



REVIEW

OPERating ON Chromatin, a Colorful Language where Context Matters

Kathryn E. Gardner¹, C. David Allis² and Brian D. Strahl^{1*}

¹Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, The University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA

²Laboratory of Chromatin Biology and Epigenetics, The Rockefeller University, New York, NY 10065, USA

Received 14 December 2010;
received in revised form
7 January 2011;
accepted 16 January 2011
Available online
25 January 2011

Edited by M. Yaniv

Keywords:

histones;
histone code;
chromatin;
post-translational
modifications;
epigenetics

Histones, the fundamental packaging elements of eukaryotic DNA, are highly decorated with a diverse set of post-translational modifications (PTMs) that are recognized to govern the structure and function of chromatin. Ten years ago, we put forward the histone code hypothesis, which provided a model to explain how single and/or combinatorial PTMs on histones regulate the diverse activities associated with chromatin (e.g., gene transcription). At that time, there was a limited understanding of both the number of PTMs that occur on histones and the proteins that place, remove, and interpret them. Since the conception of this hypothesis, the field has witnessed an unprecedented advance in our understanding of the enzymes that contribute to the establishment of histone PTMs, as well as the diverse effector proteins that bind them. While debate continues as to whether histone PTMs truly constitute a strict “code,” it is becoming clear that PTMs on histone proteins function in elaborate combinations to regulate the many activities associated with chromatin. In this special issue, we celebrate the 50th anniversary of the landmark publication of the *lac* operon with a review that provides a current view of the histone code hypothesis, the lessons we have learned over the last decade, and the technologies that will drive our understanding of histone PTMs forward in the future.

© 2011 Elsevier Ltd. All rights reserved.

*Corresponding author. E-mail address:
brian_strahl@med.unc.edu.

Abbreviations used: PTM, post-translational modification; TAF, TBP-associated factor; MS, mass spectrometry; ChIP, chromatin immunoprecipitation; bp, base pairs; SILAC, stable isotope labeling with amino acids in cell culture; modENCODE, model organism ENCyclopedia Of DNA Elements; ChIP-chip, ChIP combined with DNA microarray analysis; ChIP-seq, ChIP coupled with next-generation sequencing technology; SNAP, SILAC nucleosome affinity purification.

“Small changes modifying the distribution in time and space of the same structures are sufficient to affect deeply the form, the functioning, and the behavior of the final product.... It is always a matter of using the same elements, of adjusting them, of altering here or there, of arranging various combinations to produce new objects of increasing complexity. It is always a matter of tinkering.”

– François Jacob, “Evolution and Tinkering” (*Science* 1977)

The adult animal was in actuality the final product that François Jacob was referring to in this eloquent

statement taken from his article “Evolution and Tinkering.”¹ Yet, as chromatin biologists, we delight in the applicability of Jacob’s quote regarding the plasticity of a single template to the chromatin landscape. However, François Jacob is not best known for his theories on how patterns of gene expression affect evolution, but rather for his seminal work with Jacques Monod establishing the basis of the *lac* operon. In celebration of the 50th anniversary of François Jacob and Jacques Monod’s landmark publication on the *lac* operon,² we are honored to contribute this piece in which we reflect on how several of the scientific themes put forward by Jacob and Monod in their historic work are widely applicable to topics as diverse as chromatin biology and the histone code hypothesis.

In simplistic terms, an operon is a functional genomic unit composed of a cluster of genes that are controlled by a single regulatory element or promoter.³ Complementary genetic and biochemical studies revealed that the basic principle underlying the *lac* operon is that the coordinated expression of the genes necessary to metabolize lactose is under the control of the *lac* repressor protein and activator protein CAP, which negatively and positively control transcription of the *lac* operon, respectively.² From the pioneering studies on the *lac* operon completed by Jacob and Monod, we now know that there are three major types of regulatory DNA sequences that function in the control of gene expression in prokaryotes: (1) promoter sequences to which RNA polymerase binds; (2) operator sequences to which transcriptional repressors bind; and (3) positive control elements to which transcriptional activator proteins bind.⁴ While the *lac* operon provides a simple yet elegant mechanism by which gene expression is controlled in prokaryotes, it is unreasonable to think that such a system would adequately provide a means by which efficient regulation of gene expression could occur in eukaryotes, where DNA must be highly compacted to fit within the confines of the nuclear space. The need for differential patterns of gene expression to specify diverse types of tissues from a single genome in multicellular organisms also calls for the existence of additional regulatory mechanisms. For example, cellular identity must be faithfully maintained through cell divisions for a lifetime, despite differentiation occurring earlier during embryonic development. The plasticity of cellular differentiation and the stability of cellular memory are thought to represent *epigenetic* phenomena wherein inherited changes in phenotype occur independently of changes in the underlying DNA sequence and without the need for *trans*-factors that establish the initial programs of coordinated gene regulation. Hence, while the historic work of Jacob and Monod reveals an elegant mechanism for prokaryotic gene regulation, it is clear that more

