</tr>
<tr>
<td>KAPA mRNA HyperPrep Kit</td>
<td>Roche</td>
<td>Cat# 08098115702</td>
</tr>
<tr>
<td>TrueSeq Stranded PolyA prep</td>
<td>Illumina</td>
<td>Cat# 20020595</td>
</tr>
<tr>
<td>Accel-NGS 2S PCR-Free Library Kit (96 rxns)</td>
<td>Swift Biosciences</td>
<td>Cat#20096</td>
</tr>
<tr>
<td>Qtracker 705 Cell Labeling Kit</td>
<td>Thermo Fisher</td>
<td>Cat# Q25061MP</td>
</tr>
<tr>
<td>TaqMan Assay, primer information see Table S3</td>
<td>Sigma</td>
<td>N/A</td>
</tr>
<tr>
<td>Direct-zol RNA Miniprep</td>
<td>Zymo Research</td>
<td>Cat#R2050</td>
</tr>
<tr>
<td>SuperScript III First-Strand Synthesis SuperMix</td>
<td>Life Technologies</td>
<td>Cat#18080400</td>
</tr>
<tr>
<td>Fast SYBR Green Master Mix</td>
<td>Life Technologies</td>
<td>Cat#4385618</td>
</tr>
<tr>
<td>ProlongGold</td>
<td>Life Technology</td>
<td>Cat#P36930</td>
</tr>
<tr>
<td>Prime-It II Random Primer Labeling Kit</td>
<td>Agilent Technologies</td>
<td>Cat#300385</td>
</tr>
<tr>
<td><strong>Deposited Data</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Raw and processed data</td>
<td>This paper</td>
<td>GEO: GSE132416 (subseries: GEO: GSE132376, GEO: GSE132404, and GEO: GSE132414)</td>
</tr>
<tr>
<td>Flowcytometry data, Sanger sequencing trace files, raw images, gels, algorithm scripts</td>
<td>This paper</td>
<td><a href="https://doi.org/10.17632/6vbc6htfnf.1">https://doi.org/10.17632/6vbc6htfnf.1</a></td>
</tr>
<tr>
<td><strong>Experimental Models: Cell Lines</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mouse: SOX2-SE-TG ESCs</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Mouse: MIR290-SE-TG ESCs</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Mouse: Sox2-SE-NanogRGM ESCs</td>
<td>Steizer et al., 2015</td>
<td>N/A</td>
</tr>
<tr>
<td>Mouse: miR290-SE-NanogRGM ESCs</td>
<td>Steizer et al., 2015</td>
<td>N/A</td>
</tr>
<tr>
<td>Mouse: SOX2-SE-TG Δpeak 1-CAST ESCs</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Mouse: SOX2-SE-TG Δpeak 2-CAST ESCs</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Mouse: SOX2-SE-TG Δpeak 2-129 ESCs</td>
<td>This paper</td>
<td>N/A</td>
</tr>
<tr>
<td>Tet1-/- #19</td>
<td>Dawlaty et al., 2011</td>
<td>N/A</td>
</tr>
<tr>
<td>Tet1-/- #34</td>
<td>Dawlaty et al., 2011</td>
<td>N/A</td>
</tr>
<tr>
<td>Tet2-/- KO</td>
<td>Dawlaty et al., 2013</td>
<td>N/A</td>
</tr>
<tr>
<td>Tet1-/- Tet2-/- #26</td>
<td>Dawlaty et al., 2013</td>
<td>N/A</td>
</tr>
</tbody>
</table>

(Continued on next page)
### Table of REAGENT or RESOURCE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1-/- Tet2-/- #51</td>
<td>Dawlaty et al., 2013</td>
<td>N/A</td>
</tr>
<tr>
<td>Tet1-/- Tet2-/- Tet3 -/- #26</td>
<td>Dawlaty et al., 2014</td>
<td>N/A</td>
</tr>
<tr>
<td>Tet1-/- Tet2-/- Tet3 -/- #29</td>
<td>Dawlaty et al., 2014</td>
<td>N/A</td>
</tr>
</tbody>
</table>

