Gene regulation in embryonic development

Berg, P.R. van den

1 Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs

This chapter is based on:
Yuelin Song, Patrick R van den Berg, Styliani Markoulaki, Frank Soldner, Alessandra Dall'Agnese, Jonathan E Henninger, Jesse Drotar, Nicholas Rosenau, Malkiel A Cohen, Richard A Young, Stefan Semrau, Yonatan Stelzer, Rudolf Jaenisch. "Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs". In: *Molecular cell* 0.0 (2019), 905–920.e6. DOI: 10.1016/j.molcel.2019.06.045

Chapter preface

The following chapter is a near verbatim reproduction of Song et al. [1] (supplementary figures, however, are not reproduced and for these I refer to in the original article). This chapter demonstrates that DNA methylation is a highly dynamic process in steady-state mouse embryonic stem cells (ESCs). Furthermore, is it shown that DNA methylation of super-enhancer loci correlates strongly with gene expression, which makes the observed fluctuations highly relevant for ESC biology. I contributed significantly to making the connection between DNA methylation and gene expression.

To measure DNA methylation of specific loci in live cells, fluorescent reporter lines were created. By using F1 hybrid cell lines, it was possible to create independent reporters (with distinguishable fluorophores) for the paternal and maternal allele, respectively. After careful characterization of the methylation dynamics, we set out to study the effect on gene expression. To that end, cell populations sorted on the reporter signals were profiled by RNA-seq. My task was to analyze the RNA-seq data. The specific challenge was to design an analysis pipeline that distinguishes reads from the two maternal and paternal allele, which differ in only a few nucleotides. Moreover I assisted by reanalysing single cell WGBS (scWGBS) data and validating the expression levels of *Sox2* in these sorted populations using single-molecule FISH (smFISH).

Patrick van den Berg

Abstract

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 methyltransferase 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 preimplantation embryos, revealing that methylation heterogeneity occurs *in vivo*.

1.1 Introduction

Tissue-specific differential methylation regions (T-DMRs) have been found to strongly associate with low CpG density and inter-genic enhancers [2, 3, 4, 5, 6], and the vast majority of cell-type specific DNA methylation changes occur at distal regulatory elements [7, 8]. whole-genome bisulfite sequencing (WGBS) data indicate a low but detectable level of DNA methylation at T-DMRs overlapping active enhancers [9, 10, 11, 12, 13, 14, 8]. Recent single cell WGBS (scWGBS) data from mouse emrbyonic 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 [15, 16, 17, 18, 19, 20]. 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 [21, 13, 22]. While genome-wide epigenetic profiling provided insights into the relationship between DNA methylation, histone marks, and TFs and coactivators binding [13, 23, 24], 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 [25, 26]. 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 utilized this system at two pluripotency SEs, *Sox2* and *Mir290* SE, in ESCs. Both SEs overlap with ESC-specific DMRs, which display consistently low levels of methylation, indicating potential heterogeneity [27, 28, 19, 29, 8]. We targeted RGMs to both alleles of the two SEs in F1 129xCasteneous (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.

1.2 Results

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

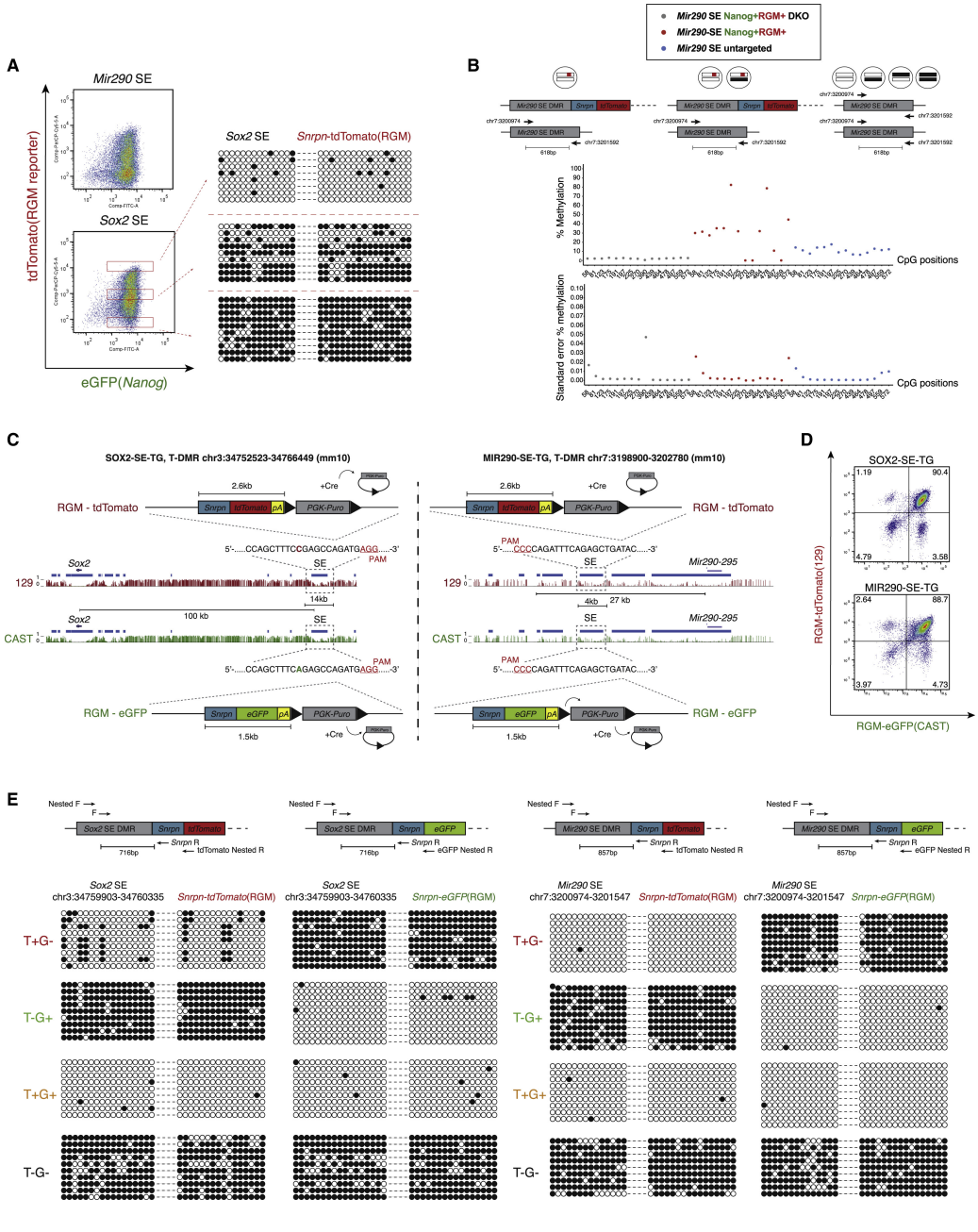
Sox2 and *Mir290* SEs reside on chromosome 3 and 7, respectively. Both SEs overlap with T-DMRs, which are hypo-methylated in ESCs but become *de novo* methylated upon differentiation [25]. The T-DMR of the *Sox2* SE is located about 100 kb upstream of the *Sox2* gene, whereas the *Mir290* SE, consisting of hypo-methylated DMR constituents interspersed by small hyper-methylated regions, is proximal to the *Mir290-295* cluster (Song et al. [1] 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 (Song et al. [1] Figure S1B) [27, 28, 29, 8]. 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 [20] 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 (Song et al. [1] 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 [16, 17, 18, 30, 31, 20], 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 [25]. 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 FACS (Figure 1A). Sorting cells based on florescence intensity, followed by bisulfate 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 hyper-methylation occurred on both the targeted and the untargeted alleles in the pluripotent ESC population ($Nanog^+$), indicating that rare allelic methylation exists among cells (Song et al. [1] 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 Song et al. [1] 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 129xCastaneus F1 hybrid ESCs with allele-specific RGM reporters and generated two cell lines, *Sox2*-$129^{\text{SE-RGM-tdTomato}}$/*Sox2*-$\text{CAST}^{\text{SE-RGM-eGFP}}$ (abbreviated below as SOX2-SE-TG) and *Mir290*-$129^{\text{SE-RGM-tdTomato}}$/*Mir290*-$\text{CAST}^{\text{SE-RGM-eGFP}}$ (abbreviated below as MIR290-SE-TG) (Figure 1C and Song et al. [1] Figure 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 single nucleotide polymorphisms (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

Figure 1 *(following page)*. 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 Song et al. [1] Figure S1.


