Chromatin features and the epigenetic regulation of pluripotency states in ESCs

Wee-Wei Tee* and Danny Reinberg*

ABSTRACT
In pluripotent stem cells, the interplay between signaling cues, epigenetic regulators and transcription factors orchestrates developmental potency. Flexibility in gene expression control is imparted by molecular changes to the nucleosomes, the building block of chromatin. Here, we review the current understanding of the role of chromatin as a plastic and integrative platform to direct gene expression changes in pluripotent stem cells, giving rise to distinct pluripotent states. We will further explore the concept of epigenetic asymmetry, focusing primarily on histone stoichiometry and their associated modifications, that is apparent at both the nucleosome and chromosome-wide levels, and discuss the emerging importance of these asymmetric chromatin configurations in diversifying epigenetic states and their implications for cell fate control.

KEY WORDS: Nucleosomes, Epigenetic regulators, Chromatin, Pluripotency

Introduction
Embryonic stem cells (ESCs) possess the remarkable abilities of self-renewal and differentiation. They have the capacity to generate differentiated cells comprising all three embryonic germ layers. This pluripotent capacity makes them an excellent differentiated cells comprising all three embryonic germ layers. This pluripotent capacity persists only transiently for a few days in vivo; however, when explanted in vitro, further development is halted and pluripotency can be captured and propagated indefinitely in the form of ESCs.

Mouse ESCs were first isolated from the ICM compartment of blastocysts in 1981 (Evans and Kaufman, 1981), and are one of the earliest and best-studied prototypes of pluripotent cells. The derivation of human ESCs, also from the ICM of explanted blastocyst prior to implantation, came almost 20 years later (Thomson et al., 1998). Since then, different stem cells corresponding to distinct embryonic precursors and pluripotency sub-states in vivo have been described. They differ from the conventional ESC pluripotent state in their developmental origins, transcription factor and signaling requirements, as well as their epigenetic configurations (Table 1). For example, unlike mouse ESCs, conventional human ESCs exhibit a pronounced tendency for X-chromosome inactivation in female cells, and are generally less amenable to genetic manipulation (Buicer and Geijser, 2010; Hanna et al., 2010b). In this regard, it was postulated that conventional human ESCs might bear a greater resemblance to the mouse epiblast stem cells (EpiSCs), which originate from a developmentally more advanced post-implantation epiblast that has undergone X chromosome inactivation and cannot efficiently form germline chimeras when injected into blastocysts (Brons et al., 2007; Tesar et al., 2007). This diminished potency may be due to the differences in transcriptional and chromatin constituency between mouse EpiSCs and ESCs (Song et al., 2012).

The ability to derive stem cells of different molecular and phenotypic characteristics at different developmental stages in the mouse embryo has led to the idea that the pluripotent state is not invariant, but rather a continuum of states that can be modulated by extrinsic signaling cues, both in vivo and in vitro (Pera and Tam, 2010). Numerous studies have delineated the signaling principles central to the establishment and maintenance of these different pluripotency states (Ng and Surani, 2011). For example, the LIF (leukemia inhibitory factor)/STAT3 (signal transducer and activator of transcription 3) and FGF (fibroblast growth factor)/ERK (extracellular signal-regulated kinase; MAPK1)
Table 1. Different states of pluripotency

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Origin</th>
<th>Teratoma formation in mouse</th>
<th>Blastocyst chimerism in mouse</th>
<th>Extra-embryonic lineages</th>
<th>X chromosome status</th>
<th>Oct4 distal enhancer</th>
<th>DNA methylation, imprinting and H3K27me3</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>ESC – conventional (mouse and human)</td>
<td>Pre-implantation epiblast</td>
<td>Yes</td>
<td>Yes (mouse)</td>
<td>Inefficient (human)</td>
<td>No</td>
<td>Both active in females</td>
<td>Active</td>
<td>Prominent occupancy on lineage-specifying genes</td>
</tr>
<tr>
<td>ESC – groundstate (mouse and human)</td>
<td>Pre-implantation epiblast</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Both active in females</td>
<td>Both active in females</td>
<td>Active</td>
<td>Active</td>
</tr>
<tr>
<td>EpiSC (mouse)</td>
<td>Post-implantation epiblast</td>
<td>Yes</td>
<td>Inefficient</td>
<td>No</td>
<td>One inactivated</td>
<td>NA</td>
<td>Inactive</td>
<td>NA</td>
</tr>
<tr>
<td>EG (mouse)</td>
<td>Primordial germ cells</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Both active in females</td>
<td>NA</td>
<td>Active</td>
<td>NA</td>
</tr>
<tr>
<td>Male GSC (mouse)</td>
<td>Testis</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>NA</td>
<td>NA</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>Hex+ ESC (mouse)</td>
<td>ESC in 2i/LIF</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ND</td>
<td>NA</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>2C-ESC (mouse)</td>
<td>ESC in serum/LIF</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>ND</td>
<td>NA</td>
<td>ND</td>
<td>NA</td>
</tr>
<tr>
<td>rESC (mouse)</td>
<td>EpiSC reprogrammed to naive pluripotency</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Both active in females</td>
<td>NA</td>
<td>Active</td>
<td>NA</td>
</tr>
<tr>
<td>iPSC (mouse and human)</td>
<td>Somatic cells via reprogramming</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Both active in females</td>
<td>Both active in females</td>
<td>Active</td>
<td>Active</td>
</tr>
</tbody>
</table>

Different states of pluripotency can be attributed to different tissues of origin as well as to perturbations in the extrinsic and intrinsic cellular environment. Different pluripotency sub-states have varying developmental potential and epigenetic status. Conventional refers to serum- and feeder-containing culture, whereas ground state refers to serum- and feeder-free culture supplemented with chemical inhibitors against specific signaling cascades.

