View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Searching Chromatin for Stem Cell Identity

Michael Buszczak[1] and Allan C. Spradling[1,*]
[1]Howard Hughes Laboratories and Embryology Department, Carnegie Institution of Washington, 3520 San Martin Drive, Baltimore, MD 21218, USA
*Contact: spradling@ciwemb.edu
DOI 10.1016/j.cell.2006.04.004

Stem cells encapsulate the fundamental problem of metazoan biology in miniature: How do cells establish and maintain their fates? Increasing evidence indicates that stem cell chromatin activates proliferation genes and represses differentiation genes. Understanding how these configurations are stabilized by Polycomb group proteins will advance our understanding of embryonic development, tissue homeostasis, regeneration, aging, and oncogenesis.

Stem cells play key roles throughout the life cycle of most multicellular eukaryotes. These cells maintain the broad potential of growing embryos and ensure that adult tissues can produce new cells when needed. But what gives stem cells their special properties? Many studies have highlighted the extrinsic intercellular signals required to maintain embryonic stem (ES) cells in culture and adult stem cells within their tissue-specific niches (reviewed in Chambers and Smith, 2004; Li and Xie, 2005). Manipulation of such molecules can dramatically expand or reduce stem cell populations, disrupt normal tissue architecture, bias differentiation decisions, and promote cancer. However, non-stem cells frequently also respond to these common developmental signals. To search for the elusive keys to stem cell identity, many researchers have turned to the study of stem cell chromatin. A number of recent reports, including two in this issue of *Cell* (Lee et al., 2006; Bernstein et al., 2006) and one in this week's *Nature* (Boyer et al., 2006), are beginning to define the chromatin organization, gene-expression profiles, and regulatory mechanisms that distinguish stem cells from their more mature counterparts.

Understanding how stem cells are specified by chromatin factors is complicated by our limited knowledge concerning which molecular processes actually determine cellular identity and why developmental potential generally declines and becomes restricted to a single cell fate prior to adulthood. Cellular identity or "state" cannot simply be equated to all the factors that govern a cell's transcriptional profile because stem cells and stable differentiated cells can all express variable gene profiles as they traverse the cell cycle and respond to environmental factors. Thus, we do not know whether cellular state is determined by some biochemically distinct subset of nuclear activities or whether it is a complex system property. Answering two questions would be helpful: Where are key chromatin factors located within stem cell genomes, and do changes in particular molecular elements correlate with changes in cell state? The new studies by Lee et al. (2006), Bernstein et al. (2006), and Boyer et al. (2006) are beginning to address these issues by mapping the location of specific chromatin proteins and histone modifications at high resolution over large segments of the genome.

Polycomb group (PcG) proteins and the modifications they catalyze represent particularly attractive candidates on which to focus such an inventory. In the fruit fly *Drosophila*, a complex interplay of repression by PcG proteins and activation by trithorax group (trxG) proteins plays a central role in maintaining epigenetic patterns of gene expression (reviewed in Ringrose and Paro, 2004). These mechanisms are substantially conserved in mammals. Intriguingly, there is also evidence that the action of PcG proteins stabilizes stem cells but goes awry in cancer cells (reviewed in Valk-Lingbeek et al., 2004). PcG proteins form two major complexes, although additional variations and developmental modulations remain to be fully understood (Kuzmichev et al., 2005). The PRC2 complex, comprising EED, EZH2, and SUZ12 in mammals, initiates gene silencing and catalyzes histone H3 methylation on lysine 27 (H3K27) at target loci (Kirmizis et al., 2004). The more complex and variable PRC1 is then recruited, in part by the presence of H3K27me3, where it helps to maintain transcriptional repression either through chromatin compaction or by interfering with transcription initiation. In *Drosophila*, *cis*-acting Polycomb response elements (PREs) that bind both to PcG and trxG proteins mediate epigenetic regulation. PRE elements often reside within the promoters of PcG targets, and collections of these *cis*-acting elements may extend throughout the gene or gene cluster (Ringrose and Paro, 2004). Corresponding regulatory sequences that mediate PcG-dependent epigenetic inheritance in mammalian genomes have not yet been identified.