A
Mir290 SE
Sox2 SE
tdTomato(RGM reporter)
eGFP(Nanog)
Sox2 SE
Snrpn-tdTomato(RGM)
B
Mir290 SE Nanog+RGM+ DKO
Mir290-SE Nanog+RGM+
Mir290 SE untargeted
% Methylation
CpG positions
Standard error % methylation
CpG positions
C
SOX2-SE-TG, T-DMR chr3:34752523-34766449 (mm10)
MIR290-SE-TG, T-DMR chr7:3198900-3202780 (mm10)
RGM - tdTomato
2.6kb
+Cre
Snrpn
tdTomato
pA
PGK-Puro
5'-.....CCAGCTTTCCGAGCCAGATGAGG.....-3'
PAM
Sox2
SE
129
CAST
14kb
100 kb
Mir290-295
4kb
27 kb
RGM - eGFP
eGFP
1.5kb
D
SOX2-SE-TG
MIR290-SE-TG
RGM-tdTomato(129)
RGM-eGFP(CAST)
1.19
90.4
4.79
3.58
2.64
88.7
3.97
4.73
E
Nested F
F
Sox2 SE DMR
Snrpn
tdTomato
eGFP
Snrpn R
716bp
857bp
tdTomato Nested R
eGFP Nested R
Sox2 SE
chr3:34759903-34760335
Snrpn-tdTomato(RGM)
Snrpn-eGFP(RGM)
Mir290 SE
chr7:3200974-3201547
T+G-
T-G+
T+G+
T-G-

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 (Song et al. [1] Figure S1C) and in our BS-PCR analysis on both targeted and wild-type alleles (Figure 1A, 1B, and Song et al. [1] Figure 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^- populations represent two extreme methylation states of the intrinsic epigenetic heterogeneity at both SEs (Song et al. [1] 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 when exiting pluripotency (Song et al. [1] Figure S1G).

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

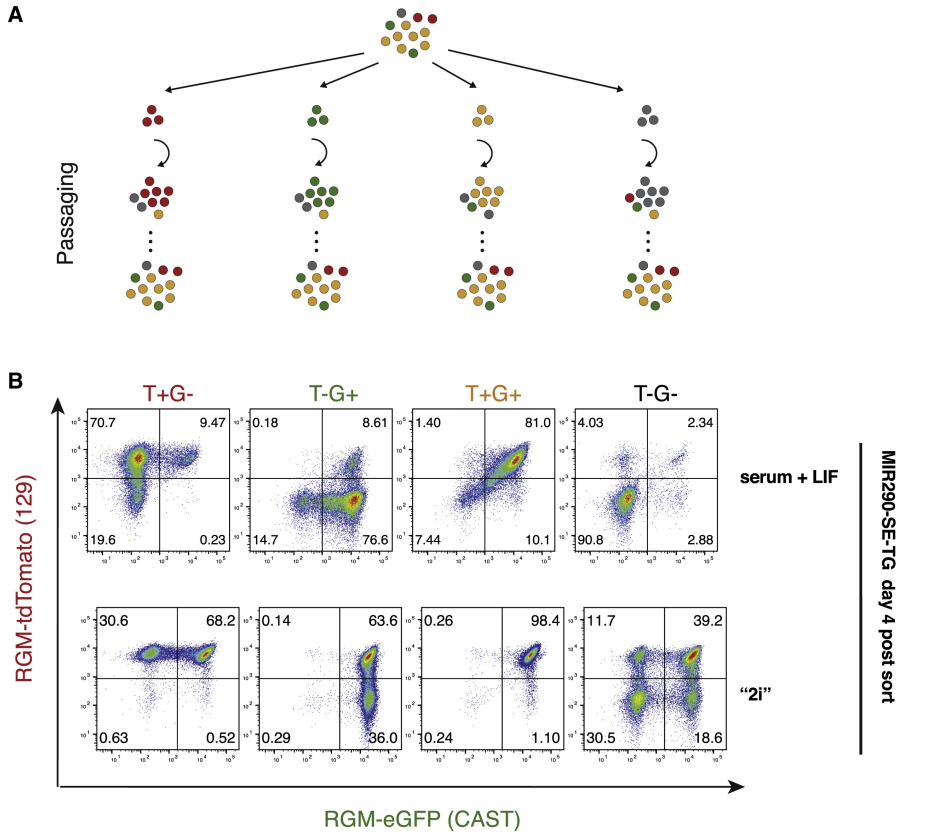
To gain insights into the origin of DNA methylation heterogeneity, we FACS sorted equal numbers of the four populations from both reporter cell lines and monitored the RGM activity upon passaging in serum + LIF medium (Figure 2A). Figure 2B (serum + LIF) and Song et al. [1] Figure S2A show that the SE 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 at the *Mir290* SE (Figure 2B and 2C; Song et al. [1] Figure S2B). Demethylation of T^-G^- population of SOX2-SE-TG in "2i", however, is slower over the long term than that in serum + LIF, possibly due to impaired cell division as shown in the later part of this article. The observed DNA methylation difference between "2i" and serum + LIF is consistent with the extensive global demethylation induced in "2i" by downregulation of *de novo* and maintenance methyltransferases [32, 33, 34, 35, 36].

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 (Song et al. [1] 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 (Figure 3A and Song et al. [1] Figure 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 (Song et al. [1] 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 [37]. 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 (Figure 3B and 3C). This suggests that cell proliferation-driven passive demethylation is responsible for SE demethylation. To confirm this observation genetically, we transfected $129^{\text{SE-RGM-tdTomato}}$ $T^{-}G^{+}$ cells with Cas9 and sgRNAs against genes encoding the maintenance enzymes DNMT1/UHRF1, which upon down-regulation 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 $129^{\text{SE-RGM-tdTomato}}$ allele demethylation (changes of the fraction of $T^{+}G^{+}$ cells) after disruption of these genes. When *Dnmt1* or *Uhrf1* were disrupted, the $129^{\text{SE-RGM-tdTomato}}$ allele became demethylated in a substantial fraction of cells (Figure 3E). In contrast, transduction of sgRNAs against *Tet* enzymes, *Aid*, or *Tdg* had no substantial effect indicating that active demethylation is not significantly involved in SE demethylation. To confirm that the lack of methylation changes upon disruption of *Tets*, *Aid*, or *Tdg* was not due to inefficient Cas9-sgRNA transfection, we further compared the demethylation kinetics of the $129^{\text{SE-RGM-tdTomato}}$ allele in single clones harboring homozygous *Tdg* and *Aid* frameshift mutations (Song et al. [1] Figure S3D) with that of wild-type cells and observed no difference (Song et al. [1] Figure S3E). In addition, DNA methylation levels, as quantified by pyro-sequencing, did not reveal a significant difference among *Tet1*, *2*, and *3* single-knockout, *Tet1*, *2* double-knockout, *Tet1*, *2*, *3* triple-knockout ESCs, and the isogenic wild-type cells [38, 39, 40] (Song et al. [1] Figure S3C). Given the rapid proliferation of ESCs, our data are consistent with the notion that locus-specific DNA methylation at both SEs is subjected to intrinsically dynamic changes at the allelic level in each cell due to un-

Figure 2 *(following page).* SE DNA Methylation Heterogeneity Is Created by Dynamic Switching of Methylation States. (A) Experiment setup for monitoring SE DNA methylation dynamics. Yellow cells: T+G+; gray cells: T-G-; red cells: T+G-; green cells: T-G+. (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 Song et al. [1] Figure S2.

