Epigenetic Regulation of Pluripotency and Differentiation

Michael J. Boland, Kristopher L. Nazor, Jeanne F. Loring

Abstract: The precise, temporal order of gene expression during development is critical to ensure proper lineage commitment, cell fate determination, and ultimately, organogenesis. Epigenetic regulation of chromatin structure is fundamental to the activation or repression of genes during embryonic development. In recent years, there has been an explosion of research relating to various modes of epigenetic regulation, such as DNA methylation, post-translational histone tail modifications, noncoding RNA control of chromatin structure, and nucleosome remodeling. Technological advances in genome-wide epigenetic profiling and pluripotent stem cell differentiation have been primary drivers for elucidating the epigenetic control of cellular identity during development and nuclear reprogramming. Not only do epigenetic mechanisms regulate transcriptional states in a cell-type-specific manner but also they establish higher order genomic topology and nuclear architecture. Here, we review the epigenetic control of pluripotency and changes associated with pluripotent stem cell differentiation. We focus on DNA methylation, DNA demethylation, and common histone tail modifications. Finally, we briefly discuss epigenetic heterogeneity among pluripotent stem cell lines and the influence of epigenetic patterns on genome topology. (Circ Res. 2014;115:311-324.)

Key Words: epigenomics ■ methylation ■ stem cell, pluripotent
chromatin remodeling (the histone code), the complex interplay of different epigenetic modes is beginning to be revealed. Epigenetic modes that involve DNA or histone modifications include the enzymes that catalyze the particular modification (writers), proteins that recognize and bind the modification (readers), and the enzymes that remove the modification (erasers). For example, histone 3 lysine 27 trimethylation (H3K27me3) is catalyzed by the histone methyltransferase (HMTase) enhancer of zeste homolog 2 (EZH2), read by chromobox homolog 7 (CBX7), and erased by the lysine-specific demethylase UTX. Readers possess characteristic recognition motifs, such as the methyl-CpG binding domain for 5mC and the bromodomain for lysine acetylation. Several different domains recognize methylated lysines or arginines in a residue/modification-specific manner. They include chromodomains, tudor, WD40 repeat, and plant homeodomain (PHD) finger domains. Figure 1 shows a list of common readers, writers, and erasers pertaining to the epigenetic marks reviewed here.

In this review, we provide an overview of DNA methylation and histone modifications and their crosstalk on how they affect specific regulatory elements, namely, promoters and enhancers. We will then review available knowledge pertaining to epigenetic dynamics during the developmental processes of differentiation and X-chromosome inactivation (XCI), including the role that epigenetics plays on nuclear organization and genome topology. Where appropriate, we provide examples pertaining to cardiovascular research and direct the reader to more in-depth reviews of the topics we discuss.

**DNA Methylation**

DNA methylation is a heritable, yet reversible, epigenetic modification that plays a central role in transcriptional repression, suppression of retrogene transposition, genomic imprinting, XCI, and higher order chromatin organization. DNA methyltransferases (DNMTs) catalyze the transfer of a methyl group from cofactor S-adenosylmethionine to carbon 5 of the cytosine ring to generate 5mC. In mammals, the de novo DNMTs (DNMT3A, DNMT3B, and cofactor DNMT3L) are responsible for establishing postfertilization genomic methylation patterns, whereas DNMT1, together with its cofactor ubiquitin-like with PHD and ring finger domains 1 (Figure 1), localizes to the replication fork to maintain methylation patterns on the newly synthesized DNA strand during replication. Although the minimal sequence requirement for DNA methylation is that the target cytosine typically resides within a CpG dinucleotide, sequence preferences have been identified for the de novo DNMTs that seem to be conserved from mice to humans.

In mouse development, after fertilization and before pronuclear fusion, zygotic genome activation and the first-cell division, the paternal genome is demethylated by enzymatic modification of 5mC (discussed later), whereas the maternal genome is passively demethylated during subsequent rounds of replication via nuclear exclusion of an oocyte-specific Dnmt1 isoform. Thus, the totipotent zygote is essentially devoid of DNA methylation except at imprinted regions. The genome is gradually remethylated during subsequent cleavage divisions that generate the morula and early blastocyst. The genomes of PSCs derived from the inner cell mass/epiblast are highly methylated. Erasure of gametic methylation patterns of replication via nuclear exclusion of an oocyte-specific Dnmt1 isoform. Thus, the totipotent zygote is essentially devoid of DNA methylation except at imprinted regions. The genome is gradually remethylated during subsequent cleavage divisions that generate the morula and early blastocyst. The genomes of PSCs derived from the inner cell mass/epiblast are highly methylated. Erasure of gametic methylation patterns of replication via nuclear exclusion of an oocyte-specific Dnmt1 isoform. Thus, the totipotent zygote is essentially devoid of DNA methylation except at imprinted regions. The genome is gradually remethylated during subsequent cleavage divisions that generate the morula and early blastocyst. The genomes of PSCs derived from the inner cell mass/epiblast are highly methylated.

**Nonstandard Abbreviations and Acronyms**

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>5caC</td>
<td>5-carboxylcytosine</td>
</tr>
<tr>
<td>5fC</td>
<td>5-formylcytosine</td>
</tr>
<tr>
<td>5mC</td>
<td>5-methylcytosine</td>
</tr>
<tr>
<td>5hmC</td>
<td>5-hydroxymethylcytosine</td>
</tr>
<tr>
<td>BER</td>
<td>base excision repair</td>
</tr>
<tr>
<td>CTCF</td>
<td>CCCTC-binding factor</td>
</tr>
<tr>
<td>DMRs</td>
<td>differentially methylated regions</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>ESC</td>
<td>embryonic stem cells</td>
</tr>
<tr>
<td>hESC</td>
<td>human embryonic stem cells</td>
</tr>
<tr>
<td>hPSC</td>
<td>human pluripotent stem cells</td>
</tr>
<tr>
<td>iPSC</td>
<td>induced pluripotent stem cells</td>
</tr>
<tr>
<td>LAD</td>
<td>lamina-associated domains</td>
</tr>
<tr>
<td>PRC</td>
<td>polycomb repressive complex</td>
</tr>
<tr>
<td>PSC</td>
<td>pluripotent stem cells</td>
</tr>
<tr>
<td>RNAPII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>TDG</td>
<td>thymine-DNA glycosylase</td>
</tr>
</tbody>
</table>

