Chromatin priming of genes in development: Concepts, mechanisms and consequences

Constanze Bonifer and Peter N. Cockerill
Institute of Cancer and Genomic Sciences, College of Medicine and Dentistry, University of Birmingham, Birmingham, UK

(Received 19 December 2016; revised 19 January 2017; accepted 21 January 2017)

During ontogeny, cells progress through multiple alternate differentiation states by activating distinct gene regulatory networks. In this review, we highlight the important role of chromatin priming in facilitating gene activation during lineage specification and in maintaining an epigenetic memory of previous gene activation. We show that chromatin priming is part of a hugely diverse repertoire of regulatory mechanisms that genes use to ensure that they are expressed at the correct time, in the correct cell type, and at the correct level, but also that they react to signals. We also emphasize how increasing our knowledge of these principles could inform our understanding of developmental failure and disease. Copyright © 2017 The Authors. Published by Elsevier Inc. on behalf of ISEH - International Society for Experimental Hematology. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Concepts

Timing of gene activation during cell differentiation involves the interaction of multiple cis-regulatory elements

All development in metazoans starts with a single totipotent cell that divides, activates differential sets of genes, and produces many types of differentiated progeny with alternate restricted lineage fates. Such cells eventually give rise to distinct subsets of functionally specified cells that comprise the organism. However, to replenish those cells exiting the proliferative state and entering terminal differentiation, stem cell systems have evolved that have acquired the ability to either self-renew or differentiate. All of these processes are under strict transcriptional control and are driven by cell-type-specific, ubiquitous transcription factors (TFs), which recruit epigenetic regulators that modify the chromatin template to facilitate the assembly of the transcription machinery. Differential gene expression takes place in topologically separate domains within individual chromosomes (topology-associated domains or TADs) that are dynamically reorganized in development [1,2]. The developmental control of gene expression is therefore a highly dynamic process that involves both the cell-type-specific expression of TFs and the establishment of specific chromatin landscapes presenting a defined target for such factors. In the last few years, using single-gene and system-wide studies, we have started to gain insights into the molecular mechanisms that regulate these processes.

The vast majority of genes expressed in any one specific cell type, which represent about half of all genes, are also active in most cell types and are controlled by generic mechanisms acting at gene promoters at the level of mRNA expression. We commonly refer to these genes as “housekeeping” genes because they typically regulate basic cellular functions such as metabolism and replication. However, genes that are specific for certain cell types or that encode crucial lineage determining regulators are under strict developmental control. These genes tend to use multiple cis-regulatory elements, ensuring that their expression is properly controlled at each developmental stage. Much of what we know about developmental regulation...
comes from detailed studies of the temporal control of gene expression of well-defined model genes during ontogeny, many of them in the hematopoietic system. These include the beta-globin gene cluster, which consists of several genes that are expressed in a developmentally controlled fashion, with the embryonic globin genes being expressed first and the adult genes being expressed after birth, and the homeobox (Hox) gene clusters, which encode regulators of body patterning that are highly conserved across multiple species and have an expression that is also under strict developmental control. An extraordinary feature that these gene clusters have in common is the fact that, in each case, the order of activation of the individual genes during ontogeny mirrors their order on the chromosome, and the elements that define these patterns are contained within these loci [3,4]. If the specific order of genes is disturbed, then the temporal activation of the individual genes is also altered [5,6], indicating that the linear organization of genes is critical for their correct developmental control.

The erythroid-specifically expressed beta-globin locus is regulated by a far upstream cis-regulatory region, the locus control region (LCR) [7], which is rendered highly accessible to both TFs and nucleases. Such elements are generally referred to as DNase-hypersensitive sites (DHSs) [8]. The LCR interacts physically in a developmentally controlled fashion with downstream genes in the cluster to drive the differential expression of embryonic and adult globin genes [9,10]. Later experiments identified a relay of stage-specifically expressed TFs that bind to both the LCR and to downstream cis-regulatory elements that define the timing of the stage-specific expression of each gene [11,12]. These studies taught us that the cell-stage-specific activation of a gene locus in development is a stepwise and hierarchical process that involves multiple cis-regulatory elements physically interacting with each other. This principle also held up for tissue-specific genes that are not organized in gene clusters or do not contain a discrete LCR, such as the myeloid-specifically expressed chicken lysozyme locus [13], the B-cell-specific Pax5 locus [14], and the human alpha-globin locus [15–17]. These genes are all regulated by complex sets of interacting regulatory elements that recruit multiple interacting TFs and cofactors and are differently active during development or drive expression in multiple tissues. At the level of chromatin organization, the interacting cis-regulatory elements form active chromatin hubs or domains [9,18] involving multiple DNA elements within active gene loci. Such clusters of regulatory elements have recently been renamed “super enhancers” [19].

