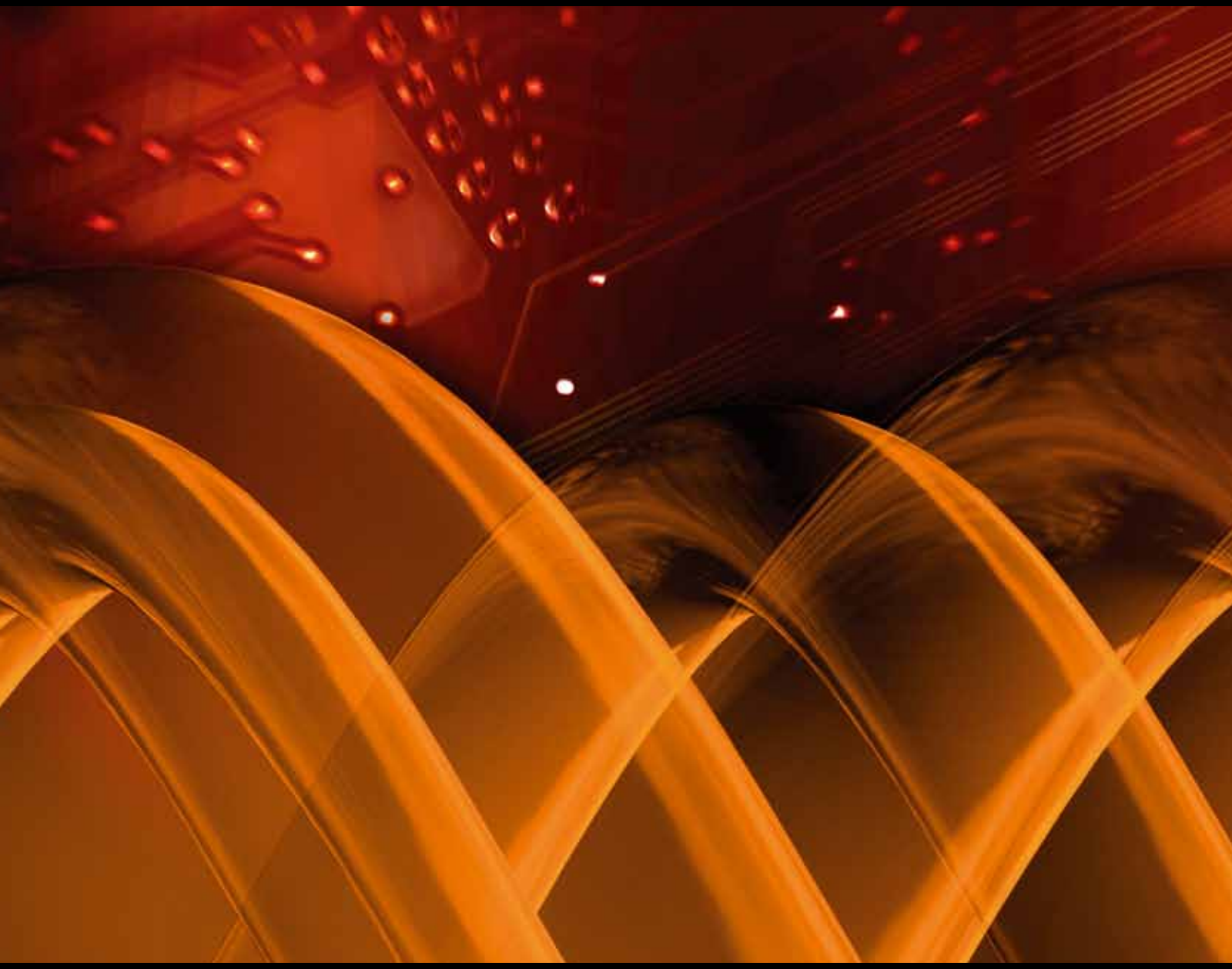
Comparative and Functional Genomics

EPIGENETIC CONTROL OF REPROGRAMMING AND CELLULAR DIFFERENTIATION

Guest Editors: Lucia Latella, Daniela Palacios, Sonia Forcales, and Pier Lorenzo Puri



Epigenetic Control of Reprogramming and Cellular Differentiation

Comparative and Functional Genomics

Epigenetic Control of Reprogramming and Cellular Differentiation

Guest Editors: Lucia Latella, Daniela Palacios, Sonia Forcales, and Pier Lorenzo Puri

Editorial Board

Contents

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 538639, 2 pages
doi:10.1155/2012/538639

Editorial

Epigenetic Control of Reprogramming and Cellular Differentiation

Lucia Latella,[1,2] Daniela Palacios,[1] Sonia Forcales,[3] and Pier Lorenzo Puri[1,4]

[1] *Dulbecco Telethon Institute (DTI), IRCCS Fondazione Santa Lucia and European Brain Research Institute, 00143 Rome, Italy*
[2] *National Research Council, Institute of Translational Pharmacology, via Fosso del Cavaliere 100, 00133 Rome, Italy*
[3] *Institute of Predictive and Personalized Medicine of Cancer (IMPPC), 08916 Badalona, Spain*
[4] *Sanford-Burnham Institute for Medical Research, La Jolla, CA 92037, USA*

Correspondence should be addressed to Pier Lorenzo Puri, plpuri@dti.telethon.it

Received 7 August 2012; Accepted 7 August 2012

Recently, numerous interdisciplinary approaches that include integration of bioinformatics, epigenetics, and molecular and cellular biology have started to shed light on the complex dynamics that regulate cellular processes such as cell "stemness," reprogramming, and differentiation. From the sequencing of the first complete genomes to the recent high-throughput genomic and proteomic approaches, a huge amount of information has been generated to improve our understanding of how the eukaryotic genome works to activate and coordinate specific programs of gene expression.

The focus of this review is to provide an overview of the epigenetic regulatory mechanisms controlling crucial cellular fate decisions, in particular during differentiation and reprogramming. A special emphasis has been given to skeletal myogenesis, as it provides an amenable model for in vitro and in vivo studies.

Y. Ito et al. in "*A system approach and skeletal myogenesis*" provide an overview of the most common approaches currently used to address transcriptional regulation. They discuss deep sequencing and microarray-based methods, cell-based high throughput assays, and the generation of *in situ* gene expression databases as potent tools to dissect complex networks of gene expression. Using skeletal muscle as a paradigm, they highlight the main discoveries obtained through the different methods and underline the advantages of an integrated approach to obtain a full picture of cellular differentiation.

An increasing number of studies suggest an epigenetic component in the development of many human diseases. In the review article "*Epigenetic alterations in muscular disorders,*" Dr. C. Lanzuolo discusses how altered epigenetic mechanisms contribute to human pathology and in particular to the development and progression of neuromuscular disorders. Muscular dystrophies arise from either mutations of proteins that play a role in the integrity of the sarcolemma—the plasma membrane that wraps the myofibers in skeletal muscles—or are due to a dysfunction within the cell nucleus. Dr. C. Lanzuolo focuses on the so-called nuclear muscular dystrophies and she analyzes the possible contribution of an important group of epigenetic modifiers, the polycomb-group (PcG) complexes, to the pathogenesis of these diseases. The dissection of the aberrant epigenetic profiles associated to several human pathologies has contributed to the design of experimental therapies with epigenetic drugs.

The review "*Tackling skeletal muscle cells epigenome in the next-generation sequencing era*" by R. Fittipaldi and G. Caretti points to the significance of genome-wide study of the epigenetic modifications and the transcription factors that are activated during skeletal muscle differentiation. The authors discuss on the new genome-wide technologies to monitor the epigenetic landscape that characterizes myoblast-to-myotube transition in order to obtain innovative information on molecular mechanisms orchestrating biological processes, such myogenic differentiation.

The fine regulation of a myogenic gene such as *myogenin* is discussed in "*Turning on myogenin in muscle: A paradigm for understanding mechanisms of tissue-specific gene expression*" by H. Faralli and F. J. Dilworth. Transcription factors, cofactors, and epigenetic modifications contribute

to regulate muscle-specific gene expression during development. In particular, the authors highlight the fact that a tissue-specific factors such as MyoD cooperate with more ubiquitously expressed proteins (Six4, Mef2, and Pbx1) to establish a transcriptionally permissive environment to ensure proper spatiotemporal *myogenin* expression.

The current challenge of epigenetics is how to analyse and extract biological knowledge from the large volumes of data produced with new high-throughput technologies. In this regard, the paper of S. Althammer et al., "*Predictive models of gene regulation from high-throughput epigenomics data*" describes a computational framework using integrative tools and machine learning algorithms that facilitate the systematic analysis of high-throughput sequencing epigenetic data for integrative studies. Their analysis shows a different epigenetic code of expression for intron-less and intron-containing genes, with more prominent differences for genes with low GC content (LGC) in their promoter, showing for instance that at the promoter regions of LGC intron-less genes, H3K36me3 and H3K4me1 are the most informative marks while in polyadenylation sites of expressed genes, the H3K36me3 signal is much weaker than in intron-containing genes. Based on epigenetic data, their model predicts very strongly gene expression between two conditions; therefore, this approach can be applied to study gene expression in different developmental stages, disease states, or treatments.

Of tremendous relevance for its therapeutic potential, is the possibility to manipulate cell identity. In the review titled "*The stability of the induced epigenetic programs*", M. J. Barrero discusses the epigenetic changes that occur during differentiation and the way back upon cellular reprogramming emphasizing on the thin regulation of pluripotent and lineage-specific genes expression to ensure a correct differentiation during development and in adult life.

"*The epigenetic regulation of B lymphocyte differentiation, transdifferentiation, and reprogramming*" by B. Barneda-Zahonero et al. reviews the hierarchical network of transcription factors that mediate the epigenetic signature regulating the transcriptome during B cell development. B lymphocytes are highly specialized terminally differentiated cells responsible for generating high-affinity antibodies, providing the humoral immunity against pathogens. Every day the human body generates millions of B lymphocytes from precursor hematopoietic stem cells by a differentiation process that is a tightly regulated. This multistep differentiation process, vital to immunity, combines the successive expression of specific transcriptional factors and associated epigenetic changes to restrict the developmental potential of lymphoid progenitors to the B cell pathway by repressing B-lineage-inappropriate genes, while simultaneously promoting B cell development by activating B-lymphoid-specific genes.

Lucia Latella
Daniela Palacios
Sonia Forcales
Pier Lorenzo Puri

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 759407, 7 pages
doi:10.1155/2012/759407

Review Article

A Systems Approach and Skeletal Myogenesis

Yoshiaki Ito,[1,2] Tomohiro Kayama,[1,3] and Hiroshi Asahara[1,2]

[1] *Department of Systems BioMedicine, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Bunkyo, Tokyo 113-8510, Japan*
[2] *Department of Systems BioMedicine, National Research Institute for Child Health and Development, Setagaya, Tokyo 157-8535, Japan*
[3] *Department of Orthopaedic Surgery, The Jikei University School of Medicine, Minatoku, Tokyo 105-8461, Japan*

Correspondence should be addressed to Hiroshi Asahara, asahara.syst@tmd.ac.jp

Received 21 March 2012; Revised 27 May 2012; Accepted 28 June 2012

Academic Editor: Daniela Palacios

Skeletal myogenesis depends on the strict regulation of the expression of various gene subsets. Therefore, the understanding of genome wide gene regulation is imperative for elucidation of skeletal myogenesis. In recent years, systems approach has contributed to the understanding of various biological processes. Our group recently revealed the critical genome network of skeletal myogenesis by using a novel systems approach combined with whole-mount *in situ* hybridization (WISH) database, high-throughput screening, and microarray analysis. In this paper, we introduce our systems approach for understanding the myogenesis regulatory network and describe the advantages of systems approach.

1. Introduction

Skeletal muscle is indispensable for any moving function of the human body, and abnormality of the skeletal muscle causes great disability in affected people. It is therefore important to understand the mechanism of skeletal myogenesis so that it may form a basis for disease treatment.

Almost all skeletal muscles in the body derive from dermomyotome or myotome in somites. The myotome and dermomyotome contain myogenic progenitor cells that evolve into skeletal muscles, aggregates of myofibers, in the whole body. During skeletal myogenesis, myofibers form from myogenic progenitors, where distinct subsets of genes are activated or repressed and form a complex molecular network of interdependent pathways [1–3]. These processes are mainly regulated by the muscle-specific basic helix-loop-helix (bHLH) transcription factors, MyoD, Myf5, Myogenin (Myog), and Mrf4. Analysis of null mice of these genes suggested that MyoD and Myf5 play a role in determining the myogenic progenitors to myoblasts [4]. Myog is important in differentiation from myoblasts to myotubes [5, 6], and Mrf4 is important in both determination and differentiation [7]. The described transcription factors are class II (tissue-specific) bHLH transcription factors capable of either homodimerization or heterodimerization with class I bHLH factors, such as E-proteins HEB/HTF4, E2-2/ITF-2, and E12/E47 [8]. All bHLH dimers bind to an E-box, a consensus sequence comprised of the sequence CANNTG. Id proteins have been identified to act as myogenic antagonists by directly binding to E-proteins and/or muscle-specific bHLH proteins, blocking their ability to bind E-boxes and activate transcription at muscle-specific promoters [9–11]. IdmRNAs are detected in proliferating skeletal muscles and are downregulated in differentiated muscle cultures [9, 12]. This downregulation was thought to be important for skeletal muscle formation; however, the mechanism of Id repression has not been understood for almost 20 years.

Recently, we revealed the Id downregulation mechanism in myogenesis by our own systems approach combined with WISH database, high-throughput screening, and microarray analysis [13]. Systems approach, a systematic study using various comprehensive analyses such as high-throughput sequencing technologies, genome wide cell-based assays, and bioinformatics, has allowed us to expand our knowledge of life phenomenon. We have reviewed studies that have utilized systems approach. In addition to this, we describe our own

systems approach and how this has helped in understanding skeletal myogenesis.

2. Deep Sequencing and Array-Based Approaches

High-throughput sequencing technologies allow high-resolution, genome wide investigation of epigenetic conditions. For instance, mapping of open chromatin regions, histone modifications, and DNA methylation across a whole genome is now feasible, and whole transcripts including noncoding RNAs (ncRNAs) can be identified via RNA sequencing.

These high-throughput sequencing-based technologies and microarray-based ChIP chip analyses are used in various fields, and there have been reports on embryonic stem (ES) cells. Meissner et al. analyzed genome-scale DNA-methylation profiles and histone methylation patterns of mouse ES cells and differentiated cells by using high-throughput bisulphite sequencing and ChIP-sequence [14]. This revealed that DNA methylation patterns are better correlated with histone methylation patterns than with the underlying genome sequence context and that methylation of CpGs is one of dynamic epigenetic marks during differentiation particularly in regulatory regions outside of core promoters [14]. Also, Bock et al. analyzed DNA methylation patterns and gene expression of 20 human ES cell lines and 12 human iPS cell lines, identifying epigenetic and transcriptional similarity of ES and iPS cells [15]. Bernstein et al. mapped Polycomb-associated Histone H3 Lysine 27 trimethylation (H3K27me3) and Trithorax-associated Histone H3 Lysine 4 trimethylation (H3K4me3) across the whole genome in mouse ES cells by ChIP-chip analysis [16]. H3K27me3 is an epigenetic mark that mediates gene silencing, whereas H3K4me3 occurs in nucleosomes found in the promoter regions of actively transcribed genes. They identified a specific modification pattern consisting of large regions of H3K27me3 harboring smaller regions of H3K4me3. It has been proposed that this "active" and "repressive" modification pattern represents genes specifically designed to initiate transcription, and this active state is thought to be essential for the developmental potential of ES cells. Pan et al. also mapped H3K27me3 and H3K4me3 across the whole genome in human ES cells [17]. The vast majority of H3K27me3 colocalized on genes modified with H3K4me3 as within mouse ES cells. These commodified genes displayed low expression levels and were enriched in developmental gene function. Another significant set of genes lacked both modifications, also expressed at low levels in ES cells, but was enriched for gene function in physiological responses rather than development. Commodified genes change expression levels rapidly during differentiation, but so do a substantial number of genes in other modification categories. Pluripotency-associated genes such as SOX2, OCT4, and NANOG shifted from modification by H3K4me3 alone to colocalization of both modifications as they were repressed during differentiation. These data revealed that H3K27me3 modifications change during early differentiation, both relieving existing repressive domains and imparting new ones, and that colocalization with H3K4me3 is not restricted to pluripotent cells. High-throughput sequencing technologies are also used in the studies on genome-wide ncRNA expression analysis. Calabrese et al. analyzed short RNA expression in Dicer-positive and Dicer-knockout mouse ES cells [18]. From quantification of miRNA levels, they estimated that there are 130,000 5′ phosphorylated short RNAs per ES cell. 15% of these RNAs are generated independently of the Dicer gene, presumed breakdown products of mRNAs, which are low in abundance and consist of highly repetitive sequences. The remaining 85% of 5′ phosphorylated ES cell short RNAs consist of miRNAs or miRNA-like species that depend on Dicer for biogenesis. The majority of ES cell miRNAs appear to be generated by six distinct loci, four of which have been implicated in cell cycle control or oncogenesis. At a depth of sequencing that approaches the total number of 5′ phosphorylated short RNAs per cell, miRNAs appeared to be Dicer's only substrate. These studies identified genome-wide epigenetic marks and gene expression in ES cells. They have obtained data revealing the characteristics of ES cells and also incidentally discovered "active" and "repressive" histone comodification patterns. This had been possible due to a genome-wide analysis, thus indicating the importance of such approach.

The systems approach is also beneficial to reveal the regulatory network of skeletal myogenesis. Myogenesis is orchestrated through a series of transcriptional controls regulated by the myogenic bHLH factors. Several groups performed ChIP-chip analysis to identify targets of myogenic regulatory factors [1, 2]. These analyses indicated overlapping of distinct targets of MyoD and Myog suggesting the mechanism of sequential expression during myogenesis. At early myogenesis, MyoD is sufficient for activation of the expression, and these genes are expressed immediately after MyoD induction. On the other hand, during late myogenesis, MyoD initiates only regional histone modification. Myog does not bind without MyoD, and the expression of late genes requires both MyoD and Myog. In recent years, genome-wide MyoD target profiling using ChIP-sequence analysis has been reported [19]. High-throughput sequencing technology-based ChIP-sequence analysis suggested over 20,000 MyoD-binding sites, greater than with the array-based ChIP-chip analysis [19]. This analysis identified that MyoD was constitutively bound to thousands of sites in both myoblasts and myotubes and that the genome wide MyoD binding was related with regional histone acetylation [19]. This suggests that myogenic master regulator MyoD genome widely acts to alter the epigenome in myoblasts and myotubes. Gagan et al. also performed high-throughput sequencing-based analysis, to find that MyoD binds to the microRNA-378 (miR-378) gene locus and induces transactivation and chromatin remodeling [20]. This activated miRNA directly downregulates the MyoR, a MyoD antagonist, and promotes myogenesis [20].

Genome-wide target gene analyses are also performed in transcription regulators other than the myogenic bHLH factors. Lagha et al. performed ChIP-chip analysis of the transcription factor Pax3 [21], which is essential for ensuring myogenic potential and survival of the progenitors [22]. Pax3

binds to a sequence 3′ of the Fgfr4 gene that directs Pax3-dependent expression at sites of myogenesis in transgenic mouse embryos. The activity of this regulatory element is also partially dependent on E-boxes, targets of the myogenic regulatory factors, which are expressed as progenitor cells entering the myogenic program. Other FGF signaling components, notably Sprouty1, are also regulated by Pax3. These results provide new insight into the Pax-initiated regulatory network that modulates stem cell maintenance versus tissue differentiation. Soleimani et al. performed ChIP-seq analysis of Pax3 and Pax7 [23]. These transcription factors regulate stem cell function in skeletal myogenesis, but little is known about the molecular mechanism of their distinct roles. The genome-wide binding-site analysis combined with gene expression data indicates that both Pax3 and Pax7 bind identical DNA motifs and jointly activate a large number of genes involved in muscle stem cell function. In adult myoblasts, Pax7 binds to many more sites than the number of genes it regulates. In spite of a significant overlap in their transcriptional network, Pax7 regulates distinct set of genes involved in the acceleration of proliferation and inhibition of myogenic differentiation. Moreover, they showed that Pax7 has a higher binding affinity to the homeodomain-binding motif relative to Pax3, suggesting that the differences in DNA binding contribute to the observed functional difference between Pax3 and Pax7 binding in myogenesis. Mousavi et al. performed ChIP-seq of Polycomb group (PcG) protein Ezh1 and mRNA-seq in skeletal muscle cells [24]. This study provides evidence for genome-wide association of Ezh1 complex with active epigenetic mark (H3K4me3), RNA polymerase II (Pol II), and mRNA production. Although Ezh2, a paralog of Ezh1, is a known trigger for transcription repression by catalyzing the addition of methyl groups onto H3K27 [25], these findings reveal another role for PcG complex in promoting mRNA transcription.

The genome-wide approach also contributes to further understanding of the epigenetic regulation in skeletal myogenesis. Asp et al. examined changes in the chromatin landscape during myogenesis by ChIP-seq analyses of several key histone marks (H3K9Ac, H3K18Ac, H4K12Ac, H2Bub, H3K4me1, H3K4me2, H3K4me3, H3K27me3, and H3K36me3) and RNA polymerase II in mouse myoblasts and myotubes [26]. Using the data, they identified novel regulatory elements flanking the *Myog* gene that act as a key differentiation-dependent switch in myogenesis. *Myog* gene is targeted by PRC2-mediated H3K27 methylation, and its expression is suppressed in myoblasts. Depletion of Suz12, a component of PRC2 complex that regulates H3K27 methylation, led to the loss of PRC2 and H3K27me3 on Myog, resulting in premature and enhanced gene induction. This histone mark could represent part of a methylation-acetylation differentiation switch, determining the timing of expression of *Myog* and therefore terminal differentiation. Vethantham et al. also performed ChIP-seq analyses of H2Bub, H3K4me3, and H3K79me3 during myogenesis [27]. Ubiquitylation of H2B on lysine 120 (H2Bub) is associated with active transcriptional elongation. H2Bub has been implicated in histone crosstalk and is generally thought to be a prerequisite for trimethylation of H3K4 and H3K79 in both yeast and mammalian cells. The genome-wide analysis of epigenetic marks identified dynamic loss of H2Bub in the differentiated state. Moreover, they found that the H2B ubiquitin E3 ligase, RNF20, was depleted from chromatin in differentiated myotubes, indicating that recruitment of this protein to genes significantly decreases during myogenesis. Furthermore, they observed retention and gaining of H3K4 trimethylation on multiple genes in the absence of H2Bub. The Set1 H3K4 trimethylase complex was efficiently recruited to a subset of genes in myotubes in the absence of H2Bub, suggesting that H3K4me3 in the absence of H2Bub in myotubes is mediated via Set1.

Trapnell et al. performed RNA-seq analysis in mouse myoblast cell line representing a differentiation time series [28]. They detected 13,692 known transcripts and 3,724 previously unannotated transcripts. Analysis of transcript expression over the time series revealed complete switches in the dominant transcription start site or splice isoform in 330 genes, along with more subtle changes in further 1,304 genes.

Overall, deep sequencing or array-based approaches have been shown to be of benefit in identifying the molecular network and novel effectors in diverse biological processes. In skeletal myogenesis, these approaches revealed comprehensive target genes of myogenic transcription factors, novel myogenic factors and the characteristics of myoblasts and myotubes, which could not be identified by conventional approaches.

3. Cell-Based High-Throughput Assay

Currently, multiple studies have demonstrated comprehensive and cell-based functional screening. Generally, screening for signals activating gene expression consists of examining potential transcription factor-binding sequences in a specific promoter using bioinformatics and reporter assays. If a factor's potential recognition motif is unknown, one-hybrid or South-western screening can be used to identify molecules directly associated with the specific sequence. However, these methods are limited to identifying direct targets only. On the other hand, cell-based reporter assays using a comprehensive set of cDNAs in an expression library allow high-throughput screening not only for direct transcriptional regulators but also for other factors, such as cell-signaling molecules, receptors, and growth factors. Chanda et al. performed a reporter assay-based approach that used about 20,000 annotated cDNAs in the investigation of activator protein-1 (AP-1) signal transduction pathway and identified novel factors of AP-1 mediated growth and mitogenic response pathway [29]. Fiscella et al. performed high-throughput assay using a unique library of cDNAs encoding predicted secreted and transmembrane domain-containing proteins [30]. Supernatants from mammalian cells transiently transfected with this library were incubated with primary T cells and T cell lines in several high-throughput assays including reporter and cytokine secretion assay. This identified a T cell factor, TIP (T cell immunomodulatory protein), which does not show any homology to proteins with known function. However, treatment of primary human and murine T cells with TIP resulted in the secretion of IFN-γ, TNF-α, and IL-10, whereas

in vivo TIP had a protective effect in a mouse acute graft-versus-host disease (GVHD) model. Konig et al. performed a systematic approach combined with genome-wide siRNA analysis and searched the human interactome database, to uncover multiprotein virus-host interactions that are likely to regulate the early steps of HIV infections [31].

In the myogenesis study, we performed cell-based high-throughput transfection assay to identify activation factors of RP58, a critical myogenesis regulator as described in the latter section.

4. *In Situ* Gene Expression Database

Microarray analysis is a powerful tool to identify the working genes in individual cells or tissues. However, this analysis is unlikely to detect gene expression restricted to small areas. In contrast, *in situ* hybridization can identify temporal and spatial gene expression patterns. The systematic *in situ* hybridization database contributes to detailed information for the spatial regulation of gene expression. Gray et al. mapped the expression of 1174 transcription factors in the brain of developing mice using section *in situ* hybridization [32]. Also, Lein et al. described an anatomically comprehensive digital atlas containing the expression patterns of around 20,000 genes using automated high-throughput procedures for *in situ* hybridization in the adult mouse brain [33]. These databases describe the anatomical organization of the brain and provide a primary data resource for a wide variety of further studies regarding brain organization and its function.

The Edinburgh Mouse Atlas Gene-Expression Database (EMAGE) is a large-scale database of *in situ* gene expression patterns of about 16,000 genes in the developing mouse embryo [34–37]. Domains of expression from raw data images are spatially transferred into a set of standard 3D virtual mouse embryos at different stages of development. Anatomy ontology is also used to describe sites of expression, which allows data to be queried using text-based methods. The GenitoUrinary Development Molecular Anatomy Project (GUDMAP) is also a database of *in situ* gene expression patterns in mouse embryos [38, 39]. GUDMAP includes whole-mount and section *in situ* hybridization data of over 3,000 genes and microarray gene expression data of microdissected, laser-captured, and FACS-sorted components of the developing mouse genitourinary (GU) system. These *in situ* gene expression databases provide more detailed information on the spatial regulation of gene expression and allow identification of discrete clusters of transcribed genes. They serve as a useful source for research in developmental biology.

5. Our Systems Approach Revealed the MyoD-Mediated Ids Repression Mechanism

We constructed a unique systems approach and applied it for elucidation of myogenesis molecular network. First, we created our own *in situ* gene expression database. To identify and characterize effectors of the transcriptional network regulating developmental processes, we developed a web-based comprehensive WISH database for transcriptional regulators using E9.5, 10.5, and 11.5 mouse embryos [13]. We prepared 1520 digoxigenin-labeled RNA probes from cDNA libraries. Using WISH results, we annotated gene expression patterns of each gene and constructed a database, termed "EMBRYS" (http://embrys.jp/embrys/html/MainMenu.html), covering these 3 embryonic days. Using this database, we identified 43 transcription regulators showing myogenic expression pattern in the limb bud. Among those, transcription repressor *RP58* was identified as a novel transcription factor expressed in myogenesis [13]. The analysis of *RP58* knockout mice revealed that this gene is critical for myogenesis [13]. This database EMBRYS is also useful to identify regulators of another tissue development. Indeed, we also identified that *Mohawk homeobox* gene is a critical regulator of tendon differentiation by using the database [40].

The WISH database EMBRYS identified a novel transcriptional factor RP58 as a critical regulator of myogenesis. To identify the molecular network anchored by RP58, we investigated the upstream events that promote RP58 expression by cell-based expression vector library transfection assay [13]. We utilized around 6000 arrayed and addressable cDNA clones, which allowed systematic, efficient, and unbiased screening of cDNA encoding factors that could activate the RP58 promoter. A highly conserved RP58 genomic region was inserted in front of luciferase gene in the reporter vector. This was then transfected in 293T cells with expression vector library, and luciferase assay was performed [13]. The high-throughput transfection assay identified myogenic bHLH factor MyoD as a direct transcription activator of RP58 [13].

RP58 has been reported to bind to the specific DNA sequence (A/C)ACATCTG(G/T)(A/C) [41] and is associated with Dnmt3a and Hdac1 [42]. These reports suggest that RP58 can bind to the promoter region of its target genes and repress transcription activity. To identify the repression targets of RP58, we performed microarray analysis and bioinformatics screening by RP58 binding sequence and identified Id2 and Id3 as RP58 repression targets [13].

Our systems approach combined with WISH database construction, high-throughput transfection assay, and microarray analysis identified a critical regulatory network of myogenesis (Figure 1). WISH database identified a novel myogenic regulator, RP58. High-throughput transfection screening and microarray analysis identified a MyoD-activated regulatory loop by RP58-mediated repression of myogenic bHLH factor inhibitors Id2 and Id3. In myoblasts, Ids are expressed and inhibit the myogenic bHLH factors. During myogenesis, RP58 is promoted by MyoD and represses Id transcription. Myogenesis then progresses by myogenic bHLH factor-mediated activation of muscle-specific genes (Figure 2). The repression mechanism of Ids had been unclear for almost 20 years, and this new finding indicates the importance of this systems approach.

6. Conclusion

A genome-wide systematic approach using high-throughput sequencing technologies, cell-based transfection assays, or

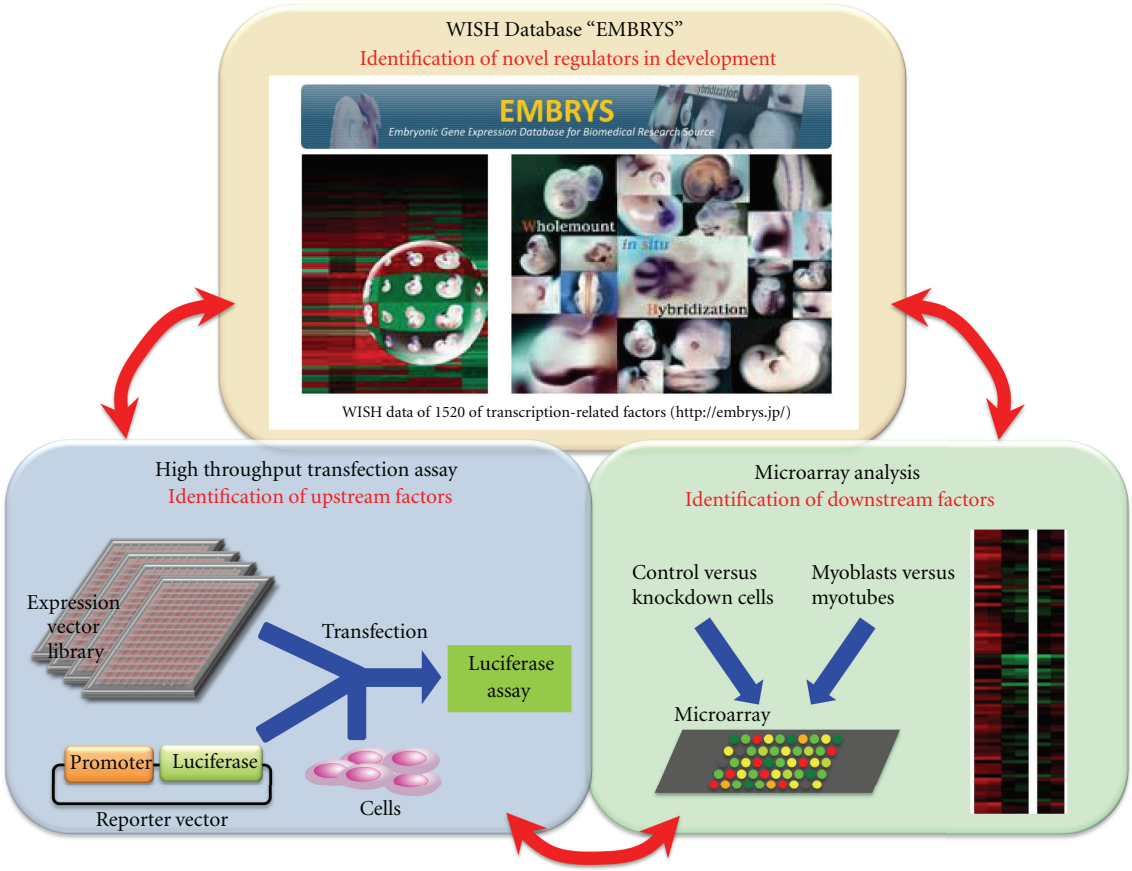


Figure 1: Scheme of our systems approach. WISH database EMBRYS identified a novel myogenesis regulator RP58. High-throughput transfection assay and microarray analysis identified upstream and downstream factors of RP58. This multicombined approach is useful for elucidation of molecular network in the developmental process.

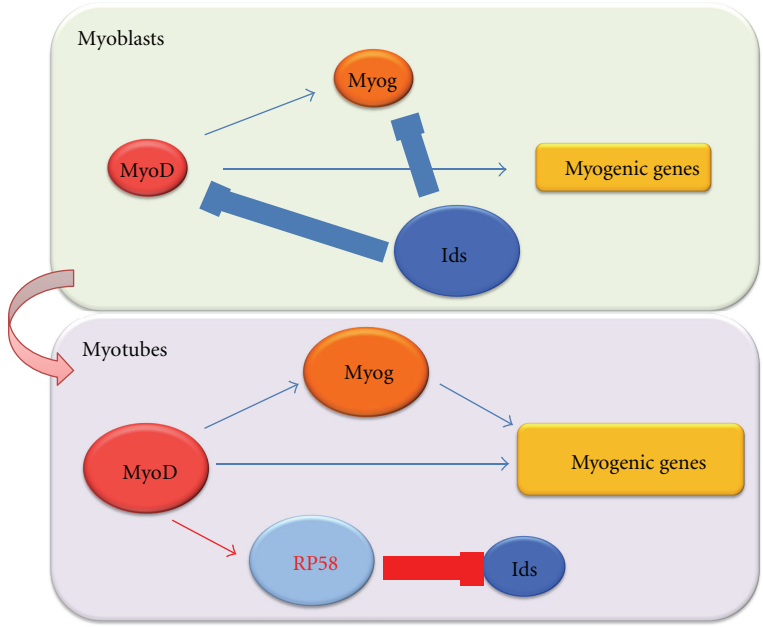


Figure 2: Proposed myogenesis regulatory network by our systems approach in myoblasts; Id proteins are expressed and inhibit the myogenic bHLH factors. During myogenic differentiation, RP58 is promoted by MyoD and represses the Id transcription. Muscle specific genes are then activated by myogenic bHLH factors.

construction of gene expression pattern database is contributing to understanding the mechanisms of various life phenomena. These methods have also been shown to be useful in studying skeletal myogenesis. High-throughput sequencing-based technologies showed genome-wide target genes of myogenesis regulators and epigenetic modification in skeletal myogenesis. We also identified a novel myogenesis network regulated by RP58 using the multicombined approach. Although the myogenesis study using systems approach is still at its early stages, the systems approach will enable further understanding of myogenesis in the future.

References

[1] A. Blais, M. Tsikitis, D. Acosta-Alvear, R. Sharan, Y. Kluger, and B. D. Dynlacht, "An initial blueprint for myogenic differentiation," *Genes & Development*, vol. 19, pp. 553–569, 2005.

[2] Y. Cao, R. M. Kumar, B. H. Penn et al., "Global and gene-specific analyses show distinct roles for Myod and Myog at a common set of promoters," *The EMBO Journal*, vol. 25, pp. 502–511, 2006.

[3] B. H. Penn, D. A. Bergstrom, F. J. Dilworth, E. Bengal, and S. J. Tapscott, "A MyoD -generated feed-forward circuit temporally patterns gene expression dining skeletal muscle differentiation," *Genes and Development*, vol. 18, no. 19, pp. 2348–2353, 2004.

[4] M. A. Rudnicki, P. N. J. Schnegelsberg, R. H. Stead, T. Braun, H. H. Arnold, and R. Jaenisch, "MyoD or Myf-5 is required for the formation of skeletal muscle," *Cell*, vol. 75, no. 7, pp. 1351–1359, 1993.

[5] Y. Nabeshima, K. Hanaoka, M. Hayasaka et al., "*Myogenin*gene disruption results in perinatal lethality because of severe muscle defect," *Nature*, vol. 364, pp. 532–535, 1993.

[6] P. Hasty, A. Bradley, J. H. Morris et al., "Muscle deficiency and neonatal death in mice with a targeted mutation in the *myogenin* gene," *Nature*, vol. 364, pp. 501–506, 1993.

[7] L. Kassar-Duchossoy, B. Gayraud-Morel, D. Gomès et al., "Mrf4 determines skeletal muscle identity in Myf5: myod double-mutant mice," *Nature*, vol. 431, no. 7007, pp. 466–471, 2004.

[8] C. A. Berkes and S. J. Tapscott, "MyoD and the transcriptional control of myogenesis," *Seminars in Cell & Developmental Biology*, vol. 16, no. 4-5, pp. 585–595, 2005.

[9] B. Chen, B. H. Han, X. H. Sun, and R. W. Lim, "Inhibition of muscle-specific gene expression by Id3: requirement of the C-terminal region of the protein for stable expression and function," *Nucleic Acids Res*, vol. 25, no. 2, pp. 423–430, 1997.

[10] K. Langlands, X. Yin, G. Anand, and E. V. Prochownik, "Differential interactions of Id proteins with basic-helix-loop-helix transcription factors," *Journal of Biological Chemistry*, vol. 272, no. 32, pp. 19785–19793, 1997.

[11] I. N. Melnikova, M. Bounpheng, G. C. Schatteman, D. Gilliam, and B. A. Christy, "Differential biological activities of mammalian Id proteins in muscle cells," *Experimental Cell Research*, vol. 247, no. 1, pp. 94–104, 1999.

[12] R. Benezra, R. L. Davis, D. Lockshon, D. L. Turner, and H. Weintraub, "The protein Id: a negative regulator of helix-loop-helix DNA binding proteins," *Cell*, vol. 61, no. 1, pp. 49–59, 1990.

[13] S. Yokoyama, Y. Ito, H. Ueno-Kudoh et al., "A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58," *Developmental Cell*, vol. 17, no. 6, pp. 836–848, 2009.

[14] A. Meissner, T. S. lsen, H. Gu et al., "Genome-scale DNA methylation maps of pluripotent and differentiated cells," *Nature*, vol. 454, pp. 766–770, 2008.

[15] C. Bock, E. Kiskinis, G. Verstappen et al., "Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines," *Cell*, vol. 144, no. 3, pp. 439–452, 2011.

[16] B. E. Bernstein, T. S. Mikkelsen, X. Xie et al., "A bivalent chromatin structure marks key developmental genes in embryonic stem cells," *Cell*, vol. 125, no. 2, pp. 315–326, 2006.

[17] G. Pan, S. Tian, J. Nie et al., "Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells," *Cell Stem Cell*, vol. 1, no. 3, pp. 299–312, 2007.

[18] J. M. Calabrese, A. C. Seila, G. W. Yeo, and P. A. Sharp, "RNA sequence analysis defines Dicer's role in mouse embryonic stem cells," *Proceedings of the National Academy of Sciences of the USA*, vol. 104, no. 46, pp. 18097–18102, 2007.

[19] Y. Cao, Z. Yao, D. Sarkar et al., "Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming," *Developmental Cell*, vol. 18, no. 4, pp. 662–674, 2010.

[20] J. Gagan, B. K. Dey, R. Layer, Z. Yan, and A. Dutta, "MicroRNA-378 targets the myogenic repressor MyoR during myoblast differentiation," *Journal of Biological Chemistry*, vol. 286, no. 22, pp. 19431–19438, 2011.

[21] M. Lagha, J. D. Kormish, D. Rocancourt et al., "Pax3 regulation of FGF signaling affects the progression of embryonic progenitor cells into the myogenic program," *Genes & Development*, vol. 22, pp. 1828–1837, 2008.

[22] M. Buckingham and F. Relaix, "The role of pax genes in the development of tissues and organs: *Pax3* and *Pax7* regulate muscle progenitor cell functions," *Annual Review of Cell and Developmental Biology*, vol. 23, pp. 645–673, 2007.

[23] V. D. Soleimani, V. G. Punch, Y. Kawabe et al., "Transcriptional dominance of *Pax7* in adult *Myogenesis* is due to high-affinity recognition of homeodomain motifs," *Developmental Cell*, vol. 22, no. 6, pp. 1208–1220, 2012.

[24] K. Mousavi, H. Zare, A. H. Wang, and V. Sartorelli, "Polycomb protein Ezh1 promotes RNA polymerase II elongation," *Molecular Cell*, vol. 45, no. 2, pp. 255–262, 2012.

[25] J. A. Simon and R. E. Kingston, "Mechanisms of polycomb gene silencing: knowns and unknowns," *Nature Reviews Molecular Cell Biology*, vol. 10, pp. 697–708, 2009.

[26] P. Asp, R. Bluma, V. Vethanthama et al., "Genome-wide remodeling of the epigenetic landscape during myogenic differentiation," *Proceedings of the National Academy of Sciences of the USA*, vol. 108, no. 22, pp. E149–E158, 2011.

[27] V. Vethantham, Y. Yanga, C. Bowmana et al., "Dynamic loss of H_2B ubiquitylation without corresponding changes in H_3K_4 trimethylation during myogenic differentiation," *Molecular and Cellular Biology*, vol. 32, no. 6, pp. 1044–1055, 2012.

[28] C. Trapnell, B. A. Williams, G. Pertea et al., "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nature Biotechnology*, vol. 28, pp. 511–515, 2010.

[29] S. K. Chanda, S. White, A. P. Orth et al., "Genome-scale functional profiling of the mammalian AP-1 signaling pathway," *Proceedings of the National Academy of Sciences of USA*, vol. 100, pp. 12153–12158, 2003.

[30] M. Fiscella, J. W. Perry, B. Teng et al., "TIP, a T-cell factor identified using high-throughput screening increases survival

in a graft-versus-host disease model," *Nature Biotechnology*, vol. 21, pp. 302–307, 2003.

[31] R. Konig, Y. Zhou, D. Elleder et al., "Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication," *Cell*, vol. 135, no. 1, pp. 49–60, 2008.

[32] P. A. Gray, H. Fu, P. Luo et al., "Mouse brain organization revealed through direct genome-scale TF expression analysis," *Science*, vol. 306, no. 5705, pp. 2255–2257, 2004.

[33] E. S. Lein, M. J. Hawrylycz, N. Ao et al., "Genome-wide atlas of gene expression in the adult mouse brain," *Nature*, vol. 445, pp. 168–176, 2007.

[34] R. A. Baldock, J. B. L. Bard, A. Burger et al., "EMAP and EMAGE: a framework for understanding spatially organized data," *Neuroinformatics*, vol. 1, no. 4, pp. 309–325, 2003.

[35] L. Richardson, S. Venkataraman, P. Stevenson et al., "EMAGE mouse embryo spatial gene expression database: 2010 update," *Nucleic Acids Research*, vol. 38, supplement 1, pp. D703–D709, 2010.

[36] S. Venkataraman, P. Stevenson, Y. Yang et al., "EMAGE-edinburgh mouse atlas of gene expression: 2008 update," *Nucleic Acids Research*, vol. 36, supplement 1, pp. D860–D865, 2008.

[37] J. H. Christiansen, Y. Yang, S. Venkataraman et al., "EMAGE: a spatial database of gene expression patterns during mouse embryo development," *Nucleic Acids Research*, vol. 34, no. 1, pp. D637–D641, 2006.

[38] S. D. Harding, C. Armit, J. Armstrong et al., "The GUDMAP database—an online resource for genitourinary research," *Development*, vol. 138, pp. 2845–2853, 2011.

[39] A. P. McMahon, B. J. Aronow, D. R. Davidson et al., "GUDMAP: the genitourinary developmental molecular anatomy project," *Journal of the American Society of Nephrology*, vol. 19, no. 4, pp. 667–671, 2008.

[40] Y. Ito, N. Toriuchia, T. Yoshitakaa et al., "The *Mohawk* homeobox gene is a critical regulator of tendon differentiation," *Proceedings of the National Academy of Sciences of USA*, vol. 107, no. 23, pp. 10538–10542, 2010.

[41] K. Aoki, G. Meng, K. Suzuki et al., "RP58 associates with condensed chromatin and mediates a sequence-specific transcriptional repression," *The Journal of Biological Chemistry*, vol. 273, pp. 26698–26704, 1998.

[42] F. Fuks, W. A. Burgers, N. Godin, M. Kasai, and T. Kouzarides, "Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription," *EMBO Journal*, vol. 20, no. 10, pp. 2536–2544, 2001.

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 564381, 10 pages
doi:10.1155/2012/564381

Review Article

Epigenetic Regulation of B Lymphocyte Differentiation, Transdifferentiation, and Reprogramming

Bruna Barneda-Zahonero,[1] Lidia Roman-Gonzalez,[1] Olga Collazo,[1] Tokameh Mahmoudi,[2] and Maribel Parra[1]

[1] *Cellular Differentiation Group, Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), Avenida Gran Via s/n km 2.7, 08907 L'Hospitalet de Llobregat, Barcelona, Spain*
[2] *Department of Biochemistry, Erasmus University Medical Center, Ee622, P.O.Box 2040, 3000 CA Rotterdam, The Netherlands*

Correspondence should be addressed to Maribel Parra, mparra@idibell.cat

Received 7 February 2012; Revised 21 May 2012; Accepted 6 July 2012

Academic Editor: Sonia Vanina Forcales

B cell development is a multistep process that is tightly regulated at the transcriptional level. In recent years, investigators have shed light on the transcription factor networks involved in all the differentiation steps comprising B lymphopoiesis. The interplay between transcription factors and the epigenetic machinery involved in establishing the correct genomic landscape characteristic of each cellular state is beginning to be dissected. The participation of "epigenetic regulator-transcription factor" complexes is also crucial for directing cells during reprogramming into pluripotency or lineage conversion. In this context, greater knowledge of epigenetic regulation during B cell development, transdifferentiation, and reprogramming will enable us to understand better how epigenetics can control cell lineage commitment and identity. Herein, we review the current knowledge about the epigenetic events that contribute to B cell development and reprogramming.