The DNA methylation machinery works in conjunction with other modes of epigenetic regulation to regulate gene expression via local chromatin structure and higher order genomic topology. Methyl-DNA binding proteins (MBD1, MBD2, and MBD3; methyl-CpG binding protein 2 [MeCP2]; and ZBTB33 [Kaiso]) bind 5mC and recruit histone deacetylases, repressive histone methyltransferases, and other chromatin remodeling proteins, such as ATP-dependent chromatin remodeling complexes, to repress transcription actively. The genomic location of DNA methylation affects its role on transcription and higher order chromatin structure. The highly and constitutively methylated centromeric and pericentromeric DNA satellite repeats play a large role in heterochromatin organization at these regions. Indeed, the heterochromatic centromeres from many chromosomes aggregate to form structures called chromocenters that play roles in overall nuclear architecture. Many studies have demonstrated that promoter CpG methylation is inversely correlated with gene expression. Genes regulated by methylation usually contain a low density of promoter CpG sites. Most low CpG density promoters are methylated in ESCs and subsequently demethylated and expressed in a lineage- or cell-type-specific manner during differentiation. Areas termed CpG islands,
Non-CpG Methylation

Non-CpG methylation (CpH, where H=A, T, or C) is enriched in PSCs and lost on differentiation in all cell types/tissues except the brain, where it accumulates in neurons during postnatal and adolescent development, coinciding with periods of synaptogenesis and synaptic pruning. It is far less abundant than CpG methylation in somatic cells, constituting only 0.02% of total 5mC in somatic cells, but in hESCs, 25% of 5mC are in CpH. The most prevalent non-CpG methylation is found at CpA dinucleotides with mCpT and mCpC constituting a small fraction of CpH methylation. Non-CpG methylation seems to be catalyzed by DNMT3A and DNMT3B. The functions of non-CpG methylation remain elusive, but high-resolution global-profiling techniques suggest that it may have a myriad of functions. For example, non-CpG and CpG methylation have been implicated in the regulation of RNA splicing in ESCs and neurons, respectively.

5mC, 5hmC

which are regions of high CpG density found within or near proximal promoters or transcription start sites, are typically devoid of DNA methylation.

Recent studies suggest that promoter hypomethylation coupled with methylation within gene bodies strongly correlates with expression. Differentially methylated regions (DMRs), generally found at regulatory elements such as enhancers and promoters, display lineage- or cell-type-specific methylation patterns. DMRs can be used to predict whether a cell belongs to a certain lineage or tissue based on the genes they regulate. As a general rule, the extent of genomic methylation, both non-CpG and CpG methylation have been implicated in the regulation of RNA splicing in ESCs and neurons, respectively.
**Oxidation of 5mC and DNA Demethylation**

Recently, it was discovered that the ten eleven translocation (TET) family of 2OG-Fe(II) dioxygenases catalyze the hydroxylation of 5mC to generate 5hmC. It was first characterized in cerebellar Purkinje neurons, where it is up to 40% as abundant as 5mC. The genomes of PSCs also possess high levels of 5hmC. At low resolution, 5hmC and 5mC exhibit mutually exclusive localization patterns: centromeric and pericentromeric regions tend to have higher concentrations of 5mC, whereas 5hmC is more enriched on chromosome arms. High-resolution global mapping demonstrates an anticorrelation between 5mC and 5hmC levels at regulatory elements. Moreover, 5hmC signatures are enriched at sites of DNaseI hypersensitivity, which are indicative of genomic regions bound by regulatory proteins, whereas 5mC is generally depleted at sites of DNA–protein interaction. 5hmC is particularly enriched at bivalent domains (discussed later), near transcription start sites, at CpG-rich proximal promoter regions, and at active and poised enhancers. Interestingly, 5hmC is generally depleted at sites of DNA–protein interaction. 5hmC is particularly enriched at bivalent domains (discussed later), near transcription start sites, at CpG-rich proximal promoters, and at active and poised enhancers. Interestingly, both 5mC and 5hmC are enriched within the gene bodies of actively transcribed genes.

To date, there are 3 identified TET isoforms that play different roles during development. TET3 primarily functions early in embryogenesis, where it rapidly converts 5mC to 5hmC in the male pronucleus. The other 2 isoforms, TET1 and TET2, not only share some targets but also have mutually exclusive functions during development. Depletion of Tet1 or Tet2 in ESCs results in globally reduced levels of 5hmC but has no discernible effect on pluripotency. Tet1 or Tet2 single knockouts display different differentiation defects—Tet1 knockout differentiation is skewed toward trophectoderm (in agreement with in vitro differentiation data using Tet1 knockdown mouse ESCs [mESCs]), and Tet2 knockout animals possess dysregulated hematopoietic stem cells. The majority of double Tet1/Tet2 knockout animals die at birth; however, a small percentage of double knockout animals are phenotypically normal and fertile but possess imprinting abnormalities. Implicating TET proteins and 5hmC in the proper maintenance of genomic imprints. These results suggest that TET1 and TET2 have nonoverlapping albeit mild functions during development: TET1 helps maintain pluripotency by suppression of trophectoderm differentiation, and TET2 maintains normal homeostasis in the hematopoietic lineage.

Originally it was thought that 5hmC was merely an intermediate in the demethylation pathway, but its putative role in gene regulation has grown, along with increased controversy as to whether MeCP2 is able to recognize and bind 5hmC. Much emphasis has been placed on MeCP2’s putative role in reading 5hmC because of the high levels of both MeCP2 and 5hmC in the brain, and MeCP2’s role in the neurodevelopmental autistic disorder, Rett Syndrome. Again, there are conflicting reports as to whether MeCP2 is able to recognize and bind 5hmC. Mellén et al. identify 5hmC as a high-affinity substrate of MeCP2 and observe MeCP2 and 5hmC colocalization at genes expressed in the brain. However, Szulwach et al. demonstrate an inverse relationship between MeCP2 and 5hmC levels in the brain. Moreover, other reports suggest that 5hmC is a poor physical substrate for MeCP2 in vitro. Although recombinant MeCP2 may possess low affinity for 5hmC in vitro, its affinity in vivo may be enhanced by either accessory proteins or by post-translational modifications. It will be important to rectify these differences and to identify molecules that read 5hmC and its oxidized derivatives further.

