Published in final edited form as:

Curr Opin Genet Dev. 2010 October ; 20(5): 519–526. doi:10.1016/j.gde.2010.06.010.

Pioneer Factors in Embryonic Stem Cells and Differentiation

Stephen T Smale

Department of Microbiology, Immunology, and Molecular Genetics, 6-730 MRL, 675 Charles E. Young Drive South, University of California, Los Angeles, CA 90095, USA

Abstract

Most studies of tissue-specific and developmental stage-specific transcription have focused on the DNA motifs, transcription factors, or chromatin events required for the active transcription of a gene in cells in which the gene is expressed, or for its active or heritable silencing in non-expressing cells. However, accumulating evidence suggests that, in multicellular eukaryotes, enhancers or promoters for tissue-specific genes interact with pioneer transcription factors in embryonic stem cells and at other early stages of development, long before the genes are transcribed. These early interactions, which can lead to the presence of unmethylated CpG dinucleotides, histone modification signatures, and/or chromatin remodeling, may carry out different functions at different classes of genes.

Introduction

Early studies of eukaryotic protein-coding genes revealed that properly initiated and regulated transcription depends on a core promoter, a distal promoter, and one or more distant control regions (e.g. enhancers, silencers, or locus control regions). Core promoter elements, such as the TATA box, Inr, and DPE, interact with components of the general transcription machinery, catalyze the formation of a transcription pre-initiation complex containing RNA polymerase II and several general transcription factors, and dictate the location of the transcription start site. In contrast, the distal promoter and distant control regions interact with sequence-specific DNA-binding proteins (i.e. transcription factors) that determine the cell types and physiological states in which the gene is actively transcribed. A general view is that efficient transcription occurs in cells in which a defined set of essential transcription factors and co-regulatory proteins are present and active.

A variety of experimental approaches have been used to identify transcription factors and co-regulatory proteins that contribute to the regulation of tissue-specific genes. Some key factors were identified on the basis of their expression patterns during development or in different tissues. Others were identified on the basis of their ability to bind DNA motifs in promoters or enhancers required for regulated transcription in cell lines or in an organism. Genetic and functional approaches have been equally valuable for the identification of factors that contribute to regulated transcription.

Although much has been learned about the factors and events needed for active transcription of tissue-specific genes in expressing cell types, relatively little is known about the developmental cascade of events that occur at specific loci prior to transcriptional activation.

Corresponding author: Smale, Stephen T (smale@mednet.ucla.edu).

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Although tissue-specific transcription factors can sometimes activate their target genes when ectopically expressed in related or unrelated cell types [1,2] a growing body of evidence suggests that efficient activation of each tissue-specific gene does not simply require the presence of a set of essential factors. Instead, the activation of each tissue-specific gene may require a defined cascade of protein-DNA interactions and chromatin transitions that begins during the earliest stages of development. Furthermore, the reprogramming of a differentiated cell to a pluripotent state may require the re-setting of each locus, such that it becomes associated with a defined set of factors and chromatin signatures that provide competence for transcriptional activation upon differentiation. In this article, I will discuss emerging insight into these molecular events.

Pioneer factor interactions in multipotent progenitor cells

It has long been known that protein-DNA interactions and chromatin changes can be observed at transcription control regions for tissue-specific genes early in development, long before the gene is transcribed. Studies of a hepatocyte-specific enhancer for the mouse *Alb1* (albumin) gene provided the first clear evidence of these early interactions. Genomic footprinting experiments initially revealed two protein-DNA interactions at the *Alb1* enhancer in multipotent definitive endoderm from day 9.5 embryos [3•,4]. The genomic footprinting technique will only reveal interactions that are present at most or all alleles in a cell population, demonstrating that the *Alb1* enhancer is usually or always bound by a limited number of proteins in definitive endoderm, before hepatic specification and *Alb1* activation. It was proposed that these early interactions may initiate chromatin decondensation and make the enhancer and downstream promoter susceptible to the binding of other transcription factors required for *Alb1* activation in hepatocytes [3•]. On the basis of this proposal, the term “pioneer factor” was coined to describe transcription factors that are among the first to access a tissue-specific locus at an early stage of development and initiate the cascade of events that culminates in transcriptional activation.