Signaling pathways converge onto the chromatin
Chromatin is the basic regulatory unit of the eukaryotic genetic material. It comprises repeating arrays of nucleosomes, each consisting of 147 bp of DNA wrapped around a histone octamer made up of two molecules of each histone: H2A, H2B, H3 and H4. Histones are classified as either canonical or variant, depending on their primary sequence and mode of deposition during the cell cycle. They are subjected to various post-translational modifications and demarcate different transcriptional domains in the genome (Fig. 1). The different combinations in which these modifications occur can significantly expand the regulatory properties of histone, beyond that of a structural scaffold upon which the DNA is wrapped (Kouzarides, 2007). This can occur in at least two ways: first, by directly altering the biophysical properties of the nucleosome particle through steric changes in histone-histone (within the same or neighboring nucleosomes) and/or histone-DNA interactions; and second, via the recruitment of effector proteins that directly recognize the modifications (Taverna et al., 2007).
Fig. 1. See next page for legend.
Fig. 1. Distinct transcriptional complexes and chromatin regulators mark active pluripotent and repressed developmental genes in ESCs.

Epigenetic marks (colored hexagons, see key) are bestowed by specific chromatin modifying proteins (colored ovals). Same color pairs denote an interaction, e.g. the H3K4me3 mark (green star) is imposed by ML2 and SET1 complexes (green ovals). (A) Hallmarks of an active transcriptional state include the presence of H3K4me3, histone H3/H4 acetylations such as H3K27ac (imposed by CBP/p300), and the assembly of RNA polymerase II (RNAPII) and associated regulatory proteins at the core promoter. Serine 5 (Ser 5P) phosphorylation of RNAPII is associated with transcription initiation, whereas serine 2 phosphorylation (Ser 2P) is associated with transcription elongation (Adelman and Lis, 2012). Active enhancers are marked by H3K27ac, H3K4me1, ELL3, RNAPII and enhancer-transcribed RNAs (eRNAs), whereas cohesin and mediator complexes act to regulate enhancer-promoter looping (Calo and Wysocka, 2013). ELL3 facilitates the assembly of SEC on the promoters of active genes for robust gene activation (Lin et al., 2013). Histone variants H2A.Z and H3.3 are enriched at promoters and enhancers, and the formation of a H2A.Z/H3.3 'hybrid' nucleosome may contribute to the dynamic nucleosome turnover at these sites. Other chromatin remodeling complexes such as esBAF, TIP60, CHD1/7 (Young, 2011) and PAD4 (Christophorou et al., 2014) are also involved in setting up the active transcriptional state. (B) Developmental gene promoters are marked by the presence of both H3K4me3 and H3K27me3. NuRD-mediated histone deacetylases promote H3K27me3 deposition by PRC2, whereas SETD8B1, a H3K9 histone methyltransferase, further represses several developmental genes (Young, 2011). RNAPII-Ser5P is evident on these developmental promoters, whereas RNAPII-Ser2P is not. Short non-coding RNAs may be transcribed at the promoter-proximal regions, potentially leading to the recruitment of PRC2 (Kanhere et al., 2010; Voigt et al., 2013). PRC1 reinforces gene repression through histone H2A ubiquitylation as well as via other non-enzymatic mechanisms, including nucleosome compaction and interference with the transcription process (Simon and Kingston, 2013). Developmental gene enhancers are marked by the presence of H3K27me3 and H3K4me1, but not H3K27ac. Despite the presence of CBP/p300, these enhancers are incapable of promoting gene activation in pluripotent cells, but will do so during differentiation, upon the loss of enhancer H3K27me3 marks (Rada-Iglesias et al., 2011). (C) Silent genes generally show reduced nucleosome dynamics, meaning they are more 'compact', and show less extensive transcriptional regulator binding. Silent genes are repressed by chromatin regulators that methylate histone H3K9 and/or H3K27me3. Promoter DNA sequences may also be methylated by DNA methyltransferases, DNMTs, giving rise to 5-methyl-cytosine (5mC). CHD1/7, chromodomains helicase DNA-methyl binding protein 1/7; ELL3, elongation factor RNA polymerase II-like 3; esBAF, ES cell-specific SWI/SNF-like ATP-dependent chromatin remodeling complex; HAT, histone acetyltransferase; HMT, histone methyltransferase; ML, mixed-lineage leukemia; PAD14, peptidyl arginine deiminase type IV; PRC1/2, polycomb repressive complex 1/2; SEC, super elongation complex; SET1A/B, SET domain-containing 1A/B; SETDB1, SET domain, bifurcated 1; TIP60, K(lysine) acetyltransferase 5; TFs, transcription factors.