1. Introduction

Hematopoietic stem cells (HSCs) give rise to mature B cells through the sequential differentiation of lymphoid progenitor cells. Long-term HSCs (LT-HSCs) have the ability to self-renew and reconstitute the entire immune system by differentiating into short-term HSCs (ST-HSCs). ST-HSCs differentiate into multipotent progenitors (MPPs) that then branch into common myeloid progenitors (CMPs) and lymphoid-primed multipotent progenitors (LMPPs). CMPs further differentiate into erythrocytes and megakaryocytes, whereas LMPPs retain the capability to give rise to myelomonocytic or lymphoid lineages [1, 2]. LMPPs become common lymphoid progenitors (CLPs) [3], which have the potential to differentiate into B and T lymphocytes as well as natural killer (NK) cells [4, 5]. Once committed to the lymphoid lineage, further differentiation steps lead to the formation of pro-B and pre-B cells, which are the early B cell precursors for immature B cells, the terminally differentiated plasma cells and germinal-center B cells (Figure 1).

Every step in B cell development is characterized by the activation of the specific genetic program characteristic of the new intermediate/progenitor generated and the repression/extinction of the genetic program of the previous cellular state. To achieve this, the different differentiation steps are tightly regulated at the transcriptional level. In recent years, the theory of the existence of networks of lineage-specific and identity-transcription factors responsible for establishing particular genomic landscapes has gained credence [6]. In the case of lymphocyte development, the transcription factors Ikaros and PU.1 are critical for the cellular commitment of LMPPs to the lymphoid lineage [2]. Subsequently, early B cell specification depends on the action of E2A, EBF, and FOXO1, whereas Pax5 is required for proper B cell development and for maintaining B cell identity [7–12]. Finally, during later developmental stages, the transcriptional repressors Bcl6 and Blimp-1 are crucial for the generation of germinal-center B cells and plasma cells, respectively [13–17] (Figure 1).

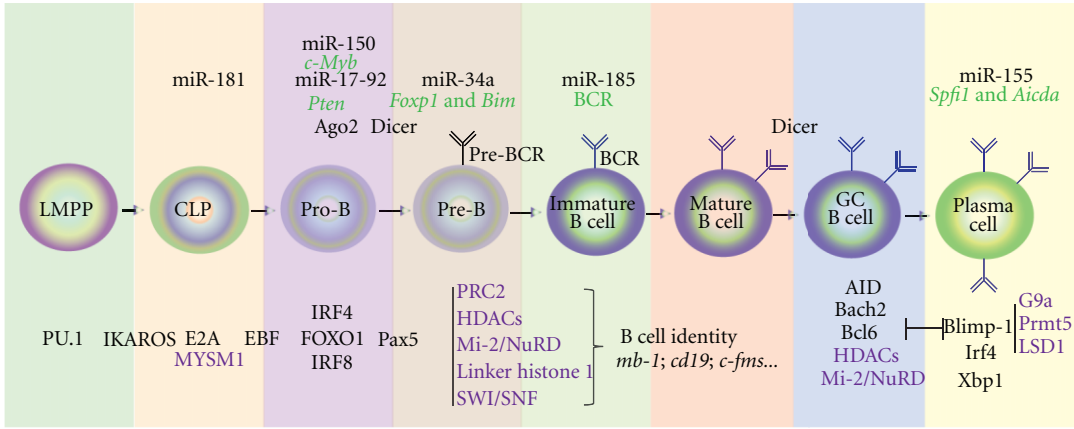


Figure 1: Scheme for B cell development. Successive stages of B cell differentiation and the key transcription factors and epigenetic regulators involved are shown. The epigenetic regulators that cooperate with specific transcription factors at every cell differentiation step are in purple. MicroRNA transcript targets are in green.

The picture of the hierarchical network of transcription factors that mediate the epigenetic signature needed to regulate the specific transcriptome of the fate of B-cells during their development has begun to emerge [18–20]. For example, Pax5, whose expression is induced by E2A and EBF, recruits chromatin-remodeling, histone-modifying and transcription-factor complexes to its target genes to activate the transcription of B cell-specific genes, and to silence lineage-inappropriate genes [19]. Extensive efforts have been made to elucidate the epigenetic mechanisms underlying the gene rearrangements of various components of the B cell receptor (BCR) [21–23]. Thus, epigenetic regulation is a critical event in B lymphocyte development. The relevance of transcription factors to the establishment and maintenance of cell-lineage identity has also been demonstrated in cellular reprogramming experiments [24–27]. The epigenetic mechanisms involved in the reprogramming and transdifferentiation of B cells have also been a focus of study in recent years.

Nucleosomes are the basic unit of the chromatin. They comprise 147 bp of DNA wrapped around a histone core, which contains two copies each of H2A, H2B, H3, and H4. This core is important for establishing interactions between nucleosomes and within the nucleosome itself [28]. Depending on the epigenetic modifications on the histone tails and in the DNA, chromatin can adopt different structural conformations that are correlated with its active, permissive (primed), or repressive status. The four main mechanisms by which epigenetic regulation occurs are DNA methylation, histone modification, chromatin remodeling, and regulation of gene expression by the action of noncoding RNAs. The methylation of cytosine residues at CpG dinucleotides (methyl-CpG), which is generally associated with transcriptional repression, is accomplished via the action of DNA methyltransferases (DNMTs) [29]. Methyl-CpG-mediated transcriptional repression can be explained by two nonmutually exclusive molecular mechanisms. First, methylation of DNA can interfere with the accessibility and recruitment of transcription factors to their DNA-binding sites. Second, DNA methylation results in the recruitment of methyl-CpG-binding proteins (MeCPs and MBDs) in association with corepressor complexes. Both mechanisms lead to the transcriptional silencing of the methylated genes [29]. The posttranslational modification of histones is another important epigenetic regulatory mechanism. Histones can be posttranslationally modified by a variety of enzymatic modifications, including acetylation, methylation, phosphorylation, sumoylation, and ubiquitination among others [30]. While acetylation is generally considered to be a mark of transcriptional activation, histone methylation can result in either transcriptional activation or repression, depending on the residue that is modified. In this regard, acetylation of histone H3 on lysine 9, 14, and or 18 (H3K9ac, H3K14ac, H3K18ac) is associated with transcriptional activation and considered "histone active marks." In the case of histone methylation, di- and tri-methylation of histone H3 at lysine 4 (H3K4me2, H3K4me3) are associated with transcriptional activation and therefore considered an active mark, whereas trimethylation of H3 at lysine 27 (H3K27me3) is found to be enriched at silenced genes and considered to be a repressive histone mark [28, 30]. Trimethylation of histone H3 at lysine 9 (H3K9me3) has also been characterized as a mark of transcriptional repression [30]. However, different reports suggest that it can also represent transcriptional activity [31, 32]. Another mechanism of epigenetic regulation involves the action of chromatin remodelers, which are multi-subunit complexes that use the energy from ATP hydrolysis to change the location or conformation of nucleosomes, resulting in increased or decreased DNA accessibility [28]. Chromatin-remodeling complexes can be divided into four groups, characterized by core ATPase subunits. Based on the defining ATPase, they are referred to as the SWI/SNF, ISWI, CHD, and INO80 families of remodelers [28]. Finally, microRNAs (miRNAs), a type of small noncoding RNAs, have been shown to anneal to 3′UTR of cognate mRNAs, leading to mRNA instability and/or the inhibition of translation, thereby making it possible to modulate the proteome of the cell [29].

In this papre we will summarize the recent advances in our understanding of the epigenetic mechanisms controlling B cell development and reprogramming.

2. B Cell Development: Early Specification towards the Lymphoid Lineage

When cells are at the LMPP stage, two transcription factors, Ikaros and PU.1, play critical roles in the early cellular specification towards the lymphoid lineage. Mice homozygous for a germline mutation in the Ikaros DNA-binding domain present a block at early lymphocyte development and therefore lack lymphocyte progenitors, T and B lymphocytes, as well as natural killer cells [33, 34]. More recently, Ikaros was shown to be a crucial transcription factor for the commitment of LMPPs into CLPs, clearly demonstrating its key role in the early cellular decision to undergo lymphocyte development [2]. LMPPs derived from Ikaros-null mice lack B cell potential and do not express *Flt3*, *Il-7r*, *Rag1* and *Rag2*, which are important genes for lymphoid commitment [2]. Mechanistically, Ikaros can either activate or repress transcription of target genes, depending on the recruitment of coactivators or corepressors. For example, in T cells, Ikaros has been shown to recruit corepressor or chromatin remodeling complexes in order to either repress or activate specific targets [35–37]. However, how Ikaros mediates the epigenetic regulation of its target genes during the differentiation of LMPPs into CLPs remains to be elucidated.

Likewise, the transcription factor PU.1 is crucial for the commitment of LMPPs to the lymphoid lineage. Strikingly, PU.1 is also required for the generation of GMPs and macrophages. In fact, mice deficient for PU.1 die around birth and lack B, T, NK and myelomonocytic cells [38, 39]. The promiscuity of PU.1 in regulating gene expression in different cell types raised the general question of what the mechanism of action is of a given transcription factor in different cell types. In this regard, Heinz et al. recently identified the genomewide binding sites of PU.1 in splenic B cells, macrophages and B cell progenitors [40]. They found that PU.1 cooperates with cell-type-specific transcription factors to activate the cisregulatory elements required for the development of a particular cell type. For example, in CLPs and pro-B cells, E2A induces PU.1 binding at B cell-specific genomic sites that contain closely located PU.1 and E2A binding motifs [40]. In addition, PU.1 binding initiates nucleosome remodeling, followed by H3K4me enrichment at many specific genomic regions [40]. These data could lead us to speculate that cooperation between PU.1 and Ikaros might be crucial for the activation of specific genes required to specify LMPP into CLPs. Also, the identity of Ikaros and PU.1 epigenetic partners remains unknown. This matter awaits investigation.

3. B Cell Development: Early B Cell Commitment

B cell development is characterized by the generation of the BCR, which consists of a heavy and a light immunoglobulin chain, IgH and IgL, respectively. The expression of the BCR subunits *VpreB*, *λ5*, and *mb-1* (*Cd79a*), and the initiation of D-J rearrangements at the IgH locus defines early B cell commitment [41]. The specification of CLPs in the B cell lineage requires two transcription factors, E2A and EBF1, which have been shown to activate the expression of genes essential for the formation of pro-B cells [42]. E2A and EBF knockout mouse models are phenotypically similar, and both transcription factors are considered to play key roles in initiating B lymphopoiesis. E2A-deficient mice show arrested B cell development at the pre-pro-B cell stage with compromised D-J rearrangements at the IgH locus and a lack of expression of *Rag1*, *mb-1*, *Iv*, *λ5*, *Cd19*, and *Pax5* genes [7–9]. More recently, it was shown that conditional deletion of E2A in pre-B cells did not result in a complete loss of expression of its target genes, indicating the involvement of E2A in the early steps of B cell commitment [43]. Similar to E2A, EBF is also known to play a crucial role in initiating B cell development. Mice lacking EBF do not express *Rag1*, *Rag2*, *mb-1*, *B29* (*Igβ*), *λ5*, *VpreB*, *cd19*, or *Pax5* genes [10]. Recent studies have also implicated the transcription factor FOXO-1 in early B lymphopoiesis. FOXO-1-deficient mice also show a developmental block at the pro-B cell stage [11]. Moreover, it has been reported that FOXO-1 regulates *Rag1* and *Rag2* expression [44].

Recent evidence indicates that the network of transcription factors Pax5, E2A and EBF also cooperate to regulate their target genes. For example, E2A, EBF, and Pax5 coordinate epigenetic events that lead to the expression of *mb-1*, which encodes the Ig*α* subunit of the pre-BCR and BCR [45]. *mb-1* is methylated at CpG dinucleotides in HSCs and is gradually demethylated during B cell commitment correlating with its pattern of expression [21]. EBF and E2A contribute to the CpG demethylation and nucleosomal remodeling of the *mb-1* promoter, an event necessary for its transcriptional activation by Pax5. ATP-dependent chromatin remodeling complexes have also been implicated in EBF and Pax5-mediated regulation of the *mb-1* gene [21]. Knockdown of Brg1 and Brm, the catalytic subunits of the SWI/SNF chromatin-remodeling complex interfere with EBF and Pax5-mediated activation of *mb-1*. In contrast, knockdown of Mi-2, the catalytic subunit of the Mi-2/NuRD chromatin-remodeling complex, enhances chromatin accessibility and demethylation of the *mb-1* promoter and its transcription in response to both transcription factors [21]. These results are consistent with a model in which the SWI/SNF and Mi-2/NuRD chromatin remodeling complexes play antagonistic regulatory roles to enable or limit the reprogramming of target genes by EBF and Pax5 during B cell development [21]. The B-cell-specific gene *Cd19* is another example of a gene that is epigenetically regulated during early B cell development. *Cd19* encodes a cell surface protein that participates in signal transduction mechanisms via the BCR and pre-BCR. Chromatin remodeling at the upstream enhancer sequences of *Cd19* occurs in multipotent progenitors [22]. This chromatin remodeling has been shown to facilitate the recruitment of E2A to this locus followed by EBF and Pax5 recruitment [22]. Interestingly, the *Cd19* promoter is transcriptionally activated only after Pax5

binding. In this context, Mercer et al. recently reported that the monomethylation of H3K4 (H3K4me) at the enhancer regions of cell lineage-specifying genes is the main epigenetic mark, which is associated with their specific expression pattern throughout the lymphoid differentiation program [46]. Taken together, these reports provide clear examples of how B cell lineage-specific transcription factors cooperatively mediate the epigenetic regulation of target genes during B lymphopoiesis.

The recent advances in ultrasequencing technologies are helping to draw a global picture of how the networks of transcription factors modify the chromatin of their target genes. The laboratory of Cornelis Murre, using a ChIP-seq experimental approach, has elucidated how the network of transcription factors E2A, EBF and FOXO-1 orchestrates B cell commitment [18]. They found that during the transition of pre-pro-B cell to pro-B cells, E2A-associated genes become monomethylated at lysine 4 on H3 (H3K4me), a mark mainly found on gene enhancer elements. Subsequently, EBF and FOXO1 are involved in the enrichment of active histone modifications such as H3K4me3 on B-cell-specifying genes, such as *Pax5* [18]. Recently, Treiber and colleagues have shed light on the EBF-mediated epigenetic regulation of its target genes [47]. They classified EBF targets as activated, repressed, or primed genes. They observed that, in pro-B and pre-B cells, the "activated" genes are enriched in H3K4me3 and H3 acetylation active marks and show low levels of the repressive mark H3K27me3 [47]. In contrast, the "repressed" genes show the opposite pattern of histone modifications. The "primed" genes are enriched in the gene enhancer mark H3K4me in pre-B and pro-B cells and enriched in H3K4me3 and H3 acetylation in mature B cells [47]. The identification of the epigenetic regulators recruited by transcription factors to mediate gene expression changes during B lymphopoiesis remains to be addressed.

Other epigenetic marks, such as ubiquitination of Histone H2A, have proved to play a role in early B cell development. Jiang et al. pointed out that the histone H2A deubiquitinase MYSM1 is an important factor in B cell development [48]. *Mysm1* knockout mice show a drastic decrease in the number of B cells in the bone marrow, peripheral blood, and lymph nodes [48]. The authors concluded that MYSM1 antagonizes the action of the polycomb repressive complex 1 (PRC1) on the *Ebf1* promoter, enabling lineage-specific transcription factors, such as E2A, to be recruited to the *Ebf1* locus and to induce its transcription [48].

Early B cell development is also known to be regulated by microRNAs. Mice deficient in Ago2, which encodes a protein essential for microRNA biogenesis and function, display a block in B cell development at the pro-B cell stage [49]. Consistent with this, specific deletion of Dicer in pro-B cells, which abolishes the entire miRNA network in B cells, results in a complete block of B cell differentiation at the transition from pro-B to pre-B-cells [50]. Another study reported that miR-181, one of the approximately 100 microRNAs known to be expressed in mouse bone marrow cells, is more abundant in the B cell lineage than in other cell types [51]. Transplantation of multipotent hematopoietic progenitors overexpressing miR-181 into lethally irradiated mice resulted in an increase in the number of B cells [51]. Thus, miR-181 appears to target and repress the transcripts of critical genes involved in generating B cells. A similar experimental approach was used to show that another microRNA, miR-150, which is expressed in mature B and T cells, can block B cell differentiation at the pro-B cell stage when expressed prematurely [52]. Accordingly, the laboratory of Klaus Rajewsky reported that miR-150 plays a role during B cell differentiation through its action on *c-Myb* expression [53]. Other miRNAs have been associated with the early development of B cells. For instance, miR-34a ablation results in a developmental block at the pre-B cell stage, and miR-17-92 knockout mice exhibit a block in pro-B cells [54, 55]. They regulate the *Foxbp1* and *Bim* and *PTEN* genes, respectively, which are known to have a role in B cell differentiation [54, 55]. Recently, Kuchen et al. have elucidated the microRNAome during lymphopoiesis at the genome-wide scale, leading to the identification of miRNAs that are primed for expression at different stages of differentiation [56]. They reported that miRNA expression is tightly regulated by epigenetic modifications. In particular, they showed that the repressive mark H3K27me3 is associated with the gene silencing of lineage-inappropriate miRNA during lymphopoiesis [56]. However, they also observed that active epigenetic regulation by the presence of H3K4me also occurs in some of the microRNAs "primed" to be expressed. On the basis of the restrictive expression and abundance of miRNAs during B cell lineage specification, miR-320, miR-191, miR-139 and miR28 appear to be potential regulators of B cell differentiation [56]. The transcripts targeted by key miRNAs for the early differentiation of B cells remains to be identified.

4. B Cell Development: Pax5 in the Maintenance of B Cell Identity

The transcription factor Pax5 is essential for maintaining the fate of B cells and is therefore considered to be "the guardian of B cell identity" [57]. Its expression gradually increases in a stepwise manner during B cell development. Pax5 expression is first detected at the early pro-B cell stage and maintained up to the mature B cell stage. Pax5 knockout mice show a block in B cell development at the pro-B stage [12]. Pax5−/− pro-B cells express both E2A and EBF transcription factors, as well as their target genes. In contrast, E2A−/− and EBF−/− derived cells do not express Pax5. Collectively, these data indicate that Pax5 is a target for both transcription, factors. The laboratory of Meinrad Busslinger has shed light on the molecular mechanisms involved in the gradual expression of Pax5 during B cell development. In particular, they have identified an enhancer in the *Pax5* locus, which in combination with the promoter, recapitulates B lymphoid Pax5 expression [58]. Interestingly, the *Pax5* enhancer is silenced by DNA methylation in embryonic stem cells, while it becomes activated in multipotent hematopoietic progenitors. The presence of consensus binding sites for the transcription factors PU.1, IRF4, IRF8, and NF-*k* B within the *Pax5* enhancer suggests that these transcription factors play a role in sequential enhancer activation in hematopoietic

progenitors and during B cell development [58]. At the onset of pro-B cell development the transcription factor EBF1 induces chromatin remodeling at the *Pax5* promoter region. In non-B cells, Polycomb group proteins repress the *Pax5* promoter region [58].

In addition to the epigenetic regulation of its expression during B cell development, Pax5 induces the establishment of a B cell-specific transcription program that is associated with the suppression of inappropriate genes of alternative lineages, thereby ensuring its role in maintaining B cell identity and differentiation. Using gene expression microarrays and genome-wide ChIP-on-chip experimental approaches, the laboratories of Busslinger and Nutt have described the complex gene regulatory network regulated by Pax5 during B lymphopoiesis [59–61]. These studies have identified genes that are activated or repressed by Pax5 in wildtype pro-B cells. Pax5-activated genes appear to encode transcription factors and key proteins involved in B cell signaling, adhesion, migration, antigen presentation and germinal-center B cell formation [59, 61]. However, Pax5-repressed genes encode secrete proteins, cell adhesion molecules, signal transducers and nuclear proteins that are specific to erythroid, myeloid, and T cell lineages [59, 61]. Pax5-activated genes in pro-B cells were found to be enriched with epigenetically active marks, including H3K9ac, H3K4me2 and H3K4me3 [60]. Importantly, in Pax5-deficient pro-B cells, these active histone marks were dramatically reduced or lost, indicating that Pax5 is essential for guaranteeing the active chromatin structure at its target genes. These findings demonstrate that Pax5 is a master regulator of B cell identity, which, in conjunction with epigenetic regulators, coordinates a B-cell-specific target gene transcription program. Recently, McManus and colleagues have described the epigenetic mechanisms mediated by Pax5 during B lymphopoiesis [19]. By using a ChIP-on-chip analysis, they have identified Pax5 target genes in committed pro-B cells. The authors also apply a proteomic approach to identify Pax5 interacting partners. They found that Pax5 interacts with the members of the SWI/SNF chromatin remodeling complex Brg1. BAF57 and BAF170. They also reported that PAX5 recruits the NCoR1 repressor complex with its associated HDAC3 activity to repressed its target genes [19]. This study has provided novel important insight into the regulatory network and epigenetic regulation, by which Pax5 directly controls B-cell commitment at the onset of B lymphopoiesis.

The mechanism by which Pax5 mediates transcriptional repression of targets has also been informatively examined using a candidate gene approach. One of the important target genes repressed by Pax5 in B cells is the colony-stimulating factor receptor 1 gene (*csf1r* or *c-fms*), a gene essential for macrophage development. *Csf1r* is expressed at low levels in HSCs and downregulated in all nonmacrophage cell types. In HSCs, MPPs, CMPs, and CLPs the *Csf1r* promoter is bound by transcription factors and its chromatin structure in an active conformation [62]. However, the *Csf1r* gene is silenced during B cell differentiation. Interestingly, an intronic antisense transcription unit that is differentially regulated during lymphopoiesis overlaps with regions of de novo DNA methylation in B cells, highlighting DNA methylation as a mechanism for *Csf1r* silencing during B cell development. Despite being silenced, *Csf1r* chromatin remains in a poised or primed conformation even in mature B cell stages. Importantly, *Csf1r* expression can be reactivated by conditional deletion of the transcription factor Pax5 [62]. Pax5 was shown to bind the *Csf1r* gene directly, resulting in loss of RNA polymerase II recruitment and binding of myeloid transcription factors at cisregulatory elements [63]. Finally, Pax5 in conjunction with linker histone H1 also coordinates DNA methylation and histone modifications in the 3′ regulatory region of the immunoglobulin heavy chain locus and thus epigenetically regulates the IgH locus [64].

5. B Cell Development: Terminal Differentiation

The completion of V(D)J recombination and expression of the BCR on the surface of B cells marks the beginning of antigen-dependent B cell development. From this point, B cells undergo terminal differentiation dependent on signals emanating from the BCR after antigen triggering [65]. Peripheral B cells, without antigen-mediated signaling, are in a resting state [66]. Once activated, they either initiate the germinal center (GC) reaction or differentiate into antibody-secreting plasma cells. Entry into the GC reaction is regulated by Bcl6, whereas the generation of antibody-secreting plasma cells is controlled by Blimp-1. Bcl6 and Blimp-1 both act as transcriptional repressors and work in a mutually exclusive manner [67, 68].

After antigen triggering, Bcl6 is upregulated in some B cells that then enter the GC reaction [13–15]. In contrast, cells in which Bcl6 is not upregulated undergo differentiation into plasma cells [69, 70]. From a mechanistic angle, Bcl6 has been shown to interact with the chromatin remodeling complex Mi-2/NuRD in GC B cells, leading to the repression of specific genes that are characteristic of plasma cells [71, 72]. This Mi-2/NURD-mediated repression requires the recruitment of histone deacetylases HDAC1 and HDAC2 [71, 72].

After activation of GC B cells Bcl-6 expression is downregulated in association with the expression of its target gene Blimp-1. Once expressed, Blimp-1 represses the gene expression program of mature B cells, thereby promoting plasma cell differentiation [16, 17]. Mechanistically, Blimp-1 exerts its repressive transcriptional activity by recruiting regulators and coordinating epigenetic modifications at its target genes. PRD1-BF1, the human orthologue of Blimp-1, silences the interferon beta gene in response to viral infection by recruiting the histone methyltransferase (HMTase) G9a to the interferon-beta promoter, resulting in H3K9me [73]. Blimp-1 has also been found in a complex with the arginine histone methyl transferase Prmt5, although the functional significance of this interaction in B cells is not clear [74]. The histone lysine demethylase LSD1 has also been shown to interact with Blimp-1 [75]. Chromatin immunoprecipitation (ChIP) experiments indicated that Blimp-1 and LSD1 share some target genes leading to a more accessible chromatin structure [75]. Importantly, disruption of the Blimp-1-LSD1 interaction resulted in attenuated antibody secretion of the

cells, highlighting the functional relevance of this interaction for B cell function.

In the last few years, additional transcription factors have emerged as being involved in B cell terminal differentiation. It has been reported that IRF4 and XBp1 control the maintenance of plasma cell identity. IRF4 is responsible for BLIMP-1 induction and, in conjunction with XBp1, determines the fate of the plasma cell. The network of transcription factors Pax5, Bach2, and Bcl6 direct B cell development into germinal center cells. It has been shown that Pax5 induces Bach2 expression after B cell activation, which in turn cooperates with Blc6 to repress Blimp-1 expression promoting activation-induced cytidine deaminase (AID) expression and antibody class switch [76, 77].

MicroRNAs are also involved in the terminal differentiation of B cells. Peripheral B cells in transit to their final maturation can give rise to two functionally distinct peripheral populations: follicular (FO) or marginal zone (MZ) B cells. FO versus MZ fate decision is functionally coupled to BCR signaling and it has been suggested that B cells bearing BCRs with autoreactive specificities are preferentially driven into a MZ fate [78]. In 2010, Belver and colleagues generated conditional Dicer-deficient mice at later stages of B cell development [79]. They observed that miRNA metabolism is important for such developmental stage since these mice presented an impairment in the generation of follicular B cells and an overrepresentation of marginal zone B cells. Accordingly, another phenotypic feature of these mice was the presence of high titers of autoreactive antibodies [79]. They identified miR185 as an important factor for the correct BCR-mediated development of B cells.

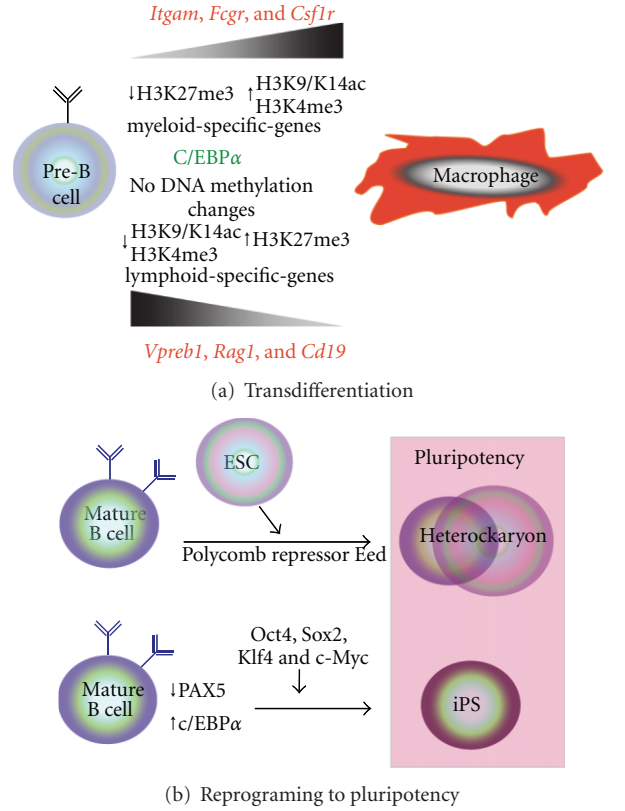


Figure 2: Transdifferentiation and reprogramming of B cells. (a) Ectopic expression of C/EBP in pre-B cells induces their transdifferentiation into macrophages. Epigenetic changes during the process are shown. (b) B cells can be reprogrammed to pluripotency by fusion with ESCs (heterokaryon) or by transgenic induction of Oct4, Sox2, Klf4 and c-Myc (iPS).

6. B Cell Reprogramming and Transdifferentiation

Since 1987, when the possibility of reprogramming specialized cells by the expression of a linage-specific transcription factor was first reported, many studies have tried to understand the molecular mechanisms that control all the processes involved. Due to the high developmental complexity that characterizes the hematopoietic system, it constitutes a model system with which study cell reprogramming and transdifferentiation in greater depth. In 1995, it was reported that overexpression of the erythroid lineage-specific transcription factor GATA-1 in myeloid leukemia cells induced their reprogramming into the megakaryocytic/erythroid lineage [24]. Subsequently, Nutt et al. reported that Pax5-defective pro-B cells differentiated into functional macrophages, granulocytes, natural killer cells, osteoclasts, and dendritic cells when specific cytokines were added to the culture medium [26]. Some years later, using knock-in and lineage-tracing technologies in mice, Xie and colleagues were able to demonstrate *in vivo* reprogramming of intrasplenic mature B cells into macrophages by the overexpression of the myeloid transcription factor C/EBPα [80]. More recently, the same laboratory generated a robust reprogramming system in which murine pre-B cells were converted into functional macrophages by the overexpression of C/EBPα [25] (Figure 2). This cellular conversion has been considered a transdifferentiation event since it is irreversible and does not require the retrodifferentiation of pre-B cells to previous progenitor stages [81]. Using the cellular system generated in Graf's laboratory, Radríguez-Ubreva and colleagues performed a high-throughput methylation analysis to study changes in DNA methylation during the transdifferentiation of pre-B cells into macrophages [82]. Surprisingly, they did not find any significant changes in DNA methylation during cellular conversion. However, they were able to identify the expected histone modifications in the genes that had previously been described to be upregulated or downregulated during the process. In particular, they reported an increase in the enrichment of the active histone marks H3K9/K14ac and H3K4me3, at the promoters of upregulated macrophage-specific genes, whereas a reduction of these modifications was observed in the B-cell-specific downregulated genes. In contrast, the repressive mark H3K27me3 was found to be enriched in the B cell downregulated genes and reduced in the upregulated macrophage-specific genes [82] (Figure 2). This study suggests that histone regulators are able to overcome the repressive effect of DNA methylation in

macrophage-specific genes in the converted cells. It also establishes an important difference from the process of reprogramming towards pluripotency in which promoter DNA demethylation plays a crucial role.

In this regard, Hanna and colleagues demonstrated that pro-B and pre-B cells can be reprogrammed into induced pluripotent stem (iPS) cells by the expression of the transcription factors Oct4, Sox2, Klf4, and c-Myc [27]. Interestingly, the expression of the four factors in mature B cells does not result in the reprogramming of mature B cells to pluripotency. They found that expression of c/EBPα in conjunction with the four "reprogramming" factors is necessary to generate iPS cells [27] (Figure 2). iPS cell lines derived from immature and mature B cells show promoter demethylation of the stem cell markers Oct4 and Nanog, whereas both promoters are heavily methylated in the original B cells. Finally, in mature B cells the promoter region of *Pax5* shows high and low levels of enrichment for the active mark H3K4me3 and the repressive mark H3K27me3, respectively. Conversely, equivalent enrichment of both histone modifications was observed in iPS lines derived from mature B cells [27]. This study raises the challenging question of how B lymphocytes at different developmental stages differ in their epigenetic landscape and how one factor can overcome this divergence to allow reprogramming into pluripotent cells.

A number of studies using experimental heterokaryons, in which a somatic cell is reprogrammed towards pluripotency by fusion with mouse embryonic stem (ES) cells, have also been used to reprogram B lymphocytes into pluripotent cells. The laboratory of Amanda Fisher has shown that when mouse ES cells are fused with human B lymphocytes the expression of human pluripotent-associated genes is rapidly induced [83]. Recently, the same group has elucidated some of the epigenetic mechanisms underlying this reprogramming process. They showed that deletion of Eed, Suz12, Ezh2, and Ring1A/B, which are members of either the polycomb repressor complex PRC1 or PRC2, in mouse ES cells abolishes their capacity to induce human B lymphocyte reprogramming towards pluripotency [84] (Figure 2).

7. Concluding Remarks

The impressive advances in genome-wide methods and the latest generation of ultrasequencing techniques are opening up new, and challenging lines of research focused on the elucidation of the epigenetic mechanisms underlying B cell differentiation and reprogramming. Many questions remain to be answered. Is there a specific "epigenetic signature" for the different cellular states comprising B cell development? How can lymphoid-specific transcription factors orchestrate the epigenetic machinery at different genes and genome regions to facilitate the choice to differentiate into a particular cellular lineage? Is the expression of epigenetic regulators lineage-specific? Epigenetic modification analyses, genome-wide RNA and ChIP-Seq studies, quantitative proteomics, and systematic functional studies offer us the opportunity to obtain high-quality measurements that will provide us with a draft of the "epigenetic-transcriptional" program that controls B cell development and reprogramming. Finally, conditional gene inactivation in mice will reveal the role of specific epigenetic regulators during B cell development. Thus, new regulatory networks connecting epigenetic and transcription factors seem likely to be revealed in the context of B lymphopoiesis in the future.

Acknowledgments

This work was supported by a grant from the Spanish Ministry of Science and Innovation (MICIIN) (BFU2008-00410) (to M.Parra) and a Marie Curie IRG Grant from the European Commission (FP7-MIRG-CT-2007-208119) (to M.Parra). M. Parra was supported by a Ramon y Cajal contract from the Spanish Ministry of Science and Innovation (MICIIN). L.Roman-Gonzalez was supported by an IDIBELL Ph.D. fellowship. T.Mahmoudi was supported by an Erasmus M.C. Fellowship.

References

[1] J. Adolfsson, R. Månsson, N. Buza-Vidas et al., "Identification of Flt3$^+$ lympho-myeloid stem cells lacking erythro-megakaryocytic potential: a revised road map for adult blood lineage commitment," *Cell*, vol. 121, no. 2, pp. 295–306, 2005.

[2] T. Yoshida, S. Yao-Ming Ng, J. C. Zuniga-Pflucker, and K. Georgopoulos, "Early hematopoietic lineage restrictions directed by Ikaros," *Nature Immunology*, vol. 7, no. 4, pp. 382–391, 2006.

[3] H. Igarashi, S. C. Gregory, T. Yokota, N. Sakaguchi, and P. W. Kincade, "Transcription from the RAG1 locus marks the earliest lymphocyte progenitors in bone marrow," *Immunity*, vol. 17, no. 2, pp. 117–130, 2002.

[4] M. Kondo, I. L. Weissman, and K. Akashi, "Identification of clonogenic common lymphoid progenitors in mouse bone marrow," *Cell*, vol. 91, no. 5, pp. 661–672, 1997.

[5] C. Cobaleda and M. Busslinger, "Developmental plasticity of lymphocytes," *Current Opinion in Immunology*, vol. 20, no. 2, pp. 139–148, 2008.

[6] G. Natoli, "Maintaining cell identity through global control of genomic organization," *Immunity*, vol. 33, no. 1, pp. 12–24, 2010.

[7] G. Bain, E. C. R. Maandag, D. J. Izon et al., "E2A proteins are required for proper b cell development and initiation of immunoglobulin gene rearrangements," *Cell*, vol. 79, no. 5, pp. 885–892, 1994.

[8] G. Bain, E. C. Robanus Maandag, H. P. J. Te Riele et al., "Both E12 and E47 allow commitment to the B cell lineage," *Immunity*, vol. 6, no. 2, pp. 145–154, 1997.

[9] Y. Zhuang, P. Soriano, and H. Weintraub, "The helix-loop-helix gene E2A is required for B cell formation," *Cell*, vol. 79, no. 5, pp. 875–884, 1994.

[10] H. Lin and R. Grosschedl, "Failure of B-cell differentiation in mice lacking the transcription factor EBF," *Nature*, vol. 376, no. 6537, pp. 263–267, 1995.

[11] H. S. Dengler, G. V. Baracho, S. A. Omori et al., "Distinct functionss for the transcription factor Foxo1 at various stages of B cell differentiation," *Nature Immunology*, vol. 9, no. 12, pp. 1388–1398, 2008.

[12] P. Urbanek, Z. Q. Wang, I. Fetka, E. F. Wagner, and M. Busslinger, "Complete block of early B cell differentiation and

altered patterning of the posterior midbrain in mice lacking Pax5/BSAP," *Cell*, vol. 79, no. 5, pp. 901–912, 1994.

[13] D. Allman, A. Jain, A. Dent et al., "BCL-6 expression during B-cell activation," *Blood*, vol. 87, no. 12, pp. 5257–5268, 1996.

[14] A. L. Dent, A. L. Shaffer, X. Yu, D. Allman, and L. M. Staudt, "Control of inflammation, cytokine expression, and germinal center formation by BCL-6," *Science*, vol. 276, no. 5312, pp. 589–592, 1997.

[15] B. H. Ye, G. Cattoretti, Q. Shen et al., "The BCL-6 proto-oncogene controls germinal-centre formation and Th2- type inflammation," *Nature Genetics*, vol. 16, no. 2, pp. 161–170, 1997.

[16] A. L. Shaffer, K. I. Lin, T. C. Kuo et al., "Blimp-1 orchestrates plasma cell differentiation by extinguishing the mature B cell gene expression program," *Immunity*, vol. 17, no. 1, pp. 51–62, 2002.

[17] R. Sciammas and M. M. Davis, "Modular nature of blimp-1 in the regulation of gene expression during B cell maturation," *Journal of Immunology*, vol. 172, no. 9, pp. 5427–5440, 2004.

[18] Y. C. Lin, S. Jhunjhunwala, C. Benner et al., "A global network of transcription factors, involving E2A, EBF1 and Foxo1, that orchestrates B cell fate," *Nature Immunology*, vol. 11, no. 7, pp. 635–643, 2010.

[19] S. McManus, A. Ebert, G. Salvagiotto et al., "The transcription factor Pax5 regulates its target genes by recruiting chromatin-modifying proteins in committed B cells," *EMBO Journal*, vol. 30, no. 12, pp. 2388–2404, 2011.

[20] N. Novershtern, A. Subramanian, L. N. Lawton et al., "Densely interconnected transcriptional circuits control cell states in human hematopoiesis," *Cell*, vol. 144, no. 2, pp. 296–309, 2011.

[21] H. Gao, K. Lukin, J. Ramírez, S. Fields, D. Lopez, and J. Hagman, "Opposing effects of SWI/SNF and Mi-2/NuRD chromatin remodeling complexes on epigenetic reprogramming by EBF and Pax5," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 27, pp. 11258–11263, 2009.

[22] K. Walter, C. Bonifer, and H. Tagoh, "Stem cell-specific epigenetic priming and B cell-specific transcriptional activation at the mouse Cd19 locus," *Blood*, vol. 112, no. 5, pp. 1673–1682, 2008.

[23] I. H. Su and A. Tarakhovsky, "Epigenetic control of B cell differentiation," *Seminars in Immunology*, vol. 17, no. 2, pp. 167–172, 2005.

[24] H. Kulessa, J. Frampton, and T. Graf, "GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts," *Genes and Development*, vol. 9, no. 10, pp. 1250–1262, 1995.

[25] L. H. Bussmann, A. Schubert, T. P. Vu Manh et al., "A robust and highly efficient immune cell reprogramming system," *Cell Stem Cell*, vol. 5, no. 5, pp. 554–566, 2009.

[26] S. L. Nutt, B. Heavey, A. G. Rolink, and M. Busslinger, "Commitment to the B-lymphoid lineage depends on the transcription factor Pax5," *Nature*, vol. 401, no. 6753, pp. 556–562, 1999.

[27] J. Hanna, S. Markoulaki, P. Schorderet et al., "Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency," *Cell*, vol. 133, no. 2, pp. 250–264, 2008.

[28] C. R. Clapier and B. R. Cairns, "The biology of chromatin remodeling complexes," *Annual Review of Biochemistry*, vol. 78, pp. 273–304, 2009.

[29] S. Guil and M. Esteller, "DNA methylomes, histone codes and miRNAs: tying it all together," *International Journal of Biochemistry and Cell Biology*, vol. 41, no. 1, pp. 87–95, 2009.

[30] T. Kouzarides, "Chromatin modifications and their function," *Cell*, vol. 128, no. 4, pp. 693–705, 2007.

[31] J. K. Wiencke, S. Zheng, Z. Morrison, and R. F. Yeh, "Differentially expressed genes are marked by histone 3 lysine 9 trimethylation in human cancer cells," *Oncogene*, vol. 27, no. 17, pp. 2412–2421, 2008.

[32] C. R. Vakoc, S. A. Mandat, B. A. Olenchock, and G. A. Blobel, "Histone H3 lysine 9 methylation and HP1γ are associated with transcription elongation through mammalian chromatin," *Molecular Cell*, vol. 19, no. 3, pp. 381–391, 2005.

[33] K. Georgopoulos, M. Bigby, J. H. Wang et al., "The Ikaros gene is required for the development of all lymphoid lineages," *Cell*, vol. 79, no. 1, pp. 143–156, 1994.

[34] J. H. Wang, A. Nichogiannopoulou, L. Wu et al., "Selective defects in the development of the fetal and adult lymphoid system in mice with an Ikaros null mutation," *Immunity*, vol. 5, no. 6, pp. 537–549, 1996.

[35] S. Y. M. Ng, T. Yoshida, and K. Georgopoulos, "Ikaros and chromatin regulation in early hematopoiesis," *Current Opinion in Immunology*, vol. 19, no. 2, pp. 116–122, 2007.

[36] J. Kim, S. Sif, B. Jones et al., "Ikaros DNA-binding proteins direct formation of chromatin remodeling complexes in lymphocytes," *Immunity*, vol. 10, no. 3, pp. 345–355, 1999.

[37] J. Zhang, A. F. Jackson, T. Naito et al., "Harnessing of the nucleosome-remodeling-deacetylase complex controls lymphocyte development and prevents leukemogenesis," *Nature Immunology*, vol. 13, no. 1, pp. 86–94, 2012.

[38] S. R. McKercher, B. E. Torbett, K. L. Anderson et al., "Targeted disruption of the PU.1 gene results in multiple hematopoietic abnormalities," *EMBO Journal*, vol. 15, no. 20, pp. 5647–5658, 1996.

[39] E. W. Scott, M. C. Simon, J. Anastasi, and H. Singh, "Requirement of transcription factor PU.1 in the development of multiple hematopoietic lineages," *Science*, vol. 265, no. 5178, pp. 1573–1577, 1994.

[40] S. Heinz, C. Benner, N. Spann et al., "Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities," *Molecular Cell*, vol. 38, no. 4, pp. 576–589, 2010.

[41] M. Ye and T. Graf, "Early decisions in lymphoid development," *Current Opinion in Immunology*, vol. 19, no. 2, pp. 123–128, 2007.

[42] M. Busslinger, "Transcriptional control of early B cell development," *Annual Review of Immunology*, vol. 22, pp. 55–79, 2004.

[43] A. S. Lazorchak, J. Wojciechowski, M. Dai, and Y. Zhuang, "E2A promotes the survival of precursor and mature B lymphocytes," *Journal of Immunology*, vol. 177, no. 4, pp. 2495–2504, 2006.

[44] R. H. Amin and M. S. Schlissel, "Foxo1 directly regulates the transcription of recombination-activating genes during B cell development," *Nature Immunology*, vol. 9, no. 6, pp. 613–622, 2008.

[45] H. Maier, R. Ostraat, H. Gao et al., "Early B cell factor cooperates with Runx1 and mediates epigenetic changes associated with mb-1 transcription," *Nature Immunology*, vol. 5, no. 10, pp. 1069–1077, 2004.

[46] E. Mercer, Y. Lin, C. Benner et al., "Multilineage priming of enhancer repertoires precedes commitment to the B and myeloid cell lineages in hematopoietic progenitors," *Immunity*, vol. 35, no. 3, pp. 413–425, 2011.

[47] T. Treiber, E. M. Mandel, S. Pott et al., "Early B cell factor 1 regulates B cell gene networks by activation, repression, and transcription- independent poising of chromatin," *Immunity*, vol. 32, no. 5, pp. 714–725, 2010.

[48] X.-X. Jiang, Q. Nguyen, Y. Chou et al., "Control of B cell development by the histone H2A deubiquitinase MYSM1," *Immunity*, vol. 35, no. 6, pp. 883–896, 2011.

[49] D. O'Carroll, I. Mecklenbrauker, P. P. Das et al., "A Slicer-independent role for Argonaute 2 in hematopoiesis and the microRNA pathway," *Genes and Development*, vol. 21, no. 16, pp. 1999–2004, 2007.

[50] S. B. Koralov, S. A. Muljo, G. R. Galler et al., "Dicer ablation affects antibody diversity and cell survival in the B lymphocyte lineage," *Cell*, vol. 132, no. 5, pp. 860–874, 2008.

[51] C. Z. Chen, L. Li, H. F. Lodish, and D. P. Bartel, "MicroRNAs modulate hematopoietic lineage differentiation," *Science*, vol. 303, no. 5654, pp. 83–86, 2004.

[52] B. Zhou, S. Wang, C. Mayr, D. P. Bartel, and H. F. Lodish, "miR-150, a microRNA expressed in mature B and T cells, blocks early B cell development when expressed prematurely," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 17, pp. 7080–7085, 2007.

[53] C. Xiao, D. P. Calado, G. Galler et al., "MiR-150 controls B cell differentiation by targeting the transcription factor c-Myb," *Cell*, vol. 131, no. 1, pp. 146–159, 2007.

[54] A. Ventura, A. G. Young, M. M. Winslow et al., "Targeted deletion reveals essential and overlapping functions of the miR-17~92 family of miRNA clusters," *Cell*, vol. 132, no. 5, pp. 875–886, 2008.

[55] R. M. O'Connell, D. S. Rao, A. A. Chaudhuri, and D. Baltimore, "Physiological and pathological roles for microRNAs in the immune system," *Nature Reviews Immunology*, vol. 10, no. 2, pp. 111–122, 2010.

[56] S. Kuchen, W. Resch, A. Yamane et al., "Regulation of MicroRNA expression and abundance during lymphopoiesis," *Immunity*, vol. 32, no. 6, pp. 828–839, 2010.

[57] C. Cobaleda, A. Schebesta, A. Delogu, and M. Busslinger, "Pax5: the guardian of B cell identity and function," *Nature Immunology*, vol. 8, no. 5, pp. 463–470, 2007.

[58] T. Decker, M. Pasca di Magliano, S. McManus et al., "Stepwise activation of enhancer and promoter regions of the B cell commitment gene Pax5 in early lymphopoiesis," *Immunity*, vol. 30, no. 4, pp. 508–520, 2009.

[59] A. Delogu, A. Schebesta, Q. Sun, K. Aschenbrenner, T. Perlot, and M. Busslinger, "Gene repression by Pax5 in B cells is essential for blood cell homeostasis and is reversed in plasma cells," *Immunity*, vol. 24, no. 3, pp. 269–281, 2006.

[60] A. Schebesta, S. McManus, G. Salvagiotto, A. Delogu, G. A. Busslinger, and M. Busslinger, "Transcription factor Pax5 activates the chromatin of key genes involved in B cell signaling, adhesion, migration, and immune function," *Immunity*, vol. 27, no. 1, pp. 49–63, 2007.

[61] C. Pridans, M. L. Holmes, M. Polli et al., "Identification of Pax5 target genes in early B cell differentiation," *Journal of Immunology*, vol. 180, no. 3, pp. 1719–1728, 2008.

