Dynamic Regulation and Functions of Locus-specific DNA Methylation

By

Yuelin Song

B.S. Biological Sciences
Zhejiang University, 2014

Submitted to the Department of Biology
In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

At the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Signature of Author ……...
Department of Biology
May, 2020

Certified by…….....
Rudolf Jaenisch
Professor of Biology
Thesis Supervisor

Accepted by…….....
Mary Gehring
Associate Professor of Biology; Member, Whitehead Institute
Co-Chair, Biology Graduate Committee

Dynamic Regulation and Functions of Locus-specific DNA Methylation

By
Yuelin Song

Abstract

The role and regulation of DNA methylation at various genetic elements have gathered tremendous interest over decades. The methylomes of many cell types have been described, revealing a dynamic and tissue-specific pattern of DNA methylation (tissue-specific differentially methylated regions, T-DMRs) in the distal regulatory elements, such as enhancers. The formation of T-DMRs still remain mysterious, however, one of their interesting features observed in mouse ES cells (mESCs) is the low-to-intermediate levels of average DNA methylation resulted from inter-cellular epigenetic heterogeneity. Given the transcriptional repressive role of DNA methylation at promoters, such non-zero levels of enhancer methylation is interesting to characterize. Prior to this thesis, a reporter for genomic DNA methylation (RGM) has been developed in the Jaenisch lab, when targeted into T-DMRs of interest, the surrounding locus-specific DNA methylation will be reported as on-and-off of fluorescent signals in single cells. We further modified RGM to investigate the regulation of DNA methylation at pluripotency super-enhancers *Sox2* and *MiR290* at single allele level in mESCs. We found that enhancer DNA methylation is surprisingly dynamic with two alleles independently being demethylated and methylated within days. Such dynamics is the basis of epigenetic and transcriptional heterogeneity and is coupled with changes in histone modifications and transcription factor binding. Furthermore, epigenetic heterogeneity was also observed in the developing pre-implantation embryos. Our work provided a paradigm to functionally investigate locus-specific DNA methylation in heterogenous tissues in diseases and development.

The regulation of locus-specific DNA methylation is highly context dependent and sensitive to the environment. Our understanding of how locus-specific DNA methylation is regulated *in vivo* is still restricted to a few genomic elements. The appendix of this thesis attempts to generate an animal model to expand the scope of research on DNA methylation to retroelement-associated metastable epialleles.

Thesis Supervisor: Rudolf Jaenisch
Title: Professor of Biology and Member of Whitehead Institute

Dedication

I dedicate this thesis to my parents who have always been unconditionally supporting my pursuit of knowledge and happiness.

Acknowledgements

The completion of this thesis wouldn't have been possible without the tremendous support and help I received over years from my mentors and colleagues.

First foremost, I would like to thank my thesis advisor Dr. Rudolf Jaenisch, whose work I've always admired since college and I almost knew that I would join his lab when applying to graduate school. What I appreciate the most for being a graduate student in the Jaenisch lab is the great scientific freedom. Rudolf has always encouraged me when I took challenges and believed in me when things didn't work as expected. I thank Rudolf for supporting me to propose projects, independently design experiments and collaborate outside of the lab, and providing me the opportunities to attend and speak at scientific conferences. I enjoyed writing my first scientific paper with Rudolf. He was very patient and edited all my smallest typos rounds after rounds. He taught me how to convey messages clearly and effectively. The journey of publishing was never a smooth sail, but thanks to Rudolf's strong support, I was able to get through the most frustrated moments and continue to grow as a scientist. Rudolf's acuity to science and the ability of immediately pointing at critical things always amaze me, and I am constantly inspired by his genuine interest and excitement for science. I especially want to thank Rudolf for his generous support when I wanted to explore alternative careers while in graduate school, without which I wouldn't be so sure about my continuing pursuit for academia. Apart from science, Rudolf has a passion for life that is so infectious and I love listening about his hardcore hiking adventures and exchanging with my amateur camping experiences, all of which little things made my PhD years very enjoyable. Rudolf, thank you for being a great mentor.

I would like to thank another two postdoc colleagues who mentored me: Dr. Yonatan Stelzer and Dr. Frank Soldner. Yonatan taught me all the baby steps in DNA methylation research and mouse genetics by hand. Frank is so knowledgeable and experienced, without whose help I wouldn't be able to get through many of my experimental crises and drainages of ideas. The daily coffee chats with Yonatan and Frank on science, politics, life or literally everything were so much fun and inspiring. The unforgettable intense scientific arguments among us were probably one of the biggest drivers for me to stay in science, as they made me realize how much we love and care about what we do. Both of them are extremely smart and insightful scientists, whom I deeply admire and learnt greatly from.

I also would like to thank formal graduate student Dr. Chikdu Shivalila, who taught me cloning, gene targeting, and importantly, how to enjoy graduate school. Many thanks to Dr. Malkiel Cohen for his help throughout my PhD and the free drives to the WI retreat every year. I would like to thank Ruthie Flannery for her "mouse-mommy-like" attentive care for our animal facility and her help in my never-ending genotyping requests. I would like to thank Dr. Styliani Markoulaki, Jesse Drotar, Nick Rosenau, and Dina Rooney for their support in mouse experiments, Raaji Alagappan for her strong sense of responsibility to care for the whole lab, Dongdong Fu for her fast and reliable work on tissue-processing, and Gerry Kemske, Robert Burger and Carrie Garrett-Engele for making our lab running smoothly. I thank all current and past Jaenisch lab members for their help and the supportive environment they contribute to create. I would also like to thank Dr. Richard Young, Dr. Stefan Semrau and their students and postdocs for the great collaboration experience.

I would like to thank my thesis committee, Dr. Phil Sharp and Dr. Tyler Jacks, for their feedbacks during my thesis committee meetings, preliminary exams, and their strong support in my academic career.

I am forever grateful to my undergraduate mentors, Dr. Yingjie Wang and Dr. Binghui Shen. Both of them care deeply for nurturing the next generation of Chinese scientists. When they started their joint lab at Zhejiang University after returning from the US, they gave me my own bench and interesting projects to explore despite the fact that I was only a sophomore. They both helped me to connect with US research labs through their network after they learnt that I would like to pursue my PhD in the US. With the help from Dr. Binghui Shen, I had the opportunity to conduct summer research at Dr. Kun-Liang Guan's lab at USCD working with a PhD student on different isoforms of PKC; with the help from Dr. Yingjie Wang, I was able to have the research experience in Dr. Robert G. Roeder's lab at the Rockefeller University under the guidance of Dr. Zhanyun Tang and Dr. Weiyi Chen on chromatin biology. Both experiences were so valuable and rare for international students at that time, and they shaped my scientific interest in epigenetics. I thank Dr. Guan and Dr. Roeder labs for the training and their generous support in my graduate school application.

Finally, to my parents, thank you for giving me unconditional supports for whichever path I choose in life. Thank you for being my biggest champions. You are the reason why I can enjoy all the freedom and privileges, and be who I am today.

Table of Contents

CHAPTER 1

CHAPTER 2

CHAPTER 3

APPENDIX

Chapter 1. Introduction.

1.1 The Spatial and Temporal Landscape of Genomic DNA CpG Methylation in Mouse Embryonic Stem Cells

5'methyl-cytosine in the context of DNA CpG dinucleotides (referred as DNA methylation below) has been proposed as an inheritable epigenetic mark regulating gene expression and cell fate since 1975[1]. However, the specific roles of genomic DNA methylation across different genomic elements still remain to be elucidated. In the vertebrate genome most CpGs exist in CpG-islands (CGI, CpG rich regions with approx. 1kb), which are protected from DNA methylation in somatic cells[2,3]. CpG-islands residing in more than half of the genome tend to be associated with transcriptional start sites (TSSs) of house-keeping and developmental genes[2]. The rest of the genome is depleted of CpGs but the remaining non-CGI CpGs are heavily methylated (>70%)[4], especially at gene bodies, transposable elements and gene deserts[3]. Non-coding regulatory elements, such as enhancer and insulators have also low CpG densities. These regions tend to have more variable levels of DNA methylation in a tissue-specific manner[5-8] (**Figure 1**).

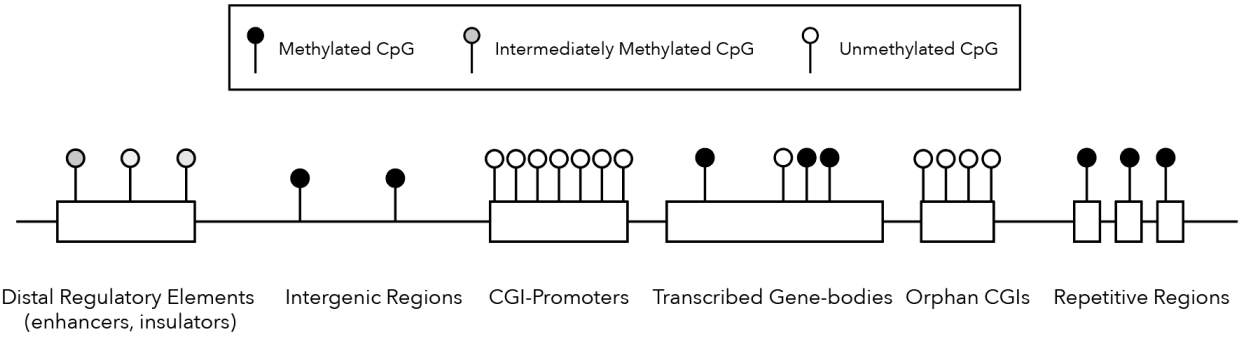

Figure 1. The Genomic Distribution of CpGs and CpG DNA Methylation.

DNA methylation is essential to mammalian development and most dynamic at early developmental stages. There are two waves of DNA methylation reset throughout mammalian development (**Figure 2**). The first one is a global erasure of DNA methylation except for genomic imprinted regions in the paternal and maternal genomes post-fertilization. This wave of reset brings the embryonic genome to a hypomethylated state at the blastocyst stage (E3.5-4.5 in mouse and E5-6 in human)[9,10]. Upon implantation, global *de novo* methylation occurs in the epiblast, derived from the inner cell mass of the blastocyst, which gives rise to the embryo proper[10-12]. Tissue-specific regulatory regions undergo demethylation during embryonic development or adult stem cell differentiation[6,13,14]. The second reset occurs in the developing primordial germ cells (PGCs), where parent-of-origin DNA methylation is removed after which germ-line- and sex-specific methylation patterns are established at different times in male and female embryos[12,15-17].

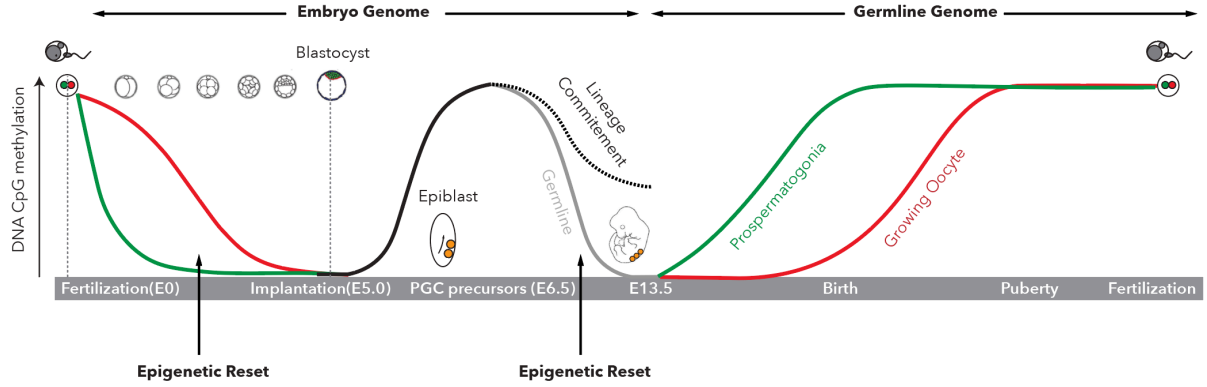

Figure 2. DNA Methylation Reprogramming (Resets) During Development

Mouse embryonic stem cells (mESCs) are isolated from the inner cell mass at the blastocyst stage and can be maintained at the pluripotent state in different media conditions *in-vitro*[18-20]. mESCs possess pluripotent signatures of genomic DNA methylation, which is extensively studied in the context of gene regulation as well as

cell fate decisions[21]. The total level of DNA methylation of mESCs is similar to that of liver and kidney cells *in vivo*[22]. Different from many other precursor cells as well as conventionally grown human ESCs, mESCs can tolerate global loss of methylation when depleted of all methyl-transferases and still remain pluripotent gene signatures[23-27]. These DNA methylation deficient mESCs are capable to proliferate however fail to differentiate, indicating that DNA methylation is dispensable for maintaining the pluripotent state but essential to cells assuming a specific lineage in mouse[28]. CGI-promoters of pluripotency genes are usually hypomethylated in mESCs, such as those of *Oct4* and *Nanog*[29,30]. Immunoprecipitation of methylated DNA combined with DNA microarrays (mDIP) revealed that promoter methylation contributes to silencing of some developmental genes in ESCs[31]. Nevertheless, a majority of CGI-associated promoters of developmental genes in mESCs remain unmethylated as shown using reduced representation bisulfite sequencing (RRBS) that specifically enriches for CpG dense regions[32]. Whole-genome bisulfite sequencing (WGBS) sufficiently covering comparatively less-studied and hypermethylated CpG-poor regions showed that, in certain tissues or cell types, short stretches of DNA are hypomethylated, which constitutes differentially methylated regions (DMRs)[5,33,34]. mESC-specific DMRs identified in WGBS studies by comparing to other somatic tissues are predominantly found at distal regulatory regions of pluripotency genes[5,13,17,35,36]. Interestingly, unlike CGI-associated proximal promoters that are usually protected from DNA methylation, tissue-specific DMR (T-DMR)-associated enhancers show low-to-intermediate levels of DNA methylation[5]. Methylation of these regions is more dynamically regulated during development than CGI promoters, which is consistent with the role of enhancers in instructing tissue-specific gene expression[6,13,37,38]. However, the identities and biological functions of tissue-specific DMRs, especially causal links to gene regulations and cell states, still remain to be elucidated.

1.2 Enzymatic Regulations of DNA Methylation Dynamics in ESCs

The processes of methylation and demethylation concertedly regulate methylation levels in ESCs (**Figure 3**). The mouse and human genomes both encode 5 DNMTs: DNMT1, DNMT2, DNMT3A and DNMT3B, and DNMT3L, although DNMT2 mainly has tRNA methyl-transferase activity[39]. DNMT3L has no enzymatic activity but is essential in facilitating DNMT3A-mediated *de novo* methylation in germ-cells and also protects DNMT3A2 from degradation in ESCs[40]. *De novo* methylation is enzymatically carried out by DNMT3A and DNMT3B, which is essential for establishing the methylation pattern in the developing embryo[41-43]. In ESCs, the predominantly active isoforms are DNMT3A, DNMT3A2, and DNMT3B1, the deletion of which result in progressive loss of methylation[42]. DNMT3A and DNMT3B have overlapping and distinct genomic targets in ESCs and both are excluded from active enhancers and promoters[44]. DNMT3B targets actively transcribed gene bodies, consistent with intragenic methylation's role in promoting transcriptional efficiency[44]. DNMT3A appears to target distal promoters and *de novo* methylates pluripotency enhancers upon differentiation[42,45,46]. Maintenance methylation is mainly mediated by DNMT1 in partnership with UHRF1, although it has been shown that DNMT3s also contributes to DNA methylation maintenance[47,48]. DNMT1 is targeted to the replication fork and interacts with PCNA, preferentially methylating hemi-methylated DNA[49]. Passive demethylation is caused by the absence of maintenance methylation activity and it critically contributes to the methylation dynamics throughout development. Genetic ablation of DNMT1 and UHRF1 leads to rapid demethylation of ESCs[23,50-52]. The post-fertilization maternal genome undergoes demethylation by excluding oocyte-specific DNMT1o from the nucleus[53]. In PGC epigenetic reprogramming, demethylation was shown to be achieved by nuclear extrusion or transcriptional repression of UHRF1 while

PGCs undergo several rounds of division[54]. Primed mESCs cultured in LIF/serum transition into the naïve state resembling the ICM when switched to "2i" (GSKβi/MEKi) media with the global demethylation during this process being mainly achieved by passive demethylation as a consequence of DNMT1/UHRF1 downregulation[55,56]. Passive demethylation also contributes to lineage-specific demethylation, however, how it is regulated in a locus-specific manner and to what degree it matters in comparison to active demethylation has not been fully elucidated[57].

Active or active-passive demethylation is initiated by TET enzymes converting 5′ methyl-cytosine (5mC) into 5′ hydroxymethyl-cytosine (5hmC). 5hmC can be further oxidized to 5′ formalcytosine (5fC) and 5′ carboxylcytosine (5caC) by TET enzymes, or deaminated by AID/APOBEC family members into 5′ hydroxymethyluracil (5mU). 5mC can also be directly deaminated by AID/APOBEC into thymidine, which together with 5caC and 5mU are recognized by DNA base-excision repair pathway (BER) members such as TDG and subsequently replaced by newly synthesized cytosine[58]. Active demethylation has been shown to be present in many cell types and plays an important role during development[58-60]. Compared to the maternal genome where demethylation is mainly passive, the paternal genome undergoes TET3-mediated active demethylation post-fertilization with much faster kinetics[61]. Another example supporting the mechanism of active demethylation is the observation of cycles of methylation and demethylation within 100min periods in response to environmental stimuli in the absence of cell division[62,63]. Post-mitotic neurons also show high levels of 5hmC and activity-dependent active demethylation mediated by TET1 and APOBEC1 was observed in adult mouse brain[64,65].

5hmC inhibits the maintenance methylation machinery and therefore causes passive demethylation in the subsequence cell division (active-passive demethylation, **Figure 4**). An example of active-passive demethylation can be found in PGC

reprogramming, where genomic imprints are erased. 5hmC accumulates after initial active demethylation and is then diluted in a replication-dependent manner[60]. In mESCs, both TET1 and 2 are highly expressed and 5hmC can be readily detected. TET1 is enriched at CGI promoters and TET2 mostly at actively transcribed gene bodies and enhancers[66,67]. TET3 is expressed at low level and contributes to ~2% of 5hmC[68]. Similar to DNMTs, loss of all TET enzymes does not affect ESC maintenance but impairs cell differentiation. TET triple-knockout mice show gastrulation defects causing embryonic lethalality[69]. Genetic ablation of the downstream BER component TDG revealed wide-spread accumulations of 5fC and 5caC at distal regulatory elements and poised promoters. Although the exact function of oxidized 5mC in transcriptional regulation is still unclear, they could be bound by other proteins and mediate local changes of chromatin[59].

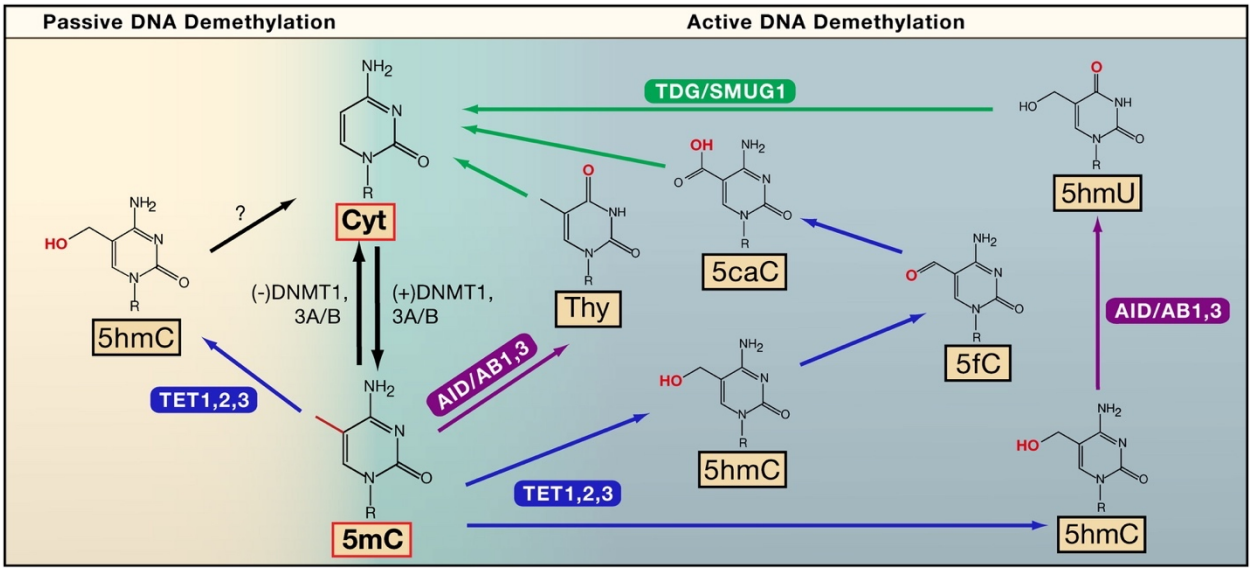

Figure 3. DNA Methylation and Demethylation Pathways (Active or Passive). Bhutani et al. *Cell*. 2011.

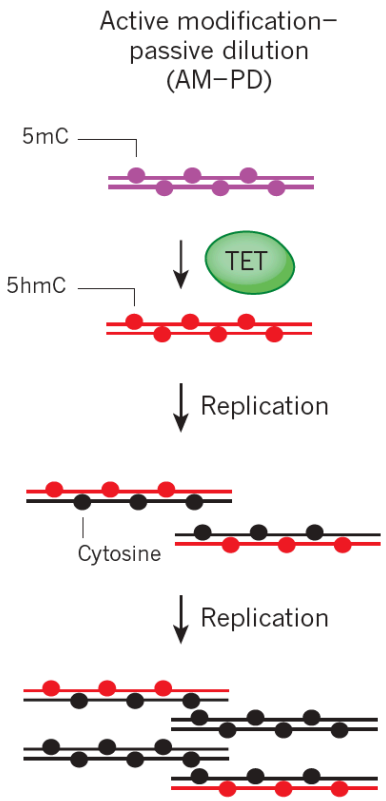

Figure 4. Active-passive Demethylation (AM-PD: active modification-passive demethylation). Kohli et al. *Nature*. 2013.

1.3 DNA Methylation in Transcriptional Regulation at Different Genetic Elements

It is challenging to generalize a universal rule of DNA methylation in gene regulation at different genomic elements. The function of DNA methylation is most extensively characterized at proximal promoters, where it serves as a stable silencing mark[3,70]. Around 50% CGIs are associated with TSSs and remain hypomethylated at house-keeping and developmental genes, and many of the CGIs remain hypomethylated even after transcription becomes inactive[70]. However, during development there are small subsets of CGI promoters become *de novo* methylated, which usually leads to long-term gene silencing[2]. In ESCs, promoter DNA methylation

contributes to long-term silencing of germline-specific genes, such as *Dazl*, whereas promoters of highly expressed pluripotency genes such as *Oct4, Nanog, Dppa4, Tdgf1* are unmethylated[31,71-73]. The mechanism of DNA methylation mediating promoter silencing involves concerted alterations of chromatin structure as well as histone modifications[74,75]. As ESCs differentiate, promoters of other lineage-specific genes become *de novo* methylated after gene silencing occurs, leading to long-term repression via heterochromatin formation, subsequently restricting pluripotency and promoting lineage commitment[76]. Similarly, in *X* chromosome inactivation, DNA methylation at gene promoters occurs several days post-inactivation and maintains repression of silenced genes on the inactivated *X* (*Xi*) whereas "escaper" genes have lower methylation at promoters[77-79]. Therefore, DNA methylation at promoters serves to stably maintain but likely not to initiate gene silencing.

In contrast, DNA methylation at gene bodies is not associated with repression in mammals. Gene-bodies are in general CpG poor and therefore are highly methylated[37,70,80]. Methylated gene bodies are associated with increased expression levels in cancer and somatic cells[81-83], and perturbation of gene body methylation using demethylating agents such as 5'azacytidine leads to down-regulation of gene expression[81]. Some "orphan CGIs" are also found at gene-bodies and thought to possess potential intra-genic alternative promoter activities[84]. "Orphan CGIs" are more methylated than promoter CGIs (20~30% vs 3% methylation)[2,84,85]. Mechanistically, DNA methylation at gene bodies does not block elongation despite repressive histone modifications such as H3K9me3 in cancer cells[86]. In mESCs DNA methylation even facilitates transcriptional elongation via preferential binding by histone variant H2A.B at gene bodies not promoters[87]. On the other hand, due to its inhibitory role in transcriptional initiation, DNA methylation at gene bodies could also enhance

transcriptional efficiency by suppressing alternative promoter usage, retroelement activation and regulatory anti-sense ncRNA production[88-90].