DNA demethylation has both fascinated and frustrated scientists for decades. The most dramatic example of DNA demethylation occurs early in development when the paternal genome is rapidly demethylated before pronuclear fusion and the first-cell division occurs. Other examples of demethylation are more localized, such as the derepression of genes during germ layer formation, lineage specification and terminal differentiation, and the activity-dependent epigenetic remodeling events related to synaptic plasticity and memory formation.

The precise role of active, or enzymatic, DNA demethylation has been the subject of speculation for years. Recent studies indicate that active demethylation involves a DNA repair event, such as base excision repair (BER). However, none of the glycosylases that initiate BER (thymine-DNA glycosylase [TDG], MBD4, or SMUG1) possess high affinity for 5mC as a substrate. Therefore, for demethylation to proceed through a BER-based mechanism, 5mC must be modified into a form that is efficiently recognized by the BER glycosylases. Spontaneous deamination of 5mC to thymidine yields a T-G mismatch, and deamination of 5hmC results in a 5-hydroxymethylcytosine (5hmC-G) mismatch, both of which are recognized by either of 2 mammalian BER glycosylases; TDG or MBD4. TDG and MBD4 have both been implicated in DNA demethylation, and recent studies have defined a clear role for TDG in BER-mediated DNA demethylation. Deamination of 5hmC by activation-induced deaminase (AID) results in demethylation of reporter constructs containing 5hmC in mammalian cells, and AID has been shown to be important for the demethylation events that occur during somatic cell reprogramming to pluripotency.

The precise role of AID in active DNA demethylation remains unclear because recently it has been shown that recombinant AID displays little activity toward 5mC and no activity toward 5hmC as substrates although unidentified cofactors may enhance AID substrate affinity in vivo.

Currently, experimental evidence suggests a model of active DNA demethylation that can be summarized as follows (Figure 2): TET-mediated hydroxylation of 5mC generates 5hmC, which is recognized by a complex containing TDG and AID (and possibly GADD45). AID then deaminates 5hmC to
enzymatic cascade catalyzed by multiple proteins. In the original studies documenting rapid demethylation, it seemed that the paternal genome was actively demethylated because the antibody-based methods used to visualize 5mC did not recognize 5hmC. It has more recently been demonstrated that 5mC in the paternal genome is rapidly converted to 5hmC by TET3, which is then passively lost during subsequent rounds of replication.75–77

**Histone Code**

Post-translational modification of histone tails results in a combinatorial readout that affects gene expression through the regulation of local chromatin structure.14 A comprehensive review of all histone tail modifications, which include acetylation, methylation, citrullination (also known as deimination), phosphorylation, ubiquitination, sumoylation, and biotinylation, is beyond the scope of this article. The precise role of histone modifications, such as phosphorylation, sumoylation, and biotinylation, and their effects on chromatin structure and transcriptional state remain unclear and warrant further investigation.78 Interestingly, citrullination, the enzymatic conversion of arginine to citrulline, has recently been implicated in the establishment and maintenance of pluripotency.79 Here, we will focus on the 5 most commonly studied modifications of histone 3 (H3), specifically H3 lysine 4 methylation (H3K4me), H3 lysine 9 methylation (H3K9me), H3 lysine 27 acetylation (H3K27ac), H3 lysine 27 methylation (H3K27me), and H3 lysine 36 methylation (H3K36me3) because they have been shown to correlate with and predict chromatin/transcriptional states of regulatory elements in many different cell types accurately.

High-resolution, genome-wide studies have elucidated common themes of gene regulation by modified histones. Genomic regulatory regions that respond to developmental and environmental stimuli (ie, enhancers and promoters) are marked by histone modifications that confer transcriptionally permissive (euchromatin) or repressive (heterochromatin) chromatin states, which are mediated by the Trithorax group proteins and polycomb group proteins, respectively. Common euchromatin modifications are H3K4me3, H3K9ac, H3K27ac, and H3K36me3. They are found primarily at active enhancers (H3K9ac and H3K27ac), promoters (H3K4me3), and within the bodies of actively transcribed genes (H3K36me3).80 H3K4me3 displays a punctate localization pattern within 1 to 2 kb of active promoters that contain CpG islands.28 It promotes transcription by recruiting nucleosome remodeling complexes and histone acetylases.

The 2 most studied modifications found at repressed chromatin are H3K9me3 and H3K27me3, generally localized to constitutive and facultative heterochromatin, respectively. The histone methyltransferase that catalyzes H327me3 is enhancer of zeste homolog 2 (EZH2), which is a member of the multigene subunit polycomb repressive complex 2 (PRC2) that also contains embryonic ectoderm development, suppressor of zeste 12, and retinoblastoma binding proteins 4 and 7. PRC2 is targeted to genomic regions in response to developmental cues where it catalyzes H3K27me3. This modification recruits another multiprotein complex, PRC1, which contains members...
of the chromobox family that recognize H3K27me3. Also included in PRC1 is the ring finger protein, RING1B, an ubiquitin ligase that catalyzes monoubiquitination of histone H2A, a modification that impedes RNA polymerase II (RNAPII) elongation, resulting in transcriptional repression. Originally identified in Drosophila, the suppressor of variegation 3 to 9 family of histone methyltransferases, which do not associate with PRC complexes, catalyzes trimethylation of H3K9. This modification is heavily enriched at centromeric and pericentromeric DNA where, together with the DNA methylation machinery, it mediates the constitutive higher order heterochromatin structure necessary for mitotic spindle assembly.

PRCs play many roles during the determination of cell identity. Abolishment of either PRC1 or PRC2 activity in ESCs results in differentiation, albeit skewed toward ectoderm and endoderm, respectively. However, double knockouts (Eed<sup>−/−</sup>, Ring1B<sup>−/−</sup>) completely lose the ability to initiate differentiation. These results indicate that the 2 complexes have both redundant and nonoverlapping functions during development and the maintenance of PSC identity. Furthermore, in an excellent example of epigenetic crosstalk, PRC2 has been physically coupled to long noncoding RNA regulation of chromatin structure during cardiomyocyte differentiation.