[62] H. Tagoh, A. Schebesta, P. Lefevre et al., "Epigenetic silencing of the c-fms locus during B-lymphopoiesis occurs in discrete steps and is reversible," *EMBO Journal*, vol. 23, no. 21, pp. 4275–4285, 2004.

[63] H. Tagoh, R. Ingram, N. Wilson et al., "The mechanism of repression of the myeloid-specific c-fms gene by Pax5 during B lineage restriction," *EMBO Journal*, vol. 25, no. 5, pp. 1070–1080, 2006.

[64] V. Giambra, S. Volpi, A. V. Emelyanov et al., "Pax5 and linker histone H1 coordinate DNA methylation and histone modifications in the 3′ regulatory region of the immunoglobulin heavy chain locus," *Molecular and Cellular Biology*, vol. 28, no. 19, pp. 6123–6133, 2008.

[65] K. Rajewsky, "Clonal selection and learning in the antibody system," *Nature*, vol. 381, no. 6585, pp. 751–758, 1996.

[66] M. Kraus, M. B. Alimzhanov, N. Rajewsky, and K. Rajewsky, "Survival of resting mature B lymphocytes depends on BCR signaling via the Igα/β heterodimer," *Cell*, vol. 117, no. 6, pp. 787–800, 2004.

[67] K. L. Calame, K. I. Lin, and C. Tunyaplin, "Regulatory mechanisms that determine the development and function of plasma cells," *Annual Review of Immunology*, vol. 21, pp. 205–230, 2003.

[68] T. Honjo, M. Muramatsu, and S. Fagarasan, "Aid: how does it aid antibody diversity?" *Immunity*, vol. 20, no. 6, pp. 659–668, 2004.

[69] R. Reljic, S. D. Wagner, L. J. Peakman, and D. T. Fearon, "Suppression of signal transducer and activator of transcription 3-dependent B lymphocyte terminal differentiation by BCL-6," *Journal of Experimental Medicine*, vol. 192, no. 12, pp. 1841–1847, 2000.

[70] A. L. Shaffer, X. Yu, Y. He, J. Boldrick, E. P. Chan, and L. M. Staudt, "BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control," *Immunity*, vol. 13, no. 2, pp. 199–212, 2000.

[71] N. Fujita, D. L. Jaye, M. Kajita, C. Geigerman, C. S. Moreno, and P. A. Wade, "MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer," *Cell*, vol. 113, no. 2, pp. 207–219, 2003.

[72] N. Fujita, D. L. Jaye, C. Geigerman et al., "MTA3 and the Mi-2/NuRD complex regulate cell fate during B lymphocyte differentiation," *Cell*, vol. 119, no. 1, pp. 75–86, 2004.

[73] I. Gyory, J. Wu, G. Fejér, E. Seto, and K. L. Wright, "PRDI-BF1 recruits the histone H3 methyltransferase G9a in transcriptional silencing," *Nature Immunology*, vol. 5, no. 3, pp. 299–308, 2004.

[74] K. Ancelin, U. C. Lange, P. Hajkova et al., "Blimp1 associates with Prmt5 and directs histone arginine methylation in mouse germ cells," *Nature Cell Biology*, vol. 8, no. 6, pp. 623–630, 2006.

[75] S. T. Su, H. Y. Ying, Y. K. Chiu, F. R. Lin, M. Y. Chen, and K. I. Lin, "Involvement of histone demethylase LSD1 in blimp-1-mediated gene repression during plasma cell differentiation," *Molecular and Cellular Biology*, vol. 29, no. 6, pp. 1421–1431, 2009.

[76] A. Muto, K. Ochiai, Y. Kimura et al., "Bach2 represses plasma cell gene regulatory network in B cells to promote antibody class switch," *EMBO Journal*, vol. 29, no. 23, pp. 4048–4061, 2010.

[77] S. L. Nutt, N. Taubenheim, J. Hasbold, L. M. Corcoran, and P. D. Hodgkin, "The genetic network controlling plasma cell differentiation," *Seminars in Immunology*, vol. 23, no. 5, pp. 341–349, 2011.

[78] D. Allman and S. Pillai, "Peripheral B cell subsets," *Current Opinion in Immunology*, vol. 20, no. 2, pp. 149–157, 2008.

[79] L. Belver, V. G. de Yébenes, and A. R. Ramiro, "MicroRNAs prevent the generation of autoreactive antibodies," *Immunity*, vol. 33, no. 5, pp. 713–722, 2010.

[80] H. Xie, M. Ye, R. Feng, and T. Graf, "Stepwise reprogramming of B cells into macrophages," *Cell*, vol. 117, no. 5, pp. 663–676, 2004.

[81] A. Di Tullio, T. P. Vu Manh, A. Schubert, R. Månsson, and T. Graf, "CCAAT/enhancer binding protein α (C/EBPα)-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 41, pp. 17016–17021, 2011.

[82] J. Rodríguez-Ubreva, L. Ciudad, D. Gómez-Cabrero et al., "Pre-B cell to macrophage transdifferentiation without significant promoter DNA methylation changes," *Nucleic Acids Research*, vol. 40, no. 5, pp. 1954–1968, 2012.

[83] C. F. Pereira, R. Terranova, N. K. Ryan et al., "Heterokaryon-based reprogramming of human B lymphocytes for pluripotency requires Oct4 but not Sox2," *PLoS Genetics*, vol. 4, no. 9, Article ID e1000170, 2008.

[84] C. F. Pereira, F. M. Piccolo, T. Tsubouchi et al., "ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency," *Cell Stem Cell*, vol. 6, no. 6, pp. 547–556, 2010.

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 284786, 13 pages
doi:10.1155/2012/284786

Research Article

Predictive Models of Gene Regulation from High-Throughput Epigenomics Data

Sonja Althammer,[1] Amadís Pagès,[1] and Eduardo Eyras[1,2]

[1] *Computational Genomics, Universitat Pompeu Fabra, Dr. Aiguader 88, Barcelona, Spain*
[2] *Institució Catalana de Recerca i Estudis Avançats (ICREA), Passeig Lluís Companys 23, Barcelona, Spain*

Correspondence should be addressed to Eduardo Eyras, eduardo.eyras@upf.edu

Received 3 February 2012; Accepted 2 May 2012

Academic Editor: Sonia Vanina Forcales

The epigenetic regulation of gene expression involves multiple factors. The synergistic or antagonistic action of these factors has suggested the existence of an epigenetic code for gene regulation. Highthroughput sequencing (HTS) provides an opportunity to explore this code and to build quantitative models of gene regulation based on epigenetic differences between specific cellular conditions. We describe a new computational framework that facilitates the systematic integration of HTS epigenetic data. Our method relates epigenetic signals to expression by comparing two conditions. We show its effectiveness by building a model that predicts with high accuracy significant expression differences between two cell lines, using epigenetic data from the ENCODE project. Our analyses provide evidence for a degenerate epigenetic code, which involves multiple genic regions. In particular, signal changes at the 1st exon, 1st intron, and downstream of the polyadenylation site are found to associate strongly with expression regulation. Our analyses also show a different epigenetic code for intron-less and intron-containing genes. Our work provides a general methodology to do integrative analysis of epigenetic differences between cellular conditions that can be applied to other studies, like cell differentiation or carcinogenesis.

1. Introduction

DNA associates with histone proteins to conform the chromatin [1]. Histones generally carry posttranscriptional modifications in cells capable of modulating the expression of genes [2, 3]. For instance, there is a genome-wide relation between the histone 3 lysine 36 trimethylation (H3K36me3) and transcription activity [4, 5]. This and other epigenetic modifications are key to cellular differentiation [6] and their alterations have been associated to early stages of cellular transformation in tumors [7, 8]. The combinations of the histone modifications, which can have cooperative or opposed effects on the chromatin state, have been proposed to reflect a histone code that would determine the regulation of gene expression and the cell state [9]. High-throughput sequencing (HTS) technologies provide a very effective way to obtain information about the histone modification patterns at genome wide scale [10]. Efforts to integrate available genome-wide datasets about chromatin in various conditions are crucial towards improving our understanding of the role of epigenetics in gene regulation.

Recent publications have made progress in the definition of a histone code of gene expression by generating predictive models of transcriptional activity based on histone mark information [11–17]. They provide insights into possible mechanisms of regulation and a formal description of the postulated histone code [18, 19]. These methods generally relate the histone signals obtained from experiments of chromatin immunoprecipitation followed by HTS (ChIP-Seq) [20], with a read-out of the gene expression based on expression microarrays or HTS for mRNAs (RNA-Seq) [21]. In these approaches, the chromatin signal is generally represented in terms of read-counts or peak significance in the promoter and sometimes the gene body of genes. However, this analysis is generally based on one single condition or cell line. That is, they effectively compare the properties of different genes in a direct way, relying on the premise that signals in two different genes should be

comparable, and the accuracy of their predictive model will be dependent on the accuracy of the estimation of the significance of the ChIP-Seq signals. However, genes present many variable properties, like number of introns or the presence of CpG islands in their promoter, which may affect these measurements. For instance, recent experiments show that the splicing machinery can recruit histone-modifying enzymes and influence the chromatin state, with the consequence that genes with introns tend to have higher levels of H3K36me3 signal [22]. Thus, the comparison of genes with and without introns is not straightforward. Additionally, various other factors may affect the local density of HTS signal [23]. For instance, the tag counts from an HTS experiment will be influenced by the chromatin structure of the DNA and by shearing effects [24–26], not all regions have the same mappability [27] and there is often a GC bias in the reads [28]. These issues will reflect on differences in coverage between regions, which will be even more exacerbated for the broad signals that are obtained for histone ChIP-Seq experiments. Control samples can partly alleviate this, but their effectiveness depends very much on the sequencing depth. Thus, HTS signals from two genes are not directly comparable in general.

Here, we propose a new method to measure epigenetic signals and to relate them to expression based on the comparison between two conditions. In our approach, the same genomic locus is compared between two conditions; hence, the predictive model describes changes of gene expression in terms of relative changes in epigenetic mark densities between two conditions or cell types. Significance of these changes is calculated taking the read density into account, thereby mitigating the confounding effects mentioned earlier. Additionally, unlike a previous method that has made pairwise comparisons of epigenetic data from cell lines [17], our method considers continuous changes of the epigenetic signal densities, rather than an on-off state description. Moreover, our framework provides greater flexibility than previous approaches for the generation of computational predictive models.

To illustrate our method, we have built a model of expression regulation from epigenetic changes using data from various ENCODE cell lines [29]. In order to extend this relation, we include additional epigenetic data not considered previously, namely, HTS of DNase I hypersensitive sites (DNase-Seq) [30] and DNA methylation data [31]. Our results show a different epigenetic code for expression for intron-less and intron-containing genes, being this difference more prominent in genes with low GC content around the transcription start site. Moreover, eliminating anti-sense transcription and overlapping promoters and tails from different genes, which has not been done before, the prediction accuracy improves considerably. Furthermore, the predictive model built from one pair of cell lines performs with high accuracy in a different pair. Finally, we are able to generate a minimal code for expression regulation between two cell lines that is generic enough to correctly predict the regulatory outcome of up to 70% transcripts from a different pair of cell lines.

2. Material and Methods

2.1. Genomic Annotations. For our analyses, we used the gene set from the 7th release of the GENCODE annotation (ftp://ftp.sanger.ac.uk/pub/gencode/release_7/gencode.v7.annotation.gtf.gz), which is based on the assembly GRCh37 (hg19) and is included in the ensembl release 62 [32]. All transcripts encoded at each gene loci and the genomic region defined by them, which we name transcript loci, were considered initially. Those transcript loci from chromosome M and of biotype "pseudogene" were removed for the analysis.

We separated transcript loci into four groups; according to whether they were intron-containing (IC) or intron-less (IL), and according to whether they had a promoter with high CpG (HCG) or low CpG (LCG) content. We classified transcripts as HCG if the region of 4 kb centred on the transcription start site (TSS) overlaps at least 200 bp with a CpG island, and LCG otherwise. CpG island annotations where obtained from the UCSC Table Browser (hg19) [33]. In order to obtain balanced sets for training and testing, an equal number of up- (Up) and down- (Dw) regulated transcripts were selected from each of the four groups. These groups were taken to be as large as possible, but such that the P-value of significance (Benjamini-Hochberg corrected) for the expression change for each transcript was smaller than 0.05. Furthermore, the same number of nonregulated (Nr) transcripts was selected. These are defined to have the highest P-values and sufficient expression, that is, the density of reads measured in RPKM (reads per kilobase per million mapped reads) as defined in [21] was greater than 1 in a cell line from the pair. With this, we obtained four different sets (Table 1). As part of our analyses, we also filtered overlapping transcript loci that would make ambiguous the assignment of the marks with the correct expression change. That is, we removed the loci from both strands when they were in any of the following configurations (Supplementary Figure 1, available at doi:10.1155/2012/284786):

(i) transcript loci that overlap in opposite strands,

(ii) transcript loci whose promoters (2 kb) overlap in opposite strands,

(iii) transcript loci whose tails (2 kb) overlap in opposite strands

(iv) transcript loci with overlapping promoter (2 kb) and tail (2 kb) on the same strand,

(v) Overlapping transcript loci on the same strand but from different genes.

2.2. Datasets. We downloaded ChIP-Seq data for RNA Polymerase II (RNAPII), CCCTC-binding factor (CTCF) and various Histone marks (Table 2), data for DNase I hypersensitive sites (DNase-Seq), methylation data from reduced representation bisulfite sequencing (Methyl-RRBS) and RNA-Seq data from the ENCODE project (http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/) for four cell lines: a chronic myelogenous leukemia line (K562), a lymphoblastoid line (GM12878), a human mammary

Table 1: Each of the four sets of transcript loci considered in our analysis. The numbers correspond to the loci before (all) or after (filtered) eliminating overlapping loci (Section 2). From each set, we considered up-, down-, or nonregulated transcript loci, each corresponding to 1/3 of the indicated numbers.

Transcript-loci set	Description	Pair 1-all	Pair 1-filtered	Pair 2-all	Pair 2-filtered
HCG IC	High CG promoter and intron-containing	6510	1959	2964	792
HCG IL	High CG promoter and intron-less	105	27	24	12
LCG IC	Low CG promoter and intron-containing	6705	1767	1980	585
LCG IL	Low CG promoter and intron-less	84	30	15	15

Table 2: ENCODE data sets and cell lines used for analysis: ChIP-Seq data for RNA Polymerase II (RNAPII), CTCF and various Histone marks, data for DNase I hypersensitive sites (DNase-Seq), methylation data from reduced representation bisulfite sequencing (methyl-RRBS) and sequencing of long polyA+ whole cell RNA (RNA-Seq). For HMEC and HSMM cells RNAPII ChIP-Seq data was not available at the time of our analyses. Datasets were generated at the Broad Institute (BROAD), Cold Spring Harbor Laboratory (CSHL), University of Washington (UW), University of Texas at Austin (UT-A), and Hudson Alpha (HA).

Factor/mark	Pair 1 Cell lines		Pair 2 Cell lines	
	K562	GM12878	HSMM	HMEC
CTCF	BROAD	BROAD	BROAD	BROAD
H3K27ac	BROAD	BROAD	BROAD	BROAD
H3K27me3	BROAD	BROAD	BROAD	BROAD
H3K36me3	BROAD	BROAD	BROAD	BROAD
H3K4me1	BROAD	BROAD	BROAD	BROAD
H3K4me2	BROAD	BROAD	BROAD	BROAD
H3K4me3	BROAD	BROAD	BROAD	BROAD
H3K9ac	BROAD	BROAD	BROAD	BROAD
H4K20me1	BROAD	BROAD	BROAD	BROAD
RNAPII	UT-A	UT-A	—	—
DNase-Seq	UW	UW	UW	UW
Methyl-RRBS	HA	HA	HA	HA
RNA-Seq	CSHL	CSHL	CSHL	CSHL

epithelial line (HMEC), and a muscle myoblast line (HSMM, Table 2). We considered two pairs of comparisons, P1: K562 versus GM12878 and P2: HSMM versus HMEC. To further validate these results, we also considered a third comparison, K562 versus HSMM, P3. We selected experiments that were available in these four cell lines, except for RNAPII, which was only available in two of the selected cell lines. For all datasets, we used only reads that did not contain any uncalled bases (N). Moreover, for ChIP-Seq and DNase-Seq reads, we kept only reads with mapping quality greater than 30. The Methyl-RRBS data was filtered for positions covered by at least 10 reads. The mean methylation of a region was defined to be the proportion of methylated sites over the total number of probed sites in that region. Further, we obtained the RPKMs for the RNA-Seq data for the individual transcript loci directly from ENCODE public datasets (http://hgdownload-test.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeCshlLongRnaSeq/releaseLatest/).

For our analysis we considered for each transcript locus, a number of regions related to its exon-intron structure (Table 3). Subsequently, for each one of these regions and for each experimental dataset, the z-score for the enrichment was calculated between a pair of cell lines using Pyicos [34]. The calculation was based on 2 replicas in one condition (K562 or HSMM) and 1 replica in the other condition (GM12878 or HMEC). Further, pseudocounts and RPKM normalization were used (details in Supplementary Material). These z-scores constitute the set of attributes that were used for Machine-Learning (ML) analyses and corresponds to each region-experiment pair. As a control, random attributes were generated for each region by random sampling z-score values from all attributes for that region type.

Unless otherwise stated, accuracies of the models were measured calculating the average area under the receiver operating characteristic (ROC) curve (AUC) for a 10-fold crossvalidation. A ROC curve relates the rates of true positives (TPs) and false positives (FPs) produced by the model. The larger the area described by the ROC curve (AUC) the better the overall accuracy of the model. AUC = 1 indicates a model that predicts no false positives and all true cases correctly, and AUC = 0.5 indicates that the model is equivalent to random. In 10-fold crossvalidation, the data is split into 10 subsets and 10 evaluations are carried out iteratively, where in each iteration 9 subsets (nine-tenths of the instances) are used for training and one subset for

Table 3: Regions considered per transcript locus for the calculation of the different attributes. We defined the 13 regions based on the gene annotations from Gencode version 7 (Ensembl 62).

Type	Region	Description
Fixed-length regions	Promoter 2 kb	Region starting 2 kb upstream of the transcription start site (TSS) and ending 1 bp before the TSS;
	Promoter 5 kb	Region starting 5 kb upstream of the TSS and ending 1 bp before the TSS;
	TSS ± 2 kb	Region starting 2 kb upstream of the TSS and ending 2 kb downstream
	TSS ± 5 kb	Region starting 5 kb upstream of the TSS and ending 5 kb downstream
	pA ± 2 kb	Region starting 2 kb upstream of the pA and ending 2 kb downstream
	Tail	Region starting 1 bp after the pA and ending 2 kb downstream
Variable-length regions	First exon	Region corresponding to the first exon of the transcript locus
	First intron	Region corresponding to the first intron of the transcript locus
	GB	Gene body, that is, region between the TSS and the poly-adenylation site (pA) of an annotated transcript locus
	GB3′ss	Region between the first 3′ splice-site and the pA of an annotated transcript locus
	GB ± 1 kb	Gene body with additional 1 kb stretches up- and downstream
	GB ± 5 kb	Gene body with additional 5 kb stretches up- and downstream
	GB + 5 kb	Gene body with an additional 5kb stretch downstream of the pA

testing. This method ensures that all instances are used for the evaluation and the overall accuracy is averaged over the ten iterations, so that it represents the mean behaviour of the model.

2.3. Read Profiles around Gene Bodies. We calculated the average number of reads from different marks around gene bodies, by plotting the average number of reads in windows (−2000, +400) and (−400, +2000) around the TSS and pA, respectively. Reads from histone marks, RNAPII, and CTCF were extended to 300 bp in the 5′ to 3′ direction, whereas methyl-RRBS data was extended in either direction to 75 bp. Genes considered for the profiles were at least 400 bp long. We further filtered out pseudogenes and those loci that overlap each other (see above) and split the remaining ones into expressed (RPKM > 0) and nonexpressed (RPKM = 0) genes, resulting into 1202 IC and 1748 IL expressed genes, and 1385 IC and 746 IL nonexpressed genes. Supplementary Figures 2(A) and 2(B) show the profiles for IC and IL genes, whereas pseudogenes are shown in Supplementary Figure 2(C). Pseudogenes were also filtered for overlapping loci and for gene lengths shorter than 400 bp before they were split into 2277 IC and 3564 IL pseudogenes.

3. Results and Discussion

3.1. A Framework for Integrative Epigenetic Studies. Our computational framework addresses three fundamental tasks in the process of acquiring knowledge: data-mining, data manipulation, and data analysis, and it is comprised of the following steps: (i) an analysis pipeline to systematically identify the changes in expression and epigenetic signals between two conditions in multiple genomic regions, (ii) an automatic way to store the results in a Biomart system [34] for easy querying and filtering and (iii) a connectivity to the application WEKA [35], to allow the application of Machine-Learning (ML) methods for creating predictive models of gene regulation.

In order to relate epigenetic signals to expression regulation, our method measures signal changes between two conditions rather than the signal level in one single condition. With this methodology, relative changes of the epigenetic state can be related to each other or to the relative change of expression. By considering relative signal changes, biases from HTS are mitigated. To verify this, we checked whether selecting significant regions according to RPKM densities or z-scores from our method would be biased by the GC content. We, therefore, considered the top 10% of genes in terms of the H3K4me3 RPKM (K562) in the gene body and found a Spearman correlation of 0.34 with GC content. However, selecting the top 10% of genes according to absolute z-scores for H3K4me3, given by the comparison between K562 and GM12878, resulted in no correlation with GC content (Spearman 0.02). Thus, relating RPKM values to gene expression could result into false positives due to GC bias. When we repeated the same calculation on the 4 kb region centered on the TSS, none of the two measures, RPKMs or z-scores, showed a GC bias (correlation coefficient of −0.02 and 0.05, resp.). As H3K4me3 is mostly distributed around the TSS [10], we deduce that in this case the real signal obscures the bias, while in the gene body, where no strong signal for H3K4me3 is present, the bias dominates over the signal.

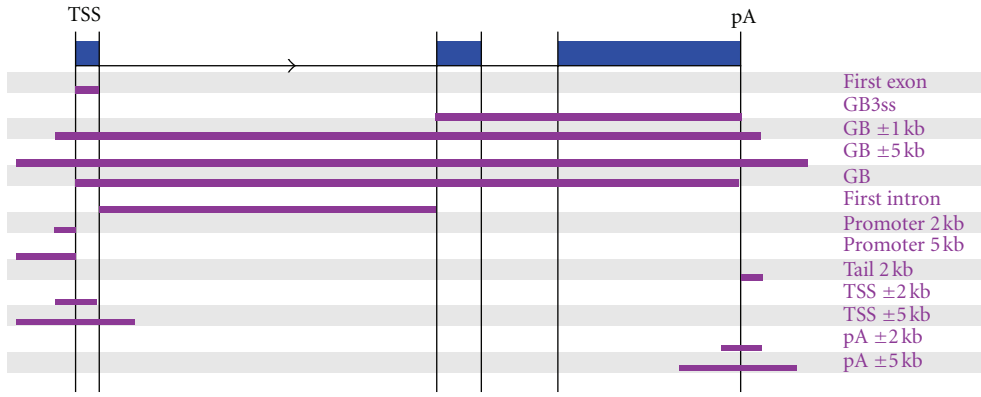


Figure 1: Graphical representation of the regions considered per transcript locus for the calculation of the different attributes. For detailed description of the regions see Table 3.

We have developed an automatic pipeline that, given a set of regions and a number of high-throughput sequencing (HTS) datasets for two conditions, can systematically calculate the log-rate of change for each region and its significance in terms of a z-score (details in Supplementary File). The datasets used are accessible through a Biomart database at http://regulatorygenomics.upf.edu/group/pages/software/. We have modified Biomart so that datasets can also be exported as ARFF (attribute-relation file format), which can be uploaded directly into the WEKA system [35], a collection of open-source machine-learning algorithms for data-mining tasks, issued under the GNU General Public License. Our system thus provides the possibility of using own custom data to train models and evaluate different ML algorithms for the study of mechanisms of gene regulation.

In order to illustrate the potential of our framework we analysed high-throughput sequencing (HTS) data from ENCODE [29] (Section 2). We started by systematically calculating the changes between cell lines in pair P1 (K562 versus GM12878) and in pair P2 (HSMM versus HMEC) for all the experiments in a variety of regions related to the transcript loci (Table 3). Most of the recently developed predictive methods use signals in the promoter region of genes or in a window around the transcription start site (TSS). We also included the gene body, as recent evidence suggests that the signal along this region will be informative as well [36]. Besides promoter, TSS, and gene body regions, we also include a region for the 1st exon, the 1st intron, and the gene body downstream of the 1st intron, which have been shown to contain relevant chromatin signatures for transcriptional regulation [22, 37, 38], and have not been used before in a predictive model. We further considered additional windows around and beyond the polyadenylation site (pA), resulting in a total of 13 different regions (Table 3, Figure 1). Accordingly, for the two pairs of cell lines P1 and P2, we had a total of $13 \times 12 = 156$ and $13 \times 11 = 143$ (as RNAPII was not available for P2) attributes per transcript locus, respectively, where each attribute is defined by the z-score of the enrichment value between the two cell lines for a region-experiment pair.

As classification value, we used expression information from RNA-Seq experiments from ENCODE in the corresponding cell lines. For each pair of cell lines, we calculated the transcripts with significant increase (Up) or decrease (Dw) of expression. In order to build a predictive model of expression that can distinguish between either type of regulation (Up or Dw) and no change, we also considered nonregulated (Nr) transcripts, defined to have sufficient expression level and no significant change in expression between the same pair of cell lines (Section 2).

Recent studies have shown that introns may influence the transcriptional regulation of genes [22, 38]. Therefore, we separated our transcripts sets according to whether they were intron-containing (IC) or intron-less (IL). Furthermore, several studies have highlighted that human promoters present different regulation according to their CG content [39–41]. Thus, we further split the sets according to whether a 4 kb region centered on the TSS overlaps with a CpG island or not, resulting in high CpG content (HCG) or low CpG content (LCG) sets (Section 2). Finally, in order to have a balanced set for training and testing, we selected from each type the same number of transcripts for each regulatory class (Table 1).

3.2. A Generic Epigenetic Code for Gene Expression Regulation. Using the datasets processed as above, we built a highly accurate and generic predictive model of gene expression changes based on epigenetic data. We tried various ML models to predict the three possible classes, up (Up), down (Dw), and nonregulated (Nr), and we decided to use a random forest model [42], as it showed the best performance using 10-fold crossvalidation (data not shown). Table 4 shows the accuracies of this model tested on intron-containing sets for various training conditions. Remarkably, we obtain a higher accuracy for the LCG set than for the HCG set (Table 4). Incidentally, CpG-related genes are quite often housekeeping genes [43], and this has been pointed out before as one of the reasons why predictive models perform differently on each set [44]. According to this, LCG transcripts should be more frequently associated to genes with differential expression (Up or Dw). This is confirmed

Table 4: We show the accuracy in terms of the area under the ROC curve (AUC) for the 10-fold cross validation for the IC transcript sets for various training conditions. The results are shown for all the transcript loci before (a) and after (b) filtering for the overlaps in opposite strands and overlaps of promoters and tails (Section 2). P1 (with RNAPII) corresponds to pair P1 with the additional RNAPII attribute, that is, the same attributes as P2 plus RNAPII. P1 and P2 denote the models for each cell line pairs with all the attributes. P1 (CFS) and P2 (CFS) denote the models for P1 and P2, respectively, where the attributes used are those that have a score 80 or higher (maximum 100) using the CFS attribute selection method independently for P1 and P2. P2 (CFS-P1) indicates that the model was trained using the data from P2 but the attributes selected using CFS on P1. P1-on-P2 indicates that the model was trained with pair P1 with all attributes and tested on pair P2. P1 (CFS)-on-P2 indicates that the model was trained with pair P1 with only selected attributes and tested on pair P2.

(a) Before filtering

Attributes	HCG-IC				LCG-IC			
	Up	Dw	Nr	Average	Up	Dw	Nr	Average
P1 (with RNAPII)	0.8	0.79	0.74	0.78	0.82	0.87	0.78	0.83
P1	0.79	0.79	0.74	0.77	0.83	0.86	0.76	0.82
P1 (CFS)	0.8	0.79	0.74	0.78	0.82	0.86	0.76	0.81
P2	0.85	0.83	0.81	0.83	0.9	0.88	0.83	0.87
P2 (CFS-P1)	0.85	0.83	0.8	0.83	0.9	0.88	0.83	0.87
P1-on-P2	0.83	0.77	0.63	0.74	0.88	0.83	0.71	0.81
P1(CFS)-on-P2	0.83	0.8	0.57	0.73	0.88	0.84	0.74	0.82

(b) After filtering

Attributes	HCG-IC				LCG-IC			
	Up	Dw	Nr	Average	Up	Dw	Nr	Average
P1 (with RNAPII)	0.79	0.84	0.76	0.8	0.85	0.9	0.81	0.86
P1	0.79	0.82	0.75	0.79	0.86	0.89	0.76	0.84
P1 (CFS)	0.79	0.81	0.73	0.78	0.84	0.9	0.77	0.84
P2	0.89	0.88	0.85	0.87	0.92	0.91	0.85	0.89
P2 (CFS-P1)	0.87	0.87	0.84	0.86	0.92	0.92	0.86	0.9
P1-on-P2	0.89	0.87	0.7	0.82	0.92	0.89	0.79	0.87
P1(CFS)-on-P2	0.85	0.82	0.68	0.78	0.91	0.89	0.81	0.87

in our analysis, as we found that the performance was always higher for the prediction of Up and Dw loci than for nonregulated transcripts (Table 4). For intron-less (IL) loci, we found the opposite behaviour, that is, HCG-IL has higher accuracy than LCG-IL (Supplementary Table 1).

Interestingly, training a model for the first pair with (Table 4(a), P1 (with RNAPII)) or without RNAPII data (Table 4(a), P1) yields very similar accuracy for all sets, which suggests that the information provided by RNAPII is redundant with the histone data for prediction. Indeed, looking at the pairwise correlations of all marks for P1, separated per region and per transcript set (Figure 2 and Supplementary Figure 3), we observe a high correlation of the z-scores for RNAPII with most of the other signals (H3K36me3, DNase-Seq, CTCF, H3K4me2, H3K9ac, H3K27ac, and H3K4me3).

With the aim of obtaining a minimal set of attributes that are sufficient to attain high prediction accuracy, we applied correlation-based feature selection (CFS) [45]. This method works by iteratively testing subsets of attributes, retaining those that best correlate with the class values (Up, Dw, or Nr) and removing those that have high redundancy. In this way, a minimal set of nonredundant attributes with optimal performance is selected. We applied CFS to the data from both pairs of cell lines and selected attributes that were selected in at least 80% of the validation rounds (Table 4(a), P1(CFS), and P2(CFS)). Interestingly, CFS provided attributes related to all the regions (Supplementary Table 2(A)), indicating that histone marks along all regions of the transcript locus may be relevant for regulation. Additionally, the prediction accuracy did not suffer, while the model is simplified by removing redundant attributes (Table 4(a), P1(CFS)).

With the aim of obtaining a generic epigenetic code of expression regulation, we decided to compare the attributes obtained from P1 with the attributes obtained for a second pair of cell lines (P2). Although CFS applied to both pairs, P1 and P2, yields a different set of optimal attributes, with only between 26% and 50% of coincidences between them (Supplementary Table 2), a model built on P2 with the attributes selected from P1 shows a high accuracy, which is comparable to the original model on P1 (Table 4(a) P2(CFS-P1)). That is, qualitatively, the attributes relevant for one pair of cell lines seem to be also relevant for the other one.

To test the generality of the model also in quantitative terms, that is, in terms of the actual numerical model, we applied directly on P2 the model built from P1. However, this test across pairs did not achieve an accuracy as high as before (Table 4(a), P1-on-P2 and P1(CFS)-on-P2). We hypothesized that the reduction of accuracy in the test across pairs could be due to differences in the homogeneity of cell lines, which would produce a very variable pattern of

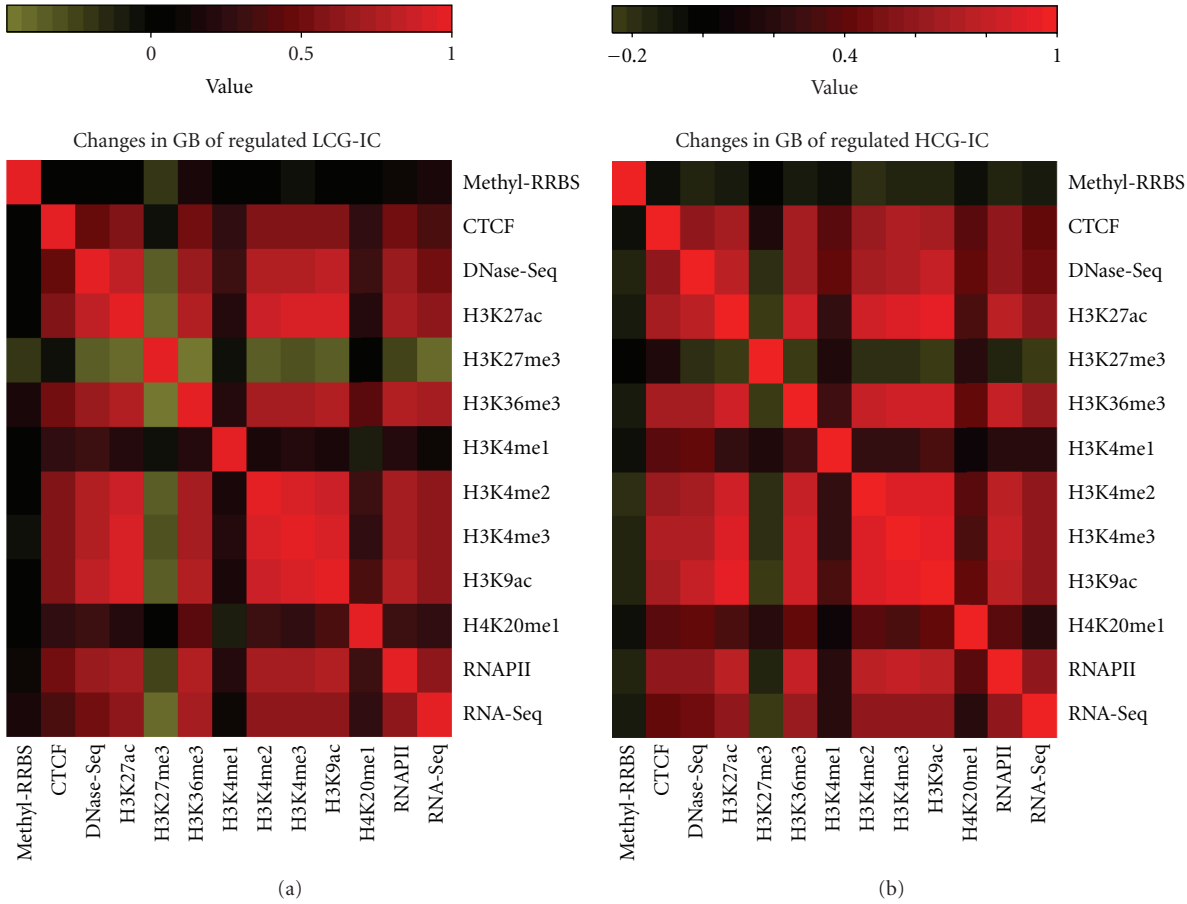


Figure 2: Pairwise correlations of marks and expression changes in gene bodies. Heatmaps are shown for regulated genes from the filtered intron-containing (IC) sets for low (LCG) (a) and high (HCG) (b) CpG promoters. The color represents the value of the Pearson correlation coefficient between the zscores for every pair of attributes. Both panels use the same scale, indicated above. For expression (RNA-Seq), the z-scores of the Up and Dw transcript loci were used to calculate the correlation.

signals. Alternatively, this lack of reproducibility could stem from the overlap of the gene body, promoters or tails from transcript loci from different genes, especially in the opposite strand, which would make ambiguous the association of the epigenetic signal change to a specific expression change. Accordingly, we removed from the training set those transcripts loci where the signal in one region could not be unambiguously assigned (Section 2, Supplementary Figure 1), thereby generating filtered sets for training and testing (Table 1). Interestingly, after removing these cases we observe a consistent increase in the accuracy of the prediction in all groups (Table 4(b)), with 60–78% of the instances correctly classified (Table 5).

To further confirm our results, we considered a third pair comparison: K562 versus HSMM or Pair 3 (P3). Supplementary Table 3 shows that accuracies for P3 are similar to those in P1 and P2, with higher accuracy for LCG loci, as found for the other pairs. As before, the AUC increases when loci are filtered (Section 2). Moreover, as shown before for P1 and P2, after filtering, the model trained on P1 gives similar prediction accuracy when applied to P3.

Despite the consistency of the models, there is still a fraction of instances that are incorrectly classified, that is, false positives. To understand why these instances cannot be correctly classified, we examined the the z-score distribution corresponding to the best separating attributes for up, down and nonregulated genes in LCG-IC and HCG-IC. Supplementary Figure 4 shows that the distributions of z-scores for the false positives in each class, Up, Dw, or Nr, show almost no differences between each other, as opposed to the true positives, which show a clear separation. Thus, there is a subset of loci where the changes in the studied marks are not sufficient to explain the change of expression.

We further explored whether the signals in one single region would be sufficient to predict the expression outcome. Accordingly, for each region, we selected the common attributes from pairs P1 and P2 with CFS score $\geq 80\%$ (Supplementary Table 4). Interestingly, the marks selected for a single region give a prediction accuracy that is comparable to that obtained with attributes from all regions (Supplementary Table 5). The highest accuracy was achieved using gene body ± 5 kb, which is not surprising as it overlaps all the other regions. Interestingly, the 2 kb

Table 5: Correctly classified instances in each transcript subset. Sets are filtered to avoid overlapping gene bodies, promoters or tails from transcript loci from different genes in the same or opposite strands (Section 2). Attribute selection has been applied to each pair: P1 (CFS) and P2 (CFS), for each of the subsets of intron-containing loci, high (HCG) or low (LCG) CG content promoter. The attribute sets correspond to the ones from Table 4(b): P1 (CFS) denotes the model for P1, where the attributes used are those that have a score 80 or higher (maximum 100) using the CFS attribute selection method. P2 (CFS-P1) indicates that the model was trained using the data from P2 but the attributes selected using CFS on P1. P1 (CFS)-on-P2 indicates that the model was trained with pair P1 with only selected attributes and tested on pair P2.

Attributes	Transcript loci set	Instances in total	Correctly classified instances
P1 (CSF)	LCG-IC	1767	1185 (67.06%)
	HCG-IC	1959	1182 (60.34%)
P2 (CSF-P1)	LCG-IC	585	454 (77.60%)
	HCG-IC	792	577 (72.85%)
P1 (CSF)-on-P2	LCG-IC	585	410 (70.09%)
	HCG-IC	792	445 (56.19%)

region downstream of the pA turns out to have a high predictive power, achieving an AUC of 0.89 for upregulated IC-LCG transcripts based only on the signals for H3K27me3 and H3K36me3. Remarkably, one single mark in the region pA ±2 kb is enough to predict upregulated genes with high accuracy (AUC = 0.85 and 0.81 for Up in IC-LCG and IC-HCG transcripts, resp.). This is consistent with the enrichment of H3K36me3 found previously in a region around the pA for active genes [10]. As before, the models achieve higher AUCs for LCGs than for HCGs.

3.3. The Relative Contribution of Marks to the Epigenetic Code. With the aim to find the most relevant attributes that appear to determine the regulation of expression, we calculated the information gain (IG) [46] for all attributes in the subsets HCG-IC and LCG-IC on pair P1 for the unfiltered and the filtered sets (Table 1). The higher the IG value, the better the attribute can separate the three classes: Up, Dw, and Nr. As a control, we generated random attributes for each region, obtained by random sampling z-score values from all attributes in that region. In Figure 3 and Supplementary Figure 5 we show how attributes rank in terms of IG within each region. Although the ranking is very similar before and after filtering transcript loci, we found an overall increase in IG values, indicating that the filtering step improves the specificity of the regulatory code. We found that for all subsets, H3K36me3 is the most informative attribute around the pA site and in gene body associated regions, whereas H3K27ac and H3K9ac are the most informative in the promoter region, which agrees with previous analyses [47]. These two acetylation marks are in fact among the most informative marks in the promoter, around the TSS and in 1st intron and 1st exon regions. Interestingly, H3K36me3 is more informative in the 1st intron than in the 1st exon, which agrees with recent results relating H3K36me3 with splicing of the first intron [22]. Although methylation data shows anticorrelation with expression change in the promoter of HCG loci (Supplementary Figure 6), we observe a modest contribution in the gene body to expression regulation (Figures 2 and 3).

Although IG values determine how well an attribute separates the three sets, Up, Dw, and Nr, we would expect that attributes that most directly associate with expression changes should show no change for the Nr set. That is, we should expect that the enrichment z-scores for Nr should distribute around zero. Accordingly, we defined an attribute to be optimal if the absolute value of the median for the Nr distribution is smaller than 0.1 and the IG is greater than 0.05. If more than one attribute accomplish these thresholds, we considered the one with the highest IG value. Interestingly, this analysis shows that the optimal attributes for H3K36me3 and H3K4me3 correspond to the 1st intron and 1st exon, respectively (Figure 4), which could be related to their role in the coupling between splicing and transcription [22, 38]. Moreover, for H3K9ac and H3K27ac, the optimal attributes are the TSS-5 kb and Promoter-5 kb regions, respectively. DNase-Seq also presented the optimal distribution in the 1st exon, whereas CTCF and H3K4me2 were best in the GB-5 kb region.

We did not find an optimal attribute for RNAPII. Although the attribute for the gene body has minimal median for the Nr distribution and largest IG (Supplementary Figure 7(A)), it shows an enrichment for Nr similar to the Up subset, which could be due to an excess of RNAPII reads in one of the cell lines (Supplementary Figure 7(B)). We also did not find optimal attributes for Methyl-RRBS, H3K4me1, and H4K20me1. For Methyl-RRBS, this is probably due to a large proportion of sites with reads but no methylation evidence (data not shown). The most informative region with minimal median for Nr for H3K4me1 indicates an enrichment of Up in GB ±5 kb but a distribution for Dw and Nr centered on zero, indicating an asymmetry in transcriptional activation. Although H4K20me1 has been related to silent chromatin [48], the most informative of the attributes showed almost no difference between Up, Dw, and Nr subsets. The absence of an optimal attribute for H3K4me1 in GB ±5 kb and for H4K20me1 in the 1st exon might be due to an unequal distribution of reads in K562 relative to GM12878, which does not occur for H3K27me3. Finally, even though we could not find an optimal attribute for H3K27me3, the z-score distributions for the 1st exon results into a clear trend that agrees with the anticorrelation of H3K27me3 and expression (Figure 4), despite the low IG (0.03): Up genes show almost no change, whereas Dw genes

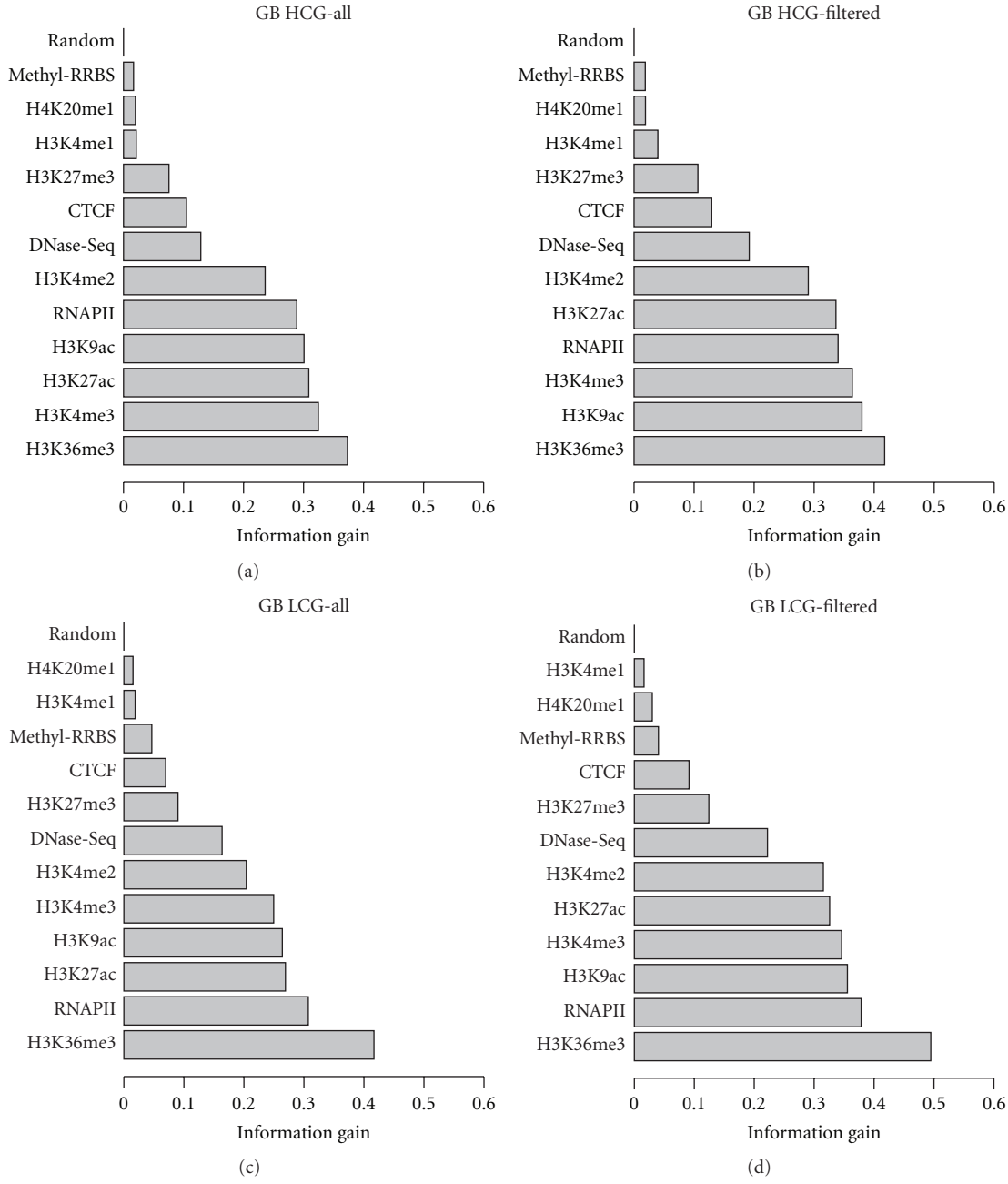


Figure 3: Information gain values measured for attributes in the gene body of intron-containing (IC) transcript loci, comparing before and after filtering loci according to overlap with transcripts from different genes (Section 2). Data is shown for high (HCG) and low (LCG) CpG promoters. Random attributes generated by random sampling z-score values from all attributes in a given region are shown as a control.

show the greatest enrichment, possibly indicating that there is an asymmetry in the pattern of this histone mark for silencing.