**Poising promoters for rapid activation and repression:**

The role of polycomb complex

There are several ways of activating genes in a stepwise fashion in development. Polycomb group proteins (PcGs) play an important role in the developmental control of gene expression and were originally identified as factors that regulate lineage specification in Drosophila and maintain different patterns of gene expression within an organism. PcGs control an extensive regulatory network functioning in many organisms, from insects to mammals, to establish repressed or poised promoter states during development. PcG-bound promoters typically bind the non-elongating form of RNA-polymerase II and exist in a bivalent state that contains the active histone H3 lysine 4 trimethylation (H3K4me3) mark together with the PcG-associated repressive H3K27me3 mark [20]. By blocking transcriptional elongation, PcG complexes can poise promoters for the rapid reactivation of transcription at the appropriate stage and time [21]. The removal of PcG complexes from cells by gene targeting therefore leads to premature gene activation [22] and a global alteration of gene expression patterns [23]. A number of system-wide studies of PcG binding at different developmental stages have shown that these complexes can display a highly dynamic binding pattern during development, with some promoters being bound and repressed by PcG at one stage, derepressed at the next stage to activate transcription transiently, and then being rebound by PcG to enforce gene silencing and block promiscuous transcription at the final stage [24–26].

**Gene loci show early signs of activation at distal elements before the onset of gene expression**

RNA synthesis from gene promoters is controlled at multiple levels, from the initial opening up of promoter regions to the recruitment of RNA-polymerase II and the process of transcriptional elongation [27]. It is now well established that developmentally regulated genes with complex regulatory regions comprising multiple DNA elements are activated in a stepwise fashion at the levels of chromatin accessibility and transcriptional activation. The concept that gene locus activation begins with chromatin priming was shown first by single-gene analyses [28–33] and confirmed later in multiple system-wide studies [11,24,34–39]. Chromatin priming represents the underlying basis for transcriptional priming in precursor cells before alternate commitment decisions. This primed state can be (but is not always) associated with low-level expression of multiple lineage-specific genes, the expression of which later becomes restricted to one lineage after terminal differentiation [40–42]. The maintenance of the primed and the committed states is in each case controlled by specific, tightly regulated TFs that become active at different stages [24,43].

Chromatin priming is associated with different features, all of which one way or another involve the action of TFs, as summarized in Figure 1. One of the earliest studies of chromatin priming was performed by the Zaret group [28] using in vivo dimethyl sulfate (DMS) foot printing to identify occupied TF-binding motifs at sequences harboring
a liver-specific albumin enhancer in precursor gut endoderm, where this gene is transcriptionally silent. The associated albumin gene is only activated after hepatic specification when a number of additional factors bind to nearby sites. An independent single-gene study examining the PU.1 locus demonstrated that this gene is marked by a selective demethylation of specific TF-binding sites at an enhancer that is essential for the activation of the gene (the upstream regulatory element [URE]). At the same time, an increase in overall DNase I accessibility was observed as a result of a hit-and-run mechanism involving the TF RUNX1, followed by the assembly of a stable factor complex and the activation of transcription together with the activation of additional transcriptional enhancers. TFs can also mediate locus derepression via DNA-demethylation by direct interference with DNA methyl transferases or by recruiting demethylating enzymatic activities such as TET enzymes. The universality of priming mechanisms involving selective demethylation, together with the early appearance of regions of accessible chromatin, TF binding, and the presence of monomethylated histone H3 lysine 4 before transcriptional activation, was confirmed by genome-wide studies.