In general, the genome of PSCs is enriched for open or transcriptionally permissive/posed chromatin and contains less heterochromatin relative to somatic cells (see below). PSCs contain more acetylated chromatin and smaller regions of H3K9me3 and H3K27me3 relative to differentiated cells. Knockout of the histone acetyltransferase MYST1 (MOZ YBF2 SAS2 TIP60 family member 1), a member of the core pluripotency network, results in enhanced heterochromatin formation and loss of differentiation potential in ESCs.

**Bivalent Domains**

Traditionally, it was thought that genes required for downstream developmental process were actively needed until recently. The discovery of bivalent domains in ESCs changed the way we view gene regulation in pluripotency and differentiation. Bivalent domains were first identified as broad H3K27me regions that also contain smaller, narrowly defined regions of H3K4me3 localized at transcriptional start sites. Bivalent domains can be further categorized based on histone modifications and polycomb group (PcG) protein occupancy. Domains that contain PRC2 only are weakly conserved and generally correspond to membrane proteins or genes of unknown function. However, domains containing both PRC1 and PRC2 complexes are highly conserved evolutionarily and show strong enrichment for many developmentally regulated transcription factors and other factors involved in morphogenesis and in cell signaling. Polycomb targets are commonly associated with CpG-rich DNA, and approximately half of PRC2 binding sites correspond to CpG islands. Interestingly, PRC2 is present at nearly all bivalent regions, whereas PRC1 occupies less than half (39%) of bivalent promoters. The majority of bivalent domains in ESCs directly overlap transcription start sites of genes that encode transcription factors and approximately half of the bivalent domains contain binding sites for ≥1 of the pluripotency transcription factors.

**Promoters**

Promoters surround transcriptional start sites and serve as the platform on which the basal transcription machinery and RNAPII assembly during transcription initiation. They are generally categorized based on CpG content/DNA methylation status and histone modifications as active, repressed, or poised (Figure 3). Active genes possess promoters enriched for H3K4me3, H3K9ac, H3K27ac, and 5hmC and are typically devoid of 5mC. Conversely, the promoters of repressed genes are marked by H3K27me3 and H3K9me3, and 5mC. Poised genes, which are either activated or repressed during development, have bivalent promoters, marked by the presence of both H3K4me3 and H3K27me3. Poised genes are generally associated with more complex expression patterns and include key developmental transcription factors, morphogens, and cell surface molecules. In addition, several bivalent promoters seem to regulate transcript for lineage-specific microRNAs.

Highly expressed genes, such as metabolic and housekeeping genes, usually contain hypomethylated CpG islands within their promoters. It is estimated that 95% of transcriptional start sites marked by H3K4me3 contain CpG islands. Promoters with low CpG content are more sensitive to regulation by DNA methylation and are typically found at genes expressed in a highly tissue-specific manner. In ESCs, only a small percentage of low CpG promoters contain significant levels of H3K4me3 and essentially none are marked by H3K27me3.

Ren et al have proposed that the epigenetic mechanisms regulating a given promoter may be influenced by the promoter’s sequence. In support of this, they find that genes preferentially expressed in ESCs and early cell types during differentiation (neural progenitor cells and mesendoderm) are CpG rich and contain CpG islands, whereas a much smaller percentage of genes containing CpG islands are expressed in tissue-specific manner. Furthermore, they find that developmentally regulated genes are regulated by promoters with high CpG content and are typically marked by H3K27me3, in contrast to tissue-specific promoters that possess low CpG density and are preferentially enriched for DNA methylation.

**Enhancers**

About 400,000 putative enhancers have been defined in the human genome based on analysis of histone modifications and proximity to transcription start sites. Enhancers are typically 200 to 300 bp in length and exhibit nucleosomal depletion when compared with flanking chromatin. They are enriched for DNase hypersensitivity sites and the histone acetyltransferase, ElA binding protein p300, suggesting an overall open chromatin structure. Indeed, most enhancers are bound by multiple transcription factors. The Mediator complex, a transcriptional coactivator, is also commonly found at enhancers. The Mediator complex is a large multi-protein complex that is necessary for RNAPII recruitment to the promoters of enhancer-regulated target genes. In ESCs, co-occupancy by Mediator proteins, Oct4, Sox2, and Nanog is predictive of an enhancer.

Enhancers can be distinguished from promoters by their unique histone modification signatures and, similar to promoters, these histone modifications can be used to classify...
them as active, poised, or repressed (Figure 3). For exam-
ple, although both promoters and enhancers typically pos-
sess H3K4me3 and repressed (red mark) chromatin. For instance, the Trithorax Group (TrxG) proteins and polycomb group (PcG) proteins, which catalyze H3K4me3 and histone 3 lysine 27 trimethylation (H3K27me3), respectively, localize to poised promoters. Poised regulatory elements are also generally devoid of 5-methylcytosine (5mC) and enriched for 5-hydroxymethylcytosine (5hmC). Poised enhancers can be distinguished from poised promoters by the presence of the enhancer-specific mark, H3K4me1, and bound Mediator complex. Chromatin loops organized by CCCTC-binding factor (CTCF) binding localizes distal poised (or active) enhancers and promoters within close physical proximity to facilitate transcription. The transition from poised to active regulatory elements involves loss of PcG localization, and demethylation and concomitant acetylation of H3K27. Gains of H3K36me3 within gene bodies are also observed on gene activation. Heterochromatin formation is associated with gene repression. This is mediated, in part, by erasure of activating histone modifications (H3K4me1, H3K27ac, and 5hmC), establishment of repressive marks, such as H3K27me3 and 5mC, and nucleosomal compaction. DNMT, DNA methyltransferase; HAT, histone acetyltransferase; HDAC, histone deacetylase; MBD, methyl-DNA binding domain protein; MED, mediator complex; PcG, polycomb group complex; POL II, RNA polymerase II; TET, ten eleven translocation dioxygenase; TF, transcription factor; TrxG, trithorax group complex.