The regulation and function of DNA methylation at enhancers are less characterized than that of promoters[3], potentially due to the fact that most enhancers locate at CpG poor regions where sufficient coverage at single-base resolutions is only achieved with WGBS and variable levels of methylation complicating the interpretation of correlations[5,13,91]. Nevertheless, locus-specific studies have shown that methylated enhancers tend to correlate with reduced transcription of cognate genes, although some enhancers still remain unmethylated even after becoming dormant[13,92]. The evidence of DNA methylation at enhancers being a negative regulator of transcription is variable. In breast cancer[92,93] a strong anti-correlation of gene expression with promoter methylation is seen, whereas in blood and skin cells this relationship is only moderate and even varies between different sub-group cells types[94]. However, enhancer methylation is becoming increasingly important as growing evidence supports the notion that T-DMRs that are only hypomethylated in certain lineages are preferentially associated with tissue-specific enhancers[5,13,95,96]. Therefore, understanding the mechanism of how T-DMRs at enhancers regulate tissue-specific gene expression is a crucial issue. The *MyoD* enhancer was shown to be hypermethylated in non-muscle cells but unmethylated in embryonic myogenic cells in a human *MyoD* enhancer transgenic mouse model. Though essential for *MyoD* tissue-specific expression, *MyoD* enhancer demethylation does not immediately cause *MyoD* expression[95], which is similar to the delayed chicken *clys* gene expression after enhancer demethylation[97]. Together these examples indicate that enhancer demethylation may be a necessary but not sufficient step for gene activation. Interestingly, mice carrying a CpG-mutant *MyoD* enhancer, which is unable to become methylated, have no aberrant ectopic *MyoD* expression, suggesting that enhancer

methylation is not required for gene silencing in this context[95]. This echoes the induced liver-specific demethylation of the *Tat* enhancer by glucocorticoids (GC). GC withdraw after initial exposure does not remethylate the enhancer, despite *Tat* expression being shut down in the absence of stimuli[98]. In other cases, aberrant enhancer hypomethylation correlating with cognate gene upregulation have been found in ER-positive tumors as well as in neurons from psychosis patients consistent with the importance of proper enhancer methylation in normal physiology[92,99]. In contrast to the delayed transcriptional activation after enhancer demethylation of *MyoD* and *clys*, demethylation of an enhancer-like Treg-specific demethylated region (TSDR) at the *Foxp3* locus is sufficient to induce *Foxp3* expression in Tregs[100,101]. However, this may not be generalizable as the TSDR is proximal to *Foxp3* promoter and contains CpG-islands which is less typical for most distal regulatory elements[5,13,102]. In mESCs, enhancer demethylation by genetic ablation of DNA methyl-transferases led to both up- and down-regulation of target gene expression, with the transcriptionally suppressive roles of DNA methylation validated at only a few loci[103].

Therefore, the answer as to the exact role and detailed steps of enhancer methylation dynamics in gene regulation is far from being conclusive and highly context-dependent[37,104]. Since enhancers are major transcription factor binding sites and are subjected to other complex epigenetic regulations to be discussed below, a more comprehensive view of all these components will facilitate a better understanding of how enhancer T-DMRs regulate gene expression.

1.4 Cross-talk Between DNA Methylation and Histone Modification

DNA methylation and histone modifications represent two parallel and complementary mechanisms in gene expression regulation. In many scenarios of promoter silencing, two processes are coordinated through direct interactions

between DNMTs and histone modifiers. For example, DNMTs can recruit the Polycomb repressive complex 2 by interacting with EZH2 to catalyze H3K27me3. H3K9 methyl-transferase G9a can recruit DNMT3s to *de novo* methylate pluripotency gene promoters. It was also shown that interactions between DNMT3s with SUV39H1 and SETDB1 mediate pericentric heterochromatinization[74]. On a different level, histone modifications can directly affect DNMT binding. For example, H3K36me3 at gene-bodies is recognized by the PWWP domain of DNMT3s and is required for intragenic DNA methylation[105]. H3K4me2 and H3K4me3 modified nucleosomes exclude DNMTs by disrupting binding from the ADD domain, thus protecting the CGI from being methylated[106]. Reciprocally, DNA methylation status can influence histone modifications. Unmethylated CGIs are bound by the CXXC domains of some H3K4 methyl-transferases as well as of CXXC finger protein 1 (CFP1), which recruits H3K4me3 methyltransferases that do not have the CXXC domain to activate CGI promoters[107,108]. In contrast, DNA methylation at a repressor domain upstream of H19 can recruit DNA methylation binding protein MeCP2, which facilitates H3K9 methylation and gene repression[109].

H3K27ac is a shared marker between active enhancers and promoters. In general, H3K27ac is inversely-correlated with DNA methylation as acetylated histones promote chromatin accessibility and transcriptionally permissivity[110,111]. Methyl-binding MeCP2, beside facilitating H3K9 methylation, is also in complex with histone deacetylase (HDAC)[112]. Histone deacetylation condenses chromatin inhibiting transcription factor access and facilitating gene silencing on methylated DNA. Treatment of live cells with HDAC inhibitor can induce global hypomethylation by altering DNMT1 nuclear dynamics and protein level[113,114]. On the other hand, knocking out DNMT1 leads to extensive changes of the H3K27ac landscape in mESCs, with both gains and losses at different enhancers, indicating a more complicated and locus-

dependent relationship between DNA methylation and H3K27ac[103]. H3K4me1 is distinguishably enriched at enhancers compared to promoters. In contrast to H3K27ac, H3K4me1-marked enhancers are usually primed to be active. The presence of H3K4me1 at enhancers often precedes nucleosome depletion and H3K27 acetylation, leaving an opportunity for enhancer activation[104]. A meta-analysis of high-throughput profiles of DNA methylation and histone modifications observed a pattern of H3K4me1 enrichment at enhancers with intermediate DNA methylation level, such as at the enhancer of *c-Myc* and *Sox2*. In ESCs, regions with low DNA methylation level has elevated H3K4me3 level and diminished H3K4me1, which seems to mark an enhancer-promoter transition. The molecular details of this correlation cannot yet be explained by any intermediate mechanisms such as binding of known MBD-containing proteins or 5hmC patterns[115]. The inconclusive understanding of how DNA methylation and enhancer histone modifications are connected is partially due to current sequencing-based technologies that use bulk-cell populations at a snapshot time-point and thus have limited temporal and locus-specific resolution. Observations yielded from the genome-wide high-throughput approaches are valuable yet mostly correlative. Therefore, it is important to utilize new systems that allow functional explanation of the various (anti-)correlations between DNA methylation and histone modifications at different genomic loci.

1.5 Cross-talk Between DNA Methylation and Transcription Factor Binding

Tissue-specific enhancers are bound by transcription factors (TFs). On one hand, the binding activity of a TF can be influenced both positively and negatively by DNA methylation. On the other hand, TF binding can recruit DNA (de)methylation machineries inducing local changes of DNA methylation. It is still challenging to

distinguish the hierarchical cross-talk between DNA methylation changes and TF binding events in transcriptional regulation.

Due to the transcriptionally suppressive role of DNA methylation at promoters and some enhancers, TF binding is traditionally thought to be inhibited by methylated DNA either directly or indirectly through methyl-binding proteins[116]. Crystal structure studies have shown that 5mC resides in the major groove of DNA, and, because the bulky methyl-group narrows the minor groove, the altered structure of the DNA can directly affect TF binding[117]. MBD-containing proteins can bind methylated DNA directly at CpGs, such as MeCP2, or to any methylated DNA independent of the sequence, such as MBD1,2,4[118]. These methyl-binding proteins are associated with protein partners that mediate heterochromatin formation and transcriptional repression. They also exclude the demethylase TET1 from methylated DNA[118,119,120] and thus indirectly inhibit TF binding to methylated DNA. However, evidence indicates that DNA methylation can both promote and inhibit TF binding. A systematic SELEX (systematic evolution of ligands by exponential enrichment) assay applied a collection of 542 full-length human TFs and DNA binding domains to CpG-methylated DNA and showed that the methylated DNA inhibited binding of many major classes of TFs including bHLH, bZIP, and ETS. In contrast, TFs containing homeodomains such as POU and NEAT were shown to preferentially bind to methylated DNA[121]. However, the forcefully expressed POU-domain containing TF Oct4 in colon carcinoma cells can only bind to DNA when it is unmethylated[122], raising the caution that the knowledge about TF-binding and DNA methylation gained through *in vitro* TF-binding motifs binding assays still requires *in vivo* validation, as nucleosomal chromatin presents a very different target for TF binding as compared to naked DNA[118]. Nevertheless, a few *in vivo* studies have yielded valuable insights on the cross-talk between DNA methylation and TF-binding in a regulated chromatin environment. By mapping DNaseI

hypersensitivity sites (DHS) in the absence of DNA methylation in DNMT triple knockout (TKO) cell , the study showed that NRF1 gained additional thousands binding sites especially in the CpG-poor distal regulatory elements, which were originally methylated in WT mESCs. The appearance of new NRF1 binding sites in TKO cells was concomitant with enrichment of H3K27ac and initiation of aberrant transcription, indicating that DNA methylation in wild-type cells was safeguarding against aberrant transcription by blocking NRF1 binding[123]. In contrast, KLF4 ChIP-bisulfite sequencing experiments have shown that KLF4 binds to two motifs differently in hESCs. KLF4 binding to one motif is preferred when the sequence is methylated whereas binding to the other motif is inhibited by methylation, providing an example of context-dependent methylation sensitivity of TF-binding[124].

The above-mentioned examples show that the DNA methylation status could act upstream excluding or attracting certain TFs. As the genomic DNA methylation pattern correlates with cell-type, locus-specific DNA methylation needs to be inherited mitotically to maintain a certain cell state. To maintain a given differentiation state DNA methylation at promoters and enhancers may serve to prevent TFs from aberrantly activating genes which are not supposed to be expressed in the given cell type, and meanwhile attracting methyl-binding repressors to secured the silenced state. Similarly, the pre-existing unmethylated promoters and enhancers need to be actively protected from *de novo* methylation for maintaining cell-type specific gene expression.

However, the chick-and-egg question is whether TF binding induces DNA methylation changes or whether the pre-existing DNA methylation state regulates TF binding affinity. The above-mentioned KLF4 is a pioneer transcription factor in iPSC programming and initiates remodeling of the chromatin, opening up pluripotency enhancers and TSSs[125]. In fact, most pioneer TFs instruct cell-fate transitions and a common view of tissue-specific enhancer activation is that pioneer TF binding instructs

local changes of the heterochromatin by interactions with various epigenetic modifiers and chromatin remodelers, which further leads to TF-dependent coactivator assembly and transcription. Therefore, the ability to bind methylated enhancers is an important attribute of pioneer TF function. In addition to KLF4, other pioneer TFs such as HOXA5, HOXA9, GATA3, GATA4, FOXC1, FOXK1, FOXK2, and FOXA1 have been shown *in vitro*[118] to have methyl-CpG binding activity. However, *in vivo* data for these pioneer TFs binding being sufficient to change DNA methylation is either lacking or supports the opposite view, i.e. that the pre-existing DNA methylation status determines whether the pioneer TF initiates transcription. For example, FOXA1 binds to target enhancers during neural differentiation, which is inhibited by DNA methylation contrary to previous *in vitro* observations[126,127]. An *in vivo* example of TF driving methylation changes is MEF2 and SIX binding to the upstream of the *Myogenin* promoter. *Myogenin* starts muscle-specific expression at around E8.5 following a rostro-caudal gradient. Its weak-CpG island promoter is initially methylated and bound by ZBTB38 and MBD2 to maintain the silenced state, both of which are degraded upon the onset of differentiation with the concomitant demethylation of *Myogenin* promoter leading to activation of the gene. Mutations of either MEF2 or SIX binding sites abolishes DHS formation illustrating the requirement for binding of two TFs in initiating demethylation, which precedes the activation of *Myogenin*[128,129]. Perhaps the most robust evidence of TF binding driving DNA methylation changes comes from the insulator CTCF and the transcriptional repressor REST. Stadler et al. showed CTCF binds regulatory elements those overlap with low-methylated regions (LMR). A reporter construct with or without CTCF binding motif was inserted into the same genomic locus of mESCs. The constructs carrying a mutated CTCF binding motif became methylated in mESCs whereas reduced methylation was observed in cells containing the functional CTCF binding motif construct. A construct containing pre-

methylated CTCF binding site was then further inserted into the same locus causing CTCF binding-induced local reduction of methylation *in vivo*, confirming the driver role of CTCF binding in creating LMRs[5]. Similar to CTCF, REST also occupies LMRs in ESCs and genetic deletion of REST led to increased methylation at its target LMRs. Reintroduction of REST restored the LMR, showing the necessity and sufficiency of REST binding in creating and maintaining DNA methylation patterns[5].

High-throughput *in vitro* biochemical studies as well as genome-wide epigenetic and TF binding profiling yielded valuable correlative patterns between TF binding and DNA methylation. However, the relationship between genomic DNA methylation and binding activities of many TFs in a physiologically relevant setting still awaits further mechanistic dissection, especially at enhancers, the activation of which is tissue-specific. With the development of genomic and epigenomic editing technologies, manipulations of TF binding sites or of locus-specific methylation status endogenously should bring a clearer picture of this mechanistic hierarchy.

1.6 Super-Enhancers and the Mediator Complex

Classical enhancers are defined as DNA elements bound by transcription factors (TF), which activate transcription of a gene independently of distance or orientation with respect to the gene[130]. Super-enhancers are a large cluster of enhancers that often have unusually high levels of coactivator binding, such as the Mediator complex, and display active histone modifications, such as H3K27ac[131]. In mESCs, super-enhancers are defined by either of the following criteria: (1) the binding of all three master TFs Oct4, Sox2 and Nanog, (2) the proximity of stitched enhancer within 12.5kb, (3) a ranking of H3K27ac or MED1 ChIP-seq signal where the intensity of the enhancer is above the slope of 1[130-132]. Compared to normal enhancers, super-enhancers are large and usually locate at genes controlling cell identity with their activity being more

sensitive to the knockdown of the Mediator complex subunits. Therefore, super-enhancers tend to be associated with T-DMRs. Most super-enhancers in mESCs are associated with pluripotency genes such as *Sox2, Oct4, Nanog, Prdm14, Esrrb, Klf4* and microRNA clusters such as *Mir290-295*. Individual enhancers within a super-enhancer can cooperate in activating transcription consistent with that super-enhancers usually drive high level of target gene expression. Strong enrichment of the Mediator complex plays a central role in coordinating transcription and was initially used to identify super-enhancers in mESCs[132,133]. The Mediator complex consists of approximately 30 subunits, including MED1 to MED31, and the cyclin-dependent kinase (CDK) 8-cyclin C pair as well as several paralogs. The subunits of the Mediator complex are not fixed and can vary between different complex isoforms. The earliest discovered function of the Mediator complex was a "bridge" to bring TF-bound enhancers to PolII and general transcription factor (GTF) machineries to promoters by a "looping" mechanism and to form the pre-initiation complex (PIC)[134]. Besides initiation, the Mediator complex has also been shown to interact with histone modifiers, lncRNAs, transcriptional elongation, termination and RNA splicing factors[133].

Recently, an alternative view of how transcription factors interacting with DNA regulatory elements has evolved when the Mediator complex together with other chromatin associated coactivators was found to form phase-separated condensates *in vitro* and *in vivo*[135,136]. Transcriptional factor and coactivator condensates form when proteins reach a critical concentration and therefore are dependent on the density and affinity of TF-binding sites at enhancers[137]. As TF and coactivators may recruit or exclude DNA methylation machinery from binding to enhancers in a locus-dependent manner, it is unresolved how DNA methylation affects transcriptional condensates. Direct or indirect interactions between the Mediator Complex and the DNA (de) methylation machineries or methyl-binding proteins may be involved in coordinating

transcriptional activity. It was shown that Mediator CDKs have repressive role in transcription of C/EBPβ target genes, and this repressive effect is achieved in part through recruiting histone arginine methyl-transferase PRMD5. The product of H4R3me2 correlated with DNMT3A recruitment and methylation at target gene promoters[138]. The Mediator complex subunit MED14 in *Arabidopsis thaliana* was shown to promote transcription at highly methylated transposable elements with *med14-3* mutation inducing decreased non-CG DNA methylation at chromosomal pericentromeric regions[139]. In summary, it remains to be elucidated whether the Mediator complex can directly cross-talk with DNMTs, TETs or methyl-binding proteins at enhancers.

1.7 Enhancer DNA Methylation Heterogeneity

DNA methylation at a certain CpG in a single cell at a given time is a binary event with values of 0 (biallelically unmethylated), 0.5 (mono-allelically methylated), or 1 (biallelically methylated). However, many active enhancers show a range of low-level value of methylation between 0-0.5, which could be a result of averaging methylated and unmethylated alleles from bulk-cell sequencing. In a cultured cell population, for example ESCs, such continuous values of low-level methylation (0-0.5) of a specific loci indicates cell-to-cell DNA methylation heterogeneity. Recent development of single cell RRBS (scRRBS) and WGBS (scWGBS) in ESCs indeed revealed widespread locus-specific methylation heterogeneity, especially at LMR active enhancers[140-142]. Interestingly, transcription levels of ESC genes have also been shown to be heterogeneous[143] suggesting the possibility that the transcriptional level heterogeneity has an epigenetic basis. However, most sequencing-based observations are correlative lacking functional validation. As cells differentiate, decommissioning enhancers gain DNA methylation to a critical level that shuts down target gene

expression. ESCs represent a transient pluripotent state *in vivo*, and this leads to a hypothesis that the heterogenous DNA methylation level may be a specific feature of active pluripotency enhancers, which is regulated in the way to permit both demethylation and *de novo* methylation upon environmental and developmental signaling, similar to the bivalent H3K4me3 and H3K27me3 modifications at "poised" developmental enhancers[104].

The methylation heterogeneity of rapidly dividing ESCs allows a dynamic view where enhancers of any cell can adopt the methylated or unmethylated state at a given time, and the heterogeneous levels of methylation collectively constituting the low-to-intermediate level of bulk-cell methylation. Unfortunately, bisulfite sequencing even at single base resolution does not allow validation of such hypothesis, because each profile represents a snapshot with little means to trace individual cells longitudinally. In addition, many of the CpG poor enhancers are not sufficiently covered by WGBS or scWGBS and allelic information is even harder to retrieve if SNP information is not available. Therefore, a different approach such as using reporters that allows real-time resolution is needed to validate the aforementioned hypothesis as well as to further investigate the causal link between epigenetic and transcriptional level of heterogeneity.

1.8 Genomic DNA Methylation Reporter for Locus-specific Studies in Single Cells

To date, many technologies have been adopted to study DNA methylation. Those measuring genomic average methylation level include WGBS and RRBS, methylated DNA immunoprecipitation (MeDIP), mass spectrometry and HPLC related methods. Single-cell bisulfite sequencing can reveal intercellular differences of genome-wide methylation. Combined-bisulfite restriction analysis (COBRA) and other site-specific methylation-sensitive restriction digestion analysis, and pyro-sequencing,

on the other hand, can generate sensitive quantification of DNA methylation at a locus-specific level[144]. However, neither of these methods can provide single-cell, locus-specific and real-time resolution of DNA methylation at the same time, especially *in vivo*.

Previously, Stelzer et al. in the Jaenisch lab generated a genomic reporter of DNA methylation (RGM) with the aim to trace dynamic changes of methylation locus-specifically in single cells. The RGM consists of a methylation sensitive *Snrpn* minimal promoter driving the expression of a fluorescent protein. The reporter construct can be readily inserted into T-DMR of interest and the methylation status of the *Snrpn* promoter follows that of the surrounding genomic loci. If unmethylated, the RGM drives fluorescent protein expression in a single cell (**Figure 5**). The faithfulness of the RGM in reporting endogenous genomic DNA methylation has been validated in different genetic elements such as CGI promoters[145], imprinting control regions (ICR)[146], retro-transposable elements (unpublished), and enhancers[145]. The fluorescent signal allows flow cytometry assisted sorting (FACS) to purify cells with a defined DNA methylation state at a particular locus, as well as real-time monitoring of the changes of DNA methylation both *in vitro* and *in vivo*. Especially for rare populations of cells that exhibit different methylation patterns different from the majority, the resolution given by RGM is superior to current existing sequencing-based technologies. For example, parent-of-origin imprinting is thought to be faithfully maintained in adult somatic tissues[147]. Using RGM, loss of imprinting pattern at the *Dlk1-Dio* IG DMR was found in multiple adult somatic tissues, which also leads to corresponding cell-type specific allelic transcription of genes controlled by this ICR[146]. Compared to other methylation analysis methods, RGM provides continuous tracing of locus-specific methylation changes in live cells on the allelic level. The basis, regulation and functional impacts of DNA methylation at tissue-specific enhancers and its intricate relationship with histone

modification, TF binding and transcription still remain to be elucidated at the mechanistic level, and the RGM system provides an attractive approach to address this question in single cells and on individual alleles.

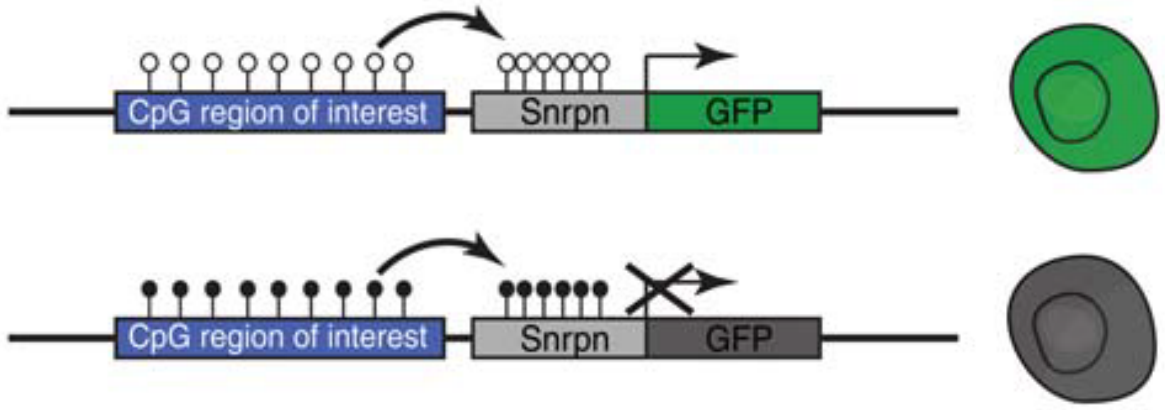

Figure 5. RGM Reporter Translates Genomic DNA Methylation Information into Fluorescent Signal in Single Cells. Stelzer et al. *CSH Symposia on Quantitative Biology*, 2015.

1.9 Why Is Enhancer DNA Methylation Heterogenous and What Does It Mean Functionally?

Analysis of bulk-WGBS and scWGBS data all pointed to the fact that enhancer methylation is heterogeneous and that the LMRs created by such heterogeneity being an enhancer-specific feature has functional impact on transcription. To test this hypothesis and to further investigate its functional implication requires separation of cells with heterogenous methylation status at enhancers, which is a task the RGM can fulfill. Mouse ESCs are developmentally equivalent cells which have shown DNA methylation heterogeneity at enhancer-associated T-DMRs. The high homologous recombination rate of mESCs allows efficient gene targeting and the self-renewal and differentiation potentials allow examination of enhancer DNA methylation during

dynamic cell-fate changes. As two alleles of enhancers could have non-synchronized activity and epigenetic states, I created allele-specific RGM targeted mESCs and transgenic animals at pluripotency super-enhancers *Sox2* and *Mir290*. The aim of my thesis is to answer the following questions: (1) Is DNA methylation heterogeneity the molecular basis of the low-to-intermediate level of enhancer methylation in mESCs? (2) If so, what creates DNA methylation heterogeneity at super-enhancers. (3) What is the functional impact of enhancer methylation heterogeneity and does it exist in the developing mouse embryos?

References

1 Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226-232 (1975).

2 Deaton, A. M. & Bird, A. CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010-1022, doi:10.1101/gad.2037511 (2011).