Although priming of gene regulatory regions in precursor cells is widespread, a recent study challenged the concept that individual enhancers invariably have to be primed at an early stage in order to be fully active later in development. That study defined priming on the basis of the presence of H3K4me1 and found that only a fraction of enhancers in precursor cells remained active in mature cells, indicating that the enhancer repertoire is re-shaped dynamically during differentiation. However, a caveat of most genome-wide studies is that they do not examine actual TF occupancy, which is a better predictor of enhancer activity. Goode et al. did measure TF binding during the differentiation of mouse embryonic stem cells to macrophages and found a number of TF-bound elements that were associated with low or absent gene expression. Interestingly, this study noted that only a very small number of distal elements preexist in a PcG-associated repressed state, indicating that chromatin priming of enhancers does not typically involve PcG binding. However, specific developmentally regulated TFs bound at active enhancers are likely to lead to the eviction of polycomb complexes from PcG-bound poised promoter regions, as exemplified in the alpha-globin locus.

**Mechanisms**

**Chromatin priming involves TF binding and is important for the correct developmental regulation of genes**

The question now arises, how important is chromatin priming for the developmental control of gene expression? Very few studies have actually addressed the issue at a functional level by following the developmental activation of genes in a context in which priming is disturbed. The reason for this is that such studies require the editing of endogenous genes because the analysis of transgenes may be misleading due to copy number, gene orientation, and genomic position effects (e.g., see Ellis et al. or Calero-Nieto et al.). One study followed the differentiation of embryonic stem cells into macrophages to examine the kinetics of developmental upregulation of the *Spi1* (Pu.1) gene in cells harboring homozygous mutations within an ETS-binding
site in the Spi1 URE [51]. The E26 transformation-specific (ETS) family factor FLI1 binds to the URE before the specification of hematopoietic cells, before the onset of Spi1 transcription in multipotent progenitor cells, and PU.1 subsequently replaces FLI1 as part of an autoregulatory loop. Interestingly, the mutation of the FLI1/PU.1-binding site did not change the level of Spi1 transcription in macrophages. However, the time course of activation was remarkably different, with the mutated allele lagging behind in the upregulation of gene expression compared with the wild-type allele in normal cells. The molecular explanation for this phenomenon is the presence of additional enhancers in the Spi1 locus that are activated later in differentiation by later-acting TFs (in this case, C/EBPα) that act independently of PU.1 and push up gene expression to wild-type levels [44]. A similar phenomenon was also seen at the Pinx1 gene, which is regulated by a switch from the early-acting TF GATA2 to the later-acting TF GATA1. The deletion of a primed enhancer which binds GATA2 results in a delay in the activation of the gene during erythroid differentiation up to a point when a second enhancer becomes active that binds GATA1 [12]. A fascinating example of how chromatin priming can be used to create functionally different cells is seen in the neuronal development of Caenorhabditis elegans. Here, the priming of an early enhancer by Notch signaling dictates the developmental timing of regulation of the isy-6 miRNA locus: it is upregulated earlier in those neuronal precursor cells that received the signal compared with those that did not, thus leading to a functional left–right asymmetry in otherwise morphologically symmetric neurons [52].

However, deregulation of timing of expression after the deletion of a primed enhancer is not seen at all loci. The deletion of the beta-globin LCR in its native context affects expression levels, but not the timing of developmental expression of the different globin genes regulated by TFs binding to the individual genes [53]. These experiments demonstrate that the order of the expression of TFs activating different enhancers dictates when and whether chromatin priming takes place. If this order is altered, cell differentiation is disturbed, as shown by experiments that switched the order of expression of two essential regulators of hematopoietic development, C/EBPα and GATA2 [54]. Together with the findings of the Matthias group [26], these experiments demonstrate that the activation of different enhancers and their cooperation over multiple differentiation stages dictates the kinetics of upregulation of specific genes.