sophisticated means of gene regulation involving components that do more than engage the DNA template alone are necessary for processes such as cellular memory in multicellular eukaryotes.

On the basis of many insightful studies on chromosome structure, we know that in eukaryotes, DNA is assembled on a histone scaffold to form chromatin.⁵ The nucleosome core particle, or fundamental repeating unit of chromatin, consists of approximately 147 bp of DNA wrapped around an octamer containing one tetramer of histones H3 and H4 (two copies each) and two histone H2A–H2B dimers.^{5–8} Nucleosomes are packaged into progressively higher-order structures to ultimately form chromosomes. Chromatin structure largely affects DNA-templated processes such as transcription, thus necessitating that access to DNA be tightly controlled to allow factors that function in such processes to make appropriate contacts with the DNA template itself.^{5,9} Post-translational modifications (PTMs) to the histone proteins themselves can significantly affect the levels of chromatin compaction by creating generally condensed “heterochromatic” or more open “euchromatic” regions, and therefore provide a means by which rapid and localized access to DNA can be accomplished.^{10,11} Additionally, other well-studied mechanisms, such as ATP-dependent chromatin remodeling and the exchange of primary sequence histone variants, introduce meaningful variation into the chromatin polymer, “tinkering” in such a way that one relatively stable genome can give rise to the demands of multicellular development.^{12–14}

The “histone code hypothesis”: the first 10 years

In 2000, we proposed what has come to be commonly referred to as the “histone code hypothesis,” which, in its original form, posits that “multiple histone modifications, acting in a combinatorial or sequential fashion on one or multiple histone tails, specify unique downstream functions.”¹⁵ Parallels to François Jacob’s quote from “Evolution and Tinkering” are readily apparent. The same fixed set of amino acids that make up the histone proteins have the potential of being post-translationally modified within the chromatin template, where distinct spatiotemporal patterns of modifications ultimately shape functional outcome. One of the more striking phenomena predicted by such a code is that subtle variations to the same template can result in vastly different outcomes, especially in the context of regulation of gene expression.

At the time that we proposed the histone code hypothesis, we had a limited understanding of the true breadth of the number and type of PTMs that exist on histone residues either on the unstructured N-terminal tails that protrude from the nucleosomal surface or within the structured globular domains.

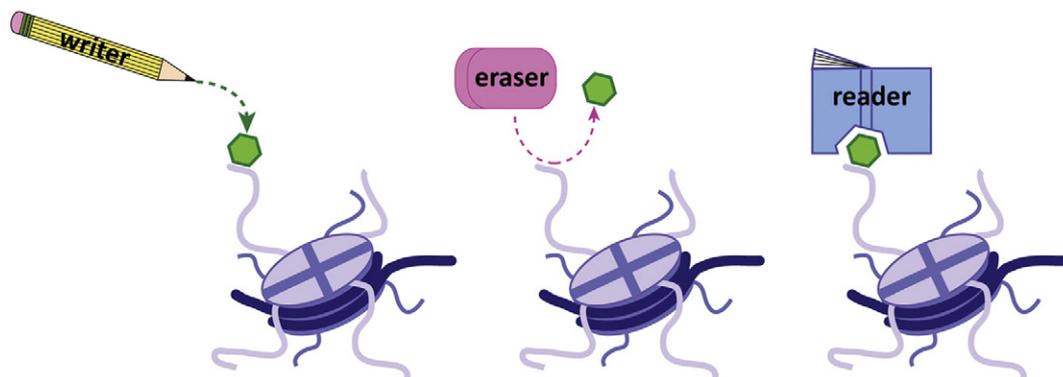
Table 1. Histone modification types and the interacting domains that “read” them