### Experimental Models: Organisms/Strains

- **Mouse: SOX2-SE-TT:** This paper N/A
- **Mouse: SOX2-SE-GG:** This paper N/A
- **Mouse: MIR290-SE-TT:** CAST/EiJ /129/ BDF1/C57BL/6 This paper N/A
- **Mouse: MIR290-SE-GG:** CAST/EiJ /129/ BDF1/C57BL/6 This paper N/A
- **Mouse: NSG NOD.Cg-PrkdcscidIl2rgtm1Wjl/SzJ:** Jackson Laboratory 005557
- **Mouse: CD1®IGS:** Charles River 022

### Oligonucleotides

- **sgRNA for targeting and knockout, see Table S1** This paper N/A
- **Primers for bisulfite PCR and pyro-sequencing, see Table S2** This paper N/A
- **Primers for TaqMan Assay and qRT-PCR, see Table S3** This paper N/A
- **Primers for ChIP-qPCR, see Table S4** This paper N/A
- **Primers for reporter ESC line, KO cell line, and mouse genotyping, see Table S5** This paper N/A

- **Stellaris® DesignReady FISH Probes, Sox2** LGC Biosearch Technologies Cat# VSMF-3075-5-BS
- **DNA FISH probe for miR290 SE** Sabari et al., 2018 N/A

### Recombinant DNA

- **miR290-SE-RGM-tdTomato targeting vector** Stelzer et al., 2015 N/A
- **miR290-SE-RGM-eGFP targeting vector** This paper N/A
- **Sox2-SE-RGM-tdTomato targeting vector** Stelzer et al., 2015 N/A
- **Sox2-SE-RGM-eGFP targeting vector** This paper N/A
- **pTurbo-Cre** GenBank AF334827
- **px330-BFP-sgRNA** Stelzer et al., 2015 N/A

### Software and Algorithms

- **MACS2 (ChIP-seq algorithms)** Zhang et al., 2008 [https://github.com/taoliu/MACS/wiki](https://github.com/taoliu/MACS/wiki)
- **Samtools** Li et al., 2009 [http://samtools.sourceforge.net](http://samtools.sourceforge.net)
- **STAR (v.2.5.3.a)** Dobin et al., 2013 [https://github.com/alexdobin/STAR](https://github.com/alexdobin/STAR)
- **SNPsplit (v0.3.2)** Babraham Bioinformatics [https://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/](https://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/)
- **RSEM (v1.2.31)** N/A [https://deweylab.github.io/RSEM/](https://deweylab.github.io/RSEM/)
- **PANTHER** Mi et al., 2005; Thomas et al., 2003 [http://pantherdb.org/](http://pantherdb.org/)
- **Image J** Schneider et al., 2012 [https://imagej.net](https://imagej.net)
- **FlowJo** N/A [https://www.flowjo.com](https://www.flowjo.com)
- **Python scripts** This paper; Mendeley Data [https://doi.org/10.17632/6vbc6htfnf.1](https://doi.org/10.17632/6vbc6htfnf.1)
- **MATLAB scripts** Raj et al., 2008 N/A
- **IGV** Broad Institute [https://software.broadinstitute.org/software/ igv](https://software.broadinstitute.org/software/ igv)
- **Bismark v0.21.0** Babraham Bioinformatics [https://www.bioinformatics.babraham.ac.uk/projects/bismark](https://www.bioinformatics.babraham.ac.uk/projects/bismark)
LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Rudolf Jaenisch (jaenisch@wi.mit.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