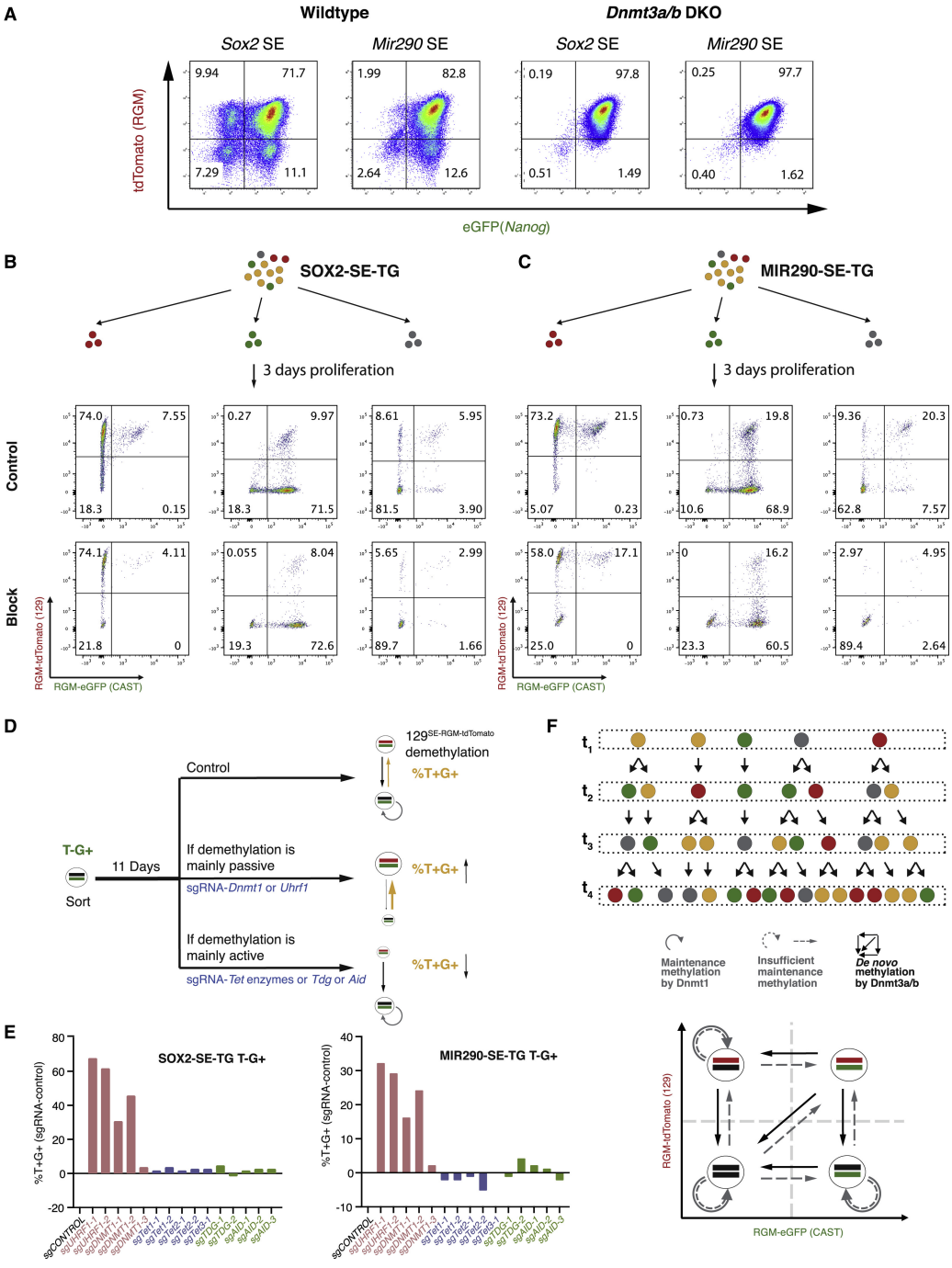

A
Passaging
B
T+G-
T-G+
T+G+
T-G-
70.7
9.47
19.6
0.23
0.18
8.61
14.7
76.6
1.40
81.0
7.44
10.1
4.03
2.34
90.8
2.88
serum + LIF
30.6
68.2
0.63
0.52
0.14
63.6
0.29
36.0
0.26
98.4
0.24
1.10
11.7
39.2
30.5
18.6
"2i"
MIR290-SE-TG day 4 post sort
RGM-tdTomato (129)
RGM-eGFP (CAST)
C
T-G+
T+G-
T-G-
T+G+
Percentage
100
50
0
Day 0
Day 4
Day 6
Day 8
Day 13
Day 11
serum + LIF
"2i"
MIR290-SE-TG

synchronized cell division and passive DNA demethylation, which leads to heterogeneous SE methylation at a snapshot sampling time (t1, …, t4, Figure 3F, top). The steady-state of such dynamic heterogeneity reflects a balance between *de novo* methylation dependent on both DNMT3A and DNMT3B and passive demethylation during rapid cell proliferation (Figure 3F, bottom).

1.2.3 TF Binding at SE Promotes Demethylation and Inhibits *De Novo* Methylation

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 [41, 42, 43, 22]. The *Sox2* SE harbors multiple enrichment sites for the master TFs OCT4 and NANOG in ESCs [44] (Figure 4A, top). We deleted enrichment sites for the two TFs (peak 1 for NANOG and 2 for both NANOG and OCT4) at the *Sox2* SE DMR on either the $129^{\text{SE-RGM-tdTomato}}$ or the $\text{CAST}^{\text{SE-RGM-eGFP}}$ 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; Song et al. [1] Figure S4A). We sorted the T^-G^- and T^+G^+ populations from these clones and monitored the 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 rates or demethylation rates, respectively (Figure 4C). We found that both the $129^{\text{SE-RGM-tdTomato}}$ and the $\text{CAST}^{\text{SE-RGM-eGFP}}$ 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

Figure 3 *(following page)*. The Dynamics of SE DNA Methylation Is Driven by De Novo Methylation and Passive Demethylation during Cell Proliferation. (A) Elimination of the population with methylated SEs in Dnmt3a and Dnmt3b DKO v6.5-*Nanog*-eGFP ESCs with the RGM-tdTomato reporter targeted mono-allelically at either the *Sox2* or the *Mir290* SE. (B and C) Demethylation of sorted T+G-, T-G+, and T-G- cells from (B) SOX2-SE-TG and (C) *MIR290-SE-TG* cells with and without thymidine block. (D) Expected changes in the T+G+ cell percentage for each demethylation mechanism upon CRISPR/Cas9-mediated gene disruptions. Changes in the percentage of T+G+ cells indicate the rate of demethylation on the 129SE-RGM-tdTomato allele. (E) Relative changes in the T+G+ percentage upon transfecting sgRNAs against enzymes involved in DNA demethylation, as compared to cells transfected with the same vector without sgRNA (sgControl). (F) A model for the origin of locus-specific DNA methylation heterogeneity. See also Song et al. [1] Figure S3.

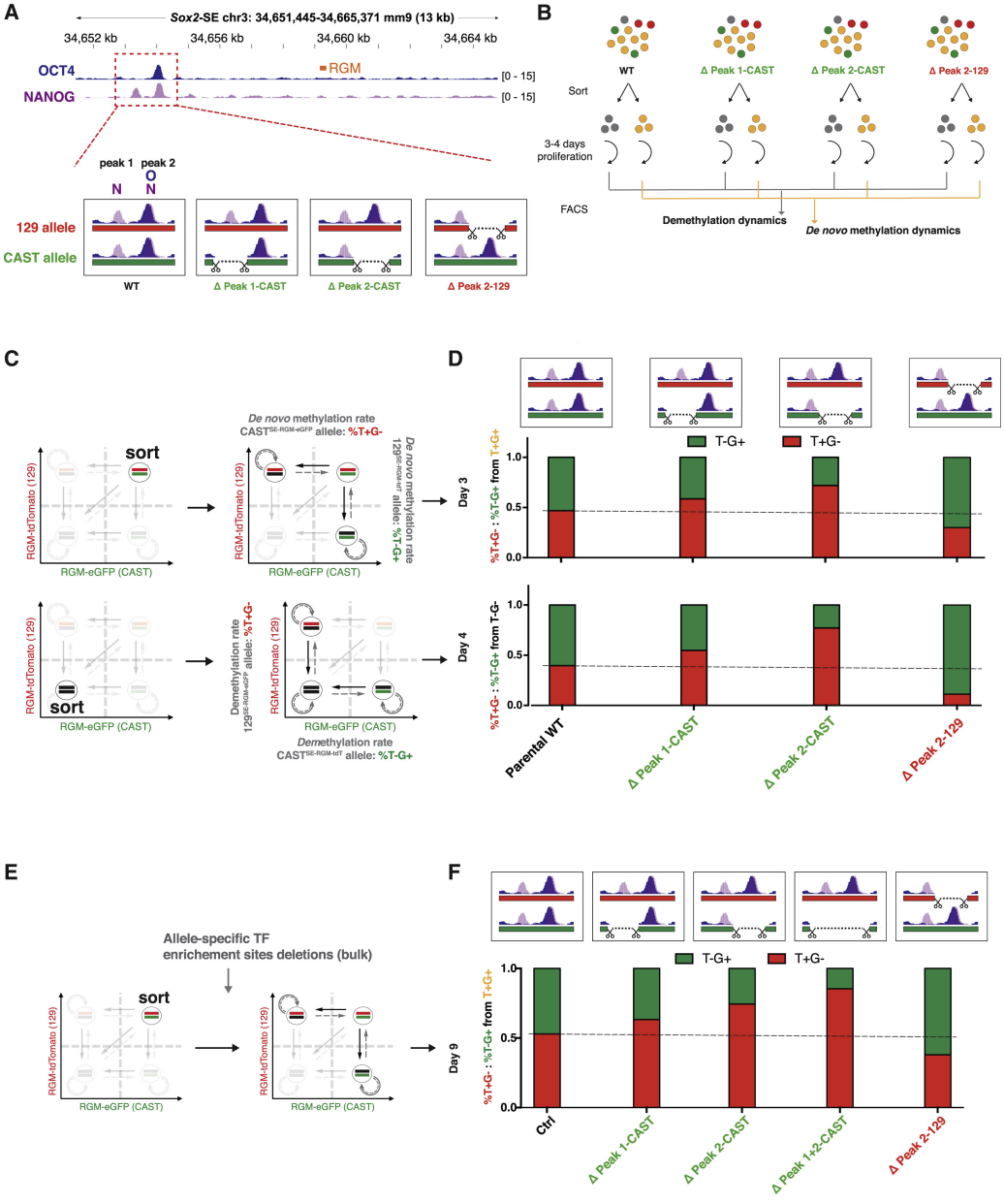

A
Wildtype
Dnmt3a/b DKO
Sox2 SE
Mir290 SE
Sox2 SE
Mir290 SE
9.94
71.7
7.29
11.1
1.99
82.8
2.64
12.6
0.19
97.8
0.51
1.49
0.25
97.7
0.40
1.62
tdTomato (RGM)
eGFP(Nanog)
B
SOX2-SE-TG
3 days proliferation
C
MIR290-SE-TG
3 days proliferation
Control
Block
RGM-tdTomato (129)
RGM-eGFP (CAST)
D
129SE-RGM-tdTomato demethylation
%T+G+
Control
T-G+
11 Days
Sort
If demethylation is mainly passive
sgRNA-Dnmt1 or Uhrf1
If demethylation is mainly active
sgRNA-Tet enzymes or Tdg or Aid
E
SOX2-SE-TG T-G+
MIR290-SE-TG T-G+
%T+G+ (sgRNA-control)
sgCONTROL
F
t1
t2
t3
t4
Maintenance methylation by Dnmt1
Insufficient maintenance methylation
De novo methylation by Dnmt3a/b
RGM-tdTomato (129)
RGM-eGFP (CAST)

BS-PCR followed by Sanger sequencing on sorted cells from ΔPeak 1-CAST and ΔPeak 2-129 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 (Song et al. [1] 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).