Embryonic Stem Cells

ES cells arise from the inner cell mass of the mammalian blastocyst and can be maintained in culture in a pluripotent state. Their uniformity and potential value for regenerative medicine make them an attractive model for analyzing stem cell chromatin. The genes encoding several transcription factors including OCT4, NANOG, and SOX2 are turned on shortly after fertilization and are required both to specify inner-cell-mass cells and to derive ES

cell lines (reviewed in Chambers and Smith, 2004). PRC2 activity appears to be required as well because loss of *Ezh2* or *Suz12* leads to a marked loss of cell proliferation in the inner cell mass and early embryonic lethality. Moreover, attempts to generate ES cells with mutations in *Ezh2* have been unsuccessful. These observations suggest that a distinct cellular identity develops in ES progenitor cells and that PRC2 is required to maintain that fate.

New insight into the nature of this state comes from mapping the location of OCT4, NANOG, and SOX2 binding sites throughout the ES cell genome at high resolution (Boyer et al., 2005). Each transcription factor occupies the promoters of more than 1000 genes, including many that are required for cell growth and division. About 350 target genes are bound by all three factors, and these include many genes encoding other transcription factors and chromatin modification enzymes. About half of these common loci are actively transcribed in ES cells, including the *Oct4*, *Nanog*, and *Sox2* genes themselves, as well as BMP and JAK/STAT signaling components that the cells likely require to respond to growth factors in the culture medium. Many of the bound inactive genes encode transcription factors expressed in differentiating cells that might be antagonistic to pluripotency. Thus, OCT4, NANOG, and SOX2 are high-level transcription regulators that cooperate with other factors to establish a stem cell chromatin state in ES cell progenitors by activating or repressing multiple target genes.

These findings suggest that the role of PcGs may be no different in ES cells than in more differentiated cells where their function was first studied. PcG proteins may simply maintain the program of gene repression established in ES cell precursors by OCT4, NANOG, and SOX2. If so, PcG proteins should be found at the sites of inactive differentiation-promoting genes. Genome localization studies of a PRC2 component, *Suz12*, now strongly support this prediction (Lee et al., 2006). SUZ12 binds at approximately 1800 sites in the human ES cell genome that are usually also associated with EED and H3K27me3. These targets frequently correspond precisely to gene promoters. Moreover, the target loci are candidate differentiation genes because most lack bound polymerase II and are produced at lower levels in ES cells than in more differentiated cells. As predicted, nearly all of the repressed genes co-occupied by OCT4, SOX2, and NANOG are also associated with SUZ12.

Now, Boyer et al. (2006) present further evidence that PRC1 and PRC2 act to keep differentiation genes silent in ES cells. PRC1 components (Phc2, Rnf2) and PRC2 components (Suz12, Eed) as well as H3K27me3 were all localized at the promoters of 512 genes in mouse ES cells, many of which encode homeodomain proteins and other putative differentiation factors. In *eed* mutant ES cells (where PRC2 and H3K27 methylation is disrupted), the expression of 87% of a sample of these target genes was increased, compared to only 13% of control genes. Moreover, in neural derivatives of ES cells, there was increased expression of a subset of the target genes encoding known neural specification factors, whereas PcG binding and H3K27me3 levels at the promoters of these genes was reduced. Thus, PcG proteins make an essential contribution to the ES cell state by repressing the premature expression of differentiation genes in a manner flexible enough to be reversed later by gene-specific and lineage-specific signals.

Conserved DNA Elements and PcG-Mediated Repression in ES Cells

The typography of SUZ12 binding in human ES cells reported by Lee et al. (2006) is reminiscent of PcG binding to *Drosophila* target genes in several respects (see Ringrose and Paro, 2004). For example, in both organisms, PcG binding sites at target genes sometimes extend from the promoter far into the gene. In clusters of repressed genes, such as the four human *HOX* gene clusters, SUZ12 was found to bind to contiguous regions of up to 100 kb in human ES cells (Lee et al., 2006). In *Drosophila*, study of such sites led to the identification of PRE sequences, which display a weak sequence consensus and act in *cis* to mediate epigenetic inheritance by interacting with PcG and trxG proteins. The PRC2 target sites that have now been identified in ES cells provide hints that *cis*-acting DNA elements may also be involved in PcG-mediated repression in mammals. Mammalian genomes are known to contain many highly conserved noncoding DNA sequence elements of unknown function (Woolfe et al., 2005). Such elements are associated with many of the regions bound by SUZ12. It should be interesting to learn whether any of these sequences can confer PcG-mediated regulation in *cis*.