Expanded states of pluripotency

The reduced level of DNA methylation and altered PRC2 occupancy in the ‘2i-LIF’ mouse ESC model may be indicative of the epigenetic foundation of groundstate pluripotency in vitro. These cells are also thought to parallel the early epiblast cells of the iCM in vivo, wherein the DNA methylation level is lowest following the development of the zygote to blastocyst (Smith et al., 2012), concomitant with imprinted X chromosome reactivation in female embryos (Mak et al., 2004; Okamoto et al., 2004). Interestingly, a recent finding demonstrates that a small fraction of mouse ESCs grown under this specific culture regime co-expresses markers of both embryonic and extra-embryonic lineages (termed Hex+ ESCs), and are able to generate both trophectoderm and epiblast lineages when reintroduced into embryos (Morgan et al., 2013). These cells clearly exhibit an expanded cellular potency in a chimeric setting, when compared with conventional ESCs, which do not typically contribute towards the trophectoderm lineage. The possibility that alleviation of particular restrictive epigenetic modifications by ‘2i-LIF’ may provoke expansion of cellular plasticity is consistent with the notion that lineage specification is generally accompanied by stabilizing epigenetic configurations. However, a second study provided a surprising twist. It was observed that conventional mouse ESCs grown in the presence of serum and LIF could transit through a previously undescribed stem cell state that resembles that of the totipotent 2-cell (2C) blastomeres in vivo (Macfarlan et al., 2012). The authors reported a striking upregulation of retrotransposons, such as endogenous retroviruses (ERVs), that accompanied the expression of 2C-specific protein-coding genes. Interestingly, many of the latter generated chimeric transcripts linked to the retrotransposons, suggesting that they had co-opted the regulatory elements of neighboring retroelements for coordinated expression, a property also likely exploited in the human ESC transcriptional circuitry (Kunaro et al., 2010). The expression of ERVs continues to persist in a sub-population of in vitro cultured ESCs, termed 2-cell-like ESCs (2C-ESCs). Although these 2C-ESCs are rare and transitory in nature, most, if not all, of the ESCs cultured in vitro are capable of passing through this peculiar state at some point. These 2C-ESCs have a transcriptional and epigenetic profile similar to that of the...
Epigenetic regulation underlies transcriptional heterogeneity in stem cells

Epigenetic regulators play a central role in controlling transcriptional output and variability. These regulators include the chromatin remodeling complexes, such as SWI/SNF and NuRD, that act to regulate nucleosome spacing and dynamics, as well as histone-modifying complexes such as histone deacetylases (HDACs) and the PRCs. Studies using different model organisms have shown that many of these chromatin regulators are involved in regulating transcriptional variability (Lehner et al., 2006; Raj et al., 2010). In mice, haploinsufficiencies in Dnm3a and Trim28 have been shown to increase phenotypic variation in inbred littermates (Whitelaw et al., 2010). In the case of serum-cultured ESCs, heterogeneous expression of pluripotency factors is postulated to regulate phenotypic plasticity (Chambers et al., 2007; Hayashi et al., 2008; Toyooka et al., 2008). However, it is noteworthy that this phenomenon is less pronounced in the ‘groundstate’ ESCs (Wray et al., 2011). At the molecular level, heterogeneous gene expression must be imparted by differential modulation of chromatin architecture (Hayashi et al., 2008). Although unrestrained transcriptional fluctuations are likely to be deleterious to cell viability, it is conceivable that cells may benefit from some degree of gene expression heterogeneity as a means to increase phenotypic variance during lineage specification and for survival fitness in general (Huang, 2009). In this regard, it is perhaps of no coincidence that a similar repertoire of epigenetic regulators is involved in the transition of pluripotent ESCs into the 2C totipotent-like state.

The equilibrium and frequency of sub-state switching enforced by signaling pathways must invariably feed back into events that occur at the level of the enhancer and promoter. For example, studies have shown that the disassembly of promoter nucleosomes is rate limiting for gene expression (Boeger et al., 2008), and that competition between nucleosome and promoter proximal occupation of RNA polymerase II can fine-tune gene expression (Gilchrist et al., 2010). These and other studies indicate that the assembly of the transcriptionally competent chromatin state is a highly dynamic process that can be regulated on several fronts, thus affording various opportunities for the fine-tuning of gene expression. In serum-cultured ESCs, the NuRD complex appears to be especially important for promoting transcriptional heterogeneity of pluripotency genes and lineage commitment (Reynolds et al., 2012a). NuRD harbors both chromatin remodeling and histone deacetylase activities, mediated through different constituents in the complex (Zhang et al., 1998, 1999). Situated at the heart of this complex is methyl-CpG binding protein 3 (MBD3), which serves as a scaffold for the assembly of other components (Kaji et al., 2006, 2007). Notably, loss of MBD3, and hence NuRD, function in mouse ESCs leads to elevated expression of select pluripotency genes that impedes lineage commitment (Reynolds et al., 2012a). Several of these NuRD-repressed pluripotency genes, e.g. Tbx3 (T-box 3), Klf4 (Kruppel-like factor 4) and Foxd3 (forkhead box D3) are also targets of the PRCs (Walker et al., 2010). Notably, in a separate study by Brooks et al., these genes are classified as ‘PRC2 active’, based on their apparent active transcription yet paradoxical presence of PRCs on their promoters in the context of a heterogenous ESC population (Brookes et al., 2012). Further analysis revealed an independent association of PRCs with the elongative form of RNA polymerase II on these genes (Fig. 1), suggesting that these two components may exist separately within a cell (i.e. on different alleles) or in distinct cell populations. It was also postulated that the heterogenous expression of key pluripotency genes, such as Nanog might impact on this dynamic (Brookes et al., 2012).
Mechanistically, NuRD-mediated deacetylation of H3K27 may facilitate the recruitment of PRC2 to repressed lineage genes, including these ‘PRC2-active’ genes (Brookes et al., 2012; Reynolds et al., 2012b). In this instance, it will also be of interest to assess the impact of NuRD on the global nucleosome occupancy in ESCs, particularly at enhancers and promoters where NuRD resides (Reynolds et al., 2012b; Whyte et al., 2012; Yildirim et al., 2011). This may reveal further insights into how the interplay between nucleosome remodeling and RNA polymerase II pausing (Gilchrist et al., 2010) may contribute to the dynamic gene expression regulation in ESCs (Henikoff, 2008).