1.2.4 DNA Methylation Decreases MED1 Association with SE, 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 [45, 46, 47, 48]; 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 [49]. 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 Song et al. [1] 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

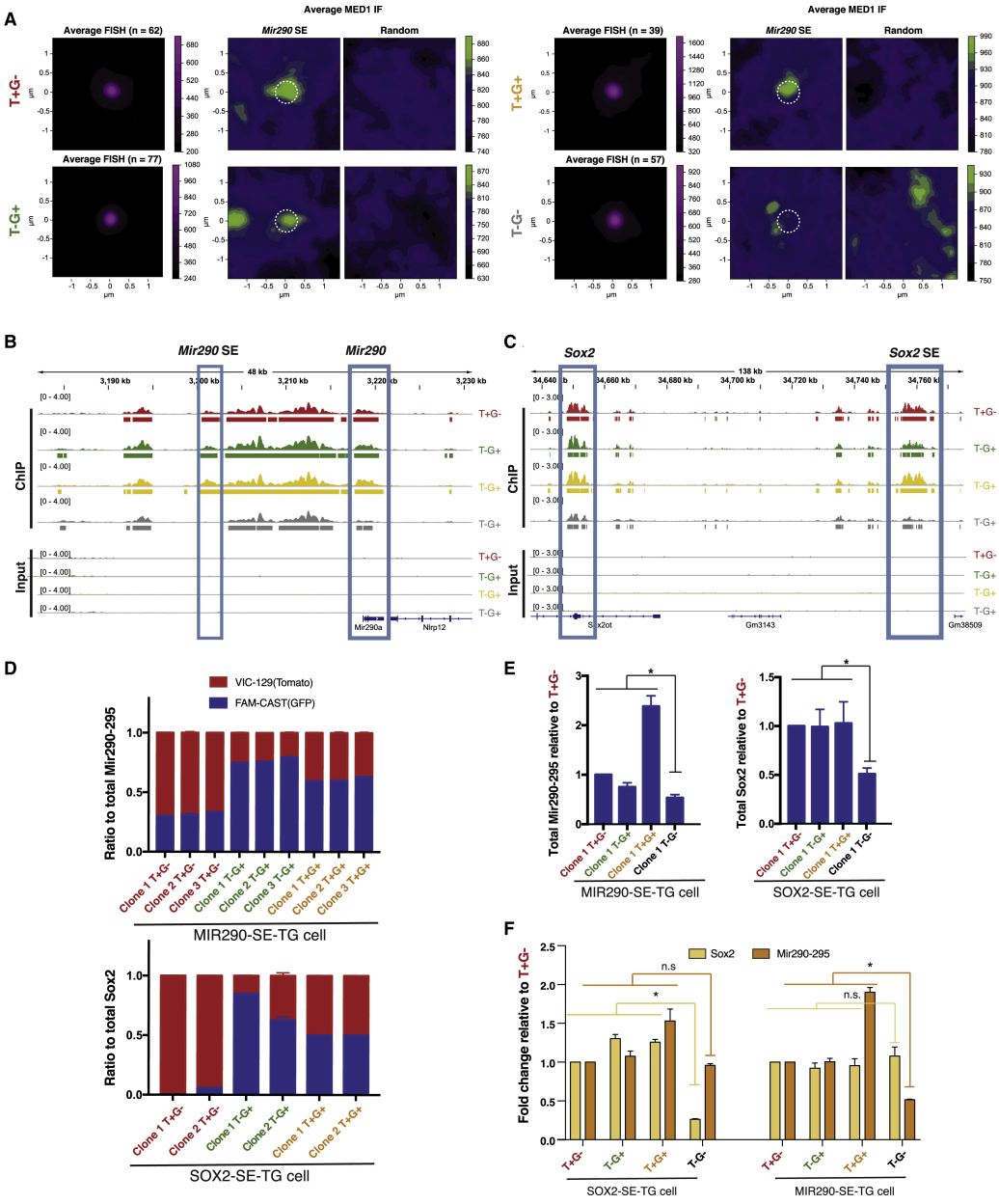
Figure 4 *(following page).* TF Binding at SEs Promotes Demethylation and Inhibits De Novo Methylation. (A) Top, schematic representation of TF enrichment sites (based on the ChIP-seq data of NANOG [pink track, peak 1 and 2] and OCT4 [blue track, peak 2]. ENCODE: ENCSR779CZG and ENCSR392DGA) relative to the RGM targeted site (orange). Bottom, allele-specific deletions of individual peaks after overlapping NANOG (N) and OCT4 (O) ChIP tracks. Red: $129^{\text{SE-RGM-tdTomato}}$ allele, green: $\text{CAST}^{\text{SE-RGM-eGFP}}$ 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-G0 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 Song et al. [1] Figure S4.


A
Sox2-SE chr3: 34,651,445-34,665,371 mm9 (13 kb)
34,652 kb
34,656 kb
34,660 kb
34,664 kb
OCT4
NANOG
RGM
[0 - 15]
[0 - 15]
peak 1 peak 2
N O N
129 allele
CAST allele
WT
Δ Peak 1-CAST
Δ Peak 2-CAST
Δ Peak 2-129
B
WT
Δ Peak 1-CAST
Δ Peak 2-CAST
Δ Peak 2-129
Sort
3-4 days proliferation
FACS
Demethylation dynamics
De novo methylation dynamics
C
sort
RGM-tdTomato (129)
RGM-eGFP (CAST)
De novo methylation rate
CAST SE-RGM-eGFP allele: %T+G-
129 SE-RGM-tdT allele: %T+G+
Demethylation rate
129 SE-RGM-eGFP allele: %T+G-
Demethylation rate
CAST SE-RGM-tdT allele: %T-G+
D
Day 3
Day 4
T-G+
T+G-
%T+G- : %T-G+ from T+G-
%T+G- : %T-G+ from T-G-
1.0
0.5
0.0
Parental WT
Δ Peak 1-CAST
Δ Peak 2-CAST
Δ Peak 2-129
E
Allele-specific TF enrichement sites deletions (bulk)
sort
RGM-tdTomato (129)
RGM-eGFP (CAST)
F
Day 9
T-G+
T+G-
%T+G- : %T-G+ from T+G-
1.0
0.5
0.0
Ctrl
Δ Peak 1-CAST
Δ Peak 2-CAST
Δ Peak 1+2-CAST
Δ Peak 2-129

unmethylated. Since the Mediator complex interacts with both the SE and the promoter [50], 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 (Figure 5B and 5C, *Sox2* SE and *Mir290* SE boxes; Song et al. [1] Figure S5B, enhancer panels) as well as allele-specific H3K27ac enrichment (Song et al. [1] Figure S5C, enhancer panels). As expected, a decrease in H3K27ac was also observed at promoters residing on the same chromosome with the methylated SE (Figure 5B and 5C, *Sox2* and *Mir290* boxes, and Song et al. [1] Figures S5B and S5C, promoter panels) but not at adjacent regions (Song et al. [1] 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-

Figure 5 *(following page).* 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) Peak calling from H3K27ac ChIP-seq of 4 sorted populations from *MIR290-SE-TG*. *Mir290* SE and *Mir290-295* cluster are boxed in blue. Peak values are normalized using RPKM (reads per million) with a 10-bp bin size. (C) Peak calling from H3K27ac ChIP-seq of 4 sorted populations from SOX2-SE-TG. *Sox2* SE and *Sox2* gene are boxed in blue. Peak values are normalized using RPKM (reads per million) with a 10-bp bin size. (D) Allele-specific expression of *Mir290-295* pri-miRNA (top) and *Sox2* mRNA (bottom) in 3 sorted populations, with VIC-TaqMan probe detecting the 129SE-RGM-tdTomato allele, and FAM-TaqMan probe detecting the $CAST^{SE\text{-}RGM\text{-}eGFP}$ allele in both SE cases. Independently targeted clones for each SE were used as biological replica. Data are represented as mean ± SD. (E) Fold change of total *Mir290-295* pri-miRNA (left) and total *Sox2* mRNA (right) from the 4 sorted populations normalized to that of the T+G- population. Independently targeted clones for each SE were used as biological replica. Data are represented as mean ± SD. (F) Quantification of *Mir290-295* expression on sorted SOX2-SE-TG cells compare to *Sox2* expression (left) and quantification of *Sox2* expression on sorted *MIR290-SE-TG* cells compare to *Mir290-295* expression (right). Data are represented as mean ± SD. See also Song et al. [1] Figure S5.