Further insight into the possible roles of *PcG* and *trxG* genes at conserved mammalian sequence elements comes from detailed mapping of the H3K27me3 and H3K4me3 methylation marks in mouse ES cells (Bernstein et al., 2006). About 200 genomic regions (including the four *Hox* gene clusters) contain most of the highly conserved noncoding elements (HCNEs) (Woolfe et al., 2005). When 61 large domains that encompass HCNEs were analyzed, large regions ($\sim$10 kb) modified by H3K27me3 and containing shorter segments of H3K4me3 within them were preferentially observed in HCNE-rich regions. These structures, termed "bivalent domains," were often associated with transcription-factor genes or other developmentally significant genes whose expression is low in ES cells compared to more differentiated cells. Another interesting sequence feature of bivalent domains is a tendency to lack transposon-derived sequences. About 50% of the bivalent domains contain a predicted binding site for OCT4, NANOG, or SOX2 based on the data of Boyer et al. (2005). These observations further support a role for *cis*-acting sequence elements in mediating a chromatin-based ES cell state that is maintained by the regulated action of *PcG* and *trxG* genes.

Transcription and Epigenetic Inheritance

The observation that an interplay of *PcG* and *trxG* genes at promoters may be important in maintaining epigenetic states is consistent with our knowledge of epigenetic inheritance in *Drosophila*. Continued association with trxG proteins is required to maintain the activity of particular *Hox* genes within the Hox clusters (reviewed by Ringrose and Paro, 2004). Transcription of PREs has been found to counteract silencing (Schmitt et al., 2005). Indeed, trithorax group proteins may function by binding to PREs and facilitating such transcription.

Recently, evidence in support of a simple molecular mechanism of epigenetic inheritance based on transcription has been reported (Mito et al., 2005). Transcribed genomic regions replace H3-containing nucleosomes, which are only produced during S phase, with nucleosomes containing the ubiquitously expressed histone variant H3.3. Biochemical differences between the two classes of nucleosomes could serve to maintain patterns of transcription in daughter cells. PcG and trxG proteins might use this method to maintain transcription patterns at PREs and thus to maintain states of gene activity (Mito et al., 2005).

Adult Stem Cells

How similar is the cellular state of adult stem cells to that of ES cells? Like ES cells, female germline stem cells in *Drosophila* express growth and proliferation genes and repress differentiation genes (Kai et al., 2005). Transcripts could not be detected from any of the *HOX* genes or from virtually any of the zygotic transcription factors that are expressed during embryonic differentiation. This program also may not represent a default state. Two zinc-finger protein isoforms with activator or repressor activity that are encoded by the *ovo* locus may be involved, as *ovo* is required for female germ cell survival (see Bielinska et al., 2005). Like ES cells in culture and mouse germ cells in vivo, *Drosophila* germline stem cells differentiate in the absence of ongoing BMP signaling (see Li and Xie, 2005). Adult germline stem cells may be biologically closer to ES cells than other adult stem cells, however (Guan et al., 2006).

Do other adult stem cells exhibit a similar cellular state to ES cells and germline stem cells that must be maintained by the action of PcG proteins? These questions have been difficult to answer experimentally due to problems with purifying these rare cells. However, this model might explain the requirement of multiple stem cell types for PcG proteins (reviewed in Valk-Lingbeek et al., 2004). Several adult mouse stem cells, including hematopoietic and neural stem cells, require the PRC1 component BMI-1. In mice lacking *Bmi-1*, hematopoietic stem cells are formed during development but are not maintained later in life. Likewise, *Bmi-1* does not appear to be needed to form neural stem cells but instead is required for their self-renewal during later life (Molofsky et al., 2003; Leung et al., 2004). Forced expression of *Bmi-1* expands the number of multipotent hematopoietic progenitors and is associated with the development of medulloblastomas. These observations could be explained if PcG-mediated repression is weakened in the absence of Bmi-1, leading to premature stem cell differentiation and loss.