3 Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484-492, doi:10.1038/nrg3230 (2012).

4 Feng, S. H. *et al.* Conservation and divergence of methylation patterning in plants and animals. *P Natl Acad Sci USA* **107**, 8689-8694, doi:10.1073/pnas.1002720107 (2010).

5 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-495, doi:10.1038/nature10716 (2011).

6 Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477-481, doi:10.1038/nature12433 (2013).

7 Elliott, G. *et al.* Intermediate DNA methylation is a conserved signature of genome regulation. *Nat Commun* **6**, doi:Artn 636310.1038/Ncomms7363 (2015).

8 West, A. G., Gaszner, M. & Felsenfeld, G. Insulators: many functions, many mechanisms. *Genes Dev* **16**, 271-288, doi:10.1101/gad.954702 (2002).

9 Cockburn, K. & Rossant, J. Making the blastocyst: lessons from the mouse. *J Clin Invest* **120**, 995-1003, doi:10.1172/JCI41229 (2010).

10 Messerschmidt, D. M., Knowles, B. B. & Solter, D. DNA methylation dynamics during epigenetic reprogramming in the germline and preimplantation embryos. *Genes Dev* **28**, 812-828, doi:10.1101/gad.234294.113 (2014).

11 Zernicka-Goetz, M., Morris, S. A. & Bruce, A. W. Making a firm decision: multifaceted regulation of cell fate in the early mouse embryo. *Nat Rev Genet* **10**, 467-477, doi:10.1038/nrg2564 (2009).
12 Smallwood, S. A. & Kelsey, G. De novo DNA methylation: a germ cell perspective. *Trends Genet* **28**, 33-42, doi:10.1016/j.tig.2011.09.004 (2012).
13 Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* **45**, 1198-U1340, doi:10.1038/ng.2746 (2013).
14 Reizel, Y. *et al.* Postnatal DNA demethylation and its role in tissue maturation. *Nat Commun* **9**, 2040, doi:10.1038/s41467-018-04456-6 (2018).
15 Saitou, M., Kagiwada, S. & Kurimoto, K. Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139**, 15-31, doi:10.1242/dev.050849 (2012).
16 Smith, Z. D. *et al.* DNA methylation dynamics of the human preimplantation embryo. *Nature* **511**, 611-615, doi:10.1038/nature13581 (2014).
17 Seisenberger, S. *et al.* The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Mol Cell* **48**, 849-862, doi:10.1016/j.molcel.2012.11.001 (2012).
18 Keller, G. Embryonic stem cell differentiation: emergence of a new era in biology and medicine. *Genes Dev* **19**, 1129-1155, doi:10.1101/gad.1303605 (2005).
19 Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156, doi:10.1038/292154a0 (1981).
20 Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523, doi:10.1038/nature06968 (2008).
21 Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat Rev Genet* **14**, 204-220, doi:10.1038/nrg3354 (2013).
22 Biniszkiewicz, D. *et al.* Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. *Mol Cell Biol* **22**, 2124-2135, doi:10.1128/mcb.22.7.2124-2135.2002 (2002).
23 Tsumura, A. *et al.* Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805-814, doi:10.1111/j.1365-2443.2006.00984.x (2006).
24 Liao, J. *et al.* Targeted disruption of DNMT1, DNMT3A and DNMT3B in human embryonic stem cells. *Nat Genet* **47**, 469-478, doi:10.1038/ng.3258 (2015).
25 Trowbridge, J. J., Snow, J. W., Kim, J. & Orkin, S. H. DNA Methyltransferase 1 Is Essential for and Uniquely Regulates Hematopoietic Stem and Progenitor Cells. *Blood* **114**, 163-163 (2009).
26 Fan, G. P. *et al.* DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *J Neurosci* **21**, 788-797 (2001).

27 Jackson-Grusby, L. *et al*. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat Genet* **27**, 31-39 (2001).
28 Panning, B. & Jaenisch, R. DNA hypomethylation can activate Xist expression and silence X-linked genes. *Gene Dev* **10**, 1991-2002, doi:DOI 10.1101/gad.10.16.1991 (1996).
29 Hattori, N. *et al*. Epigenetic control of mouse Oct-4 gene expression in embryonic stem cells and trophoblast stem cells. *J Biol Chem* **279**, 17063-17069, doi:10.1074/jbc.M309002200 (2004).
30 Imamura, M. *et al*. Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *Bmc Dev Biol* **6**, doi:Artn 3410.1186/1471-213x-6-34 (2006).
31 Fouse, S. D. *et al*. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3K4/K27 trimethylation. *Cell Stem Cell* **2**, 160-169, doi:10.1016/j.stem.2007.12.011 (2008).
32 Meissner, A. *et al*. Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766-U791, doi:10.1038/nature07107 (2008).
33 Slieker, R. C. *et al*. Identification and systematic annotation of tissue-specific differentially methylated regions using the Illumina 450k array. *Epigenet Chromatin* **6**, doi:Artn 2610.1186/1756-8935-6-26 (2013).
34 Song, F. *et al*. Tissue specific differentially methylated regions (TDMR): Changes in DNA methylation during development. *Genomics* **93**, 130-139, doi:10.1016/j.ygeno.2008.09.003 (2009).
35 Leung, D. *et al*. Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *P Natl Acad Sci USA* **111**, 6690-6695, doi:10.1073/pnas.1322273111 (2014).
36 Kobayashi, H. *et al*. Contribution of Intragenic DNA Methylation in Mouse Gametic DNA Methylomes to Establish Oocyte-Specific Heritable Marks. *Plos Genet* **8**, doi:ARTN e100244010.1371/journal.pgen.1002440 (2012).
37 Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Reviews Genetics* **13**, 484-492, doi:10.1038/nrg3230 (2012).
38 Ong, C. T. & Corces, V. G. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics* **12**, 283-293, doi:10.1038/nrg2957 (2011).
39 Lyko, F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nature Reviews Genetics* **19**, 81-92, doi:10.1038/nrg.2017.80 (2018).
40 Veland, N. *et al*. DNMT3L facilitates DNA methylation partly by maintaining DNMT3A stability in mouse embryonic stem cells. *Nucleic Acids Res* **47**, 152-167, doi:10.1093/nar/gky947 (2019).

41 Kaneda, M. *et al*. Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429**, 900-903, doi:10.1038/nature02633 (2004).
42 Chen, T. P., Ueda, Y., Dodge, J. E., Wang, Z. J. & Li, E. Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol Cell Biol* **23**, 5594-5605, doi:10.1128/Mcb.23.16.5594-5605.2003 (2003).
43 Okano, M., Bell, D. W., Haber, D. A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257, doi:Doi 10.1016/S0092-8674(00)81656-6 (1999).
44 Baubec, T. *et al*. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243-U278, doi:10.1038/nature14176 (2015).
45 Gu, T. P. *et al*. DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome biology* **19**, doi:ARTN 8810.1186/s13059-018-1464-7 (2018).
46 Petell, C. J. *et al*. An epigenetic switch regulates de novo DNA methylation at a subset of pluripotency gene enhancers during embryonic stem cell differentiation. *Nucleic Acids Res* **44**, 7605-7617, doi:10.1093/nar/gkw426 (2016).
47 Walton, E. L., Francastel, C. & Velasco, G. Maintenance of DNA methylation Dnmt3b joins the dance. *Epigenetics-Us* **6**, 1373-1377, doi:10.4161/epi.6.11.17978 (2011).
48 Feng, J. *et al*. Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat Neurosci* **13**, 423-U437, doi:10.1038/nn.2514 (2010).
49 Vilkaitis, G., Suetake, I., Klimasauskas, S. & Tajima, S. Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J Biol Chem* **280**, 64-72, doi:10.1074/jbc.M411126200 (2005).
50 Bostick, M. *et al*. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317**, 1760-1764, doi:10.1126/science.1147939 (2007).
51 Jeltsch, A. & Jurkowska, R. Z. New concepts in DNA methylation. *Trends Biochem Sci* **39**, 310-318, doi:10.1016/j.tibs.2014.05.002 (2014).
52 Piccolo, F. M. & Fisher, A. G. Getting rid of DNA methylation. *Trends Cell Biol* **24**, 136-143, doi:10.1016/j.tcb.2013.09.001 (2014).
53 Cardoso, M. C. & Leonhardt, H. DNA methyltransferase is actively retained in the cytoplasm during early development. *J Cell Biol* **147**, 25-32, doi:DOI 10.1083/jcb.147.1.25 (1999).

54 Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *Embo Journal* **32**, 340-353, doi:10.1038/emboj.2012.331 (2013).
55 Marks, H. *et al*. The Transcriptional and Epigenomic Foundations of Ground State Pluripotency. *Cell* **149**, 590-604, doi:10.1016/j.cell.2012.03.026 (2012).
56 von Meyenn, F. *et al*. Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells (vol 62, pg 848, 2016). *Mol Cell* **62**, 983-983, doi:10.1016/j.molcel.2016.06.005 (2016).
57 Oda, M., Oxley, D., Dean, W. & Reik, W. Regulation of lineage specific DNA hypomethylation in mouse trophectoderm. *PloS one* **8**, e68846, doi:10.1371/journal.pone.0068846 (2013).
58 Bhutani, N., Burns, D. M. & Blau, H. M. DNA demethylation dynamics. *Cell* **146**, 866-872, doi:10.1016/j.cell.2011.08.042 (2011).
59 Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* **18**, 517-534, doi:10.1038/nrg.2017.33 (2017).
60 Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472-479, doi:10.1038/nature12750 (2013).
61 Oswald, J. *et al*. Active demethylation of the paternal genome in the mouse zygote. *Curr Biol* **10**, 475-478, doi:10.1016/s0960-9822(00)00448-6 (2000).
62 Kangaspeska, S. *et al*. Transient cyclical methylation of promoter DNA. *Nature* **452**, 112-115, doi:10.1038/nature06640 (2008).
63 Metivier, R. *et al*. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45-50, doi:10.1038/nature06544 (2008).
64 Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423-434, doi:10.1016/j.cell.2011.03.022 (2011).
65 Mellen, M., Ayata, P. & Heintz, N. 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. *Proc Natl Acad Sci U S A* **114**, E7812-E7821, doi:10.1073/pnas.1708044114 (2017).
66 Hon, G. C. *et al*. 5mC Oxidation by Tet2 Modulates Enhancer Activity and Timing of Transcriptome Reprogramming during Differentiation. *Mol Cell* **56**, 286-297, doi:10.1016/j.molcel.2014.08.026 (2014).
67 Huang, Y. *et al*. Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A* **111**, 1361-1366, doi:10.1073/pnas.1322921111 (2014).

68 Lu, F., Liu, Y., Jiang, L., Yamaguchi, S. & Zhang, Y. Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev* **28**, 2103-2119, doi:10.1101/gad.248005.114 (2014).
69 Dawlaty, M. M. *et al*. Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Developmental cell* **29**, 102-111, doi:10.1016/j.devcel.2014.03.003 (2014).
70 Bird, A. P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-213, doi:10.1038/321209a0 (1986).
71 Farthing, C. R. *et al*. Global mapping of DNA methylation in mouse promoters reveals epigenetic reprogramming of pluripotency genes. *Plos Genet* **4**, e1000116, doi:10.1371/journal.pgen.1000116 (2008).
72 De Smet, C., Lurquin, C., Lethe, B., Martelange, V. & Boon, T. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* **19**, 7327-7335 (1999).
73 Kelly, T. K. *et al*. H2A.Z maintenance during mitosis reveals nucleosome shifting on mitotically silenced genes. *Molecular cell* **39**, 901-911, doi:10.1016/j.molcel.2010.08.026 (2010).
74 Cedar, H. & Bergman, Y. Linking DNA methylation and histone modification: patterns and paradigms. *Nature Reviews Genetics* **10**, 295-304, doi:10.1038/nrg2540 (2009).
75 Newell-Price, J., Clark, A. J. L. & King, P. DNA methylation and silencing of gene expression. *Trends Endocrin Met* **11**, 142-148, doi:Doi 10.1016/S1043-2760(00)00248-4 (2000).
76 Mohn, F. *et al*. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Molecular cell* **30**, 755-766, doi:10.1016/j.molcel.2008.05.007 (2008).
77 Lock, L. F., Takagi, N. & Martin, G. R. Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* **48**, 39-46, doi:10.1016/0092-8674(87)90353-9 (1987).
78 Sharp, A. J. *et al*. DNA methylation profiles of human active and inactive X chromosomes. *Genome Res* **21**, 1592-1600, doi:10.1101/gr.112680.110 (2011).
79 Beard, C., Li, E. & Jaenisch, R. Loss of methylation activates Xist in somatic but not in embryonic cells. *Genes Dev* **9**, 2325-2334, doi:10.1101/gad.9.19.2325 (1995).
80 Ball, M. P. *et al*. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells (vol 27, pg 361, 2009). *Nat Biotechnol* **27**, 485-485, doi:10.1038/nbt0509-485b (2009).

81 Yang, X. J. *et al*. Gene Body Methylation Can Alter Gene Expression and Is a Therapeutic Target in Cancer. *Cancer Cell* **26**, 577-590, doi:10.1016/j.ccr.2014.07.028 (2014).
82 Rauch, T. A., Wu, X. W., Zhong, X., Riggs, A. D. & Pfeifer, G. P. A human B cell methylome at 100-base pair resolution. *P Natl Acad Sci USA* **106**, 671-678, doi:10.1073/pnas.0812399106 (2009).
83 Arechederra, M. *et al*. Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer (vol 9, 3164, 2018). *Nat Commun* **9**, doi:ARTN 397610.1038/s41467-018-06482-w (2018).
84 Maunakea, A. K. *et al*. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* **466**, 253-257, doi:10.1038/nature09165 (2010).
85 Illingworth, R. S. *et al*. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *Plos Genet* **6**, e1001134, doi:10.1371/journal.pgen.1001134 (2010).
86 Nguyen, C. T., Gonzales, F. A. & Jones, P. A. Altered chromatin structure associated with methylation-induced gene silencing in cancer cells: correlation of accessibility, methylation, MeCP2 binding and acetylation. *Nucleic Acids Res* **29**, 4598-4606, doi:DOI 10.1093/nar/29.22.4598 (2001).
87 Chen, Y. B., Chen, Q., McEachin, R. C., Cavalcoli, J. D. & Yu, X. C. H2A.B facilitates transcription elongation at methylated CpG loci. *Genome research* **24**, 570-579, doi:10.1101/gr.156877.113 (2014).
88 Neri, F. *et al*. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72-+, doi:10.1038/nature21373 (2017).
89 Aporntewan, C. *et al*. Hypomethylation of Intragenic LINE-1 Represses Transcription in Cancer Cells through AGO2. *PloS one* **6**, doi:ARTN e1793410.1371/journal.pone.0017934 (2011).
90 Tufarelli, C. *et al*. Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* **34**, 157-165, doi:DOI 10.1038/ng1157 (2003).
91 Lister, R. *et al*. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-322, doi:10.1038/nature08514 (2009).
92 Fleischer, T. *et al*. DNA methylation at enhancers identifies distinct breast cancer lineages. *Nat Commun* **8**, doi:Artn 137910.1038/S41467-017-00510-X (2017).
93 Aran, D. & Hellman, A. DNA Methylation of Transcriptional Enhancers and Cancer Predisposition. *Cell* **154**, 11-13, doi:10.1016/j.cell.2013.06.018 (2013).
94 Bock, C. *et al*. DNA Methylation Dynamics during In Vivo Differentiation of Blood and Skin Stem Cells. *Mol Cell* **47**, 633-647, doi:10.1016/j.molcel.2012.06.019 (2012).

95 Brunk, B. P., Goldhamer, D. J. & Emerson, C. P. Regulated demethylation of the myoD distal enhancer during skeletal myogenesis. *Dev Biol* **177**, 490-503, doi:DOI 10.1006/dbio.1996.0180 (1996).
96 Schmidl, C. *et al*. Lineage-specific DNA methylation in T cells correlates with histone methylation and enhancer activity. *Genome research* **19**, 1165-1174, doi:10.1101/gr.091470.109 (2009).
97 Tagoh, H. *et al*. Dynamic reorganization of chromatin structure and selective DNA demethylation prior to stable enhancer complex formation during differentiation of primary hematopoietic cells in vitro. *Blood* **103**, 2950-2955, doi:10.1182/blood-2003-09-3323 (2004).
98 Thomassin, H., Flavin, M., Espinas, M. L. & Grange, T. Glucocorticoid-induced DNA demethylation and gene memory during development. *EMBO J* **20**, 1974-1983, doi:10.1093/emboj/20.8.1974 (2001).
99 Pai, S. *et al*. Differential methylation of enhancer at IGF2 is associated with abnormal dopamine synthesis in major psychosis. *Nat Commun* **10**, doi:UNSP 204610.1038/s41467-019-09786-7 (2019).
100 Polansky, J. K. *et al*. DNA methylation controls Foxp3 gene expression. *Eur J Immunol* **38**, 1654-1663, doi:10.1002/eji.200838105 (2008).
101 Kim, H. P. & Leonard, W. J. CREB/ATF-dependent T cell receptor-induced FoxP3 gene expression: a role for DNA methylation. *J Exp Med* **204**, 1543-1551, doi:10.1084/jem.20070109 (2007).
102 Floess, S. *et al*. Epigenetic control of the foxp3 locus in regulatory T cells. *PLoS Biol* **5**, e38, doi:10.1371/journal.pbio.0050038 (2007).
103 King, A. D. *et al*. Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. *Cell Rep* **17**, 289-302, doi:10.1016/j.celrep.2016.08.083 (2016).
104 Calo, E. & Wysocka, J. Modification of Enhancer Chromatin: What, How, and Why? *Mol Cell* **49**, 825-837, doi:10.1016/j.molcel.2013.01.038 (2013).
105 Teissandier, A. & Bourc'his, D. Gene body DNA methylation conspires with H3K36me3 to preclude aberrant transcription. *Embo Journal* **36**, 1471-1473, doi:10.15252/embj.201796812 (2017).
106 Ooi, S. K. T. *et al*. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448**, 714-U713, doi:10.1038/nature05987 (2007).
107 Thomson, J. P. *et al*. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082-U1162, doi:10.1038/nature08924 (2010).

108 Clouaire, T. *et al*. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Gene Dev* **26**, 1714-1728, doi:10.1101/gad.194209.112 (2012).
109 Fuks, F. *et al*. The Methyl-CpG-binding protein MeCP2 links DNA methylation to histone methylation. *J Biol Chem* **278**, 4035-4040, doi:10.1074/jbc.M210256200 (2003).
110 Creyghton, M. P. *et al*. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *P Natl Acad Sci USA* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).
111 El-Osta, A. & Wolffe, A. P. DNA methylation and histone deacetylation in the control of gene expression: Basic biochemistry to human development and disease. *Gene Expression* **9**, 63-75 (2000).
112 Jones, P. L. *et al*. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-191, doi:Doi 10.1038/561 (1998).
113 Arzenani, M. K. *et al*. Genomic DNA Hypomethylation by Histone Deacetylase Inhibition Implicates DNMT1 Nuclear Dynamics. *Mol Cell Biol* **31**, 4119-4128, doi:10.1128/Mcb.01304-10 (2011).
114 Cervoni, N. & Szyf, M. Demethylase activity is directed by histone acetylation. *J Biol Chem* **276**, 40778-40787, doi:DOI 10.1074/jbc.M103921200 (2001).
115 Sharifi-Zarchi, A. *et al*. DNA methylation regulates discrimination of enhancers from promoters through a H3K4me1-H3K4me3 seesaw mechanism. *Bmc Genomics* **18**, doi:ARTN 96410.1186/s12864-017-4353-7 (2017).
116 Tate, P. H. & Bird, A. P. Effects of DNA Methylation on DNA-Binding Proteins and Gene-Expression. *Curr Opin Genet Dev* **3**, 226-231, doi:Doi 10.1016/0959-437x(93)90027-M (1993).
117 Tippin, D. B. & Sundaralingam, M. Nine polymorphic crystal structures of d(CCGGGCCCGG), d(CCGGGCCm(5)CGG), d(Cm(5)CGGGCCm(5)CGG) and d(CCGGGCC(Br)(5)CGG) in three different conformations: Effects of spermine binding and methylation on the bending and condensation of A-DNA. *J Mol Biol* **267**, 1171-1185, doi:DOI 10.1006/jmbi.1997.0945 (1997).
118 Zhu, H., Wang, G. H. & Qian, J. Transcription factors as readers and effectors of DNA methylation. *Nature Reviews Genetics* **17**, 551-565, doi:10.1038/nrg.2016.83 (2016).
119 Fatemi, M. & Wade, P. A. MBD family proteins: reading the epigenetic code. *J Cell Sci* **119**, 3033-3037, doi:10.1242/jcs.03099 (2006).
120 Ludwig, A. K. *et al*. Binding of MBD proteins to DNA blocks Tet1 function thereby modulating transcriptional noise. *Nucleic Acids Res* **45**, 2438-2457, doi:10.1093/nar/gkw1197 (2017).

121 Yin, Y. M. *et al.* Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, doi:ARTN eaaj223910.1126/science.aaj2239 (2017).
122 You, J. S. *et al.* OCT4 establishes and maintains nucleosome-depleted regions that provide additional layers of epigenetic regulation of its target genes. *P Natl Acad Sci USA* **108**, 14497-14502, doi:10.1073/pnas.1111309108 (2011).
123 Domcke, S. *et al.* Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* **528**, 575-+, doi:10.1038/nature16462 (2015).
124 Hu, S. *et al.* DNA methylation presents distinct binding sites for human transcription factors. *Elife* **2**, e00726, doi:10.7554/eLife.00726 (2013).
125 Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes Dev* **28**, 2679-2692, doi:10.1101/gad.253443.114 (2014).
126 Serandour, A. A. *et al.* Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res* **21**, 555-565, doi:10.1101/gr.111534.110 (2011).
127 Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470-484, doi:10.1016/j.cell.2010.10.012 (2010).
128 Palacios, D., Summerbell, D., Rigby, P. W. J. & Boyes, J. Interplay between DNA Methylation and Transcription Factor Availability: Implications for Developmental Activation of the Mouse Myogenin Gene. *Mol Cell Biol* **30**, 3805-3815, doi:10.1128/Mcb.00050-10 (2010).
129 Marchal, C. & Miotto, B. Emerging concept in DNA methylation: role of transcription factors in shaping DNA methylation patterns. *J Cell Physiol* **230**, 743-751, doi:10.1002/jcp.24836 (2015).
130 Pott, S. & Lieb, J. D. What are super-enhancers? *Nat Genet* **47**, 8-12, doi:10.1038/ng.3167 (2015).
131 Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-947, doi:10.1016/j.cell.2013.09.053 (2013).
132 Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-319, doi:10.1016/j.cell.2013.03.035 (2013).
133 Yin, J. W. & Wang, G. The Mediator complex: a master coordinator of transcription and cell lineage development. *Development* **141**, 977-987, doi:10.1242/dev.098392 (2014).
134 Bourbon, H. M. *et al.* A unified nomenclature for protein subunits of Mediator complexes linking transcriptional regulators to RNA polymerase II. *Mol Cell* **14**, 553-557, doi:DOI 10.1016/j.molcel.2004.05.011 (2004).

135 Cho, W. K. *et al*. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412-415, doi:10.1126/science.aar4199 (2018).
136 Sabari, B. R. *et al*. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, doi:ARTN eaar395810.1126/science.aar3958 (2018).
137 Shrinivas, K. *et al*. Enhancer Features that Drive Formation of Transcriptional Condensates. *Mol Cell* **75**, 549-+, doi:10.1016/j.molcel.2019.07.009 (2019).
138 Tsutsui, T. *et al*. Mediator complex recruits epigenetic regulators via its two cyclin-dependent kinase subunits to repress transcription of immune response genes. *J Biol Chem* **288**, 20955-20965, doi:10.1074/jbc.M113.486746 (2013).
139 Bourguet, P. *et al*. A role for MED14 and UVH6 in heterochromatin transcription upon destabilization of silencing. *Life Sci Alliance* **1**, e201800197, doi:10.26508/lsa.201800197 (2018).
140 Guo, H. *et al*. Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. *Nature protocols* **10**, 645-659, doi:10.1038/nprot.2015.039 (2015).
141 Smallwood, S. A. *et al*. Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. *Nature methods* **11**, 817-820, doi:10.1038/nmeth.3035 (2014).
142 Guo, F. *et al*. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell research* **27**, 967-988, doi:10.1038/cr.2017.82 (2017).
143 Singer, Z. S. *et al*. Dynamic heterogeneity and DNA methylation in embryonic stem cells. *Molecular cell* **55**, 319-331, doi:10.1016/j.molcel.2014.06.029 (2014).
144 Stelzer, Y. & Jaenisch, R. Monitoring Dynamics of DNA Methylation at Single-Cell Resolution during Development and Disease. *Cold Spring Harbor symposia on quantitative biology* **80**, 199-206, doi:10.1101/sqb.2015.80.027334 (2015).
145 Stelzer, Y., Shivalila, C. S., Soldner, F., Markoulaki, S. & Jaenisch, R. Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell* **163**, 218-229, doi:10.1016/j.cell.2015.08.046 (2015).
146 Stelzer, Y. *et al*. Parent-of-Origin DNA Methylation Dynamics during Mouse Development. *Cell Rep* **16**, 3167-3180, doi:10.1016/j.celrep.2016.08.066 (2016).
147 Ferguson-Smith, A. C. Genomic imprinting: the emergence of an epigenetic paradigm. *Nat Rev Genet* **12**, 565-575, doi:10.1038/nrg3032 (2011).