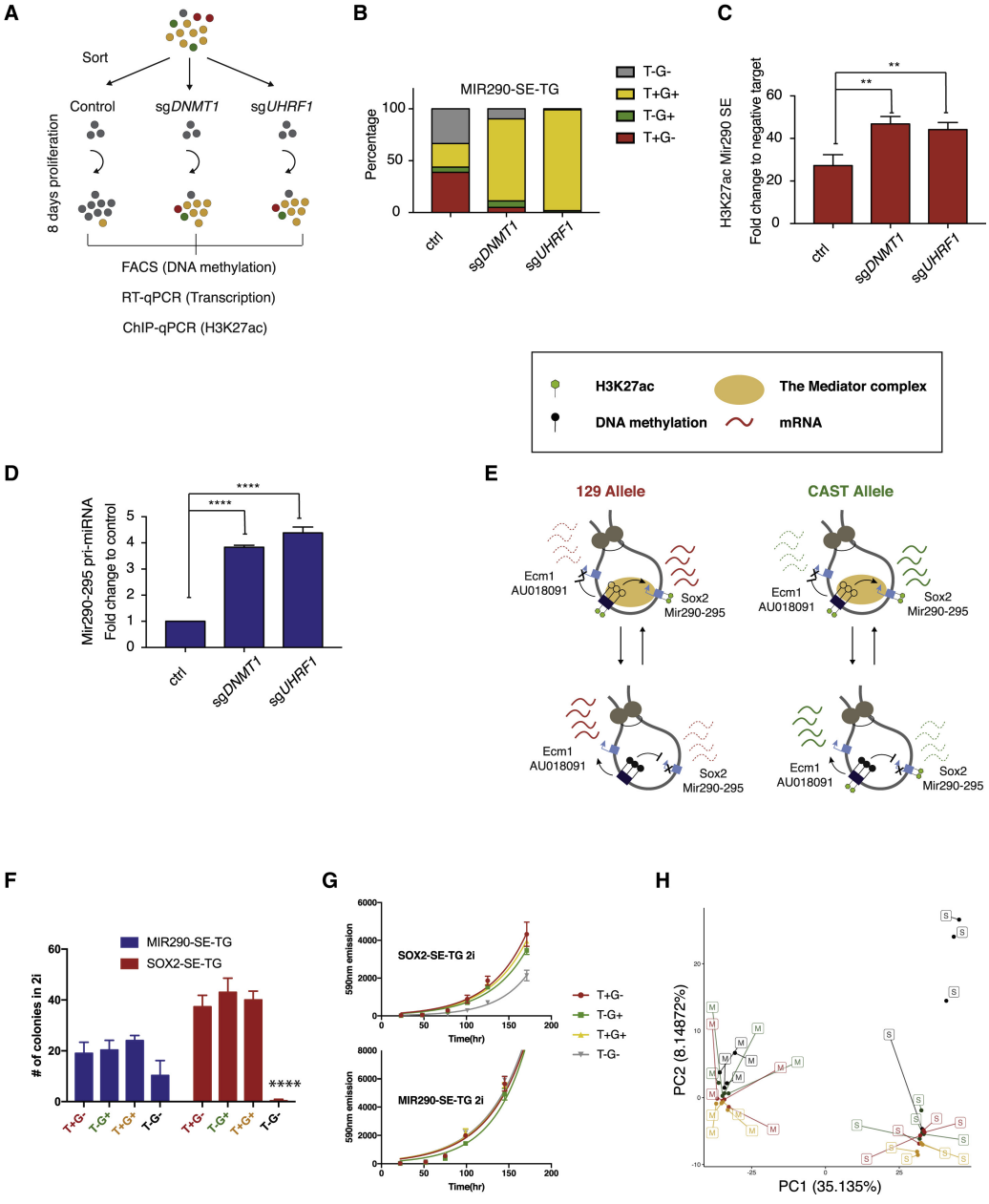

A
Average MED1 IF
Average FISH (n = 62)
Mir290 SE
Random
Average FISH (n = 39)
Average FISH (n = 77)
Average FISH (n = 57)
T+G-
T+G+
T-G+
T-G-
B
Mir290 SE
Mir290
ChIP
Input
Mir290a
Nlrp12
C
Sox2
Sox2 SE
Sox2ot
Gm3143
Gm38509
D
VIC-129(Tomato)
FAM-CAST(GFP)
Ratio to total Mir290-295
MIR290-SE-TG cell
Ratio to total Sox2
SOX2-SE-TG cell
E
Total Mir290-295 relative to T+G-
Total Sox2 relative to T+G-
MIR290-SE-TG cell
SOX2-SE-TG cell
F
Sox2
Mir290-295
Fold change relative to T+G-
n.s
SOX2-SE-TG cell
MIR290-SE-TG cell

molecule RNA FISH (smFISH) could detect a slight decrease of *Sox2* transcripts (Song et al. [1] 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* 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 (Song et al. [1] Figure S6A). As expected, *Mir290-295* and *Sox2* both appeared in the

Figure 6 ***(following page).*** **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 Song et al. [1] Figures S5-S7.


A
Sort
Control
sgDNMT1
sgUHRF1
8 days proliferation
FACS (DNA methylation)
RT-qPCR (Transcription)
ChIP-qPCR (H3K27ac)
B
MIR290-SE-TG
Percentage
T-G-
T+G+
T-G+
T+G-
ctrl
sgDNMT1
sgUHRF1
C
H3K27ac Mir290 SE
Fold change to negative target
ctrl
sgDNMT1
sgUHRF1
H3K27ac
The Mediator complex
DNA methylation
mRNA
D
Mir290-295 pri-miRNA
Fold change to control
ctrl
sgDNMT1
sgUHRF1
E
129 Allele
CAST Allele
Ecm1
AU018091
Sox2
Mir290-295
F
MIR290-SE-TG
SOX2-SE-TG
# of colonies in 2i
T+G-
T-G+
T+G+
T-G-
G
SOX2-SE-TG 2i
MIR290-SE-TG 2i
590nm emission
Time(hr)
H
PC2 (8.14872%)
PC1 (35.135%)

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 hypermethylation strongly correlated with upregulations of both anti-sense transcripts, whereas SE hypo-methylation correlated with inhibition (Song et al. [1] 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 [51].

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).

1.2.5 *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 [52]. Long-term culture of MIR290-SE-TG and SOX2-SE-TG cells in "2i" after passaging from serum + LIF media, though favoring T^+G^+ population decreased but did not abolish heterogeneity completely (Song et al. [1] Figure S6D). The persistence of all four populations in both reporter cell 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 [53, 44, 54, 55], 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 (Figure 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 (Figure 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 (Song et al. [1] 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 (Song et al. [1] 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 (Song et al. [1] 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 SE.

1.2.6 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 [56] and *Mir290-295* expression significantly upregulates at the 4-cell stage [57]. To investigate changes in DNA methylation of the two SEs at single-cell and allelic resolution, we generated transgenic mice homozygous for the $129^{\text{SE-RGM-tdTomato}}$ allele or the $\text{CAST}^{\text{SE-RGM-eGFP}}$ allele and obtained 2–4 cell embryos carrying one $129^{\text{SE-RGM-tdTomato}}$ allele and one $\text{CAST}^{\text{SE-RGM-eGFP}}$ allele by mating animals homozygous for RGM-eGFP or RGM-tdTomato (Figure 7A). The two 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 trophectoderm (TE) [52, 58, 59]. 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 at single-cell resolution (Song et al. [1] Figure S7B). A long-term membrane bound dye (Cy5) was used to track the injected cells (Figure 7D; Song et al. [1] Figure S7C). Figure 7D (SOX2-SE-TG cells) and Song et al. [1] 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).

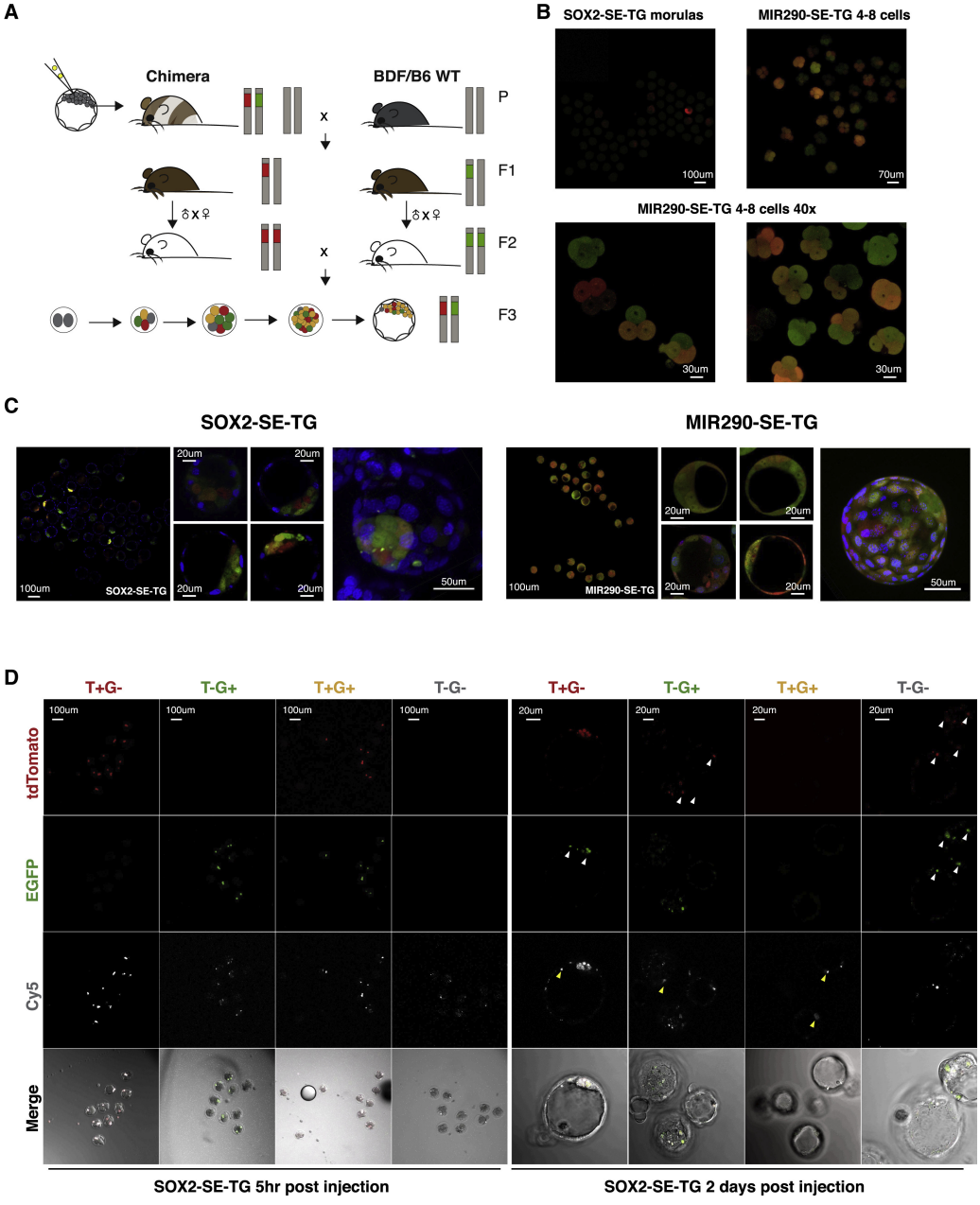
In summary, our data indicate that dynamic DNA methylation exists at active SEs in early preimplantation embryos creating locus-specific epigenetic heterogeneity, recapitulating and extending our observations in ESCs *in vitro*.