**Maintaining an epigenetic memory of previous transcriptional activation**

In addition to the finding that chromatin priming is important for the correct temporal regulation of gene expression, a number of recent studies have shown that it is vital for establishing a transcriptional memory of previous gene activation [55]. Early studies by the Grange group [56] used in vivo DMS foot printing to demonstrate that the induction of the tyrosine aminotransferase (Tat) gene by glucocorticoids in the liver led to a demethylation of a binding site for the liver-specific TF FOXA1 within 2 days that persisted even after steroid removal. Once demethylation had occurred, subsequent reactivation of the genes occurred with much more rapid kinetics. A similar phenomenon was seen in macrophages, a cell type that is highly responsive to external stimuli such as proinflammatory cytokines or bacterial lipopolysaccharide (LPS) [57]. Stimulation of macrophages with LPS leads to activation of a number of enhancers from the silent chromatin state (inducible enhancers), which is concordant with the binding of inducible TFs. Importantly, once the stimulus is removed, factor binding on such enhancers is lost, but the formation of an active enhancer complex leaves an active chromatin mark behind that allows for a faster reactivation of associated genes after restimulation.

The importance of maintaining a memory of previous gene activation is paramount for cell types within the adaptive immune system. Both B and T cells retain a memory of previous encounters with specific antigens by reverting back to a quiescent state after the immune system has cleared an infection, but then patrolling the body as memory cells to respond much more rapidly once the same antigen is encountered again [58–61]. This feature of the adaptive immune system is the basis for vaccination [62].

Prior to their first cycle of activation during an immune response, naive T cells exist in a quiescent state with small nuclei and barely any cytoplasm. It was found that the chromatin landscape at the IL3/CSF2 locus in naive T cells is more similar to the inactive structure seen in the thymus than it is to the active structure seen in recently activated T cells [63–65]. The lack of chromatin modifications was consistent with an inability to activate these genes in naive T cells. However, once an appropriate immune response has been triggered by antigen-presenting cells, dormant T cells undergo a dramatic transformation by increasing their nuclear volume and activating a large battery of genes as they enter a phase of rapid proliferation. The transformation of naive T cells to T blast cells is driven by extensive chromatin remodeling [66,67], which leaves hundreds of immune response genes in a primed state [64,68,69]. Once these genes have been primed in activated T cells, they remain primed when T cells revert to the quiescent state as memory T cells [64,70]. Memory T cells have the ability to reactivate genes faster and in response to weaker stimuli than naive T cells [58,64,71–76].

The basis for the rapid recall response was until recently far from clear because the transcriptional network in resting-state memory T cells is not very different from that seen in naive T cells [70]. However, a recent study identified a simple mechanism that goes a long way toward accounting for the underlying basis of the rapid recall
response in both recently activated T cells and memory T cells [64]. This study performed extensive integrated analyses of accessible chromatin, together with chromatin immunoprecipitation studies and genome editing, to investigate the molecular basis and significance of chromatin priming in T cells [64]. We found that previously activated T cells acquire and stably maintain ~3000 primed DHSs that are closely associated with inducible immune response genes and the inducible enhancers that control them. As summarized in Figure 2, these primed DHSs exist as open chromatin regions occupied by constitutively expressed factors such as ETS-1 and RUNX1, which are also expressed in naive T cells, but only gain the opportunity to bind to these regions during the T blast cell transformation process, when inducible factors and remodelers first open up these primed DHSs [64]. TFs such as activating protein 1 (AP-1), the activity of which is signaling dependent (inducible), are implicated in the initial opening up of primed DHSs but, once established, RUNX1 and ETS-1 appear to be largely sufficient to maintain primed DHSs in both rapidly dividing T cells and quiescent circulating memory T cells in the absence of additional stimuli [64]. The primed DHSs identified in recently activated T cells do not typically function as transcriptional enhancers and function instead to maintain domains of accessible chromatin-carrying histone H3K4me2 and H3K27ac modifications. There is evidently a requirement for the presence of primed DHSs and active chromatin domains at inducible loci because the inducible enhancers found in association with primed DHSs can be induced in previously activated T cells but not in naive T cells [64,65]. This concept was confirmed by demonstrating that chromatin accessibility at the IL3 locus was diminished following the targeted deletion of a priming element located next to an inducible enhancer, leading to a reduction in the speed of IL3 transcriptional activation.