Alternatively, *Bmi-1* might play a distinct role, with or without other PRC1 members, that is specifically required by stem cells. One idea is that *Bmi-1* is needed to repress the expression of the *Ink4a/Arf* cell-cycle repressor locus, which otherwise would serve as a safety factor by placing an upper limit on cell proliferation (reviewed in Valk-Lingbeek et al., 2004). Other *Bmi-1* targets probably also exist because the effects of *Bmi-1* mutation on stem cells are not completely reversed by deleting the *Ink4a/Arf* gene (Bruggeman et al., 2005). Regardless of a possible special function of the BMI-1 protein in adulthood, general repression of differentiation by PcG proteins may contribute to maintaining adult stem cells. Recently, it was shown that proteins associated with chromatin remodeling complexes are required to sustain both germline and epithelial stem cells in *Drosophila* (Xi and Xie, 2005). Remodeling complexes are known to interact with some PcG proteins.

Counteracting PcG-Mediated Repression in Stem Cell Daughters

If PcG proteins maintain the stem cell state, mechanisms must exist to alter their activity to allow cellular differentiation. Downstream of the *Drosophila* male germline stem cell, a special system of alternative basal transcription factors antagonizes PRC1 action (Chen et al., 2005). Shortly after departing their niche, young male germ cells induce testis-specific isoforms of the general transcription factors known as TATA-associated factors (TAFs). TAFs are known to form complexes with PcG proteins, and testis-specific TAFs (tTAFs) are required to express key genes that promote spermatogenesis. Many tTAF targets are bound by Polycomb (PC), both at the promoter (where tTAFs bind) and extending farther downstream. As primary spermatocytes develop prior to meiosis, there is an inverse relationship between tTAF binding and PC binding. Eventually, substantial amounts of PC protein, as well as the PRC1 components Polyhomeotic (PH) and dRING, leave the site of tTAF target genes and accumulate in the nucleolus. Relief of repression requires the action of *trithorax* (*trx*), because trimethylated H3K4 marks accumulate at tTAF-dependent target genes, and both target gene expression and chromatin modifications are reduced in *trx* mutants (Chen et al., 2005). Special TAFs are also produced in the mammalian testis (Pointud et al., 2003), suggesting that similar events may occur in mammalian germ cells.

A Derepression Model of Cell-Fate Determination

These observations suggest a simple model in which cell fates are determined by the selective derepression of a subset of differentiation genes. Key high-level transcription factors such as OCT4, NANOG, and SOX2

would program the embryonic stem cell state by establishing an initial hierarchy of active and inactive genes. Polycomb group genes would then maintain the transcriptional repression of differentiation genes within the stem cells and early progenitors until lineage-appropriate genes become activated via the action of intercellular signals and trxG proteins. Three new studies now provide evidence supporting selective derepression of PcG target genes during the differentiation of muscle, nerve, and other lineages (Lee et al., 2006; Bernstein et al., 2006; Boyer et al., 2006). Each type of differentiated cell might maintain in a silent state many of the genes originally repressed in the ES cell. Alternatively, the initial changes might lead to so many additional induced chromatin changes that evidence of a common initial state might be obscured. Despite its excessive simplicity, however, this model suggests that, to a first-order approximation, the targets of OCT4, NANOG, and SOX2 in ES cells represent the "cell state" genes of the mammalian embryo.

Do Stem Cells Have a Special Requirement for PcG Genes?

All cell fates, not just stem cells, may depend on the continued operation of *PcG* and *trxG* genes. Such a requirement is indicated by classic studies of transdetermination, in which errors in cell fate occur at a greatly elevated frequency in *Drosophila* imaginal disc cells that have been stimulated to grow after wounding. The frequency of transdetermination is greatly increased in PcG mutant flies (Lee et al., 2005). Transdetermined cells often exhibit downregulation of *PcG* genes through the action of the Jun kinase signaling pathway, which is strongly induced by wounding. The downregulation of *PcG* genes by JNK signaling appears to be conserved in mammals, suggesting that this pathway may facilitate cell-fate switching in a wide range of organisms (Lee et al., 2005).