1.3 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 [60, 61, 62]. 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

Figure 7 *(following page).* 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 preimplantation 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 10X low magnification, 40X 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 compare to 5 h post-injection. Channels were adjusted for brightness and contrast for optimal visibility. See also Song et al. [1] Figure S7.


A
Chimera
BDF/B6 WT
P
F1
F2
F3
♂x♀
B
SOX2-SE-TG morulas
MIR290-SE-TG 4-8 cells
MIR290-SE-TG 4-8 cells 40x
100um
70um
30um
C
SOX2-SE-TG
MIR290-SE-TG
20um
50um
D
T+G-
T-G+
T+G+
T-G-
tdTomato
EGFP
Cy5
Merge
SOX2-SE-TG 5hr post injection
SOX2-SE-TG 2 days post injection

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 [63, 49]. 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 [64, 65]. 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 (Figure 6F and 6G; Song et al. [1] 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 SE. *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 [66]. 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 [67, 68].

1.4 STAR Methods

1.4.1 Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit poly-clonal anti-H3K27ac	Abcam	Cat#ab4729; RRID:http://antibodyregistry.org/AB_2118291
Rabbit poly-clonal anti-MED1	Abcam	Cat#ab64965; RRID:http://antibodyregistry.org/AB_1142301
Chemicals, Peptides, and Recombinant Proteins		
Puromycin	Sigma Aldrich	Cat#P7255
Thymidine	Sigma Aldrich	Cat#T1895
Retinoic acid	Sigma Aldrich	Cat#R2625
LIF recombinant protein	House-made	N/A
MAPK inhibitor PD0325901	Stemgent	Cat#04-0006-10
GSK-3β inhibitor CHIR99021	Stemgent	Cat#04-0004-10
Critical Commercial Assays		
Zymoclean Gel DNA Recovery Kit	Zymo Research	Cat#D4002
X-tremeGENE 9 DNA Transfection Reagent	Sigma Aldrich	Cat#6365809001
Xfect ESC Transfection Reagent	Clontech	Cat#631320
AlamarBlue Cell Viability Reagent	Bio-Rad	BUF012A
NEBNext®Ultra™DNA Library Prep Kit for Illumina	NEB	Cat#E7370S
NEBNext®Multiplex Oligos for Illumina®	NEB	Cat#E7335S
KAPA mRNA HyperPrep Kit	Roche	Cat# 08098115702
TrueSeq Stranded PolyA prep	Illumina	Cat# 20020595
Accel-NGS 2S PCR-Free Library Kit (96 rxns)	Swift Biosciences	Cat#20096
Qtracker™705 Cell Labeling Kit	Thermo Fisher	Cat# Q25061MP
TaqMan Assay, primer information see table S3	Sigma	N/A
Direct-zol RNA Miniprep	Zymo Research	Cat#R2050
SuperScript III First-Strand Synthesis SuperMix	Life Technologies	Cat#18080400
Fast SYBR Green Master Mix	Life Technologies	Cat#4385618
ProlongGold	Life Technologie	Cat#P36930
Prime-It II Random Primer Labeling Kit	Agilent Technologies	Cat#300385

Continued on next page

Continued from previous page

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw and processed data	This paper	GEO: GSE132416 (subseries: GEO: GSE132376, geo: GSE132404, and geo: GSE132414)
Flowcytometry data, Sanger sequencing trace files, raw images, gels, algorithm scripts	This paper	https://doi.org/10.17632/6vbc6htfnf.1
Experimental Models: Cell Lines		
Mouse: SOX2-SE-TG ESCs	This paper	N/A
Mouse: MIR290-SE-TG ESCs	This paper	N/A
Mouse: Sox2-SE-NanogRGM ESCs	[25]	N/A
Mouse: miR290-SE-NanogRGM ESCs	[25]	N/A
Mouse: SOX2-SE-TG Δpeak 1-CAST ESCs	This paper	N/A
Mouse: SOX2-SE-TG Δpeak 2-CAST ESCs	This paper	N/A
Mouse: SOX2-SE-TG Δpeak 2-129 ESCs	This paper	N/A
Tet1-/- #19	[38]	N/A
Tet1-/- #34	[38]	N/A
Tet2 -/- KO	[39]	N/A
Tet1-/- Tet2-/- #26	[39]	N/A
Tet1-/- Tet2-/- #51	[39]	N/A
Tet1-/- Tet2-/- Tet3 -/- #26	[40]	N/A
Tet1-/- Tet2-/- Tet3 -/- #29	[40]	N/A
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		N/A

Continued on next page

Continued from previous page

REAGENT or RESOURCE	SOURCE	IDENTIFIER
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	[49]	N/A
Recombinant DNA		
miR290-SE-RGM-tdTomato targeting vector	[25]	N/A
miR290-SE-RGM-eGFP targeting vector	This paper	N/A
Sox2-SE-RGM-tdTomato targeting vector	[25]	N/A
Sox2-SE-RGM-eGFP targeting vector	This paper	N/A
pTurbo-Cre	GenBank	AF334827
px330-BFP-sgRNA	[25]	N/A
Software and Algorithms		
MACS2 (ChIP-seq algorithms)	[69]	https://github.com/taoliu/MACS/wiki
Samtools	[70]	http://samtools.sourceforge.net/
BWA	[71]	http://bio-bwa.sourceforge.net/
deepTools 3.0.2	[72]	http://deeptools.readthedocs.io/en/
STAR (v.2.5.3.a)	[73]	https://github.com/alexdobin/STAR
DESeq2 (v1.18.1)	[74]	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
SNPsplit (v0.3.2)	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/SNPsplit/
RSEM (v1.2.31)	N/A	https://deweylab.github.io/RSEM/
PANTHER	[75, 76]	http://pantherdb.org/
Image J	[77])	https://imagej.net/
FlowJo	N/A	https://www.flowjo.com/
PyroMark Q48 Autoprep	QIAGEN	http://www.qiagen.com/us/
Python scripts	This paper; Mendeley Data	https://doi.org/10.17632/6vbc6htfnf.1
MATLAB scripts	[78]	N/A
IGV	Broad Institute	https://software.broadinstitute.org/software/igv
Bismark v0.21.0	Babraham Bioinformatics	https://www.bioinformatics.babraham.ac.uk/projects/bismark

1.4.2 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).

1.4.3 Experimental Model and Subject Details

ESC cell lines

ESC cell culture and proliferation assays All cells were cultured at 37°C with 5% CO_2. 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; 17502048), 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% CO_2 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-tdTomato, Mir290-SE-RGM-eGFP, Sox2-SE-RGM-tdTomato, 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 [38, 39, 40]. 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(129) and sgTFBS-*Sox2*-SE-3(Both); Δ Peak 1+2-CAST: sgTFBS-*Sox2*-SE-1(CAST) and sgTFBS-*Sox2*-SE-3(CAST). All sgRNA sequences are listed in Song et al. [1] 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-um 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 (DPC). 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-tdTomato (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-tdTomato (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 (Song et al. [1] 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 25-50μl KSOM media droplets (CytoSpring # KO102) covered by mineral oil in a 37°C 5% CO_2 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% CO_2. 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.

1.4.4 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 ^{32}P 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 [25]. 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 Song et al. [1] 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 0.25μM 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.

1.4.5 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, Song et al. [1] Table S3) targeting *Sox2* and *Mir290-295* pri-mRNA SNPs. Tukey's multiple comparison (****P<0.0001, **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 Song et al. [1] 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 [79], 2μg 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 Song et al. [1] 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-miRNA 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 (Benjamin-Hochberg procedure).