Figure 3. Epigenetic control and remodeling of regulatory elements during development. A, Poised enhancers/promoters are typically found in pluripotent stem cells. They possess epigenetic modifications indicative of both transcriptionally active (green marks) and repressed (red mark) chromatin. For instance, the Trithorax Group (TrxG) proteins and polycomb group (PcG) proteins, which catalyze H3K4me3 and histone 3 lysine 27 trimethylation (H3K27me3), respectively, localize to poised promoters. Poised regulatory elements are also generally devoid of 5-methylcytosine (5mC) and enriched for 5-hydroxymethylcytosine (5hmC). Poised enhancers can be distinguished from poised promoters by the presence of the enhancer-specific mark, H3K4me1, and bound Mediator complex. Chromatin loops organized by CCCTC-binding factor (CTCF) binding localizes distal poised (or active) enhancers and promoters within close physical proximity to facilitate transcription. B, The transition from poised to active regulatory elements involves loss of PcG localization, and demethylation and concomitant acetylation of H3K27. Gains of H3K36me3 within gene bodies are also observed on gene activation. C, Heterochromatin formation is associated with gene repression. This is mediated, in part, by erasure of activating histone modifications (H3K4me1, H3K27ac, and 5hmC), establishment of repressive marks, such as H3K27me3 and 5mC, and nucleosomal compaction. DNMT, DNA methyltransferase; HAT, histone acetyltransferase; HDAC, histone deacetylase; MBD, methyl-DNA binding domain protein; MED, mediator complex; PcG, polycomb group complex; POL II, RNA polymerase II; TET, ten eleven translocation dioxygenase; TF, transcription factor; TrxG, trithorax group complex.

Enhancers are highly variable between lineages and cell types and confer cell-type–specific or lineage-restricted gene expression.88,94,96 Interestingly, only a small percentage of either class I or II enhancers overlap with CpG islands, and it is estimated that 94% of lineage-restricted enhancers are CpG poor and depleted for 5mC.88,89 Furthermore, most lineage-
restricted enhancers have low CpG density and are marked by H3K27ac in ESCs and cells from the early germ layers. These same enhancers are enriched for H3K27me3 and DNA methylation in terminally differentiated cells, which negatively correlates with expression of their target genes.88

Super enhancers are a recently identified class of enhancers that play a major role in determining cellular identity.97,98 They are characterized as large genomic regions (≥50 kb) possessing enhancer-like features: enrichment of H3K4me1, H3K27ac, Mediator complex, and transcription factors. However, unlike typical enhancers, they contain unusually high levels of Mediator complex proteins and transcription factors and display enrichment for cohesin and Nipbl (Nipped-B homolog), a protein involved in enhancer–promoter communication.97 The ~200 super enhancers are often observed as clusters of smaller enhancers enriched for transcription factor motifs involved in cell identity.97,98 Therefore, genes regulated by super enhancers are highly cell type specific and are usually expressed at higher levels than genes regulated by typical enhancers. For example, nearly all genes associated with ESC identity (eg, POU5F1 [OCT4], SOX2, NANOG, and KLF4) are regulated by super enhancers.98
Epigenetic Dynamics During Differentiation

The epigenome of PSCs is dramatically different from that of somatic cells. When comparing histone modification and DNA methylation patterns between differentiated cells and ESCs, Hawkins et al.\(^9\) estimate that approximately one third of the genome differs in chromatin structure. Many of these differences can be attributed to differential transcription factor binding and chromatin state remodeling during differentiation.\(^9\)\(^6\)\(^9\)

For example, differentiated cells have larger regions marked by H3K9me3 and H3K27me3 when compared with ESCs.\(^9\) Most of these H3K9me3 and H3K27me3 domains are significantly larger (2–3-fold) in fibroblasts and T lymphocytes when compared with ESCs, whereas H3K36me3 domains are similar between ESCs and differentiated cells.\(^9\)\(^6\)\(^9\)\(^9\) These large expanded H3K27me3 domains are remodeled to an ESC-like state during reprogramming.\(^9\)\(^6\) This suggests that the expansion of H3K27me3 domains accompanies cellular differentiation.

Bivalent domains in ESCs often transition to H3K27me3 domains that expand in lineage-committed cells such that H3K27me3 affects ≈40% of the genome in differentiated cells compared with only ≈8% of the ESC genome.\(^10\) Differentiation results in a progressive, global H3K27me3 enrichment that occurs in a cell-type–specific manner. These results identified Polycomb-mediated repression as an important mechanism for cell fate determination and lend support to the long-standing idea that differentiation results in an increasingly restrictive chromatin landscape.\(^8\)\(^8\)\(^9\)\(^9\)\(^6\)\(^9\)\(^6\) The majority of bivalent domains observed in undifferentiated ESCs resolve to a monovalent status in cells committed to a given lineage.\(^9\)\(^6\)\(^6\) Bivalent domains found in promoters encoding transcription factors typically resolve to either H3K4me3 only or H3K27me3 only in a lineage-specific manner.\(^9\)\(^6\)\(^6\) A clear example of this can be found during cardiomyocyte differentiation from ESCs. Many key transcription factors, such as GATA6, GATA4, NKX2.5, HAND2, TBX5, and signaling molecules (BMP2 and WNT5A) involved in cardiac development gradually lose H3K27me3 and gain H3K4me3 concomitantly with transcription in a stage-specific manner as differentiation proceeds.\(^10\)\(^10\)\(^10\) Conversely, later stages of cardiac differentiation require EZH2–mediated epigenetic repression of progenitor-specific transcription factors for cardiomyocyte maturation.\(^10\)\(^10\)\(^10\) Loss of H3K4me1/me3 is usually accompanied by the gain of DNA methylation.\(^9\)\(^6\)

Transcription factor motifs are particularly enriched for gains of DNA methylation during germ layer specification. For example, the number of transcription factor motifs that gain DNA methylation are ≈2-fold greater in the definitive endoderm than in either the early ectoderm or the mesoderm.\(^9\)\(^6\) Loss of transcription factor occupancy during differentiation may trigger DNA methylation; for example, putative regulatory elements downstream of the proneural gene DBX1 (developing brain homeobox 1) gain DNA methylation in the endoderm and mesoderm but not in the ectoderm where this gene is expressed.\(^9\)\(^6\)