3.4. The Effect of Introns in the Epigenetic Code. A number of specific histone modifications have been related to the cotranscriptional splicing of introns [22, 38]. We, therefore, hypothesized that there should be relevant differences in the histone modifications between IC and IL loci. We thus compared the most informative attributes between intron-containing (IC) and intron-less (IL) loci (Figure 5 and Supplementary Figure 8). As there was many more IC than IL loci, we selected a subset of loci from IC of the same size as IL and compared the IG values for attributes related to fixed-length regions (Table 3). For HCG loci, although we found almost no differences when we ranked the attributes according to IG, there is an overall reduction of the IG values in IL genes. Strikingly, we found that for LCG loci the IG becomes very small for most of the attributes. For

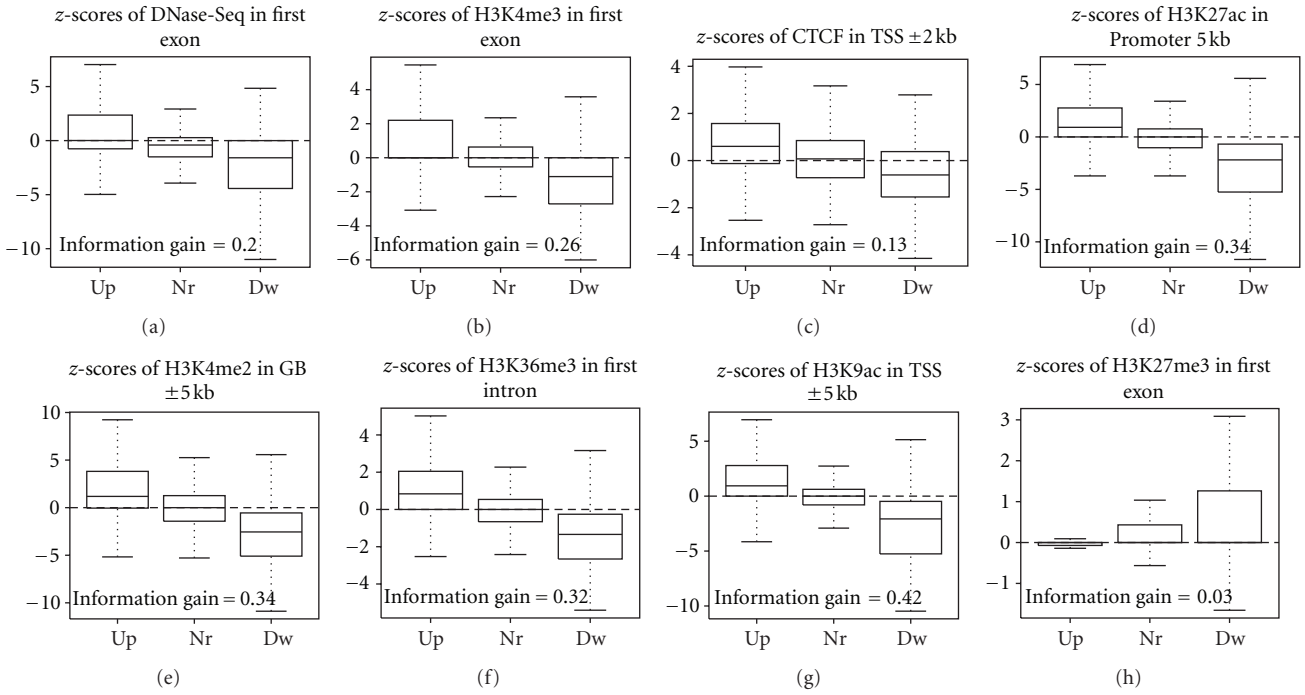


Figure 4: Distribution of z-scores for up- (Up), down- (Dw), and non- (Nr) regulated genes for the optimal attributes for each experiment, calculated by maximizing the Information Gain and minimizing the absolute value of the median for the z-score distribution of the Nr subset. The y-axis shows the z-score corresponding to the enrichment of the attribute. These distributions correspond to the set of LCG-IC loci of Pair1.

instance, in the promoter region, most of the attributes that are informative for LCG-IC loci do not contribute at all in LCG-IL; and H3K36me3, which is considered most relevant downstream of the TSS, and H3K4me1, which is not generally associated to an active TSS, become the most informative attribute for LCG-IL loci. Similarly, in the tail regions most of the attributes that are informative for LCG-IC loci do not contribute for LCG-IL loci, where the IG values are very low. In contrast, the tail region behaves more similarly for HCG-IC and HCG-IL, in terms of ranking and IG value. To further explore the differences in regulation between IC and IL genes, we calculated the profiles of reads for each mark in filtered transcript loci (Section 2). The profiles show large differences between expressed and nonexpressed IC genes (Supplementary Figure 2(A)) and confirm some of the already established locations of the marks relative to the loci. We also observe a striking difference of H3K36me3 read density around the pA in expressed versus nonexpressed IC genes, with a higher density around expressed genes. For IL genes, however, the signal is much weaker. This could be due to the fact that single exon genes tend to occur in families; hence, read mappability may be an issue. However, only 2% of IL genes overlap with low mappability regions, as classified at the UCSC Genome Browser. Nonetheless, we still observe differences between expressed and nonexpressed IL genes (Supplementary Figure 2(B)). For DNA methylation, we observe higher densities upstream of nonexpressed compared to expressed IL genes, consistent with earlier findings [7, 49]. However, we hardly see differences in DNA methylation for IC genes, which appear to be generally less methylated upstream of the TSS and more methylated downstream of the pA. The profile of pseudogenes, which have been excluded from the study of the expression code, are also shown in Supplementary Figure 2(C). Interestingly, although IC pseudogenes have much lower coverage of reads, they have similar profiles to the filtered IC genes, except for the transcription related signals: H3K4me1, RNAPII and H3K36me3, which show almost no signal, indicating nearly absent transcription.

4. Conclusions

A current challenge in epigenetics is how to extract biological knowledge from large volumes of data produced with new high-throughput technologies. Integrative tools and Machine-Learning (ML) algorithms are crucial to this aim. In this article, we have described a novel computational framework for the integration of high-throughput sequencing (HTS) epigenetic data that facilitates the generation and testing of quantitative models of gene regulation. Our methodology proposes a new way to relate epigenetic signals to expression using the comparison of the same locus between two conditions, instead of comparing loci to each other in a single condition, which can be affected by various biases. Three novel aspects of our methodology are that it (1) considers continuous values for the change in epigenetic signals, (2) it explores the enrichment of signals in multiple

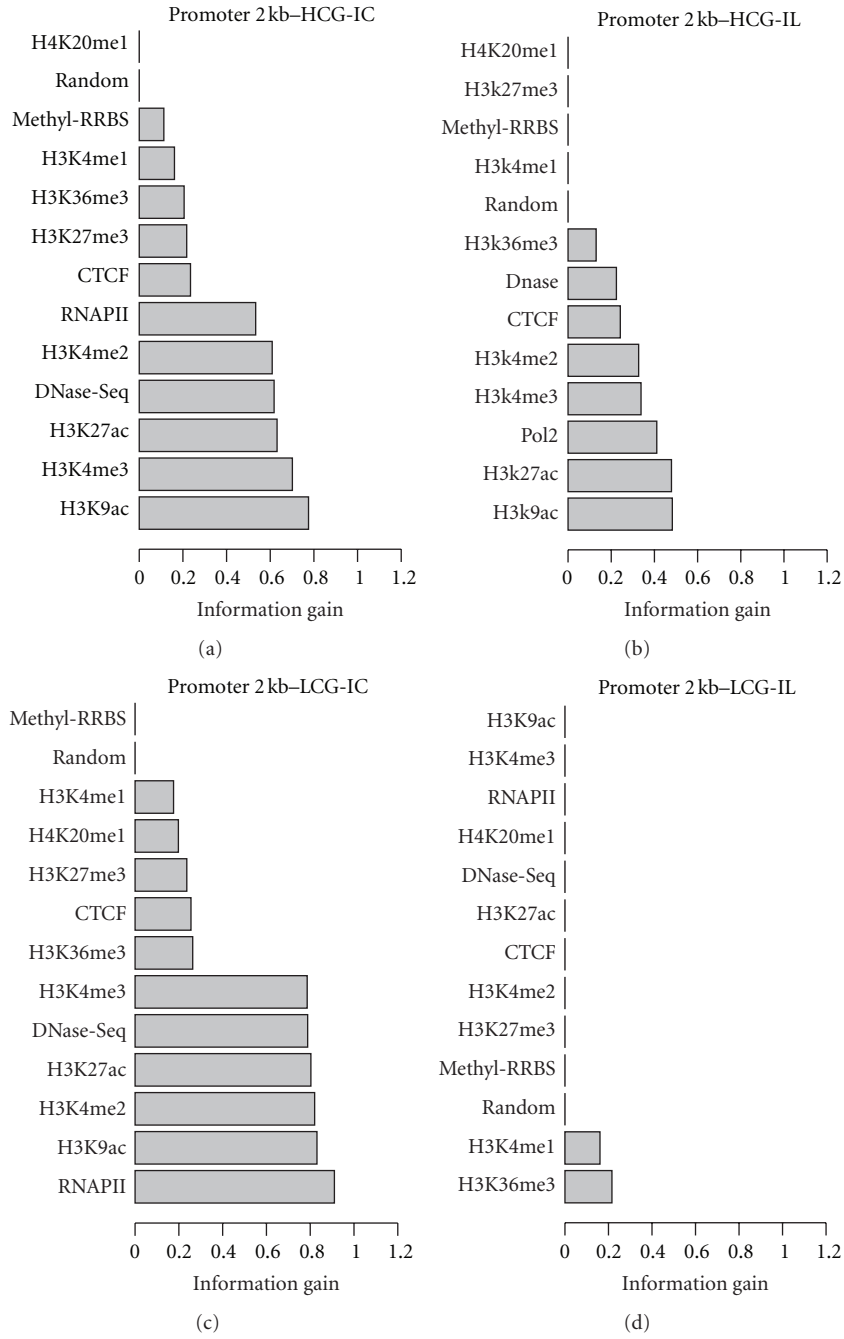


Figure 5: Information gain values measured for attributes in the 2 kb promoter region, comparing intron-less (IL) genes with intron-containing (IC) genes before filtering transcripts (Section 2). The compared sets were taken to be of the same size (105 transcript loci for HCGs and 84 transcript loci for LCGs).

regions and (3) it can be applied to any HTS data type in two conditions.

We have shown the effectiveness of this methodology by building a predictive model of gene expression regulation based on epigenetic information for a pair of cell lines from the ENCODE project. The processed data used to build the models in this paper is available as a Biomart database at http://regulatorygenomics.upf.edu/group/pages/software/ Our quantitative models can predict whether a gene shows expression differences (up or down) or no difference between two cell lines. The relevant attributes and the accuracy for each model vary according to whether transcript loci have high CpG-content promoters (HCG) or not (LCG) and whether they contain introns (IC) or not (IL). These differences indicate that the histone signals are very heterogeneous and that regulation depends strongly on the actual structural properties of promoters and genes. Our analyses also indicate that there is high redundancy in the histone code, as different

groups of attributes from different regions can explain a similar number of regulatory events.

Additionally, we have taken into account a fact largely overlooked in previous publications, which is that a considerable number of gene loci overlap with each other [50] at promoter and tail regions, or over their gene bodies, either on the same or on opposite strands. Accordingly, previous models of expression based on histone marks have this confounding effect, since the strand-less ChIP-Seq signal cannot be unambiguously associated to the regulation of a specific gene. Interestingly, when we removed these overlapping genes, the prediction accuracy improves considerably and the predictive model built from one pair of cell lines performs with high accuracy in a second pair of different cell lines. We conclude that removing these overlapping loci allows us to build a more general epigenetic code for expression regulation. This is further confirmed by our analysis of the information gain (IG), which shows that attributes can separate better the three regulatory classes after the overlapping loci are removed. Notably, this filtering does not change the ranking of IG values, hence although we improve the quantitative description of the histone code, the qualitative description does not change. The IG analysis confirms the role of some of the histone marks, like H3K9ac and H3K27ac, in the promoter and around the transcription start site in expression regulation as described before in the literature; and uncovers new regions, like the first intron for H3K36me3, the first exon for H3K4me3, and downstream of the polyadenylation site for H3K36me3, where changes in these marks associate strongly with expression regulation. The role of these marks in the first exon and intron indicates a general role in the coupling between splicing and transcription, as recently shown in the literature. In this direction, we also explored the patterns of epigenetic changes between intron-containing (IC) and intron-less (IL) loci and found that IC loci contain more epigenetic information and can therefore be better characterised. These differences are more remarkable between high (HCG) and low CpG promoters (LCG), which suggest that the type of promoter might influence the epigenetic changes that take place in cotranscriptional splicing [22]. Alternatively, this could indicate that these loci have a distinct mode of regulation, possibly by other marks that have not been considered in this study.

The epigenetic signals analysed in this study provide a strong prediction power for expression regulation. However, there is still a proportion of genes for which their change in expression cannot be explained by the changes of the studied signals. In any case, the associations found do not necessarily imply causality or a direct functional effect, as the effect of a given histone mark may be context dependent and may occur through the action of other factors. Nonetheless, the models described reflect the complex network of gene regulation and provide some of the generic features of this network. Our methodology provides an effective way to integrate the continuous changes in epigenetic signals between different conditions. Applying this approach to datasets with more histone modifications and transcription factors will help expanding and characterizing further this complex regulatory network. In particular, the application of our approach to different developmental stages, disease states, or treatments, will help uncovering the epigenetic mechanisms responsible for cellular differentiation and carcinogenesis.

Acknowledgments

The authors would like to thank Juan González-Vallinas, Eneritz Agirre, Nicolas Bellora, João Curado, and Alba Jené for useful discussions and Julien Lagarde for the ENCODE data support. This work was supported by Grants BIO2011-23920 and CSD2009-00080 from the Spanish Ministry of Science and by the Sandra Ibarra Foundation. S. Althammer was supported by an FI grant from the Generalitat de Catalunya.

References

[1] R. D. Kornberg and J. O. Thomas, "Chromatin structure: oligomers of the histones," *Science*, vol. 184, no. 4139, pp. 865–868, 1974.

[2] B. Li, M. Carey, and J. L. Workman, "The Role of chromatin during transcription," *Cell*, vol. 128, no. 4, pp. 707–719, 2007.

[3] J. Mellor, P. Dudek, and D. Clynes, "A glimpse into the epigenetic landscape of gene regulation," *Current Opinion in Genetics and Development*, vol. 18, no. 2, pp. 116–122, 2008.

[4] D. K. Pokholok, C. T. Harbison, S. Levine et al., "Genome-wide map of nucleosome acetylation and methylation in yeast," *Cell*, vol. 122, no. 4, pp. 517–527, 2005.

[5] A. A. Joshi and K. Struhl, "Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to pol II elongation," *Molecular Cell*, vol. 20, no. 6, pp. 971–978, 2005.

[6] R. Lister, M. Pelizzola, Y. S. Kida et al., "Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells," *Nature*, vol. 471, no. 7336, pp. 68–73, 2011.

[7] L. Ellis, P. W. Atadja, and R. W. Johnstone, "Epigenetics in cancer: targeting chromatin modifications," *Molecular Cancer Therapeutics*, vol. 8, no. 6, pp. 1409–1420, 2009.

[8] M. Kulis and M. Esteller, "DNA methylation and cancer," *Advances in Genetics*, vol. 70, pp. 27–56, 2010.

[9] T. Jenuwein and C. D. Allis, "Translating the histone code," *Science*, vol. 293, no. 5532, pp. 1074–1080, 2001.

[10] A. Barski, S. Cuddapah, K. Cui et al., "High-resolution profiling of histone methylations in the human genome," *Cell*, vol. 129, no. 4, pp. 823–837, 2007.

[11] H. Yu, S. Zhu, B. Zhou, H. Xue, and J. D. J. Han, "Inferring causal relationships among different histone modifications and gene expression," *Genome Research*, vol. 18, no. 8, pp. 1314–1324, 2008.

[12] G. Hon, B. Ren, and W. Wang, "ChromaSig: a probabilistic approach to finding common chromatin signatures in the human genome," *PLoS Computational Biology*, vol. 4, no. 10, article e1000201, 2008.

[13] B. Van Steensel, U. Braunschweig, G. J. Filion, M. Chen, J. G. Van Bemmel, and T. Ideker, "Bayesian network analysis of targeting interactions in chromatin," *Genome Research*, vol. 20, no. 2, pp. 190–200, 2010.

[14] R. Karlić, H. R. Chung, J. Lasserre, K. Vlahoviček, and M. Vingron, "Histone modification levels are predictive for gene expression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 7, pp. 2926–2931, 2010.

[15] J. Ernst and M. Kellis, "Discovery and characterization of chromatin states for systematic annotation of the human genome," *Nature Biotechnology*, vol. 28, no. 8, pp. 817–825, 2010.

[16] C. Cheng and M. Gerstein, "Modeling the relative relationship of transcription factor binding and histone modifications to gene expression levels in mouse embryonic stem cells," *Nucleic Acids Research*, vol. 40, no. 2, pp. 553–568, 2012.

[17] J. Ernst, P. Kheradpour, T. S. Mikkelsen et al., "Mapping and analysis of chromatin state dynamics in nine human cell types," *Nature*, vol. 473, no. 7345, pp. 43–49, 2011.

[18] T. Jenuwein and C. D. Allis, "Translating the histone code," *Science*, vol. 293, no. 5532, pp. 1074–1080, 2001.

[19] B. D. Strahl and C. D. Allis, "The language of covalent histone modifications," *Nature*, vol. 403, no. 6765, pp. 41–45, 2000.

[20] G. Robertson, M. Hirst, M. Bainbridge et al., "Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing," *Nature Methods*, vol. 4, no. 8, pp. 651–657, 2007.

[21] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, "Mapping and quantifying mammalian transcriptomes by RNA-Seq," *Nature Methods*, vol. 5, no. 7, pp. 621–628, 2008.

[22] S. F. de Almeida, A. R. Grosso, F. Koch et al., "Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36," *Nature Structural and Molecular Biology*, vol. 18, pp. 977–983, 2011.

[23] M.-S. Cheung, T. A. Down, I. Latorre, and J. Ahringer, "Systematic bias in high-throughput sequencing data and its correction by BEADS," *Nucleic Acids Research*, vol. 39, no. 15, article e103, 2011.

[24] L. Teytelman, B. Özaydin, O. Zill et al., "Impact of chromatin structures on DNA processing for genomic analyses," *PLoS ONE*, vol. 4, no. 8, article e6700, 2009.

[25] R. K. Auerbach, G. Euskirchen, J. Rozowsky et al., "Mapping accessible chromatin regions using Sono-Seq," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 35, pp. 14926–14931, 2009.

[26] P. J. Park, "ChIP-seq: advantages and challenges of a maturing technology," *Nature Reviews Genetics*, vol. 10, no. 10, pp. 669–680, 2009.

[27] J. Rozowsky, G. Euskirchen, R. K. Auerbach et al., "PeakSeq enables systematic scoring of ChIP-seq experiments relative to controls," *Nature Biotechnology*, vol. 27, no. 1, pp. 66–75, 2009.

[28] J. C. Dohm, C. Lottaz, T. Borodina, and H. Himmelbauer, "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing," *Nucleic Acids Researchearch*, vol. 36, no. 16, article e105, 2008.

[29] R. M. Myers, J. Stamatoyannopoulos, M. Snyder, I. Dunham, R. C. Hardison, B. E. Bernstein et al., "A user's guide to the encyclopedia of DNA elements (ENCODE)," *PLoS Biology*, vol. 9, no. 4, article e1001046, 2011.

[30] A. P. Boyle, S. Davis, H. P. Shulha et al., "High-resolution mapping and characterization of open chromatin across the genome," *Cell*, vol. 132, no. 2, pp. 311–322, 2008.

[31] A. Meissner, A. Gnirke, G. W. Bell, B. Ramsahoye, E. S. Lander, and R. Jaenisch, "Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis," *Nucleic Acids Research*, vol. 33, no. 18, pp. 5868–5877, 2005.

[32] P. Flicek, M. R. Amode, D. Barrell, K. Beal, S. Brent, D. Carvalho-Silva et al., "Ensembl 2012," *Nucleic Acids Research*, vol. 40, pp. 84–90, 2012.

[33] T. R. Dreszer, D. Karolchik, A. S. Zweig, A. S. Hinrichs, B. J. Raney, and R. M. Kuhn, "The UCSC Genome Browser database: extensions and updates 2011," *Nucleic Acids Research*, vol. 40, pp. 918–923, 2012.

[34] A. Kasprzyk, "BioMart: driving a paradigm change in biological data management," *Database (Oxford)*, vol. 2011, article bar049, 2011.

[35] E. Frank, M. Hall, L. Trigg, G. Holmes, and I. H. Witten, "Data mining in bioinformatics using Weka," *Bioinformatics*, vol. 20, no. 15, pp. 2479–2481, 2004.

[36] S. A. Hoang, X. Xu, and S. Bekiranov, "Quantification of histone modification ChIP-seq enrichment for data mining and machine learning applications," *BMC Research Notes*, vol. 4, article 288, 2011.

[37] J. T. Huff, A. M. Plocik, C. Guthrie, and K. R. Yamamoto, "Reciprocal intronic and exonic histone modification regions in humans," *Nature Structural and Molecular Biology*, vol. 17, no. 12, pp. 1495–1499, 2010.

[38] R. J. Sims III, S. Millhouse, C. F. Chen et al., "Recognition of Trimethylated Histone H3 Lysine 4 Facilitates the Recruitment of Transcription Postinitiation Factors and Pre-mRNA Splicing," *Molecular Cell*, vol. 28, no. 4, pp. 665–676, 2007.

[39] P. Carninci, A. Sandelin, B. Lenhard et al., "Genome-wide analysis of mammalian promoter architecture and evolution," *Nature Genetics*, vol. 38, no. 6, pp. 626–635, 2006.

[40] S. Saxonov, P. Berg, and D. L. Brutlag, "A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 5, pp. 1412–1417, 2006.

[41] E. Valen and A. Sandelin, "Genomic and chromatin signals underlying transcription start-site selection," *Trends in Genetics*, vol. 27, no. 11, pp. 475–485, 2011.

[42] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[43] J. Schug, W. P. Schuller, C. Kappen, J. M. Salbaum, M. Bucan, and C. J. Stoeckert, "Promoter features related to tissue specificity as measured by Shannon entropy," *Genome Biology*, vol. 6, no. 4, p. R33, 2005.

[44] Z. Zhang and M. Q. Zhang, "Histone modification profiles are predictive for tissue/cell-type specific expression of both protein-coding and microRNA genes," *BMC Bioinformatics*, vol. 12, article 155, 2011.

[45] M. Hall, *Correlation-based feature selection for machine learning [Ph.D. thesis]*, Department of Computer Science, Waikato University, Hamilton, New Zealand, 1999.

[46] T. Mitchell, *Machine Learning*, The Mc-Graw-Hill Companies, Inc., 1997.

[47] Z. Wang, C. Zang, J. A. Rosenfeld et al., "Combinatorial patterns of histone acetylations and methylations in the human genome," *Nature Genetics*, vol. 40, no. 7, pp. 897–903, 2008.

[48] J. K. Sims, S. I. Houston, T. Magazinnik, and J. C. Rice, "A trans-tail histone code defined by monomethylated H4 Lys-20 and H3 Lys-9 demarcates distinct regions of silent chromatin," *The Journal of Biological Chemistry*, vol. 281, no. 18, pp. 12760–12766, 2006.

[49] A. Bird, "The essentials of DNA methylation," *Cell*, vol. 70, no. 1, pp. 5–8, 1992.

[50] S. Katayama, Y. Tomaru, T. Kasukawa et al., "Molecular biology: antisense transcription in the mammalian transcriptome," *Science*, vol. 309, no. 5740, pp. 1564–1566, 2005.

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 836374, 10 pages
doi:10.1155/2012/836374

Review Article

Turning on Myogenin in Muscle: A Paradigm for Understanding Mechanisms of Tissue-Specific Gene Expression

Herve Faralli[1] and F. Jeffrey Dilworth[1,2]

[1] *Regenerative Medicine Program, Sprott Center for Stem Cell Research, Ottawa Hospital Research Institute, Ottawa, ON, Canada K1H 8L6*
[2] *Department of Cellular and Molecular Medicine, University of Ottawa, Ottawa, ON, Canada K1H 8M5*

Correspondence should be addressed to F. Jeffrey Dilworth, jdilworth@ohri.ca

Received 14 February 2012; Accepted 23 April 2012

Academic Editor: Lucia Latella

Expression of the *myogenin* (*Myog*) gene is restricted to skeletal muscle cells where the transcriptional activator turns on a gene expression program that permits the transition from proliferating myoblasts to differentiating myotubes. The strict temporal and spatial regulation on *Myog* expression in the embryo makes it an ideal gene to study the developmental regulation of tissue-specific expression. Over the last 20 years, our knowledge of the regulation of *Myog* expression has evolved from the identification of the minimal promoter elements necessary for the gene to be transcribed in muscle, to a mechanistic understanding of how the proteins that bind these DNA elements work together to establish transcriptional competence. Here we present our current understanding of the developmental regulation of gene expression gained from studies of the *Myog* gene.

1. Introduction

The diploid human genome encodes the genes required to establish the ~200 different cell types that make up the body. Each of these different cell types can be defined by the complement of genes that they express. These cell-specific gene expression programs are established through spatially and temporally defined signals from hormones, cytokines, and growth factors that modulate transcription factor activity. Once established, these gene expression programs must then be transmitted to daughter cells through epigenetic mechanisms. Studies in *Drosophila* have identified Trithorax (TrxG) and Polycomb (PcG) group proteins as the mediators of this epigenetic cellular memory [1]. However, the PcG and TrxG proteins display relatively ubiquitous expression and therefore cannot work in isolation to mediate temporal and spatial regulation of gene expression. Thus, in order to understand how tissue specific patterns of gene expression are established we must examine how the TrxG and PcG proteins work with the transcriptional machinery in specific cells to modulate expression of a particular gene.

The skeletal muscle-specific gene *myogenin* (*Myog*) is a key developmental regulator for skeletal muscle formation and is one of the better studied tissue-specific genes. The *Myog* gene encodes a transcription factor of the basic-helix-loop-helix (bHLH) protein family. Displaying expression that is highly restricted, both temporally and spatially, Myog transcripts are first detected in the primary myotome of the developing mouse embryo at around day E9 [2, 3]. Myog then continues to be expressed in all the newly formed skeletal muscle of the trunk and the limb bud during embryonic myogenesis before being downregulated in the mature muscle fiber. The importance of Myog expression in the developing embryo is highlighted by the fact that knockout mice fail to form myofibers [4, 5]. This phenotype is consistent with studies in cultured cell systems showing that Myog is not expressed in the proliferating myoblast, but is regulated early in the terminal differentiation process where it is required to turn on the muscle gene expression program [6]. Myog is also expressed in regenerating adult myofibers where its expression is induced 4-5 days after muscle damage [7]. However, the role of Myog in the differentiation process in regenerating

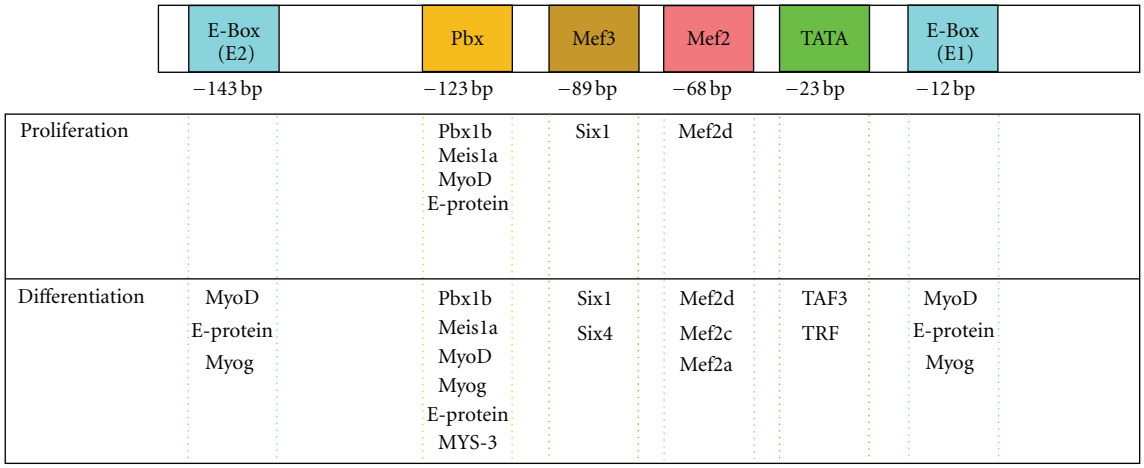


Figure 1: Conserved DNA binding elements within the *Myog* promoter. Conserved DNA binding elements within the proximal promoter (−184 to +33 bp) that have been characterized for a role in regulating *Myog* expression include: E-Box (E1 and E2—blue), Pbx element (orange), Mef3 (red), Mef2 (pink), and the TATA Box (green). Transcription factors that are known to bind at each of the elements in either proliferative or differentiation myoblasts are summarized.

muscle appears to be less critical as conditional knockout in adult muscle does not show a regeneration defect [8]. The alternate pathway that permits adult muscle differentiation in the absence of Myog has not yet been established. Thus, Myog is expressed at critical points in development where it plays an essential role in embryonic myofiber formation and facilitates regeneration of damaged muscle. This highly restricted temporal and spatial expression of the gene makes it an ideal model to study developmentally regulated gene expression. This paper will discuss what we have learned from 20 years of studying the regulation of Myog expression and the questions that remain to be answered.

2. The *Myog* Locus

In mice, Myog is transcribed from a gene that is 2.5 kb in length on chromosome 1. Splicing of the three exons coded within this gene gives rise to an mRNA of 1.5 kb length. The fact that there are no splicing variants or known alternate transcription start sites further simplifies its study. Transgenic mouse studies in the early 1990s by the Rigby and Olson groups were key to defining the minimal promoter region required to ensure expression of Myog in the myotome during embryonic myogenesis [9, 10]. Using a LacZ reporter driven by *Myog* regulatory elements, low levels of expression could be observed in muscle using a construct containing the −130 to +18 bp region of the promoter. While the level of LacZ expression from the construct was relatively weak, these experiments clearly established that this short fragment of DNA was sufficient to ensure both the proper temporal and spatial expression of the *Myog* gene. This region of the *Myog* promoter contains several evolutionarily conserved DNA binding elements that are very well characterized (see Figure 1). These include the TATA Box (TFIID or TAF3/TRF3), Mef2 site (Mef2A, Mef2C, or Mef2D), Mef3 site (Six1 or Six4), Pbx (Pbx1 or MSY3), and an E-Box (MyoD/E-protein, Myf5/E-Protein, or *Myog*/E-protein). The roles of these elements in the regulation of Myog expression will be discussed below.

Another point to be taken away from these transgenic studies is the fact that additional DNA elements beyond the −130 to +18 bp sequence are required for high-level expression of the reporter construct suggesting the presence of an enhancer element somewhere between the −1092 to −340 bp of *Myog* gene [9, 10]. Interestingly, this region does not appear to be evolutionarily conserved through mammals but we cannot rule out that it could contain a murine-specific enhancer. Comparative analysis of the *Myog* gene from different genomes (Figure 2) show that three additional uncharacterized enhancers may exist at −4.5 kb, −5.5 kb, and −6.5 kb upstream of the transcription start site (TSS). In addition to high conservation across species, these putative enhancers are marked by acetylation of H3K27 (H3K27ac) and DNase I hypersensitivity in human skeletal muscle cells (ENCODE/BROAD [11]). Furthermore, these sites are marked by the additional enhancer enriched-epigenetic modification Histone H3 lysine 4 monomethylation (H3K4me1) in mouse myoblasts [12]. Analysis of genome wide studies show that the −4.5 kb (Enh3) and −6.5 kb (Enh1) enhancers, but not the −5.5 kb (Enh2), are bound by MyoD in differentiating myoblasts (see Figure 2) [13]. Interestingly, Myog binding is observed at these same two putative enhancers in myotubes differentiated for 60 hr, while only the −4.5 kb enhancer is bound at 24 hr of differentiation (unpublished observation from ENCODE/CalTech data). Thus, the *Myog* locus contains three elements that appear to possess several enhancer-like characteristics and show some unique aspects of regulation. Further studies will be necessary to elucidate their roles in regulating/enhancing Myog expression during myogenesis. One possibility is that these additional regulatory regions could permit fine-tuning of Myog expression in specific muscles. Examples

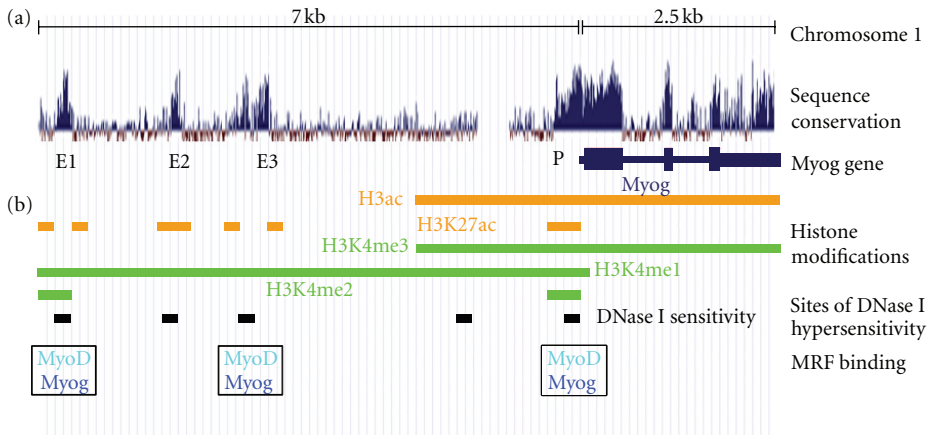


Figure 2: Characteristics of the *Myog* locus in differentiating muscle. (a) Conservation across mammalian species is plotted across the *Myog* locus representing the region from −7.0 kb to +2.5 kb for the transcription start site. The map shows the well-characterized proximal promoter (P) and three predicted enhancers (E1, E2, and E3) that remain uncharacterized. (b) Summary of *myogenin* locus characteristics as identified from high-throughput studies of muscle cells. Regions enriched for total Histone H3 acetylation or for acetylation at histone H3 lysine 27 (H3K27ac) are shown in orange. DNase I hypersensitive sites are shown in black. Regions marked by methylation of histone H3 lysine 4 (H3K4me1, H3K4me2, and H3K4me3) are shown in green. The positioning of MyoD and Myog binding sites as identified by ChIP-Seq studies are shown.

of alternate regulation of Myog in different muscles have been reported. Indeed, during late stage of embryogenesis (days E16.5 to E19.5), innervation of the *extensor digiform longus* muscle leads to downregulation of both MyoD and Myog [14]. In contrast, innervation of the *soleus* muscle leads to a downregulation of MyoD expression, while Myog expression level stays the same [14]. Alternatively, these uncharacterized regulatory regions could be responsible for modulating Myog expression at distinct temporal stages such as embryonic versus adult myogenesis, or even precise stages of embryonic development. Evidence suggesting differential transcriptional regulation of the *Myog* gene during embryonic myogenesis is provided from studies showing that the Mef2 binding element is required for expression of Myog in the developing limb bud and a subset of somites at day E11.5, but not at day E12.5 [9, 10]. This would suggest that *Myog* requires the activity of multiple different transcriptional regulators to ensure a precise temporal and spatial regulation of gene expression. It remains to be determined how these three highly conserved DNA regions at the *Myog* locus contribute to regulating the expression of Myog in muscle development and regeneration.

3. DNA Bound Transcription Factors That Modulate *Myog* Expression

The primary DNA sequence of the proximal (−130 to +18 bp) region of the *Myog* promoter region has been extensively studied, and multiple conserved binding elements have been characterized (see Figure 1). Indeed, each of these promoter elements appears to be crucial to the proper expression of Myog in the embryo. Initial studies focused on the cluster of elements that include the E-Box, TATA Box, and Mef2 element [9, 10]. More recently an important role for the Mef3 and Pbx binding elements have been shown for Myog expression [15, 16].

The TATA box is required for the binding of TFIID that directs the assembly of the general transcriptional machinery at the promoter region. It has also been shown to bind the TAF3/TRF3 transcriptional complex to the *Myog* promoter [17]. Studies of the minimal *Myog* promoter driving expression of a reporter gene in chick myoblasts show that deletion of this element within the context of the proximal promoter completely blocks its expression [9].

The E-box (E1) present between the TATA box and the transcription start site is the binding site for myogenic bHLH protein complexes, including MyoD/E-protein and Myog/E-protein heterodimers. Mutation of the E1 E-box in the context of the proximal promoter led to a block of Myog expression in the mouse myotome during development [10]. It is interesting to note that the extension of the promoter to generate a fragment running from −180 bp to +18 bp restored expression of the reporter gene in the myotome even when the E1 E-box was mutated [10]. This finding is important as the slightly longer construct contains a second E-box (E2), and suggests that the exact positioning of the MyoD binding site is not essential to the promoter function in establishing muscle-specific gene expression. However, the E2 E-box is not conserved in humans, suggesting that the E1 E-box is likely the more important MyoD binding site mediating muscle development.

The Mef2 binding element in the *Myog* promoter is bound by members of the Mef2 family of transcription factors—including Mef2a, Mef2c, and Mef2d. Mutation of the Mef2 binding element in the context of the proximal promoter driving expression of the reporter gene blocks the activation in both chick myoblasts and fibroblasts undergoing myogenic conversion [9]. *In vivo*, the Mef2

binding element is required for the activation of the LacZ reporter construct in the somites posterior to somite 7 (but not the most rostral somites) of day E10.5 mouse embryos in the context of a −1565 to +18 bp promoter construct [10, 18]. This result suggests that the Mef2 binding element is required for activation of the *Myog* in some developmental contexts, but that activation of the *Myog* gene can also be achieved through alternative binding elements.

The Mef3 binding element serves as a binding site for the Six family of transcription factors—including Six1 and Six4. Mutation of the Mef3 site in the context of the −184 to +18 bp *Myog* promoter has also been shown to be crucial to the proper expression of the reporter gene in the developing embryo [15]. Consistent with this finding, knockout of *Six1* in mice leads to an impaired primary myogenesis, muscle hypoplasia, and decreased endogenous Myog in the limb buds [19].

The Pbx binding element (or *myog*HCE) has been shown to serve as a binding site for at least two different proteins—Pbx-Meis heterodimers [16] and Pbx-MSY3 [20]. Studied in the context of the proximal promoter, Pbx-Meis heterodimers bind the Pbx binding element in proliferating myoblasts. The binding of the heterodimer then facilitates the targeting of MyoD to the *Myog* promoter through a tethering mechanism [16]. This promoter element (−130 to +18 bp) has not been studied in transgenic mice. However, transgenic studies using a −1092 to +18 bp reporter construct with a mutated Pbx binding site have shown that this element is not required for proper Myog expression in the developing embryo [20]. Interestingly, transgenic studies using a −1092 to +18 bp reporter construct containing the mutated Pbx binding site displayed persistent Myog expression in postnatal muscle [20]. The persistent expression of Myog in the adult myofibers has been attributed to the loss of MSY-3/Pbx complex binding at the promoter. Thus, the Pbx binding element plays two separate roles in myogenesis—firstly Pbx-Meis binding in order to target MyoD to the promoter and initiate gene expression, and secondly to permit Pbx-MSY3 binding that mediates downregulation of the gene later in development.

4. Co-Regulators That Modulate *Myog* Expression

4.1. Repression of the Myog Promoter in Proliferating Myoblasts. In proliferating myoblasts, the *Myog* gene exists in a transcriptionally repressed state. Though not yet expressed, fluorescence *in situ* hybridization (FISH) has shown that the repressed *Myog* locus is localized to the nuclear lumen in myoblasts [21]. However, the locus is marked by hypermethylation of the DNA suggesting a transcriptionally repressive chromatin environment [22, 23]. Examination of the primary DNA sequence within the *Myog* locus shows a relatively low density of CpG residues [22] suggesting that the role of this methylation might be distinct from classical repressive mechanisms mediated by methylated CpG islands [24]. Nevertheless, the modification of cytosine nucleotides within the *Myog* promoter appears to play a role in the negative regulation of transcription [22, 23, 25]. Studies in the developing embryo (day E9.5) show that a reporter gene containing the *Myog* proximal promoter (−192 to +58 bp) is more extensively methylated in anterior somites that have not yet expressed Myog compared to posterior somites that do express Myog [23]. Similarly, Fuso et al. document a strong of methylation of cytosine residues across a region consisting of −1092 bp to the transcription start site of *Myog* in growing myoblasts [22]. What remains unclear is whether this methylation within the promoter is mediated by *de novo* methyltransferase activity (DMNT3a/DNMT3b) targeted to the *Myog* promoter or maintenance methyltransferase activity (DNMT1) during DNA replication.

While in many cases DNA methylation is thought to prevent binding of transcription factors to DNA [24], that does not appear to be the case for the repression of Myog in proliferating myoblasts. Indeed, binding of Pbx1 [16], MyoD [13], and Six1 [26] is observed at the *Myog* proximal promoter in growing myoblasts suggesting that the presence of DNA methylation is not inhibitory to the targeting of these important factors to the loci. Instead, the methylation of DNA within the proximal promoter appears to be essential to the recruitment of the transcriptional repressor CIBZ which directly binds isolated methylated CpG sequences [25]. The mechanism by which CIBZ participates in the repression of the *Myog* gene remains unclear but it is interesting to note that knockdown of the methyl-binding protein leads to transcriptional activation of the promoter in the absence of CpG demethylation [25]. Thus it appears that methylation of the CpG poor *Myog* promoter region helps repress expression through the recruitment of the CIBZ protein (see Figure 3).

In addition to DNA methylation the posttranslational modifications of histones play an important role in maintaining a transcriptionally repressive environment at the *Myog* promoter. Among the marks that are known to play a role in repressing the *Myog* gene are the methylation of histone H3 at lysine 9 (H3K9) and lysine 27 (H3K27). Indeed, the repressed *Myog* promoter has been shown to be marked by both dimethyl-H3K9 (H3K9me2) and trimethyl-H3K9 (H3K9me3) in proliferating myoblasts [27, 28]. The presence of these marks at the *Myog* promoter has been attributed to the H3K9 methyltransferase KMT1A/Suv39h1 [27, 28] which is targeted to the locus through an interaction with MyoD [29] (see Figure 3). The recruitment of KMT1A to the *Myog* promoter is modulated by the phosphorylation of MyoD by p38γ MAPK [30]. The importance of this recruitment is highlighted by the fact that the knockdown of KMT1A/Suv39h1 in growing myoblasts leads to a precocious activation of Myog [29]. In addition KMT1A/Suv39h1, the H3K9 methyltransferase G9a also associates with the *Myog* promoter in repressive conditions [31], though its contribution to establishing the repressive chromatin state has yet to be elucidated. Instead, it has been shown that in proliferating myoblasts, G9a associates with MyoD at the *Myog* promoter to mediate a methylation of the muscle regulatory factor [31]. This methylation of MyoD antagonizes a competing acetylation by pCAF that is required to facilitate recruitment of additional coactivators to the gene [32, 33]. Interestingly, myoblasts that express exogenous G9a continue to display

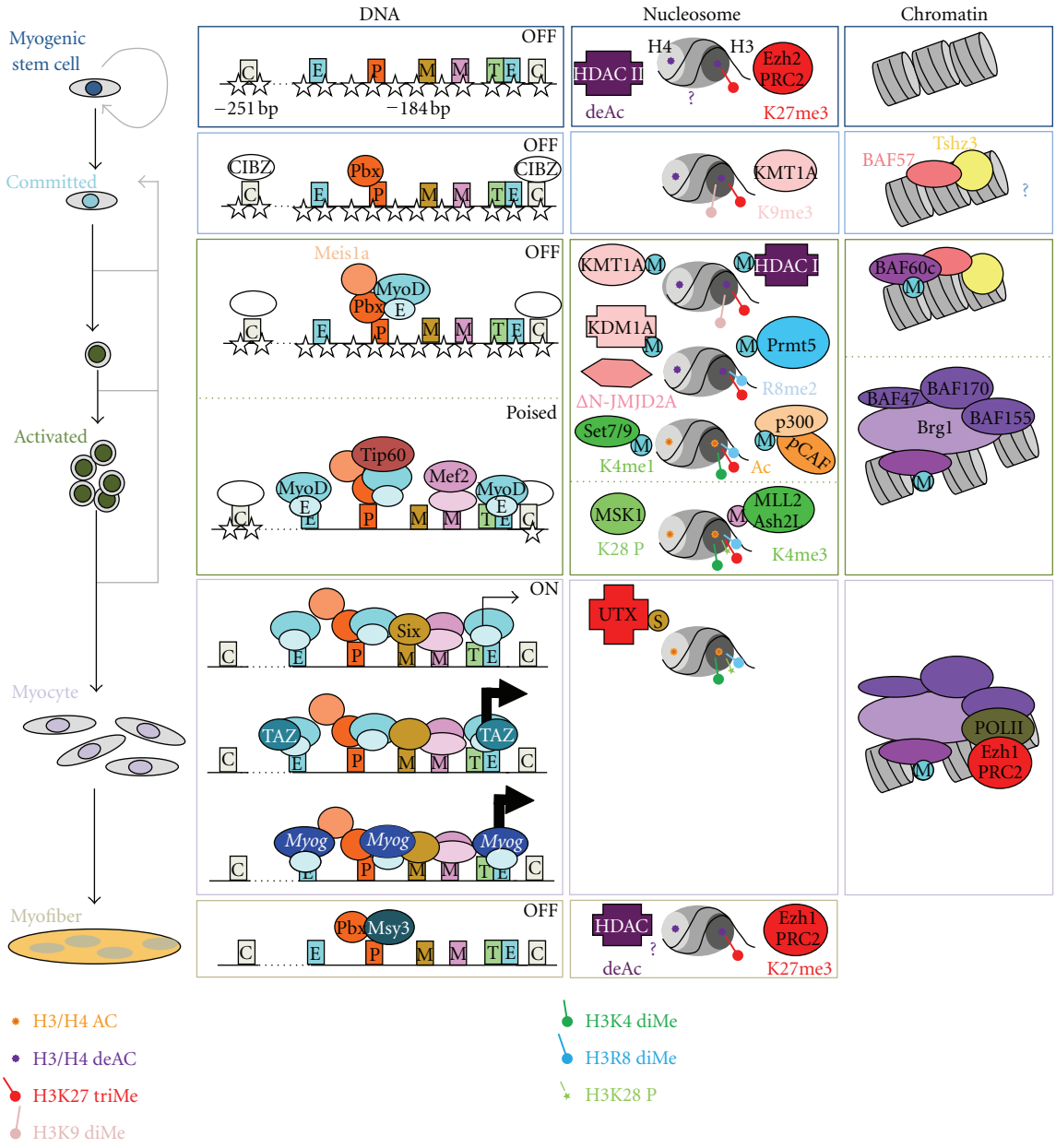


Figure 3: Model for the regulation of *Myog* gene expression. The model summarizes current knowledge of the regulation of *Myog* gene expression. See main body of the text for details. The first column highlights the different steps of myogenic differentiation. The muscle stem cell/progenitors (dark blue cell) undergo commitment to the muscle lineage (light blue cell), then proliferate (dark green cell), differentiate into myocytes (purple cell), and fuse with each other to form myofibers, the functional unit of the muscle. The second column highlights the role of DNA binding transcription factors at the *myogenin* promoter. The boxes correspond to the DNA binding site for the specific transcription factors: CCGG repeat (C, grey); E-box (E, blue); Pbx1b (P, orange) Mef3 (M, red); Mef2 (M, pink); TATA box (T, green). DNA methylation is represented by white stars and the transcription factors which bind the DNA elements are coded by colored circle. Transcriptional repression and activation are indicated by OFF and ON, respectively. The light black arrows indicate low-level transcription, while the thick black arrows represent strong transcription. The third column represents the regulation of *Myog* through histone modifications. The nucleosome is represented by circles: grey (Histone H4), medium grey (the Histone H2A/H2B core), dark grey (Histone H3). Deacetylation of the lysine is represented by pink stars (deAC), while acetylation (Ac) is indicated by orange stars. Modifications of histone H3 map as colored lollipops indicates: monomethylation (Me1), dimethylation (Me2), trimethylation (Me3), and phosphosrylation (P) of specific amino acids. The histone modifying enzymes are represented by colored circles (methyltransferase, acetyltransferases and kinases) or colored crosses (deacetylases and demethylases). The smaller circles represent the transcription factors: MyoD (blue), Mef2 (pink), and Six (brown). The fourth column represents regulation of chromatin structure. Nucleosomes are represented by the black DNA wrapped around grey cylinders (histone octamers): heterochromatin (tightly packed nucleosomes) and euchromatin (spaced nucleosomes). The SWI/SNF subunits are represented as colored circles. The smaller circles represent the transcription factor MyoD (blue).

an H3K9me2 in differentiation conditions and prevent myogenesis [31]. It is not clear whether this continued marking of the promoter by H3K9me2 is a cause or consequence of the impaired differentiation. Thus, the contribution of G9a to the establishment of the H3K9me2 mark at the *Myog* promoter remains to be investigated.