Chapter 2. Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs.

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

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

[1]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA

[2]Biology Department, Massachusetts Institute of Technology, Cambridge, MA 02142, USA

[3]Department of Molecular Cell Biology, Weizmann Institute of Science, 76100 Rehovot, Israel

[4]Leiden Institute of Physics, Leiden University, 2300 RA Leiden, the Netherlands

[5]Lead Contact

Published as:

Song Y, van den Berg PR, Markoulaki S, Soldner F, Dall'Agnese A, Henninger JE, Drotar J, Rosenau N, Cohen MA, Young RA, Semrau S, Stelzer Y, Jaenisch R. Dynamic Enhancer DNA Methylation as Basis for Transcriptional and Cellular Heterogeneity of ESCs. *Mol Cell*. 2019 Aug 15. pii: S1097-2765(19)30502-7

Author contribution:

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.

Summary

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

Introduction

Tissue-specific differential methylation regions (T-DMRs) have been found to strongly associate with low CpG density and inter-genic enhancers (Ehrlich et al., 2016; Fleischer et al., 2017; Izzi et al., 2016; Jones, 2012; Rinaldi et al., 2016), and the vast majority of cell-type specific DNA methylation changes occur at distal regulatory elements (Luo et al., 2018; Stadler et al., 2011). Whole-genome bisulfite sequencing (WGBS) data indicate a low but detectable level of DNA methylation at T-DMRs overlapping active enhancers (Elliott et al., 2015; Heyn et al., 2016; Hon et al., 2013; Jiang et al., 2015; King et al., 2016; Shull et al., 2016; Stadler et al., 2011). Recent single-cell WGBS (scWGBS) data from mouse embryonic stem cells (ESCs) and the early mouse embryo suggest that the variable low-to-intermediate DNA methylation levels

found at enhancer regions in bulk-cell measurements are largely due to averaging signals across cells with heterogeneous methylation states (Cheow et al., 2015; Guo et al., 2013, 2015, 2017; Rulands et al., 2018; Smallwood et al., 2014). However, due to the static snapshot view of sequencing-based methods, it has been difficult to define the basis, regulation, and functional impact of DNA methylation heterogeneity on gene expression and cellular states.

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

Our recently developed Reporter of Genome Methylation (RGM) allows tracing of locus-specific DNA methylation based on the on-and-off of a fluorescent signal in single cells in real time, and has been shown to faithfully reflect the endogenous DNA methylation states at multiple genomic loci (Stelzer et al., 2015, 2016). This system allows for robustly tracking locus-specific DNA methylation at enhancer regions and for functionally dissecting the hierarchy of epigenetic events that regulate enhancer activity and cellular states, overcoming the challenges faced by bulk measurements or

sequence-based methods. We utilized this system at two pluripotency super-enhancers (SEs), Sox2 and Mir290 SEs, in ESCs. Both SEs overlap with ESC-specific DMRs, which display consistently low levels of methylation, indicating potential heterogeneity (Kobayashi et al., 2012; Leung et al., 2014; Rulands et al., 2018; Seisenberger et al., 2012; Stadler et al., 2011). We targeted RGMs to both alleles of the two SEs in F1 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.

Results

DNA methylation at the Sox2 and Mir290 SEs is heterogeneous at the allelic level

Sox2 and Mir290 SEs reside on chromosome 3 and 7, respectively. Both SEs overlap with T-DMRs, which are hypo-methylated in ESCs but become de novo methylated upon differentiation (Stelzer et al., 2015). The T-DMR of the Sox2 SE is located about 100 kb upstream of the Sox2 gene, whereas the Mir290 SE, consisting of hypo-methylated DMR constituents interspersed by small hyper-methylated regions, is proximal to the Mir290-295 cluster (Figure S1A). WGBS of ESCs indicates that the Sox2

and Mir290 SE DMRs have overall DNA methylation levels higher than that of hypo-methylated promoters of highly expressed genes in ESCs, such as Gapdh and Oct4, but lower than that of imprinting control regions or retroelements, which are monoallelically and hyper-methylated, respectively (Figure S1B) (Kobayashi et al., 2012; Leung et al., 2014; Seisenberger et al., 2012; Stadler et al., 2011). This low-to-intermediate level of methylation at both SEs in bulk cell WGBS suggests that they are hypermethylated in a small population of cells. Re-analysis of published scWGBS data (Smallwood et al., 2014) revealed that the T-DMRs of both SEs belong to the 5% regions with the most variable DNA methylation level compared to other regions of chromosome 7 or chromosome 3 (Figure S1C), further supporting the presence of rare cells with hypermethylated SE DMRs.

Consistent with published scWGBS studies reporting heterogeneity in the wild-type genome (Guo et al., 2013, 2015, Guo et al., 2017; Hu et al., 2016; Singer et al., 2014; Smallwood et al., 2014), we previously observed methylation heterogeneity in ESCs with the endogenous Nanog tagged with eGFP and RGM-tdTomato reporter inserted mono-allelically into the Sox2 or Mir290 SE DMRs (Stelzer et al., 2015). The heterogeneity at these two specific loci was manifested by the bi-modal distribution of RGM activity in Nanog positive (Nanog+) pluripotent cells as seen in fluorescence-activated cell sorting (FACS) (Figure 1A). Sorting cells based on florescence intensity, followed by bisulfite PCR (BS-PCR) and sequencing, validated that RGM methylation strictly correlates with the endogenous methylation in both regions (Figure 1A). Analyzing the Sox2 SE revealed that 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 (Figure S1D). The rare methylated alleles were also detected at the Mir290 SE by high-throughput sequencing of BS-PCR

amplicons from the wild-type allele. Figure 1B shows that, comparing to Dnmt3a/b double-knockout cells (described later in Figure S3A), we found methylation at the Mir290 SE in non-manipulated wild-type ESCs as well as on the untargeted allele in the Nanog+RGM+ ESCs. These results indicate that SE DNA methylation heterogeneity is created by allele-specific hypermethylation in rare ESC populations independent of RGM targeting. To track DNA methylation heterogeneity on each allele, we targeted the Mir290 and the Sox2 SE independently in 129xCastaneus F1 hybrid ESCs with allele-specific RGM reporters and generated two cell lines, Sox2-129SE-RGM-tdTomato/Sox2-CASTSE-RGM-eGFP (abbreviated below as SOX2-SE-TG) and Mir290-129SE-RGM-tdTomato/Mir290-CASTSE-RGM-eGFP (abbreviated below as MIR290-SE-TG) (Figures 1C and S1E) allowing to visualize the SE locus-specific DNA methylation state at allelic and single-cell resolution. These cell lines also enabled dissection of allelic functional output of SE methylation states by distinguishing the two alleles based on the abundance of 129 or CAST allele-specific SNPs at both the DNA and the mRNA level.

The initial FACS analysis detected a small fraction of single-positive (T+G–, T–G+) as well as double-negative (T–G–) cells in both cell lines, though the majority of cells were double-positive (T+G+) (Figure 1D), consistent with the heterogeneity reported in scWGBS data by others (Figure S1C) and in our BS-PCR analysis on both targeted and wild-type alleles (Figures 1A, 1B, and S1D). To confirm that the RGM reporter activity faithfully reflected the allele-specific endogenous DNA methylation state, we sorted the four populations and performed allele-specific BS-PCR followed by Sanger sequencing of the DMRs upstream of the reporters. Figure 1E shows that the reporter activities on both alleles were consistent with the DNA methylation levels of the genomic SE regions and the inserted RGMs in all sorted populations. Quantitative

pyro-sequencing further confirmed that T+G+ and T–G– populations represent two extreme methylation states of the intrinsic epigenetic heterogeneity at both SEs (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 (Figure S1G).

Dynamic allele-specific SE DNA methylation is regulated by de novo methylation and passive demethylation during cell proliferation.

To gain insights into the origin of DNA methylation heterogeneity, we FACS sorted equal numbers of the four populations from both reporter cell lines and monitored the RGM activity upon passaging in serum + LIF medium (Figure 2A). Figure 2B (serum + LIF) and Figure S2A show that the SE DNA methylation states in the four sorted populations were not stable but highly dynamic with each allele independently switching the RGM on-and-off over the course of only a few days. This indicates that the observed SE DNA methylation heterogeneity is a result of fast dynamic and reversible switching of allelic DNA methylation states. When sorted cells were passaged and cultured in "2i" (GSKi and MAPKi) medium, the kinetics of the transitions between different methylation states was significantly altered with slowed de novo methylation for both SEs and an initial acceleration of demethylation at the Mir290 SE (Figures 2B and 2C; 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 (Choi et al., 2017; Leitch et al., 2013; Sim et al., 2017; von Meyenn et al., 2016; Yagi et al., 2017).

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

DNA demethylation can occur either passively in rapidly dividing cells, caused by inhibition of DNMT1 or by active removal of the methyl group mediated by Tet enzymes and base excision repair (BER) pathways (Wu and Zhang, 2017). To assess whether demethylation of the SEs involved active or passive mechanisms, we analyzed whether DNA demethylation would be affected in cells upon delaying cell-cycle progression using thymidine block. In all three populations carrying at least one methylated allele, the kinetics of demethylation upon thymidine block was significantly decreased upon 3 days in culture (Figures 3B and 3C). This suggests that cell proliferation-driven passive demethylation is responsible for SE demethylation. To confirm this observation genetically, we transfected 129SE-RGM-tdTomato T–G+ cells with Cas9 and single-guide RNAs (sgRNAs) against genes encoding the maintenance

enzymes DNMT1/UHRF1, which upon downregulation would lead to genome-wide passive dilution of methylation. In addition, we used sgRNAs against enzymes implicated in mediating active demethylation (Tets/Tdg/Aid). Figure 3D shows the predicted outcomes of 129SE-RGM-tdTomato allele demethylation (changes of the fraction of T+G+ cells) after disruption of these genes. When Dnmt1 or Uhrf1 were disrupted, the 129SE-RGM-tdTomato allele became demethylated in a substantial fraction of cells (Figure 3E). 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 129SE-RGM-tdTomato allele in single clones harboring homozygous Tdg and Aid frameshift mutations (Figure S3D) with that of wild-type cells and observed no difference (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 (Dawlaty et al., 2011, 2013, 2014) (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 unsynchronized 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).

TF binding at SEs 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 (Feldmann et al., 2013; Maurano et al., 2015; Yin et al., 2017; Zhu et al., 2016). The Sox2 SE harbors multiple enrichment sites for the master TFs OCT4 and NANOG in ESCs (Hnisz et al., 2013) (Figure 4A, top). We deleted enrichment sites for the two TFs (peak 1 for NANOG and 2 for both NANOG and OCT4) at the Sox2 SE DMR on either the 129SE-RGM-tdTomato or the CASTSE-RGM-eGFP allele using sgRNAs against allele-specific SNPs (Figure 4A, bottom) and generated ESC clones harboring allele-specific peak deletions (ΔPeak 1-CAST, ΔPeak 2-CAST, and ΔPeak 2-129 clones; Figure S4A). We sorted the T–G– and T+G+ populations from these clones and monitored the 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 129SE-RGM-tdTomato and the CASTSE-RGM-eGFP allele exhibited a faster de novo methylation rate after deletion of its TF enrichment sites as compared to the intact wild-type allele (Figure 4D, top), indicating higher susceptibility to de novo methylation upon loss of TF binding. Similarly, the allele that had its TF enrichment site deleted showed a slower demethylation rate than the wild-type allele, indicating less resistance to maintenance methylation upon loss of TF binding (Figure 4D, bottom). To confirm that the observed RGM activity changes correspond to changes in DNA methylation, we performed BS-PCR followed by Sanger sequencing on sorted cells from ΔPeak 1-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

(Figure S4B). The TF binding effect on methylation dynamics was seen not only in cloned cells but also in sorted T+G+ cell population transfected with allele-specific sgRNAs against TF enrichment sites (Figure 4E). Consistent with the single-cell clone analyses, the allele with TF enrichment site deletion showed a faster de novo methylation rates than the wild-type allele that was not targeted by the sgRNAs (Figure 4F).

DNA methylation decreases MED1 association with SEs, enhancer-promoter H3K27ac, and in-cis transcription of the target genes

We investigated whether the rapid changes in SE DNA methylation would dynamically affect target gene transcription. Promoter DNA methylation has long been associated with stable silencing of gene expression (Deaton and Bird, 2011; Dor and Cedar, 2018; Schübeler, 2015; Smith and Meissner, 2013); in comparison, enhancer methylation's role in transcription is less well characterized. The Mediator complex has been shown to be dynamically involved in phase-separated condensates concentrating at SEs for transcription of key cell-identity genes (Sabari et al., 2018). Since SE DNA methylation is dynamically changing, we investigated whether different allelic methylation states affect association of MED1 condensates with the Mir290 SE. We performed DNA FISH at the Mir290 SE locus and MED1 immunostaining on sorted cell populations. Figure 5A and Figure S5A show that MED1 was not enriched at the methylated Mir290 SE as T–G– cell populations did not have DNA FISH foci that overlapped with MED1 enrichment as compared to cells in which at least one Mir290 SE was unmethylated. Since the Mediator complex interacts with both the SE and the promoter (Whyte et al., 2013), a loss of MED1 enrichment upon SE DNA methylation may affect promoter activity as well. We therefore performed H3K27ac chromatin immunoprecipitation sequencing (ChIP-seq) as a proxy epigenetic mark defining active enhancers and

promoters on four sorted populations from both reporter cell lines. H3K27ac was significantly reduced at both methylated SE regions, as measured by total (Figures 5B and 5C, Sox2 SE and Mir290 SE boxes; Figure S5B, enhancer panels) as well as allele-specific H3K27ac enrichment (Figure S5C, enhancer panels). As expected, a decrease in H3K27ac was also observed at promoters residing on the same chromosome with the methylated SE (Figures 5B and 5C, Sox2 and Mir290 boxes, and Figures S5B and S5C, promoter panels) but not at adjacent regions (Figure S5B, adjacent regions panels). This demonstrates that SE methylation affects the promoter H3K27ac level, likely through a loss of enhancer-promoter communication.

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

This indicates that the DNA methylation state of the two SEs switches independently of each other.

To determine whether SE methylation has a causal role in suppressing enhancer-promoter H3K27ac and transcription, we transfected Cas9-sgRNAs targeting Dnmt1 and Uhrf1 and removed DNA methylation in sorted T–G– MIR290-SE-TG cells to induce rapid passive demethylation (Figure 6A). Figure 6B shows that cells deficient for Dnmt1 or Uhrf1 displayed significantly faster demethylation resulting in a higher proportion of T+G+ cells as compared to the control. Both acetylation of H3K27 at the SE (Figure 6C) and Mir290-295 expression (Figure 6D) were significantly increased upon Dnmt1/Uhrf1 disruptions, as measured by ChIP-qPCR and qRT-PCR from the same cultures, respectively. This suggests that change in DNA methylation directly regulates SE function and transcription in cis.

Since correlating abundance in RNA allele-specific SNPs with allele-specific RGM activities allows distinguishing direct targets regulated in cis by the SE methylation status versus expression changes caused by secondary effects, we searched additional genomic targets on the same chromosomes that are directly regulated by SE methylation by allele-specific RNA sequencing (RNA-seq) analysis on sorted populations. We quantified allele-specific expression of genes on chromosome 3 (for MIR290-SE-TG) and chromosome 7 (for SOX2-SE-TG) in single positive cells and calculated the ratio between expressions from the allele with an unmethylated SE over that of the other allele with a methylated SE. We plotted this ratio of each gene calculated in T–G+ cells as the x-axis value and the ratio calculated in T+G– cells as the y-axis value (Figure S6A). As expected, Mir290-295 and Sox2 both appeared in the upper right corner as they were in cis directly suppressed by allelic SE methylation.

Surprisingly, two antisense transcripts relative to Sox2 and Mir290-295, Ecm1 and AU018091, respectively, were oppositely regulated by allele-specific Sox2 or Mir290 SE methylation: SE hyper-methylation strongly correlated with upregulations of both anti-sense transcripts, whereas SE hypo-methylation correlated with inhibition (Figures S6B and S6C). This result shows that direct transcriptional targets of SE methylation are highly specific with possibly opposite effects on some cis-regulated genes. Though the detailed mechanism of such regulation remains to be elucidated, Ecm1 was upregulated in Sox2 SE deletion cells (Gagnon et al., 2014).

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

Sox2 and Mir290 SE methylation heterogeneities have different biological impacts on ESC state

Culture in "2i" medium has been shown to only allow naïve pluripotent cells to proliferate (Nichols and Smith, 2009). Long-term culture of MIR290-SE-TG and SOX2-SE-TG cells in "2i" after passaging from serum + LIF media, though favoring T+G+ population decreased but did not abolish heterogeneity completely (Figure S6D). The persistence of all four populations in both reporter 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 (Calabrese et al., 2007; Hnisz et al., 2013; Jaenisch and Young, 2008; Thomson et al., 2011), raising the possibility that allelic transcriptional heterogeneity caused by SE methylation heterogeneity may lead to co-existing heterogeneous cellular states of ESCs. In "2i"

media, SOX2-SE-TG T–G– cells exhibited significantly impaired colony-forming ability and proliferation (Figures 6F and 6G). However, under the same condition, the heterogeneous DNA methylation at the Mir290 SE did not lead to any obvious changes of ESCs, despite the slight colony formation disadvantage of MIR290-SE-TG T–G– cells (Figures 6F and 6G). We further explored the functional differences among populations in vivo by injecting sorted cells to form teratomas. Surprisingly, despite the significant growth disadvantage of SOX2-SE-TG T–G– population, they were able to contribute to all three germ layers in teratoma formation assays with no obvious contribution bias towards any germ layer compared to SOX2-SE-TG T+G+, MIR290-SE-TG T+G+, and T–G– cells (Figure S6E). This indicates that ESCs with biallelic methylation at the Sox2 SE are still pluripotent. However, when examined at the molecular level, these cells were distinct from other populations in principal-component analysis on difference in the 5% most highly variably expressed genes (Figure 6H) and 17,000 uniquely distinct H3K27ac enrichment peaks in ChIP-seq (Figure S6F). GO analysis on RNA-seq revealed that the SOX2-SE-TG T–G– population preferentially expressed genes in differentiation-related pathways as compared to the MIR290-SE-TG T–G– population (Figure S7A). The epigenetic and transcriptional differences of SOX2-SE-TG T–G– cells indicate that these cells downregulate Sox2 expression and are prone to differentiate but not as yet committed to a certain fate. Our results are consistent with the notion that pluripotent ESC are heterogeneous as reflected by the dynamic allelic DNA methylation of key pluripotency SEs.

DNA methylation is dynamic at both SEs in blastocysts, while exhibiting spatial-temporal differences in pre-implantation embryos

In vivo, both Sox2 and Mir290-295 are expressed in preimplantation embryos. As reported previously Sox2 expression increases between the morula and the blastocyst

stage (Mistri et al., 2018) and Mir290-295 expression significantly upregulates at the 4-cell stage (Medeiros et al., 2011). To investigate changes in DNA methylation of the two SEs at single-cell and allelic resolution, we generated transgenic mice homozygous for the 129SE-RGM-tdTomato allele or the CASTSE-RGM-eGFP allele and obtained 2–4 cell embryos carrying one 129SE-RGM-tdTomato allele and one CASTSE-RGM-eGFP allele by mating animals homozygous for RGM-eGFP or RGM-tdTomato (Figure 7A). The two 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) (Nichols and Smith, 2009; Paikari et al., 2017; Wicklow et al., 2014). Heterogeneous SE DNA methylation was consistent with the established spatial expression pattern of the two genes in blastocysts (Figure 7C). We further investigated whether the observed methylation heterogeneity was due to dynamic allelic methylation state switching in vivo. We sorted the four populations from SOX2-SE-TG and MIR290-SE-TG ESCs, injected each population into 8-cell stage wild-type CD1-IGS host embryos, and cultured embryos for 2 days to monitor de novo methylation or demethylation at single-cell resolution (Figure S7B). A long-term membrane bound dye (Cy5) was used to track the injected cells (Figures 7D; Figure S7C). Figure 7D (SOX2-SE-TG cells) and Figure S7C (MIR290-SE-TG cells) show that, at the blastocyst stage, injected T–G– cells demethylated the SE as they turned on the RGMs on either or both alleles and became single positive or T+G+ cells (T–G– columns, white arrows). Demethylation also was observed in injected single-positive cells as the originally methylated allele at injection became unmethylated and cells became T+G+ (T+G– and T–G+ columns, white arrows). Similarly, dynamic de novo methylation was observed in vivo, as injected T+G+ cells shut down RGM activities on either or both

alleles (T+G+ columns, yellow arrows) and single-positive cells became T–G– cells (T–G+ and T–G+ columns, yellow arrows).

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

Discussion

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

developmentally identical states as for the Mir290 SE. (5) Finally, we show that dynamic DNA methylation is not only seen in cultured ESCs but also in preimplantation embryos. Allele-specific RGM reporters targeted to the endogenous Sox2 and the Mir290 SEs allowed us to trace DNA methylation both in vitro and in vivo. Detailed analyses showed that the low levels of DNA methylation of the Sox2 and the Mir290 SE are due to the presence of a small fraction of cells with hypermethylated SE alleles. The methylation heterogeneity in these cells results from highly dynamic and reversible switching between allelic DNA methylation states. Because the RGM reporter allowed isolation of cells with defined allele-specific SE DNA methylation states, we were able to demonstrate that dynamic changes in SE DNA methylation are tightly anti-correlated in cis with enhancer-promoter H3K27ac levels. This is likely due to disruption of enhancer-promoter interactions consistent with the Mediator complex condensates showing decreased association at the methylated Mir290 SE. The Mediator complex and its unit MED1 have been shown previously to form condensates with liquid-like properties, which allows dynamic interactions with TFs and the transcription apparatus (Cho et al., 2018; Sabari et al., 2018). Our study shows that DNA methylation can affect these transcriptional condensates. Given the dynamic state switching of allelic SE DNA methylation as well as the dynamic nature of MED1 condensate formation, it is highly likely that one process mediates the other. We also show that SE DNA methylation can have opposing effects on transcription of different genes located on the same chromosome: the direct target genes Sox2 and Mir290-295 were repressed, while the antisense genes Ecm1 and AU018091 were activated by SE methylation. By removing DNA methylation at the Mir290 SE through Dnmt1/Uhrf1 deletion, we showed that changes in SE DNA methylation is a dynamic process actively regulating its transcriptional activity. By enabling sorting for a particular epigenetic state and combined with allelic expression analyses, we demonstrate that dynamic DNA

methylation serves as an epigenetic basis for allelic heterogeneity in gene expression and that dynamic DNA methylation at SEs is a likely mechanism for dynamic random monoallelic transcription seen in mammalian cells (Deng et al., 2014; Reinius and Sandberg, 2015). However, it warrants further exploration to establish the causal link between allelic epigenetic and transcriptional heterogeneity in vivo.

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

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

heterogeneous tissues and complex biological processes, such as development and diseases (Heyn and Esteller, 2012; Robertson, 2005).

Methods

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 [69,148,149]. 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 Table S1.

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 (Table S5).