ESC cell lines
ESC cell culture and proliferation assays
All cells were cultured at 37°C with 5% CO₂. 129xCAST or v6.5 mouse male ESCs were cultured on irradiated mouse embryonic fibroblasts (MEFs) with standard ESCs medium: (500 ml) DMEM supplemented with 10% FBS (HyClone), 10 mg recombinant leukemia inhibitory factor (LIF), 0.1 mM beta-mercaptoethanol (Sigma-Aldrich), penicillin/streptomycin, 1 mM L-glutamine, and 1% nonessential amino acids (all from Invitrogen). For experiments in 2i culture conditions, ESCs were cultured on gelatin-coated plates with N2B27 + 2i + LIF medium containing: (500 ml), 240 ml DMEM/F12 (Invitrogen; 11320), 240 ml Neurobasal media (Invitrogen; 21103), 5 ml N2 supplement (Invitrogen; 1750248), 10 ml B27 supplement (Invitrogen; 17504044), 10 mg recombinant LIF, 0.1 mM beta-mercaptoethanol (Sigma-Aldrich), penicillin/streptomycin, 1 mM L-glutamine, and 1% nonessential amino acids (all from Invitrogen), 50 mg/ml BSA (Sigma), PD0325901 (Stemgent, 1 mM), and CHIR99021 (Stemgent, 3 mM). For measuring cell proliferation, AlamarBlue Cell Viability Reagent (Bio-Rad, BUF012A) was added to cell culture and incubated at 37°C with 5% CO₂ and emission at 590nm was monitored every 50hrs. At each sampling time point, relative changes in cell numbers were compared to 0hr after sorting.

Generating biallelically targeted reporter cell lines
To generate SOX2-SE-TG and MIR290-SE-TG reporter cell lines, targeting vectors (Mir290-SE-RGM-tdTTomato, Mir290-SE-RGM-eGFP, Sox2-SE-RGM-tdTTomato, Sox2-SE-RGM-eGFP), and CRISPR/Cas9 were transfected into ESCs using Xfect ESC Transfection Reagent (Clontech, Cat#631320), according to the provider’s protocol. Forty-eight hours following transfection, cells were selected for puromycin resistance (Sigma Aldrich, Cat#P7255) and plated on MEF feeder plates. Single colonies were further analyzed for proper and single integration by Southern blot and Junction PCR analysis. PGK-Puromycin resistance cassette were looped out by overexpression of Cre recombinase (pTurbo-Cre, GenBank accession number AF334827) and followed by Southern blot validation.

ESCs with CRISPR-Cas9-mediated deletions
Tet-enzyme single-, double- and triple knockouts were generated and described previously (Dawlaty et al., 2011, 2013, 2014). sgRNA sequences are cloned into px330-BFP vector under U6 promoter. px330-BFP-sgRNA vectors were transfected into pre-plated ESC cells using Xfect ESC Transfection Reagent, according to the provider’s protocol. For analysis in populations, cells were sorted for BFP 48 hours post-transfection and cultured on MEF feeder plates. For single clone analysis, cells were genotyped using Southern blot or TA cloning of PCR products of CRISPR targeting site from each allele followed by sequencing. For Dnmt3a and Dnmt3b, Aid and Tdg single knockouts, single clones with frame-shifting indels were selected for further analysis; for TF binding site deletions, single clones have allele-specific entire peak site deletions were selected for further analysis. For TF enrichment site deletion experiments, sgRNA pairs for generating deletion are transfected as following: Δpeak 1-CAST: sgTFBS-Sox2-SE-1(CAST) and sgTFBS-Sox2-SE-2(CAST); Δpeak 2-CAST: sgTFBS-Sox2-SE-2(CAST) and sgTFBS-Sox2-SE-3(CAST); Δpeak 2-129: sgTFBS-Sox2-SE-2(CAST) and sgTFBS-Sox2-SE-3(CAST). All sgRNA sequences are listed in Table S1.