| Modification types | Residue(s) modified | Reader domain(s) |
|--------------------|---------------------|--|
| Unmodified lysine | Lysine | PHD |
| Acetylation | Lysine | Bromo |
| Methylation | Lysine/arginine | Ankyrin, Chromo, HEAT, MBT, PHD, Tudor, PWWP, WD40 |
| Phosphorylation | Serine/threonine | 14-3-3, BIR, BRCT |
| Ubiquitylation | Lysine | ? |
| Sumoylation | Lysine | ? |
| ADP-ribosylation | Lysine | ? |
| Citrullination | Arginine | ? |
| Butyrylation | Lysine | ? |
| Propionylation | Lysine | ? |
| Glycosylation | Serine/threonine | ? |

Acetylation and phosphorylation were the best-characterized modifications at that time, with multiple sites and several of the enzymes responsible for their placement and removal having been identified. However, investigations on the dynamics of histone methylation were in their infancy. Only a handful of sites modified by methylation were known, and the function of histone methylation was largely unclear, primarily because the enzyme systems responsible for the steady-state balance of methyl marks (histone methyltransferases and demethylases) were not yet identified and the intricacies associated with a modification that could exist in multiple states (mono-, di-, or trimethyl) complicated studies. Insight into other modifications was even more rudimentary. Today, we know that a number of PTMs exist, including acetylation, methylation, phosphorylation, ubiquitylation, sumoylation, ADP-ribosylation, proline isomerization, citrullination, butyrylation, propionylation, and glycosylation (Table 1).^{11,16,17} Numerous studies using both biochemical and genetic approaches have revealed many of the enzymes that are responsible for the placement or removal of

these modifications on specific amino acid residues on histones as well as non-histone proteins. While the functional significance of some of these modifications remains to be determined, the collective field of chromatin biology has made great strides toward identifying the biological consequence of the others. For example, modifications can disturb contacts between histones in contiguous nucleosomes or histones with DNA, resulting in the alteration of higher-order chromatin structure. Specifically, acetylation of lysine residues on histone tails neutralizes the basic charge of the residue on which it occurs, thereby disrupting histone contacts with other histones and/or DNA and, in turn, chromatin compaction.⁹ While it had been known that histone modifications such as methylation did not disrupt nucleosomal contacts by altering the charge of the modified residue, we now know that specialized domains within effector proteins facilitate recognition and binding to methyl marks in a defined state on specific residues to mediate downstream effects. Domains characterized thus far as being able to bind to methylated residues include chromodomains, tudor domains, PHD fingers, MBT domains, Ankyrin repeats, PWWP domains, HEAT domains, and WD40 repeats (Table 1).^{18–22} Other domains that recognize and bind to specifically modified histone forms have also been characterized. For instance, bromodomains can bind to acetylated lysine residues, whereas 14-3-3, BRCT, and BIR domains can bind to phosphorylated threonine and serine residues (Table 1).^{19,23}

The chromatin-modifying enzymes that facilitate alterations to the chromatin landscape by placing, removing, or interpreting modifications to establish variable states have recently been more generally referred to as writers, erasers, and readers, respectively, of the histone code (Fig. 1). Returning to the idea of tinkering with chromatin, we are now in a position to appreciate the true potential of a “toolkit”²⁴ of writers, erasers, and readers of the

**Fig. 1.** Toolkit for modifying the chromatin template. Schematic illustrating the concept that writers place PTMs on histone proteins (left), erasers remove such modifications from histone proteins (middle), and readers function to interpret these covalent modifications (right) to mediate diverse downstream processes.

histone code in the establishment of proper spatio-temporal patterns of modifications necessary for cellular identity and function. At defined points, writers place marks on defined histone residues, which are in turn interpreted by readers harboring specialized domains that facilitate recognition and binding to the specific mark of interest to drive the progression of a specific biological phenomenon. At a time when such signaling needs to be terminated, erasers are recruited to their defined target(s) to remove the mark, thereby ending the associated functional outcome of the previously defined reader. Admittedly, the situation is made vastly more complicated by the fact that particular amino acid residues can house more than one type of modification (this is largely true for lysine residues, which can be methylated, acetylated, ubiquitylated, or sumoylated), and that some enzymes can write, erase, or read more than one modification. Moreover, one mark can often recruit multiple effector proteins.^{25,26} Such complications, however, support the general notion of tinkering with combinatorial patterns of PTMs to control proper recruitment of effector proteins or complexes in which they reside.