The protein-DNA interactions observed at the *Alb1* enhancer in endoderm are carried out by FoxA1 and GATA-4. FoxA1 contains a winged helix DNA-binding domain and GATA-4 binds DNA through two C4 zinc fingers [5,6]. The hypothesis that these proteins play an important “pioneering” role at the enhancer was supported by evidence that FoxA1 binds more stably to a nucleosome core particle than to naked DNA, and that FoxA1 binding to nucleosomal DNA can facilitate GATA-4 binding [7]. FoxA1 and GATA-4 not only can bind nucleosomal DNA, but they also have the ability to alter the local conformation of nucleosomes assembled *in vitro* in the presence of linker histones [7]. These activities appear to be unusual for sequence-specific DNA-binding proteins, suggesting that FoxA1 may be representative of a class of proteins that can access nucleosomal DNA and initiate chromatin changes that serve as an initial step toward transcriptional activation. As additional evidence that Fox proteins carry out functions that are fundamentally different from those of most other transcription factors, FoxA1 was found to be relatively immobile in the nucleus when analyzed by fluorescence recovery after photobleaching (FRAP) [8•].

Another well-studied tissue-specific gene that exhibits protein-DNA interactions at its control regions early in development, and prior to transcriptional activation, is the chicken lysozyme gene. Transcription of this gene begins at an early stage of myelopoiesis and continues through development, with substantially increased transcription following the activation of mature macrophages [9]. Several DNase I hypersensitive sites have been identified in the lysozyme locus that appear or disappear at specific stages of development and therefore are likely to coincide with regulatory regions [9,10]. Transgenes containing deletions of specific control regions have revealed a complex regulatory strategy at different developmental stages [11]. A fine structure analysis revealed similarities between the chromatin structure observed in

hematopoietic progenitor cells that do not express the lysozyme gene and mature lysozyme-expressing cells; these similarities suggest that changes in chromatin structure required for activation take place in the early progenitors prior to transcriptional activation [12]. The binding of specific transcription factors has been observed at lysozyme regulatory regions in the hematopoietic progenitors, with factor binding accompanied by reductions in DNA methylation and linker histones in the vicinity of the binding sites [13-15].

Recently, genome-wide studies performed with murine macrophages and B cells have provided intriguing evidence that the developmentally regulated transcription factor PU.1, along with a limited number of other lineage-determining factors, plays a broad role in the specification of enhancers for genes that will eventually be activated in these lineages. In one study, ChIP-Seq with antibodies directed against the p300 co-activator and histone H3K4me1, which is preferentially observed at enhancers [16•], was used to identify enhancers that are induced following LPS stimulation of mature macrophages [17•]. Activation of the enhancers by LPS was accompanied by inducible p300 association and inducible binding of NF- κ B and other factors involved in LPS-induced transcription. Surprisingly, however, a high percentage of the enhancers were occupied by PU.1 and exhibited a broad window of H3K4me1 in the unstimulated macrophages. The PU.1 binding sites coincided with a low nucleosome density, suggesting that PU.1 binding during macrophage development catalyzes nucleosome eviction or re-positioning at the enhancers. Importantly, ectopic expression of PU.1 was sufficient for inducing the H3K4me1 mark at a subset of these enhancers.

A second ChIP-Seq study suggested that PU.1 marks enhancers for B cell-specific and myeloid-specific genes at early stages of development in combination with a small number of other factors that have previously been implicated in specification of the B-cell and myeloid lineages [18•]. The initial ChIP-Seq experiments performed with PU.1 antibodies revealed that PU.1 associates with a large number of DNA regions marked by H3K4me1 in the B-cell and myeloid lineages, with PU.1 association in each lineage coinciding with transcription of the linked genes. Motif analysis revealed that the regions occupied by PU.1 contained nearby binding sites for other key factors involved in either B-cell or myeloid development, such as E2A, EBF, or C/EBP, with additional experiments demonstrating the importance of these factors for enhancer marking. Consistent with the results of Ghisletti et al. [17•], factor binding and the H3K4me1 mark appeared to specify the potential for transcriptional activation, with transcription often occurring in mature cells only following the activation of other factors that are induced by environmental stimuli. PU.1 and the other developmental regulatory proteins implicated in enhancer marking in these studies have long been known to carry out key roles at early stages of development by directly activating genes that must be expressed for development to proceed [18-21]. These new results suggest that a second critical function of these factors is to broadly mark enhancers for genes that will be activated at later stages of development or in mature terminally differentiated cells in response to environmental stimuli.

Marking of tissue-specific promoters and enhancers in ESC

Extending the notion that protein-DNA interactions and chromatin signatures mark tissue-specific genes long before the genes are transcribed, evidence has emerged that enhancers and/or promoters for many tissue-specific and developmental stage-specific genes are first marked, or specified, in pluripotent ESC.