Animals
Blastocyst injections and generation of reporter mice
Blastocyst injections were performed using (C57BL/6xDBA) B6D2F1 (Charles River) or CD1 (Charles River) host embryos. In brief, 6-7-week old B6D2F1 females were hormone primed by an intraperitoneal (i.p.) injection of pregnant mare serum gonadotropin (PMS, EMD Millipore) followed 46 hr later by an injection of human chorionic gonadotropin (hCG, VWR). Embryos were harvested at the morula stage and cultured in a CO₂ incubator overnight. To obtain tetraploid (4n) blastocysts, electrofusion was performed at approximately 44–47 h post hCG using a BEX LF-301 cell fusion device (Protech International Inc., Boerne, TX). On the day of the injection, groups of embryos were placed in drops of M2 medium using a 16-μm diameter injection pipet (CytoSpring). Approximately ten cells were injected into the blastocoel cavity of each embryo using a Piezo micromanipulator (Prime Tech). Approximately 20 blastocysts were subsequently transferred to each recipient female; the day of injection was considered as 2.5 days postcoitum (MPC). Male chimera mice were mated to C57BL/6 females and the ones that gave birth to agouti pups (F1) have germ-line transmitted CASTX129 ESC. Mice were handled in accordance with institutional guidelines and approved by the Committee on Animal Care (CAC) and Department of Comparative Medicine (DCM) of Massachusetts Institute of Technology.

Mouse mating scheme and genotyping
All mouse F1 mice heterozygous for either the SE-RGM-tdTTomato (abbreviated as SOX2-SE-T0 or MIR290-SE-T0) or the SE-RGM-eGFP allele (abbreviated as SOX2-SE-G0 or MIR290-SE-G0) were obtained by mating germ-line transmitted chimeras to C57BL/6 females. F2 mice homozygous for either SE-RGM-tdTTomato (abbreviated as SOX2-SE-TT or MIR290-SE-TT) or the SE-RGM-eGFP
allele (abbreviated as SOX2-SE-GG or MIR290-SE-GG) were generated by inbreeding (SOX2-SE-T0 x SOX2-SE-T0, SOX2-SE-G0 x SOX2-SE-G0, MIR290-SE-T0 x MIR290-SE-T0, MIR290-SE-G0 x MIR290-SE-G0). Mice are genotyped by PCR the 5’junction of the SE RGM: SOX2-SE-F (or MIR290-SE-F) with tdTomato-R for the RGM-tdTomato allele, SOX2-SE-F (or SOX2-SE-F) with eGFP-R for the RGM-eGFP allele, and SOX2-SE-F (or MIR290-SE-F) with SOX2-SE-R (or MIR290-SE-R) for the wild-type allele (Table S5).

**Confocal imaging of live pre-implantation embryos**

2-cell embryos were obtained from mating SOX2-SE-TT or MIR290-SE-TT females hormone primed step-wise with PMS and hCG to SOX2-SE-GG or MIR290-SE-GG males, respectively, or the opposite mating strategy (SOX2-SE-GG or MIR290-SE-GG females to SOX2-SE-TT or MIR290-SE-TT males, respectively). 2-cell embryos were flushed out from the oval-duct by M2 media with BSA (CytoSpring # m2113) 48hrs post mating. The embryos were then cultured in 25-50ul KSOM media droplets (CytoSpring # KO102) covered by mineral oil in a 37°C/5% CO2 incubator. Embryos will become blastocysts at E3.5. For monitoring methylation dynamics in vivo, ESCs were cultured in serum + LIF, pre-plated and sorted based on RGM activity before injection. 2-3 cells were injected into 8-cell stage CD1 host embryos and cultured in M2 media with BSA at 37°C in 5% CO2. Images were taken by a Zeiss LSM 710 Laser Scanning Confocal microscope. Images were taken using either 10x or 40x water lenses and saved in LSM format. Channels for eGFP (excitation 488nm), tdTomato (excitation 594nm), Cy5 (excitation 633nm), and Hoechst 33342 (excitation 405nm) were merged into image composites.