The aforementioned group of NuRD-PRC2 co-target genes also make up part of the extended network of pluripotent transcription factors that function to stabilize the core OCT4, NANOG and SOX2 regulatory circuitry (Walker et al., 2011). Thus, an important role of NuRD in ESCs may be to modulate expression of these ‘extended’ pluripotency genes, thereby destabilizing the pluripotency network and lowering the threshold required for cells to respond to differentiation cues and exit the self-renewing state. Consistent with this, depletion of MBD3 during iPSC generation yields unprecedented reprogramming efficiency, effectively eliminating stochastic effects in this otherwise inefficient process (Rais et al., 2013); however this effect may be context-dependent (dos Santos et al., 2014). In addition, UTX [lysine (K)-specific demethylase 6A; KDM6A], a histone H3K27me3 demethylase, is required during iPSC generation to remove trimethyl marks from H3K27me3 in the case of early expressed pluripotency genes (Mansour et al., 2012). Notably, the loss of PRC2 and LIN53, a component of the NuRD complex, is an absolute pre-requisite for triggering germ to somatic cell conversion in C. elegans (Patel et al., 2012; Tursun et al., 2011), which collectively points to an important role of these transcriptional repressors in safeguarding lineage fidelity. However, it is interesting to note that the loss of PRC2 apparently impedes iPSC generation even when using the optimized homogeneous ‘secondary’ reprogramming cells (Buganim et al., 2012; Onder et al., 2012; Pereira et al., 2010), suggesting that loss of PRC2 may result in additional deleterious effects during nuclear reprogramming.

The ESC epigenetic landscape

Genome-wide charting of DNA and histone modifying enzymes, along with their associated modifications in stem cells and differentiating tissues have given rise to chromatin-modification ‘maps’, which may be used to identify DNA elements of regulatory importance (Gifford et al., 2013; Stergachis et al., 2013; Xie et al., 2013; Zhu et al., 2013). This wealth of epigenetic information on the chromatin is further organized spatially, in a three-dimensional manner in relation to the nuclear structure (Jin et al., 2013; Phillips-Cremins et al., 2013). Several epigenetic features have been extensively described in ESCs (Young, 2011) and some of these are highlighted in Fig. 1.

Bivalent chromatin domains

The preponderance of the bivalent H3K4me3-H3K27me3 chromatin domain is arguably one of the more distinctive epigenetic characteristics described in ESCs and has received considerable attention (Azuara et al., 2006; Bernstein et al., 2006). Bivalent domains are historically described as chromatin stretches comprising oligonucleosomes that harbor both the activating H3K4me3 and the repressive H3K27me3 marks. At the time of the discovery, no distinction was made as to whether the marks co-exist on a single nucleosome particle or on different nucleosomes. In addition, it is important to note that bivalent domains are not exclusive to stem cells as they also occur in differentiated tissues, albeit at lower frequency. While the factors involved in the establishment of this bivalent state are well defined, the exact function of this chromatin configuration remains debatable, in part due to a lack of an appropriate genetic experimental model (reviewed by Voigt et al., 2013). Towards this end, two groups have recently identified mixed-lineage leukemia 2 (MLL2) as the enzyme that directs H3K4me3 deposition on bivalent promoters (Denissov et al., 2014; Hu et al., 2013a). Whereas loss of MLL2 abolished H3K4me3 specifically on bivalent promoters, it did not significantly alter the transcriptional kinetics of these genes upon retinoic acid treatment. This calls into question the purported role of bivalency in priming gene expression (Box 1), and also highlights a redundant mechanism involving other H3K4 methyltransferases, such as MLL1. It was also proposed that MLL2 engages unmethylated CpG islands and functions primarily as a pioneer H3K4 methyltransferase to mark bivalent promoters, with the presence of PRCs somehow blocking the engagement of SET1 H3K4 methyltransferases known to deposit H3K4me3 at active genes (Denissov et al., 2014). However, it will be important to further examine the gene expression kinetics of Mll2-null as well as Mll2/2 double-knockout ESCs in other physiological settings that are representative of development in vivo, e.g. in the context of germ layer differentiation. This is pertinent given that both Mll2 and PRC2 knockout mice are embryonic lethal, with PRC2 mutants showing defects during gastrulation (reviewed by Voigt et al., 2013), and that Mll2-null embryoid bodies are defective in select HoxB gene inductions (Glaser et al., 2006).

Bivalent domains have also been detected in mammalian gametes (Ng et al., 2013; Sachs et al., 2013), although it remains unclear whether the marks co-exist on the same nucleosome. Nevertheless, the retention of residual H3K4me3-K27me3-marked nucleosomes in the sperm genome and the observation that genes marked by higher levels of

**Box 1. Are bivalently marked asymmetric nucleosomes functionally relevant?**

The existence of asymmetrically modified nucleosomes may afford greater flexibility in gene expression regulation. Although inherently attractive, this hypothesis still requires rigorous testing on several fronts. Some of these key questions listed below may be addressed using either in vitro reconstituted transcription assays or genome-editing tools to engineer a ‘reporter’ for bivalent promoters in vivo. Examples of the latter include the recently described chromatin in vivo assay (CIA), a powerful tool that allows for the specific perturbation of histone modification patterns in a locus-specific manner (Hathaway et al., 2012), as well as the use of programmable transcription activator-like effector (TALE) repeat fusions (Maeder et al., 2013; Mendenhall et al., 2013) and/or CRISPR-mediated editing (Chen et al., 2013a; Cong et al., 2013).