We appreciate that the “histone code hypothesis,” as originally articulated by us in 2000, evolved into an influential review on the function(s) of covalent histone modifications. We acknowledge that this hypothesis, and extensions of it, rests heavily on the foundation of many biologists and biochemists who were dedicated to the general view that chromatin was going to be much more than a passive way to package the genome. However, because of the rapid pace of research in chromatin biology and the complexity associated with chromatin modifications such as those mentioned above, we must continually refine how we define the histone code. In fact, the mere existence of a code in the first place has been a point of contention.²⁷ Beyond discussions in the field as to whether a strict histone code truly exists, there is also debate over whether it is most appropriate to define it as “code” in which definite combinations lead to an absolute outcome (as exemplified by the genetic code). Some see it more in terms of a “language,” where complex combinatorial patterns of modifications form words that ultimately give rise to a vocabulary of histone crosstalk.²⁸ Others yet prefer to think of it more specifically in terms of an “epigenetic code” that is defined by combinations of histone PTMs that are predictive of, and necessary for, expression patterns of differentiation and developmental-specific genes.²⁹ On the other hand, it has been argued that histone modifications are not truly “epigenetic,” as the nature of their heritability (a requisite condition to be defined in the classical sense of epigenetic) is questionable,³⁰ thereby disputing the appropriateness of an “epigenetic code.” At some point, the question of how exactly to define the histone code

becomes somewhat rhetorical, as at their very essence, all definitions ultimately seem to convey the same fundamental principle that histone PTMs act in concert to elicit downstream biological outcomes. Here, we reflect on the many forms the “histone code hypothesis” has come to take since the time of its inception a decade ago, and suggest that individual definitions may not be mutually exclusive of one another, but are perhaps instead complementary.

Transcribing the “histone code”: chicken or egg?

Although applicable to a diverse set of cellular processes, the histone code is most commonly considered in the context of transcription regulation. Within this realm, there has been much debate as to whether a putative code formed by combinatorial modifications can formally regulate transcription itself or, rather, if patterns of modifications are generally associated with a particular transcriptional state. On one side is the argument that genes are not necessarily regulated by chromatin modifications *per se*, but rather are regulated by specific DNA-binding proteins that recruit activating and repressive complexes to genomic loci to modulate transcriptional activity. According to this line of reasoning, the histone-modifying machinery is recruited by canonical transcriptional activators and repressors (as would be defined in the classical sense by Jacob and Monod), and the placement of modifications by these enzymes then contributes to transcription by creating a more or less permissive chromatin environment for the further recruitment of downstream factors that regulate transcription. In support of this idea, it has long been known that histone acetylation is associated with active genes³¹ and functions to facilitate the disruption of higher-order chromatin structure prior to gene activation.⁹ Thus, one would argue that it is the action of the activators that directly determines transcriptional output, and that the targeting of acetylation to histones via activators that bind to specific upstream activating sequences functions to make the chromatin environment more permissive for transcriptional regulation.^{9,32} In an analogous fashion, binding of transcriptional repressors to upstream repressive sequences facilitates recruitment of histone deacetylase enzymes to chromatin, which in turn remove acetyl marks to contribute to transcriptional repression through chromatin compaction.^{32–34} By this argument, modifications are thereby associated with gene activation and/or silencing (much like RNA polymerase II is associated with active genes), but do not formally regulate transcription itself. An extension of this position would be that chromatin modifications themselves do not intrinsically regulate gene expression alone because an element of targeting or recruitment is necessary (in other words, how do the enzymes know

where to place the marks?). Once set, PTMs putatively function in transcriptional regulation by promoting or excluding the binding of elements that directly function in regulation (i.e., activators and/or repressors).