Confocal imaging of live pre-implantation embryos

2-cell embryos were obtained from mating SOX2-SE-TT or MIR290-SE-TT females hormone primed step-wise with PMS and hCG to SOX2-SE-GG or MIR290-SE-GG males, respectively, or the opposite mating strategy (SOX2-SE-GG or MIR290-SE-GG females to SOX2-SE-TT or MIR290-SE-TT males, respectively). 2-cell embryos were flushed out from the oval-duct by M2 media with BSA (CytoSpring # m2113) 48hrs post mating. The embryos were then cultured in 25-50ul KSOM media droplets (CytoSpring # KO102) covered by mineral oil in a 37°C 5% 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.

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 [145]. Pyro-seq of all bisulfite converted genomic DNA samples were performed with PyroMark Q48 Autoprep (QIAGEN) according to the manufacturer's instructions. Primers used for BS-PCR and pyro-sequencing are listed in Table S2.

Retinoic acid differentiation

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

Double thymidine block

10-20k cells/per well were plated onto 12-well plates after sorting with media containing 2.5mM thymidine for 12hrs. Blocking was released by washing twice with

PBS and culturing in serum + LIF mouse ES media for 9hrs. Cells were then again blocked with 2.5mM thymidine for 14hrs and FACS analyses were done 6hrs post release.

qRT-PCR and TaqMan assays

Total mRNA was extracted from ESCs using Direct-zol RNA Miniprep (Zymo Research, Cat#R2050) after pre-plating for elimination of MEF feeders, treated with DNase A defined amount of mRNA reverse-transcribed into cDNA using SuperScript III First-Strand Synthesis SuperMix (Life Technologies, Cat#18080400) using random hexamers. Total expression of transcripts were quantified by qRT-PCR using Fast SYBR Green Master Mix (Life Technologies), and allele-specific transcripts are quantified by TaqMan Assay customized probes (Sigma, Table S3) targeting *Sox2* and *Mir290-295* pri-mRNA SNPs. Tukey's multiple comparison (****P<0.0001, **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 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 [150], 2ug of anti-H3K27ac antibody (abcam ab4729) was used for precipitation. Eluted DNA was quantified using real-time qPCR with Fast SYBR Green Master Mix. Each ChIP-qPCR was repeated 3 times. Enrichment was calculated using as percentage of input. Statistical differences between samples are calculated with two-way ANOVA (alpha = 0.05), followed by Tukey's multiple comparison (****P<0.0001, **P<0.01). Primers used for detecting positive and negative control sequences, and SE targets are listed in Table S4.

H3K27ac ChIP-Seq and analysis

ChIP samples of 4 same-culture-sorted populations from SOX2-SE-TG and MIR290-SE-TG, respectively, were validated for positive and negative targets using qPCR. Libraries

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

RNA-Seq and analysis

For each reporter cell line, 2 independently targeted clones are independently sorted twice, generating 2 biological replica x 2 experimental replica = 4 replica in total. Stranded mRNA libraries were prepared using KAPA HyperPrep (SOX2-SE-TG) and TrueSeq Stranded PolyA prep (MIR290-SE-TG). mRNA libraries were sequenced on Illumina HiSeq 2500. Allele-specific RNA expression was quantified with a custom pipeline. In short, raw fastq files are aligned to a consensus genome using STAR (v.2.5.3.a). The reference transcriptome includes the Mir290-295 pri-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 [136]. 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 l = 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 l x l 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 [141], and number of methylated counts was assumed to be a binomial random variable.

Acknowledgement

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 HD045022, 1U19AI131135, 5R01MH104610, and 1R01GM114864.

Figure 1. DNA Methylation at the Sox2 and Mir290 SEs Is Heterogeneous at the Allelic Level

(A) Left, DNA methylation heterogeneity at both the Sox2 and the Mir290 SE in v6.5-Nanog-eGFP ESC where the RGM-tdTomato reporter was mono-allelically targeted. Right, BS-PCR followed by sequencing of the Sox2 SE in different populations of the bimodal distribution.

(B) Average methylation percentage and standard errors were quantified from high-throughput sequencing of BS-PCR amplicons of the Mir290 SE wild-type alleles in Dnmt3a/b double-knockout ESCs, in Nanog+RGM+ ESCs and in untargeted wild-type ESCs. BS-PCRs were amplified allele-specifically as illustrated from potential epigenetic states indicated above. Standard error was estimated assuming number of methylated counts as a binomial random variable.

(C) Targeting strategy for generating SOX2-SE-TG and MIR290-SE-TG ESCs using CRISPR/Cas9 and targeting vectors. Methylation tracks from (Stadler et al., 2011) were used as the genome reference with blue bars highlighting the DMRs of the two SEs. Red tracks, 129 allele; green tracks, CAST allele.

(D) FACS analysis of CASTx129 F1 ESC clones targeted with allele-specific RGMs at either the Mir290 or the Sox2 SE.

(E) Allele-specific BS-PCR of the SEs with RGM (Snprn-tdTomato or Snprn-eGFP) in single PCR amplicons followed by Sanger sequencing in sorted cells from both SOX2-SE-TG and MIR290-SE-TG.

See also Figure S1.

Figure 2. 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 Figure S2.

Figure 3. 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 double-knockout (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 Figure S3.

Figure 4. TF Binding at SEs Promotes Demethylation and Inhibits De Novo Methylation

(A) Top, schematic representation of TF enrichment sites (based on the ChIP-seq data of NANOG [pink track, peak 1 and 2] and OCT4 [blue track, peak 2].

ENCODE: ENCSR779CZG and ENCSR392DGA) relative to the RGM targeted site (orange). Bottom, allele-specific deletions of individual peaks after overlapping

NANOG (N) and OCT4 (O) ChIP tracks. Red: 129SE-RGM-tdTomato allele, green: CASTSE-RGM-eGFP allele. Scissors illustrate sgRNA targeting sites. ChIP-seq value is presented as fold-change-over-control.

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

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

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

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

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

See also Figure S4.

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

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

(B) 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 CASTSE-RGM-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 Figure S5.

Figure 6. DNA Methylation Directly Suppresses SE Activity and Affects ESC State

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

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

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

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

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

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

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

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

See also Figures S5–S7.

Figure 7. DNA Methylation Is Dynamic at Both SEs in Blastocysts while Exhibiting Spatial-Temporal Differences in Pre-implantation Embryos

(A) Mating scheme for generating SOX2-SE-TG and MIR290-SE-TG mice and heterozygous pre-implantation embryos genetically carry 129SE-RGM-tdTomato and CAST-SE-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 103 low magnification, 403 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 Figure S7.

References

Calabrese, J.M., Seila, A.C., Yeo, G.W., and Sharp, P.A. (2007). RNA sequence analysis defines Dicer's role in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA 104, 18097–18102.

Cheow, L.F., Quake, S.R., Burkholder, W.F., and Messerschmidt, D.M. (2015). Multiplexed locus-specific analysis of DNA methylation in single cells. Nat. Protoc. 10, 619–631.

Cho, W.K., Spille, J.H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I.I. (2018). Mediator and RNA polymerase II clusters associate in transcription dependent condensates. Science 361, 412–415.

Choi, J., Huebner, A.J., Clement, K., Walsh, R.M., Savol, A., Lin, K., Gu, H., Di Stefano, B., Brumbaugh, J., Kim, S.Y., et al. (2017). Prolonged Mek1/2 suppression impairs the developmental potential of embryonic stem cells. Nature 548, 219–223.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C., and Jaenisch, R. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell 9, 166–175.

Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev. Cell 24, 310–323.

Dawlaty, M.M., Breiling, A., Le, T., Barrasa, M.I., Raddatz, G., Gao, Q., Powell, B.E., Cheng, A.W., Faull, K.F., Lyko, F., and Jaenisch, R. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. Dev. Cell 29, 102–111.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. 25, 1010–1022.

Deng, Q., Ramsko¨ ld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNAseq reveals dynamic, random monoallelic gene expression in mammalian cells. Science 343, 193–196.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21.

Dor, Y., and Cedar, H. (2018). Principles of DNA methylation and their implications for biology and medicine. Lancet 392, 777–786.

Ehrlich, K.C., Paterson, H.L., Lacey, M., and Ehrlich, M. (2016). DNA Hypomethylation in Intragenic and Intergenic Enhancer Chromatin of Muscle-Specific Genes Usually Correlates with their Expression. Yale J. Biol. Med. 89, 441–455.

Elliott, G., Hong, C., Xing, X., Zhou, X., Li, D., Coarfa, C., Bell, R.J.A., Maire, C.L., Ligon, K.L., Sigaroudinia, M., et al. (2015). Intermediate DNA methylation is a conserved signature of genome regulation. Nat. Commun. 6. Published online February 18, 2015.

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., and Sch€ubeler, D. (2013). Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. PLoS Genet. 9, e1003994.

Fleischer, T., Tekpli, X., Mathelier, A., Wang, S., Nebdal, D., Dhakal, H.P., Sahlberg, K.K., Schlichting, E., Børresen-Dale, A.L., Borgen, E., et al.; Oslo Breast Cancer Research Consortium (OSBREAC) (2017). DNA methylation at enhancers identifies distinct breast cancer lineages. Nat. Commun. 8, 1379.

Gagnon, J.A., Valen, E., Thyme, S.B., Huang, P., Akhmetova, L., Pauli, A., Montague, T.G., Zimmerman, S., Richter, C., and Schier, A.F. (2014). Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS ONE 9, e98186.

Guo, H., Zhu, P., Wu, X., Li, X., Wen, L., and Tang, F. (2013). Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. Genome Res. 23, 2126–2135.

Guo, H., Zhu, P., Guo, F., Li, X., Wu, X., Fan, X., Wen, L., and Tang, F. (2015). Profiling DNA methylome landscapes of mammalian cells with single-cell reduced-representation bisulfite sequencing. Nat. Protoc. 10, 645–659.

Guo, F., Li, L., Li, J., Wu, X., Hu, B., Zhu, P., Wen, L., and Tang, F. (2017). Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. Cell Res. 27, 967–988.

Heyn, H., and Esteller, M. (2012). DNA methylation profiling in the clinic: applications and challenges. Nat. Rev. Genet. 13, 679–692.

Heyn, H., Vidal, E., Ferreira, H.J., Vizoso, M., Sayols, S., Gomez, A., Moran, S., Boque-Sastre, R., Guil, S., Martinez-Cardus, A., et al. (2016). Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. Genome Biol. 17, 11.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-Andre´ , V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell 155, 934–947.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat. Genet. 45, 1198–1206.

Hu, Y., Huang, K., An, Q., Du, G., Hu, G., Xue, J., Zhu, X., Wang, C.Y., Xue, Z., and Fan, G. (2016). Simultaneous profiling of transcriptome and DNA methylome from a single cell. Genome Biol. 17, 88.

Izzi, B., Pistoni, M., Cludts, K., Akkor, P., Lambrechts, D., Verfaillie, C., Verhamme, P., Freson, K., and Hoylaerts, M.F. (2016). Allele-specific DNA methylation reinforces PEAR1 enhancer activity. Blood 128, 1003–1012.

Jaenisch, R., and Young, R. (2008). Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. Cell 132, 567–582.

Jiang, R., Jones, M.J., Chen, E., Neumann, S.M., Fraser, H.B., Miller, G.E., and Kobor, M.S. (2015). Discordance of DNA methylation variance between two accessible human tissues. Sci. Rep. 5, 8257.

Jin, B., Li, Y., and Robertson, K.D. (2011). DNA methylation: superior or subordinate

in the epigenetic hierarchy? Genes Cancer 2, 607–617.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. 13, 484–492.

King, A.D., Huang, K., Rubbi, L., Liu, S., Wang, C.Y., Wang, Y., Pellegrini, M., and Fan, G. (2016). Reversible Regulation of Promoter and Enhancer Histone Landscape by DNA Methylation in Mouse Embryonic Stem Cells. Cell Rep. 17, 289–302.

Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., et al. (2012). Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. PLoS Genet. 8, e1002440.

Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., et al.; Roadmap Epigenomics Consortium (2015). Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. 20, 311–316.

Lesch, B.J., Dokshin, G.A., Young, R.A., McCarrey, J.R., and Page, D.C. (2013). A set of genes critical to development is epigenetically poised in mouse germ cells from fetal stages through completion of meiosis. Proc. Natl. Acad. Sci. USA 110, 16061–16066.

Leung, D., Du, T., Wagner, U., Xie, W., Lee, A.Y., Goyal, P., Li, Y., Szulwach, K.E., Jin, P., Lorincz, M.C., and Ren, B. (2014). Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. Proc. Natl. Acad. Sci. USA 111, 6690–6695.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics 25, 2078–2079.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550.

Luo, C., Hajkova, P., and Ecker, J.R. (2018). Dynamic DNA methylation: In the right place at the right time. Science 361, 1336–1340. Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. 12, 1184–1195.

Medeiros, L.A., Dennis, L.M., Gill, M.E., Houbaviy, H., Markoulaki, S., Fu, D., White, A.C., Kirak, O., Sharp, P.A., Page, D.C., and Jaenisch, R. (2011). Mir-290-295 deficiency in mice results in partially penetrant embryonic lethality and germ cell defects. Proc. Natl. Acad. Sci. USA 108, 14163–14168.

Mi, H., Lazareva-Ulitsky, B., Loo, R., Kejariwal, A., Vandergriff, J., Rabkin, S., Guo, N., Muruganujan, A., Doremieux, O., Campbell, M.J., et al. (2005). The PANTHER database of protein families, subfamilies, functions and pathways. Nucleic Acids Res. 33, D284–D288.

Mistri, T.K., Arindrarto, W., Ng, W.P., Wang, C., Lim, L.H., Sun, L., Chambers, I., Wohland, T., and Robson, P. (2018). Dynamic changes in Sox2 spatio-temporal expression promote the second cell fate decision through Fgf4/Fgfr2 signaling in preimplantation mouse embryos. Biochem. J. 475, 1075–1089.

Nichols, J., and Smith, A. (2009). Naive and primed pluripotent states. Cell Stem Cell 4, 487–492.

Paikari, A., D Belair, C., Saw, D., and Blelloch, R. (2017). The eutheria-specific miR-290 cluster modulates placental growth and maternal-fetal transport. Development 144, 3731–3743.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nature Methods 5, 877–879.

Ramirez, F., Dundar, F., Diehl, S., Gruning, B.A., and Manke, T. (2014). deepTools: a flexible platform for exploring deep-sequencing data. Nucleic Acids Res. 42, W187–W191.

Reinius, B., and Sandberg, R. (2015). Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. Nat. Rev. Genet. 16, 653–664.

Rinaldi, L., Datta, D., Serrat, J., Morey, L., Solanas, G., Avgustinova, A., Blanco, E., Pons, J.I., Matallanas, D., Von Kriegsheim, A., et al. (2016). Dnmt3a and Dnmt3b Associate with Enhancers to Regulate Human Epidermal Stem Cell Homeostasis. Cell Stem Cell 19, 491–501.

Robertson, K.D. (2005). DNA methylation and human disease. Nat. Rev. Genet. 6, 597–610.

Rulands, S., Lee, H.J., Clark, S.J., Angermueller, C., Smallwood, S.A., Krueger, F., Mohammed, H., Dean, W., Nichols, J., Rugg-Gunn, P., et al. (2018). Genome-Scale Oscillations in DNA Methylation during Exit from Pluripotency. Cell Syst. 7, 63–76.

Sabari, B.R., Dall'Agnese, A., Boija, A., Klein, I.A., Coffey, E.L., Shrinivas, K., Abraham, B.J., Hannett, N.M., Zamudio, A.V., Manteiga, J.C., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science 361, 361.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods 9, 671–675. Schbubeler, D. (2015). Function and information content of DNA methylation. Nature 517, 321–326.

Schultz, M.D., He, Y., Whitaker, J.W., Hariharan, M., Mukamel, E.A., Leung, D., Rajagopal, N., Nery, J.R., Urich, M.A., Chen, H., et al. (2015). Human body epigenome maps reveal noncanonical DNA methylation variation. Nature 523, 212–216.

Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. Mol. Cell 48, 849–862.

Shull, A.Y., Luo, J.F., Pei, L.R., Lee, E.J., Liu, J.M., Choi, J., Awan, F.T., and Shi, H.D. (2016). DNA Hypomethylation within B-Cell Enhancers and Super Enhancers Reveal a Dependency on Immune and Metabolic Mechanisms in Chronic Lymphocytic Leukemia. Blood 128, 1049.

Sim, Y.J., Kim, M.S., Nayfeh, A., Yun, Y.J., Kim, S.J., Park, K.T., Kim, C.H., and Kim, K.S. (2017). 2i Maintains a Naive Ground State in ESCs through Two Distinct Epigenetic Mechanisms. Stem Cell Reports 8, 1312–1328.

Singer, Z.S., Yong, J., Tischler, J., Hackett, J.A., Altinok, A., Surani, M.A., Cai, L., and Elowitz, M.B. (2014). Dynamic heterogeneity and DNA methylation in embryonic stem cells. Mol. Cell 55, 319–331.

Smallwood, S.A., Lee, H.J., Angermueller, C., Krueger, F., Saadeh, H., Peat, J., Andrews, S.R., Stegle, O., Reik, W., and Kelsey, G. (2014). Single-cell genome-wide bisulfite sequencing for assessing epigenetic heterogeneity. Nat. Methods 11, 817–820.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. 14, 204–220.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Scholer, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480, 490–495.

Stelzer, Y., Shivalila, C.S., Soldner, F., Markoulaki, S., and Jaenisch, R. (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. Cell 163, 218–229.

Stelzer, Y., Wu, H., Song, Y., Shivalila, C.S., Markoulaki, S., and Jaenisch, R. (2016). Parent-of-Origin DNA Methylation Dynamics during Mouse Development. Cell Rep. 16, 3167–3180.

Suzuki, H.I., Young, R.A., and Sharp, P.A. (2017). Super-Enhancer-Mediated RNA Processing Revealed by Integrative MicroRNA Network Analysis. Cell 168, 1000–1014.

Thomas, P.D., Campbell, M.J., Kejariwal, A., Mi, H., Karlak, B., Daverman, R., Diemer, K., Muruganujan, A., and Narechania, A. (2003). PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 13, 2129–2141.

Thomson, M., Liu, S.J., Zou, L.N., Smith, Z., Meissner, A., and Ramanathan, S. (2011). Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. Cell 145, 875–889.

von Meyenn, F., Iurlaro, M., Habibi, E., Liu, N.Q., Salehzadeh-Yazdi, A., Santos, F., Petrini, E., Milagre, I., Yu, M., Xie, Z., et al. (2016). Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. Mol. Cell 62, 983.

Weigel, C., Veldwijk, M.R., Oakes, C.C., Seibold, P., Slynko, A., Liesenfeld, D.B., Rabionet, M., Hanke, S.A., Wenz, F., Sperk, E., et al. (2016). Epigenetic regulation of

diacylglycerol kinase alpha promotes radiation induced fibrosis. Nat. Commun. 7, 10893.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319.

Wicklow, E., Blij, S., Frum, T., Hirate, Y., Lang, R.A., Sasaki, H., and Ralston, A. (2014). HIPPO pathway members restrict SOX2 to the inner cell mass where it promotes ICM fates in the mouse blastocyst. PLoS Genet. 10, e1004618.

Wilson, S., and Filipp, F.V. (2018). A network of epigenomic and transcriptional cooperation encompassing an epigenomic master regulator in cancer. NPJ Syst. Biol. Appl. 4, 24.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. 18, 517–534.

Yagi, M., Kishigami, S., Tanaka, A., Semi, K., Mizutani, E., Wakayama, S., Wakayama, T., Yamamoto, T., and Yamada, Y. (2017). Derivation of ground state female ES cells maintaining gamete-derived DNA methylation. Nature 548, 224–227.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science 356, 356.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

Zhu, H., Wang, G., and Qian, J. (2016). Transcription factors as readers and effectors of DNA methylation. Nat. Rev. Genet. 17, 551–565.

Ziller, M.J., Gu, H., Muller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013).

Charting a dynamic DNA methylation landscape of the human genome. Nature 500, 477–481.

Supplementary Information

Supplementary Figure S1. DNA methylation at the Sox2 and Mir290 SE is heterogeneous at the allelic level, Related to Figure 1. (A) The genomic views of *Sox2* and *Mir290* SEs and the distances to their primary target genes. Hypo-methylated DMRs were highlighted with blue bars. Methylation tracks are from [5], OSN (OCT4, SOX2, and NANOG) and MED1 ChIP-Seq data are from Young Lab. (B) Comparison of SE low-methylated DMRs to other genomic loci across four mouse ESC WGBS libraries [5,17,35,36]. For each locus, the distributions of methylation levels of CpGs within the DMR were plotted as individual columns, with each column representing data from one published library. Tukey's multiple comparisons test was performed in comparing column-means (****P<0.0001, **P<0.01, *P<0.05). (C) Top, violin plots of DNA methylation level at *Sox2* SE and *Mir290* SE from Smallwood et al. scBS-seq data. Each column represents a window over 3000bp. Bottom, coefficient of variation (CV) of CpG DNA methylation levels across single cells. Red bars indicate the 95th percentile for the CVs along the whole chromosome. Arrows indicate where RGM were targeted in all experiments later in this study. (D) BS-PCR followed by Sanger sequencing. ESCs targeted with RGM-tdTomato at one allele was sorted based on the reporter activity. BS-PCR was performed using allele-specific primers in List 2 as illustrated. Cell illustration: RGM-tdTomato (red box), methylated allele (filled), unmethylated allele (open); wild-type allele is indicated by absence of the RGM (red box on the targeted allele). (E) Southern blot genotyping for the selected MIR290-SE-TG and SOX2-SE-TG clones with a single integration on each allele. The HindIII digested fragments containing the tdTomato reporter from the 129$^{SE\text{-}RGM\text{-}tdTomato}$ allele was detected by the tdTomato probe and fragments from the CAST$^{SE\text{-}RGM\text{-}eGFP}$ allele was detected by the eGFP probes. (F) Heatmap of percentage of methylation from pyro-sequencing of the *Sox2* (5 consecutive CpGs) and *Mir290* (2 groups of 3 consecutive CpGs, separated by

the dotted line) SE DMR in DKO ESCs, sorted double positive (T+G+) and double negative (T-G-) cells, and untargeted wild-type ESCs. Primers used for amplifying pyro-sequencing template are listed in primer list 2. (G) DNA methylation dynamics in the SOX2-SE-TG and the MIR290-SE-TG cells after 4 days of differentiation in the N2B27 medium supplemented with 0.25μM retinoic-acid (RA).

Supplementary Figure S2. SE DNA methylation heterogeneity is created by dynamic switching of methylation states, Related to Figure 2. (A) FACS analysis of allelic DNA methylation dynamics at the *Mir290* and the *Sox2* SE in the serum + LIF medium over 13 days, and (B) in "2i" medium over 11 days.

Supplementary Figure S3. The dynamics of SE DNA methylation is driven by de novo methylation and passive demethylation during cell proliferation, Related to Figure 3. (A) Top, RT-PCR on mRNA from *Dnmt3a/b* single- or double catalytic domain knockout clones. Clones showed similar size of cDNA PCR products (labeled in bold) as the wild-type cells are due to one or more alleles having small frame-shift genomic indels unresolved by agarose gel electrophoresis. Bottom, Topo-cloned genomic PCR sequences from clones labeled in bold above to confirm frame-shift disruption. PAMs of sgRNA are highlighted in red, sgRNA in blue (*Dnmt3a* reference) or yellow (*Dnmt3b* reference), and exons are in orange. (B) DNA methylation dynamics of one SE allele in the wild type, *Dnmt3a* or *Dnmt3b* single, and DKO ESCs over 14 days. All clones were derived from the same parental clone for each SE case. Cells were sorted at day 0 based on RGM "on" (red) or "off" (green), and changes in each population were monitored at each passaging for over 14 days. The results of the sorted RGM "on" and "off" population pairs originally from the same culture are overlapped in each FACS histogram plot. (C) Quantification of DNA methylation level of the endogenous *Sox2* or *Mir290* SE locus in ESCs with indicated genotypes (*Tet* enzymes knockouts or *Dnmts* knockouts) using pyro-sequencing. Each bar represents the methylation level of a CpG. Wild-type clones (WT) and *Tet* enzymes knockout clones were from the same parental ES clone as described in[69,148,149]. (D) Genomic PCR followed by topo-cloning and sequencing of *Tdg* and *Aid* KO clones. Yellow highlights: exons; pink highlights: introns; CCT (*Tdg* clones) and AGG (*Aid* clone) are used as PAM sequences. All deletions are validated as frame-shifting. (E) Top, experiment design and predicted changes if the mechanism of demethylation is mainly active; bottom, DNA methylation dynamics of T-G+ cells transitioning to double-positive T+G+ cells from *Tdg* or *Aid* knockout (KO) single SOX2-SE-TG clones as compared to their respective sister wild-type clones (WT) for the *Sox2* SE.