**Key questions**

- How do the kinetics of demethylases differ for (a)symmetrically modified nucleosomes?
- What is the impact of an (a)symmetric configuration on nucleosome turnover/stability?
- Does the spatial inequality of post-translational modifications affect the binding affinities of effector proteins that can potentially recognize combinations of post-translational modifications?
- Does the distribution and composition of (a)symmetrically modified nucleosomes change in response to developmental stimuli and/or cell cycle progression, as in the case of heterotypic H2A-H2A.Z nucleosomes?
- Are (a)symmetrically modified nucleosomes differentially distributed on regulatory elements, e.g. promoters versus enhancers, and does this exert a differential impact on gene expression priming?
- Does perturbing ‘bivalency’ in vivo affect the kinetics and inheritance of gene expression patterns during development?
Chromatin organization in pluripotent cells

In ESCs, the bivalent chromatin domains are embedded in a wider context of a highly dynamic and accessible chromatin landscape, which is postulated to constrain PRC2 occupancy and H3K27me3 spreading (Zhu et al., 2013). Indeed, it is well established that pluripotent cells have a more dynamic chromatin structure compared with their differentiated counterparts (Gasper-Maia et al., 2011). For example, undifferentiated ESCs contain a lower abundance of constitutive heterochromatin, marked by fewer DAPI-dense foci (Efroni et al., 2008). Pluripotent chromatin is also marked by strategic deposition of histone variants at key regulatory elements, as well as a dynamic binding of structural chromatin proteins such as HP1 (Meshorer et al., 2006). The regulated displacement of linker histones H1 at active regulatory regions also contributes towards chromatin decondensation (Christophorou et al., 2014). These chromatin features are likely to be important for attaining the transcriptionally pervasive pluripotent state (Efroni et al., 2008). Notably, some of these chromatin features are also observed in pluripotent cells in vivo, such as the ICM (Ahmed et al., 2010), and are recapitulated during nuclear reprogramming (Fussner et al., 2011; Mattout et al., 2011).

The plasticity of ESCs is often attributed to hyperdynamic chromatin features. However, whether permissive chromatin structure is instructive for pluripotency or merely a consequence thereof, remains an open question. Emerging evidence suggests it is likely a combination of both. For example, the extent of heterochromatin compaction in ESCs and ICM cells inversely correlates with NANOG expression (Fussner et al., 2011). More importantly, loss of OCT4 is sufficient to trigger ectopic formation of silenced heterochromatin domains in the ICM cells (Ahmed et al., 2010). Using OCT4 as a further example, an additional study showed that the differential kinetics of OCT4 binding in the early mouse embryo might contribute toward lineage allocation, and it was postulated that this differential binding might be regulated by the state of chromatin accessibility (Plachta et al., 2011). This study thus implies a role for local chromatin structure in regulating transcription factor occupancy. Interestingly, a reduced level of OCT4 in mouse ESCs apparently results in enhanced self-renewal, comparable with that observed in ‘2i-LIF’ (Karwacki-Neisius et al., 2013). Somewhat counter-intuitively, this overall reduction in OCT4 level is accompanied by an increase in OCT4 and NANOG occupancies at pluripotency-associated enhancers. This phenomenon reflects either an active redistribution of OCT4 to key regulatory sites and/or an inherent tighter, or additive association of OCT4 at these specific genomic regions. On a speculative note, it is plausible that certain chromatin features may underlie these ‘attractor’ sites, and in the case of the early embryos, may contribute in part to the differential kinetics of OCT4-chromatin binding (Plachta et al., 2011). One such chromatin feature may be the presence of architectural proteins such as cohesin. It is well documented that transcription factor binding often occurs in a clustered fashion that is orchestrated by the cohesin complex. Depletion of cohesin results in reduced DNA accessibility and transcription factor binding, indicating a causative role for cohesin in promoting the formation of these transcription factor clusters (Yan et al., 2013). Cohesin may thus help to confer robustness in gene expression by helping to stabilize highly occupied cis-regulatory modules (Faure et al., 2012) (Fig. 1). This is consistent with the findings in ESCs that pluripotency factors such as OCT4, NANOG and KLF4 interact and cooperate with cohesin complex in a highly coordinated fashion to access and regulate expression of target loci (Apostolou et al., 2013; Wei et al., 2013). As noted previously, it will be important to assess transcription factor binding occupancy in the context of local chromatin features, and how global changes in nucleosome dynamics correlate with transcription heterogeneity and developmental specification (Li et al., 2012; Teif et al., 2012).

The binary relationship between transcription factor binding and open chromatin may be overly simplistic. Recent evidence suggests that OCT4 itself can function as pioneer factor, readily accessing regions of closed chromatin structure (Soufi et al., 2012). In fact this property is not merely restricted to pluripotency factors, but also exhibited by numerous lineage-specific regulators (Zaret and Carroll, 2011). Although pioneer factors can bind to nucleosomal DNA, additional determinants are likely in place to direct target specificity in a context-dependent manner. A recent study aimed at understanding the role of ASCL1 in somatic cell-neuron transdifferentiation identified a ‘trivalent’ chromatin feature that may help guide this pioneer factor to target regions (Wapinski et al., 2013). Notably, this trivalent signature, comprising H3K9me3, H3K4me1 and H3K27ac, putatively within a single nucleosome, is present only in cell types that are amenable to neuronal transdifferentiation, perhaps serving as a means to identify other recipient cells that are permissive for neural reprogramming. This work raises the exciting possibility that certain chromatin features can instruct transcription factor-target specificity, paving the way for future research to assess the universality of such a ‘chromatin-guided’ mechanism.