Trimethylation of H3K27 (H3K27me3) is a very well-characterized histone modification that has been shown to mark developmentally regulated genes to maintain them in a transcriptionally silent state [34]. Consistent with the fact that it displays a strict temporal and spatial regulation of transcription, the *Myog* locus is marked by H3K27me3 in proliferating myoblasts [12, 35], and nonmuscle (erythroleukemia—K562) cells (H.F. and F.J.D, unpublished observation based on available data from ENCODE/University of Washington). However the exact delimitation of the regions of the *Myog* locus marked by H3K27me3 in myoblasts varies between reports depending on whether the chromatin immunoprecipitation experiments were performed under native or cross-linked conditions [12, 35]. Nevertheless, both studies clearly demonstrate a role for H3K27me3 in maintaining repression of Myog gene expression. Furthermore, these studies establish that this repressive H3K27me3 mark is mediated by the PcG protein Ezh2 which is a component of the PRC2 complex. Although the mechanism by which the PRC2 complex is targeted to the *Myog* promoter is not known, the Ezh2 protein has been shown to associate with a region at −1500 bp upstream of the TSS [12] as well as the proximal promoter [35, 36] (see Figure 3). The functional importance of the PRC2 complex to maintaining the repressed state at the *Myog* gene was demonstrated by knockdown of Suz12—a subunit of the PRC2 complex critically required to methylate H3K27. Loss of Suz12 in growing myoblasts leads to a loss of H3K27me3 at the −1500 bp region of the gene and results in expression of Myog under proliferative conditions [12]. Thus, the temporal regulation of Myog expression is clearly regulated through the activity of the PRC2 complex and its associated H3K27 methyltransferase activity.

The repression of Myog expression is also associated with the removal of transcriptionally permissive histone modifications. In particular, the *Myog* promoter is known to be targeted by histone deacetylase (HDAC) enzymes that are responsible for removing acetyl groups from lysines within the Histone H3 and H4 of the nucleosome. Targeting of the HDAC enzymes to the *Myog* promoter occurs through a direct interaction with MyoD [28, 29, 37, 38]. Interestingly, Mef2 proteins are also able to interact with class II HDAC (HDAC4 and HDAC5) enzymes [39]. This suggests the possibility that MyoD and Mef2 proteins might cooperate to ensure efficient recruitment, and tight association of HDACs with the *Myog* promoter to mediate repression in proliferating myoblasts. Interestingly, studies using innervated muscle show that Mef2 forms a complex with Dach2, MITR (HDAC9) and class I HDACs to mediate the downregulation of Myog expression [40, 41]. It remains to be determined whether the same group of proteins acts to repress Myog expression in proliferating myoblasts.

Lastly, recent studies have shown that two subunits of the ATP-dependent chromatin remodeling complex SWI/SNF are associated with the repressed *Myog* promoter [42, 43]. Indeed, BAF60c and BAF57 are shown to associate with the *Myog* promoter in proliferating myoblasts. However, this association occurs as a subcomplex, as BRG1/BRM and other core SWI/SNF subunits are not associated with the repressed *Myog* promoter [43] (see Figure 3). The role of these SWI/SNF subunits at the promoter in the absence of the remodeling activity is not clear. In the case of BAF57, this subunit was shown to directly participate in the repression of myogenesis through its association with the zinc-finger protein Tshz3 [42]. While it is not clear if BAF60c is required for the repression of Myog expression, this subunit has been shown to be recruited by MyoD to the promoter to permit efficient assembly of the functional SWI/SNF remodeling complex upon signals that mediate terminal muscle differentiation [43]. Future studies should provide us with insight into the identity of the additional components of BAF57/BAF60c containing complex, and the role of this group of proteins in maintaining repression of the *Myog* promoter.

4.2. Activation of the Myog Promoter in Differentiating Myotubes. Under conditions permissive to terminal myogenesis, expression of the *Myog* gene is activated relatively early in the developmental program. In differentiating C2C12 myoblasts, a change in DNA methylation status can be observed as early as 2 hrs after induction of differentiation, though 24 hrs is required to see complete demethylation [22]. Studies using transgenic mice expressing a reporter construct with a −192 to +58 *Myog* promoter element suggest that Six1 and Mef2 binding is required for the demethylation of DNA at this locus [23]. However, Six1 and Mef2 binding itself is not likely sufficient for recruitment of a DNA demethylase activity since these proteins are bound at the *Myog* promoter in proliferating myoblasts [26]—where the gene is methylated. Thus, a signal-dependent event is likely to be required to mediate the recruitment of this yet unidentified DNA demethylase enzyme.

Among the different signaling pathways that are activated during induction of myogenic differentiation the best characterized is that of the p38 MAPK signaling pathway. The use of small molecule inhibitors first showed that p38 MAPK signaling is required for myoblast differentiation and cell fusion [44]. A direct link to Myog gene regulation was demonstrated when Perdiguero et al. demonstrated that activation of the p38 signaling pathway in proliferating myoblasts lead to an activation of Myog expression [45]. The p38α MAPK is responsible for this activation of transcription, where it phosphorylates several proteins involved in regulating Myog expression. These p38α MAPK targets include the E-proteins E12/E47 (which facilitates dimerization with MyoD—[46]), the SWI/SNF subunit BAF60c (which allows the incorporation of the BAF60c subunit into the core SWI/SNF complex [43]), and Mef2 proteins (which allows for their interaction with Ash2L/MLL2 protein complexes [47]). Thus, p38 MAPK signaling plays a key role in assembling the factors necessary for establishing the transcriptionally permissive

promoter. Other signaling pathways that have been implicated in activating transcription at the *Myog* promoter include calcineurin [48] and AKT signaling [49].

While the activation of differentiation promoting signaling pathways permits the assembly of transcriptional activators at the *Myog* promoter, several key repressors of transcription are downregulated early in terminal muscle differentiation. Indeed, both H3K27 methyltransferase Ezh2 and the H3K9 methyltransferase G9a undergo decreased expression at the onset of terminal myogenesis [12, 31, 50]. Importantly, the loss of G9a in the muscle cells allows MyoD to become acetylated at lysine 99, 102, and 104 [31], a modification that is established by the MyoD-dependent targeting of the pCAF acetyltransferases to the *Myog* promoter [32, 33]. This acetylation of MyoD then allows for a stabilization of the interaction between the DNA transcriptional activator and a second acetyltransferase p300 [32, 33, 51]. Once associated with the *Myog* promoter, p300 then mediates the acetylation of nucleosomes through the modification of specific lysines within histone H3 and H4 to create a transcriptionally permissive environment (see Figure 3). Further contributing to the acetylation of histones within the *Myog* promoter is the acetyltransferase Tip60 that is recruited to the locus via a direct interaction with MyoD [52].

In addition to the acetylation of histones observed at the *Myog* promoter in the early stages of differentiation, a change in nucleosome methylation is also observed. This includes both the removal of transcriptionally repressive histone marks, as well as the depositing of transcriptionally permissive modifications. As mentioned above, the transcriptionally repressed *Myog* promoter is marked by both H3K9me2/3 and H3K27me3 in proliferating myoblasts. The removal of the H3K9 methyl marks is mediated through the activity of the histone demethylase KDM1A/LSD1A [53] using a mechanism that appears to be facilitated by the activity of ΔN-JMJD2A [54]. Indeed, while the ΔN-JMJD2A isoform is a variant that lacks demethylase activity due to truncation, its presence at the *Myog* promoter is required to observe loss of the repressive H3K9me3 mark. KDM1A and ΔN-JMJD2A are proposed to be recruited to the *Myog* promoter through interactions with MyoD and Mef2, respectively [53, 54], suggesting a synergy between the two transcriptional activators in converting the promoter to a transcriptionally permissive state. MyoD further promotes the departure of the H3K9 methyl mark at *Myog* through the recruitment of the Set7/9 methyltransferase to the locus [55] (see Figure 3). The Set7/9 enzyme is responsible for the establishment of the monomethylation of histone H3 lysine 4 (H3K4me1) (see Figure 3). Accumulation of this H3K4me1 mark is important to the activation of Myog, as the presence of methylation on histone H3 at positions K4 and K9 is mutually exclusive [56]. Thus, the H3K4me1 mark acts to ensure that spurious H3K9me3 activity is prevented from repressing *Myog* transcription.

The association of MyoD with the *Myog* promoter is also responsible for the recruitment of the PRMT5 arginine methyltransferase to the locus [57]. The PRMT5 methyltransferase is responsible for establishing dimethylation of arginine 8 of histone 3 (H3R8me2) at the *Myog* promoter, a mark required for transcriptional activation during differentiation [57]. Though it remains to be determined whether the presence of H3R8me2 acts to sterically hinder the H3K9 methyltransferases, the Imbalzano group have clearly established that this histone modification is required for stable association of the SWI/SNF chromatin remodeling complex at the *Myog* promoter [57] (see Figure 3). As mentioned above, the phosphorylation of the MyoD-associated BAF60c subunit on the repressed promoter by p38α MAPK allows for incorporation of subunit into the core complex [43]. This association between MyoD and BAF60c would allow an initial targeting of the SWI/SNF complex to the *Myog* promoter. The association of the SWI/SNF complex with the promoter is then likely stabilized through the interaction of specific subunits with H3R8me2 and acetylated histone H3 and H4 marks [57, 58]. Thus, MyoD plays a highly important role in establishing nucleosome positioning at the *Myog* promoter.

The recruitment of the SWI/SNF complex to the *Myog* promoter is critical to the activation of gene expression [58]. Indeed, studies suggest that MyoD does not bind to the E-boxes when it associates the transcriptionally repressed *Myog* promoter [16]. Instead, it appears to be tethered to the promoter through an interaction with the DNA-bound Pbx1 protein [16] as the MyoD is likely sterically hindered from recognizing its E-box by the presence of a nucleosome (see Figure 3). It is thus proposed that the Pbx1-tethered MyoD protein facilitates the recruitment of the SWI/SNF complex onto the *Myog* promoter to permit a reorganization of the nucleosomes that would in turn facilitate the binding of the transcriptional activator to the E1 E-box element [58].

MyoD binding to the proximal promoter is also critical to the recruitment of the basal transcriptional machinery. A direct interaction between MyoD and TAF3 allows the recruitment of the TRF3/TAF3 complex to the TATA box of the *Myog* promoter in differentiation conditions [59] (see Figure 3). Furthermore, MyoD has been shown to interact directly with the TFIIB subunit of the preinitiation complex [60]. The binding of these factors to the *Myog* promoter is then likely to permit the complete preinitiation complex to form at the promoter. Thus it appears that MyoD has the ability to recruit most of the factors necessary to activate transcription from the *Myog* promoter.

The final step in the activation of the *Myog* promoter appears to be the removal of the transcriptionally repressive H3K27me3 mark, and the establishment of the transcriptionally permissive H3K4me3 mark within the gene (for review see [61]). The efficient removal of the repressive H3K27me3 mark from the *Myog* gene is mediated though several parallel events, though the order in which they occur remains unclear. The most straightforward mechanism for removal of the repressive histone marks is nucleosome exchange. Indeed, it has recently been shown that the chromatin within the promoter of the *Myog* gene undergoes a shift from an Histone H3.1 containing nucleosome to a nucleosome containing the variant Histone H3.3 [62] (see Figure 3). This exchange occurs through the activity of the histone chaperone HIRA which is targeted to the *Myog* promoter by Mef2 proteins and results in an erasure of

the repressive histone mark [62]. A second mechanism at work on the *Myog* promoter is the phosphorylation of Histone H3 serine 28 (H3S28P) by the Msk1 kinase that occurs upon differentiation [36] (see Figure 3). This phosphorylation event inhibits the association between Ezh2-containing PRC2 complexes and favors the binding of the Ezh1-containing PRC2 complexes [36] that show a much weaker H3K27 methyltransferase activity [63]. It is not known how the Msx1 kinase is targeted to the *Myog* promoter to mediate this exchange of PRC2 complexes. Finally, the decreased levels of H3K27me3 are also established through the recruitment of the histone demethylase UTX/KDM6A [35]. This TrxG protein is recruited to the *Myog* promoter through an association with the Six4 transcriptional activator [35] (see Figure 3). The presence of UTX ensures an active removal of any H3K27me3 present prior to gene activation, or caused by the continued presence of Ezh1-containing PRC2 complexes at the *Myog* promoter after expression has been initiated. Thus, multiple mechanisms appear to be working together to mediate efficient removal of the transcriptionally repressive H3K27me3 mark.

Once the PcG mediated H3K27me3 mark is removed, the TrxG protein containing Ash2L/MLL2 methyltransferase complex targets the *Myog* promoter to establish the transcriptionally permissive H3K4me3 mark that permits high levels of gene expression [47]. Importantly, the recruitment of the Ash2L/MLL2 methyltransferase complex to the *Myog* promoter is mediated by Mef2d in a process that is dependent upon activation of the p38 MAPK signaling cascade [47]. Indeed, the blocking of the p38 signaling in differentiating myotubes leads to the formation of a transcriptionally poised promoter (containing the RNA Pol II, p300, and acetylated histones) though no transcription is observed [47]. While this TrxG-mediated methylation of H3K4 is crucial for high-level expression of Myog, recent studies have shown that the PcG protein Ezh1 must also be present at the *Myog* promoter for transcription to occur [64]. This paper demonstrated that Ezh1 binding at the *Myog* promoter facilitates the recruitment of RNA Pol II to mediate transcription of *Myog* and suggests a previously unappreciated role for PcG proteins in the activation of gene expression. Thus, the *Myog* promoter appears to be regulated by Polycomb and Trithorax group proteins through both antagonistic and synergistic modes of action.

Once activated, the expression of Myog permits the differentiating myoblasts to undergo terminal myogenesis and fuse to form myofibers. In the mature myofiber, Myog expression is eventually downregulated. The mechanisms that lead to this downregulation are poorly understood. However, studies have implicated the proteins MSY3 [20] and the Ezh1-containing PRC2 complex [36] in this process. Future studies which focus on the complement of proteins bound at the *Myog* locus in the late stages of myofiber formation should provide insight into the mechanism by which this muscle-specific gene is downregulated to maintain its strict spatial and temporal expression pattern.

5. Conclusions

The control of *Myog* gene expression during myogenesis has become an important paradigm for understanding mechanisms that drive tissue-specific gene expression. While many genes that display specific temporal and spatial patterns of gene expression require regulatory regions (enhancers) that lie far outside their proximal promoter, we expect that many of the principles that modulate their transcription to be common to those elucidated on the *Myog* promoter. In particular, we highlight the fact that tissue-specific factors such as MyoD must cooperate with more ubiquitously expressed proteins (Six4, Mef2, and Pbx1) to establish a transcriptionally permissive environment within the gene. Though the MyoD-Six1/4-Mef2-Pbx1 axis of transcription factors is important to the proper developmental expression of Myog, it is well established that additional combinations of transcription factors work synergistically in the activation of other muscle-specific genes. With the recent advance in high-throughput analysis of transcription factor binding and chromatin structure analysis, we expect that evolving transcriptional network models will provide us with important new insight into the many other axes of transcription factors that modulate expression of the specific genes that make up the muscle-specific gene expression program.

Acknowledgments

The authors would like to thank Dr. Arif Aziz for helpful discussion and critical reading of the paper. Work in the Dilworth lab is supported by grants from the Canadian Institutes of Health Research. F. J. Dilworth holds a Canada Research Chair in Epigenetic Regulation of Transcription.

References

[1] H. W. Brock and C. L. Fisher, "Maintenance of gene expression patterns," *Developmental Dynamics*, vol. 232, no. 3, pp. 633–655, 2005.

[2] D. Sassoon, G. Lyons, W. E. Wright et al., "Expression of two myogenic regulatory factors myogenin and MyoD1 during mouse embryogenesis," *Nature*, vol. 341, no. 6240, pp. 303–307, 1989.

[3] M. Buckingham and D. Montarras, "Skeletal muscle stem cells," *Current Opinion in Genetics and Development*, vol. 18, no. 4, pp. 330–336, 2008.

[4] P. Hasty, A. Bradley, J. H. Morris et al., "Muscle deficiency and neonatal death in mice with a targeted mutation in the myogenin gene," *Nature*, vol. 364, no. 6437, pp. 501–506, 1993.

[5] J. M. Venuti, J. H. Morris, J. L. Vivian, E. N. Olson, and W. H. Klein, "Myogenin is required for late but not early aspects of myogenesis during mouse development," *Journal of Cell Biology*, vol. 128, no. 4, pp. 563–576, 1995.

[6] M. G. Cusella-De Angelis, G. Lyons, C. Sonnino et al., "MyoD, myogenin independent differentiation of primordial myoblasts in mouse somites," *Journal of Cell Biology*, vol. 116, no. 5, pp. 1243–1255, 1992.

[7] P. Zhao, S. Iezzi, E. Carver et al., "Slug is a novel downstream target of MyoD. Temporal profiling in muscle regeneration," *Journal of Biological Chemistry*, vol. 277, no. 33, pp. 30091–30101, 2002.

[8] E. Meadows, J. M. Flynn, and W. H. Klein, "Myogenin regulates exercise capacity but is dispensable for skeletal muscle regeneration in adult mdx mice," *PLoS ONE*, vol. 6, no. 1, Article ID e16184, 2011.

[9] D. G. Edmondson, T. C. Cheng, P. Cserjesi, T. Chakraborty, and E. N. Olson, "Analysis of the myogenin promoter reveals an indirect pathway for positive autoregulation mediated by the muscle-specific enhancer factor MEF-2," *Molecular and Cellular Biology*, vol. 12, no. 9, pp. 3665–3677, 1992.

[10] S. P. Yee and P. W. J. Rigby, "The regulation of myogenin gene expression during the embryonic development of the mouse," *Genes and Development A*, vol. 7, no. 7, pp. 1277–1289, 1993.

[11] O. Ram, A. Goren, I. Amit et al., "Combinatorial patterning of chromatin regulators uncovered by genome-wide location analysis in human cells," *Cell*, vol. 147, pp. 1628–1639, 2011.

[12] P. Asp, R. Blum, V. Vethantham et al., "Genome-wide remodeling of the epigenetic landscape during myogenic differentiation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 22, pp. E149–E158, 2011.

[13] Y. Cao, Z. Yao, D. Sarkar et al., "Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming," *Developmental Cell*, vol. 18, no. 4, pp. 662–674, 2010.

[14] C. H. Washabaugh, M. P. Ontell, S. H. Shand, N. Bradbury, J. A. Kant, and M. Ontell, "Neuronal control of myogenic regulatory factor accumulation in fetal muscle," *Developmental Dynamics*, vol. 236, no. 3, pp. 732–745, 2007.

[15] F. Spitz, J. Demignon, A. Porteu et al., "Expression of myogenin during embryogenesis is controlled by six/sine oculis homeoproteins through a conserved MEF3 binding site," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 95, no. 24, pp. 14220–14225, 1998.

[16] C. A. Berkes, D. A. Bergstrom, B. H. Penn, K. J. Seaver, P. S. Knoepfler, and S. J. Tapscott, "Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential," *Molecular Cell*, vol. 14, no. 4, pp. 465–477, 2004.

[17] M. D. E. Deato and R. Tjian, "Switching of the core transcription machinery during myogenesis," *Genes and Development*, vol. 21, no. 17, pp. 2137–2149, 2007.

[18] T. C. Cheng, M. C. Wallace, J. P. Merlie, and E. N. Olson, "Separable regulatory elements governing myogenin transcription in mouse embryogenesis," *Science*, vol. 261, no. 5118, pp. 215–218, 1993.

[19] C. Laclef, G. Hamard, J. Demignon, E. Souil, C. Houbron, and P. Maire, "Altered myogenesis in Six1-deficient mice," *Development*, vol. 130, no. 10, pp. 2239–2252, 2003.

[20] L. Berghella, L. De Angelis, T. De Buysscher et al., "A highly conserved molecular switch binds MSY-3 to regulate myogenin repression in postnatal muscle," *Genes and Development*, vol. 22, no. 15, pp. 2125–2138, 2008.

[21] J. Yao, R. D. Fetter, P. Hu, E. Betzig, and R. Tjian, "Subnuclear segregation of genes and core promoter factors in myogenesis," *Genes and Development*, vol. 25, no. 6, pp. 569–580, 2011.

[22] A. Fuso, G. Ferraguti, F. Grandoni et al., "Early demethylation of non-CpG, CpC-rich, elements in the myogenin 5′-flanking region: a priming effect on the spreading of active demethylation?" *Cell Cycle*, vol. 9, no. 19, pp. 3965–3976, 2010.

[23] D. Palacios, D. Summerbell, P. W. J. Rigby, and J. Boyes, "Interplay between DNA methylation and transcription factor availability: implications for developmental activation of the mouse Myogenin gene," *Molecular and Cellular Biology*, vol. 30, no. 15, pp. 3805–3815, 2010.

[24] A. M. Deaton and A. Bird, "CpG islands and the regulation of transcription," *Genes and Development*, vol. 25, no. 10, pp. 1010–1022, 2011.

[25] Y. Oikawa, R. Omori, T. Nishii, Y. Ishida, M. Kawaichi, and E. Matsuda, "The methyl-CpG-binding protein CIBZ suppresses myogenic differentiation by directly inhibiting myogenin expression," *Cell Research*, vol. 21, pp. 1578–1590, 2011.

[26] Y. Liu, A. Chu, I. Chakroun, U. Islam, and A. Blais, "Cooperation between myogenic regulatory factors and SIX family transcription factors is important for myoblast differentiation," *Nucleic Acids Research*, vol. 38, no. 20, pp. 6857–6871, 2010.

[27] C. L. Zhang, T. A. McKinsey, and E. N. Olson, "Association of class II histone deacetylases with heterochromatin protein 1: potential role for histone methylation in control of muscle differentiation," *Molecular and Cellular Biology*, vol. 22, no. 20, pp. 7302–7312, 2002.

[28] A. Mal and M. L. Harter, "MyoD is functionally linked to the silencing of a muscle-specific regulatory gene prior to skeletal myogenesis," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 4, pp. 1735–1739, 2003.

[29] A. K. Mal, "Histone methyltransferase Suv39h1 represses MyoD-stimulated myogenic differentiation," *The EMBO Journal*, vol. 25, no. 14, pp. 3323–3334, 2006.

[30] M. A. Gillespie, F. Le Grand, A. Scimè et al., "p38-γ-dependent gene silencing restricts entry into the myogenic differentiation program," *Journal of Cell Biology*, vol. 187, no. 7, pp. 991–1005, 2009.

[31] B. M. Ling, N. Bharathy, T. K. Chung et al., "Lysine methyltransferase G9a methylates the transcription factor MyoD and regulates skeletal muscle differentiation," *Proceedings of the National Academy of Sciences of the United State*, vol. 109, pp. 841–846, 2012.

[32] V. Sartorelli, P. L. Puri, Y. Hamamori et al., "Acetylation of MyoD directed by PCAF is necessary for the execution of the muscle program," *Molecular Cell*, vol. 4, no. 5, pp. 725–734, 1999.

[33] F. J. Dilworth, K. J. Seaver, A. L. Fishburn, S. L. Htet, and S. J. Tapscott, "In vitro transcription system delineates the distinct roles of the coactivators pCAF and p300 during MyoD/E47-dependent transactivation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 32, pp. 11593–11598, 2004.

[34] R. Cao, L. Wang, H. Wang et al., "Role of histone H3 lysine 27 methylation in polycomb-group silencing," *Science*, vol. 298, no. 5595, pp. 1039–1043, 2002.

[35] S. Seenundun, S. Rampalli, Q. C. Liu et al., "UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis," *The EMBO Journal*, vol. 29, no. 8, pp. 1401–1411, 2010.

[36] L. Stojic, Z. Jasencakova, C. Prezioso et al., "Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells," *Epigenetics Chromatin*, vol. 4, article 16, 2011.

[37] A. Mal, M. Sturniolo, R. L. Schiltz, M. K. Ghosh, and M. L. Harter, "A role for histone deacetylase HDAC1 in modulating the transcriptional activity of MyoD: inhibition of the myogenic program," *The EMBO Journal*, vol. 20, no. 7, pp. 1739–1753, 2001.

[38] S. Iezzi, M. Di Padova, C. Serra et al., "Deacetylase inhibitors increase muscle cell size by promoting myoblast recruitment and fusion through induction of follistatin," *Developmental Cell*, vol. 6, no. 5, pp. 673–684, 2004.

[39] J. Lu, T. A. McKinsey, C. L. Zhang, and E. N. Olson, "Regulation of skeletal myogenesis by association of the MEF2 transcription factor with class II histone deacetylases," *Molecular Cell*, vol. 6, no. 2, pp. 233–244, 2000.

[40] A. Méjat, F. Ramond, R. Bassel-Duby, S. Khochbin, E. N. Olson, and L. Schaeffer, "Histone deacetylase 9 couples neuronal activity to muscle chromatin acetylation and gene expression," *Nature Neuroscience*, vol. 8, no. 3, pp. 313–321, 2005.

[41] H. Tang, P. Macpherson, M. Marvin et al., "A histone deacetylase 4/myogenin positive feedback loop coordinates denervation-dependent gene induction and suppression," *Molecular Biology of the Cell*, vol. 20, no. 4, pp. 1120–1131, 2009.

[42] H. Faralli, E. Martin, N. Core et al., "Teashirt-3, a novel regulator of muscle differentiation, associates with BRG1-associated factor 57 (BAF57) to inhibit myogenin gene expression," *Journal of Biological Chemistry*, vol. 286, no. 26, pp. 23498–23510, 2011.

[43] S. V. Forcales, S. Albini, L. Giordani et al., "Signal-dependent incorporation of MyoD-BAF60c into Brg1-based SWI/SNF chromatin-remodelling complex," *The EMBO Journal*, vol. 31, pp. 301–316, 2011.

[44] A. Zetser, E. Gredinger, and E. Bengal, "p38 mitogen-activated protein kinase pathway promotes skeletal muscle differentiation: participation of the MEF2C transcription factor," *Journal of Biological Chemistry*, vol. 274, no. 8, pp. 5193–5200, 1999.

[45] E. Perdiguero, V. Ruiz-Bonilla, L. Gresh et al., "Genetic analysis of p38 MAP kinases in myogenesis: fundamental role of p38α in abrogating myoblast proliferation," *The EMBO Journal*, vol. 26, no. 5, pp. 1245–1256, 2007.

[46] F. Lluís, E. Ballestar, M. Suelves, M. Esteller, and P. Muñoz-Cánoves, "E47 phosphorylation by p38 MAPK promotes MyoD/E47 association and muscle-specific gene transcription," *The EMBO Journal*, vol. 24, no. 5, pp. 974–984, 2005.

[47] S. Rampalli, L. Li, E. Mak et al., "p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation," *Nature Structural and Molecular Biology*, vol. 14, no. 12, pp. 1150–1156, 2007.

[48] B. B. Friday, P. O. Mitchell, K. M. Kegley, and G. K. Pavlath, "Calcineurin initiates skeletal muscle differentiation by activating MEF2 and MyoD," *Differentiation*, vol. 71, no. 3, pp. 217–227, 2003.

[49] C. Serra, D. Palacios, C. Mozzetta et al., "Functional interdependence at the chromatin level between the MKK6/p38 and IGF1/PI3K/AKT pathways during muscle differentiation," *Molecular Cell*, vol. 28, no. 2, pp. 200–213, 2007.

[50] G. Caretti, M. Di Padova, B. Micales, G. E. Lyons, and V. Sartorelli, "The polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation," *Genes and Development*, vol. 19, no. 6, pp. 2627–2638, 2005.

[51] P. L. Puri, V. Sartorelli, X. J. Yang et al., "Differential roles of p300 and PCAF acetyltransferases in muscle differentiation," *Molecular Cell*, vol. 1, no. 1, pp. 35–45, 1997.

[52] J. W. Kim, S. M. Jang, C. H. Kim et al., "Tip60 regulates myoblast differentiation by enhancing the transcriptional activity of MyoD via their physical interactions," *The FEBS Journal*, vol. 278, pp. 4394–4404, 2011.

[53] J. Choi, H. Jang, H. Kim, S. T. Kim, E. J. Cho, and H. D. Youn, "Histone demethylase LSD1 is required to induce skeletal muscle differentiation by regulating myogenic factors," *Biochemical and Biophysical Research Communications*, vol. 401, no. 3, pp. 327–332, 2010.

[54] L. Verrier, F. Escaffit, C. Chailleux, D. Trouche, and M. Vandromme, "A new isoform of the histone demethylase JMJD2A/KDM4A is required for skeletal muscle differentiation," *PLoS Genetics*, vol. 7, no. 6, Article ID e1001390, 2011.

[55] Y. Tao, R. L. Neppl, Z. P. Huang et al., "The histone methyltransferase Set7/9 promotes myoblast differentiation and myofibril assembly," *The Journal of Cell Biology*, vol. 194, pp. 551–565, 2011.

[56] H. Wang, R. Cao, L. Xia et al., "Purification and functional characterization of a histone H3-lysine 4-specific methyltransferase," *Molecular Cell*, vol. 8, no. 6, pp. 1207–1217, 2001.

[57] C. S. Dacwag, Y. Ohkawa, S. Pal, S. Sif, and A. N. Imbalzano, "The protein arginine methyltransferase Prmt5 is required for myogenesis because it facilitates ATP-dependent chromatin remodeling," *Molecular and Cellular Biology*, vol. 27, no. 1, pp. 384–394, 2007.

[58] I. L. de La Serna, Y. Ohkawa, C. A. Berkes et al., "MyoD targets chromatin remodeling complexes to the myogenin locus prior to forming a stable DNA-bound complex," *Molecular and Cellular Biology*, vol. 25, no. 10, pp. 3997–4009, 2005.

[59] M. D. E. Deato, M. T. Marr, T. Sottero, C. Inouye, P. Hu, and R. Tjian, "MyoD targets TAF3/TRF3 to activate myogenin transcription," *Molecular Cell*, vol. 32, no. 1, pp. 96–105, 2008.

[60] H. Heller and E. Bengal, "TFIID (TBP) stabilizes the binding of MyoD to its DNA site at the promoter and MyoD facilitates the association of TFIIB with the preinitiation complex," *Nucleic Acids Research*, vol. 26, no. 9, pp. 2112–2119, 1998.

[61] A. Aziz, Q. C. Liu, and F. J. Dilworth, "Regulating a master regulator: establishing tissue-specific gene expression in skeletal muscle," *Epigenetics*, vol. 5, no. 8, pp. 691–695, 2010.

[62] J. H. Yang, Y. Song, J. H. Seol et al., "Myogenic transcriptional activation of MyoD mediated by replication-independent histone deposition," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 1, pp. 85–90, 2011.

[63] R. Margueron, G. Li, K. Sarma et al., "Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms," *Molecular Cell*, vol. 32, no. 4, pp. 503–518, 2008.

[64] K. Mousavi, H. Zare, A. H. Wang, and V. Sartorelli, "Polycomb protein Ezh1 promotes RNA polymerase II elongation," *Molecular Cell*, vol. 45, pp. 255–262, 2012.

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 256892, 12 pages
doi:10.1155/2012/256892

Review Article

Epigenetic Alterations in Muscular Disorders

Chiara Lanzuolo

CNR Institute of Cellular Biology and Neurobiology, IRCCS Santa Lucia Foundation, Via Del Fosso di Fiorano 64, 00143 Rome, Italy

Correspondence should be addressed to Chiara Lanzuolo, chiara.lanzuolo@inmm.cnr.it

Received 2 February 2012; Revised 11 April 2012; Accepted 19 April 2012

Academic Editor: Daniela Palacios

Epigenetic mechanisms, acting via chromatin organization, fix in time and space different transcriptional programs and contribute to the quality, stability, and heritability of cell-specific transcription programs. In the last years, great advances have been made in our understanding of mechanisms by which this occurs in normal subjects. However, only a small part of the complete picture has been revealed. Abnormal gene expression patterns are often implicated in the development of different diseases, and thus epigenetic studies from patients promise to fill an important lack of knowledge, deciphering aberrant molecular mechanisms at the basis of pathogenesis and diseases progression. The identification of epigenetic modifications that could be used as targets for therapeutic interventions could be particularly timely in the light of pharmacologically reversion of pathological perturbations, avoiding changes in DNA sequences. Here I discuss the available information on epigenetic mechanisms that, altered in neuromuscular disorders, could contribute to the progression of the disease.

1. Introduction

Although every cell within our body bears the same genetic information, only a small subset of genes is transcribed in a given cell at a given time. The distinct gene expression of genetically identical cells is responsible for cell phenotype and depends on the epigenome, which involve all structural levels of chromosome organization from DNA methylation and histone modifications up to nuclear compartmentalization of chromatin [1–5]. Enormous progress over the last few years in the field of epigenetic regulation indicated that the primary, monodimensional structure of genetic information is insufficient for a complete understanding of how the networking among regulatory regions actually works. The contribution of additional coding levels hidden in the three-dimensional structure of the chromosome and nuclear structures appears to be a fundamental aspect for the control of the quality and stability of genetic programs. Damage or perturbation of epigenetic components may lead to deviations from a determined cellular program, resulting in severe developmental disorders and tumour progression [6, 7]. Moreover, for human complex diseases, the phenotypic differences and the severity of the disease observed among patients could be attributable to inter-individual epigenomic variation. Unravelling the intricacies of the epigenome will be a complex process due to the enormity and dynamic nature of the epigenomic landscape but is essential to gain insights into the aetiology of complex diseases.

2. The Complexity of the Epigenome

The epigenome consists of multiple mechanisms of transcriptional regulation that establish distinct layers of genome organization and includes covalent modification of DNA and histones, packaging of DNA around nucleosomes, higher-order chromatin interactions, and nuclear positioning [4]. The first layer of epigenetic control is the DNA methylation, an heritable epigenetic mark typically associated with a repressed chromatin state [8], which seems to play a role, together with other histone modifications, in preventing gene reactivation [9]. Vertebrate genomes are predominantly methylated at cytosine of the dinucleotide sequence CpG (for a review see [3]). Despite the high level of CpG methylation, some regions of mammalian genomes are refractory to this modification [10]. These regions, called CpG islands, contain high levels of CpG dinucleotides [11] and localize at or

near gene promoters [12], suggesting a strong correlation between differential methylation of CpG islands and flanking promoter activity. From the mechanistic point of view, DNA methylation can inhibit gene expression by blocking the access of transcriptional activators to their binding site on DNA or by recruiting chromatin modifying activities to DNA (for a review see [3]). For long time, DNA methylation was considered as a stable epigenetic mark. However, recently it has been shown that methylated cytosines could be converted to 5-hydroxymethylcytosines (5hmeC) by Tet (Ten eleven Translocation) family proteins [13–15] and the generation of 5hmeC is a necessary intermediate step preceding active demethylation of DNA [16]. The second level of epigenetic regulation occurs through posttranslational histone modification. Histone proteins assemble into a complex that associates with DNA forming the elementary unit of chromatin packaging: the nucleosome. The amino and carboxy termini of the histones (histone tails), protruding from the nucleosome, play an essential role in controlling gene expression, being the target for posttranscriptional modifications, including acetylation, methylation, phosphorylation, ubiquitylation, biotinylation, and several others (for a review see [2, 17]). Multiple histone modifications can also coexist on the same tail, dictating specific biological readouts [18–24]. In addition to histone modifications, a fraction of chromatin contains one or more variant isoforms of the canonical histones that can be incorporated into specific regions of the genome throughout the cell cycle and are essential for the epigenetic control of gene expression and other cellular responses (for a review see [25]). Combinatorial histone modifications and variants play an important role in folding nucleosomal arrays into higher-order chromatin structures, creating local structural and functional diversity and delimiting chromatin subdomains then subjected to a specific protein environment. Chromatin higher-order structures established at DNA level give signals that are recognized by specific binding proteins that in turn influence gene expression and other chromatin functions [1, 26]. This represents an additional layer of epigenetic gene regulation and includes factors, such as transcriptional repressors or activators, that recognizing specific chromatin patterns regulate the folding or modulate the activity of RNA Polymerase II (Pol II).

The topological organization of chromatin and the association of regulatory elements with specific components of the eukaryotic nucleus is another parameter to be considered in the complexity of the epigenetic information. It is now clear that specific chromosomal conformations, mediated by cis-trans interactions, are associated with distinct transcriptional states in many organisms, allowing the establishment of chromatin boundaries between promoters and regulatory element (for a review see [27, 28]). The nuclear localization also influences gene expression, regulating its access to specific machinery responsible for specific functions, such as transcription or replication [29, 30]. In addition, due to its highly dynamic nature, the genome moves in the nucleus driving specific genomic regions toward nuclear compartments defined by a high concentration of specific factors and substrates that facilitate more efficient biological reactions [31]. This constant motion plays also a role in coordinating the expression of coregulated genes, separated by longer chromosomal regions or located on different chromosomes [32].

The evolutionarily conserved Polycomb group of proteins (PcG) are multiprotein complexes that play a central role during development [1]. The most characterized PcG-encoded protein complexes are Polycomb Repressive Complex 1 (PRC1) and 2 (PRC2). Three other complexes were characterized in *Drosophila*, PHO-repressive complex (PhoRC), dRing-associated factors (dRAF) complex, and Polycomb repressive deubiquitinase (PR-DUB), and their components have orthologues in mammals [33, 34]. PcG complexes mediate gene silencing by regulating different levels of chromatin structures. Biochemical studies revealed that Enhancer of zeste 2 (EZH2), the Histone Methyl Transferase (HMTase) subunit of PRC2, marks lysine 27 of histone H3 [35–38] and PRC1 complex monoubiquitylates Lys 119 of histone H2A [39]. Moreover, the H3K27me3 mark constitutes a docking site for the chromodomain present in PRC1 components [35], determining a sequential recruitment of PRC complexes, although recent chromatin profiling studies evidenced that PRC1 and PRC2 also have targets independent of each other [40, 41]. Examination of the localization of PcG proteins in the nucleus has revealed that they are organized into distinct domains called Polycomb or PcG bodies, which are often localized, closed to pericentric heterochromatin [42]. PcG targets are frequently localized in PcG bodies in the tissue where they are repressed, suggesting that such nuclear localization may be required for efficient silencing [43, 44]. However, the number of PcG bodies is less than the number of PcG target genes, implying that several PcG targets share the same body. FISH studies together with Chromosome Conformation Capture (3C) analysis have confirmed this coassociation [43–46] and revealed that PcG-dependent higher-order structures organization is conserved in mammals [47–49]. The characteristic feature of the PcG memory system is inheritability of gene expression patterns throughout the cell cycle, ensured by the PcG capability to bind its own methylation mark [50, 51] and specific cell cycle-dependent dynamics [52–55]. Besides their extensively described role in development, in the last years emerging evidence has shown PcG involvement in several other biological processes, such as X chromosome inactivation, differentiation, and reprogramming (reviewed in [56–58]). The highly variability of PcG functions and the fine quantitative and qualitative tuning of their activities is generated by the association of different PcG proteins and their coregulators in a combinatorial fashion and/or by the regulation of their recruitment at specific chromatin sites (reviewed in [1, 59, 60]). One recent example is a genomewide study of TET complex localization, in murine Embryonic Stem (ES) cells. This complex, responsible for 5hmeC generation, colocalizes with a subpopulation of Polycomb-repressed genes, contributing to gene transcription control [61–63].

3. Muscle Diseases

Skeletal muscles are composed by multiple aligned multinucleated cells, the myofibers, wrapped in a plasma membrane

called sarcolemma. Inside the sarcolemma and all around the myofibers, there is a specialized cytoplasm, the sarcoplasm, that contains the usual subcellular elements [64]. A plethora of structural molecules and cellular proteins connecting all the fibers components together with specialized signalling pathways and transcription factors are required for a correct muscle formation and function. Dysfunction or lack of any component of the skeletal muscle could lead to a muscular disorder, the muscular dystrophy (MD), clinically characterized by muscle weakness and skeletal muscle degeneration [64]. Some dystrophies arise from mutations of molecules that play a role outside the nucleus while other dystrophies derive from dysfunction of the nucleus or its membrane. The nonnuclear dystrophies include Duchenne MD (DMD), Becker MD (BMD), and all MD affecting proteins working in the sarcoplasm. DMD is the most severe form of muscular dystrophy and is caused by mutations that preclude the production of the essential cytoskeletal muscle protein dystrophin, which anchors proteins from the internal cytoskeleton to a complex of proteins (dystrophin-associated protein complex, DAPC) on the membrane of muscle fibers [65]. This interaction is important for the structural stabilization of the sarcolemma [66]. Interestingly, recent reports highlighted the influence of epigenetic mechanisms regulating histone deacetylation (HDAC) pathways in the development of this disease [67–69] and the reversion of some DMD-associated phenotypes in presence of inhibitors of HDACs [70, 71].

The nuclear dystrophies include all MD generated by a dysfunction of nuclear membrane (laminopathies) or by expansion or contraction of nucleotide repeats, not necessarily contained in a coding region, which affect nuclear function. Myotonic dystrophy is the most common MD in adult and is a complex multisystemic inherited muscle degenerative disorder caused by a pathogenic expansion of microsatellite repeats within noncoding elements of dystrophia myotonica protein kinase (*DMPK*) or zinc finger protein 9 (*ZNF9*) genes [72]. These expansions, although transcribed into RNA, do not affect the protein-coding region of any other gene. However, it has been shown that transcripts accumulate in the nucleus and interfere with protein families that regulate alternative splicing during development [64, 73]. In this paper, I will describe the contribution of epigenetic mechanisms mediated by Polycomb group of proteins to two human nuclear muscular dystrophies, facioscapulohumeral muscular dystrophy (FSHD), and laminopathies.

4. Polycomb Group of Protein as Epigenetic Regulators of Muscle Differentiation

PcG proteins regulate large numbers of target genes, primarily those involved in differentiation and development [74–80]. During cell differentiation the progressive restriction of the developmental potential and increased structural and functional specialization of cells ensure the formation of tissues and organs [57]. Myogenesis is a multistep process that starts with the commitment of multipotent mesodermal precursor cells. Upon appropriate stimuli these cells differentiate and fuse into multinucleated myotubes, giving rise to the myofibers. In mammals, PcG proteins are primarily involved in muscle differentiation by binding and repressing muscle-specific gene regulatory regions in undifferentiated myoblasts to prevent premature transcription. During myogenesis progression, PcG binding and H3K27me3 are lost at muscle-specific loci, resulting in appropriate muscle gene expression [81–84]. Interestingly, artificial modulation of EZH2 levels, either by depletion or overexpression, consistently affects normal muscle differentiation, accelerating or delaying, respectively, muscle cell fate determination [82, 83, 85]. Although emerging evidence suggested a key role for epigenetic mechanisms in muscular diseases [68, 71, 86–89], the precise contribution of Polycomb proteins to the pathology and progression remains largely unexplored.

5. Facioscapulohumeral Muscular Dystrophy (FSHD)

Facioscapulohumeral muscular dystrophy (FSHD) is a frequent (1 : 15.000) dominant autosomal miopathy that is characterized by progressive, often asymmetric weakness and wasting of facial (facio), shoulder, and upper arm (scapulohumeral) muscles [90]. Monozygotic twins with different penetrance of FSHD have been described, suggesting a strong epigenetic contribution to the pathology [91, 92]. Genetically, FSHD1, one of the two forms of FSHD, is caused by a contraction of the highly polymorphic D4Z4 macrosatellite repeat in chromosome 4q [93]. In the general population, this repeat array varies between 11 and 100 units of 3,3 kb each, ordered head to tail [94]. Most patients with FSHD1 present a partial deletion of the D4Z4 array, which leaves 1–10 units on the affected allele [93]. Although a linear negative correlation between repeat size and clinical severity has not been observed, some findings indicated that smaller D4Z4 arrays result in earlier disease onset and enhanced severity in patients [95–97]. Interestingly, at least one D4Z4 unit is necessary to develop FSHD, as monosomy of 4q does not cause the disease [98]. In addition to polymorphism associated with D4Z4 repeat number, two allelic variants of the 4q subtelomere, termed 4qA and 4qB, have been identified. These variants differ for the presence of a β satellite repeat immediately distal to the D4Z4 array on 4qA allele [99]. Whereas 4qA and 4qB chromosomes are almost equally common in the population, FSHD arises mainly from 4qA haplotype [99–102]. D4Z4 repeat arrays are not restricted to chromosome 4q, but homologous sequences have been identified on many chromosomes [103]. In particular, the subtelomere of chromosome 10q is almost identical to the region in 4q containing D4Z4 repeats, containing highly homologous and equally polymorphic repeat arrays [104, 105]. However, chromosome 10 with less than 11 repeat units does not cause FSHD1 [106], suggesting that the chromatin environment associated with chromosome 4q and/or 4q-specific DNA sequences could contribute to FSHD development. In agreement with this observation, the relatively gene-poor region flanking D4Z4

repeats on chromosome 4q contains two attractive candidates that have been characterized for their contribution to disease development: FRG1 (FSHD Region Gene 1) and the double-homeobox transcription factor DUX4. FRG1 is highly conserved in both vertebrates and invertebrates and it has been found overexpressed in some FSHD samples [107, 108]. Moreover, transgenic mice overexpressing FRG1 develop, selectively in the skeletal muscle, pathologies with physiological, histological, ultrastructural, and molecular features that mimic human FSHD [109]. However, FRG1 overexpression in FSHD samples is not a uniform finding [110, 111] and thus the contribution of the FRG1 gene to the FSHD phenotype needs further validation. Although some evidence suggests a role for FRG1 in pre-mRNA splicing [109, 112, 113], to date the mechanism of action and the role of FRG1 in FSHD onset and development is largely unknown.