Supplementary Figure S4. Transcription factor binding at SEs promotes demethylation and inhibits de novo methylation, Related to Figure 4. (A) Southern blot genotyping of the representative single cell clones harboring TF enrichment site deletions on either allele. Southern blot was done using HindIII and BrsGI double digestion followed by probing with tdTomato (T) or eGFP (G) coding sequences. (B) Allele-specific BS-PCR followed by Sanger sequencing at the *Sox2* SE in sorted T+G- cells form ΔPeak 1-CAST and T-G+ cells from ΔPeak 2-129 clones. Primers used for PCR amplification are the same as the ones used in Figure 1F.

Supplementary Figure S5. DNA methylation decreases MED1 association at SE, enhancer-promoter H3K27ac, and in-cis transcription of the target gene, Related to Figure 5. (A) Representative images of DNA FISH (Mir290 SE) and MED1 IF co-localization in MIR290-SE-TG T+G+, T-G+ and T+G- populations. (B) Quantification of H3K27ac ChIP-seq 100bp-bin FPKM values for SE, promoter, and adjacent regions for 4 sorted populations from MIR290-SE-TG (top) and SOX2-SE-TG (bottom). Genome coordinates are from genome assembly mm10. Data points within 1.5xinterquartile range are represented as median with 25-75th percentile R boxplots. (C) Percentages of CAST-specific H3K27ac ChIP-seq reads out of total reads at the SE and the promoter regions for *Mir290-295* (top) and *Sox2* (bottom). SNP positions with read counts > 3 were accepted for allele-specific read quantification. Genome coordinates are from genome assembly mm10. Data points within 1.5xinterquartile range are represented as median with 25-75th percentile R boxplots. (D) Representative images of smFISH of *Sox2* mRNA in the SOX2-SE-TG cells with heterogeneous DNA methylation. (E) Quantification of total *Sox2* transcripts in double negative (T-G-), single-positive (T+G- or T-G+) and double positive cells.

Supplementary Figure S6. DNA methylation of both SEs regulates in-cis transcription of the target genes and affects ES cell state at the Sox2 SE, Related to Figure 5, 6. (A) Highly specific *in-cis* regulation on *Sox2* (left) and *Mir290-295* (right) expression in single-positive cells. X-axis value represents the ratio of rlog2 of transcripts from the CAST (unmethylated) allele versus the 129 (methylated) allele in the T-G+ population, and y-axis value represents the ratio of rlog2 of transcripts from the 129 (unmethylated) allele versus the CAST (methylated) allele in the T+G- population. The majority of genes were not affected by the SE methylation state in an allele-specific manner, as they appeared in the center of the plot. Genes appeared in the upper left and lower right corners were differentially regulated due to different genetic backgrounds of CAST and 129 alleles. Transcription of *Ecm1* and *AU018091* correlates *in-cis* with presence of DNA methylation at the *Sox2* SE and the *Mir290* SE, respectively. (B) Genomic view of allele-specific FPKM of *Mir290-295* versus *AU018091* and (C) *Sox2* versus *Ecm1* from the respective T+G- and T-G+ populations of MIR290-SE-TG and SOX2-SE-TG cells. (D) DNA methylation dynamics and persistence of allelic heterogeneity of SOX2-SE-TG (top) and MIR290-SE-TG (bottom) cells transitioned from serum + LIF medium and cultured long-term in "2i" media. (E) Contribution to three germ layers by double positive and double negative cells of both MIR290-SE-TG and SOX2-SE-TG in teratoma formation assays. (F) Hierarchical clustering of H3K27ac ChIP-seq peaks 3000bp up- and downstream of peak centers from sorted T+G-, T-G+, T+G+, and T-G- ESCs. Peaks uniquely enriched in the SOX2-SE-TG T-G- population were framed with black bars.

Supplementary Figure S7. DNA methylation is dynamic at both SEs in blastocysts, while exhibiting spatial-temporal differences in pre-implantation embryos, Related to Figure 6, 7. (A) GO analysis of the differentially regulated genes (T+G+ vs. T-G-) in both SOX2-SE-TG and MIR290-SE-TG RNA-seqs. All expressed genes in the respective cell lines were used as the reference backgrounds (FDR < 5%). Listed pathways were ranked top 23. (B) Experimental setup to investigate dynamic SE DNA methylation *in vivo*. Four populations were sorted from SOX2-SE-TG and MIR290-SE-TG were injected into 8-cell stage CD1-IGS host embryos. Embryos were cultured in M2 media and imaged at indicated time points in Figure 7D and S7C. (C) Tracking *Mir290* SE DNA methylation dynamics *in vivo*. Columns are injected population 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 *de novo* methylation, at 2 days post-injection compare to 5hr post-injection. Channels were adjusted for brightness and contrast for optimal visibility.

Table S1. sgRNAs

sgDnmt1-1	CTCGGCTGAGTCGGCTGCAG
sgDnmt1-2	CTAACGTGGCTTCTCTCTGC
sgDnmt1-3	GAAGGCTGGAGTAGGATCCA
sgDnmt3A-1	GTGGGAACAACAACTGCTGC
sgDnmt3A-2	CTGGTGGAATGCACTGCAGA
sgDnmt3B-1	AGCTGAGGATGCCAAGCTGC
sgDnmt3B-1	CCGTTCGACTTGGTGATTGG
sgUhrf1-1	TCCCATCCATAGTTCGAACC
sgUhrf1-2	CAGCAAGTACGCTCCTGCAG
sgTet1-1	ACCACGTCTACTGCAGTCCA

sgTet1-2	TCTTTGTTGGGTACCTTCTC
sgTet2-1	TTCTGAAAAGGACAGGTACC
sgTet2-2	TGGAGGCAGCTGCAAGCTTG
sgTet3	AAAAGCGGGGCTTCTTAAGG
sgTDG-1	GCACCAGGTAGTTCTTACAT
sgTDG-2	TTGTTCTTTCAGCTATTCTC
sgAID-1	GGCGCGCGGTGAAAATCCTC
sgAID-2	TGCGGAGACTGCACCGCGCT
sgAID-3	CGCTGGAGACCGATATGGAC
sgTFBS-Sox2-SE-1(CAST)	TAGTTCTGTGTGTGTGCGCC
sgTFBS-Sox2-SE-2(CAST)	CCTCTTTGGGGGAGGGGTGG
sgTFBS-Sox2-SE-2(129)	TCTTTGGGGGAGGGGTGTGG
sgTFBS-Sox2-SE-3(CAST)	AGGGAAGTTTTTTCAAGACC
sgTFBS-Sox2-SE-3(Both)	TTTAAAAGTTTTTCAAAACA
sgRNA for targeting Sox2-SE 129 allele	CCAGCTTTCCGAGCCAGATG
sgRNA for targeting Sox2-SE CAST allele	CCAGCTTTCAGAGCCAGATG
sgRNA for targeting RGM into miR290-SE	GTATCAGCTCTGAAATCTGC
sgRNA for targeting RGM into Sox2-SE (CAST)	CCAGCTTTCAGAGCCAGATG
sgRNA for targeting RGM into Sox2-SE (129)	CCAGCTTTCCGAGCCAGATG

Table S2. Primers for Bisulfite PCR

Sox2-SE Nested BS-F	GTGGTTGTTGTGTTTAGTATGTGGG
Sox2-SE BS-F	GTAGATGTAGGAATTATTTTTGGTGTTTT
miR290-SE Nested BS-F	GTGATATTGTGTTTTGGGGAGAAAG
miR290-SE BS-F	GTTTTGGGGAGAAAGTTTTGTTATTAAAG

Snrpn Nested BS-R	CCCTTACTCACCATACTAACAAAATCC
Snrpn BS-R	ACAAACCCAACTAACCTTCCTC
tdTomato Nested BS-R	AACGCATAAACTCTTTAATAACCTCCTC
eGFP Nested BS-R	CTCGACCAAAATAAACACCACCCC
Sox2-SE Pyro-PCR-F	TGATTATAGGGAAGTGGGAGAATTTT
Sox2-SE Pyro-PCR-R	ACCTAAATTCCTTAAACCCTATTTCA
Sox2-SE Pyro-seq-F	GGGAGAATTTTTTTTTGGAG
miR290-SE Pyro-PCR-F1	AGGTGGTTTTGTTAGTTTGTTT
miR290-SE Pyro-PCR-R1	AAACCATTCACCACCACATT
miR290-SE Pyro-seq-F1	AGTAAGAGAGAAAAGATTTTATT
miR290-SE Pyro-PCR-F2	ATTTTGGTTGGGTGGAGT
miR290-SE Pyro-PCR-R2	CCCAAACAACTTTCTCTACCTTCA
miR290-SE Pyro-seq-F2	GTGGAGTAGAAGGTTT

Table S3. Primers for qRT-PCR

miR290-pri-RNA-F	ACCTGGCTCCTAGCACAAACA
miR290-pri-RNA-R	GGGCTATTGTAAAGCCAAAAGGTA
Sox2-F	GCGGAGTGGAAACTTTTGTCC
Sox2-R	CGGGAAGCGTGTACTTATCCTT
Gapdh-F	AGGTCGGTGTGAACGGATTTG
Gapdh-R	TGTAGACCATGTAGTTGAGGTCA
eGFP-F	ACGTAAACGGCCACAAGTTC
eGFP-R	AAGTCGTGCTGCTTCATGTG
tdTomato-F	GCTGAAGGGCGAGATCCA
tdTomato-R	GTGGGAGGTGATGTCCAGCTT

TaqMan Sox2 context sequence	GGAGGGGTGCAAAAAGAGGAGAGTA[G/A]AAAAATCT GATAATGCTCAAAAGGA
TaqMan Sox2 F	TGGACTGCGAACTGGAGAAG
TaqMan Sox2 R	CGTTTCGCTGCGGAGATTTTTTT
TaqMan Sox2 reporter – FAM	AGGAGAGTAAGAAAAAT
TaqMan Sox2 reporter – VIC	AGGAGAGTAGGAAAAAT
TaqMan miR290 context sequence	TTTCCTTCAGGTTGGAGTGANTTTT[A/G]GTTCTCCACG TTCTTTCCTCCTGGT
TaqMan miR290 F	GGTTTCCTTCAGGTTGGAGTGA
TaqMan miR290 R	GCGGTCCAGACGTTAAAACATTT
TaqMan miR290 reporter – FAM	ACGTGGAGAACTCAAAA
TaqMan miR290 reporter – VIC	ACGTGGAGAACTTAAAA

Table S4. Primers for ChIP-qPCR

Oct4-distal enhancer (positive target) F	CCCGAAGCTGACTTTGAACTCATG
Oct4-distal enhancer (positive target) R	GCTACAACCTCCCCACACC
Chr19 gene desert (negative target) F	TGGTTCCACACACATCTCCG
Chr19 gene desert (negative target) R	CAGCCGAACCAGGAACTCAT
Sox2-SE H3K27ac ChIP F	CCCGAAGCTGACTTTGAACTCATG
miR290-SE H3K27ac ChIP F	GCAATCTGAAGGCAGAGAAAGCTG
RGM ChIP R	GAATGCTTGAGCATTCCTACTGCG

Table S5. Primers for Genotyping

Sox2-SE-F	AACCCAGGCAGACTGCAAAC
Sox2-SE-R	GAGAGAAGGGAAGAGAGGAGCT
Mir290-SE-F	GCAATCTGAAGGCAGAGAAAGCTG
Mir290-SE-R	AGAATTCCCACTGAGTCTCCTGC
tdTomato-R	AACTCTTTGATGACCTCCTCG
eGFP-R	CTGAACTTGTGGCCGTTTAC
Dnmt3a-RT-F	TTTGATGGGATTGCTACAGGGC
Dnmt3a -RT-R	CACGGGGTTAGACTCAAGAAATCG
Dnmt3b-RT-F	TGGAATTGCAACGGGGTACTTG
Dnmt3b -RT-R	CATCACTGGGTTACATGCCAGG
Dnmt3a-genomic-F	TGTTGGGTCTGTTTGCTCACTG
Dnmt3a-genomic-R	GATGACCACCACACCTCTTTGAAC
Dnmt3b-genomic-F	GGGCAGTTGGAGGTATAATTCAGG
Dnmt3b-genomic-R	GTTCTCAGATTAAGCCACACCCTC
Tdg-genomic-F	ACCTCCCTAATCTCCTGTTACCTG
Tdg-genomic-R	GAGACATTCAAGGGAGACCATG
Aid-genomic-F	CACAACAGCACTGAAGCAGC
Aid-genomic-R	CTTCGTGGTCAGAGTTAGGTCC

Chapter 3. Future Directions

3.1 Developmental Impacts of DNA Methylation Heterogeneity by Epigenetic Lineage Tracing

Chapter 2 has shown that DNA methylation of pluripotency super-enhancers is heterogeneous in the developing embryo, and such epigenetic heterogeneity is responsible for transcriptional level differences between ESCs *in vitro*, especially in the case of *Mir290-295*, where monoallelic methylation reduces its expression by half compared to the biallelically unmethylated cells. *Mir290-295* is broadly expressed in the trophectoderm and the ICM and presents a great degree of allelic DNA methylation heterogeneity. It is therefore interesting to investigate whether cells with different levels of super-enhancer DNA methylation eventually contribute to different cell types in the placenta or the embryo proper. However, since the DNA methylation status is dynamically changing, determining the critical developmental stage where such heterogeneity matters is crucial. Combining conditional genomic recombination with the RGM system would provide a powerful tool for epigenetic lineage tracing, which can help to answer this question[144,151]. Since two alleles of super-enhancers behave independently of each other and pluripotency super-enhancers are initially unmethylated pre-implantation, two inducible recombinase systems are needed. The modified *E. coli* dihydrofolate reductase protein (DHFR) is rapidly degraded in the absence of inducer TMP. Compared to tamoxifen used in the traditional inducible ER-fusion protein system, TMP is easy to administrate and readily diffusible across tissues *in vivo*[152]. Therefore, a *Snrpn-DHFR-Cre* can be targeted on one allele releasing tdTomato expression from a *loxP-stop-loxP-tdTomato* cassette when this allele is unmethylated, and a *Snrpn-DHFR-Flp* on the other allele inducing eGFP expression from an *FRT-stop-FRT-eGFP* cassette when that allele is unmethylated. Both

fluorophore cassettes can be targeted into the either alleles of the safe-harbor *ROSA26* locus (Figure 1). At the blastocyst stage of a developing embryo (E3.5-4.5), when high degree of DNA methylation heterogeneity was observed in the experiments in Chapter 2, TMP can be administrated to the pregnant mouse, and analyzing embryos at different post-implantation stage will likely yield interesting results as to whether early-stage allelic DNA methylation heterogeneity at pluripotency super-enhancers regulates the differentiation potential of cells.

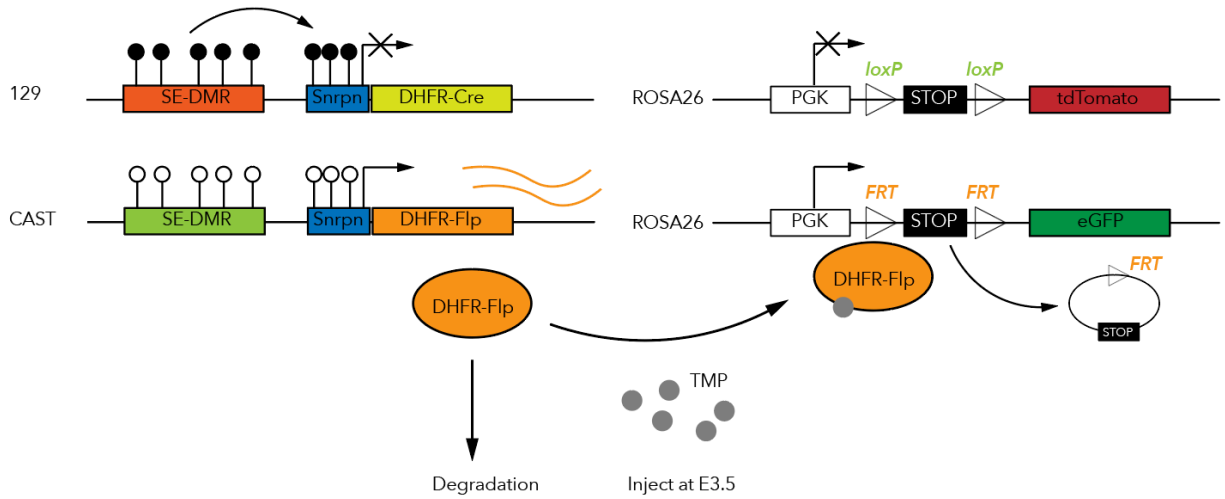

Figure 1. Allele-specific Epigenetic Lineage Tracing.

3.2 Exploring the Upstream Signals Leading to Aberrant DNA Methylation in Diseases

The above-mentioned epigenetic lineage tracing approach can be modified and applied to another interesting biological question: what upstream signals regulate tissue-specific methylation patterns and whether dysregulation of this process during development underlies the initiation of disease, such as cancer or autoimmune diseases?

On one hand, it has been shown that tissue-specific demethylation at regulatory elements such as enhancers is important for cell maturation and function in the hematopoietic, hepatic, neuronal, and muscular systems[14,65,153,154], indicating that failure of demethylation could led to differentiation defects. Many tumor suppressor genes are improperly silenced in cancers which chromatin profile also resembles undifferentiated stem cells[155]. Mutations of *IDH1* and *IDH2* are associated with acute myeloid leukemia (AML), which causes accumulation of D-2-hydroxyglutarate, a competitive inhibitor of TET enzymes. Impaired TET enzyme activities leads to global hypermethylation and impairs differentiation of hematopoietic cells, which causes pro-leukemic transcriptional programs[156]. Gene-body hypermethylation, on the opposite, is linked to increased oncogenic expression in liver cancer[83]. During liver development, insulin signaling is responsible for demethylating ~40% of the hepatocyte-specific DMRs around 15 weeks postnatally[14]. raising the question of whether dysregulation of developmental-related tissue-specific demethylation can be responsible for oncogenic hypermethylation in the cell origin of cancer, and what perturbations in the environment and signal transduction are pathological. In fact, the immune system responses rapidly to environmental signals with DNA methylation changes and cell differentiation, and mis-regulation of these processes can also lead to blood malignancy[157].

On the other hand, abnormal DNA hypomethylation is even more wide-spread in cancer tissues compared to counterpart normal tissues than hypermethylation, especially around repetitive regions, germline-specific and oncogenes. Genome instability and inappropriately high levels of oncogene expression provide two plausible functional explanations of DNA hypomethylation in tumor initiation. However, what are the upstream signals causing such hypomethylation is unclear in most cases[158].

Due to the heterogeneity of somatic tissues, dysregulation of DNA methylation occurring in a small subset of cells could be sufficient for tumorigenesis[159]. DNA methylation is very heterogeneous in tumor tissues and understanding this heterogeneity is crucial for cancer diagnosis and treatment, as drug resistance could be attributed be rare populations with unique epigenetic features in addition to acquired secondary mutations. Epigenetic lineage tracing tool that reports abnormal hypomethylation and hypermethylation can help to understand the initiation of tumors and to isolate cells presenting DNA methylation features associated with the drug-resistance phenotype. This requires careful selection of T-DMRs which function and methylation status throughout normal development is comprehensively understood. For tracing hypomethylation, current version of RGM or an inducible *Snprn-Cre* can be utilized to mark the cells. However, for tracing abnormal hypermethylation, an opposite circuit, which includes *Snprn* driving a transcriptional repressor such as the KRAB domain may be tested.

3.3 Metabolic Regulation of DNA Methylation in Early Embryonic Development

DNA methylation is carried out by DNMTs transferring the methyl-group to the 5th carbon on cytosines using S-adenosylmethionine (SAM) as the methyl donor and forming by-product S-adenosylhomocysteine (SAH). SAH is a potent inhibitor of DNMTs and the balance of SAM/SAH regulates methyltransferase reactions. SAM is generated from methionine and SAH can be metabolized into homocysteine and fuel back to methionine production by coupling with the folate cycle. The methionine cycle (SAM production) and the folate cycle together constitutes one carbon metabolism which has been shown to be regulated by many upstream signals, including dietary nutrients[160]. Mutations in one-carbon metabolism enzymes have been linked to dysregulation of DNA and histone methylation and contribute to diseases, such as liver

cancer[161,162]. DNA methylation during early embryonic development is highly dynamic, and deficiencies of one carbon metabolism could lead to alterations of epigenetic dynamics that has long-lasting effect[163]. Many genome-wide profiles of DNA methylation dynamics during early embryonic development have been generated, however the corresponding dynamics of one carbon metabolism and its influence on cells' future developmental potential are not well characterized *in vivo*. Therefore, the pre-implantation embryos of the transgenic mice described in Chapter 2 could be potentially used as a system to test the effect methyl-nutrient availability on T-DMR dynamics by manipulating the methyl-nutrient levels on the *in-vitro* developing embryos and monitoring the subsequent RGM signal changes. In the Mir290-295 super-enhancer case, it will be interesting to further study how DNA methylation dynamics is influenced by nutrient availability in the developing placenta by transplanting manipulated embryos back to pseudo-pregnant mice.

3.4 DNA Methylation and Transcriptional Condensates

In Chapter 2, we showed that DNA methylation at the *Mir290* super-enhancer affect the formation of MED1 condensates. Since phase-separated condensate contents exchange as well as enhancer DNA methylation are both dynamic, it will be interesting to investigate whether two dynamic processes are interconnected, and if so, which one is upstream of the other. Another interesting question is to investigate whether other intrinsically disordered chromatin-associated proteins that have been shown to form phase-separated condensates also show correlations with the DNA methylation of the *cis*-regulatory elements, or the DNA methylation and demethylation machineries.

Concluding Remarks

DNA methylation was discovered in 1948, not long after the discovery of DNA being the genetic material[164], and the function of DNA methylation in regulating gene expression was proposed in 1975[1,165]. With the advancement of high-throughput sequencing technologies at both bulk and single cell levels, the research community is generating high-resolution methylation maps together with many other "omic" profiles of different cell types, developmental stages, and diseases at an unprecedented pace. Now with these "maps" in hand, it is time to find the bread crumbles and dive deep into the unexplored areas to fully understand how locus- and cell-type specific DNA methylation is regulated by environmental cues, signaling pathways, and its intricate cross-talks with a broad range of epigenetic regulators in the physiologically relevant contexts.

Reference

1 Stelzer, Y. & Jaenisch, R. Monitoring Dynamics of DNA Methylation at Single-Cell Resolution during Development and Disease. Cold Spring Harbor symposia on quantitative biology 80, 199-206, doi:10.1101/sqb.2015.80.027334 (2015).

2 Kretzschmar, K. & Watt, F. M. Lineage tracing. Cell 148, 33-45, doi:10.1016/j.cell.2012.01.002 (2012).

3 Iwamoto, M., Bjorklund, T., Lundberg, C., Kirik, D. & Wandless, T. J. A general chemical method to regulate protein stability in the mammalian central nervous system. Chem Biol 17, 981-988, doi:10.1016/j.chembiol.2010.07.009 (2010).

4 Orlanski, S. et al. Tissue-specific DNA demethylation is required for proper B-cell differentiation and function. P Natl Acad Sci USA 113, 5018-5023, doi:10.1073/pnas.1604365113 (2016).

5 Lucarelli, M., Fuso, A., Strom, R. & Scarpa, S. The dynamics of myogenin site-specific demethylation is strongly correlated with its expression and with muscle differentiation. J Biol Chem 276, 7500-7506, doi:DOI 10.1074/jbc.M008234200 (2001).

6 Reizel, Y. et al. Postnatal DNA demethylation and its role in tissue maturation. Nat Commun 9, 2040, doi:10.1038/s41467-018-04456-6 (2018).

7 Mellen, M., Ayata, P. & Heintz, N. 5-hydroxymethylcytosine accumulation in postmitotic neurons results in functional demethylation of expressed genes. Proc Natl Acad Sci U S A 114, E7812-E7821, doi:10.1073/pnas.1708044114 (2017).

8 Ohm, J. E. et al. A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat Genet 39, 237-242, doi:10.1038/ng1972 (2007).