Breaking nucleosome symmetry and diversifying epigenetic regulation

Role of histone variants

Given the importance of nucleosome dynamics in gene regulation, what then are the molecular attributes that affect nucleosome turnover and/or remodeling? Among many factors, histone variants may be particularly well poised to fulfill such a role. Unlike canonical histones such as H2A and H3.1/2, which are primarily expressed during S phase and deposited in a strict DNA replication-dependent manner, histone variants such as H2A.Z and H3.3 are expressed and deposited throughout the cell cycle, uncoupled from DNA synthesis (Campos and Reinberg, 2009). Histone variants are also subjected to different post-translational modifications and as such could be selectively deployed to alter local epigenetic configuration through replacement of canonical histones at any point during the cell cycle (Hake et al., 2006; Loyola et al., 2006; McKittrick et al., 2004). When incorporated, histone variants can alter the overall stability of the nucleosome particle, regulating nucleosome turnover at key regulatory elements in the genome (Bonisch and Hake, 2012; Chen et al., 2013; Henikoff, 2008; Jin and Felsenfeld, 2007). Some histone variants have also evolved highly specialized functions, whereas others are expressed in a tissue-specific manner (Talbert and Henikoff, 2010). Interestingly, in certain cases the regulatory function of histone variants can be compensated by upregulation of other related histone isoforms, indicative of functional redundancy (Lin et al., 2000; Montellier et al., 2013). In fact, a recent report suggested that canonical histones bearing particular post-translational modification combinations could compensate for the lack of a particular histone variant, highlighting remarkable plasticity in the process (Montellier et al., 2013). In this case, depletion of a testis-specific H2B variant, TH2B (histone cluster 1, H2ba; Hist1h2ba), during spermatogenesis induced compensatory mechanisms via the
upregulation of canonical H2B, as well as via the additional installation of specific post-translational modifications on histones H3 and H4, to drive the completion of histone-protamine exchange. It is likely that such crosstalk between histone genes is necessary to ensure robust completion of key biological processes and therefore is presumably employed in other developmental settings. Along this line, a similar mechanism may also operate in mouse ESCs depleted of H3.3. Notably, despite the prevalent deposition of histone H3.3 at key regulatory and genomic sites, only modest transcriptional abnormalities were observed in H3.3-depleted mutant ESCs that successfully differentiated into all three germ layers during teratoma formation (Banaszynski et al., 2013). This highlights the possibility that, as in the case of TH2B, canonical H3 (or post-translational modification changes on other histones) might rescue, in part, the lack of H3.3 at affected loci (Banaszynski et al., 2013; Goldberg et al., 2010).

The engagement of histone variants significantly expands the flexibility of epigenetic control in maintaining pluripotency as well as during development. Consistent with the crucial importance of histone variants in modulating gene expression is their frequent presence at regulatory elements in the genome, coinciding with promoters and enhancers, many of which are also ‘hotspots’ of transcription factor binding. For example in mouse ESCs, H2A.Z occupies both active and inactive regions, including distal regulatory elements (Hu et al., 2013b; Ku et al., 2012). Notably, H2A.Z depletion leads to reduced chromatin accessibility, concomitant with impaired MLL, PRC2 and OCT4 targeting (Hu et al., 2013b). This has led to the view that H2A.Z may act in a general manner to modulate local chromatin accessibility, thereby facilitating efficient targeting of transcriptional activating or repressing complexes to their respective sites in the genome. As mentioned previously, H3.3 also maps to both active and inactive regulatory elements, although, unlike H2A.Z, the loss of H3.3 deposition specifically impacts targeting of the polycomb, but not that of the trithorax machinery (Banaszynski et al., 2013).

Regions enriched for H2A.Z/H3.3 often coincide with regions of nucleosome depletion (Chen et al., 2013c; He et al., 2010), indicative of active nucleosome exchange and/or remodeling (Fig. 1). The observation that H2A.Z, when in complex with H3.3, constitutes a highly unstable nucleosome in vivo is relevant here, highlighting that a key regulatory role for H2A.Z might be to amplify the intrinsic instability of H3.3-containing nucleosomes (Jin and Felsenfeld, 2007) (see Fig. 3). This ‘hybrid’ configuration, as initially referred to by the authors to draw attention to the co-existence of two distinct histone variants in the same nucleosome particle, maps to nucleosome-free regions of active promoters and other regulatory regions (Jin et al., 2009). Although these studies were performed in somatic cells, the findings very likely also pertain to ESCs, given that both cell types share common features with respect to genomic regions co-occupied by H2A.Z and H3.3.