Aberrant production of DUX4, the gene present in the D4Z4 array, was detected in both FSHD1 and FSHD2 muscle biopses [114], suggesting that D4Z4 could affect disease progression [115]. However, D4Z4 array has a complex transcriptional profile that includes sense and antisense transcripts and RNA processing [116]. The DUX4 mRNA is generated by transcription of the last, most distal, unit of the array, including a region named pLAM, which contains a polyadenylation signal, necessary for DUX4 transcript stabilization [115]. The absence of this polyadenylation signal on chromosome 10 suggests its involvement in FSHD development [100]. The DUX4 pre-mRNA can be alternatively spliced [116] and there has been found a DUX4 mRNA isoform encoding for the full-length protein, expressed in FSHD muscle, whereas healthy subjects present an alternative splicing mRNA encoding for a truncated protein [114].

DUX4 RNA and protein levels have been arguments of debate in the field for several years. Previous works demonstrated a proapoptotic function for DUX4 [117] and DUX4 overexpression was found to have dramatically toxic effect on cell growth [118]. On the other hand extremely low levels of DUX4 were found in FSHD muscles raising some doubts on the role of this gene in FSHD development [114, 119]. In a recent report, overexpression of DUX4 mRNA in human primary myoblasts followed by gene expression analysis showed deregulation of several genes involved in RNA splicing and processing, immune response pathways, and gametogenesis [119]. These genes were found aberrantly expressed in both FSHD1 and FSHD2 muscles while a partial recovery of the repressed state occurs upon depletion of endogenous DUX4 mRNA. Although no direct evidence was presented about the role of deregulated genes in the FSHD development, these findings suggest that critical DUX4 protein and RNA levels could be responsible for gene transcription deregulation in FSHD [119].

Aside putative genes involved in the FSHD development there is a general consensus in the field in supporting the view that epigenetic mechanisms are important players in FSHD, affecting the severity of the disease, its rate in progression, and the distribution of muscle weakness [120, 121]. Increasing evidence suggested that, in patients, chromatin conformation of FSHD locus is altered at multiple levels, from DNA methylation up to higher-order chromosome structures, resulting in perturbation of heterochromatic gene silencing in the subtelomeric domain of the long arm of chromosome 4. As stated previously, DNA methylation is associated with gene silencing and defects in methylation are generally associated with deregulation of transcriptional programs and disease [6]. D4Z4 is overall very GC-rich, having characteristics of CpG islands [122], and in healthy subjects is methylated, while contracted D4Z4 is always associated with an hypomethylation [123, 124]. Interestingly, FSHD2 patients, which phenotypically show FSHD though lacking D4Z4 contractions, display general D4Z4 hypomethylation [123], indicating an important epigenetic condition necessary to develop or generate the disease.

Combination of posttranslation histone modifications establishes a specific code that recruits nuclear factors responsible of several functions such as transcriptional or replication control. The D4Z4 repeat array is enriched of two repressive marks: trimethylation of lysine 9 or 27 of histone H3 (H3K9me3 and H3K27me3, resp.). The first, generally associated with constitutive heterochromatin, is deposited by the histone methyltransferase SUV39 and is responsible for HP1 repressor recruitment [125]. H3K27me3 is characteristic of facultative heterochromatin, is deposited by the PRC2 subunit EZH2, and in turn recruits PRC1 and PRC2 to establish transcriptionally repressed domains. It has been shown that H3K27me3 and the two Polycomb proteins YY1 and EZH2 are bound to D4Z4 and FRG1 promoter in myoblasts [107, 108] and are reduced during myogenic differentiation [108]. Interestingly, DNA association studies, by using 3C technologies [126], revealed that D4Z4 physically interacts with FRG1 promoter and this DNA loop is reduced upon differentiation. These epigenetic signatures dynamics during myogenesis are accompanied by a gradual upregulation of FRG1 [108]. Conversely, in FSHD1 myoblasts the D4Z4-FRG1 promoter interaction is reduced and FRG1 expression is anticipated during differentiation, suggesting an alteration of epigenetic signatures dynamics occurring when the differentiation starts. Notably, H3K27me3 can still be detected by ChIP at D4Z4 repeats in FSHD1 myoblasts, although by 3D immuno-FISH it was found specifically reduced on D4Z4 on 4q chromosome in FSHD1 myoblasts compared with controls [108]. This apparent inconsistency is justified by the extensive duplication of D4Z4 sequences in the human genome and the limitation of ChIP assay to distinguish specific 4q D4Z4 repeat. In addition to the complex heterochromatic features found at D4Z4 locus, there has been shown the presence of histone marks associated with transcriptional activation in the first proximal D4Z4 unit of the array, such as acetylation of histone H4 and dimethylation of Lys 4 of histone H3 [110, 125]. This could reflect the complexity of bidirectional transcriptional activity at the locus and could suggest the potential presence of noncoding RNA that further regulate the transcription.

As stated before, epigenetic chromatin regulation depends also on appropriate intranuclear positioning. Most nuclear events do not occur randomly in the nucleoplasm, rather regulatory proteins are spatially clustered in specific

territories, and the position of chromosomal region in the nucleus influences its transcriptional activity. The 4q subtelomere is preferentially localized in the nuclear periphery in both controls and FSHD patients [127, 128], and this localization is evolutionary conserved [129]. In FSHD1 cells, this localization depends on a sequence within D4Z4 unit that tethers the subtelomere in the nuclear periphery in a CTCF and Lamin-A-dependent manner [130]. Although intranuclear positioning of 4q subtelomere does not change during muscle differentiation, when several epigenetic modifications take place [108], it has been shown that the nuclear periphery localization in controls and FSHD1 cells can be directed by different sequences, proximal or within D4Z4 repeat, respectively. This suggests that the nuclear environment of FSHD locus in normal or affected subjects could be different and could contribute to the disease development [130].

In summary, the epigenetic analysis suggests that probably the presence of more than ten D4Z4 repeats provides a physiological heterochromatization and repression of the subtelomeric region, due to the saturating levels of epigenetic repressors. In this view, less than ten D4Z4 repeats could be considered as *border line* genotype, because the correct heterochromatin formation is not ensured, determining a predisposition to the disease and also explaining the high variability in disease severity even in the same genetic background. This hypothesis is reinforced by the evidence that patients with less than 3 repeats have more chances to develop FSHD1 disease and that asymptomatic carriers of D4Z4 deletion are increasingly evident in FSHD [131]. Another complex issue about FSHD is the requirement for at least one D4Z4 repeat for the development of the disease, suggesting a *gain of function* effect, where the presence of an aberrant transcription of coding or noncoding RNA or dysregulated binding of epigenetic factors recruited by the D4Z4 array could be necessary for disease development. Systematic analysis of epigenetic modifications across the entire genome in FSHD1 and FSHD2 patients will be crucial to dissect epigenetic mechanisms acting specifically on D4Z4 locus and involved in FSHD pathogenesis and progression.

6. Laminopathies

The nuclear scaffold (or nuclear matrix) is the network of fibers found inside a cell nucleus. The lamina is the major component of nuclear matrix and is constituted by a complex meshwork of proteins closely associated with the inner nuclear membrane [132]. In vertebrates, lamins have been divided into A and B types, based on sequence homologies. All A-type lamins, A, C, C2, and Δ10, are encoded by alternative splicing of a single gene (LMNA) while two major mammalian B-type lamins, B1 and B2, are encoded by different genes (LMNB1 and LMNB2) [133]. All major lamins terminate with a CAAX-box that is involved in numerous posttranslational modifications including the farnesylation of the cysteine, removal of the-AAX, and carboxymethylation of the cysteine [134]. These modifications are thought to be important for the efficient targeting of the lamins to the inner nuclear membrane [135]. Moreover, Lamin A is further processed by the zinc metalloproteinase, Zmpste24/FACE1, which catalyzes the removal of additional 15 residues from Lamin A C-terminus including the farnesylated and carboxymethylated cysteine [136]. Expression of the A- and B-type lamins is developmentally regulated in mammals, resulting in cell type-specific complements of lamins [137]. In the last years, genome wide studies describing lamin bound chromosomal regions were focalized specifically on B-type [138, 139]. However, it is becoming increasingly evident that A-type lamins are scaffolds for proteins that regulate DNA synthesis, responses to DNA damage, chromatin organization, gene transcription, cell cycle progression, cell differentiation, cancer invasiveness, and epigenetic regulation of chromatin [140–143]. In line with this observation, lamin distribution in the nucleus is type specific, with Lamin B being predominantly present at inner nuclear membrane and Lamin A also present in lower concentrations, throughout the nucleoplasm [144], suggesting, for the latter, a role beyond the maintenance of mechanical stability of the nucleus. Genetic studies confirmed this hypothesis, showing that A- or B-type lamin mutations have different impacts on organisms. Mutations in genes encoding B-type lamins are not frequently connected to diseases in human and Lamin B1 null mice die during early postnatal life with severe defects in their lung and bones [145] while mice lacking Lamin B2 die shortly after birth with severe brain abnormalities. Taken together, these findings indicate that B-type lamins play a structural role in the nucleus essential for cell and tissue function. On the other hand, mice lacking A-type lamins have apparently normal embryonic development [146], but postnatal growth is delayed and they develop abnormalities of cardiac and skeletal muscle. This is in line with studies in human, where a large number of mutations of Lamin A/C (LMNA) were found, causing a wide range of human disorders, including lipodystrophy, neuropathies, autosomal dominant Emery-Dreifuss muscular dystrophy (EDMD), and progeria. The latter includes Hutchinson-Gilford progeria syndrome (HGPS), atypical Werner syndrome, restrictive dermopathy, and mandibuloacral dysplasia type A (MADA) [147]. Collectively, these degenerative disorders with a wide spectrum of clinical phenotypes are known as the laminopathies. To date, despite the identification of several mutations on Lamin A causing these disorders, it is difficult to correlate phenotype to genotype in laminopathies. It is still unclear how specific mutations result in a particular tissue-specific laminopathy phenotype [148] or why a single mutation in Lamin A gene can result in different phenotypes [149]. This suggests an involvement of the individual epigenetic background to the disease. Studies in HGPS cells confirmed this hypothesis finding several epigenetic alterations. In particular there has been shown a decrease of the heterochromatin mark H3K9me3 in pericentric regions and a downregulation of the PRC2 component EZH2, accompanied by a loss of H3K27me3 on the inactive X chromosome (Xi), which leads to some decondensation of the Xi [150]. Notably, it is not clear if observed epigenetic defects are cause or consequence of the irreversible cascade of cellular mecha-

nisms dysfunction accompanying HGPS progression. Most inherited LMNA mutations in humans cause disorders that selectively affect striated muscle, determining decreased levels of A-type lamins. This was confirmed in the Lmna null mice, which develop abnormalities of cardiac and skeletal muscle reminiscent of those seen in human subjects [146]. Remarkably, in humans, decreased lamin A levels observed in some laminopathies could also be dependent on dominant negative effect caused by an aberrant form of Lamin A. Indeed, overexpression in transgenic mice of a human lamin A variant responsible for Emery-Dreifuss muscular dystrophy determines severe heart damage [151]. Several hypotheses have been proposed to explain molecular mechanisms underlying muscular dystrophies caused by lamin A mutations. The current model suggests that the prolonged exposure to mechanical stress of muscle cells determines the tissue-specific degeneration observed in laminopathies. This model takes into consideration only the structural role of lamin A, neglecting its functional role in chromatin organization and gene expression control. There is a growing body of evidence indicating that several signalling pathways, such as pRb, MyoD, Wnt-β catenin, and TGF-β, are altered in laminopathies [152, 153]. The Rb-MyoD-crosstalk is one of the most described pathways altered in laminopathies. MyoD is a master transcription factor of muscle differentiation that activates muscle-specific genes. Its levels are modulated by dephosphorilated pRb, which takes part in the acetylation and expression of MyoD [154]. Lamin A controls Rb levels favouring its dephosphorilation [155]. Thus in the absence of Lamin A the level of hypophosphorilated Rb and consequently the level of MyoD are reduced, determining a defect in muscle cells' differentiation [156]. This was confirmed by a decreased number of MyoD positive nuclei observed in skeletal muscle from laminopathy patients [157]. Given its role in muscle-specific genes regulation, PcG protein could be involved in aberrant gene expression observed in lamin A defective background. Several indirect evidences support this hypothesis indicating a potential crosstalk between PcG proteins and Lamin A. As mentioned previously the nuclear positioning of the PcG-regulated FSHD locus, responsible for the described neuromuscular disorder, is altered in human Lamin A/C null cells [127]. However, while the role of PcG proteins in governing local chromatin higher-order structures was extensively addressed [44, 47, 49], it is still unknown if they also control the chromosomal position in the nucleus and if the peripheral localization of FSHD is dependent on PcG proteins. Recently, it has been suggested that nuclear position of PcG proteins could be crucial for muscle differentiation [158]. In this work, Wang and colleagues have shown that the localization of PRC2 complex at the nuclear periphery is mediated by the myogenic regulator, Msx1, and is required for a correct repression of Msx1 target genes. This localization occurs in myoblasts and is necessary for a proper muscle differentiation [158]. The importance of chromatin architecture dynamics during muscle differentiation was further confirmed by studies performed by Mattout et al. in *C. elegans* [159]. Using ablation of the unique lamin gene in worm they found that lamin is necessary for perinuclear positioning of heterochromatin. Then, to test the physiological relevance of this association in developing animals, they monitored tissue-specific changes in nuclear position of specific genomic regions in worms that express a dominant mutant form of lamin, which mimics the human Emery-Dreifuss muscular dystrophy. They found that in lamin defective background, muscle-specific genes are not able to relocalize from the nuclear periphery to a more internal location and this determines loss of muscle integrity. Although there has been extensively shown the crucial role of Polycomb proteins in mediating nuclear chromatin architecture, to date no evidence supports a direct involvement of PcG in muscle genes relocalization during normal differentiation. Further studies are needed to determine if physiological epigenetic dynamics that ensure a correct myogenesis are altered in lamin defective background and the role of Polycomb proteins in this process.

7. Conclusions

In the last years the study of the epigenome and its role in human disease progression has attracted considerable interest. The insurgence of epigenetic deregulation in human pathologies suggests that specific diseases might benefit from epigenetic-targeted therapies and this type of drug therapy is becoming a reality in clinical settings [160, 161]. Notably, epigenetic variation could arise as a consequence of the disease. Distinguishing epigenetic variations causing or contributing to the disease process is not straightforward but is nevertheless crucial to elucidate the functional role of the disease-associated epigenetic variation and to optimize their utility in terms of diagnostics or therapeutics. Recent advances in genomic technologies, by the expanding use of next-generation DNA sequencing (ChIP-seq) to assess the genomic distribution of histone modifications, histone variants, DNA methylation, and epigenetic factors, will be helpful to study human disease-associated epigenetic variation at genomewide level. Combined with appropriate statistical and bioinformatic tools [162], these methods will give us a more complete picture of all the loci that are epigenetically altered, although they will not resolve the *cause or consequence* issue. Then, the functional characterization of the variety of epigenetic modifications at specific loci could provide insight into the function of these modifications in normal development and in subsequent transition to disease states. These studies could ultimately lead to the future development of more effective epigenetic-based therapies, although treatment with these classes of drugs should be carefully examined to determine whether the therapeutic benefits outweigh the potential adverse effects [163, 164].

Acknowledgments

The author would like to thank Valerio Orlando for his mentorship, encouragement, and for kindly providing lab space; and Federica Lo Sardo and Beatrice Bodega for constructive criticisms about this manuscript. C. Lanzuolo is supported by the Italian Ministry of Research and University (Futuro in Ricerca RBFR106S1Z_001).

References

[1] C. Beisel and R. Paro, "Silencing chromatin: comparing modes and mechanisms," *Nature Reviews Genetics*, vol. 12, no. 2, pp. 123–135, 2011.

[2] E. I. Campos and D. Reinberg, "Histones: annotating chromatin," *Annual Review of Genetics*, vol. 43, pp. 559–599, 2009.

[3] A. M. Deaton and A. Bird, "CpG islands and the regulation of transcription," *Genes and Development*, vol. 25, no. 10, pp. 1010–1022, 2011.

[4] C. Lanzuolo and V. Orlando, "The function of the epigenome in cell reprogramming," *Cellular and Molecular Life Sciences*, vol. 64, no. 9, pp. 1043–1062, 2007.

[5] Y. S. Mao, B. Zhang, and D. L. Spector, "Biogenesis and function of nuclear bodies," *Trends in Genetics*, vol. 27, no. 8, pp. 295–306, 2011.

[6] S. B. Baylin and P. A. Jones, "A decade of exploring the cancer epigenome—biological and translational implications," *Nature Reviews Cancer*, vol. 11, pp. 726–773, 2011.

[7] S. R. Bhaumik, E. Smith, and A. Shilatifard, "Covalent modifications of histones during development and disease pathogenesis," *Nature Structural and Molecular Biology*, vol. 14, no. 11, pp. 1008–1016, 2007.

[8] Z. Siegfried, S. Eden, M. Mendelsohn, X. Feng, B. Z. Tsuberi, and H. Cedar, "DNA methylation represses transcription in vivo," *Nature Genetics*, vol. 22, no. 2, pp. 203–206, 1999.

[9] Y. Q. Feng, R. Desprat, H. Fu et al., "DNA methylation supports intrinsic epigenetic memory in mammalian cells," *PLoS Genetics*, vol. 2, no. 4, article e65, 2006.

[10] A. Bird, M. Taggart, M. Frommer, O. J. Miller, and D. Macleod, "A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA," *Cell*, vol. 40, no. 1, pp. 91–99, 1985.

[11] R. S. Illingworth and A. P. Bird, "CpG islands—"a rough guide"," *FEBS Letters*, vol. 583, no. 11, pp. 1713–1720, 2009.

[12] R. S. Illingworth, U. Gruenewald-Schneider, S. Webb et al., "Orphan CpG Islands Identify numerous conserved promoters in the mammalian genome," *PLoS Genetics*, vol. 6, no. 9, article e1001134, 2010.

[13] S. Kriaucionis and N. Heintz, "The nuclear DNA base 5-hydroxymethylcytosine is present in purkinje neurons and the brain," *Science*, vol. 324, no. 5929, pp. 929–930, 2009.

[14] M. Tahiliani, K. P. Koh, Y. Shen et al., "Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1," *Science*, vol. 324, no. 5929, pp. 930–935, 2009.

[15] S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers, and Y. Zhang, "Role of tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification," *Nature*, vol. 466, no. 7310, pp. 1129–1133, 2010.

[16] P. Zhang, L. Su, Z. Wang, S. Zhang, J. Guan et al., "The involvement of 5-hydroxymethylcytosine in active DNA demethylation in mice," *Biology of Reproduction*, vol. 86, no. 4, p. 104, 2012.

[17] A. Munshi, G. Shafi, N. Aliya, and A. Jyothy, "Histone modifications dictate specific biological readouts," *Journal of Genetics and Genomics*, vol. 36, no. 2, pp. 75–88, 2009.

[18] M. Altaf, N. Saksouk, and J. Côté, "Histone modifications in response to DNA damage," *Mutation Research*, vol. 618, no. 1-2, pp. 81–90, 2007.

[19] O. Bell, M. Schwaiger, E. J. Oakeley et al., "Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing," *Nature Structural and Molecular Biology*, vol. 17, no. 7, pp. 894–900, 2010.

[20] B. E. Bernstein, M. Kamal, K. Lindblad-Toh et al., "Genomic maps and comparative analysis of histone modifications in human and mouse," *Cell*, vol. 120, no. 2, pp. 169–181, 2005.

[21] M. L. Eaton, J. A. Prinz, H. K. MacAlpine, G. Tretyakov, P. V. Kharchenko, and D. M. MacAlpine, "Chromatin signatures of the Drosophila replication program," *Genome Research*, vol. 21, no. 2, pp. 164–174, 2011.

[22] G. J. Filion, J. G. van Bemmel, U. Braunschweig, W. Talhout, and J. Kind, "Systematic protein location mapping reveals five principal chromatin types in Drosophila cells," *Cell*, vol. 143, pp. 212–224, 2010.

[23] V. Krishnan, M. Z. Y. Chow, Z. Wang et al., "Histone H4 lysine 16 hypoacetylation is associated with defective DNA repair and premature senescence in Zmpste24-deficient mice," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 30, pp. 12325–12330, 2011.

[24] D. Schübeler, D. M. MacAlpine, D. Scalzo et al., "The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote," *Genes and Development*, vol. 18, no. 11, pp. 1263–1271, 2004.

[25] S. Henikoff and K. Ahmad, "Assembly of variant histones into chromatin," *Annual Review of Cell and Developmental Biology*, vol. 21, pp. 133–153, 2005.

[26] B. Schuettengruber, A. M. Martinez, N. Iovino, and G. Cavalli, "Trithorax group proteins: switching genes on and keeping them active," *Nature Reviews Molecular Cell Biology*, vol. 12, pp. 799–814, 2011.

[27] E. Splinter and W. de Laat, "The complex transcription regulatory landscape of our genome: control in three dimensions," *EMBO Journal*, vol. 30, pp. 4345–4355, 2011.

[28] G. Li and D. Reinberg, "Chromatin higher-order structures and gene regulation," *Current Opinion in Genetics and Development*, vol. 21, no. 2, pp. 175–186, 2011.

[29] P. Fraser and W. Bickmore, "Nuclear organization of the genome and the potential for gene regulation," *Nature*, vol. 447, no. 7143, pp. 413–417, 2007.

[30] C. Lanctôt, T. Cheutin, M. Cremer, G. Cavalli, and T. Cremer, "Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions," *Nature Reviews Genetics*, vol. 8, no. 2, pp. 104–115, 2007.

[31] E. Soutoglou and T. Misteli, "Mobility and immobility of chromatin in transcription and genome stability," *Current Opinion in Genetics and Development*, vol. 17, no. 5, pp. 435–442, 2007.

[32] S. Schoenfelder, I. Clay, and P. Fraser, "The transcriptional interactome: gene expression in 3D," *Current Opinion in Genetics and Development*, vol. 20, no. 2, pp. 127–133, 2010.

[33] T. Klymenko, B. Papp, W. Fischle et al., "A polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities," *Genes and Development*, vol. 20, no. 9, pp. 1110–1122, 2006.

[34] J. C. Scheuermann, A. G. de Ayala Alonso, K. Oktaba et al., "Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB," *Nature*, vol. 465, no. 7295, pp. 243–247, 2010.

[35] R. Cao, L. Wang, H. Wang et al., "Role of histone H3 lysine 27 methylation in polycomb-group silencing," *Science*, vol. 298, no. 5595, pp. 1039–1043, 2002.

[36] B. Czermin, R. Melfi, D. McCabe, V. Seitz, A. Imhof, and V. Pirrotta, "Drosophila enhancer of Zeste/ESC complexes

have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites," *Cell*, vol. 111, no. 2, pp. 185–196, 2002.

[37] A. Kuzmichev, K. Nishioka, H. Erdjument-Bromage, P. Tempst, and D. Reinberg, "Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of zeste protein," *Genes and Development*, vol. 16, no. 22, pp. 2893–2905, 2002.

[38] J. Müller, C. M. Hart, N. J. Francis et al., "Histone methyltransferase activity of a Drosophila Polycomb group repressor complex," *Cell*, vol. 111, no. 2, pp. 197–208, 2002.

[39] H. Wang, L. Wang, H. Erdjument-Bromage et al., "Role of histone H2A ubiquitination in Polycomb silencing," *Nature*, vol. 431, no. 7010, pp. 873–878, 2004.

[40] M. Ku, R. P. Koche, E. Rheinbay et al., "Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains," *PLoS Genetics*, vol. 4, no. 10, article e1000242, 2008.

[41] S. Schoeftner, A. K. Sengupta, S. Kubicek et al., "Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing," *EMBO Journal*, vol. 25, no. 13, pp. 3110–3122, 2006.

[42] D. Cmarko, P. J. Verschure, A. P. Otte, R. van Driel, and S. Fakan, "Polycomb group gene silencing proteins are concentrated in the perichromatin compartment of the mammalian nucleus," *Journal of Cell Science*, vol. 116, no. 2, pp. 335–343, 2003.

[43] F. Bantignies, V. Roure, I. Comet et al., "Polycomb-dependent regulatory contacts between distant hox loci in drosophila," *Cell*, vol. 144, no. 2, pp. 214–226, 2011.

[44] C. Lanzuolo, V. Roure, J. Dekker, F. Bantignies, and V. Orlando, "Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex," *Nature Cell Biology*, vol. 9, no. 10, pp. 1167–1174, 2007.

[45] F. Cléard, Y. Moshkin, F. Karch, and R. K. Maeda, "Probing long-distance regulatory interactions in the Drosophila melanogaster bithorax complex using Dam identification," *Nature Genetics*, vol. 38, no. 8, pp. 931–935, 2006.

[46] T. Sexton, E. Yaffe, E. Kenigsberg, F. Bantignies, and B. Leblanc, "Three-dimensional folding and functional organization principles of the Drosophila genome," *Cell*, vol. 148, pp. 458–472, 2012.

[47] S. Kheradmand Kia, P. Solaimani Kartalaei, E. Farahbakhshian, F. Pourfarzad, and M. von Lindern, "EZH2-dependent chromatin looping controls INK4a and INK4b, but not ARF, during human progenitor cell differentiation and cellular senescence," *Epigenetics & Chromatin*, vol. 2, no. 1, p. 16, 2009.

[48] D. Noordermeer, M. Leleu, E. Splinter, J. Rougemont, and W. De Laat, "The dynamic architecture of Hox gene clusters," *Science*, vol. 334, pp. 222–225, 2011.

[49] V. K. Tiwari, K. M. McGarvey, J. D. F. Licchesi et al., "PcG proteins, DNA methylation, and gene repression by chromatin looping," *PLoS Biology*, vol. 6, no. 12, pp. 2911–2927, 2008.

[50] K. H. Hansen, A. P. Bracken, D. Pasini et al., "A model for transmission of the H3K27me3 epigenetic mark," *Nature Cell Biology*, vol. 10, no. 11, pp. 1291–1300, 2008.

[51] R. Margueron, N. Justin, K. Ohno et al., "Role of the polycomb protein EED in the propagation of repressive histone marks," *Nature*, vol. 461, no. 7265, pp. 762–767, 2009.

[52] S. Chen, L. R. Bohrer, A. N. Rai et al., "Cyclin-dependent kinases regulate epigenetic gene silencing through phosphorylation of EZH2," *Nature Cell Biology*, vol. 12, no. 11, pp. 1108–1114, 2010.

[53] S. Kaneko, G. Li, J. Son et al., "Phosphorylation of the PRC2 component Ezh2 is cell cycle-regulated and up-regulates its binding to ncRNA," *Genes and Development*, vol. 24, no. 23, pp. 2615–2620, 2010.

[54] C. Lanzuolo, F. Lo Sardo, A. Diamantini, and V. Orlando, "PcG complexes set the stage for epigenetic inheritance of gene silencing in early S phase before replication," *PLoS Genetics*, vol. 7, article e1002370, 2011.

[55] C. Lanzuolo, F. Lo Sardo, and V. Orlando, "Concerted epigenetic signatures inheritance at PcG targets through replication," *Cell Cycle*, vol. 11, no. 7, pp. 1296–1300, 2012.

[56] N. Brockdorff, "Chromosome silencing mechanisms in X-chromosome inactivation: unknown unknowns," *Development*, vol. 138, pp. 5057–5065, 2011.

[57] C. Prezioso and V. Orlando, "Polycomb proteins in mammalian cell differentiation and plasticity," *FEBS Letters*, vol. 585, no. 13, pp. 2067–2077, 2011.

[58] C. L. Fisher and A. G. Fisher, "Chromatin states in pluripotent, differentiated, and reprogrammed cells," *Current Opinion in Genetics and Development*, vol. 21, no. 2, pp. 140–146, 2011.

[59] R. Margueron and D. Reinberg, "The Polycomb complex PRC2 and its mark in life," *Nature*, vol. 469, no. 7330, pp. 343–349, 2011.

[60] M. C. Trask and J. Mager, "Complexity of polycomb group function: diverse mechanisms of target specificity," *Journal of Cellular Physiology*, vol. 226, no. 7, pp. 1719–1721, 2011.

[61] K. Williams, J. Christensen, M. T. Pedersen et al., "TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity," *Nature*, vol. 473, no. 7347, pp. 343–349, 2011.

[62] H. Wu, A. C. D'Alessio, S. Ito et al., "Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells," *Nature*, vol. 473, no. 7347, pp. 389–394, 2011.

[63] H. Wu, A. C. D'Alessio, S. Ito et al., "Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells," *Genes and Development*, vol. 25, no. 7, pp. 679–684, 2011.

[64] K. E. Davies and K. J. Nowak, "Molecular mechanisms of muscular dystrophies: old and new players," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 10, pp. 762–773, 2006.

[65] K. Matsumura, F. M. S. Tome, H. Collin et al., "Expression of dystrophin-associated proteins in dystrophin-positive muscle fibers (revertants) in Duchenne muscular dystrophy," *Neuromuscular Disorders*, vol. 4, no. 2, pp. 115–120, 1994.

[66] J. M. Ervasti and K. J. Sonnemann, "Biology of the Striated Muscle Dystrophin-Glycoprotein Complex," *International Review of Cytology*, vol. 265, pp. 191–225, 2008.

[67] C. Colussi, C. Mozzetta, A. Gurtner et al., "HDAC2 blockade by nitric oxide and histone deacetylase inhibitors reveals a common target in Duchenne muscular dystrophy treatment," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 49, pp. 19183–19187, 2008.

[68] D. Cacchiarelli, J. Martone, E. Girardi et al., "MicroRNAs involved in molecular circuitries relevant for the duchenne muscular dystrophy pathogenesis are controlled by the dystrophin/nNOS pathway," *Cell Metabolism*, vol. 12, no. 4, pp. 341–351, 2010.

[69] D. Cacchiarelli, T. Incitti, J. Martone et al., "MiR-31 modulates dystrophin expression: new implications for Duchenne muscular dystrophy therapy," *EMBO Reports*, vol. 12, no. 2, pp. 136–141, 2011.

[70] G. C. Minetti, C. Colussi, R. Adami et al., "Functional and morphological recovery of dystrophic muscles in mice treated with deacetylase inhibitors," *Nature Medicine*, vol. 12, no. 10, pp. 1147–1150, 2006.

[71] S. Consalvi, V. Saccone, L. Giordani, G. Minetti, C. Mozzetta, and P. L. Puri, "Histone deacetylase inhibitors in the treatment of muscular dystrophies: epigenetic drugs for genetic diseases," *Molecular Medicine*, vol. 17, no. 5-6, pp. 457–465, 2011.

[72] M. Gomes-Pereira, T. A. Cooper, and G. Gourdon, "Myotonic dystrophy mouse models: towards rational therapy development," *Trends in Molecular Medicine*, vol. 17, no. 9, pp. 506–517, 2011.

[73] J. W. Day and L. P. W. Ranum, "RNA pathogenesis of the myotonic dystrophies," *Neuromuscular Disorders*, vol. 15, no. 1, pp. 5–16, 2005.

[74] L. A. Boyer, K. Plath, J. Zeitlinger et al., "Polycomb complexes repress developmental regulators in murine embryonic stem cells," *Nature*, vol. 441, no. 7091, pp. 349–353, 2006.

[75] A. P. Bracken, N. Dietrich, D. Pasini, K. H. Hansen, and K. Helin, "Genome-wide mapping of polycomb target genes unravels their roles in cell fate transitions," *Genes and Development*, vol. 20, no. 9, pp. 1123–1136, 2006.

[76] T. I. Lee, R. G. Jenner, L. A. Boyer et al., "Control of developmental regulators by polycomb in human embryonic stem cells," *Cell*, vol. 125, no. 2, pp. 301–313, 2006.

[77] N. Nègre, J. Hennetin, L. V. Sun et al., "Chromosomal distribution of PcG proteins during Drosophila development," *PLoS Biology*, vol. 4, no. 6, article e170, 2006.

[78] Y. B. Schwartz, T. G. Kahn, D. A. Nix et al., "Genome-wide analysis of Polycomb targets in Drosophila melanogaster," *Nature Genetics*, vol. 38, no. 6, pp. 700–705, 2006.

[79] S. L. Squazzo, H. O'Geen, V. M. Komashko et al., "Suz12 binds to silenced regions of the genome in a cell-type-specific manner," *Genome Research*, vol. 16, no. 7, pp. 890–900, 2006.

[80] B. Tolhuis, I. Muijrers, E. De Wit et al., "Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster," *Nature Genetics*, vol. 38, no. 6, pp. 694–699, 2006.

[81] P. Asp, R. Blum, V. Vethantham et al., "Genome-wide remodeling of the epigenetic landscape during myogenic differentiation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 22, pp. E149–E158, 2011.

[82] G. Caretti, M. Di Padova, B. Micales, G. E. Lyons, and V. Sartorelli, "The Polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation," *Genes and Development*, vol. 18, pp. 2627–2638, 2004.

[83] A. H. Juan, R. M. Kumar, J. G. Marx, R. A. Young, and V. Sartorelli, "Mir-214-dependent regulation of the polycomb protein Ezh2 in skeletal muscle and embryonic stem cells," *Molecular Cell*, vol. 36, no. 1, pp. 61–74, 2009.

[84] L. Stojic, Z. Jasencakova, C. Prezioso, A. Stutzer, and B. Bodega, "Chromatin regulated interchange between polycomb repressive complex 2 (PRC2)-Ezh2 and PRC2-Ezh1 complexes controls myogenin activation in skeletal muscle cells," *Epigenetics Chromatin*, vol. 4, article 16, 2011.

[85] A. H. Juan, A. Derfoul, X. Feng et al., "Polycomb EZH2 controls self-renewal and safeguards the transcriptional identity of skeletal muscle stem cells," *Genes and Development*, vol. 25, no. 8, pp. 789–794, 2011.

[86] M. Cesana, D. Cacchiarelli, I. Legnini, T. Santini, and O. Sthandier, "A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA," *Cell*, vol. 147, pp. 358–369, 2011.

[87] M. V. Neguembor and D. Gabellini, "In junk we trust: repetitive DNA, epigenetics and facioscapulohumeral muscular dystrophy," *Epigenomics*, vol. 2, no. 2, pp. 271–287, 2010.

[88] D. Palacios, C. Mozzetta, S. Consalvi et al., "TNF/p38α/polycomb signaling to Pax7 locus in satellite cells links inflammation to the epigenetic control of muscle regeneration," *Cell Stem Cell*, vol. 7, no. 4, pp. 455–469, 2010.

[89] A. Mattout, B. L. Pike, B. D. Towbin, E. M. Bank, A. Gonzalez-Sandoval, and et al., "An EDMD mutation in C. elegans lamin blocks muscle-specific gene relocation and compromises muscle integrity," *Current Biology*, vol. 21, pp. 1603–1614, 2011.

[90] R. Tawil and S. M. Van Der Maarel, "Facioscapulohumeral muscular dystrophy," *Muscle and Nerve*, vol. 34, no. 1, pp. 1–15, 2006.

[91] R. Tawil, D. Storvick, T. E. Feasby, B. Weiffenbach, and R. C. Griggs, "Extreme variability of expression in monozygotic twins with FSH muscular dystrophy," *Neurology*, vol. 43, no. 2, pp. 345–348, 1993.

[92] R. C. Griggs, R. Tawil, M. McDermott, J. Forrester, D. Figlewicz, and B. Weiffenbach, "Monozygotic twins with facioscapulohumeral dystrophy (FSHD): implications for genotype/phenotype correlation," *Muscle and Nerve*, vol. 18, no. 2, pp. S50–S55, 1995.

[93] J. C. T. Van Deutekom, C. Wijmenga, E. A. E. Van Tienhoven et al., "FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit," *Human Molecular Genetics*, vol. 2, no. 12, pp. 2037–2042, 1993.

[94] J. C. T. van Deutekom, E. Bakker, R. J. L. F. Lemmers et al., "Evidence for subtelomeric exchange of 3.3 kb tandemly repeated units between chromosomes 4q35 and 10q26: implications for genetic counselling and etiology of FSHD1," *Human Molecular Genetics*, vol. 5, no. 12, pp. 1997–2003, 1996.

[95] P. W. Lunt, P. E. Jardine, M. C. Koch et al., "Correlation between fragment size at D4F104S1 and age at onset or at wheelchair use, with a possible generational effect, accounts for much phenotypic variation in 4q35-facioscapulohumeral muscular dystrophy (FSHD)," *Human Molecular Genetics*, vol. 4, no. 5, pp. 951–958, 1995.

[96] E. Ricci, G. Galluzzi, G. Deidda, S. Cacurri, and L. Colantoni, "Progress in the molecular diagnosis of facioscapulohumeral muscular dystrophy and correlation between the number of KpnI repeats at the 4q35 locus and clinical phenotype," *Annals of Neurology*, vol. 45, pp. 751–757, 1999.

[97] R. Tawil, J. Forrester, R. C. Griggs et al., "Evidence for anticipation and association of deletion size with severity in facioscapulohumeral muscular systrophy," *Annals of Neurology*, vol. 39, no. 6, pp. 744–748, 1996.

[98] R. Tupler, A. Berardinelli, L. Barbierato et al., "Monosomy of distal 4q does not cause facioscapulohumeral muscular dystrophy," *Journal of Medical Genetics*, vol. 33, no. 5, pp. 366–370, 1996.

[99] R. J. L. F. Lemmers, P. De Kievit, L. Sandkuijl et al., "Facioscapulohumeral muscular dystrophy is uniquely associated with one of the two variants of the 4q subtelomere," *Nature Genetics*, vol. 32, no. 2, pp. 235–236, 2002.

[100] R. J. L. F. Lemmers, P. J. Van Der Vliet, R. Klooster et al., "A unifying genetic model for facioscapulohumeral muscular dystrophy," *Science*, vol. 329, no. 5999, pp. 1650–1653, 2010.

[101] R. J. F. L. Lemmers, M. Wohlgemuth, R. R. Frants, G. W. Padberg, E. Morava, and S. M. Van Der Maarel, "Contractions of D4Z4 on 4qB subtelomeres do not cause facioscapulohumeral muscular dystrophy," *American Journal of Human Genetics*, vol. 75, no. 6, pp. 1124–1130, 2004.

[102] N. S. T. Thomas, K. Wiseman, G. Spurlock, M. MacDonald, D. Üstek, and M. Upadhyaya, "A large patient study confirming that facioscapulohumeral muscular dystrophy (FSHD) disease expression is almost exclusively associated with an FSHD locus located on a 4qA-defined 4qter subtelomere," *Journal of Medical Genetics*, vol. 44, no. 3, pp. 215–218, 2007.

[103] R. Lyle, T. J. Wright, L. N. Clark, and J. E. Hewitt, "The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes," *Genomics*, vol. 28, no. 3, pp. 389–397, 1995.

[104] E. Bakker, C. Wijmenga, R. H. A. M. Vossen et al., "The FSHD-linked locus D4F104S1 (p13E-11) on 4q35 has a homologue on 10qter," *Muscle and Nerve*, vol. 18, no. 2, pp. S39–S44, 1995.

[105] G. Deidda, S. Cacurri, P. Grisanti, E. Vigneti, N. Piazzo, and L. Felicetti, "Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter," *European Journal of Human Genetics*, vol. 3, no. 3, pp. 155–167, 1995.

[106] R. J. L. F. Lemmers, P. de Kievit, M. van Geel et al., "Complete allele information in the diagnosis of facioscapulohumeral muscular dystrophy by triple DNA analysis," *Annals of Neurology*, vol. 50, no. 6, pp. 816–819, 2001.

[107] D. Gabellini, M. R. Green, and R. Tupler, "Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle," *Cell*, vol. 110, no. 3, pp. 339–348, 2002.

[108] B. Bodega, G. D. Ramirez, F. Grasser et al., "Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation," *BMC Biology*, vol. 7, article 41, 2009.

[109] D. Gabellini, G. D'Antona, M. Moggio et al., "Facioscapulohumeral muscular dystrophy in mice overexpressing FRG1," *Nature*, vol. 439, no. 7079, pp. 973–977, 2006.

[110] G. Jiang, F. Yang, P. G. M. van Overveld, V. Vedanarayanan, S. van der Maarel, and M. Ehrlich, "Testing the position-effect variegation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q," *Human Molecular Genetics*, vol. 12, no. 22, pp. 2909–2921, 2003.

[111] R. J. Osborne, S. Welle, S. L. Venance, C. A. Thornton, and R. Tawil, "Expression profile of FSHD supports a link between retinal vasculopathy and muscular dystrophy," *Neurology*, vol. 68, no. 8, pp. 569–577, 2007.

[112] S. van Koningsbruggen, R. W. Dirks, A. M. Mommaas et al., "FRG1P is localised in the nucleolus, Cajal bodies, and speckles," *Journal of Medical Genetics*, vol. 41, no. 4, article e46, 2004.

[113] S. van Koningsbruggen, K. R. Straasheijm, E. Sterrenburg et al., "FRG1P-mediated aggregation of proteins involved in pre-mRNA processing," *Chromosoma*, vol. 116, no. 1, pp. 53–64, 2007.

[114] L. Snider, L. N. Geng, R. J. Lemmers et al., "Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene," *PLoS Genetics*, vol. 6, no. 10, article e1001181, 2010.

[115] M. Dixit, E. Ansseau, A. Tassin et al., "DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 46, pp. 18157–18162, 2007.

[116] L. Snider, A. Asawachaicharn, A. E. Tyler et al., "RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy," *Human Molecular Genetics*, vol. 18, no. 13, pp. 2414–2430, 2009.

[117] V. Kowaljow, A. Marcowycz, E. Ansseau et al., "The DUX4 gene at the FSHD1A locus encodes a pro-apoptotic protein," *Neuromuscular Disorders*, vol. 17, no. 8, pp. 611–623, 2007.

[118] D. Bosnakovski, S. Lamb, T. Simsek et al., "DUX4c, an FSHD candidate gene, interferes with myogenic regulators and abolishes myoblast differentiation," *Experimental Neurology*, vol. 214, no. 1, pp. 87–96, 2008.

[119] L. N. Geng, Z. Yao, L. Snider, A. P. Fong, and J. N. Cech, "DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy," *Developmental Cell*, vol. 22, pp. 38–51, 2012.

[120] D. S. Cabianca and D. Gabellini, "FSHD: copy number variations on the theme of muscular dystrophy," *Journal of Cell Biology*, vol. 191, no. 6, pp. 1049–1060, 2010.

[121] S. M. Van der Maarel, R. Tawil, and S. J. Tapscott, "Facioscapulohumeral muscular dystrophy and DUX4: breaking the silence," *Trends in Molecular Medicine*, vol. 17, no. 5, pp. 252–258, 2011.

[122] J. E. Hewitt, R. Lyle, L. N. Clark et al., "Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy," *Human Molecular Genetics*, vol. 3, no. 8, pp. 1287–1295, 1994.

[123] J. C. de Greef, R. J. Lemmers, B. G. van Engelen et al., "Common epigenetic changes of D4Z4 in contraction-dependent and contraction-independent FSHD," *Human Mutation*, vol. 30, no. 10, pp. 1449–1459, 2009.

[124] P. G. M. Van Overveld, R. J. F. L. Lemmers, G. Deidda et al., "Interchromosomal repeat array interactions between chromosomes 4 and 10: a model for subtelomeric plasticity," *Human Molecular Genetics*, vol. 9, no. 19, pp. 2879–2884, 2000.

[125] W. Zeng, J. C. De Greef, Y. Y. Chen et al., "Specific loss of histone H3 lysine 9 trimethylation and HP1γ/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD)," *PLoS Genetics*, vol. 5, no. 7, article e1000559, 2009.

[126] J. Dekker, K. Rippe, M. Dekker, and N. Kleckner, "Capturing chromosome conformation," *Science*, vol. 295, no. 5558, pp. 1306–1311, 2002.

[127] P. S. Masny, U. Bengtsson, S. A. Chung et al., "Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease?" *Human Molecular Genetics*, vol. 13, no. 17, pp. 1857–1871, 2004.

[128] R. Tam, K. P. Smith, and J. B. Lawrence, "The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres," *Journal of Cell Biology*, vol. 167, no. 2, pp. 269–279, 2004.

[129] B. Bodega, M. F. Cardone, S. Müller et al., "Evolutionary genomic remodelling of the human 4q subtelomere (4q35.2)," *BMC Evolutionary Biology*, vol. 7, article 39, 2007.

[130] A. Ottaviani, C. Schluth-Bolard, S. Rival-Gervier et al., "Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF," *EMBO Journal*, vol. 28, no. 16, pp. 2428–2436, 2009.

[131] I. Scionti, G. Fabbri, C. Fiorillo, G. Ricci, and F. Greco, "Facioscapulohumeral muscular dystrophy: new insights from compound heterozygotes and implication for prenatal genetic counselling," *Journal of Medical Genetics*, vol. 49, pp. 171–178, 2012.

[132] G. Patrizi and M. Poger, "The ultrastructure of the nuclear periphery. The Zonula Nucleum Limitans," *Journal of Ultrasructure Research*, vol. 17, no. 1-2, pp. 127–136, 1967.

[133] F. Lin and H. J. Worman, "Structural organization of the human gene encoding nuclear lamin A and nuclear lamin C," *The Journal of Biological Chemistry*, vol. 268, no. 22, pp. 16321–16326, 1993.

[134] A. E. Rusiñol and M. S. Sinensky, "Farnesylated lamins, progeroid syndromes and farnesyl transferase inhibitors," *Journal of Cell Science*, vol. 119, no. 16, pp. 3265–3272, 2006.

[135] G. Krohne, I. Waizenegger, and T. H. Hoger, "The conserved carboxy-terminal cysteine of nuclear lamins is essential for lamin association with the nuclear envelope," *Journal of Cell Biology*, vol. 109, no. 5, pp. 2003–2011, 1989.

[136] S. G. Young, L. G. Fong, and S. Michaelis, "Prelamin A, Zmpste24, misshapen cell nuclei, and progeria—new evidence suggesting that protein farnesylation could be important for disease pathogenesis," *Journal of Lipid Research*, vol. 46, no. 12, pp. 2531–2558, 2005.