Inactive genes that appear to be marked in ESC can be divided into two fundamentally distinct classes (Figure 1). One class was initially identified through ChIP-chip analysis of histone modifications in ESC [22-25]. The histone H3K4me3 modification, typically observed at the promoters of active genes, was found to co-localize with the repressive histone H3K27me3 mark at the promoters of many inactive genes that encode regulators of early developmental

decisions (Figure 1A). Promoters containing these bivalent histone modification domains typically contain CpG islands [26••,27••], which are found at the promoters of approximately 70% of mammalian genes, including most housekeeping genes and a subset of regulated genes. Consistent with the presence of the H3K27me3 mark, polycomb complexes, which deposit this mark, were found to be associated with the bivalent histone modification domains [22-27]. Depletion of polycomb proteins resulted in moderately increased transcription. These results support the hypothesis that genes encoding key developmental regulators are poised for activation due to the presence of H3K4me3, but polycomb complexes are recruited to the promoters in pluripotent cells to deposit H3K27me3; this modification promotes the assembly of a silent chromatin structure to prevent transcription until the H3K27me3 mark is removed during differentiation.

The second class of inactive genes that appear to be marked in ESC includes tissue-specific genes that are not involved in developmental decisions [28]. The promoters for these genes typically contain a low CpG content and are not marked by bivalent domains (Figure 1B). Representative genes within this class exhibit high CpG methylation in their promoters, and their promoters exhibit no evidence of either active or repressive histone modifications [29•]. However, their enhancers can be marked in ESC by protein-DNA interactions, reduced CpG methylation, and/or active histone modifications.

An initial report of enhancer marking at a typical tissue-specific gene focused on the mouse pre-B-cell-specific $\lambda 5/VpreB$ locus (*Igll1/Vpreb1*). ChIP scanning through the locus revealed hyperacetylation of histone H3 at a distant site that corresponds to an enhancer [30,31]. In a multipotent progenitor line and in pre-B cells, the hyperacetylation gradually broadened through the locus, suggesting that the region hyperacetylated in ESC represents a nucleation site for chromatin changes that provide competence for transcriptional activation in pre-B cells.

More recently, windows of unmethylated CpG dinucleotides have been observed in ESC at the liver-specific enhancer for the *Alb1* gene, and also at tissue-specific enhancers for the thymocyte-specific *Ptcra* gene and macrophage/dendritic cell-specific *Il12b* gene [29•,32•]. At the *Ptcra* enhancer, the unmethylated window was accompanied by histone H3 acetylation, but at the *Alb1* and *Il12b* enhancers, no enrichment in histone modifications associated with either active or inactive chromatin was observed [29•]. The presence of unmethylated CpGs provided initial evidence that transcription factors may be bound to the enhancers in ESC; factor binding could block DNA methylation during propagation of the ESC lines, although active demethylation cannot be ruled out.

At the *Alb1* enhancer, low methylation was observed in multiple ESC lines at a single CpG [29•]. This dinucleotide was located within one of the key binding sites for FoxA1, suggesting that a Fox family member occupies the enhancer. FoxA1 is not expressed in ESC, but another Fox family member, FoxD3, is highly expressed and has been implicated in a pluripotency network [33-35]. Chromatin immunoprecipitation (ChIP) revealed that FoxD3 binds the *Alb1* enhancer in ESC [29•]. Furthermore, siRNA knockdown of *Foxd3* led to a rapid increase in methylation at the *Alb1* enhancer CpG. Since FoxD3 expression is downregulated during differentiation of ESC into endoderm, these results suggest that FoxD3 gains initial access to the enhancer in ESC and may serve as a place-holder for FoxA1, which binds after gastrulation [3] (Figure 2). The precise functional significance of FoxD3 binding remains to be established.

At the *Ptcra* enhancer, the DNA motifs that are needed for establishment of the unmethylated window in ESC were identified using an assay in which enhancer-promoter-reporter plasmids were methylated in vitro by the Sss1 CpG methylase and then stably transfected into an ESC line [29•,32•]. An unmethylated window was consistently observed at the stably transfected *Ptcra* enhancer in clonal ESC lines, revealing that factors can gain access to the methylated

promoter and either prevent maintenance methylation or drive active demethylation. Analysis of a series of enhancer mutations revealed that an Sp1 binding motif and an adjacent E box are essential for establishment of the unmethylated window. When either motif was disrupted, the transfected plasmid remained fully methylated. The finding that an Sp1 site is important for establishment of the unmethylated window at this tissue-specific enhancer is interesting to consider in light of earlier evidence that Sp1 binding is essential for the establishment of an unmethylated state at CpG island promoters during embryogenesis [36,37].