One counterargument that could be made in response to the aforementioned view of the histone code whereby chromatin-modifying machinery is recruited by transcriptional activators or repressors would be that histone modifications are a prerequisite for recruitment of certain elements of the transcriptional machinery. For example, two TBP-associated factor (TAF) subunits of the transcription factor complex TFIID have been shown to bind directly to histone PTMs, which would suggest that modification of histone proteins is necessary for binding of the transcriptional machinery. The double bromodomain of TAF1, the largest subunit of TFIID, binds preferentially to diacetylated histone H4.³⁵ TAF3 harbors a PHD finger that is selective for binding to trimethylated lysine 4 of histone H3 (H3K4me3), and loss of this chromatin mark results in reduced TFIID association with and transcriptional activity from certain promoters,³⁶ providing support for the role of histone PTMs as a requisite component in the recruitment of transcription factors.

Despite the seemingly opposite lines of reasoning regarding the role of histone modifications in transcriptional regulation, we maintain that the nature of the histone code may not necessarily be as clear-cut as histone PTMs functioning solely as a consequence of or as a prerequisite for recruitment of the canonical transcriptional machinery. It is likely that both arguments hold true in their own rights with respect to transcription (as well as other DNA-templated processes), and that possibly no absolute rule exists favoring either position over the other, thereby necessitating examination of such codes on an individual basis. It is, therefore, perhaps more judicious to focus our discussion on the histone code in the context of how it more generally contributes to the physical organization of eukaryotic genomes. Three major principles have developed during the evolution of the histone code hypothesis over the past 10 years: (1) interactions between histone modifications are not limited to a single tail; (2) a single mark can recruit more than one protein; and (3) proteins acting alone or in the context of a macromolecular complex can contain multiple domains to facilitate binding to chromatin (Fig. 2). At the time when the histone code hypothesis was put forward, we had a relatively limited scope of the existent histone PTMs, the combinations in which they exist, and how they affect downstream functionality. That marks located in close proximity to one another oftentimes exhibit functional interplay was demonstrated by examples such as phosphorylation of serine 10 of histone H3 (H3S10ph) reducing the affinity of the chromodo-

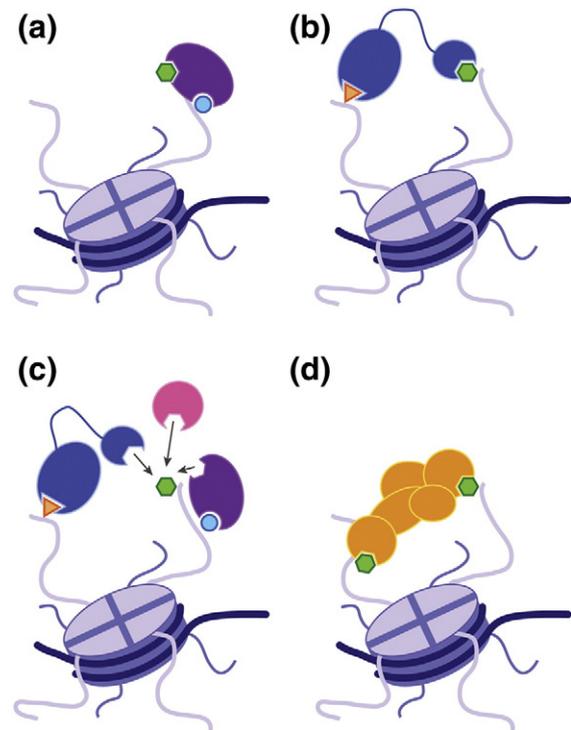


Fig. 2. Mechanisms of histone-recognition modules binding their target modification. Binding of specialized domains to histone PTMs can occur in *cis*, where contact is made to a series of modifications on the same histone tail (a), or in *trans*, where contacts are made to distinct modifications across histone tails (b). Often, a single modification can serve as a docking site for more than one protein, in which secondary signals (e.g., other PTMs) may serve to dictate which protein is recruited to the specific mark (c). Proteins acting alone (a and b) or in the context of a macromolecular complex (d) can harbor multiple domains capable of facilitating chromatin recognition and binding. For clarity, no attempts have been made to depict histone recognition between nucleosomes in either the same or distinct polynucleosome fibers, but these modes of binding recognition are also likely (reviewed by Ruthenburg *et al.*³⁷).