[137] J. L. V. Broers, B. M. Machiels, H. J. H. Kuijpers et al., "A- and B-type lamins are differentially expressed in normal human tissues," *Histochemistry and Cell Biology*, vol. 107, no. 6, pp. 505–517, 1997.

[138] L. Guelen, L. Pagie, E. Brasset et al., "Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions," *Nature*, vol. 453, no. 7197, pp. 948–951, 2008.

[139] H. Pickersgill, B. Kalverda, E. De Wit, W. Talhout, M. Fornerod, and B. Van Steensel, "Characterization of the Drosophila melanogaster genome at the nuclear lamina," *Nature Genetics*, vol. 38, no. 9, pp. 1005–1014, 2006.

[140] J. L. V. Broers, F. C. S. Ramaekers, G. Bonne, R. Ben Yaou, and C. J. Hutchison, "Nuclear lamins: laminopathies and their role in premature ageing," *Physiological Reviews*, vol. 86, no. 3, pp. 967–1008, 2006.

[141] B. R. Johnson, R. T. Nitta, R. L. Frock et al., "A-type lamins regulate retinoblastoma protein function by promoting subnuclear localization and preventing proteasomal degradation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, pp. 9677–9682, 2004.

[142] V. L. R. M. Verstraeten, J. L. V. Broers, F. C. S. Ramaekers, and M. A. M. van Steensel, "The nuclear envelope, a key structure in cellular integrity and gene expression," *Current Medicinal Chemistry*, vol. 14, no. 11, pp. 1231–1248, 2007.

[143] N. D. Willis, T. R. Cox, S. F. Rahman-Casañs et al., "Lamin A/C is a risk biomarker in colorectal cancer," *PLoS ONE*, vol. 3, no. 8, article e2988, 2008.

[144] P. Hozak, A. M. J. Sasseville, Y. Raymond, and P. R. Cook, "Lamin proteins form an internal nucleoskeleton as well as a peripheral lamina in human cells," *Journal of Cell Science*, vol. 108, no. 2, pp. 635–644, 1995.

[145] L. Vergnes, M. Péterfy, M. O. Bergo, S. G. Young, and K. Reue, "Lamin B1 is required for mouse development and nuclear integrity," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 28, pp. 10428–10433, 2004.

[146] T. Sullivan, D. Escalante-Alcalde, H. Bhatt et al., "Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy," *Journal of Cell Biology*, vol. 147, no. 5, pp. 913–919, 1999.

[147] B. C. Capell and F. S. Collins, "Human laminopathies: nuclei gone genetically awry," *Nature Reviews Genetics*, vol. 7, no. 12, pp. 940–952, 2006.

[148] I. Landires, J. M. Pascale, and J. Motta, "The position of the mutation within the LMNA gene determines the type and extent of tissue involvement in laminopathies [1]," *Clinical Genetics*, vol. 71, no. 6, pp. 592–596, 2007.

[149] J. Scharner, V. F. Gnocchi, J. A. Ellis, and P. S. Zammit, "Genotype-phenotype correlations in laminopathies: how does fate translate?" *Biochemical Society Transactions*, vol. 38, no. 1, pp. 257–262, 2010.

[150] D. K. Shumaker, T. Dechat, A. Kohlmaier et al., "Mutant nuclear lamin A leads to progressive alterations of epigenetic control in premature aging," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 23, pp. 8703–8708, 2006.

[151] Y. Wang, A. J. Herron, and H. J. Worman, "Pathology and nuclear abnormalities in hearts of transgenic mice expressing M371K lamin A encoded by an LMNA mutation causing Emery-Dreifuss muscular dystrophy," *Human Molecular Genetics*, vol. 15, no. 16, pp. 2479–2489, 2006.

[152] V. Andrés and J. M. González, "Role of A-type lamins in signaling, transcription, and chromatin organization," *Journal of Cell Biology*, vol. 187, no. 7, pp. 945–957, 2009.

[153] S. Marmiroli, J. Bertacchini, F. Beretti et al., "A-type lamins and signaling: the PI 3-kinase/Akt pathway moves forward," *Journal of Cellular Physiology*, vol. 220, no. 3, pp. 553–564, 2009.

[154] P. L. Puri, S. Iezzi, P. Stiegler et al., "Class I histone deacetylases sequentially interact with MyoD and pRb during skeletal myogenesis," *Molecular Cell*, vol. 8, no. 4, pp. 885–897, 2001.

[155] J. H. Van Berlo, J. W. Voncken, N. Kubben et al., "A-type lamins are essential for TGF-β1 induced PP2A to dephosphorylate transcription factors," *Human Molecular Genetics*, vol. 14, no. 19, pp. 2839–2849, 2005.

[156] R. L. Frock, B. A. Kudlow, A. M. Evans, S. A. Jameson, S. D. Hauschka, and B. K. Kennedy, "Lamin A/C and emerin are critical for skeletal muscle satellite cell differentiation," *Genes and Development*, vol. 20, no. 4, pp. 486–500, 2006.

[157] Y. E. Park, Y. K. Hayashi, K. Goto et al., "Nuclear changes in skeletal muscle extend to satellite cells in autosomal dominant Emery-Dreifuss muscular dystrophy/limb-girdle muscular dystrophy 1B," *Neuromuscular Disorders*, vol. 19, no. 1, pp. 29–36, 2009.

[158] J. Wang, R. M. Kumar, V. J. Biggs, H. Lee, Y. Chen et al., "The Msx1 homeoprotein recruits polycomb to the nuclear periphery during development," *Developmental Cell*, vol. 21, pp. 575–588, 2011.

[159] A. Mattout, B. L. Pike, B. D. Towbin, E. M. Bank, and A. Gonzalez-Sandoval, "An EDMD mutation in C. elegans lamin blocks muscle-specific gene relocation and compromises muscle integrity," *Current Biology*, vol. 21, no. 19, pp. 1603–1614, 2011.

[160] L. Giacinti, P. Vici, and M. Lopez, "Epigenome: a new target in cancer therapy," *Clinica Terapeutica*, vol. 159, no. 5, pp. 347–360, 2008.

[161] C. Ptak and A. Petronis, "Epigenetics and complex disease: from etiology to new therapeutics," *Annual Review of Pharmacology and Toxicology*, vol. 48, pp. 257–276, 2008.

[162] S. Pepke, B. Wold, and A. Mortazavi, "Computation for ChIP-seq and RNA-seq studies," *Nature methods*, vol. 6, no. 11, pp. S22–S32, 2009.

[163] J. Peedicayil, "Pharmacoepigenetics and pharmacoepigenomics," *Pharmacogenomics*, vol. 9, no. 12, pp. 1785–1786, 2008.

[164] B. Claes, I. Buysschaert, and D. Lambrechts, "Pharmacoepigenomics: discovering therapeutic approaches and biomarkers for cancer therapy," *Heredity*, vol. 105, no. 1, pp. 152–160, 2010.

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 434529, 9 pages
doi:10.1155/2012/434529

Review Article

The Stability of the Induced Epigenetic Programs

Maria J. Barrero

Center of Regenerative Medicine in Barcelona, Aiguader 88, 7th floor, 08003 Barcelona, Spain

Correspondence should be addressed to Maria J. Barrero, mjbarrero@cmrb.eu

Received 31 January 2012; Accepted 3 April 2012

Academic Editor: Lucia Latella

For many years scientists have been attracted to the possibility of changing cell identity. In the last decades seminal discoveries have shown that it is possible to reprogram somatic cells into pluripotent cells and even to transdifferentiate one cell type into another. In view of the potential applications that generating specific cell types in the laboratory can offer for cell-based therapies, the next important questions relate to the quality of the induced cell types. Importantly, epigenetic aberrations in reprogrammed cells have been correlated with defects in differentiation. Therefore, a look at the epigenome and understanding how different regulators can shape it appear fundamental to anticipate potential therapeutic pitfalls. This paper covers these epigenetic aspects in stem cells, differentiation, and reprogramming and discusses their importance for the safety of in vitro engineered cell types.

1. Introduction

The genome is organized into particular chromatin structures that have specific roles both in maintaining the overall structure and in gene expression. The fundamental unit of chromatin is the nucleosome, composed of two copies each of four core histones, H2A, H2B, H3, and H4, wrapped by 146 bp of DNA. The recruitment of linker histone H1 and other structural proteins can lead to further condensation and the of higher-order structures, which play additional roles in the organization of chromosomes. Chromatin offers a physical barrier to the efficient recruitment and processivity of the RNA Polymerase II (Poll l) and thus impedes gene transcription [1].

The extent of chromatin condensation is subject to regulation. The N-terminal tails of histones are relatively accessible to enzymatic modifications such as acetylation, methylation, phosphorylation, ubiquitination, and sumoylation. Furthermore, the cytosine residues of DNA can be modified by methylation and hydroxymethylation. These modifications can influence the degree of condensation of chromatin per se or/and facilitate the recruitment of structural or effector proteins, such as remodeling complexes, that directly affect the condensation of chromatin.

Certain areas of the genome are organized into heavily condensed chromatin structures, such as centromeric regions, and offer little room for transcriptional regulation. These areas are enriched in H3K9 methylation and marked by the presence of structural proteins such as HP1 (heterochromatin protein 1), which contribute to maintain high levels of condensation that play mainly structural roles in the organization of chromosomes. However, other regions of the genome are enriched in genes that are silenced but that can be active in certain situations or in different cell types. Although the mechanisms of gene silencing might be heterogeneous and gene specific, overall these areas are occupied by the Polycomb complex and marked with H3K27me3. Genes encoding many developmental regulators are located in such regions.

Tissue specific genes and developmental regulators are thus subject to intense regulation. The mechanisms leading to transcriptional activation or repression are presumably gene specific and highly influenced by the transcription factors bound at the regulatory regions of a particular gene at a given time. Extensive genomewide studies have been pursued in an effort to correlate transcriptional competence and histone modifications. This rationale is the basis of the "histone code" that postulates that the particular combination of histone modifications present at a given genomic region acts as a code to specify gene activity [2]. However, although certain modifications are strongly correlated with transcriptional activation or repression, it is often difficult to predict from the presence of a single mark the transcriptional status of a gene and even more difficult to envision the

predisposition of genes to become activated or repressed. While some silent genes can be activated by certain signals others remain permanently silent and refractory to stimulation. This property is displayed in cell-specific ways and defines both cell identity and plasticity. Certain cell types, such as stem cells, have very plastic chromatin that makes them extremely sensitive to environmental signals. As cells differentiate, particular genes become silent with a consequent loss of regulatory potential.

2. The Epigenetic Landmarks of ES Cells

Embryonic stem (ES) cells are derived from the inner cell mass (ICM) of the preimplantation embryo and are characterized by their ability to self-renew and to give rise to virtually any cell type of the adult organism, a property called pluripotency.

A considerable amount of effort has been devoted to identify the network of transcription factors that control these two unique properties. As a result, a core regulatory network governed by the transcription factors Oct4, Sox2, and Nanog has been identified. These three factors are able to stimulate the expression of each other and also to control self-renewal and pluripotency through different mechanisms. First, they bind to the regulatory regions of genes involved in self-renewal and stimulate their transcriptional activity. Second, they can also occupy the regulatory regions of critical genes involved in development and differentiation and presumably contribute to maintain these genes in a silenced but poised state for activation during differentiation, which constitutes the basis of pluripotency [40]. How these factors mediate these two apparently opposite functions at the two types of genes is intriguing. Accumulating evidence suggests that despite the presence of self-renewal factors in both types of genes the chromatin complexes that assemble on these genes are completely different [6, 13, 20, 21].

Embryonic stem cells can be maintained in vitro in the presence of signalling molecules such as LIF, FGF, or TGFb. While mouse ES cells are dependent on LIF, human ES cells depend on the presence of FGF, which seems to sustain a pluripotent state that resembles mouse stem cells derived from the epiblasts. Overall, despite the differences between species, all these complex and still quite unexplored signalling events converge into two main responses: (1) maintenance of very high rates of transcription of genes that belong to the pluripotency network and (2) "poising" of developmental genes.

Overall, ES cells display high rates of transcriptional activity compared to differentiated cells [15], which is presumably devoted to the maintenance of high expression of pluripotency genes. Accordingly, ES cells express high levels of general transcription factors (GTFs) and of certain complexes involved in transcriptional activation such as the ATP-remodeling BAF complex and the Mediator complex [15, 21]. Moreover, differences in the expression levels of different subunits of these complexes lead to the formation of unique complexes that differ in subunit composition and potentially in function from differentiated cells [12, 21]. The relevance of these complexes is further supported by the reported loss of self-renewal caused by the depletion of the remodelling factors Chd1 or Brg1 [15, 16] and the Mediator subunit Med12 [21] in mouse ES cells. Moreover, a recent genomewide RNAi screening revealed the involvement of the chromatin remodelling complex INO80, the Mediator complex, and TBP-associated factors (TAFs) in human ES cell biology [14]. Since some of these factors have been found to cooccupy the regulatory regions of self-renewal genes with Oct4, Sox2, or Nanog [21] it is possible that they act as cofactors that contribute to support the high levels of transcription mediated by these pluripotency transcription factors. Accordingly, Brg1, subunits of the INO80 complex, and Chd1 have also been identified as part of the Oct4 interacting network [41, 42].

ESC chromatin presents structural peculiarities compared to differentiated cells. Heterochromatin appears more relaxed, perhaps due to the fact that the proteins involved in the formation of heterochromatin such as HP1 and linker histone H1 display hyperdynamic interactions with chromatin [43]. ES cells also display unique modification patterns, referred to as bivalent domains, at the regulatory regions of developmental genes. These are characterized by the presence of large regions of H3K27me3 harboring smaller regions of H3K4me3 around the transcriptional start site. The coexistence of these two antagonistic marks has been suggested to play a role in silencing developmental genes in ES cells while keeping them poised for activation upon initiation of specific developmental pathways [7, 44]. Bivalent genes are further enriched in CpG islands that in ES cells are nonmethylated. The enzyme Tet1, which is highly expressed in ES cells, has been suggested to maintain DNA in a hypomethylated state through the hydroxymethylation of CpGs at these domains [45, 46]. However, a recent study reports that Tet1 is dispensable for maintaining pluripotency of mouse ES cells [47]. Moreover, the function of DNA hydroxymethylation still remains obscure regarding its potential roles in protection against DNA methylation, providing docking sites for given factors or as an intermediary of DNA demethylation.

Despite being transcribed at very low levels, bivalent genes have considerable levels of transcriptionally engaged RNA Polymerase II near their transcription start sites but greatly reduced levels of productive elongating Pol ll [48]. However, Pol II at these promoters is confined to extremely proximal regions relative to the transcription start site and is in a conformation that is apparently different than the one found at bona fide paused locations of actively transcribed genes [3], suggesting that Pol II is stalled at these promoters in a unique conformation that can be referred to as "poised." Knock out of Ring1B (Table 1), the Polycomb subunit that mediates ubiquitination of histone H2A, causes the loss of ubiquitinated H2A at bivalent genes, which in turn leads to changes in Pol II conformation and the derepression of the target genes. Therefore, ubiquitination of H2A seems to play a role in restraining poised Pol II at bivalent genes [3].

The regulation and the potential role of the H3K4me3 marks at these domains remain obscure but it is likely that members of the MLL family of H3K4 methyltransferases

Table 1: Chromatin- and transcription- related complexes involved in maintaining ES cells pluripotency. The main reported activity of the complex is indicated; however, notice that specific subunits might carry out enzymatic activities different than the main described activity.

Complex	Main activity	Subunit	Reported function
Polycomb			
PRC1	H2AK119 ubiquitination	Ring1B	Restrains Pol II at bivalent genes avoiding premature differentiation [3]
		CBX7	Maintenance of pluripotency regulating PRC1 targets [4]
PRC2	H3K27 methylation	Ezh2/1	Repression of differentiation genes [5]
		Eed	Repression of differentiation genes [5–7]
		Suz12	Repression of differentiation genes and needed during differentiation [8, 9]
		JARID2	Recruitment of Polycomb to target genes [10]
		RBP2	Mediates H3K4me demethylation at bivalent genes [11]
BAF	Nucleosome remodelling	Brg1	Coactivator of the pluripotency network [12]
NuRD	Nucleosome remodelling	Mbd3	Nucleosome stabilization at bivalent domains [13]
INO80	Nucleosome remodelling	INO80	Co-activator of the pluripotency network [14]
CHD1	Nucleosome remodelling	Chd1	Co-activator of the pluripotency network [15, 16]
MLL	H3K4 methylation	UTX/jmjd3	H3K27me demethylation of bivalent domains [17]
		Dpy-30	Participates in the induction of developmental genes during differentiation [18]
		WDR5	Co-activator of the pluripotency network [19]
CoREST	Histone deacetylase	LSD1	H3K4me demethylation of bivalent domains [20]
Mediator	Transcription activation	Med12	Co-activator of the pluripotency network [21]

(Table 1) play a role in mediating this modification, while its deposition might favour the recruitment of Pol II to these domains [49]. Knockdown of the newly identified MLL subunit Dpy-30 [18] does not cause self-renewal defects, but rather defects in differentiation. However, knock down of the MLL complex core subunit WDR5 in ES cells has been reported to induce differentiation and loss of self-renewal [19]. More fully described is the essential role of the Polycomb complexes in the control of these domains (Table 1). Mouse ES cells null for specific Polycomb proteins result in decreased H3K27 methylation and show aberrantly induced expression of key developmental genes [6–8]. Interestingly, bivalent domains seem to be tightly regulated by the balance of activating and repressing activities. The Polycomb complex can mediate the recruitment of the H3K4 demethylase RBP2 to the bivalent domains to maintain the proper balance of H3K4 and H3K27 methylation in mouse ES cells [11]. Similarly, the H3K4 demethylase LSD1 is recruited to bivalent domains to regulate the levels of H3K4 methylation in human ES cells [20]. Moreover, subunits from the remodelling complexes BAF and NuRD (Brg1 and Mbd3, resp.) antagonistically control nucleosome occupancy at bivalent genes, with Brg1-mediated nucleosome loss associated with gene activation, and competing nucleosome stabilization by Mbd3 associated with gene repression. Interestingly, it has been suggested that hydroxymethyl cytosines serve to recruit the Mbd3/NURD complex to these domains [13]. Overall, bivalent domains appear governed by a complex and highly dynamic equilibrium of epigenetic activators and repressors that is likely to make them extremely sensitive to differentiation signals.

How the enzymes that maintain the bivalent domains in ES cells are recruited or stabilized at these particular regions is not fully understood. In Drosophila, the Polycomb complex is able to bind to specific DNA sequences [50]. This mechanism seems not to apply to mammalian cells, but some reports highlight the possibility that the extremely conserved distribution of CpG domains in the regulatory regions of developmental genes play a role in the recruitment of the Polycomb complex [51, 52]. Moreover, Tet1 has been reported to facilitate the chromatin binding of Polycomb components likely by decreasing DNA methylation levels at CpG-rich domains [53]. It is also possible that specific transcription factors that bind the regulatory regions of bivalent genes such as Oct4, Sox2, and Nanog contribute to Polycomb recruitment or stabilization. Accordingly, Nanog and Oct4 have been described to interact with complexes involved in transcriptional repression, including Polycomb subunits [41, 54]. Also, the transcription factor JARID2 has been suggested to participate in the recruitment of the Polycomb complex PRC2 to the regulatory regions of developmental regulators in mouse ES cells [10].

3. Epigenetic Changes during Differentiation

The in vitro differentiation of ES cells is achieved through the removal of molecules that promote self-renewal, such as LIF or FGF, and the addition of factors that induce differentiation. These changes in culture conditions lead to down-regulation of the pluripotency network and to the activation or repression of developmental genes in a germ-layer-specific fashion.

Ultimately, the physiological function of bivalent domains might be to maintain important regulatory sequences accessible to the binding of relevant transcription factors that

are activated by the differentiation signals. The regulatory areas, which were accessible at the undifferentiated stage, and which are not targeted by transcription factors, "close up" during differentiation becoming further inaccessible [55]. Therefore, differentiation to one particular lineage implies the permanent and irreversible silencing of genes involved in alternative lineages. Bivalent domains tend to resolve into methylated H3K4 alone for those genes that will become activated or methylated H3K27 alone for those that will be repressed during differentiation [56, 57]. Repression might be further reinforced by the incorporation of other repressive marks such as H3K9me3 or DNA methylation [58] ensuring the permanent silencing of developmental genes. The resolution of bivalent domains requires the coordinated action of histone lysine methyltransferases and demethylases. An elegant example is the role of H3K27 demethylates UTX and Jmjd3 in the activation of Hox genes during development [17] and in neuronal commitment [59]. Both demethylases are associated with MLL complexes [60, 61], suggesting that removal of the H3K27me3 mark and maintenance of the H3K4me3 at bivalent genes that become activated during differentiation are coordinated events. Importantly, a significant number of bivalent domains can remain unresolved and new bivalent genes might appear after differentiation [56], which might have consequences for the degree of plasticity that adult cells display.

Changes in subunit composition of chromatin-related complexes might also contribute to establishing the new epigenetic landscapes of differentiated cells. Such is the case of the Cbx subunits of the Polycomb complex. During differentiation, the expression of Cbx7 is down-regulated, while Cbx2, Cbx4, and Cbx8 are induced, leading to changes in the complex composition and properties [4]. In a similar fashion, changes in the expression of histone variants might also be involved in establishing the appropriate patterns of gene expression during differentiation. As an example, histone linker variant H1.0 is induced during differentiation and specifically recruited to the regulatory regions of pluripotency and developmental genes, contributing to their repression [62].

The silencing of the genes that belong to the pluripotency network is a critical event for proper differentiation. These genes become passively down-regulated due to the absence of LIF or FGF signalling and more actively due to the action of transcriptional repressors that are induced during differentiation. As a result, several mediators of repression are recruited to these genes, such as the methyltransferase G9a that has been reported to participate in the silencing of Oct4 by mediating methylation at H3K9 and contributing to the recruitment of HP1 and establishment of DNA methylation [63, 64]. The fact that in differentiated cells the regulatory regions of different genes of the pluripotency network are marked with different combinations of repressive modifications [58] further suggests that the mechanisms and epigenetic regulators that participate in their repression are likely to be gene specific.

4. Walking Back the Epigenetic Road during Reprogramming

Nuclear transfer experiments [23] showed for the first time that it is possible to reverse the differentiated phenotype. More recently, it became possible to generate induced pluripotent stem cells (iPSCs) from somatic cells by overexpressing specific transcription factors, most commonly Oct4, Sox2, Klf4, and c-Myc [24]. How these transcription factors impinge on the somatic cell genome to reprogram its gene expression profile is still not clear, but the low efficiency of the process suggests that somatic cells present barriers that prevent switches in cell identity. The fact that the efficiency of reprogramming can be increased by using inhibitors of DNA methyltransferases, histone methyltransferases and deacetylases [65–67] points to a critical role of chromatin as a barrier that prevents reprogramming.

Reprogramming appears to be a gradual process in which in early stages cells acquire the ability to self-renew and downregulate cell specific programs [68]. At this stage, cells can be trapped in a partially reprogrammed state in which they self-renew and continue to depend on the expression of the transgenes. A second critical phase consists of the activation of the endogenous pluripotency network, including the genes Oct4, Sox2, and Nanog. This event allows the maintenance of pluripotency in an autonomous way and independently of the transgenes. However, this stage is reached at a low frequency likely due to the inability of the transfactors to bind and activate the regulatory regions of the endogenous pluripotency genes [68]. Although the first stages of reprogramming lead to down-regulation of the expression of cell-specific genes, the complete erasure of this transcriptional memory takes place gradually after the activation of the pluripotency network [69]. Importantly, bivalent domains need to be re-established at critical developmental genes. Failure to regain this permissive status has dramatic consequences for differentiation, as found in iPSCs derived from nonhaematopoietic cells that display impaired blood-forming potential due to residual DNA methylation at loci required for differentiation into the haematopoietic lineage [70]. Moreover, the aberrant expression of bivalent genes in mouse iPSCs can be inversely correlated with their ability to give rise to viable animals by tetraploid complementation [71].

Two recent studies suggest that reprogramming factors first target regions of the genome that are in a permissive chromatin conformation in somatic cells (Figure 1). Koche et al. [72] analyzed the expression patterns and epigenetic landscapes of fibroblasts at very early stages of reprogramming. At this phase, changes in gene expression seem limited to down-regulation of the somatic specific program, even though epigenetic changes can be readily detected at promoters that are already in an open and accessible conformation. Most conspicuous early event consists of the gain of H3K4me2 at promoters that are typically marked with H3K4me3 in ES cells, such as certain pluripotency and early developmental genes. These changes are restricted to sites of high CpG density, which are devoid of DNA methylation both in fibroblasts and ES cells, and in which reprogramming

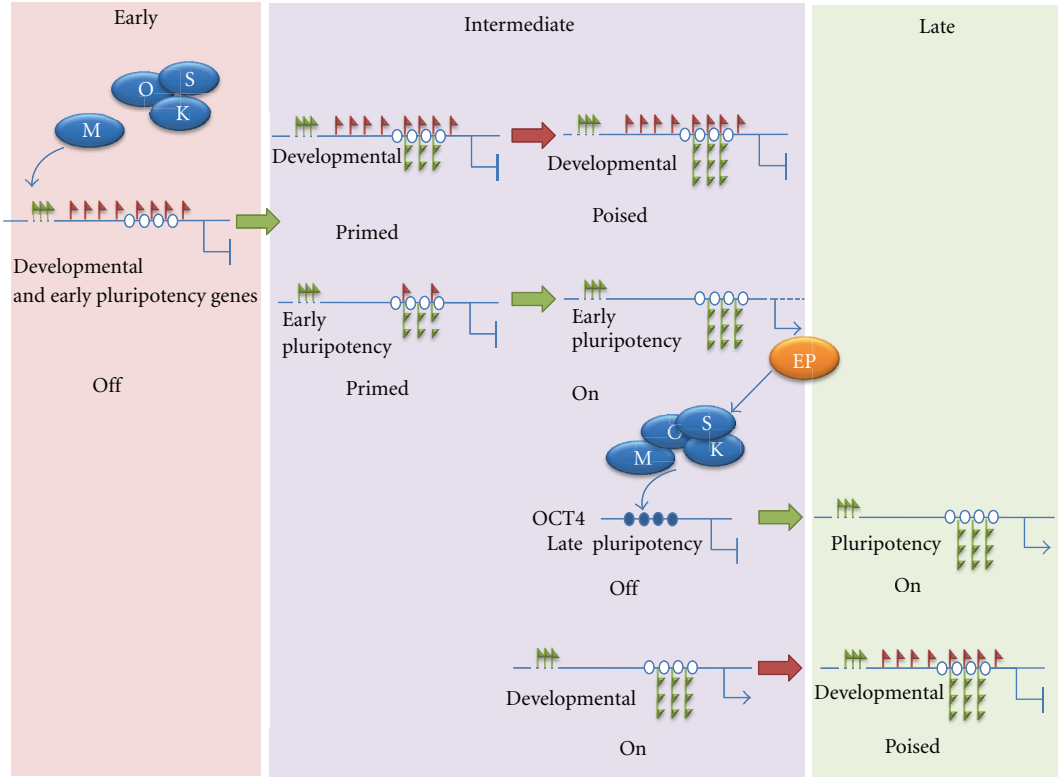


Figure 1: Epigenetic changes during reprogramming in genes containing CpG islands. Pluripotency and developmental genes have high CpG content and suffer dramatic changes during reprogramming. The reprogramming factors (OSKM, Oct4, Sox2, Klf4, and c-Myc) target preferentially the permissive enhancers of Polycomb target genes that are devoid of DNA methylation. These include silent developmental genes and perhaps pluripotency genes that respond early to the transfactors. As a result these genes gain H3K4 methylation at proximal promoters and are primed to become poised (developmental genes) or active (pluripotency genes) at latter stages. The products of early pluripotency (EP) genes might contribute to activate, together with the transfactors, late pluripotency genes marked with DNA methylation, such as Oct4. Finally, developmental genes become poised by gain of H3K27 methylation. Permissive enhancers are represented as dotted lines. Red flags denote H3K27me3. Green flags denote H3K4me1 (one flag), H3K4me2 (two flags), and H3K4me3 (three flags). Circles correspond to unmethylated (open) or methylated (filled) CpG islands.

factor regulatory motifs might be present. Using a different approach, Taberlay et al. [73] describe that developmental genes marked with H3K27me3 at their promoters contain permissive enhancers that are depleted of nucleosomes and marked with H3K4me1 in somatic cells. These enhancers are likely to be targeted by the reprogramming factors at initial stages of reprogramming.

In contrast, the gain of DNA hypermethylation typical of ES cells and the reestablishment of H3K27me3 at bivalent promoters take place late in the reprogramming process [72]. Therefore, the acquisition of the facultative heterochromatin typical of ES cells might be a late critical step that cells need to overcome during reprogramming. Indeed the silencing of tissue specific genes appears to be more important than previously thought for the process of reprogramming, as suggested by the discovery that somatic cells can be fully reprogrammed to pluripotency by overexpression of the miR-302/367 cluster alone [36].

5. Discordances between ES and IPS Cells

The process of somatic cell reprogramming generates cells with similar properties to ES cells but an important question still remains: how similar are iPS and ES cells? Studies looking at global gene expression and the epigenome suggest that they are in fact quite similar and that iPS cells are unequivocally different from the somatic cells of origin [74, 75]. However, particular differences have been identified whose significance remains to be determined [37, 76]. The comparison between ES and iPS cells is not straightforward for two main reasons: (1) ES cells themselves show variability between lines; (2) the different strategies to generate iPS cells and culture techniques used in different laboratories make it difficult to tease out as to which differences are due to variations in experimental procedures and which are intrinsic to iPS cells.

Reflecting the gradual nature of the reprogramming process, early passage iPS cells retain residual expression of

genes from the cell of origin, which has been proposed to facilitate their differentiation back into those very same cell types [69, 77]. However, the expression of these genes tends to disappear at late passages [69, 78]. Importantly, it has been described that a certain number of genes are still differentially expressed in several lines of iPS cells at late passages [76], although the consequences of this differential expression still need to be addressed.

More important is perhaps the potential differences in the epigenomes of iPS and ES cells, since this is likely to influence the ability of cells to differentiate and the quality of the final differentiated products. Regarding DNA methylation at CpGs, genome-wide analyses revealed that a significant number of developmental genes retained significant levels of DNA methylation at their regulatory regions in early passage iPS cells [79]. However, a very similar methylation profile between ES cells and iPSC was found in these genes at late passages [37]. Overall, iPS cells show a few hundred differentially methylated regions when compared to ES cells, corresponding both to somatic memory and to aberrant methylation [37, 79]. Different iPS cell lines share only a small number of differentially methylated regions, suggesting that there is significant reprogramming variability with regard to DNA methylation. However, some hot spots of shared differential methylation between lines have been found [37, 76]. More specifically, defective re-establishment of DNA methylation at particular loci that correlates with sustained expression of a few somatic genes has been reported [76]. Regarding non-CpG DNA methylation, iPS cells show hypomethylation in large regions proximal to centromeres and telomeres compared to ES cells [37]. Since the role of non-CpG DNA methylation is not clear, it is difficult to predict the potential outcome of this difference. A recent comprehensive study [74] combined analysis of global gene expression and DNA methylation in undifferentiated and differentiated cells to score for the ability of human ES and iPS cells to differentiate into certain lineages. The authors concluded that no common distinctive pattern was shared by all iPS cell analyzed and that reprogrammed cells were not functionally distinguishable from ES cells.

6. Future Goal: Assess the Epigenetic Stability to Ensure Safety

The prospect of potential uses of iPS cells in autologous therapies made researchers rush into the development of nonintegrative approaches to deliver the lowest number of factors into somatic cells [80] (Table 2). However, the idea that the absence of viral integrations would make these cells safe for therapy appears naïve if we take into account two main facts. First, iPS cells are expected to have the same drawbacks of ES cells, that is, the generation of differentiated products that are plastic enough to integrate into the damaged tissues but differentiated enough to avoid the formation of tumors. Second, iPS cells not only display epigenetic aberrancies but they also have accumulated a number of genetic mutations during reprogramming and expansion in culture [38, 39].

Table 2: Main breakthroughs regarding reprogramming and transdifferentiation of somatic cells. The need of oocytes and the low efficiency of nuclear transfer in humans have propitiated the search for alternative strategies to generate pluripotent cells. Induced pluripotent cells, initially obtained with retroviruses encoding Oct4, Sox2, Klf4 and c-Myc (OSKM), were considered unsafe for therapy due to the presence of viral integrations and the use of oncogenes Klf4 and c-Myc. Therefore, a major rush to develop non integrative methods and to avoid the use of oncogenes started. However, the finding that iPS cells have epigenetic and genetic aberrations suggests that these cells will need to be analyzed in detail before moving to the clinic.

Year	Breakthrough
1987	Fibroblast transdifferentiation to muscle cells [22]
1997	Pluripotent cells by nuclear transfer [23]
2006	Mouse iPS cells with OSKM retroviruses [24]
2007	Human iPS cells with OSKM retroviruses [25]
	IPS cells without c-Myc [26]
	iPS cells from neural stem cells with two factors [27]
2008	iPS cells with two factors and small molecules [28]
	iPS cells with non integrative viruses [29]
	Desease-specific iPS cells [30]
2009	iPS cells with proteins [31]
	iPS cells with RNA [32]
2010	Transdifferentiation of fibroblasts to neurons or to cardiomyocytes [33, 34]
	Transdifferentiation of fibroblasts to blood cells [35]
	iPS cells with miRNAs [36]
2011	iPS cells have epigenetic aberrations [37]
	iPS cells have genetic aberrations [38, 39]

Laboratories around the world have now started to generate therapeutic cell types by transdifferentiation, such as blood, neurons, or cardiomyocytes from fibroblasts [33, 81, 82]. Assessing the similarity of these cells or the ones obtained from pluripotent cells to their in vivo counterparts is not straightforward. Moreover, it is likely that these cells obtained in vitro differ significantly from those found in the body. But perhaps more important is to understand the potential adverse affects that these cells can cause. Important questions emerge. How stable are these newly induced programs? What is the probability that the transplanted cells revert to less differentiated and proliferative phenotypes? Importantly, transdifferentiation is often driven by the action of master transcriptional regulators or pioneer transcription factors able to induce whole new programs of gene expression. These master genes, which might be also regulated during early stages of reprogramming [73], are targets of Polycomb and are enriched in CpG islands that are usually DNA hypomethylated in normal tissues but become often hypermethylated in cancer [79, 83] and during aging [84]. This suggests that regions of the genome that are subject to intense regulation and that play important roles in defining cell identity might be also more likely to suffer pathological deregulation.

The answer to the above questions unequivocally lies in the epigenome and how the regulators of the epigenetic marks can secure the stability of the newly established epigenetic programs. A better characterization of the epigenetic landscape, including the identification of critical genomic areas with tendency to suffer aberrant or unstable epigenetic reprogramming and the activities involved in their regulation, will be needed in order to predict the safety of the induced cell types.

Acknowledgments

The author apologizes for those authors whose work could not be cited because of space limitation. The author is indebted to J. C. Izpisua Belmonte for his encouragement and support and to S. Malik for long fruitful discussions and critical reading of the paper. This work was partially supported by Grants RYC-2007-01510 (Ramón y Cajal program) and SAF2009-08588 from the Ministerio de Ciencia e Innovación of Spain to M. J. Barrero.

References

[1] C. Hodges, L. Bintu, L. Lubkowska, M. Kashlev, and C. Bustamante, "Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II," *Science*, vol. 325, no. 5940, pp. 626–628, 2009.

[2] T. Jenuwein and C. D. Allis, "Translating the histone code," *Science*, vol. 293, no. 5532, pp. 1074–1080, 2001.

[3] J. K. Stock, S. Giadrossi, M. Casanova et al., "Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells," *Nature Cell Biology*, vol. 9, no. 12, pp. 1428–1435, 2007.

[4] A. O'Loghlen, A. M. Muñoz-Cabello, A. Gaspar-Maia et al., "MicroRNA regulation of Cbx7 mediates a switch of *Polycomb* orthologs during ESC differentiation," *Cell Stem Cell*, vol. 10, no. 1, pp. 33–46, 2012.

[5] X. Shen, Y. Liu, Y. J. Hsu et al., "EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency," *Molecular Cell*, vol. 32, no. 4, pp. 491–502, 2008.

[6] L. A. Boyer, K. Plath, J. Zeitlinger et al., "*Polycomb* complexes repress developmental regulators in murine embryonic stem cells," *Nature*, vol. 441, no. 7091, pp. 349–353, 2006.

[7] V. Azuara, P. Perry, S. Sauer et al., "Chromatin signatures of pluripotent cell lines," *Nature cell biology.*, vol. 8, no. 5, pp. 532–538, 2006.

[8] T. I. Lee, R. G. Jenner, L. A. Boyer et al., "Control of developmental regulators by *Polycomb* in human embryonic stem cells," *Cell*, vol. 125, no. 2, pp. 301–313, 2006.

[9] D. Pasini, A. P. Bracken, J. B. Hansen, M. Capillo, and K. Helin, "The *Polycomb* group protein Suz12 is required for embryonic stem cell differentiation," *Molecular and Cellular Biology*, vol. 27, no. 10, pp. 3769–3779, 2007.

[10] D. Landeira, S. Sauer, R. Poot et al., "Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators," *Nature Cell Biology*, vol. 12, no. 6, pp. 618–624, 2010.

[11] D. Pasini, K. H. Hansen, J. Christensen, K. Agger, P. A. C. Cloos, and K. Helin, "Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and *Polycomb*-Repressive Complex 2," *Genes and Development*, vol. 22, no. 10, pp. 1345–1355, 2008.

[12] L. Ho, J. L. Ronan, J. Wu et al., "An embryonic stem cell chromatin remodeling complex, esBAF, is essential for embryonic stem cell self-renewal and pluripotency," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 13, pp. 5181–5186, 2009.

[13] O. Yildirim, R. Li, J. H. Hung et al., "Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells," *Cell*, vol. 147, no. 7, pp. 1498–1510, 2011.

[14] N. Y. Chia, Y. S. Chan, B. Feng et al., "A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity," *Nature*, vol. 468, no. 7321, pp. 316–320, 2010.

[15] S. Efroni, R. Duttagupta, J. Cheng et al., "Global transcription in pluripotent embryonic stem cells," *Cell Stem Cell*, vol. 2, no. 5, pp. 437–447, 2008.

[16] A. Gaspar-Maia, A. Alajem, F. Polesso et al., "Chd1 regulates open chromatin and pluripotency of embryonic stem cells," *Nature*, vol. 460, no. 7257, pp. 863–868, 2009.

[17] K. Agger, P. A. C. Cloos, J. Christensen et al., "UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development," *Nature*, vol. 449, no. 7163, pp. 731–734, 2007.

[18] H. Jiang, A. Shukla, X. Wang, W. Y. Chen, B. E. Bernstein, and R. G. Roeder, "Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains," *Cell*, vol. 144, no. 4, pp. 513–525, 2011.

[19] Y. S. Ang, S. Y. Tsai, D. F. Lee et al., "Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network," *Cell*, vol. 145, no. 2, pp. 183–197, 2011.

[20] A. Adamo, B. Sesé, S. Boue et al., "LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells," *Nature Cell Biology*, vol. 13, no. 6, pp. 652–659, 2011.

[21] M. H. Kagey, J. J. Newman, S. Bilodeau et al., "Mediator and cohesin connect gene expression and chromatin architecture," *Nature*, vol. 467, no. 7314, pp. 430–435, 2010.

[22] R. L. Davis, H. Weintraub, and A. B. Lassar, "Expression of a single transfected cDNA converts fibroblasts to myoblasts," *Cell*, vol. 51, no. 6, pp. 987–1000, 1987.

[23] I. Wilmut, A. E. Schnieke, J. McWhir, A. J. Kind, and K. H. S. Campbell, "Viable offspring derived from fetal and adult mammalian cells," *Nature*, vol. 385, no. 6619, pp. 810–813, 1997.

[24] K. Takahashi and S. Yamanaka, "Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors," *Cell*, vol. 126, no. 4, pp. 663–676, 2006.

[25] K. Takahashi, K. Tanabe, M. Ohnuki et al., "Induction of pluripotent stem cells from adult human fibroblasts by defined factors," *Cell*, vol. 131, no. 5, pp. 861–872, 2007.

[26] M. Nakagawa, M. Koyanagi, K. Tanabe et al., "Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts," *Nature Biotechnology*, vol. 26, no. 1, pp. 101–106, 2008.

[27] J. B. Kim, H. Zaehres, G. Wu et al., "Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors," *Nature*, vol. 454, no. 7204, pp. 646–650, 2008.

[28] D. Huangfu, K. Osafune, R. Maehr et al., "Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2," *Nature Biotechnology*, vol. 26, no. 11, pp. 1269–1275, 2008.

[29] M. Stadtfeld, M. Nagaya, J. Utikal, G. Weir, and K. Hochedlinger, "Induced pluripotent stem cells generated without viral integration," *Science*, vol. 322, no. 5903, pp. 945–949, 2008.

[30] I. H. Park, N. Arora, H. Huo et al., "Disease-specific induced pluripotent stem cells," *Cell*, vol. 134, no. 5, pp. 877–886, 2008.

[31] H. Zhou, S. Wu, J. Y. Joo et al., "Generation of induced pluripotent stem cells using recombinant proteins," *Cell Stem Cell*, vol. 4, no. 5, pp. 381–384, 2009.

[32] L. Warren, P. D. Manos, T. Ahfeldt et al., "Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA," *Cell Stem Cell*, vol. 7, no. 5, pp. 618–630, 2010.

[33] M. Ieda, J. D. Fu, P. Delgado-Olguin et al., "Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors," *Cell*, vol. 142, no. 3, pp. 375–386, 2010.

[34] T. Vierbuchen, A. Ostermeier, Z. P. Pang, Y. Kokubu, T. C. Südhof, and M. Wernig, "Direct conversion of fibroblasts to functional neurons by defined factors," *Nature*, vol. 463, no. 7284, pp. 1035–1041, 2010.

[35] E. Szabo, S. Rampalli, R. M. Risueño et al., "Direct conversion of human fibroblasts to multilineage blood progenitors," *Nature*, vol. 468, no. 7323, pp. 521–526, 2010.

[36] F. Anokye-Danso, C. M. Trivedi, D. Juhr et al., "Highly efficient miRNA-mediated reprogramming of mouse and human somatic cells to pluripotency," *Cell Stem Cell*, vol. 8, no. 4, pp. 376–388, 2011.

[37] R. Lister, M. Pelizzola, Y. S. Kida et al., "Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells," *Nature*, vol. 471, no. 7336, pp. 68–73, 2011.

[38] A. Gore, Z. Li, H. L. Fung et al., "Somatic coding mutations in human induced pluripotent stem cells," *Nature*, vol. 471, no. 7336, pp. 63–67, 2011.

[39] L. C. Laurent, I. Ulitsky, I. Slavin et al., "Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture," *Cell Stem Cell*, vol. 8, no. 1, pp. 106–118, 2011.

[40] L. A. Boyer, I. L. Tong, M. F. Cole et al., "Core transcriptional regulatory circuitry in human embryonic stem cells," *Cell*, vol. 122, no. 6, pp. 947–956, 2005.

[41] M. Pardo, B. Lang, L. Yu et al., "An expanded Oct4 interaction network: implications for stem cell biology, development, and disease," *Cell Stem Cell*, vol. 6, no. 4, pp. 382–395, 2010.

[42] D. L. C. van den Berg, T. Snoek, N. P. Mullin et al., "An Oct4-centered protein interaction network in embryonic stem cells," *Cell Stem Cell*, vol. 6, no. 4, pp. 369–381, 2010.

[43] E. Meshorer, D. Yellajoshula, E. George, P. J. Scambler, D. T. Brown, and T. Misteli, "Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells," *Developmental Cell*, vol. 10, no. 1, pp. 105–116, 2006.

[44] B. E. Bernstein, T. S. Mikkelsen, X. Xie et al., "A bivalent chromatin structure marks key developmental genes in embryonic stem cells," *Cell*, vol. 125, no. 2, pp. 315–326, 2006.

[45] K. P. Koh, A. Yabuuchi, S. Rao et al., "Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells," *Cell Stem Cell*, vol. 8, no. 2, pp. 200–213, 2011.

[46] K. Williams, J. Christensen, M. T. Pedersen et al., "TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity," *Nature*, vol. 473, no. 7347, pp. 343–348, 2011.

[47] M. M. Dawlaty, K. Ganz, B. E. Powell et al., "Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development," *Cell Stem Cell*, vol. 9, no. 2, pp. 166–175, 2011.

[48] I. M. Min, J. J. Waterfall, L. J. Core, R. J. Munroe, J. Schimenti, and J. T. Lis, "Regulating RNA polymerase pausing and transcription elongation in embryonic stem cells," *Genes and Development*, vol. 25, no. 7, pp. 742–754, 2011.

[49] M. Vermeulen, K. W. Mulder, S. Denissov et al., "Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4," *Cell*, vol. 131, no. 1, pp. 58–69, 2007.

[50] J. Simon, A. Chiang, W. Bender, M. J. Shimell, and M. O'Connor, "Elements of the *Drosophila bithorax* complex that mediate repression by *Polycomb* group products," *Developmental Biology*, vol. 158, no. 1, pp. 131–144, 1993.

[51] M. Ku, R. P. Koche, E. Rheinbay et al., "Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains," *PLoS Genetics*, vol. 4, no. 10, Article ID e1000242, 2008.

[52] A. Tanay, A. H. O'Donnell, M. Damelin, and T. H. Bestor, "Hyperconserved CpG domains underlie *Polycomb*-binding sites," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 13, pp. 5521–5526, 2007.

[53] H. Wu, A. C. D'Alessio, S. Ito et al., "Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells," *Nature*, vol. 473, no. 7347, pp. 389–393, 2011.

[54] J. Wang, S. Rao, J. Chu et al., "A protein interaction network for pluripotency of embryonic stem cells," *Nature*, vol. 444, no. 7117, pp. 364–368, 2006.

[55] G. Gargiulo, S. Levy, G. Bucci et al., "NA-Seq: a discovery tool for the analysis of chromatin structure and dynamics during differentiation," *Developmental Cell*, vol. 16, no. 3, pp. 466–481, 2009.

[56] K. Cui, C. Zang, T. Y. Roh et al., "Chromatin signatures in multipotent human hematopoietic stem cells indicate the fate of bivalent genes during differentiation," *Cell Stem Cell*, vol. 4, no. 1, pp. 80–93, 2009.

[57] T. S. Mikkelsen, M. Ku, D. B. Jaffe et al., "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells," *Nature*, vol. 448, no. 7153, pp. 553–560, 2007.

[58] R. D. Hawkins, G. C. Hon, L. K. Lee et al., "Distinct epigenomic landscapes of pluripotent and lineage-committed human cells," *Cell Stem Cell*, vol. 6, no. 5, pp. 479–491, 2010.