**Teratoma formation assays and H&E staining**

0.5-1 million sorted ESCs in serum + LIF media were 1:1 mixed with Matrigel and injected subcutaneously into the femur on both sides of the NSG mice. Tumors were taken when reaching 1cm in diameter and mice euthanized. Mice were handled in accordance with institutional guidelines and approved by the Committee on Animal Care (CAC) and Department of Comparative Medicine (DCM) of Massachusetts Institute of Technology. Tissues were dissected and fixed in 10% formalin overnight. Tissues were embedded in paraffin, sectioned, and stained for H&E.

**METHOD DETAILS**

**Southern blots**

Genomic DNA (10–15 mg) was digested with appropriate restriction enzymes overnight. Subsequently, genomic DNA was separated on a 0.8% agarose gel, transferred to a nylon membrane (Amersham) and hybridized with 32P probe labeled by Prime-It II Random Primer Labeling Kit (Agilent Technologies, Cat#300385).

**Flow cytometry**

To assess the proportion of eGFP and tdTomato in the established reporter cell lines, a single-cell suspension was filtered and assessed on the BD Aria or FACSCanto II. Compensation was achieved by using cells with either tdTomato or eGFP fluorescence. Fsc files were analyzed by FlowJo.

**Bisulfite conversion-PCR (BS-PCR) and pyro-sequencing**

Bisulfite conversion of genomic DNA, nested PCR, and sequencing was established as described previously (Stelzer et al., 2015). Pyro-seq of all bisulfite converted genomic DNA samples were performed with PyroMark Q48 Autoprep (QIAGEN) according to the manufacturer’s instructions. Primers used for BS-PCR and pyro-sequencing are listed in Table S2.

**Retinoic acid differentiation**

ESCs carrying the reporter for both Mir290 and Sox2 SE regions were sorted for Nanog-eGFP positive and RGM-tdTomato positive and plated on gelatin-coated plates in ESC medium (+LIF). The next day, cells were washed with PBS, re-suspended in basal N2B27 medium (2i medium without LIF, insulin, and the two inhibitors), and supplemented with 0.25uM retinoic acid (RA, Sigma Aldrich, Cat#R2625-50MG). Medium was replaced every other day.

**Double thymidine block**

10-20k cells/per well were plated onto 12-well plates after sorting with media containing 2.5mM thymidine for 12hrs. Blocking was released by washing twice with PBS and culturing in serum + LIF mouse ES media for 9hrs. Cells were then again blocked with 2.5mM thymidine for 14hrs and FACS analyses were done 6hrs post release.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

**qRT-PCR and TaqMan assays**

Total mRNA was extracted from ESCs using Direct-zol RNA Miniprep (Zymo Research, Cat#R2050) after pre-plating for elimination of MEF feeders, treated with DNase A defined amount of mRNA reverse-transcribed into cDNA using SuperScript III First-Strand Synthesis SuperMix (Life Technologies, Cat#18080400) using random hexamers. Total expression of transcripts were quantified by qRT-PCR using Fast SYBR Green Master Mix (Life Technologies), and allele-specific transcripts are quantified by TaqMan Assay.
customized probes (Sigma, Table S3) targeting Sox2 and Mir290-295 pri-mRNA SNPs. Tukey’s multiple comparison (**P<0.001, *P<0.01, *P<0.05). Both qRT-PCR and TaqMan assays used at least 2 independently targeted clones as biological replica. The probes and context sequences are listed in Table S3.

ChIP-qPCR
ChIP was done on 2-5 million cells of each same-culture-sorted population from both reporter cell lines as described previously (Leach et al., 2013). 2ug of anti-H3K27ac antibody (abcam ab4729) was used for precipitation. Eluted DNA was quantified using real-time qPCR with Fast SYBR Green Master Mix. Each ChIP-qPCR was repeated 3 times. Enrichment was calculated using as percentage of input. Statistical differences between samples are calculated with two-way ANOVA (alpha = 0.05), followed by Tukey’s multiple comparison (**P<0.0001, *P<0.01). Primers used for detecting positive and negative control sequences, and SE targets are listed in Table S4.