It is important to emphasize that, although the *Ptcra* enhancer exhibits moderate levels of histone H3K9/K14 acetylation in ESC, marking of the *Alb1* and *Il12b* enhancers in ESC would not have been revealed by genome-wide profiling of common histone modifications; the protein-DNA interactions at these enhancers do not lead to histone modifications associated with either active or inactive chromatin in ESC [29•]. Enhancer occupancy also did not lead to nuclease hypersensitivity [29•]. Thus, the only common strategies that would provide evidence of protein-DNA interactions at these enhancers in ESC are bisulfite sequencing analysis of DNA methylation state or ChIP analysis of individual transcription factors. Bisulfite sequencing does not require prior knowledge of the transcription factors that may be bound, but reduced CpG methylation may be observed only if CpGs are located in close proximity to the bound factors. Although reduced methylation of a CpG within the *Alb1* enhancer provided the initial evidence that a transcription factor was associated with the enhancer in ESC, it is important to note that the absence of methylation of this one CpG may have little functional relevance; instead, it may simply be a consequence of FoxD3 binding.

Functional relevance of protein-DNA interactions at tissue-specific enhancers in ESC

The functional role of the protein-DNA interactions observed at tissue-specific enhancers in ESC remains a critical unanswered question. At promoters marked by bivalent histone modification domains, the presence of H3K27me3 and the association of polycomb complexes, coupled to their loss at genes that are activated during differentiation, provides compelling evidence that the bivalent domains exist to keep genes silent but poised for activation. However, the protein-DNA interactions observed at the enhancers of typical tissue-specific genes in ESC do not coincide with repressive histone modifications and often do not coincide with histone modifications associated with active chromatin [29•]. One possibility is that the factors involved in these early interactions possess special properties that allow them to access the ESC chromatin and maintain accessibility during the earliest stages of differentiation. Since FoxD3 and FoxA1 share similar winged helix DNA-binding domains, it is reasonable to anticipate that FoxD3 will share FoxA1's capacity to bind stably to nucleosomal DNA and catalyze changes in nucleosome structure. Similarly, Sp1 and the E protein that appear to mark the *Ptcra* enhancer may help keep the enhancer competent for the binding of other factors following differentiation into the thymocyte lineage.

To fully understand the relevance of the protein-DNA interactions in ESC, it may be important to consider the fundamental differences in chromatin structure that have been observed in ESC in comparison to differentiated cells. ESC exhibit a global chromatin structure that is more “open” than that found in differentiated cells [28,38]. In particular, heterochromatin protein 1 (HP1) and core histones are associated with chromatin less tightly in ESC than in differentiated cells [39••]. Consistent with the view that chromatin is generally more accessible, low levels of transcription can be detected from many tissue-specific genes in ESC [40••,41••].

Recently, the Chd1 nucleosome remodeling protein was found to be essential for the characteristically open chromatin structure observed in ESC [42••]. Chd1 was also important for the differentiation of ESC into endoderm and for the reprogramming of somatic cells into

induced pluripotent stem cells (iPSC) [42••,43••,44]. These results support the hypothesis that the unique characteristics of ESC chromatin may be important for pluripotency.

With the unique properties of ESC chromatin in mind, one can speculate that protein-DNA interactions at tissue-specific enhancers in ESC may be necessary to prevent the enhancers from assembling into repressive chromatin structures during differentiation that may be resistant to activation. As an initial test of this hypothesis, pre-methylated plasmids containing the *Ptcra* or *Il12b* enhancer were stably transfected into differentiated thymocytes or macrophages that actively express the endogenous *Ptcra* or *Il12b* genes, respectively. Although these differentiated cells contain all of the factors needed for active transcription of the endogenous genes, the pre-methylated plasmids rapidly assembled into silent chromatin and remained fully methylated and resistant to activation [29•,32•]. Establishment of an unmethylated window occurred only after depletion of the repressive Mi-2/NuRD chromatin remodeling complex [32•]. These results contrast with those obtained in ESC, where unmethylated windows were rapidly established at the stably transfected pre-methylated enhancers, despite the absence of active transcription. These results provide initial support for the hypothesis that protein-DNA interactions in ESC or early progenitor cells may be required to prevent the assembly of a repressive chromatin structure that is resistant to activation. It follows that the reprogramming of somatic cells into iPSC may require a transition to a more open chromatin structure that is permissive to the re-establishment of marks at the enhancers for tissue-specific genes.