**Histone variants and tetramer splitting**

In addition to promoting nucleosome turnover, another interesting feature of H3.3-containing nucleosomes is that a small but significant proportion of them undergo a higher incidence of (H3.3-H4)_2 tetramer splitting in vivo compared with H3.1/2-containing nucleosomes that are highly stable (Fig. 2). In yeast, where the single H3 isoform is homologous to H3.3, (H3-H4)_2 tetramer splitting can be observed on transcriptionally active loci (Katan-Khaykovich and Struhl, 2011). The same is observed in mammalian cells, although the frequency of (H3.3-H4)_2 splitting appears to correlate more with enhancer specificity and potentially during DNA replication (Huang et al., 2013; Xu et al., 2010). Considering that ESCs have a significantly larger number of epigenetically marked enhancers compared with differentiated lineages (Stergachis et al., 2013; Whyte et al., 2013), it will be interesting to examine the frequency and specificity of (H3.3-H4)_2 tetramer splitting as a function of differentiation, and whether pluripotency and/or lineage-specific transcription factors, either acting in isolation or in collaboration with histone chaperones, may play a role in orchestrating this process. Such analysis is important given the realization that it is the chromatin state of enhancers, rather than that of promoters, that better reflects cell type-specific activities (Heintzman et al., 2009; Hnisz et al., 2013; Rada-Iglesias et al., 2011; Visel et al., 2009). Additionally, many studies have now shown that epigenetic patterning of enhancers occurs prior to cell fate specification, and that this enhancer-priming event permits differential access of signaling molecules and transcription factors that drive cell type-specific gene expression programs (Bergsland et al., 2011; Liber et al., 2010; Mullen et al., 2011; Trompouki et al., 2011). The finding that a given transcription factor binds first to the enhancer before potentiating epigenetic changes at the promoter during reprogramming, further attests to the instructive nature of enhancer marking in directing cell fate changes (Taberlay et al., 2011). Therefore, understanding the
The dynamics of tetramer splitting and identifying the players that regulate this process will reveal novel insights into how enhancers participate in cell fate specification. Indeed, if \((H3.3-H4)\) tetramer splitting is restricted to active enhancers during DNA replication, it may provide a semi-conservative means for active propagation of these instructive regulatory elements through cell divisions, an exciting possibility that warrants further testing.

**Nucleosome stoichiometry: homotypic versus heterotypic**

The aforementioned studies demonstrating the impact of nucleosome composition in gene expression are an appropriate reminder of the fundamental importance of nucleosome stoichiometry in gene regulation, a feature perhaps overshadowed by the overwhelming attention on histone tail post-translational modifications. A prevailing view is that the octameric nucleosome particle is symmetrical, comprising identical pairs of each of the four core histone proteins. This type of nucleosome composition is referred to as homotypic. However, the observation that the nucleosome core particle can actually accommodate a non-identical \(H2A-H2A.Z\) pair, a composition referred to as heterotypic, has challenged this view (Fig. 3). In the case of \(H2A-H2A.Z\) heterotypic nucleosomes, although initial structural studies did not support the feasibility of such a conformation (Suto et al., 2000), their presence has been detected *in vivo*, in organisms spanning from yeast, fly and mouse to human (Luk et al., 2010; Nekrasov et al., 2012; Viens et al., 2006; Weber et al., 2010). Importantly, as shown in mouse trophoblast stem cells, the composition of the \(H2A.Z-H2A\) nucleosome particle changes dynamically throughout the cell cycle, with heterotypic and homotypic nucleosomes localizing to different regions of the genome and assuming different functions (Nekrasov et al., 2012). Thus, heterotypic nucleosomes represent one spectrum of nucleosome asymmetry wherein the two copies of sister histones are clearly non-equivalent. It is conceivable that other histone proteins, in particular \(H3-H3.3\), may be similarly engaged given their overall similarity in amino acid composition and identical interaction surface within the tetramer (Luger et al., 1997) (Fig. 3). Importantly, the presence of a mixture of homotypic and heterotypic \(H3\) nucleosome populations can potentially influence the choice between two different modes of nucleosome deposition (Ahmad and Henikoff, 2002) and how active transcriptional states, e.g. primed enhancers, are epigenetically inherited. Understanding the precise mechanisms of how histone chaperones regulate the deposition and/or exchange of histones may reveal further insights into how the different permutations of histone stoichiometry are established (Fig. 3). Given our previous discussion on the regulatory roles of ‘hybrid’ \(H2A.Z/H3.3\) nucleosomes, it will be interesting to examine the full histone stoichiometry in this case, to assess whether sister histones are homotypic or heterotypic. If different ‘hybrid-heterotypic’ permutations were possible, then this would greatly expand the number of epigenetic conformations, in addition to the epigenetic diversity made possible by various combinations of post-translational modifications (Fig. 3).

**Asymmetrically modified mononucleosomes**

Theoretically, histone variant-imposed nucleosomal asymmetry can be further amplified by the presence of non-identical post-translational modifications carried by each sister histone (Fig. 4). Indeed, as shown by Voigt et al., asymmetrically modified nucleosomes exist *in vivo*, in both pluripotent and differentiated cells (Voigt et al., 2012). By using an affinity-purification-based mass spectrometry (MS) methodology, the authors were able to assess in a quantitative manner the distribution of post-translational modifications on \(H3\)
Epigenetic asymmetry: moving beyond the mononucleosome