9 Figueroa, M. E. et al. Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. Cancer Cell 18, 553-567, doi:10.1016/j.ccr.2010.11.015 (2010).

10 Arechederra, M. et al. Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer (vol 9, 3164, 2018). Nat Commun 9, doi:ARTN 397610.1038/s41467-018-06482-w (2018).

11 Suarez-Alvarez, B., Rodriguez, R. M., Fraga, M. F. & Lopez-Larrea, C. DNA methylation: a promising landscape for immune system-related diseases. Trends in Genetics 28, 506-514, doi:10.1016/j.tig.2012.06.005 (2012).

12 Ehrlich, M. DNA methylation in cancer: too much, but also too little. Oncogene 21, 5400-5413, doi:10.1038/sj.onc.1205651 (2002).

13 Kim, M. & Costello, J. DNA methylation: an epigenetic mark of cellular memory. Exp Mol Med 49, e322, doi:10.1038/emm.2017.10 (2017).

14 Ducker, G. S. & Rabinowitz, J. D. One-Carbon Metabolism in Health and Disease. Cell Metab 25, 27-42, doi:10.1016/j.cmet.2016.08.009 (2017).

15 Puszyk, W. M., Trinh, T. L., Chapple, S. J. & Liu, C. Linking metabolism and epigenetic regulation in development of hepatocellular carcinoma. Lab Invest 93, 983-990, doi:10.1038/labinvest.2013.94 (2013).

16 Lu, C. & Thompson, C. B. Metabolic regulation of epigenetics. Cell Metab 16, 9-17, doi:10.1016/j.cmet.2012.06.001 (2012).

17 Xu, J. & Sinclair, K. D. One-carbon metabolism and epigenetic regulation of embryo development. Reprod Fertil Dev 27, 667-676, doi:10.1071/RD14377 (2015).

18 Hotchkiss, R. D. The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175, 315-332 (1948).

19 Holliday, R. & Pugh, J. E. DNA modification mechanisms and gene activity during development. Science 187, 226-232 (1975).

20 Riggs, A. D. X inactivation, differentiation, and DNA methylation. Cytogenet Cell Genet 14, 9-25, doi:10.1159/000130315 (1975).

APPENDIX I. Metastable Epiallele Reporter Mouse Model Development for Studying Environmental Regulation of DNA Methylation

Background

Metastable epialleles (ME) have various levels of expression among genetically identical individuals due to differences in their stochastically established epigenetic states during early development[1,2]. Many MEs found in plants and mammals have been associated with transposable elements[3,4]. The *viable yellow Agouti* (A^{vy}) allele is one of the most extensively-described mammalian MEs, which arose from a spontaneous insertion of a transposable element intracisternal-A-particle (IAP) into the pseudo-exon 1A (PS1A) 100kb upstream of the TSS (exon II) of the *Agouti* gene[5]. In the wild-type *A* allele, *Agouti* expression is mostly controlled hair-cycle dependent promoters located ~80kb downstream at exons 1B, 1C, resulting in black-yellow-black banded hair pigements[6] . Loss-of-function "*a*" allele carried by C57/BL6 mice is created by a large insertion of VL30 element downstream of the exon 1C. A recombination of the VL30 element left only 40bp VL30 LTR behind, resulting in the A^{w} allele that is commonly found in 129 strains[7]. In the A^{w} allele, a ventral-specific promoter located at exon 1A (proximal downstream to PS1A) seems to be constitutively active thus mice carrying an A^{w} allele resembles wild-type agouti mice dorsally, but have a white/yellow cream colored belly[8] (Figure 1).

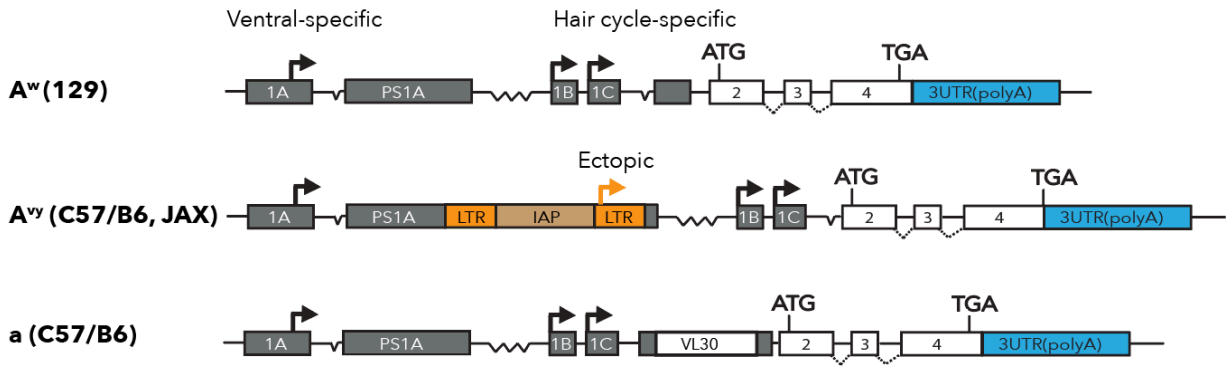

Figure 1. Configurations of Aʷ (mostly derived from 129 backgrounds), Aᵛʸ (on the C57/B6 background from Jackson Laboratory , and "a" alleles of Agouti genes.

Among all the alleles of *Agouti* gene, A^{vy} is a metastable epiallele due to the regulation of IAP. IAP is a class of LTR-based endogenous retrovirus, which are usually silenced by DNA methylation, especially at the LTRs which possess promoter activities. The silencing of IAPs is important for maintaining genome stability and aberrant transcription[9,10]. DNA methylation of the IAP-LTR on the A^{vy} allele is established during development with unknown timing and regulation. When left unmethylated, the LTR became an ectopic promoter capable of driving *Agouti* expression constitutively and non-specifically in somatic tissues, and presented on the skin with yellow hair[11,12]. When left unmethylated, the LTR became an ectopic promoter capable of driving *Agouti* expression constitutively and non-specifically in somatic tissues, and presented on the skin with yellow hair (Figure 2)[11,12]. The DNA methylation status of the A^{vy} metastable epiallele to a certain degree can be inherited trans-generationally via the maternal lineage and is modifiable by gestation methyl-nutrient levels in the developing embryos[13,14]. These phenotypes associated with the A^{vy} mice triggered as much interests as puzzles in the research community over years[12,15-17].

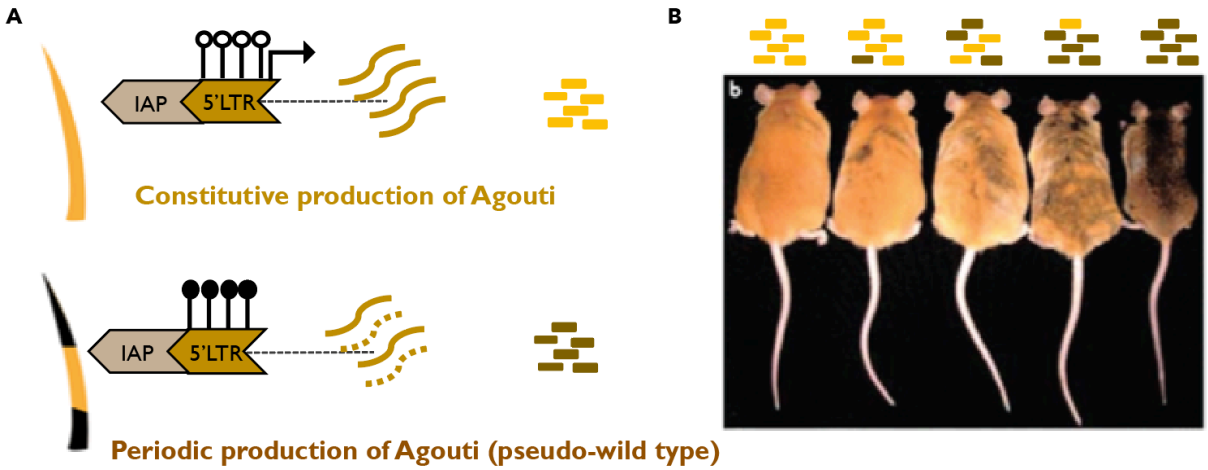

Figure 2. The *A^{vy}* Mouse Model.

(A). IAP-5'LTR at the *A^{vy}* locus serves as a constitutive ectopic promoter for *Agouti* in addition to . The LTR activity is suppressed by DNA methylation. (B) DNA Methylation is established during development stochastically, leading to a random distribution of mice with coat colors ranging from pseudoagouti to extreme yellow and mottled yellow in between.

There are ~1000 copies of IAP in the mouse genome. IAP elements are the only whole class of endogenous retroelements known to be resistant to primordial germ cell genomic demethylation[18], leaving them particularly relevant to transgenerational epigenetic inheritance and metastable epialleles[4,14,19]. The full length IAP is around 7kb and capable of autonomous retro-transposition. However, 30% IAPs have undergone mutations and recombination, leaving a solo-LTR or truncated sequences behind[20,21]. Interestingly, many truncated IAP-LTRs have been shown to lead to ectopic expression of adjacent genes[22]. Therefore, we ask if a solo-LTR is sufficient to create a metastable epiallele at the *Agouti* locus. If so, whether the epigenetic status of the LTR established during development can be trans-generationally inherited, and whether such process is sensitive to environmental perturbation, such as methyl-nutrient supplement? To study these questions, we generated and analyzed the phenotype of

transgenic mice derived from v6.5 ES cells. These mice carry a tdTomato reporter A^w alleles with or without solo-LTR targeted at the endogenous *Agouti* PS1A, where IAP of the original A^{vy} metastable epiallele resides.

Result

Generation of *$A^{w\text{-}LTR}$-tdTomato and A^w-tdTomato* Alleles to Study IAP-LTR Genomic Insertional Effect

To allow real-time analysis of *Agouti* expression and epigenetic regulation of its epiallele at embryonic stage, before coat color is an available phenotype, we engineered an P2A-tdTomato expression reporter into the endogenous locus of *Agouti* in v6.5 (129xC57/B6) ES cells by CRISPR-Cas9 mediated double-strand break followed by homology recombination (Figure 3). The correctly targeted homozygous ES cell clone (A^wT/a^wT) exhibited expected tdTomato reporting on *Agouti* expression during *in vitro* differentiation (Figure 4A). We also generated germ-line transmitted F1s by mating male chimeras to C57/B6 females. Followed with inter-crossing of A^wT/a F1s, we obtained A^wT/A^wT homozygous colonies. E14.5 embryos from A^wT/A^wT xC57/B6 (A^wT/a) mating showed physiological expression pattern of *Agouti* (Figure 4B)as reported earlier using RNA-FISH[6].

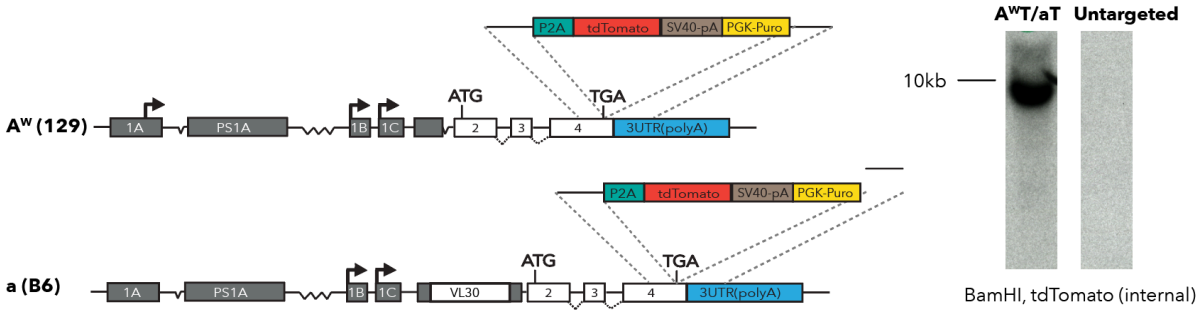

Figure 3. Genomic Targeting of Agouti-tdTomato Expression Reporter in v6.5 ES Cells (A^wT/a^wT).

Left: targeting scheme; right: Southern blot using tdTomato internal probe. Expected size ~10kb.

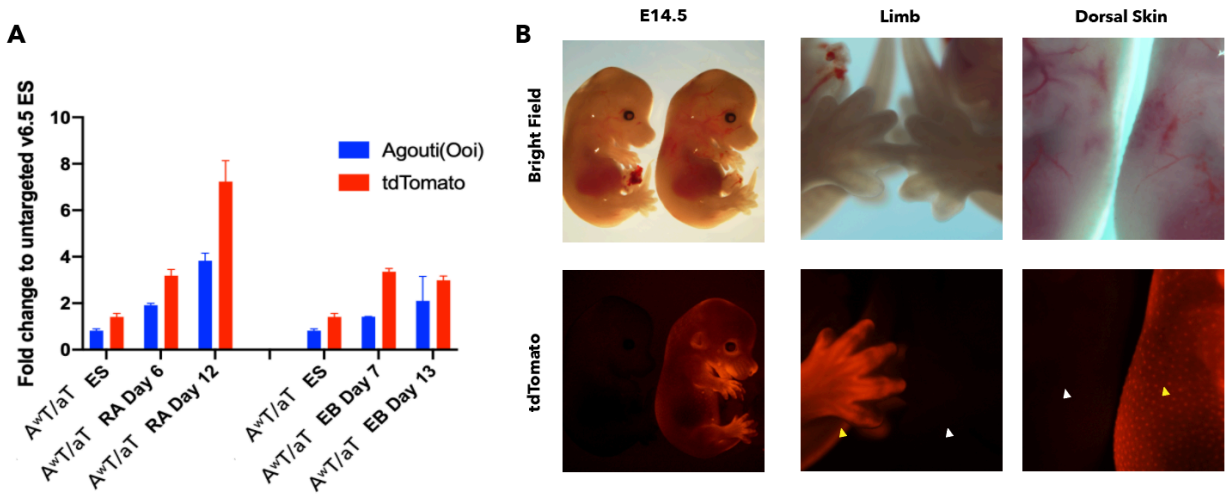

Figure 4. Validation of tdTomato Reporter on The $A^{w}T$ Allele.

(A). *In vitro* embryoid body and retinoic acid differentiation. (B) $A^{w}T/A^{w}T$ x C57/B6 offspring ($A^{w}T/a$) at E14.5.

Next, we targeted the LTR sequence derived from A^{vy} IAP into PS1A (~100kb upstream of the TSS), with exactly the same breakpoint as the A^{vy} IAP insertion and made germ-line transmitted mice from two ES clones F12 and A5 carrying the $A^{w\text{-}LTR}T$ allele (Figure 5).

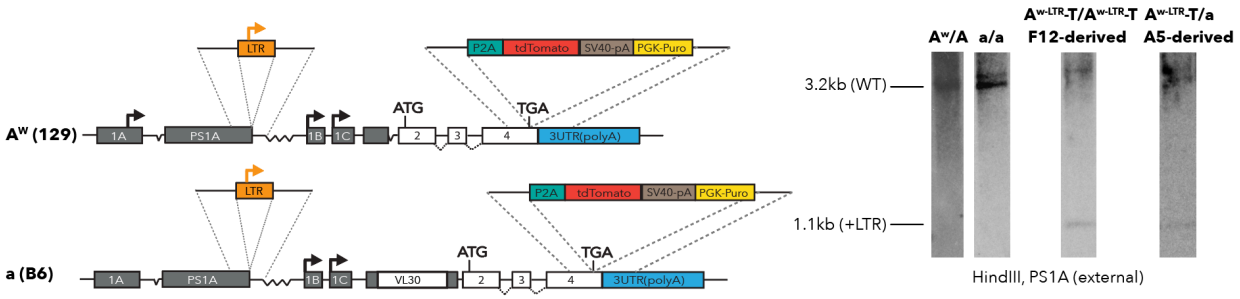

Figure 5. Genomic Targeting of LTR into PS1A. Left: Targeting scheme.

Right: Southern blot of A^{w}/a ES cell (untargetd v6.5), a/a C57/B6 mice tail, and F1 homozygous $A^{w\text{-}LTR}T/ A^{w\text{-}LTR}T$ (F12 derived) and heterozygous $A^{w\text{-}LTR}T/a$ (A5 derived) mice tails.

However, $A^{w\text{-}LTR}T/a$ animals did not recapitulate the phenotype seen in A^{vy}/a animals. Instead, mice with hypomethylated LTR in general exhibit a "black-back-yellow-belly" (YB) phenotype, whereas mice with hypermethylated LTR are have similar dorsal *Agouti* expression as wild-type A^wT/ A^wT or A^wT /a animals (Figure 6). To further confirm the phenotypic-epigenetic status relationship of the $A^{w\text{-}LTR}T$ allele, we analyzed DNA methylation levels of all YB $A^{w\text{-}LTR}T/a$ mice obtained in the first generation from chimera mating, and found all YB mice have a hypomethylated LTR (Figure 7). The YB phenotype of mice with hypomethylated LTR indicate some disruption of the dorsal *Agouti* expression, which has been shown to be dependent on hair-cycle specific 1B and 1C promoters[7,8] . As the distance between LTR insertion site (PS1A) and 1B, 1C are rather large (~100kb, Figure 1), the exact mechanism of how this LTR distally affected dorsal *Agouti* expression when unmethylated is currently unknown.

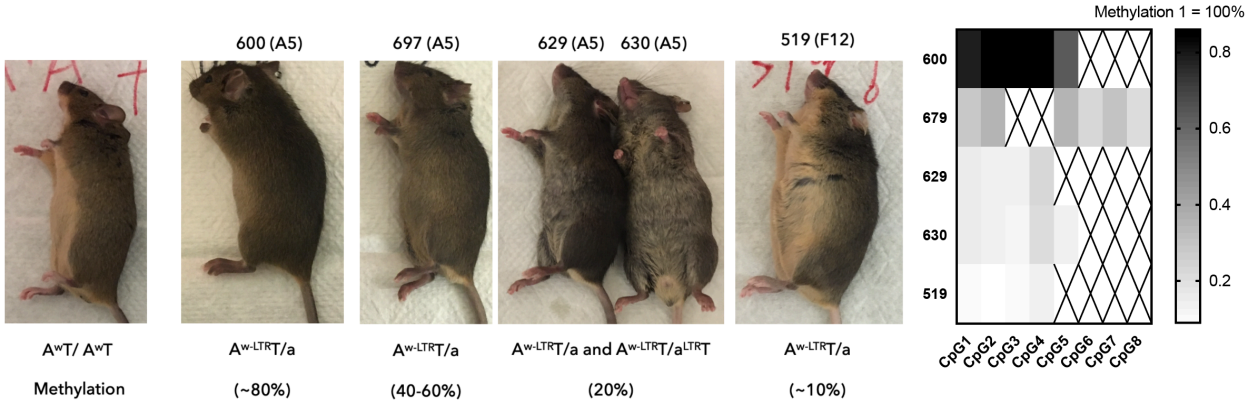

Figure 6. Phenotypes of Mice Carrying $A^{w\text{-}LTR}T$ Alleles with Different Degrees of LTR Methylation.

Mice are numbered and their ES origins are indicated in brackets (left). LTR methylation was measured by pyro-sequencing and averaged across CpGs (right). Methylation values of crossed CpGs didn't pass pyro-sequencing quality control.

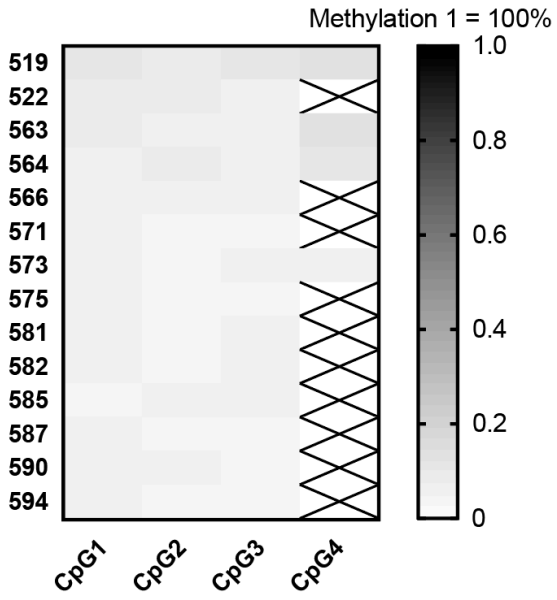

Figure 7. LTR DNA Methylation of Tail Genomic DNA from YB F1 Mice.

Only CpGs passed quality control are shaded with their respective methylation values.

DNA Methylation of LTR Is Homogenous Across Tissues within Individuals.

Since LTR methylation so far has been only measured all in tail genomic DNA, we next ask whether LTR methylation level is similar across different tissues within individual mice or exhibits tissue-specific values. We took two YB $A^{w\text{-}LTR}T$/a mice and measured LTR methylation in their brain, kidney, liver and spleen, and found these tissues exhibited similar hypomethylation (Figure 8A). Such unanimous methylation level across tissues was also observed in LTR hypermethylated mice (Figure 8B). This is consistent with observations in the A^{vy} allele and other metastable epialleles[14,23].

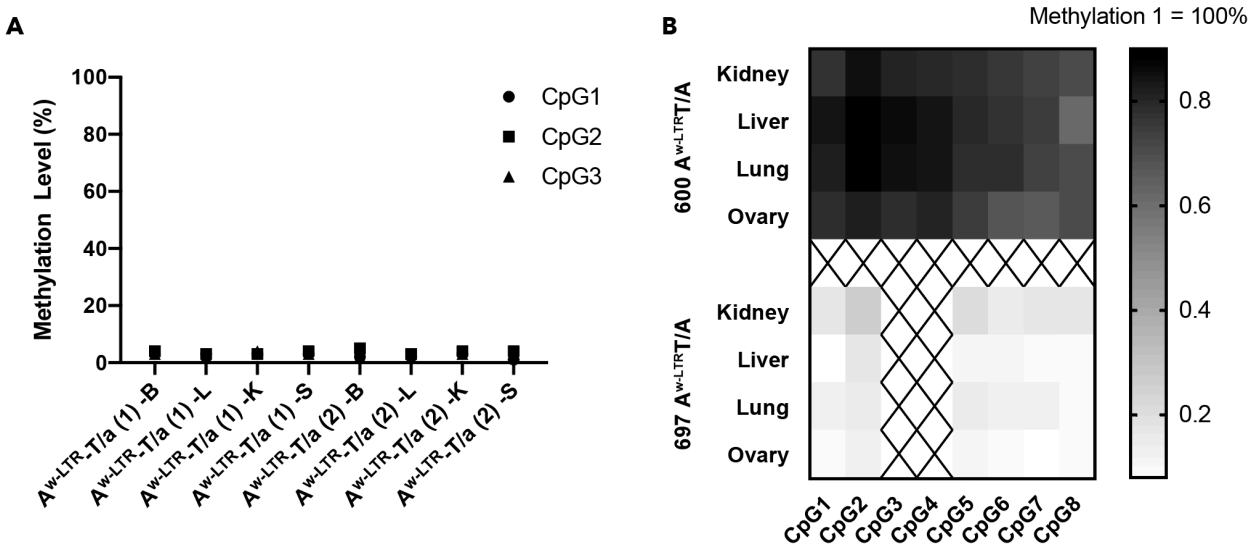

Figure 8. LTR Methylation Is Similar Across Tissues within Individual Mice.

(A). LTR methylation of brain (B), liver (L), Kidney (K) and spleen (S) of two YB (hypomethylated) mice. (B). LTR methylation of hypermethylated (600) and hypomethylated (697) mice.

Hypomethylated LTR Induces Weak Ectopic Agouti Expression, with Tissue-specificity

Since the IAP-LTR in the A^{vy} allele drives ectopic *Agouti* expression when unmethylated, we asked if an unmethylated solo LTR is sufficient to drive ectopic *Agouti* as well, especially in organ/tissue types where *Agouti* is not expressed from the wild-type $A^{w}T$ allele. We performed immunostaining against tdTomato on brain, kidney, and liver from $A^{w\text{-}LTR}T/a$ mice with hypomethylated LTRs (YB). All tissues exhibited overexpression of tdTomato in the $A^{w\text{-}LTR}T/a$ male and females with different degrees, whereas no tdTomato signal was detected in $A^{w}T/ A^{w}T$ or aT/a animals (Figure 9).