H3K27ac ChIP-Seq and analysis
ChIP samples of 4 same-culture-sorted populations from SOX2-SE-TG and MIR290-SE-TG, respectively, were validated for positive and negative targets using qPCR. Libraries of Input-Chip pairs were prepared with Accel-NGS 2S PCR-Free Library Kit (Cat#20096) and sequenced using Illumina HiSeq 2500. Raw reads were aligned to the reference genome mm10 using BWA using default parameters. Peak calling was done using MACS2. Peak intensities at SE, promoter and in-between regions are quantified and compared using bamCompare – deepTools 3.0.2 with FPKM from 10bp genomic bins of each sample. SNPs specific to 129 or CAST genomes at SE and promoters were counted from mapped raw reads and SNPs covered by more than 3 reads are accepted for quantification. Coordinates for analysis (mm10): Mir290-SE: chr7:3198900-3202780, Mir290-promoter: chr7:3215340-3221110; Sox2-SE: chr3:34752523-34766449, Sox2-promoter chr3:34649995-34652460.

RNA-Seq and analysis
For each reporter cell line, 2 independently targeted clones are independently sorted twice, generating 2 biological replica x 2 experimental replica = 4 replica in total. Stranded mRNA libraries were prepared using KAPA HyperPrep (SOX2-SE-TG) and TrueSeq Stranded PolyA prep (MIR290-SE-TG). mRNA libraries were sequenced on Illumina HiSeq 2500. Allele-specific RNA expression was quantified with a custom pipeline. In short, raw fastq files are aligned to a consensus genome using STAR (v.2.5.3.a). The reference transcriptome includes the Mir290-295 pri-mRNA or Sox2 and the RGMs on pseudo-chromosomes. After alignment SNPsplit (v0.3.2) splits the reads into four files based on single nucleotide variations (SNV). The reads were either allele specific (for CAST or S129), unassigned (if there are no SNVs present) or conflicting (if the SNVs in the read are from both alleles). The split read files were quantified using RSEM (v1.2.31) separately for each sample. Raw counts were then normalized to library using DESeq2 (v1.18.1) for each split. To obtain sample-level quantifications raw counts were summed over the splits before normalization. Differential expression analysis (DEA) was performed using DESeq2 (v1.18.1) at the level of samples. Samples were corrected for genetic clone and batch effect. GO analyses were performed using PANTHER. All expressed genes in the respective cell lines were used as the reference backgrounds. All P-values were controlled for false discovery rate (Benjamini-Hochberg procedure).

DNA smFISH and image analyses
Cells were fixed for 15 min with 4% PFA at room temperature and subsequently permeabilized in 70% EtOH overnight. Custom designed smFISH probes for Sox2 labeled with Quasar 750 (Stellaris® DesignReady FISH Probes, Cat# VSMF-3075-5-BS) were incubated with the samples at 16 hours at 30°C in hybridization buffer (100 mg/mL dextran sulfate, 25% formamidine, 2X SSC, 1 mg/mL E.coli tRNA, 1 mM vanadyl ribonucleoside complex, 0.25 mg/mL BSA). Samples were washed twice for 30 min at 30°C with wash buffer (25% formamide, 2X SSC) containing DAPI (1 μg/mL, Sigma D9542). All solutions were prepared with RNase-free water. Finally, the sections were mounted using ProlongGold (Life Technologies, P36930) and imaged two days later. Mounted samples were imaged on a Nikon Ti-Eclipse epifluorescence microscope equipped with an Andor iXON Ultra 888 EMCCD camera, using a 100X/1.45 Plan Apo Lambda oil objective (Nikon) and dedicated, custom-made fluorescence filter sets (Nikon). z-stacks with a distance of 0.3 μm between planes were collected. The number of Sox2 (mRNA) signals per cell was quantified using home-made MATLAB scripts.