In contrast, many poised enhancers become active (assessed by the loss of H3K27me3, gain of H3K27ac, and acquisition of RNAPII) during the transition from pluripotency to germ layer specification.\(^9\)\(^6\) Bruneau and colleagues have found that the majority of enhancers are poised at any given stage during cardiomyocyte differentiation. However, a small group of enhancers become active in a cell-type–specific manner. For example, the enhancers that become active in cardiac progenitors regulate genes essential for heart morphogenesis and development, whereas later during differentiation there is enrichment for genes encoding cardiac structural and functional proteins.\(^10\)

Loss of DNA methylation at regulatory elements is also common during germ layer formation and lineage commitment and is especially prevalent in ectoderm formation relative to the other germ layers.\(^27\)\(^8\) Similar to the idea of poised regulatory elements, there are examples of lineage-specific epigenetic priming during development. Specifically, many regions that are associated with fetal and adult brain development exhibit high levels of DNA methylation in ESCs. These regions are demethylated and resolve to an H3K4me1 state early during ectoderm differentiation yet remain methylated in the other lineages.\(^9\)\(^6\) Many of the genes that acquire H3K4me1 do not show immediate expression changes and, therefore, can be said to be primed for transcription at later stages. There are similar examples of epigenetic priming in cardiogenesis and lymphogenesis, suggesting that epigenetic priming is a common feature of development.\(^10\)\(^4\)\(^10\)\(^10\) Other examples of priming include distal CpG-poor regions that exhibit high DNA methylation levels in ESCs are demethylated and transition to H3K27me3 only in a lineage-specific manner. These regions are generally required at later stages of development in that particular lineage,\(^9\)\(^6\) further emphasizing the dynamic nature of histone-mediated epigenetic regulation when compared with DNA methylation.

Genes that temporally regulate cellular identity during differentiation can also be the subject to cyclic methylation events. For instance, a subset of genes necessary for neuronal progenitor fates are demethylated in neural precursors, expressed in progenitors, and then remethylated and repressed in later cell types, such as neurons.

Epigenetic Effects on Nuclear Organization and Genome Topology

At the most basic level, chromatin can be viewed as beads on a string where DNA (string) is coiled around nucleosomes (beads) with ≈165 bp periodicity to form the 11-nm fiber. Subsequent compaction generates the 30- and 300-nm fibers, ≈700-nm chromosomal domains, and ≈1400-nm mitotic chromosomes.\(^10\) During interphase, each chromosome occupies its own defined space within the nucleus, referred to as a chromosome territory. This allows nonrandom juxtapositions or physical interactions between genomic regions both in cis (intrachromosomal) and trans (interchromosomal). These interactions are characterized using chromosome conformation capture (3C) techniques and their derivatives, such as 4C, which permits interrogation of multiple genomic contacts at a single locus, and Hi-C or Chromatin Interaction Analysis by Paired-End Tag Sequencing (CHIA-PET), which have the power to identify many associations, thereby providing information on genome topology.\(^10\) Chromosomal regions of high gene density and high transcriptional activity tend to loop out of their respective chromosome territories and indiscriminately associate in trans with each other. Unlike associations of active regions, subnuclear movement of inactive regions is more locally constrained and hence
these interactions tend to be intrachromosomal within a chromosome territory core.108

Somatic nuclei possess a nuclear organization different from pluripotent cells, and these differences are largely reflected by chromatin state.109 Pluripotent cells contain less heterochromatin in general and less chromatin associated with the nuclear lamina and nucleoli than differentiated cells. Lamina-associated domains (LADs) are a class of stable chromatin domains found at the nuclear periphery that typically possess low levels of gene expression.110 There are ≈1300 LADs in mammalian cells ranging in size from 0.1 to 10 Mb. Coincident with their transcriptional status, LADs are enriched for the heterochromatic marks, H3K9me2/3 and H3K27me3, and their borders usually contain CpG islands and binding sites for the chromatin insulator, CCCTC-binding factor (CTCF).110 The interaction pattern and number of LADs seem to be cell type specific; however, some LADs are common across cell types.111 During differentiation, specific LADs containing genes destined for activation in the differentiated progeny translocate from the nuclear periphery to the interior. Conversely, genes to be repressed in the next cell type condense and are often relocated to the periphery where they interact with the nuclear lamina.111 Interestingly, genome reorganization occurs sequentially during differentiation. For example, specific genomic regions are relocated from the nuclear lamina during differentiation of ESCs to neural precursor cells.111 These regions remain spatially constant during the differentiation of neural progenitor cells to astrocytes, but a new, distinct group of LADs relocates away from the periphery during this subsequent differentiation step. Repressed genes that relocate away from the nuclear lamina become unlocked and are primed/ poised for activation later during differentiation, whereas active genes that shuttle to the nuclear lamina become locked and stably repressed.

In addition to LADs, mammalian genomes contain ≈2000 large domains called topologically associating domains, which are megabase-sized local interaction domains that inhibit spreading of heterochromatin in cis.112 They are generally bounded by invariant CTCF-binding sites, tRNAs or small interspersed elements, and contain coordinately regulated enhancers and promoters that cluster together. The CTCF-binding sites demarcate the boundaries of a large number of topologically associating domains are highly conserved between human and mouse and between cell types. Thus, they are unlikely to play a role in the establishment of cell-type–specific genomic topology during differentiation. Subregions within each topologically associating domain, however, seem to be dynamic.112 These dynamic regions may confer cell-type–specific genome organization via transcription factor complex–es bound at the regulatory elements of active genes (Figure 4).

The genome of PSCs is divided into a large complex pattern of chromatin loops bounded by CTCF-binding sites.113 These loops can be classified according to epigenetic modifications found within the loop and at the loop boundary.111 Loops are generally enriched for either active or repressive histone modifications inside the loop in a mutually exclusive manner although some loops do not seem to be enriched for any particular histone modification. Interestingly, ≈20% of loops are enriched for enhancer marks (H3K4me1/2) within the loop, promoter marks (H3K4me3) at the loop boundary, and marks for active transcription (H3K36me3), and repressed chromatin (H3K27me3) outside the loops on opposite sides. CTCF and cohesin function together to form chromatin loops that segregate heterochromatin from euchromatin.113,114 An example of chromatin dynamics that occur during differentiation can be found at the HOXA locus. In ESCs, the HOXA locus consists of bivalent chromatin.114 On differentiation, chromatin reorganization results in demarcation of a euchromatic and a heterochromatic domain at a highly conserved CTCF-binding site bound by both CTCF and cohesin.114