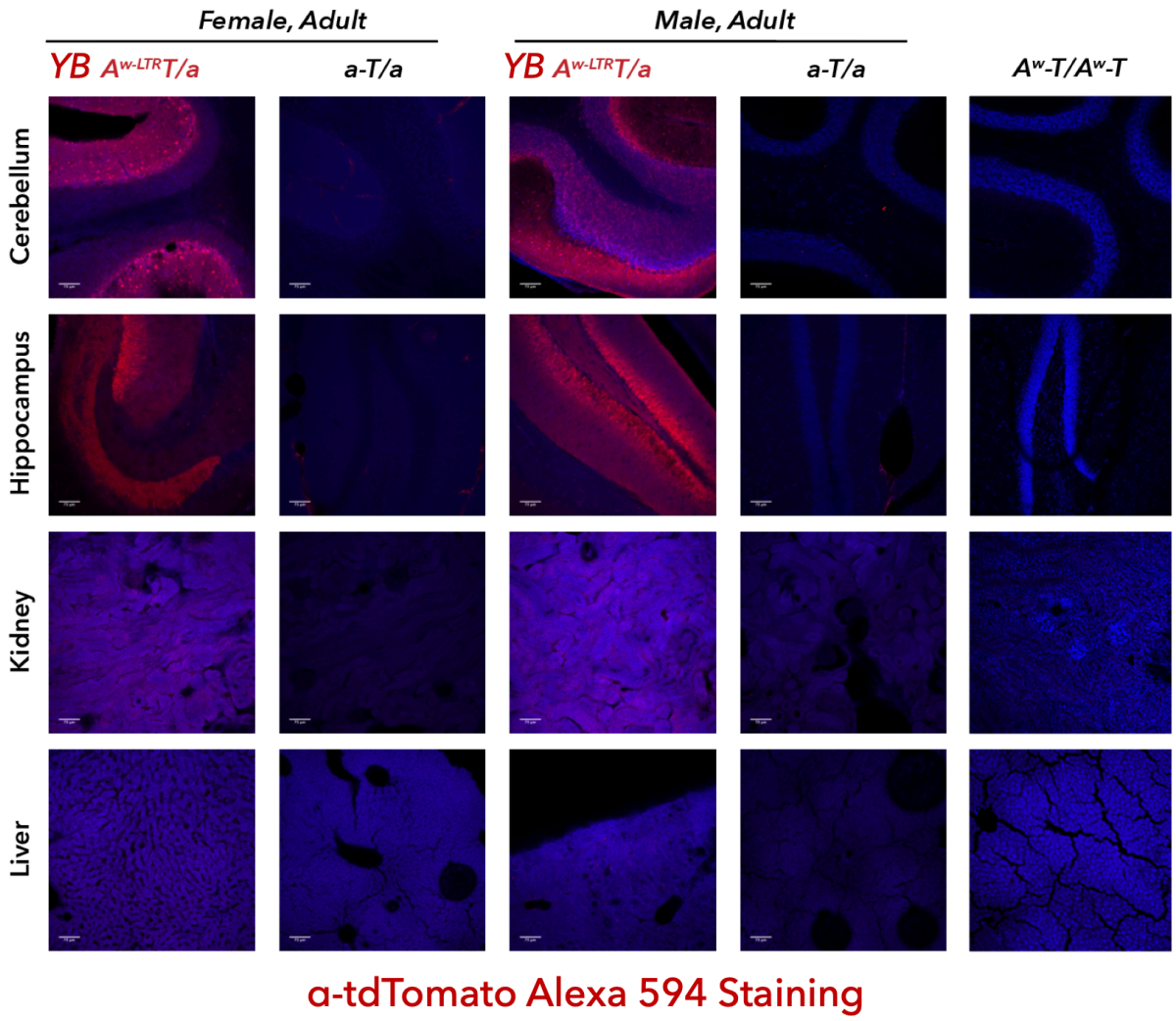

Figure 9. α-tdTomato Immunofluorescence on Mice Tissue Sections of Indicated Genotypes.

Animals are age matched. YB: black-back-yellow-belly phenotype, indicating hypomethylation at the LTR.

Next, we asked if hypermethylation at the LTR suppresses such ectopic promoter activity by directly comparing tdTomato signal single liver cells across different epigenotypes. Indeed, agouti mice ($A^{w\text{-}LTR}T$ /a with hypermethylated LTR) have almost no tdTomato$^+$ cells compared to either wild-type (A^wT/A^wT) or C57/B6J animals, whereas YB mice ($A^{w\text{-}LTR}T$ /a with hypomethylated LTR) showed various degrees of tdTomato$^+$ cells (Figure 10A).

Given the DNA methylation level within one animal is homogenous across different tissue-types, it is puzzling that the majority of the liver cells in YB animals are negative for tdTomato. We hypothesize that the LTR in tdTomato$^-$ liver cells in the YB liver should still be hypomethylated, and the LTR activity is suppressed by other epigenetic or transcriptional mechanisms. To test this, we sorted out the tdTomato$^-$ liver cells and compared the LTR methylation level to that of agouti animals ($A^{w\text{-}LTR}T$ /a with hypermethylated LTR) using pyrosequencing. Indeed, the tdTomato$^-$ liver cells from the YB mice are not hypermethylated, compared to the phenotypically identical tdTomato$^-$ liver cells from agouti mice (Figure 10B), indicating that and DNA methylation is sufficient (as all cells are tdTomato$^-$ from agouti mice), but not necessary (as tdTomato$^-$ cells from YB mice are still hypomethylated) in silencing LTR activity in the adult liver. Other silencing mechanisms may co-exist, such as histone methylation, while DNA methylation may serve as a safe guarding mechanism. The tdTomato$^+$ liver cells from YB mice may simply be "escapees" due to lack of DNA methylation at the LTR.

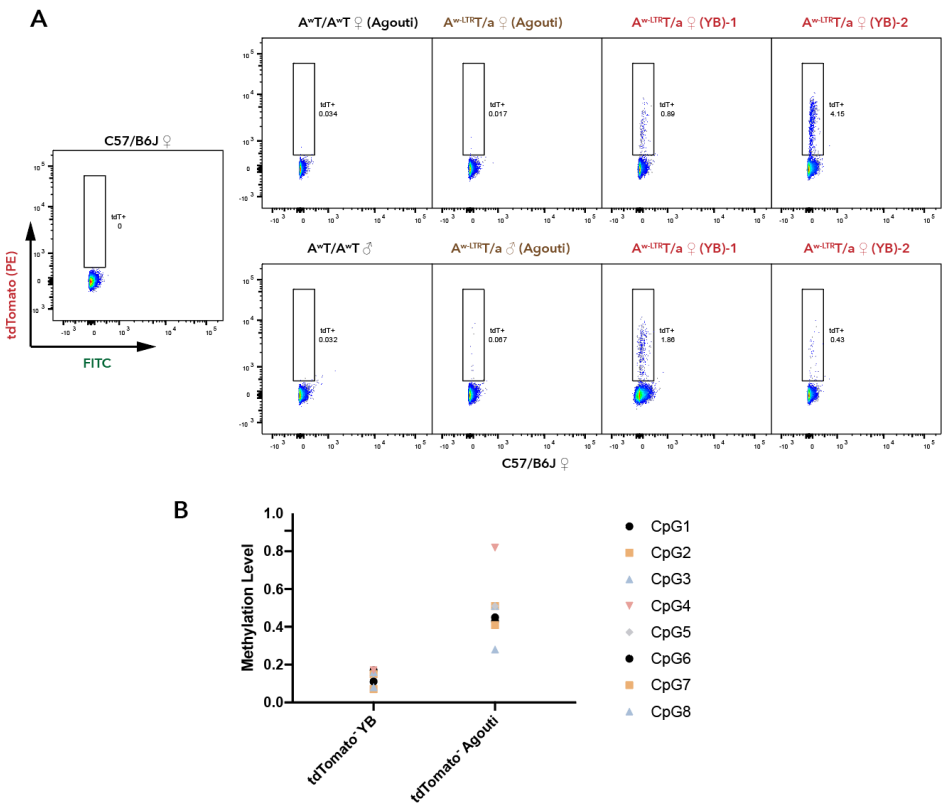

Figure 10. TdTomato Signal Across Aw-LTRT/a Mice with Hypoethylated or Hypermethylated LTRs.

(A) FACS of single liver cells of age- and sex-matched adults. Genotypes are indicated. Agouti animals have a hypermethylated LTR and YB animals have a hypomethylated LTR. (B). DNA methylation level of tdTomato$^-$ liver cells from either a YB mouse or an agouti mouse.

We wonder if the leakiness of LTR activity due to lack of LTR methylation of the A$^{w\text{-}LTR}$T allele happens in different tissue types, therefore we performed qRT-PCR analyses of both *Agouti* and tdTomato transcripts from the ovary, lung, kidney, liver and cerebellum in LTR-hypermethylated and LTR-hypomethylated animals, both of which are normalized to wildtype controls. Although hypomethylation of LTR drives *tdTomato* and *Agouti* ectopic expressions compared to wild-type A^{w}T allele and hypermethylated A$^{w\text{-}LTR}$T allele (with liver the most prominent), the absolute leakiness

of LTR is minimal compared to the expression level of house-keeping gene such as *GAPDH* and is variable dependent on the tissue type (Figure 11). In summary, LTR insertion is silenced by cell-intrinsic mechanisms, among which DNA methylation is sufficient but not necessary induced weak ectopic expression of *Agouti*, which is suppressed by DNA methylation.

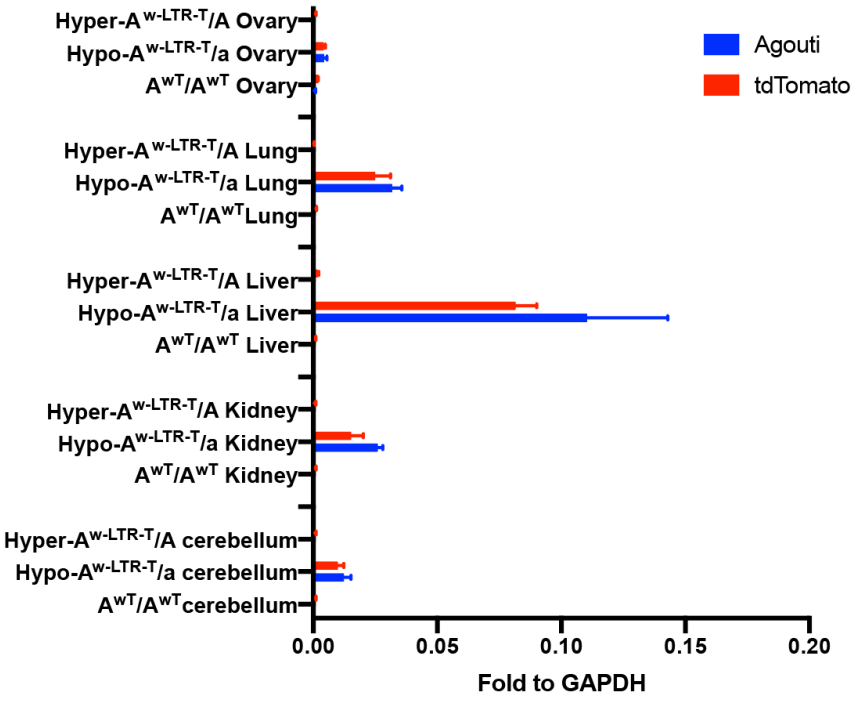

Figure 11. Hypomethylated LTR Cause LTR Leakiness of Tissue-dependent Levels.

qRT-PCR of *Agouti* and *tdTomato* expression from various tissues of LTR-hypermethyalted (Hyper), LTR-hypomethyalted (Hypo) $A^{w\text{-}LTR}T$/a mice and wild-type $A^{w}T$/ $A^{w}T$ mice.

DNA Methylation of IAP-LTR Can Be Inherited Across Generations, But Doesn't Re-establish Once Demethylated.

The presence of F1 mice with both hypomethylated and hypermethylated LTR indicate that the newly generated $A^{w\text{-}LTR}T$ allele is a metastable epiallele. However, these mice were initially derived from different ES clones. To exclude clonal effects and

further confirm that DNA methylation at the A$^{w\text{-}LTR}$T allele is metastable, we analyzed DNA methylation of all offspring from one single hypermethylated founder female (600, derived from A5 ES cells). Among all her offspring, regardless of sex, we observed both mice carrying LTRs with various methylation levels (Figure 12A). We then selected a few hypomethylated and hypermethylated offspring from 600 to breed with B6 mice for the next generation. We observed that hypomethylated LTRs are stably transmitted across the next generation, whereas hypermethylated whereas hypermethylated or lowly methylated LTRs present certain degrees of epigenetic metastability (Figure 12B). As the F1 founder of F12 ES clone-derived mice were YB to begin with (mostly with almost 0% LTR methylation), we did not observe re-gain of DNA methylation in any of the offspring from this lineage, consistent with the observation in A5 ES clone-derived mice (Figure 12C). In summary, the newly created A$^{w\text{-}LTR}$T allele can transmit DNA methylation to some degree across generations, however, is unable to re-establish DNA methylation at an inherited hypomethylated LTR. In conclusion, it did not meet the criteria of a conventionally-defined metastable epiallele[1,23].

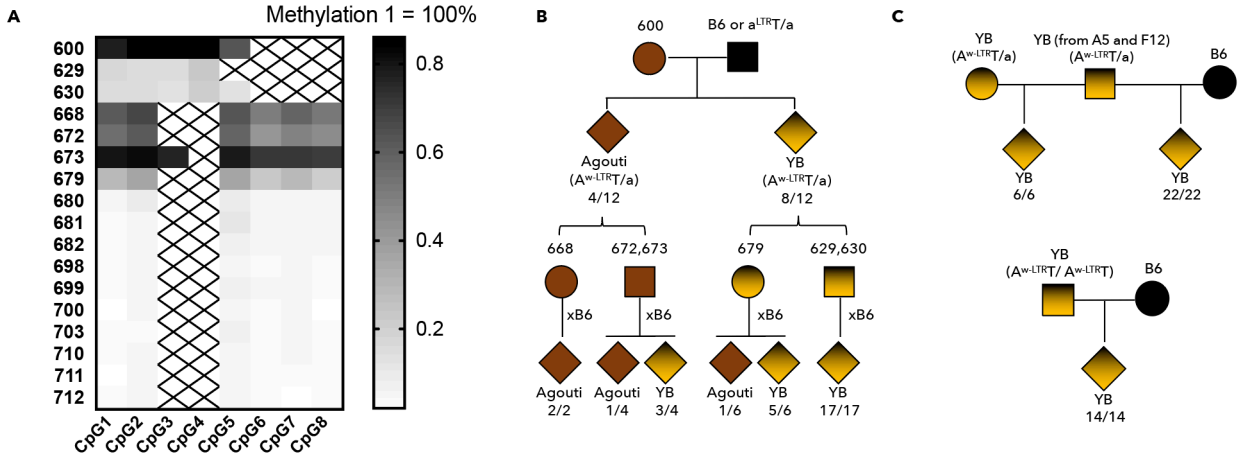

Figure 12. Epigenetic Inheritance of The A$^{w\text{-}LTR}$T allele.

(A) Methylation levels of the LTR in offspring from 600. (B) Phenotypes of the A5 ES cell-derived animals. Black-yellow gradient filling: YB phenotype (Hypomethylated LTR); black: B6; brown: Agouti (hypermethylated LTR). Squares: males, circles: females;

rhombus: all offspring. (C) Phenotypes of the F12 ES cell-derived animals. Annotation is the same as in (B).

No Major Effects of Methyl-rich Diet on Transgenerational Inheritance of LTR Methylation Have Been Observed.

As we observed no re-establishment of DNA methylation on hypomethylated LTRs across generation, we asked if supplement of methyl-rich nutrient in the maternal diet could generate stochastic de novo methylation in the offspring LTR. We subjected 7-8-week-old CD1 females with high-methyl or control diet for 2 weeks prior to mating to a YB $A^{w\text{-}LTR}T/a$ (hypomethylated LTR) males. We also include mating between CD1 and wild-type $A^{w}T/a$ males as controls. Among all the offspring inherited the $A^{w\text{-}LTR}T$ allele, we did not observe significant difference in tdTomato fluorescence intensity of the whole embryos at E15.5 (data not shown) between control and supplemented groups. More quantitatively in the developing gonads, tdTomato expression did not vary between control or supplemented groups $A^{w\text{-}LTR}T/a$ offspring, although they all showed elevated ectopic expression of tdTomato (LTR-driven) compared to wild-type $A^{w}T/a$ offspring (Figure 13). However, we cannot rule out that the lack of effect from maternal methyl-nutrient supplement is due to the low numbers of embryos sampled or the selection of organ (gonads).

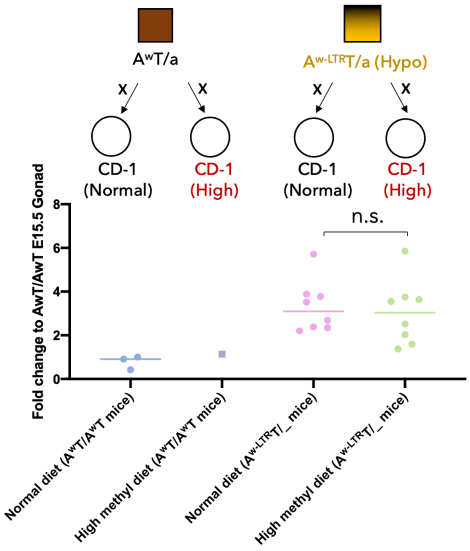

Figure 13. Maternal Diet Methyl-nutrient Supplement Did Not Significantly Alter Offspring LTR-driven tdTomato Expression in The Offspring.

In summary, to study the epigenetic effect of LTR insertion, potential origin and regulation of metastable epialleles at the cellular level during development, we generateds a targeted allele $A^{w\text{-}LTR}T$ using A^{vy}-derived LTR and *Agouti*-P2A-tdTomato expression reporter. Based on our results so far, solo-LTR insertion is not sufficient to generate a bona fide metastable epiallele mimicking A^{vy} at PS1A. However, the LTR can serve as an ectopic promoter for *Agouti*. The LTR activity can be silenced by DNA methylation, but can also possibly by other uncharacterized mechanisms, such as SETDB1-mediated histone modifications[24]. However, DNA methylation as a sufficient silencing mechanism provides stable suppression, as the lack of which causes leakiness of LTR activity and ectopic *tdTomato/Agouti* expression. Importantly, DNA methylation of the LTR is comparable across tissues within one individual, however, the ectopic promoter activity leaked by LTR hypomethylation exhibits tissue-level specificity. The DNA methylation of the LTR can be inherited from a hypermethylated parent to some offspring. However, once demethylated, the LTR cannot re-establish DNA methylation in later generations. Methyl-diet effect on modifying LTR methylation status might not

be very strong or hard to detect with small animal numbers, as seen in our preliminary results with $A^{w\text{-}LTR}T$ embryos.

References

1 Rakyan, V. K., Blewitt, M. E., Druker, R., Preis, J. I. & Whitelaw, E. Metastable epialleles in mammals. *Trends Genet* **18**, 348-351 (2002).

2 Dolinoy, D. C., Das, R., Weidman, J. R. & Jirtle, R. L. Metastable epialleles, imprinting, and the fetal origins of adult diseases. *Pediatr Res* **61**, 30R-37R, doi:10.1203/pdr.0b013e31804575f7 (2007).

3 Slotkin, R. K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**, 272-285, doi:10.1038/nrg2072 (2007).

4 Whitelaw, E. & Martin, D. I. K. Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat Genet* **27**, 361-365, doi:Doi 10.1038/86850 (2001).

5 Duhl, D. M. J., Vrieling, H., Miller, K. A., Wolff, G. L. & Barsh, G. S. Neomorphic Agouti Mutations in Obese Yellow Mice. *Nat Genet* **8**, 59-65, doi:DOI 10.1038/ng0994-59 (1994).

6 Millar, S. E., Miller, M. W., Stevens, M. E. & Barsh, G. S. Expression and Transgenic Studies of the Mouse Agouti Gene Provide Insight into the Mechanisms by Which Mammalian Coat Color Patterns Are Generated. *Development* **121**, 3223-3232 (1995).

7 Bultman, S. J. *et al.* Molecular analysis of reverse mutations from nonagouti (a) to black-and-tan (a(t)) and white-bellied agouti (Aw) reveals alternative forms of agouti transcripts. *Genes Dev* **8**, 481-490, doi:10.1101/gad.8.4.481 (1994).

8 Vrieling, H., Duhl, D. M., Millar, S. E., Miller, K. A. & Barsh, G. S. Differences in dorsal and ventral pigmentation result from regional expression of the mouse agouti gene. *Proc Natl Acad Sci U S A* **91**, 5667-5671, doi:10.1073/pnas.91.12.5667 (1994).

9 Gaudet, F. *et al.* Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. *Mol Cell Biol* **24**, 1640-1648, doi:10.1128/Mcb.24.4.1640-1648.2004 (2004).

10 Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**, 116-117, doi:Doi 10.1038/2413 (1998).

11 Dolinoy, D. C. The agouti mouse model: an epigenetic biosensor for nutritional and environmental alterations on the fetal epigenome. *Nutr Rev* **66**, S7-S11, doi:10.1111/j.1753-4887.2008.00056.x (2008).

12 Stathopoulou, A., Lucchiari, G. & Ooi, S. K. T. DNA Methylation Is Dispensable for Suppression of the Agouti viable yellow Controlling Element in Murine Embryonic Stem Cells. *PloS one* **9**, doi:ARTN e10735510.1371/journal.pone.0107355 (2014).
13 Morgan, H. D., Sutherland, H. G. E., Martin, D. I. K. & Whitelaw, E. Epigenetic inheritance at the agouti locus in the mouse. *Nat Genet* **23**, 314-318 (1999).
14 Waterland, R. A. & Jirtle, R. L. Transposable elements: Targets for early nutritional effects on epigenetic gene regulation. *Mol Cell Biol* **23**, 5293-5300, doi:10.1128/Mcb.23.15.5293-5300.2003 (2003).
15 Blewitt, M. E., Vickaryous, N. K., Paldi, A., Koseki, H. & Whitelaw, E. Dynamic reprogramming of DNA methylation at an epigenetically sensitive allele in mice. *Plos Genet* **2**, 399-405, doi:ARTN e4910.1371/journal.pgen.0020049 (2006).
16 Waterland, R. A., Tahiliani, K. G., Rached, M. T. & Travisano, M. Diet-induced hypermethylation at viable yellow agouti is not inherited transgenerationally. *Faseb J* **21**, A291-A291 (2007).
17 Cropley, J. E., Suter, C. M., Beckman, K. B. & Martin, D. I. K. Germ-line epigenetic modification of the murine A(vy) allele by nutritional supplementation. *P Natl Acad Sci USA* **103**, 17308-17312, doi:10.1073/pnas.0607090103 (2006).
18 Seisenberger, S. *et al.* The Dynamics of Genome-wide DNA Methylation Reprogramming in Mouse Primordial Germ Cells. *Mol Cell* **48**, 849-862, doi:10.1016/j.molcel.2012.11.001 (2012).
19 Waterland, R. A. *et al.* Maternal methyl supplements increase offspring DNA methylation at Axin Fused. *Genesis* **44**, 401-406, doi:10.1002/dvg.20230 (2006).
20 Dupressoir, A. & Heidmann, T. Expression of intracisternal A-particle retrotransposons in primary tumors of oncogene-expressing transgenic mice. *Oncogene* **14**, 2951-2958, doi:DOI 10.1038/sj.onc.1201148 (1997).
21 Kuff, E. L. & Lueders, K. K. The Intracisternal a-Particle Gene Family - Structure and Functional-Aspects. *Adv Cancer Res* **51**, 183-276, doi:Doi 10.1016/S0065-230x(08)60223-7 (1988).
22 Christy, R. J. & Huang, R. C. C. Functional-Analysis of the Long Terminal Repeats of Intracisternal a-Particle Genes - Sequences within the U3-Region Determine Both the Efficiency and Direction of Promoter Activity. *Mol Cell Biol* **8**, 1093-1102, doi:Doi 10.1128/Mcb.8.3.1093 (1988).
23 Kazachenka, A. *et al.* Identification, Characterization, and Heritability of Murine Metastable Epialleles: Implications for Non-genetic Inheritance. *Cell* **175**, 1717, doi:10.1016/j.cell.2018.11.017 (2018).

24 Leung, D. *et al.* Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *P Natl Acad Sci USA* **111**, 6690-6695, doi:10.1073/pnas.1322273111 (2014).