main of heterochromatin protein 1 for di- and trimethylated lysine 9 of histone H3.³⁸ At present, the chromatin field continually refines our understanding of how one modification affects the placement of another, especially in the context of how modifications on one histone tail affect the placement of marks and the recruitment of effector proteins on the other tails. A clear example of this idea is provided by studies that have demonstrated a signal cascade in which 14-3-3 is recruited to the enhancer of FOSL1 by binding to H3S10ph and subsequently recruits the histone acetyltransferase MOF, which acetylates histone H4 on lysine 16 (H4K16ac) to create a doubly modified H3S10ph/H4K16ac nucleosome.³⁹ These PTMs then function as a platform for the

bromodomain-containing protein BRD4 (which in turn recruits positive transcription elongation factor b) to activate transcription elongation, thus providing an elegant example of the numerous intricacies associated with interactions between multiple histone PTMs across multiple tails.³⁹ It is becoming increasingly clear that modifications that work together to form a putative code are not limited to a single histone tail, but are likely to span multiple tails within one nucleosome, between adjacent nucleosomes, or between nonadjacent nucleosomes that are physically located in close proximity to one another due to higher-order chromatin structure. Examples of histone crosstalk continue to evolve, and many more are likely to surface from future work, thereby shedding light on the growing complexity associated with the many permutations of a histone code.

As alluded to above, modifications of histone residues in defined states can serve as platforms for binding of more than one effector protein. For example, multiple proteins (including JMJD2A, Rag2, BPTF, Ing2, and TAF3) have all been demonstrated to bind to H3K4me3.^{36,40–45} Such promiscuity by a defined mark for multiple readers indicates that secondary levels of specification must exist. One possible explanation is that one protein can harbor multiple domains that cooperatively facilitate recognition and binding to chromatin.³⁷ For example, Tsai *et al.* have recently shown that the tandem PHD finger and bromodomain of the protein TRIM24, a co-activator of oestrogen receptor- α , bind combinatorially to unmodified H3K4 and acetylated H3K23 to facilitate chromatin recognition and contribute to oestrogen receptor- α -mediated transcription activation.⁴⁶ Alternatively, more than one histone PTM (or the recognition of unmodified histone residues with modified ones) can function in concert to form a recognition code for a single protein with multiple chromatin-binding domains or multiple proteins within a chromatin-associated complex.⁴⁷ One example of this type of nucleosomal interaction is provided by the Rpd3S histone deacetylase complex, which stably interacts with H3K36 methylated nucleosomes via recognition of H3K36 methylation by the chromodomain-containing subunit Eaf3 and H3 recognition by a PHD finger within the Rco1 subunit of this same complex.⁴⁸ Thus, it is becoming increasingly clear that the one mark/one reader (or writer or eraser in certain instances) ratio does not allow for generation of enough physically distinct relationships to sufficiently impart the degree of information necessary to mediate diverse outcomes, supporting the existence of numerous levels of complexity built into the histone code. Such complexity would allow multiple ways to tinker with the same chromatin landscape to promote diverse biological outcomes.