It has become increasing clear that interactions between transcription factors and Mediator complex bound to regulatory elements influence genome topology through their interactions with CTCF and cohesin. For example, Oct4 has been shown to interact with CTCF during the establishment of XCI and the maintenance of bivalent chromatin.114,115 Conversely, loss of Oct4 expression during differentiation permits the CTCF/cohesin-mediated heterochromatin formation at the HOXA locus114 mentioned earlier. Interactions between Mediator and cohesin at enhancers have been shown to regulate gene expression in a cell-type–specific manner by regulating the formation of chromatin loops that bring distal enhancers into close proximity with promoters of active genes.113 Interestingly, both active and poised enhancers participate in looping interactions in differentiated cells, suggesting that chromatin looping may precede and influence epigenetic priming116 during differentiation. Promoter clusters play a central role in chromosome topology but their contribution seems to not be tissue specific because promoters have an equally probable chance of interaction in ESCs and in somatic cells. Surprisingly, high-resolution analyses have determined that promoter states are largely invariant and highly stable between diverse cell types.80,90 Furthermore, the majority of promoter clusters exhibit activity in multiple cell types.99 However, enhancers contribute to genome topology in a more tissue-specific manner.95,117,118 For example, in ESCs, the SOX2 promoter interacts with a cluster of enhancers that are specifically active in ESCs, whereas a different set of enhancers is associated with the SOX2 promoter in neural progenitor cells.118,119 In general, cell type–specific enhancers are highly methylated in cells of other lineages and a direct correlation exists between DNA demethylation and enhancer

**Figure 4. Schematic of cell-type–specific genome topology.** Genetic loci (depicted as colored circles) cluster together making both intra (Chr A:Chr A) and inter (Chr A:Chr B) chromosomal interactions in a cell-type–specific manner. This establishes coregulated transcription of multigene networks that confer cellular identity. Differing topological interaction patterns are expected to be associated with different cell types, such as embryonic stem cells (cell type 1) and progenitor cells (cell type 2).
activation. Inactive enhancers are highly methylated and become demethylated in a tissue-specific manner during development. Interestingly, in addition to the cell-type-specific transcription programs regulated by enhancers, differential enhancer usage may also regulate common sets of promoters. Indeed, cell-type-specific enhancers have been shown to interact with promoter clusters common to multiple cell types, suggesting that different cell types use specific enhancers to regulate broadly expressed housekeeping genes.

During reprogramming to iPSCs, cells acquire a spatial genomic organization specific to pluripotent cells. In ESCs, genes involved in the maintenance of pluripotency (ie, Dppa5, Zfp42, Zfp281, Lefty, and Lin28) make interchromosomal contacts with the Nanog locus, suggesting that pluripotency genes aggregate together. Reprogramming-induced reactivation of the Nanog locus results in the establishment of pluripotency-associated gene contacts that are not observed at the Nanog locus in somatic cells. Furthermore, in ESCs, it has been shown that the majority of genes associated with reprogramming are spatially connected to the SOX2 locus within 1 large gene cluster, implying they may be coordinately regulated within an active structure termed a transcription factory.

Imprinting and XCI

Two specialized modes of epigenetic regulation are imprinting and XCI. Both rely on combinatorial epigenetic silencing (ie, a complex interplay of repressive histone modifications, CpG methylation, and noncoding RNA modulation of chromatin structure). Genomic imprinting is parent-of-origin allele-specific expression thought to have arisen to allocate extraembryonic nutrient resources to the developing fetus. The number of imprinted genes is a matter of debate. Estimates in mice range from ≈150 to >1000 imprinted genes, but recent work indicates that there are probably not many more than ≈200. The majority of imprinted genes contain DMRs that possess parent-of-origin-specific DNA methylation patterns (CpG and non-CpG) at enhancers and promoters that confer allele-specific expression. Many imprinted genes are expressed in the brain in a temporal manner to regulate neurodevelopment. Loss of imprinting, either through gain or loss of DNA methylation (or genetic deletion of regulatory elements/DMRs), results in biallelic gene silencing or expression that has been implicated in cancer and many neurodevelopmental disorders, including Beckwith–Wiedermann Syndrome, Prader–Willi Syndrome, and Angelman Syndrome. Genomic imprints are known to biallelic gene silencing or expression that has been implicated in cancer and many neurodevelopmental disorders, including Beckwith–Wiedermann Syndrome, Prader–Willi Syndrome, and Angelman Syndrome. Genomic imprints are known to biallelic gene silencing or expression that has been implicated in cancer and many neurodevelopmental disorders, including Beckwith–Wiedermann Syndrome, Prader–Willi Syndrome, and Angelman Syndrome.

XCI is an excellent example of many epigenetic modes acting synergistically to control gene expression via chromatin structure and chromosome topology. XCI is a means of dosage compensation between XY males and XX females. Shortly before embryo implantation, one of the X-chromosomes is randomly chosen for transcriptional repression. The earliest event in XCI is expression of XIST, a long noncoding RNA expressed from the X-inactivation center on the X-chromosome marked for inactivation. XIST then spreads along the X-chromosome beginning with regions either interacting with or in close proximity to the XIST locus followed by spreading to more distant regions. Specific domains of XIST recruit PRC2 to establish transcriptional repression. Together, PRC1 and DNA methylation maintain heritable XCI in somatic cells. In the naïve state of undifferentiated mESCs, both X-chromosomes are active (Xa/Xa). However, hPSCs generally exhibit XCI (Xa/Xi) on derivation. Recently, it has been shown that the inactive X in hPSCs often becomes reactivated over time in culture. Once reactivated, XCI is not re-established by differentiation. Erosion of XCI during time in culture can have profound effects on the use of female hPSCs for cell therapy, disease modeling, or drug screening. Careful analysis using SNP (single nucleotide polymorphism) genotyping combined with allele-specific reverse transcription quantitative polymerase chain reaction demonstrates that erosion or loss of XCI is common among late passage female hPSCs. Loss of XCI in undifferentiated hPSCs is associated with XIST repression, loss of H3K27me3, and derepression of genes silenced on the Xi. Once lost, the active state is stable and XCI cannot be re-established by differentiation.