Conclusions

The results described above provide an initial glimpse of events occurring at the earliest stages of development that help coordinate the proper regulation of tissue-specific and developmental stage-specific genes. Excluded from the above discussion are housekeeping genes, which generally contain CpG-island promoters assembled into constitutively active chromatin. Like the housekeeping genes, regulated genes involved in early developmental decisions contain CpG-island promoters, but in their inactive state, polycomb complexes deposit a histone H3K27me3 mark, which promotes assembly of a chromatin structure that keeps the genes repressed but poised for activation when differentiation is initiated.

Many other tissue-specific, developmental stage-specific, and inducible genes contain promoters with a low CpG content. In ESC, these promoters often lack common histone modifications and contain methylated CpGs. However, distant enhancers for some genes are clearly associated with transcription factors, which can lead to the presence of unmethylated CpGs and/or active histone modifications. The finding that the *Alb1* and *Ptcra* enhancers are associated with transcription factors from different families raises the possibility that many different factors may carry out important interactions with tissue-specific enhancers in ESC. During development, small sets of key developmental regulatory proteins - such as PU.1, EBF, and E2A for B-lineage genes and PU.1 and C/EBP β for myeloid-specific genes [17••,18••] - associate with large sets of tissue-specific enhancers and appear to drive nucleosome remodeling and deposition of the histone H3K4me1 mark. Then, prior to or coincident with gene activation, other factors bind the enhancers and promoters for these genes and activate transcription.

This simple model provides a framework for future investigation, which undoubtedly will reveal much more complexity. As initial evidence of the greater complexity, a recent study revealed two distinct classes of bivalent domains in ESC [45••]. More generally, the cascade of events leading to gene activation is likely to involve an intricate balance between positive and negative regulatory mechanisms that organize chromatin domains in a manner that maintains competence for eventual activation and accessibility to essential factors, while

keeping the locus inactive until an appropriate developmental stage. As one example, FoxA1, a transcriptional activator implicated in transcriptional competence, was recently found to associate with a Groucho co-repressor at its target genes in endoderm; this co-repressor interaction may help to establish and maintain an inactive chromatin structure while simultaneously maintaining competence for transcription in hepatocytes [46,47•,48•].

Future progress toward the goal of understanding how transcriptional regulation is coordinated during development will greatly benefit from ongoing improvements in genomics, as well as from the ability to combine studies of development with studies of somatic cell reprogramming. However, to fully understand developmental gene regulation, new technologies will be needed, including technologies that provide greater knowledge of the physical structure of chromatin at individual loci at various developmental stages.

Acknowledgments

Studies of pioneer factors and the marking of tissue-specific genes in ESC in the author's laboratory are supported by NIH grant R21CA137278.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Davis RL, Weintraub H, Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 1987;51:987–1000. [PubMed: 3690668]
2. Schlissel M, Voronova A, Baltimore D. Helix-loop-helix transcription factor E47 activates germ-line immunoglobulin heavy-chain gene transcription and rearrangement in a pre-T-cell line. *Genes Dev* 1991;5:1367–1376. [PubMed: 1714414]
3. Gualdi R, Bossard P, Zheng M, Hamada Y, Coleman JR, Zaret KS. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev* 1996;10:1670–1682. [PubMed: 8682297] ••The concept that transcription factors bind to tissue-specific enhancers at an early stage of development and initiate a cascade of events that culminates in transcription in mature tissues was first discussed in this paper. Genomic footprinting was used to show that the liver-specific *Alb1* enhancer is occupied by transcription factors in definitive endoderm.
4. Bossard P, Zaret KS. GATA transcription factors as potentiators of gut endoderm differentiation. *Development* 1998;125:4909–4917. [PubMed: 9811575]
5. Clark KL, Halay ED, Lai E, Burley SK. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 1993;364:412–420. [PubMed: 8332212]
6. Omichinski JG, Clore GM, Schaad O, Felsenfeld G, Trainor C, Appella E, Stahl SJ, Gronenborn AM. NMR structure of a specific DNA complex of Zn-containing DNA binding domain of GATA-1. *Science* 1993;261:438–446. [PubMed: 8332909]
7. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell* 2002;9:279–289. [PubMed: 11864602]
8. Sekiya T, Muthurajan UM, Luger K, Tulin AV, Zaret KS. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev* 2009;23:804–809. [PubMed: 19339686] ••This study uses fluorescence recovery after photobleaching to show that the FoxA1 pioneer factor is far less mobile in the nucleus than transcription factors within several other families. Reduced mobility appears to be due to stable non-specific binding to nucleosomal DNA.
9. Bonifer C, Hecht A, Saueressig H, Winter DM, Sippel AE. Dynamic chromatin: the regulatory domain organization of eukaryotic gene loci. *J Cell Biochem* 1991;47:99–108. [PubMed: 1757483]