Tinkering the “histone code hypothesis” in years to come

The key question that remains then is perhaps not one of mulling over how to best define the histone code, but rather, what form will the histone code hypothesis take over the years to come? Given the rapidity of chromatin-based research and the prominent role of chromatin in numerous DNA-based processes, research in the years to come is likely to continue along the same fruitful path of discovery that it has witnessed in the past 10 years, demonstrating additional levels of complexity by which intrinsic cellular machines tinker with the chromatin template. While studies aimed at identifying additional writers and erasers of the histone code as well as novel marks remain ever important, investigations elucidating how chromatin marks act in concert to recruit readers are of equal significance. Technological advancements and new methodologies have significantly progressed our efforts in both areas of study, and are expected to continue to do so well into the future.^{49,50} Histone PTMs have traditionally been identified by metabolic labeling, microsequencing, the generation of immunological reagents, and, more recently, mass spectrometry (MS).⁵¹ Advancements in the MS technology include the recently developed top-down method, which analyzes intact proteins samples (as opposed to the more canonical bottom-up approach where proteins are fragmented prior to analysis). Because proteins are analyzed at the whole-molecule level, top-down MS allows for identification of combinatorial patterns of modifications that exist within one histone protein.⁵² For example, top-down MS analysis has now been completed on all three human histone H3 variants (H3.1, H3.2, and H3.3), revealing complex patterns of modified H3 forms.^{53,54} Additionally, analysis of asynchronously grown HeLa cells treated with the histone deacetylase inhibitor sodium butyrate has revealed a surprising and complex number of combinatorially modified species of histones H3.2 and H4.^{53,55} Although still in its infancy, studies such as these have made it readily apparent that top-down MS analysis will be a highly utilized technique in future studies to decipher how combinatorial patterns of histone modifications contribute to the regulation of diverse biological processes.⁴⁹

Identification of the histone marks themselves and the combinatorial patterns in which they exist is not enough to understand the functional consequences of their placement. The availability of modification-specific antibodies has allowed for immunoprecipitation of DNA fragments associated with a particular mark by chromatin immunoprecipitation (ChIP). It should be formally noted that one major limitation to be kept in mind when designing and/or interpreting experiments involving ChIP is the

requirement for a high-quality antibody that can specifically recognize a defined modification state (e.g., a dimethylated but not trimethylated lysine residue). Moreover, as neighboring modifications may unpredictably affect antibody specificity, it is becoming increasingly clear that rigorous validation of antibody quality is essential for any ChIP-based analysis to effectively provide insight into the location of a particular modification in a defined state.^{56–58}

Early approaches for studying chromatin modifications on a genome-wide level utilized ChIP combined with DNA microarray analysis (ChIP-chip). More recently, ChIP coupled with next-generation sequencing technology (ChIP-seq) has provided considerable insight into the function of histone PTMs, allowing for the identification of genome-wide patterns of specific modifications as well as transcription factors and the machinery responsible for modifying the chromatin landscape under defined biological conditions.⁵⁹ Early ChIP-seq analyses mapping histone modifications in CD4⁺ T cells or mouse embryonic stem cells revealed a number of findings.^{60–62} For instance, a comparative ChIP-seq analysis of mouse embryonic stem, neural progenitor, and embryonic fibroblasts confirmed the existence of bivalent domains characterized by the colocalization of H3K4 and H3K27 trimethylation that function in cellular plasticity and commitment to a defined lineage.^{61,63} However, how widespread bivalent domains occur in various developmental contexts remains unclear and is under active investigation. Genome-wide association studies derived from ChIP-seq analyses completed to date have led many to see the histone code less as sets of definite combinations that produce an absolute outcome, but rather, more as patterns of modifications that, when in combination, tend to favor a specified outcome. In that vein, the ramifications of the histone code are correlative rather than causal in that combinatorial patterns provide a bias for a specific outcome rather than serve as an absolute mark of one. ChIP-seq analyses have, for example, revealed that, in general, higher levels of H3K9me1 and H2BK5me1 in the 5' end, H3K27me1 distributed throughout, and H3K36me3 in the 3' end of a transcribed region mark actively transcribed regions.⁶⁰ Furthermore, another study found that there is a combinatorial pattern of methylation and acetylation events on histone tails that are co-associated with each other on a significant fraction of genes within the human genome.⁶² Such studies provide important insight, in that they demonstrate that actively transcribed regions of the genome, as well as functional elements in general, bear distinct histone PTM signatures.⁶⁴ Additional studies will surely expand upon whether the histone code is characterized by a fixed set of combinatorial patterns that establish defined chromatin states (also

referred to frequently as chromatin “signatures”) or, rather, if certain combinations tend to tip the balance in favor of a certain state. For example, recent work published by the modENCODE Consortium has provided great insight into the genome-wide chromatin organization in the model organisms *Caenorhabditis elegans* and *Drosophila melanogaster*, which together have vastly advanced our understanding how various histone PTMs are associated with genomic regulatory elements in defined developmental states.^{65–67} Newcomers to this field should refer to these studies to become oriented not only to some of the principal PTMs that mark chromatin domains but also to the staggering complexities underlying the combinatorial nature with which gene bodies and regulatory elements are specified and defined in a chromatin context. Indeed, the language is colorful and must be interpreted in context, especially in a developmental setting.