DNA FISH, Med1 IF and average image analyses
DNA FISH of the Mir290 SE and IF of MED1 were done as previously described (Sabari et al., 2018). For analysis of RNA/DNA FISH with immunofluorescence, custom Python scripts were written to process and analyze 3D image data gathered in FISH and IF channels. Nuclear stains were blurred with a Gaussian filter (sigma = 2.0), maximally projected in the z plane, and clustered into 2 clusters (nuclei and background) by K-means. FISH foci were either manually called with ImageJ or automatically called using the scypl ndimage package. For automatic detection, an intensity threshold (mean + 3*standard deviation) was applied to the FISH channel. The ndimage find_objects function was then used to call contiguous FISH foci in 3D. These FISH foci were then filtered by various criteria, including size (minimum 100 voxels), circularity of a max z-projection (circularity = 4*areaperimeter2; 0.7), and being present in a nucleus (determined by nuclear mask described above). For manual calling, FISH foci were identified in maximum z-projections of the FISH channel, and the x and y coordinates were used as reference points to guide the automatic detection described above.
The FISH foci were then centered in a 3D-box (length size $i = 3.0 \, \mu m$). The IF signal centered at FISH foci for each FISH and IF pair are then combined and an average intensity projection is calculated, providing averaged data for IF signal intensity within a $i \times i$ square centered at FISH foci. As a control, this same process was carried out for IF signal centered at an equal number of randomly selected nuclear positions. These average intensity projections were then used to generate 2D contour maps of the signal intensity. Contour plots are generated using the matplotlib python package. For the contour plots, the intensity-color ranges presented were customized across a linear range of colors ($n! = 15$). For the FISH channel, black to magenta was used. For the IF channel, we used chroma.js (an online color generator) to generate colors across 15 bins, with the key transition colors chosen as black, blueviolet, medium-blue, lime. This was done to ensure that the reader’s eye could more readily detect the contrast in signal. The generated colormap was employed to 15 evenly spaced intensity bins for all IF plots. The averaged IF centered at FISH or at randomly selected nuclear locations are plotted using the same color scale, set to include the minimum and maximum signal from each plot.

**High-throughput sequencing of bisulfite PCR**

PCR amplicons were sonicated using Covarius into 150-200bp range. NEBNext® Ultra™ DNA Library Prep Kit for Illumina and NEBNext® Multiplex Oligos for Illumina® were used to construct libraries according to manufacturer’s protocol. Single barcoded library was prepared from sonicated bisulfite PCR amplicon fragments of the *Mir290* SE wildtype-allele using NEBNext® Ultra™ DNA Library Prep Kit for Illumina (NEB #E7370S) and NEBNext® Multiplex Oligos for Illumina® (Index Primers Set 1, NEB #E7335S). Libraries were sequenced with 40bp single reads, adapter trimmed, aligned and analyzed with Bismark v0.21.0 (bismark -nondirectional). CpGs with >1000 coverage were counted to generate average percentage of methylation. Methylation percentage and its standard error were estimated as described in (Smallwood et al., 2014), and number of methylated counts was assumed to be a binomial random variable.

**DATA AND CODE AVAILABILITY**

Description: [https://doi.org/10.17632/6vbc6htfnf.1](https://doi.org/10.17632/6vbc6htfnf.1)

The raw confocal, gel and film images and original fsc files have been deposited at Mendeley Data

Description: Raw and processed high-throughput sequencing data have been deposited at NCBI Gene Expression Omnibus under ID code GEO: GSE132416 (subseries: GEO: GSE132376 for H3K27ac ChIP-seq, GEO: GSE132404 for BS-seq, and GEO: GSE132414 for RNA-seq).