10. Huber MC, Graf T, Sippel AE, Bonifer C. Dynamic changes in the chromatin of the chicken lysozyme gene domain during differentiation of multipotent progenitors to macrophages. *DNA Cell Biol* 1995;14:397–402. [PubMed: 7748489]
11. Jäggle U, Müller AM, Kohler H, Bonifer C. Role of positive and negative cis-regulatory elements in the transcriptional activation of the lysozyme locus in developing macrophages of transgenic mice. *J Biol Chem* 1997;272:5871–5879. [PubMed: 9038204]
12. Kontaraki J, Chen HH, Riggs A, Bonifer C. Chromatin fine structure profiles for a developmentally regulated gene: reorganization of the lysozyme locus before trans-activator binding and gene expression. *Genes Dev* 2000;14:2106–2122. [PubMed: 10950873]
13. Lefevre P, Melnik S, Wilson N, Riggs AD, Bonifer C. Developmentally regulated recruitment of transcription factors and chromatin modification activities to chicken lysozyme cis-regulatory elements in vivo. *Mol Cell Biol* 2003;23:4386–43400. [PubMed: 12773578]
14. Lefevre P, Lacroix C, Tagoh H, Hoogenkamp M, Melnik S, Ingram R, Bonifer C. Differentiation-dependent alterations in histone methylation and chromatin architecture at the inducible chicken lysozyme gene. *J Biol Chem* 2005;280:27552–27560. [PubMed: 15923188]
15. Tagoh H, Melnik S, Lefevre P, Chong S, Riggs AD, Bonifer C. Dynamic reorganization of chromatin structure and selective DNA demethylation prior to stable enhancer complex formation during differentiation of primary hematopoietic cells in vitro. *Blood* 2004;103:2950–2955. [PubMed: 15070670]
16. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* 2007;39:311–318. [PubMed: 17277777] ••This study revealed that mammalian enhancers are commonly marked by histone H3K4me1, whereas promoters are more commonly marked by histone H3K4me3.
17. Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, Venturini E, Gregory L, Lonie L, Chew A, Wei CL, Ragoussis J, Natoli G. Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. *Immunity* 2010;32:317–328. [PubMed: 20206554] ••Genome-wide identification of LPS-induced enhancers in mouse macrophages by p300 ChIP-Seq revealed that the inducible enhancers are frequently associated with PU.1 and H3K4me1 in unstimulated cells, with p300 and LPS-induced factors associating with the enhancers after cell stimulation. PU.1 binding appeared to be responsible for chromatin remodeling and the H3K4me1 mark at the enhancers.
18. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mo Cell* 2010;38:576–589. p576. ••Genome-wide identification of PU.1 binding sites in B cells and macrophages by ChIP-Seq revealed that PU.1 commonly binds enhancers in close proximity to other key regulators of B-cell or myeloid development. Binding of these small sets of key factors appeared to be responsible for chromatin remodeling and the H3K4me1 mark at the lineage-specific enhancers.
19. Scott EW, Simon MC, Anastasi J, Singh H. Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages. *Science* 1994;265:1573–1577. [PubMed: 8079170]
20. Friedman AD. Transcriptional control of granulocyte and monocyte development. *Oncogene* 2007;26:6816–6828. [PubMed: 17934488]
21. Murre C. Regulation and function of the E2A proteins in B cell development. *Adv Exp Med Biol* 2007;596:1–7. [PubMed: 17338171]
22. Azuara V, Perry P, Sauer S, Spivakov M, Jørgensen HF, John RM, Gouti M, Casanova M, Warnes G, Merckenschlager M, Fisher AG. Chromatin signatures of pluripotent cell lines. *Nat Cell Biol* 2006;8:532–538. [PubMed: 16570078]
23. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 2006;125:315–326. [PubMed: 16630819]
24. Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, et al. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 2006;441:349–353. [PubMed: 16625203]

25. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, et al. Control of developmental regulators by polycomb in human embryonic stem cells. *Cell* 2006;125:301–313. [PubMed: 16630818]
26. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 2007;448:553–560. [PubMed: 17603471] ••This study used ChIP-Seq to examine histone modifications across the genome in ESC, neural progenitor cells, and embryonic fibroblasts. Among several key findings, bivalent histone modifications domains in ESC were found to be associated preferentially with CpG-island promoters that exhibit low transcriptional activity.
27. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, Adli M, Kasif S, Ptaszek LM, Cowan CA, Lander ES, Koseki H, Bernstein BE. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 2008;4:e1000242. [PubMed: 18974828] ••ChIP-Seq analysis of histone modifications and of specific subunits of the PRC1 and PRC2 polycomb complexes revealed that both PRC1 and PRC2 are associated with the CpG-island promoters of genes encoding key developmental regulators, whereas a second class of promoters associated only with PRC2 appears to be functionally distinct.
28. Ram EV, Meshorer E. Transcriptional competence in pluripotency. *Genes Dev* 2009;23:2793–2798. [PubMed: 20008929]
29. Xu J, Pope SD, Jazirehi AR, Attema JL, Papathanasiou P, Watts JA, Zaret KS, Weissman IL, Smale ST. Pioneer factor interactions and unmethylated CpG dinucleotides mark silent tissue-specific enhancers in embryonic stem cells. *Proc Natl Acad Sci USA* 2007;104:12377–12382. [PubMed: 17640912] •This study identified and characterized unmethylated CpG windows in ESC within the enhancers of three typical tissue-specific genes.
30. Szutorisz H, Canzonetta C, Georgiou A, Chow CM, Tora L, Dillon N. Formation of an active tissue-specific chromatin domain initiated by epigenetic marking at the embryonic stem cell stage. *Mol Cell Biol* 2005;25:1804–1820. [PubMed: 15713636]
31. Minaee S, Farmer D, Georgiou A, Sabbattini P, Webster Z, Chow CM, Dillon N. Mapping and functional analysis of regulatory sequences in the mouse lambda5-VpreB1 domain. *Mol Immunol* 2005;42:1283–1292. [PubMed: 15950724]
32. Xu J, Watts JA, Pope SD, Gadue P, Kamps M, Plath K, Zaret KS, Smale ST. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes Dev* 2009;23:2824–2838. [PubMed: 20008934] •This study demonstrated that FoxD3 binding to the *Alb1* enhancer is responsible for an unmethylated CpG observed at the enhancer in ESC. The study also showed that pluripotent cells and differentiated cells exhibit different sensitivities to the establishment of unmethylated windows at pre-methylated enhancers, and that the reprogramming of somatic cells to induced pluripotent stem cells is accompanied by the re-establishment of enhancer marks at typical tissue-specific genes.
33. Clevidence DE, Overdier DG, Tao W, Qian X, Pani L, Lai E, Costa RH. Identification of nine tissue-specific transcription factors of the hepatocyte nuclear factor 3/forkhead DNA-binding-domain family. *Proc Natl Acad Sci USA* 1993;90:3948–3952. [PubMed: 7683413]
34. Hanna LA, Foreman RK, Tarasenko IA, Kessler DS, Labosky PA. Requirement for Foxd3 in maintaining pluripotent cells of the early mouse embryo. *Genes Dev* 2002;16:2650–2661. [PubMed: 12381664]
35. Sutton J, Costa R, Klug M, Field L, Xu D, Largaespada DA, Fletcher CF, Jenkins NA, Copeland NG, Klemsz M, et al. Genesis, a winged helix transcriptional repressor with expression restricted to embryonic stem cells. *J Biol Chem* 1996;271:23126–23133. [PubMed: 8798505]
36. Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H. Sp1 elements protect a CpG island from de novo methylation. *Nature* 1994;371:435–438. [PubMed: 8090226]
37. Macleod D, Charlton J, Mullins J, Bird AP. Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island. *Genes Dev* 1994;8:2282–2292. [PubMed: 7958895]
38. Meshorer E, Misteli T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nat Rev Mol Cell Biol* 2006;7:540–546. [PubMed: 16723974]
39. Meshorer E, Yellajoshula D, George E, Scambler PJ, Brown DT, Misteli T. Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev Cell* 2006;10:105–116. [PubMed: 16723974]

16399082] ••Fluorescence recovery after photobleaching (FRAP) and biochemical methods were employed to demonstrate that HP-1, histone H1, and core histones are less tightly associated with chromatin in pluripotent cells than in differentiated cells.

40. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* 2007;130:77–88. [PubMed: 17632057] ••This study reports transcription initiation from most protein-coding genes in ESC, consistent with the view that chromatin is more generally accessible in ESC than in differentiated cells. The authors failed to observe full-length transcripts for many genes that exhibited transcription initiation, suggesting that elongation is tightly controlled at many genes.
41. Efroni S, Duttagupta R, Cheng J, Dehghani H, Hoepfner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RD, Buetow KH. Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* 2008;2:437–447. [PubMed: 18462694] ••Widespread transcription of both coding and non-coding regions of the genome was observed in ESC. Upon differentiation, transcription became much more restricted, supporting the view that the genome is more generally accessible to the transcription machinery in pluripotent cells.
42. Gaspar-Maia A, Alajem A, Polesso F, Sridharan R, Mason MJ, Heidersbach A, Ramalho-Santos J, McManus MT, Plath K, Meshorer E, Ramalho-Santos M. Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* 2009;460:863–868. [PubMed: 19587682] ••This study revealed that the nucleosome remodeling factor Chd1 is a key regulator of the open chromatin structure that characterizes pluripotent cells. Chd1 was found to be critical for pluripotency of ESC and for the generation of iPSC, providing support for the hypothesis that an open chromatin structure is important for pluripotency.
43. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006;126:663–676. [PubMed: 16904174] ••This landmark study demonstrates that induced pluripotent stem cells can be generated by the ectopic expression of genes encoding four transcription factors in somatic cells.
44. Hochedlinger K, Jaenisch R. Nuclear reprogramming and pluripotency. *Nature* 2006;441:1061–1067. [PubMed: 16810240]
45. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* 2008;4:e1000242. [PubMed: 18974828] ••This study defined two distinct classes of bivalent domains in ESC. One class is characterized by association with both PRC1 and PRC2, whereas the other class contains only PRC1. The first class is enriched in genes encoding developmental regulators and containing CpG-island promoters.
46. Wang JC, Waltner-Law M, Yamada K, Osawa H, Stifani S, Granner DK. Transducin-like enhancer of split proteins, the human homologs of *Drosophila* groucho, interact with hepatic nuclear factor 3beta. *J Biol Chem* 2000;275:18418–18423. [PubMed: 10748198]
47. Sekiya T, Zaret KS. Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. *Mol Cell* 2007;28:291–303. [PubMed: 17964267] ••This study showed that in vitro recruitment of a Groucho co-repressor to a nucleosomal array by FoxA1 and other transcription factors promotes assembly of an inaccessible chromatin structure spanning a small number of nucleosomes. Targeted inaccessibility was similarly observed upon ectopic expression in vivo.
48. Santisteban P, Recacha P, Metzger DE, Zaret KS. Dynamic expression of Groucho-related genes Grg1 and Grg3 in foregut endoderm and antagonism of differentiation. *Dev Dyn* 2010;239:980–986. [PubMed: 20108349] •The Groucho family member, Grg3, was found to be expressed in foregut endoderm, where FoxA1 is known to first bind the *Alb1* enhancer. Based on the earlier finding that FoxA1 can interact with Grg3 and promote the assembly of inaccessible chromatin, the authors speculate that Grg3 is recruited to FoxA1 targets in endoderm, where it helps maintain an inactive state while FoxA1 simultaneously maintains competence for activation in hepatocytes.

A. Inactive Genes Encoding Developmental Regulators in ESC

B. Inactive Tissue-Specific Genes in ESC

Figure 1.

Two classes of inactive genes in ESC. A. Genes encoding developmental regulators often contain CpG-island promoters that contain unmethylated CpG dinucleotides and are marked by histone H3K4me3 and histone H3K27me3 in ESC. Together, these modifications, which typically are associated with active and inactive genes, respectively, comprise a bivalent histone modification domain [22-25]). Transcription factors are presumably associated with the promoters of genes marked by bivalent domains in ESC. These factors presumably recruit polycomb complexes, which deposit the H3K27me3 modification. B. Typical tissue-specific genes often contain low CpG promoters that contain methylated CpGs in ESC. The promoters of these genes may not be associated with transcription factors or histone modifications representative of either active or inactive genes. However, windows of unmethylated CpG dinucleotides and transcription factor binding have been observed in ESC at representative tissue-specific enhancers.

Figure 2.

Properties of the mouse *Alb1* enhancer during development (adapted from [32•]). In pluripotent ESC, one CpG dinucleotide in the *Alb1* enhancer exhibits low methylation, with high methylation of all other CpGs [29•]. This CpG is within a Fox binding site and FoxD3 binds the enhancer in ESC. siRNA knockdown of *Foxd3* leads to full methylation of the CpG [32•]. In definitive endoderm, *Foxd3* is downregulated and *Foxa1* and *Foxa2* are upregulated. The Fox and GATA binding sites in the *Alb1* enhancer are occupied in endoderm by FoxA1 and/or FoxA2 and by GATA4 and/or GATA6 [3•,4]. Upon differentiation into hepatocytes, several additional factors associate with the enhancer and all CpGs within the enhancer become unmethylated. High methylation of these CpGs is observed in most other mouse tissues [29•], suggesting that the enhancer is likely to be methylated in mesoderm and ectoderm.