Taken together, these studies established the model presented in Figure 2 in which immunological memory is acquired via a hit-and-run process involving transient association of inducible factors that assist the recruitment of constitutively expressed factors, which then function to retain a transcriptional memory at inducible loci. Once established, both the primed DHSs and the adjacent enhancers can recruit inducible factors much faster within an active chromatin domain. The slow response in naive T cells is mainly due to the inaccessibility of immune response genes within condensed chromatin. Chromatin priming maintained by primed DHSs adds an additional dimension to previously defined mechanisms in memory T cells, which include the replacement of repressive PcG complexes by activating Trithorax group complexes at a subset of immune response genes [77], increased mRNA stability [78], and enhanced receptor signaling [72,79]. This priming model is reminiscent of the process termed “assisted loading” in which steroid receptors act in a transient manner to open up chromatin to enable the binding of other factors [80]. A similar hit-and-run mechanism of transcriptional memory was defined in plants, in which transient binding of a heat shock factor led to stable maintenance of active chromatin modifications [81].
Interestingly, although epigenetic priming in memory T cells is independent of repeated T-cell receptor signaling, it does require intermittent episodes of cytokine signaling to reinforce the pattern of reprogramming continually. In mice and in humans, immunological memory decays in CD4+ memory T cells in the absence of the tumor necrosis factor superfamily receptor OX40 [82,83]. This implies that intermittent signaling from receptors to inducible factors such as AP-1 is required for the stable long-term maintenance of immunological memory, which in humans lasts for decades.

An open chromatin state has to be maintained actively because, in the absence of activators, chromatin adopts an inactive state (reviewed in Obier et al. [84]). The role of PcG in maintaining gene expression patterns during cell division is well established (reviewed in Geisler and Paro [85]), but, as outlined above, these proteins play little role in the priming of distal elements. A question that is therefore intrinsically associated with chromatin priming is how a primed state is maintained during cell division when the chromatin landscape is disrupted by the replication machinery and has to be reassembled [86]. A recent study has shed light on this question by showing that highly accessible chromatin regions are retained in mitotic chromatin [87]. Such regions colocalize with TF-binding sites, indicating that factor binding may be retained. It has now been shown for a number of TFs and cofactors that they are capable of binding to mitotic chromatin and thus serve as “bookmarks” for the reassembly of factor complexes (reviewed in Kadoake and Blobel [88]).

Consequences
The studies described above have had a profound impact on our mechanistic understanding of how we envisage the genome to regulate developmental and immunological processes because they highlight the fact that the interplay between TFs and a differential chromatin landscape is an essential component of the temporal regulation of gene activation in development. The idea that the chromatin landscape has a profound influence on TF access and thus on the dynamics of gene expression was suggested a long time ago [89], but it is now clear from additional studies [55] in yeast [90,91] and mammals [92], to quote a few examples, that control of chromatin accessibility is central for the fine-tuning of gene activation. Another consequence of chromatin priming as observed in T cells is the fact that extracellular signals can influence the kinetics of gene activation profoundly via the maintenance of an open chromatin state. In addition, aberrant or aberrantly expressed TFs, as found in hematological cancers, will derail the finely tuned mechanisms that activate genes, thus leading to a block in differentiation. Last, but not least, the vast majority of sequence variants between individuals and common disease variations affect cis-regulatory elements with an immediate consequence for TF binding [93,94]. Life is robust and has therefore built in redundancy into the system by, for example by using enhancer duplications, thereby buffering overly deleterious effects of such defects [95]. However, sequence variants will inevitably also affect priming elements and thus lead to a variable and individual response to signals and alterations of the temporal regulation of gene expression. To decipher the precise molecular mechanisms of how such variations influence morphology and physiology and cause disease will be a paramount research task in the coming years.

Acknowledgments
Research in the Bonifer and Cockerill laboratories is funded by grants from Bloodwise, the Biotechnology and Biological Sciences Research Council, the Kay Kendall Leukaemia Fund, and the Medical Research Council.

References


70. Curtinsinger JM, Lins DC, Mescher MF. CD8+ memory T cells (CD44high, Ly-6C−) are more sensitive than naive cells to (CD44low, Ly-6C+) to TCR/CD8 signaling in response to antigen. J Immunol. 1998;160:3236–3243.