RNA 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 670 (Stellaris®DesignReady FISH Probes, Cat# VSMF-3075-5-BS) were incubated with the samples for 16 hours at 30°C in hybridization buffer (100 mg/mL dextran sulfate, 25% formamide, 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 [49]. 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 scipy 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 $i = 3.0$ μm). 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 x 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 [20], and number of methylated counts was assumed to be a binomial random variable.

1.4.6 Data and Code Availability

Description: 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).

1.5 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 Chilaka, 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 Saturno-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, Haiting Ma, Emile Wogram, and all of the members of the Jaenisch 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, 1U19AI131135, 5R01MH104610, and 1R01GM114864.

1.5.1 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. conducted blastocyst injections. S.S. and P.R.v.d.B. performed re-analysis of the published scWGBS data, RNA-seq analysis, and smFISH. A.D. and J.E.H. assisted with DNA FISH, IF, and quantitative image analyses. F.S. assisted with cloning, targeting, and designing of the CRISPR knockout experiments and contributed instrumentally to the writing of the manuscript. M.A.C. assisted with teratoma injection.

1.5.2 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.

Acronyms

129xCAST 129xCasteneous

DKO double-knockout

DMR differential methylation region

DNMT DNA methyltransferase

ESC emrbyonic stem cell

PCR polymerase chain reaction

RGM Reporter of Genome Methylation

RNA-seq RNA sequencing

scWGBS single cell WGBS

SE super-enhancer

smFISH single-molecule RNA FISH

SNP single nucleotide polymorphism

T-DMR tissue-specific differential methylation region

TF transcription factor

WGBS whole-genome bisulfite sequencing

1.6 References

[1] Yuelin Song et al. “Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs”. In: *Molecular cell* 0.0 (2019), 905–920.e6. DOI: 10.1016/j.molcel.2019.06.045.

[2] Kenneth C Ehrlich et al. “DNA Hypomethylation in Intragenic and Intergenic Enhancer Chromatin of Muscle-Specific Genes Usually Correlates with their Expression.” In: *The Yale Journal of Biology and Medicine* 89.4 (2016), pp. 441–455.

[3] Thomas Fleischer et al. “DNA methylation at enhancers identifies distinct breast cancer lineages”. In: *Nature Communications* 8.1 (2017), pp. 1–14. DOI: 10.1038/s41467-017-00510-x.

[4] Benedetta Izzi et al. “Allele-specific DNA methylation reinforces PEAR1 enhancer activity”. In: *Blood* 128.7 (2016), pp. 1003–1012. DOI: 10.1182/blood-2015-11-682153.

[5] Peter A Jones. “Functions of DNA methylation: islands, start sites, gene bodies and beyond”. In: *Nature Publishing Group* 13.7 (2012), pp. 484–492. DOI: 10.1038/nrg3230.

[6] Lorenzo Rinaldi et al. “Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis”. In: *Cell Stem Cell* 19.4 (2016), pp. 491–501. DOI: 10.1016/j.stem.2016.06.020.

[7] Chongyuan Luo et al. “Dynamic DNA methylation: In the right place at the right time”. In: *Science* 361.6409 (2018), pp. 1336–1340. DOI: 10.1126/science.aat6806.

[8] Michael B Stadler et al. “DNA-binding factors shape the mouse methylome at distal regulatory regions”. In: *Nature* 480.7378 (2011), pp. 490–495. DOI: 10.1038/nature10716.

[9] GiNell Elliott et al. “Intermediate DNA methylation is a conserved signature of genome regulation”. In: *Nature Communications* 6.1 (2015), pp. 1–10. DOI: 10.1038/ncomms7363.

[10] Holger Heyn et al. “Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer”. In: *Genome Biology* 17.1 (2016), p. 1198. DOI: 10.1186/s13059-016-0879-2.

[11] Gary C Hon et al. “Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues”. In: *Nature Genetics* 45.10 (2013), pp. 1198–1206. DOI: 10.1038/ng.2746.

[12] Ruiwei Jiang et al. “Discordance of DNA Methylation Variance Between two Accessible Human Tissues”. In: *Scientific Reports* 5.1 (2015), pp. 1–8. DOI: 10.1038/srep08257.

[13] Andrew D King et al. “Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells”. In: *Cell Reports* 17.1 (2016), pp. 289–302. DOI: 10.1016/j.celrep.2016.08.083.

[14] Austin Y Shull et al. *DNA Hypomethylation within B-Cell Enhancers and Super Enhancers Reveal a Dependency on Immune and Metabolic Mechanisms in Chronic Lymphocytic Leukemia*. Tech. rep. 2016.

[15] Lih Feng Cheow et al. "Multiplexed locus-specific analysis of DNA methylation in single cells". In: *Nature Protocols* 10.4 (2015), pp. 619–631. DOI: 10.1038/nprot.2015.041.

[16] H Guo et al. "Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing". In: *Genome Research* 23.12 (2013), pp. 2126–2135. DOI: 10.1101/gr.161679.113.

[17] Hongshan Guo et al. "Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing". In: *Nature Protocols* 10.5 (2015), pp. 645–659. DOI: 10.1038/nprot.2015.039.

[18] Fan Guo et al. "Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells". In: *Cell Research* 27.8 (2017), pp. 967–988. DOI: 10.1038/cr.2017.82.

[19] Steffen Rulands et al. "Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency". In: *Cell Systems* 7.1 (2018), 63–76.e12. DOI: 10.1016/j.cels.2018.06.012.

[20] Sébastien A Smallwood et al. "Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity". In: *Nature Methods* 11.8 (2014), pp. 817–820. DOI: 10.1038/nmeth.3035.

[21] B Jin et al. "DNA Methylation: Superior or Subordinate in the Epigenetic Hierarchy?" In: *Genes & Cancer* 2.6 (2011), pp. 607–617. DOI: 10.1177/1947601910393957.

[22] Heng Zhu et al. "Transcription factors as readers and effectors of DNA methylation". In: *Nature Publishing Group* 17.9 (2016), pp. 551–565. DOI: 10.1038/nrg.2016.83.

[23] Anshul Kundaje et al. "Integrative analysis of 111 reference human epigenomes". In: *Nature* 518.7539 (2015), pp. 317–330. DOI: 10.1038/nature14248.

[24] Stephen Wilson et al. "A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer". In: *npj Systems Biology and Applications* 4.1 (2018), pp. 1–10. DOI: 10.1038/s41540-018-0061-4.

[25] Yonatan Stelzer et al. "Tracing Dynamic Changes of DNA Methylation at Single-Cell Resolution". In: *CELL* 163.1 (2015), pp. 218–229. DOI: 10.1016/j.cell.2015.08.046.

[26] Yonatan Stelzer et al. "Parent-of-Origin DNA Methylation Dynamics during Mouse Development". In: *CellReports* 16.12 (2016), pp. 3167–3180. DOI: 10.1016/j.celrep.2016.08.066.

[27] Hisato Kobayashi et al. "Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks". In: *PLOS Genetics* 8.1 (2012), e1002440. DOI: 10.1371/journal.pgen.1002440.

[28] Danny Leung et al. “Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1”. In: *Proceedings of the National Academy of Sciences* 111.18 (2014), pp. 6690–6695. DOI: 10.1073/pnas.1322273111.

[29] Stefanie Seisenberger et al. “The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells”. In: *Molecular cell* 48.6 (2012), pp. 849–862. DOI: 10.1016/j.molcel.2012.11.001.

[30] Yidan Hu et al. “An estimated method of urban PM2.5 concentration distribution for a mobile sensing system”. In: *Pervasive and Mobile Computing* 25 (2016), pp. 88–103. DOI: 10.1016/j.pmcj.2015.06.004.

[31] Zakary S Singer et al. “Dynamic Heterogeneity and DNA Methylation in Embryonic Stem Cells”. In: *Molecular cell* 55.2 (2014), pp. 319–331. DOI: 10.1016/j.molcel.2014.06.029.

[32] Jiho Choi et al. “Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells”. In: *Nature* 548.7666 (2017), pp. 219–223. DOI: 10.1038/nature23274.

[33] Harry G Leitch et al. “Naive pluripotency is associated with global DNA hypomethylation”. In: *Nature Structural & Molecular Biology* 20.3 (2013), pp. 311–316. DOI: 10.1038/nsmb.2510.

[34] Ye-Ji Sim et al. “2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms”. In: *Stem Cell Reports* 8.5 (2017), pp. 1312–1328. DOI: 10.1016/j.stemcr.2017.04.001.

[35] Ferdinand von Meyenn et al. “Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells”. In: *Molecular cell* 62.6 (2016), p. 983. DOI: 10.1016/j.molcel.2016.06.005.

[36] Masaki Yagi et al. “Derivation of ground-state female ES cells maintaining gamete-derived DNA methylation”. In: *Nature* 548.7666 (2017), pp. 224–227. DOI: 10.1038/nature23286.

[37] Xiaoji Wu et al. “TET-mediated active DNA demethylation: mechanism, function and beyond”. In: *Nature Reviews Genetics* 18.9 (2017), pp. 517–534. DOI: 10.1038/nrg.2017.33.