ESCs largely divide in a symmetrical fashion with equal partitioning of genetic and epigenetic materials, giving rise to functionally equivalent daughter cells. However, a recent report documents that altering the cell polarity in ESCs can perturb this symmetry (Habib et al., 2013). When subjecting mouse ESCs to localized WNT3A signaling, the authors observed that the ESCs underwent asymmetric division instead. Remarkably, daughter cells proximal to the source of WNT3A had higher expression levels of pluripotency genes, whereas distal cells acquired differentiation hallmarks. This is reminiscent of adult stem cells in vivo that mostly undergo asymmetric cell divisions to generate a self-renewed progenitor and a daughter cell destined for differentiation (Clevers, 2005). A central question relating to this process is how is the parental epigenetic information maintained in the progenitor stem cell, but altered in the differentiating cell? Studies using Drosophila male germline stem cells (GSCs) as a model system have revealed some fascinating insights into this process. Male GSCs undergo asymmetric cell division, giving rise to a self-renewed GSC and a daughter cell gonialblast that undergoes differentiation. By using a dual-color labeling strategy to distinguish between ‘old’ and ‘new’ histones, Tran et al. discovered that pre-existing old canonical H3, but not H3.3, histones are specifically retained in the GSCs, whereas newly synthesized H3 histones are enriched in the differentiating gonialblast cells. Notably, this asymmetric mode of H3 distribution is lost when GSCs are experimentally manipulated to divide symmetrically (Tran et al., 2012). Mechanistically, this asymmetric distribution of histones can occur if (1) pre-existing H3 and newly synthesized H3 are already differentially localized onto the two sets of sister chromatids during S phase prior to mitosis, and (2) the mitotic machinery can somehow distinguish between these epigenetically distinct sister chromatids. It is unclear at this moment how the
different epigenetic states between the two sister chromatids are established and subsequently distinguished, although it is tempting to speculate that similar principles governing X chromosome inactivation and imprinting in mammals may be operating in this instance. Further insight into this elusive process may be gained from the observation that asymmetric epigenetic regulation is required to generate neuronal bilateral asymmetry in *C. elegans*. In a truly remarkable fashion, it has been shown that a single mutation in one of the 24 *C. elegans* H3 genes could abolish the left-right neuronal symmetry owing to defective H3-H4 tetramer formation and nucleosome assembly (Nakano et al., 2011). This finding led to the intriguing hypothesis that in wild-type worms, unequal loading of nucleosomes onto the leading and lagging DNA strands during DNA replication gives rise to sister chromatids marked with differing nucleosome densities at discrete loci required for neuronal specification. In this regard, the difference in nucleosome density may henceforth constitute a bona fide epigenetic feature that serves as an initial ‘mark’ to distinguish between the two sister chromatids, which is subsequently transmitted following mitosis.

Does the non-random segregation of epigenetically distinct sister chromatids also occur in mammalian stem cells? Some studies do support such an occurrence. Klar et al. reported that mouse ESCs and neuroectoderm cells, but not other cell types tested, show varied segregation patterns of chromosome 7 (Armakolas and Klar, 2006). However, a separate group showed that in a subpopulation of adult mouse skeletal muscle stem cells, apparently all sister chromatids (not just select chromosomes), could undergo biased segregation (Rocheteau et al., 2012). To what extent and significance this asymmetric inheritance of sister chromatids may regulate cell fate decisions in mammalian stem cell remains to be fully defined.

Concluding remarks

In this Review, we have discussed the emerging importance of epigenetic regulators in governing lineage fidelity and explored how the manipulation of chromatin factors can perturb the equilibrium between discrete pluripotent states. The increasing ease in genome editing and the means to engineer lineage-specific reporters in human ESCs in a high-throughput fashion have provided exciting inroads into our understanding of how epigenetic barriers are established during distinct phases of mammalian development, and how they are overcome during nuclear reprogramming (Cong et al., 2013; Poser et al., 2008; Wang et al., 2013). Mapping of histone post-translational modifications as a function of lineage specification has proven to be extremely informative, and has been the primary method of choice to assess genome-wide chromatin states. However, most studies performed to date do not provide information about the exact histone composition within the nucleosomes or distinguish between modifications on sister histones.

The view that a nucleosome is symmetrical must be revised in light of observations demonstrating the functional differences between heterotypic and homotypic nucleosomes, and how asymmetric presentation of post-translational modifications on sister histones can impact the kinetics of histone-modifying enzymes. Clearly, the extent to which these nonequivalent nucleosomal states exist *in vivo* remains to be investigated, but they may indeed be the norm rather than the exception. The continual discovery of new histone variants, e.g. through alternative splicing, adds to the growing list of possible nucleosome conformations (Bonisch et al., 2012; Rasmussen et al., 1999). Last but not least, an additional layer of complexity that warrants attention is the finding that histone modifications, as well as nucleosomes, can often be asymmetrically positioned around discrete transcription factor binding sites, with skewed enrichment on one side of the binding site, but not on the opposite end (Kundaje et al., 2012). It is unclear how this directionality impacts on gene regulation, but a clue to this may be gleaned from the recent description that a class of pioneer transcription factors apparently exerts a stronger effect in chromatin opening on one side of their motif than on the other (Sherwood et al., 2014). In conclusion, epigenetic asymmetry is manifested in different forms, from within the most fundamental nucleosome particle, to larger regulatory transcription factor binding units, as well as the inheritance of epigenetically distinct sister chromatids, collectively attesting to the remarkably complex realm of gene expression regulation.

Acknowledgements

We thank Drs Eric Campos, Matthias Stadtfeld, Maria Elena Torres-Padilla and Lynne Vales for critical reading of the manuscript.

Competing interests

The authors declare no competing financial interests.

Funding

Work in the Reinberg laboratory is supported by grants from the National Institutes of Health (NIH), New York State Stem Cell Science (NYSTEM) and the Howard Hughes Medical Institute. W.-W.T. is a New York Stem Cell Foundation-Druckenmiller Fellow. This research was supported by The New York Stem Cell Foundation. Deposited in PMC for release after 12 months.

References