So is the requirement of stem cells for *PcG* genes special or general? What explains the striking instability of the ES cell state (pluripotency) compared to the states of more mature cells, including adult stem cells? The ES cell state may be more sensitive to disruption by gene misexpression simply because a larger number of differentiation genes must be maintained in a silent state. However, it seems more likely that the process of PcG-mediated gene repression itself is more labile in early embryonic cells than in older cells. ES cell chromatin has been reported to be more dynamic than chromatin in differentiated cells (Meshorer et al., 2006). Changes in the structure of PcG complexes or the identity of interacting proteins may occur that stabilize gene repression as development proceeds. Some components of PcG complexes are known to differ between embryonic and adult cells in *Drosophila*, for example (Ringrose and Paro, 2004). Therefore, understanding the differences between embryonic and adult chromatin that control the strength of cell-state programming might allow us to further stabilize normal cellular states or facilitate the conversion of one cellular state into another.

REFERENCES

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). Cell, this issue.

Bielinska, B., Lu, J., Sturgill, D., and Oliver, B. (2005). Genetics *169*, 161–172.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Cell *122*, 947–956.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., et al. (2006). Nature. 10.1038/nature04733.

Bruggeman, S.W., Valk-Lingbeek, M.E., van der Stoop, P.P., Jacobs, J.J., Kieboom, K., Tanger, E., Hulsman, D., Leung, C., Arsenijevic, Y., Marino, S., and van Lohuizen, M. (2005). Genes Dev. *19*, 1438–1443.

Chambers, I., and Smith, A. (2004). Oncogene *23*, 7150–7160.

Chen, X., Hiller, M., Sancak, Y., and Fuller, M.T. (2005). Science *310*, 869–872.

Guan, K., Nayernia, K., Maier, L.S., Wagner, S., Dressel, R., Lee, J.H., Nolte, J., Wolf, F., Li, M., Engel, W., and Hasenfuss, G. (2006). Nature. Published online March 24, 2006. 10.1038/nature04697.

Kai, T., Williams, D., and Spradling, A.C. (2005). Dev. Biol. *283*, 486–502.

Kirmizis, A., Bartley, S.M., Kuzmichev, A., Margueron, R., Reinberg, D., Green, R., and Farnham, P.J. (2004). Genes Dev. *18*, 1592–1605.

Kuzmichev, A., Margueron, R., Vaquero, A., Preissner, T.S., Scher, M., Kirmizis, A., Ouyang, X., Brockdorff, N., Abate-Shen, C., Farnham, P., and Reinberg, D. (2005). Proc. Natl. Acad. Sci. USA *102*, 1859–1864.

Lee, N., Maurange, C., Ringrose, L., and Paro, R. (2005). Nature *438*, 234–237.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K.-i., et al. (2006). Cell, this issue.

Leung, C., Lingbeek, M., Shakhova, O., Liu, J., Tanger, E., Saremaslani, P., Van Lohuizen, M., and Marino, S. (2004). Nature *428*, 337–341.

Li, L., and Xie, T. (2005). Annu. Rev. Cell Dev. Biol. *21*, 605–631.

Meshorer, E., Yellajoshula, D., George, E., Scambler, P.J., Brown, D.T., and Misteli, T. (2006). Dev. Cell *10*, 105–116.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2005). Nat. Genet. *37*, 1090–1097.

Molofsky, A.V., Pardal, R., Iwashita, T., Park, I.K., Clarke, M.F., and Morrison, S.J. (2003). Nature *425*, 962–967.

Pointud, J.C., Mengus, G., Brancorsini, S., Monaco, L., Parvinen, M., Sassone-Corsi, P., and Davidson, I. (2003). J. Cell Sci. *116*, 1847–1858.

Ringrose, L., and Paro, R. (2004). Annu. Rev. Genet. *38*, 413–443.

Schmitt, S., Prestel, M., and Paro, R. (2005). Genes Dev. *19*, 697–708.

Valk-Lingbeek, M.E., Bruggeman, S.W., and van Lohuizen, M. (2004). Cell *118*, 409–418.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., et al. (2005). PLoS Biol. *3*, e7. Published online November 11, 2004. 10.1371/journal.pbio.0030007.

Xi, R., and Xie, T. (2005). Science *310*, 1487–1489.