[38] Meelad M Dawlaty et al. “Tet1 Is Dispensable for Maintaining Pluripotency and Its Loss Is Compatible with Embryonic and Postnatal Development”. In: *Cell Stem Cell* 9.2 (2011), pp. 166–175. DOI: 10.1016/j.stem.2011.07.010.

[39] Meelad M Dawlaty et al. “Combined Deficiency of Tet1 and Tet2 Causes Epigenetic Abnormalities but Is Compatible with Postnatal Development”. In: *Developmental Cell* 24.3 (2013), pp. 310–323. DOI: 10.1016/j.devcel.2012.12.015.

[40] Meelad M Dawlaty et al. “Loss of Tet Enzymes Compromises Proper Differentiation of Embryonic Stem Cells”. In: *Developmental Cell* 29.1 (2014), pp. 102–111. DOI: 10.1016/j.devcel.2014.03.003.

[41] Angelika Feldmann et al. “Transcription Factor Occupancy Can Mediate Active Turnover of DNA Methylation at Regulatory Regions”. In: *PLOS Genetics* 9.12 (2013), e1003994. DOI: 10.1371/journal.pgen.1003994.

[42] Matthew T Maurano et al. “Role of DNA Methylation in Modulating Transcription Factor Occupancy”. In: *Cell Reports* 12.7 (2015), pp. 1184–1195. DOI: 40_10.1016/j.celrep.2015.07.024.

[43] Yinong Yin et al. “Recent advances in oxide thermoelectric materials and modules”. In: *Vacuum* 146 (2017), pp. 356–374. DOI: 10.1016/j.vacuum.2017.04.015.

[44] Denes Hnisz et al. “Super-Enhancers in the Control of Cell Identity and Disease”. In: *CELL* 155.4 (2013), pp. 934–947. DOI: 10.1016/j.cell.2013.09.053.

[45] Aimée M Deaton et al. “CpG islands and the regulation of transcription.” In: *Genes & Development* 25.10 (2011), pp. 1010–1022. DOI: 10.1101/gad.2037511.

[46] Yuval Dor et al. “Principles of DNA methylation and their implications for biology and medicine”. In: *The Lancet* 392.10149 (2018), pp. 777–786. DOI: 10.1016/S0140-6736(18)31268-6.

[47] Dirk Schübeler. “Function and information content of DNA methylation”. In: *Nature* 517.7534 (2015), pp. 321–326. DOI: 10.1038/nature14192.

[48] Zachary D Smith et al. “DNA methylation: roles in mammalian development”. In: *Nature Publishing Group* 14.3 (2013), pp. 204–220. DOI: 10.1038/nrg3354.

[49] Benjamin R Sabari et al. “Coactivator condensation at super-enhancers links phase separation and gene control”. In: *Science* 361.6400 (2018), eaar2555. DOI: 10.1126/science.aar3958.

[50] Warren A Whyte et al. “Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes”. In: *CELL* 153.2 (2013), pp. 307–319. DOI: 10.1016/j.cell.2013.03.035.

[51] James A Gagnon et al. “Efficient Mutagenesis by Cas9 Protein-Mediated Oligonucleotide Insertion and Large-Scale Assessment of Single-Guide RNAs”. In: *PLOS ONE* 9.5 (2014), e98186. DOI: 10.1371/journal.pone.0098186.

[52] Jennifer Nichols et al. “Naive and Primed Pluripotent States”. In: *Cell Stem Cell* 4.6 (2009), pp. 487–492. DOI: 10.1016/j.stem.2009.05.015.

[53] J M Calabrese et al. “RNA sequence analysis defines Dicer’s role in mouse embryonic stem cells”. In: *Proceedings of the National Academy of Sciences* 104.46 (2007), pp. 18097–18102. DOI: 10.1073/pnas.0709193104.

[54] Rudolf Jaenisch et al. “Stem Cells, the Molecular Circuitry of Pluripotency and Nuclear Reprogramming”. In: *CELL* 132.4 (2008), pp. 567–582. DOI: 10.1016/j.cell.2008.01.015.

[55] Matt Thomson et al. “Pluripotency Factors in Embryonic Stem Cells Regulate Differentiation into Germ Layers”. In: *CELL* 145.6 (2011), pp. 875–889. DOI: 10.1016/j.cell.2011.05.017.

[56] TK Mistri et al. “Dynamic changes in Sox2 spatio-temporal expression promote the second cell fate decision through Fgf4/ Fgfr2 signaling in preimplantation mouse embryos”. In: *The Biochemical journal* 475.6 (2018), pp. 1075–1089. DOI: 10.1042/BCJ20170418.

[57] Lea A Medeiros et al. “Mir-290–295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects”. In: *Proceedings of the National Academy of Sciences* 108.34 (2011), pp. 14163–14168. DOI: 10.1073/pnas.1111241108.

[58] Alireza Paikari et al. “The eutheria-specific miR-290 cluster modulates placental growth and maternal-fetal transport”. In: *Development* 144.20 (2017), pp. 3731–3743. DOI: 10.1242/dev.151654.

[59] Eryn Wicklow et al. “HIPPO Pathway Members Restrict SOX2 to the Inner Cell Mass Where It Promotes ICM Fates in the Mouse Blastocyst”. In: *PLOS Genetics* 10.10 (2014), e1004618. DOI: 10.1371/journal.pgen.1004618.

[60] Matthew D Schultz et al. “Human body epigenome maps reveal noncanonical DNA methylation variation”. In: *Nature* 523.7559 (2015), pp. 212–216. DOI: 10.1038/nature14465.

[61] Christoph Weigel et al. “Epigenetic regulation of diacylglycerol kinase alpha promotes radiation-induced fibrosis”. In: *Nature Communications* 7.1 (2016), pp. 1–12. DOI: 10.1038/ncomms10893.

[62] Michael J Ziller et al. “Charting a dynamic DNA methylation landscape of the human genome”. In: *Nature* 500.7463 (2013), pp. 477–481. DOI: 10.1038/nature12433.

[63] Won-Ki Cho et al. “Mediator and RNA polymerase II clusters associate in transcription-dependent condensates”. In: *Science* 361.6400 (2018), pp. 412–415. DOI: 10.1126/science.aar4199.

[64] Qiaolin Deng et al. “Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells”. In: *Science* 343.6167 (2014), pp. 193–196. DOI: 10.1126/science.1245316.

[65] Björn Reinius et al. “Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation”. In: *Nature Publishing Group* 16.11 (2015), pp. 653–664. DOI: 10.1038/nrg3888.

[66] Hiroshi I Suzuki et al. “Super-Enhancer-Mediated RNA Processing Revealed by Integrative MicroRNA Network Analysis”. In: *CELL* 168.6 (2017), 1000–1014.e15. DOI: 10.1016/j.cell.2017.02.015.

[67] Holger Heyn et al. “DNA methylation profiling in the clinic: applications and challenges”. In: *Nature Publishing Group* 13.10 (2012), pp. 679–692. DOI: 10.1038/nrg3270.

[68] Keith D Robertson. “DNA methylation and human disease”. In: *Nature Publishing Group* 6.8 (2005), pp. 597–610. DOI: 10.1038/nrg1655.

[69] Yong Zhang et al. “Model-based Analysis of ChIP-Seq (MACS)”. In: *Genome Biology* 9.9 (2008), pp. 1–9. DOI: 10.1186/gb-2008-9-9-r137.

[70] H Li et al. “The Sequence Alignment/Map format and SAMtools”. In: *Bioinformatics* 25.16 (2009), pp. 2078–2079. DOI: 10.1093/bioinformatics/btp352.

[71] H Li et al. “Fast and accurate short read alignment with Burrows-Wheeler transform”. In: *Bioinformatics* 25.14 (2009), pp. 1754–1760. DOI: 10.1093/bioinformatics/btp324.

[72] Fidel Ramírez et al. “deepTools: a flexible platform for exploring deep-sequencing data”. In: *Nucleic Acids Research* 42.W1 (2014), W187–W191. DOI: 10.1093/nar/gku365.

[73] Alexander Dobin et al. “STAR: ultrafast universal RNA-seq aligner”. In: *Bioinformatics* 29.1 (2012), pp. 15–21. DOI: 10.1093/bioinformatics/bts635.

[74] Michael I Love et al. “Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2”. In: *Genome Biology* 15.12 (2014), p. 31. DOI: 10.1186/s13059-014-0550-8.

[75] H Mi. “The PANTHER database of protein families, subfamilies, functions and pathways”. In: *Nucleic Acids Research* 33.Database issue (2004), pp. D284–D288. DOI: 10.1093/nar/gki078.

[76] Paul D Thomas et al. “PANTHER: a library of protein families and subfamilies indexed by function.” In: *Genome Research* 13.9 (2003), pp. 2129–2141. DOI: 10.1101/gr.772403.

[77] Caroline A Schneider et al. “NIH Image to ImageJ: 25 years of image analysis”. In: *Nature Methods* 9.7 (2012), pp. 671–675. DOI: 10.1038/nmeth.2089.

[78] Arjun Raj et al. “Imaging individual mRNA molecules using multiple singly labeled probes”. In: *Nature Methods* 5.10 (2008), pp. 877–879. DOI: 10.1038/nmeth.1253.

[79] Bluma J Lesch et al. “A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis”. In: *Proceedings of the National Academy of Sciences* 110.40 (2013), pp. 16061–16066. DOI: 10.1073/pnas.1315204110.