Epigenetic Heterogeneity Among PSCs

Reprogramming of a somatic cell fate to a pluripotent state is largely an epigenetic process. Indeed, removal of certain epigenetic roadblocks either by chemical inhibition of DNA methylation or knockdown of MBDs during reprogramming enhances iPSC generation. In addition, inclusion of histone deacetylase inhibitors during exogenous transcription factor–based reprogramming and somatic cell nuclear transfer also increases the proficiency of PSC derivation. The 3 different types of PSCs, ESCs, iPSCs, and SCNT ESCs (derived by somatic cell nuclear transfer), display global methylation patterns that are similar to each other and different from somatic cells. The first in-depth study to examine the epigenomic differences among the 3 classes of mouse PSCs found that SCNT ESC global DNA methylation patterns were more similar to ESCs than iPSC patterns were to ESCs. This suggests that the ooplasm is more robust at reprogramming somatic genomes than transcription factor–based reprogramming. Now that human SCNT ESC lines are available, it will be interesting to compare the 3 hPSC epigenomes and contrast them with data generated from mouse-derived PSC types to determine whether conclusions drawn from this study are applicable to humans.

It is generally accepted that transcription factor–based reprogramming does not completely remodel the epigenome, at least initially. Multiple studies have shown that a certain degree of epigenetic memory is retained in early passage iPSCs, and this can lead to biased differentiation propensities...
depending on the donor cell source. This somatic memory is attributable to incomplete silencing of donor (source) cell-specific expression patterns. DNA methylation patterns account for the majority of somatic memory because histone modification patterns between iPSCs and ESCs exhibit minimal differences. During reprogramming, large regions of the genome fail to reset non-CpG methylation patterns, and many DMRs are specific to and common across iPSC lines when compared with ESCs. Recently, however, it has been demonstrated that these epigenetic memories account for only a small percentage of the variability among iPSCs and ESCs. In addition, it has been shown that expansion of iPSCs and time in culture diminishes the epigenetic differences between ESCs and iPSCs. This suggests that iPSC global epigenetic patterns stabilize over time although certain hotspots of variation remain, particularly at imprinted regions.27,142

H19, MEG3, and MEG3 are imprinted genes, such as MEG3.141 Furthermore, when DNA methylation patterns and gene expression profiles are considered together, the majority of iPSC lines cannot be distinguished from the majority of ESC lines.27,140 Together these studies suggest that the epigenetic heterogeneity observed in iPSC lines is not substantially different from the heterogeneity seen in ESC lines. Thus, when compared with ESC epigenetic patterns established during preimplantation development, it seems that iPSCs display a small amount of epigenetic variation, which can be attributed to reprogramming; however, hiPSCs and hESCs are essentially identical in their epigenetic profiles and their differentiation abilities.

Conclusions
We have discussed several of the best-understood epigenetic mechanisms and linked them to current knowledge of PSCs and cardiac lineage development wherever possible. We did not touch on nucleosome remodeling, histone variants, histone replacement, or noncoding RNA regulation of chromatin. This is a fast moving field, so our intention in this review is to provide a general background for appreciating future studies.

Epigenetics of Pluripotency and Differentiation

Epigenetic regulation of differentiation and cellular identity is critically important for the use of hPSCs and their derivatives in clinical applications. Evolving in vitro differentiation protocols increasingly incorporate knowledge about the normal processes of embryogenesis. Because protocols are fine-tuned to mimic morphogen gradients and subtle signaling events in the developing embryo, researchers are learning to develop hPSC-derived cell types that are more physiologically similar to their in vivo counterparts. Knowledge of cellular epigenetic states will allow researchers to develop cell types or combinations of cell types that faithfully represent human cells and tissues better. Ultimately, this knowledge will lead to tremendous improvements in predictive in vitro assays for drug toxicity and efficacy, which will greatly enhance the effectiveness and safety of clinical trials. Because regenerative medicine, in the form of cell replacement therapies, begins to be a viable medical treatment, epigenetic profiling will allow exquisite quality control of cells used for transplantation.

Acknowledgments
Because of the broad nature of this review and to length restrictions, we apologize to those investigators whose work was not cited. We thank the article reviewers for insightful suggestions. We thank members of the Loring lab for helpful comments and suggestions.

Sources of Funding
This work was supported by National Institutes of Health grant 5R33MH087925 and California Institute for Regenerative Medicine grants R1-01717, CL1-00502, RT1-01108, and TR1-01250 to J.F. Loring.

Disclosures
None.


36. Gowher H, Jeltsch A. Enzymatic properties of recombiant Dnmt3a DNA methyltransferase from mouse: the enzyme modifies DNA in a non-pro-


38. Guo J, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methyl-


41. Pastor WA, Pape UJ, Huang Y, et al. Genome-wide mapping of 5-hydroxy-


44. Szulwach KE, Li X, Li Y, et al. 5-hmC-mediated epigenetic dynam-

45. Wu H, Coskun V, Tao J, Xie W, Ge W, Yoshikawa K, Li E, Zhang Y, Sun YE. Dnmt3a-dependent nonpromoter DNA methylation facilitates trans-


Boland et al


Epigenetic Regulation of Pluripotency and Differentiation
Michael J. Boland, Kristopher L. Nazor and Jeanne F. Loring

Circ Res. 2014;115:311-324
doi: 10.1161/CIRCRESAHA.115.301517

Circulation Research is published by the American Heart Association, 7272 Greenville Avenue, Dallas, TX 75231
Copyright © 2014 American Heart Association, Inc. All rights reserved.
Print ISSN: 0009-7330. Online ISSN: 1524-4571

The online version of this article, along with updated information and services, is located on the World Wide Web at:
http://circres.ahajournals.org/content/115/2/311

Permissions: Requests for permissions to reproduce figures, tables, or portions of articles originally published in Circulation Research can be obtained via RightsLink, a service of the Copyright Clearance Center, not the Editorial Office. Once the online version of the published article for which permission is being requested is located, click Request Permissions in the middle column of the Web page under Services. Further information about this process is available in the Permissions and Rights Question and Answer document.

Reprints: Information about reprints can be found online at:
http://www.lww.com/reprints

Subscriptions: Information about subscribing to Circulation Research is online at:
http://circres.ahajournals.org//subscriptions/