[59] T. Burgold, F. Spreafico, F. de Santa et al., "The histone H3 lysine 27-specific demethylase Jmjd3 is required for neural commitment," *PLoS ONE*, vol. 3, no. 8, Article ID e3034, 2008.

[60] I. Issaeva, Y. Zonis, T. Rozovskaia et al., "Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth," *Molecular and Cellular Biology*, vol. 27, no. 5, pp. 1889–1903, 2007.

[61] F. de Santa, M. G. Totaro, E. Prosperini, S. Notarbartolo, G. Testa, and G. Natoli, "The histone H3 lysine-27 demethylase Jmjd3 links inflammation to inhibition of *Polycomb*-mediated gene silencing," *Cell*, vol. 130, no. 6, pp. 1083–1094, 2007.

[62] J. M. Terme, B. Sesé, L. Millán-Ariño et al., "Histone H1 variants are differentially expressed and incorporated into chromatin during differentiation and reprogramming to pluripotency," *The Journal of Biological Chemistry*, vol. 286, no. 41, pp. 35347–35357, 2011.

[63] S. Epsztejn-Litman, N. Feldman, M. Abu-Remaileh et al., "*De novo* DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes," *Nature Structural and Molecular Biology*, vol. 15, no. 11, pp. 1176–1183, 2008.

[64] N. Feldman, A. Gerson, J. Fang et al., "G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis," *Nature Cell Biology*, vol. 8, no. 2, pp. 188–194, 2006.

[65] B. Feng, J. H. Ng, J. C. D. Heng, and H. H. Ng, "Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells," *Cell Stem Cell*, vol. 4, no. 4, pp. 301–312, 2009.

[66] D. Huangfu, R. Maehr, W. Guo et al., "Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds," *Nature Biotechnology*, vol. 26, no. 7, pp. 795–797, 2008.

[67] T. S. Mikkelsen, J. Hanna, X. Zhang et al., "Dissecting direct reprogramming through integrative genomic analysis," *Nature*, vol. 454, no. 7200, pp. 49–55, 2008.

[68] R. Sridharan, J. Tchieu, M. J. Mason et al., "Role of the murine reprogramming factors in the induction of pluripotency," *Cell*, vol. 136, no. 2, pp. 364–377, 2009.

[69] J. M. Polo, S. Liu, M. E. Figueroa et al., "Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells," *Nature Biotechnology*, vol. 28, no. 8, pp. 848–855, 2010.

[70] K. Kim, A. Doi,, B. Wen et al., "Epigenetic memory in induced pluripotent stem cells," *Nature*, vol. 467, pp. 285–290, 2010.

[71] S. Boue, I. Paramonov, M. J. Barrero, and J. C. I. Belmonte, "Analysis of human and mouse reprogramming of somatic cells to induced pluripotent stem cells. What is in the plate?" *PLoS ONE*, vol. 5, no. 9, Article ID e12664, 2010.

[72] R. P. Koche, Z. D. Smith, M. Adli et al., "Reprogramming factor expression initiates widespread targeted chromatin remodeling," *Cell Stem Cell*, vol. 8, no. 1, pp. 96–105, 2011.

[73] P.C. Taberlay, T. K. Kelly, C. C. Liu et al., "*Polycomb*-repressed genes have permissive enhancers that initiate reprogramming," *Cell*, vol. 147, no. 6, pp. 1283–1294, 2011.

[74] C. Bock, E. Kiskinis, G. Verstappen et al., "Reference maps of human es and ips cell variation enable high-throughput characterization of pluripotent cell lines," *Cell*, vol. 144, no. 3, pp. 439–452, 2011.

[75] R. Lister, M. Pelizzola, R. H. Dowen et al., "Human DNA methylomes at base resolution show widespread epigenomic differences," *Nature*, vol. 462, no. 7271, pp. 315–322, 2009.

[76] Y. Ohi, H. Qin, C. Hong et al., "Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells," *Nature Cell Biology*, vol. 13, no. 5, pp. 541–549, 2011.

[77] O. Bar-Nur, H. A. Russ, S. Efrat, and N. Benvenisty, "Epigenetic memory and preferential lineage-specific differentiation in induced pluripotent stem cells derived from human pancreatic islet beta cells," *Cell Stem Cell*, vol. 9, no. 1, pp. 17–23, 2011.

[78] M. H. Chin, M. J. Mason, W. Xie et al., "Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures," *Cell Stem Cell*, vol. 5, no. 1, pp. 111–123, 2009.

[79] A. Doi, I. H. Park, B. Wen et al., "Differential methylation of tissue-and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts," *Nature Genetics*, vol. 41, no. 12, pp. 1350–1353, 2009.

[80] F. Gonzalez, S. Boue, and J. C. I. Belmonte, "Methods for making induced pluripotent stem cells: reprogramming a la carte," *Nature Reviews Genetics*, vol. 12, no. 4, pp. 231–242, 2011.

[81] E. Szabo, S. Rampalli, R. M. Risueño et al., "Direct conversion of human fibroblasts to multilineage blood progenitors," *Nature*, vol. 468, no. 7323, pp. 521–526, 2010.

[82] T. Vierbuchen, A. Ostermeier, Z. P. Pang, Y. Kokubu, T. C. Südhof, and M. Wernig, "Direct conversion of fibroblasts to functional neurons by defined factors," *Nature*, vol. 463, no. 7284, pp. 1035–1041, 2010.

[83] M. Widschwendter, H. Fiegl, D. Egle et al., "Epigenetic stem cell signature in cancer," *Nature Genetics*, vol. 39, no. 2, pp. 157–158, 2007.

[84] V. K. Rakyan, T. A. Down, S. Maslau et al., "Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains," *Genome Research*, vol. 20, no. 4, pp. 434–439, 2010.

Hindawi Publishing Corporation
Comparative and Functional Genomics
Volume 2012, Article ID 979168, 8 pages
doi:10.1155/2012/979168

Review Article

Tackling Skeletal Muscle Cells Epigenome in the Next-Generation Sequencing Era

Raffaella Fittipaldi and Giuseppina Caretti

Department of Biomolecular Sciences and Biotechnology, University of Milano, Via Celoria 26, 20133 Milan, Italy

Correspondence should be addressed to Giuseppina Caretti, giuseppina.caretti@unimi.it

Received 3 February 2012; Accepted 3 April 2012

Academic Editor: Lucia Latella

Recent advances in high-throughput technologies have transformed methodologies employed to study cell-specific epigenomes and the approaches to investigate complex cellular phenotypes. Application of next-generation sequencing technology in the skeletal muscle differentiation field is rapidly extending our knowledge on how chromatin modifications, transcription factors and chromatin regulators orchestrate gene expression pathways guiding myogenesis. Here, we review recent biological insights gained by the application of next-generation sequencing techniques to decode the epigenetic profile and gene regulatory networks underlying skeletal muscle differentiation.

1. Introduction

During developmental processes, in response to external cues, changes in chromatin involving activation and repression of transcription factors and chromatin regulators (e.g., chromatin-modifying enzymes) underlie commitment of specific cells to defined lineages. Importantly, transcription factors and chromatin modifiers are also able to alter the cellular expression program to maintain cell identity even upon removal of the initiating differentiation stimuli [1]. During development and in adulthood, cellular identity is established and maintained by finely tuned mechanisms of gene repression and expression, preserved through rounds of cell division. Differentiation processes are achieved by sculpting cell-specific epigenomes, which establish and maintain cellular diversity [2]. Epigenetic information relies on several elements, such as histone modifications, histone variants, nucleosome accessibility, DNA methylation and hydroxymethylation, and nuclear organization. These chromatin states influence access of transcription factors and enzymes to the underlying DNA [1].

2. Epigenetic Regulators of Myogenesis

Myogenic lineage commitment and execution of the terminal differentiation program relies on the activity of the paired-box transcription factors Pax3 and Pax7 and of the muscle regulatory factors (MRFs) MyoD, Myf5, MRF4, and myogenin [3]. Myogenesis is a multistep process, which restricts cell fate and commits cells to become skeletal muscle. During embryonic development, myogenic precursor cells derived from the somites express Pax3 and Pax3/Pax7 and are capable of proliferation and self-renewal [4, 5].

In response to extrinsic signals, Pax3+ cells are committed to myoblasts and undergo terminal myogenic differentiation through the transcriptional network activated by MRFs. During postnatal life, muscle growth relies on satellite cells, which are a subpopulation of somite-derived cells that reside between myofibers and the basal lamina [5]. Satellite cells adopt a quiescent state, and upon environmental cues, such as mechanical stress, injury or in pathological environment of degenerative muscle diseases, they are activated to proliferate and terminally differentiate to regenerate muscle [3]. Satellite cells encompass a population of cells that

maintains the uncommitted state and another group of cells that are committed to the myogenic lineage and will undergo myogenic differentiation. Because of these two distinct features, satellite cells are defined as bona fide adult stem cells.

MRFs achieve the task of transcriptional activation through the participation of nonmuscle restricted transcriptional activators, such as the Mef2, Six, and Runx family members [6]. The progressive elucidation of transcription factors involved in myogenesis has revealed the strong contribution of several epigenetic regulators of skeletal muscle genes transcriptional activation: chromatin modifiers such as histone acetyltransferases and a subset of histone demethylases modify histones, promoting transcription activation. Additionally, chromatin remodeling complexes facilitate nucleosome mobility to favor the access of transcription factors to chromatin regulatory regions. At the onset of differentiation, histone acetyltransferases such as p300 and PCAF are recruited to muscle specific genes by myogenic bHLH, Mef2, and SRF proteins and exert their enzymatic activity on regulatory chromatin regions and on transcription factors such as MyoD [7, 8]. Moreover, the WDR5/Ash2L/MLL2 complex is engaged to the chromatin of Myf5 and myogenin genes by Pax7 and Mef2D, respectively, to methylate histone H3 at lysine 4 and to stimulate transcription activation [9, 10].

Moreover, other sets of epigenetic regulators exert their function by repressing gene transcription: HDACs (e.g., SIRT1 and HDAC1/3/4/5) [8] and histone methylases such as Suv39h1 and the Polycomb repressive complex 2 (PRC2) complex are recruited at inactive muscle specific genes [3, 8]. Although HDACs role in regulating transcription has been traditionally viewed as mutually exclusive to HATs function, a novel model is emerging in which HDACs and HATs cooccupy active genes [11]. HDACs task on transcribed regions is to remove acetyl groups previously added by HATs and to reset the chromatin modification state in preparation of the next round of transcription. According to this model, HATs and HDACs activities are interconnected by a dynamic interplay in regulating gene transcription [11]. Although further investigation at the genomewide level is warranted to define HATs/HDACs interplay throughout myogenesis, the knowledge that MyoD, PCAF, and the histone deacetylase Sir2 are able to form a complex and associate with chromatin [12] hints for a dynamic role of HATs and HDACs in skeletal muscle specification and maintenance, as well as in their nonhistone targets modulation. Polycomb group (PcG) proteins are a set of proteins that play a prominent role in the maintenance of cellular identity and in the regulation of developmental genes. These specialized proteins assemble in different Polycomb Repressive Complexes (PRCs) that exert their repressive function by establishing a nonpermissive chromatin structure [13]. Ezh2, Suz12, and EED are the core components of PRC2, which is responsible for the di- and tri-trimethylation of histone H3 at lysine 27. This repressive mark subsequently confers to chromatin the ability to recruit the PRC1 complex that in turn catalyzes the ubiquitination of histone H2A and further compacts chromatin [14, 15]. PcGs contribution to the preservation of cell identity is achieved through the regulation of developmental genes from different lineages, both in embryonic and in lineage-committed stem cells [16]. In embryonic stem cells, PRC2 localizes to chromatin regulatory loci of developmental regulators often coexisting with H3K4 trimethylation, a histone mark associated with a transcriptionally permissive chromatin state [17–19].

Several lines of evidence showed that enhancer of Zeste homolog 2 (Ezh2), the catalytic subunit of the PRC2 complex, is recruited to a subset of muscle-specific promoters in myoblasts [15, 20, 21] to prevent premature transcription, and it is displaced from these loci throughout the differentiation process. Furthermore, Juan et al. dissected the role of PRC2 in skeletal muscle stem cells, by exploiting satellite cells derived from mice with a Pax7-dependent Ezh2 deletion. The authors conclude that Ezh2 is essential for satellite cells self-renewal, proliferation, and cell identity. Deletion of Ezh2 leads to increased levels of the cell cycle inhibitor p16 (Cdkn2a), which affects satellite cells proliferation rate. Moreover, Ezh2-null satellite cells misexpress nonmyogenic lineage genes, normally repressed by PRC2 in skeletal satellite cells [22].

In summary, PcG role in satellite cells is to prevent ectopic differentiation, thus maintaining quiescence and proliferation states and safeguarding the mis-expression of non-myogenic genes in skeletal muscle cells. Another layer of control in myogenesis and muscle homeostasis is exerted by microRNAs (miRNAs) and long noncoding RNAs (lncRNAs) [23, 24]. miRNAs attenuate transcripts level through translation repression or mRNA degradation. Several miRNAs that modulate myogenesis have been shown to impact numerous processes in various lineages, such as miR-214 and miR-26a that target Ezh2 [25, 26]. Instead, miR-1, miR-133, and miR206 are specifically expressed in cardiac and skeletal muscle under the control of MyoD, SRF, and Mef2C transcription factors, and they regulate satellite cells proliferation and myoblast differentiation [23, 27, 28]. miR-208 and miR-499 are also specifically expressed in muscle and constitute a family of intronic miRNAs, referred to as "MyomiRs" [29]. These miRNAs are encoded by introns of the myosin genes and regulate myofibers specification. In addition, long noncoding RNAs (lncRNAs) are also emerging as regulators of the myogenic differentiation program [24, 30, 31]. Among lncRNAs, competing endogenous RNAs (ceRNAs) regulate miRNAs activity. Cesana et al., discovered that the muscle-specific lncRNA linc-MD1 acts as a natural decoy for miR-133 and miR-135, preventing them from binding to Mef2C and Mastermind-like 1 mRNAs and thus hampering miRNAs mediated negative effect on protein levels [24].

In this paper we will focus on recent findings depicting the epigenetic landscape of myoblasts cells, obtained through the application of Next Generation Sequencing (NGS) technology. The most recent results presented by three independent research groups shed new light on the genomewide epigenome of myoblast cells, focusing on histone modifications signature, the genomewide location of the MyoD transcription factor and of chromatin modifiers, for example, PcG proteins.

3. The Next-Generation Sequencing "Revolution": Brief Overview of NGS-Based Methods to Study Epigenomes

In recent years, high-throughput technologies have been developed and rapidly improved to interrogate several aspects of cellular processes. Likewise, traditional approaches employed to interrogate the epigenome have undergone profound transformation since their coupling to massive parallel sequencing. As a result, approaches that investigated restricted groups of genomic loci have been adapted to genomewise analysis, with unprecedented resolution and specificity, and with dramatically decreasing sequencing costs.

Methodologies currently employed to investigate global epigenetic signature are summarized in Table 1 and briefly described as follows.

(i) 4C-Seq (circular chromosome conformation capture) is used to map long-range chromatin interactions and relies on proximity-based DNA ligation and sequencing of chromosomal regions contacting a bait DNA region [32].

Hi-C-Seq is also employed to define chromosome architecture but it is free of bias for a bait locus [33].

(ii) ChIA-PET-Seq (chromatin interaction through DNA-binding protein) combines chromatin immunoprecipitation (ChIP), proximity ligation, pair-end tag, and deep sequencing to detect chromatin interactions mediated by a protein, at the genomewide scale [34, 35].

(iii) MNase-Seq exploits the property of micrococcal nuclease (MNase) to preferentially cleave linker region DNA and the MNase-digested DNA can therefore be used to annotate nucleosomes distribution [36]. DNase sequencing on the other hand allows to map DNase I hypersensitive sites on the genome and therefore to identify accessible chromatin regions, which are not tightly compacted by nucleosomes [37]. FAIRE-Seq (formaldehyde-assisted isolation of regulatory regions) describes regulatory sequences with a reduced nucleosome content [38].

(iv) Chip sequencing (ChIP-Seq) couples the conventional chromatin immunoprecipitation technique with massive parallel sequencing, and it is used to profile histone modifications, and map transcription factors, core transcriptional machinery and chromatin modifiers recruitment in a genomewide scale. The genomewide map of generated reads provides information on the location and the intensity of the protein-DNA interaction [19, 39].

(v) Bisulfite-Seq relies on the ability of bisulfite to convert unmethylated cytosines to uracils so that massive parallel sequencing can identify cytosine methylation status at the single base pair resolution [40]. Alternatively, ChIP-based methods can be employed to enrich for methylated DNA, exploiting antibodies raised against 5-methylcytosine (MeDIP-Seq) [41] or methylated DNA-binding proteins (MBD-Seq) [42].

(vi) Furthermore, global mRNA expression profiling can now be achieved through a deep-sequencing approach (RNA-Seq) that sequences steady-state RNA (known and novel ones) in a sample, offering larger than microarrays dynamic range and sensitivity. However, a gene list is not the final step of this approach, deeper insight is gained by bionformatics resources such as gene enrichment analysis, pathways, and regulatory network analysis [43]. One key advantage of RNA-Seq over methodologies employed in the past is that it can profile mRNAs, lncRNA, and small RNAs at the same time.

4. Interpreting Transcriptional Regulation of Myogenesis

Exploiting Chip assays coupled to massive parallel sequencing, Cao et al. identified the genomewide binding of MyoD, in C2C12 myoblast cells, satellite cells, and in mouse embryonic fibroblasts (MEFs) converted to myotubes by the stable expression of MyoD [44]. As predicted, MyoD is recruited to E-boxes of genes regulated during skeletal muscle differentiation. Unexpectedly, MyoD also constitutively binds additional nonmyogenic loci, with a high density of peaks in intergenic and intronic regions, in addition to an enriched distribution in promoter regions. Of note, the ChIP-Seq data presented show high concordance of results obtained from C2C12 myoblasts cells, primary satellite cells, and MEFs cells transduced with MyoD.

Most of MyoD bound regions regulate genes expressed both in undifferentiated C2C12 and differentiated myotubes, while, in approximately 10% of MyoD targets, the binding is associated with distinct peaks in myotubes versus proliferating cells. Therefore, MyoD constitutively binds the majority of identified sites in both proliferating and differentiated C2C12 cells. Most of MyoD binding sites do not represent classical enhancer elements, since out of 25 MyoD bound regions only 10 showed enhancer activity in a transient transfection assay. Interestingly, MyoD broad binding throughout the genome correlates with chromatin regions with overall increase in their histone H4 acetylation levels in MyoD-transduced MEF cells versus control cells. Conversely, MyoD expression in MEF cells marginally impacts the distribution of the H3K4me3 marks. Thus, MyoD may play a crucial role in broadly reprogramming the epigenetic architecture of the lineage-specified cells. The latter findings suggest that MyoD association throughout the genome leads to the recruitment of histone acetyltransferases and most likely of chromatin remodeling factors [8], possibly independently of the local transcriptional regulation of the target region. Moreover, MyoD expression may deeply affect nuclear organization and genomic architecture. Early experiments in several cell lines demonstrated that MyoD is a master regulator able to convert certain cell types but not others to skeletal muscle [45, 46]. Therefore, MyoD ability to induce skeletal muscle-specific transcripts expression in nonmyogenic lineages may rely on its ability to associate a broad number of nonmyogenic target genes and to coordinate a broad reconfiguration of chromatin states. Further Hi-C-Seq experiments in MEF cells converted by MyoD overexpression and control MEF cells may reveal global spatial reorganization of the genome, triggered by MyoD-forced expression. Furthermore, ChIA-PET experiments may reveal MyoD-mediated long-range

Table 1

Method	Method description	Genomewide data	Reference
4C-Seq, Hi-C-Seq	Chromatin proximity ligation and sequencing	3-Dimensional protein-DNA interaction Long-range chromatin interaction DNA looping	[23, 24]
ChIA-PET	Chromatin interaction analysis with paired-end tag sequencing: ChIP enrichment followed by chromatin proximity ligation and sequencing	Long-range chromatin interactions mediated by a DNA binding protein DNA looping	[25, 26]
DNase-Seq	Sequencing of DNaseI-digested DNA fragments	Chromatin accessibility and nucleosome positioning Identification of nucleosome-free regulatory regions	[28]
FAIRE-Seq	Formaldehyde-assisted isolation of regulatory elements sequencing	Chromatin accessibility Genome-wide mapping of protein-free DNA	[29]
MNase-Seq	Sequencing of micrococcal nuclease-digested DNA	Genomewide mapping of nucleosome-protected DNA nucleosome positioning	[27]
ChIP-Seq	Sequencing of DNA fragments obtained from chromatin immunoprecipitation assays	Genomewide mapping of protein-DNA interactions (a) Histone posttranslational modifications (b) Histone variants (c) Transcription factors (d) Chromatin modifiers and remodelers	[19, 30]
BS-seq	BS-Seq: bisulfite sequencing	DNA methylation	[31]
MeDIP-seq	MeDIP-Seq: methylated DNA immunoprecipitation sequencing	DNA methylation	[32]
MBD-seq	ChIP with methylated DNA-binding proteins	DNA methylation	[33]
RNA-Seq	Sequencing of RNA	Expression profiling Noncoding RNA expression Novel transcripts Alternative splicing	[34]

chromatin interactions [34, 35] and disclose how MyoD affects topological myoblast architecture.

5. Histone Modifications Signature in Myoblasts and Myotubes

Asp et al. performed an initial comprehensive description of myoblasts and myotubes histone modifications profile [47]. The authors exploited the Chip technique coupled to massively parallel sequencing (ChIP-Seq) to identify the genomewide distribution of epigenetic marks and RNA polymerase II (RNA PolII). Data from ChIP-Seq experiments were merged with expression profiles data to depict a first portrait of epigenetic landscape changes underlying C2C12 myoblasts differentiation. The authors provide evidence that histone H3 acetylation at lysines 9 and histone H4 acetylation at lysine 12 are dramatically decreased on chromatin regions of differentiating myotubes, while H3K18Ac levels are reduced on constitutively active genes and on genes whose expression decreases in myotubes. Furthermore, promoters of genes upregulated in myotubes prematurely show enriched recruitment of RNA PolII and histone marks correlating with transcriptional activation in proliferating myoblasts, suggesting that they already acquire features of active chromatin, before they reach their maximal expression levels. In addition, the authors focus their attention on genes silenced both in myoblast and myotubes. Histone H3K27me3 is largely found in promoters, gene bodies, and intergenic regions. Genes marked by H3K27me3 show inverse correlation with RNA PolII recruitment. Genes marked by H3K27me3 modification can be divided into two classes: one group of genes displays H3K27me3 mark in undifferentiated state and is corresponding to a gene ontology (GO) category of genes essential for muscle formation. Consistent with previous studies [15, 22], a significant subset of these genes exhibits a reduction in H3K27me3 mark upon differentiation. Nevertheless, another subset of genes in this cluster is transcriptionally upregulated and is capable to maintain the H3K27me3 mark throughout differentiation.

Furthermore, genes involved in nonmuscle lineage commitment can be grouped in a distinct cluster characterized by the H3K27me3 mark and a transcriptionally silent state, both in myoblasts and myotubes. Notably, mesoderm-related genes, such as genes involved in myogenesis, adipogenesis, and osteogenesis are not present in this cluster. Interestingly, as shown for other lineage committed cells [18], only a small fraction (11%) of H3K27me3 genes displays a bivalent mark in myoblasts.

Because every cell lineage displays a specific genomewide profile of histone modifications, a comprehensive description of the chromatin signature of skeletal muscle cells will represent a precious resource for scientists investigating mechanisms underlying myogenic transcriptional regulation.

6. Chromatin Modifiers: Contrasting Roles of PcG Proteins in Myogenesis

Genomewide mapping of chromatin factors has provided insights into the gene networks regulated by histone acetyltransferases, HDACs, and chromatin modifiers in ESCs and hematopoietic lineages [11, 48], which have been cellular systems at the forefront in the application of next-generation sequencing technologies (NGS). More recently, Mousavi et al. employed ChIP-Seq to investigate the genome-wide occupancy of chromatin modifiers belonging to the PcG proteins in skeletal muscle cells and explored their relation to RNA Polymerase II recruitment and the distribution of the opposing histone marks H3K27me3 and H3K4me3 [49]. As anticipated, Ezh2 recruitment substantially overlaps with the H3K27me3 marked genes in C2C12 myoblasts. Unexpectedly, Ezh1, an Ezh2 paralog that has been shown to have overlapping transcriptional repressive function in ESCs [50, 51], showed overlap with only 14% of H3K27me3 marked genes. Conversely, 51% of the regions occupied by Ezh1 are found to be associated with H3K4me3. Comparison of RNA-Seq profiles and RNA polymerase II global recruitment reveals that Ezh1 enriched regions are located on actively transcribed loci. Chip-Seq experiments with antibodies recognizing the elongating form of RNA Pol II (Phospho-serine-2 RNA Pol II) and a marker of elongation (H3K36me3) in Ezh1-depleted cells revealed that Ezh1 promotes RNA transcription, regulating the elongation step. Thus, genomewide mapping analysis unveiled a novel function for a member of PcG proteins, during the myogenic differentiation process.

7. Deep Sequencing Technologies and the Study of Skeletal Muscle Diseases

NGS technologies open novel avenues in the study of myopathies: the new methodological approach will shed new light on the biological pathways involved in skeletal muscle diseases and become an essential tool to study genetic causes of myopathologies. Furthermore, it will likely become a valuable technological approach in molecular diagnosis [52].

A recent report by Geng et al. highlights the role of ChIP-Seq technology in revealing the molecular pathways modulated by the homeobox protein DUX4, a candidate gene misexpressed in facioscapulohumeral dystrophy (FSHD) [53]. The authors performed chromatin immunoprecipitation assay followed by high-throughput sequencing in human skeletal myoblasts overexpressing the isoform of DUX4 (DUX4-fl) detectable in FSHD but not in control muscles. The set of genes identified as misregulated in DUX4-overexpressing cells has been also deregulated in FSHD skeletal muscle but not in control tissues, and this finding led authors to the identification of candidate pathways contributing to a better understanding of the disease etiology. Furthermore, this set of genes represents promising biomarkers to design novel diagnostic and treatment strategies.

NGS approaches may also represent a suitable diagnostic tool for Duchenne and Becker muscular dystrophy, because of the heterogeneity in the mutational spectrum of the disease and the large size of the dystrophin gene [52]. Therefore, NGS-rooted technologies may not only be invaluable in the discovery of human disease genes but also represent an important avenue to be pursued in molecular diagnosis of myopathies in particular, as NGS protocols have been recently extended to analyze clinical samples that are formalin-fixed and paraffin-embedded [54].

8. Conclusions and Perspectives

Comprehensive understanding of the myoblast/myotube epigenome will extend our fundamental knowledge of molecular mechanisms orchestrating biological processes, such as myogenic differentiation, muscle development, maintenance, and regeneration, in physiological and myopathological conditions. Deep sequencing methodologies will assist stem cell research toward the comprehension of mechanisms underlying muscle cells identity. This knowledge has important implication for regenerative medicine, underscoring new paradigms for satellite cells biology and offering novel targets of pharmacological intervention.

Recent advances in NGS technology allow addressing several epigenetic questions at the genomewide scale (Figure 1). To date, a restricted number of epigenetic regulators of myogenesis have been examined exploiting RNA-Seq and ChIP-Seq approaches. Although we predict that additional studies are underway, further studies are warranted to define global transcriptional targets and genomewide binding sites of transcription factors that modulate myogenesis (e.g., Pax7, Myf5, Myogenin, Mef2, SRF). A comprehensive description of their transcription targets and regulatory elements will define how individual transcription factors are connected to gene regulatory networks and how they influence each other at the genome level.

Characterization of the histone modifications landscape in skeletal muscle cells can be exploited further to annotate lineage-specific functional genomic elements, such as enhancers and insulators [55–57]. Chromatin profiling can therefore become a predictive means to define regulatory regions that are selectively functional, poised, or nonfunctional in myoblasts and myotubes or throughout mouse development. Moreover, the application of 4C/Hi-C-Seq technology and ChIA-PET in skeletal muscle cells that undergo differentiation will provide structural information on chromosomes long-range chromatin interactions and chromatin compartmentalization within the nucleus.

Potential drawbacks of NGS techniques are the amount of starting material that frequently is a limiting factor. Isolation of quiescent satellite cells by cell sorting greatly reduces the number of cells one can use for RNA-Seq or ChIP-Seq experiments. Nevertheless, new protocols for chromatin immunoprecipitation assays coupled with massively parallel sequencing have been developed and optimized for limiting sample material [58]. Furthermore, third-generation sequencing platforms are now designed to sequence DNA at the single molecule and have already been used in proof of principle epigenetic studies [59].

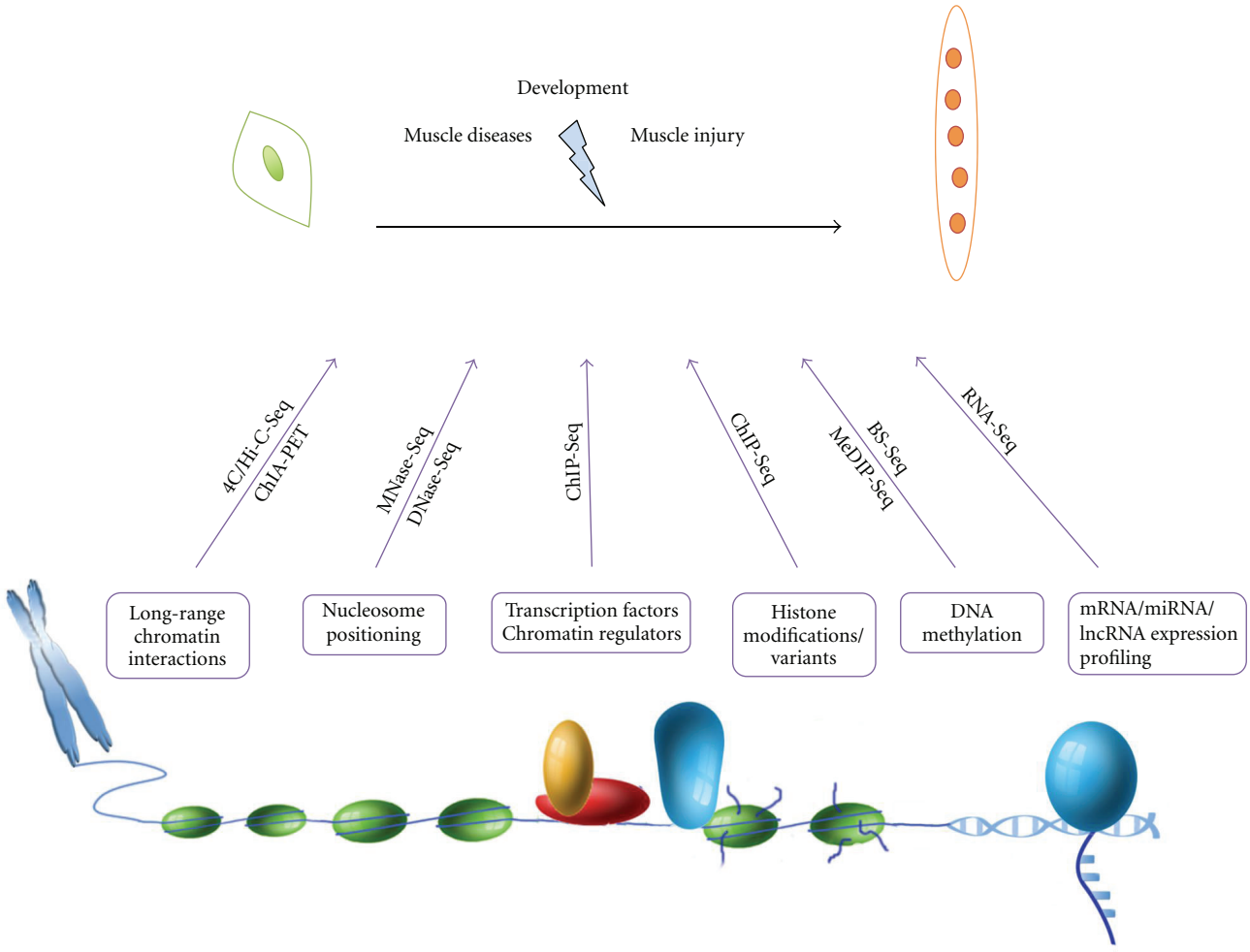


Figure 1: During development and muscle regeneration, satellite cells exit quiescence, actively proliferate, and terminally differentiate and fuse to form new muscle fibers. Within the nucleus, these steps are finely tuned by modifications in high-order chromatin structures and nucleosome accessibility, by changes in protein complexes recruited at regulatory regions, by alterations in histones marks and in the DNA methylation state. NGS offers invaluable means to explore each of these variations genomewide and to accurately identify pathways and regulatory networks underlying satellite cells activation and differentiation.

Another challenge introduced by NGS technology is the growing need for bioinformatics tools to analyze the vast amount of data generated by each experiment. Moreover, large-scale data obtained from NGS platforms need to be made available to the broad scientific community in a standardized and simple annotation form, as these data represent a precious framework to infer biological information. Bioinformatics pipelines require further refinements to enable expansion of mapping capabilities to the genome [60], improve normalization within and between samples to avoid high number of false positive differential results [61] and to better correlate results of ChIP-Seq, RNA-Seq, and whole-genome DNA methylation analysis at the single-nucleotide resolution.

Besides implementing computational methods in biological studies, successful application of NGS technologies imposes a more profound revolution related to experimental design and forces a transition from single-gene hypothesis-driven experiments to a genomewide view, in which understanding of the integration and the cross-talk of different circuitries is at the forefront. The achievements of these remarkable research capabilities lie in the combination of conventional molecular and cellular biology approaches with a systemwide genetic view, which enables us to explore so far not addressable questions and will help us to gain a more comprehensive insight into mechanisms guiding skeletal muscle cells during differentiation, its maintenance and response to pathological insult.

Acknowledgments

The support of AIRC MFAG 5386 and Marie Curie PIRGES-224833 is kindly acknowledged.

References

[1] M. J. Barrero, S. Boué, and J. C. Izpisúa Belmonte, "Epigenetic mechanisms that regulate cell identity," *Cell Stem Cell*, vol. 7, no. 5, pp. 565–570, 2010.

[2] C. Prezioso and V. Orlando, "Polycomb proteins in mammalian cell differentiation and plasticity," *FEBS Letters*, vol. 585, no. 13, pp. 2067–2077, 2011.

[3] V. Sartorelli and A. H. Juan, "Sculpting chromatin beyond the double helix: epigenetic control of skeletal myogenesis," *Current Topics in Developmental Biology*, vol. 96, pp. 57–83, 2011.

[4] M. Maroto, R. Reshef, A. E. Münsterberg, S. Koester, M. Goulding, and A. B. Lassar, "Ectopic Pax-3 activates MyoD and Myf-5 expression in embryonic mesoderm and neural tissue," *Cell*, vol. 89, no. 1, pp. 139–148, 1997.

[5] S. Tajbakhsh and G. Cossu, "Establishing myogenic identity during somitogenesis," *Current Opinion in Genetics and Development*, vol. 7, no. 5, pp. 634–641, 1997.

[6] D. Palacios and P. L. Puri, "The epigenetic network regulating muscle development and regeneration," *Journal of Cellular Physiology*, vol. 207, no. 1, pp. 1–11, 2006.

[7] V. Guasconi and P. L. Puri, "Chromatin: the interface between extrinsic cues and the epigenetic regulation of muscle regeneration," *Trends in Cell Biology*, vol. 19, no. 6, pp. 286–294, 2009.

[8] E. Perdiguero, P. Sousa-Victor, E. Ballestar, and P. Muñoz-Cánoves, "Epigenetic regulation of myogenesis," *Epigenetics*, vol. 4, no. 8, pp. 541–550, 2009.

[9] I. W. McKinnell, J. Ishibashi, F. Le Grand et al., "Pax7 activates myogenic genes by recruitment of a histone methyltransferase complex," *Nature Cell Biology*, vol. 10, no. 1, pp. 77–84, 2008.

[10] S. Rampalli, L. Li, E. Mak et al., "p38 MAPK signaling regulates recruitment of Ash2L-containing methyltransferase complexes to specific genes during differentiation," *Nature Structural and Molecular Biology*, vol. 14, no. 12, pp. 1150–1156, 2007.

[11] Z. Wang, C. Zang, K. Cui et al., "Genome-wide Mapping of HATs and HDACs Reveals Distinct Functions in Active and Inactive Genes," *Cell*, vol. 138, no. 5, pp. 1019–1031, 2009.

[12] M. Fulco, R. L. Schiltz, S. Iezzi et al., "Sir2 regulates skeletal muscle differentiation as a potential sensor of the redox state," *Molecular Cell*, vol. 12, no. 1, pp. 51–62, 2003.

[13] R. Margueron and D. Reinberg, "The Polycomb complex PRC2 and its mark in life," *Nature*, vol. 469, no. 7330, pp. 343–349, 2011.

[14] R. Eskeland, M. Leeb, G. R. Grimes et al., "Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination," *Molecular Cell*, vol. 38, no. 3, pp. 452–464, 2010.

[15] G. Caretti, M. Di Padova, B. Micales, G. E. Lyons, and V. Sartorelli, "The polycomb Ezh2 methyltransferase regulates muscle gene expression and skeletal muscle differentiation," *Genes and Development*, vol. 18, pp. 2627–2638, 2005.

[16] C. L. Fisher and A. G. Fisher, "Chromatin states in pluripotent, differentiated, and reprogrammed cells," *Current Opinion in Genetics and Development*, vol. 21, no. 2, pp. 140–146, 2011.

[17] L. A. Boyer, K. Plath, J. Zeitlinger et al., "Polycomb complexes repress developmental regulators in murine embryonic stem cells," *Nature*, vol. 441, no. 7091, pp. 349–353, 2006.

[18] T. S. Mikkelsen, M. Ku, D. B. Jaffe et al., "Genome-wide maps of chromatin state in pluripotent and lineage-committed cells," *Nature*, vol. 448, no. 7153, pp. 553–560, 2007.

[19] B. E. Bernstein, T. S. Mikkelsen, X. Xie et al., "A bivalent chromatin structure marks key developmental genes in embryonic stem cells," *Cell*, vol. 125, no. 2, pp. 315–326, 2006.

[20] D. Palacios, C. Mozzetta, S. Consalvi et al., "TNF/p38α/polycomb signaling to Pax7 locus in satellite cells links inflammation to the epigenetic control of muscle regeneration," *Cell Stem Cell*, vol. 7, no. 4, pp. 455–469, 2010.

[21] A. Blais, M. Tsikitis, D. Acosta-Alvear, R. Sharan, Y. Kluger, and B. D. Dynlacht, "An initial blueprint for myogenic differentiation," *Genes and Development*, vol. 19, no. 5, pp. 553–569, 2005.

[22] A. H. Juan, A. Derfoul, X. Feng et al., "Polycomb EZH2 controls self-renewal and safeguards the transcriptional identity of skeletal muscle stem cells," *Genes and Development*, vol. 25, no. 8, pp. 789–794, 2011.

[23] Y. Ge and J. Chen, "MicroRNAs in skeletal myogenesis," *Cell Cycle*, vol. 10, no. 3, pp. 441–448, 2011.

[24] M. Cesana, D. Cacchiarelli, I. Legnini et al., "A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA," *Cell*, vol. 147, no. 2, pp. 358–369, 2011.

[25] A. H. Juan, R. M. Kumar, J. G. Marx, R. A. Young, and V. Sartorelli, "Mir-214-Dependent Regulation of the Polycomb Protein Ezh2 in Skeletal Muscle and Embryonic Stem Cells," *Molecular Cell*, vol. 36, no. 1, pp. 61–74, 2009.

[26] C. F. Wong and R. L. Tellam, "MicroRNA-26a targets the histone methyltransferase enhancer of zeste homolog 2 during myogenesis," *Journal of Biological Chemistry*, vol. 283, no. 15, pp. 9836–9843, 2008.

[27] J. F. Chen, E. M. Mandel, J. M. Thomson et al., "The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation," *Nature Genetics*, vol. 38, no. 2, pp. 228–233, 2006.

[28] H. K. Kim, Y. S. Lee, U. Sivaprasad, A. Malhotra, and A. Dutta, "Muscle-specific microRNA miR-206 promotes muscle differentiation," *Journal of Cell Biology*, vol. 174, no. 5, pp. 677–687, 2006.

[29] E. van Rooij, D. Quiat, B. A. Johnson et al., "A family of microRNAs encoded by myosin genes governs myosin expression and muscle performance," *Developmental Cell*, vol. 17, no. 5, pp. 662–673, 2009.

[30] S. van Leeuwen and H. Mikkers, "Long non-coding RNAs: guardians of development," *Differentiation*, vol. 80, no. 4-5, pp. 175–183, 2010.

[31] G. Caretti, R. L. Schiltz, F. J. Dilworth et al., "The RNA Helicases p68/p72 and the Noncoding RNA SRA Are Coregulators of MyoD and Skeletal Muscle Differentiation," *Developmental Cell*, vol. 11, no. 4, pp. 547–560, 2006.

[32] Z. Zhao, G. Tavoosidana, M. Sjölinder et al., "Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions," *Nature Genetics*, vol. 38, no. 11, pp. 1341–1347, 2006.

[33] E. Lieberman-Aiden, N. L. Van Berkum, L. Williams et al., "Comprehensive mapping of long-range interactions reveals folding principles of the human genome," *Science*, vol. 326, no. 5950, pp. 289–293, 2009.

[34] M. J. Fullwood, M. H. Liu, Y. F. Pan et al., "An oestrogen-receptor-α-bound human chromatin interactome," *Nature*, vol. 462, no. 7269, pp. 58–64, 2009.

[35] G. Li, X. Ruan, R. K. Auerbach et al., "Extensive promoter-centered chromatin interactions provide a topological basis

for transcription regulation," *Cell*, vol. 148, no. 1-2, pp. 84–98, 2012.

[36] D. E. Schones, K. Cui, S. Cuddapah et al., "Dynamic regulation of nucleosome positioning in the human genome," *Cell*, vol. 132, no. 5, pp. 887–898, 2008.

[37] A. P. Boyle, S. Davis, H. P. Shulha et al., "High-resolution mapping and characterization of open chromatin across the genome," *Cell*, vol. 132, no. 2, pp. 311–322, 2008.

[38] P. G. Giresi and J. D. Lieb, "Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements)," *Methods*, vol. 48, no. 3, pp. 233–239, 2009.

[39] A. Barski, S. Cuddapah, K. Cui et al., "High-resolution profiling of histone methylations in the human genome," *Cell*, vol. 129, no. 4, pp. 823–837, 2007.

[40] S. J. Cokus, S. Feng, X. Zhang et al., "Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning," *Nature*, vol. 452, no. 7184, pp. 215–219, 2008.

[41] C. Bock, E. M. Tomazou, A. B. Brinkman et al., "Quantitative comparison of genome-wide DNA methylation mapping technologies," *Nature Biotechnology*, vol. 28, no. 10, pp. 1106–1114, 2010.

[42] A. B. Brinkman, F. Simmer, K. Ma, A. Kaan, J. Zhu, and H. G. Stunnenberg, "Whole-genome DNA methylation profiling using MethylCap-seq," *Methods*, vol. 52, no. 3, pp. 232–236, 2010.

[43] R. D. Hawkins, G. C. Hon, and B. Ren, "Next-generation genomics: an integrative approach," *Nature Reviews Genetics*, vol. 11, no. 7, pp. 476–486, 2010.

[44] Y. Cao, Z. Yao, D. Sarkar et al., "Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming," *Developmental Cell*, vol. 18, no. 4, pp. 662–674, 2010.

[45] H. Weintraub, R. Davis, S. Tapscott et al., "The myoD gene family: nodal point during specification of the muscle cell lineage," *Science*, vol. 251, no. 4995, pp. 761–766, 1991.

[46] H. Weintraub, S. J. Tapscott, R. L. Davis et al., "Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 86, no. 14, pp. 5434–5438, 1989.

[47] P. Asp, R. Blum, V. Vethantham et al., "Genome-wide remodeling of the epigenetic landscape during myogenic differentiation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 22, pp. E149–E158, 2011.

[48] X. Chen, H. Xu, P. Yuan et al., "Integration of external signaling pathways with the core transcriptional network in embryonic stem cells," *Cell*, vol. 133, no. 6, pp. 1106–1117, 2008.

[49] K. Mousavi, H. Zare, A. H. Wang, and V. Sartorelli, "Polycomb protein Ezh1 promotes RNA polymerase II elongation," *Molecular Cell*, vol. 45, no. 2, pp. 255–262, 2012.

[50] R. Margueron, G. Li, K. Sarma et al., "Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms," *Molecular Cell*, vol. 32, no. 4, pp. 503–518, 2008.

[51] X. Shen, Y. Liu, Y. J. Hsu et al., "EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency," *Molecular Cell*, vol. 32, no. 4, pp. 491–502, 2008.

[52] B. C. Lim, S. Lee, J.-Y. Shin et al., "Genetic diagnosis of duchenne and becker: comprehensive mutational search in a single platform," *Journal of Medical Genetics*, vol. 48, no. 11, pp. 731–736, 2011.

[53] L. N. Geng, Z. Yao, L. Snider et al., "dux4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy," *Developmental Cell*, vol. 22, pp. 38–51, 2012.

[54] H. Gu, C. Bock, T. S. Mikkelsen et al., "Genome-scale DNA methylation mapping of clinical samples at single-nucleotide resolution," *Nature Methods*, vol. 7, no. 2, pp. 133–136, 2010.

[55] N. D. Heintzman, R. K. Stuart, G. Hon et al., "Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome," *Nature Genetics*, vol. 39, no. 3, pp. 311–318, 2007.

[56] S. Cuddapah, R. Jothi, D. E. Schones, T. Y. Roh, K. Cui, and K. Zhao, "Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains," *Genome Research*, vol. 19, no. 1, pp. 24–32, 2009.

[57] C. Chu, K. Qu, F. Zhong, S. Artandi, and H. Chang, "Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions," *Molecular Cell*, vol. 44, no. 4, pp. 667–678, 2011.

[58] P. Shankaranarayanan, M. A. Mendoza-Parra, W. van Gool, L. M. Trindade, and H. Gronemeyer, "Single-tube linear DNA amplification for genome-wide studies using a few thousand cells," *Nature Protocols*, vol. 7, pp. 328–338, 2012.

[59] E. E. Schadt, S. Turner, and A. Kasarskis, "A window into third-generation sequencing," *Human Molecular Genetics*, vol. 19, no. 2, pp. R227–240, 2010.

[60] C. Trapnell, B. A. Williams, G. Pertea et al., "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation," *Nature Biotechnology*, vol. 28, no. 5, pp. 511–515, 2010.

[61] M. D. Robinson and A. Oshlack, "A scaling normalization method for differential expression analysis of RNA-seq data," *Genome Biology*, vol. 11, no. 3, article r25, 2010.