Novel methods are also being developed to characterize combinatorial patterns that facilitate binding of effector proteins as well as identify novel proteins that can bind to modified histone tails. Use of combinatorial peptide libraries based on the N-terminal histone tails has become a widely used practice to identify how the presence of additional marks enhances or weakens the affinity of an effector protein for its target binding module. Peptide libraries have been synthesized as various types of platforms, including resin-bound PTM-containing histone tail libraries and custom peptide microarrays.^{56,58,68,69} Such platforms have recently begun to be used to identify synergistic and antagonistic combinations of histone modifications that ultimately affect the binding of effectors. For example, the H3K9me2 demethylase PHF8 binds to H3K4me3/2, and hybridization of a recombinant GST-PHF8 fusion protein to a synthetic peptide array containing combinatorial modifications patterns revealed that binding to H3K4me3/2 was also achieved when peptides were acetylated at the H3K9/K14 positions.⁷⁰ While peptide libraries are advantageous for looking at how effector proteins respond to various combinatorial patterns of modifications, alternative functional technologies are being employed to screen for proteins that bind to a particular modification in an unbiased manner. Recently, a histone peptide pull-down approach paired with the SILAC proteomics technology was used to define a large-scale methyl lysine interactome.⁷¹ Extending this concept further, designer synthetic nucleosomes in which nucleosomes are reconstituted using recombinant histones harboring specific modifications states have allowed for unbiased identification of cellular proteins that bind to a specific state on a nucleosomal substrate in a technique called SNAP (SILAC nucleosome affinity purification).⁷² Because the DNA sequence and modifications of interest are user defined, one

could theoretically begin to make oligonucleosomes in which crosstalk both within and across nucleosomes can be addressed. This latter technology holds great potential for future studies in which peptides harboring several modifications are fused to multiple histone tails by native chemical ligation to reconstitute multiply modified nucleosomes to give a more complete picture of how combinatorial patterns affect binding by chromatin readers in the more physiologically relevant nucleosomal context.

Strict code versus rich language: exciting either way

At the time of inception, it is always difficult to discern how influential a hypothesis will truly be. We have been privileged to witness that François Jacob and Jacques Monod's report on the *lac* operon in the *Journal of Molecular Biology* in 1961 has revolutionized our understanding of the basic mechanisms underlying gene regulation. We are also beginning to understand the richness of the histone code hypothesis. When we posited this hypothesis, now 10 years ago, we had what in retrospect would be described as a quite limited scope of histone PTMs. One decade later, we stand in awe at how the chromatin field, and the scientific research community at large, has come together to expand this code to a scope beyond what was imaginable at the time of its conception. For example, never in our wildest dreams had we envisioned a Keystone Meeting being dedicated to the singular topic of the "Histone Code": Fact or Fiction (January 10–15, 2011 in Midway, Utah). However, it is with a sense of realism that we recognize that many obstacles remain to be overcome before we can officially declare that this code has been deciphered to its fullest potential. For example, it will be difficult to discern when saturation has been reached and all modifications have been identified, a reality complicated by the fact that organismal differences exist within the chromatin landscape. The staggering complexity of this proposed "epigenetic code" promises to keep many talented scientists busy for the next decade with many more welcome surprises along the way. Moreover, we are coming to realize that such a code may not pertain specifically to histones, but could potentially be extended to proteins in general. That proteins are modified post-translationally is by no means a novel concept, but the idea that modifications working in concert are predictive of defined downstream biological events has received more thought recently. The tumor suppressor p53 is highly regarded as the model for the existence of a more general protein code, as this protein is subject to a number of PTMs, including methylation, acetylation, phosphorylation, and ubiquitylation.⁷³ The observation that modifications, such as acetyla-

tion, correlate with the stabilization and activation of p53,⁷⁴ in concert with the idea that one modification can enhance or preclude the placement of another, supports a more general mechanism in which modifications are tightly linked to p53 function in an analogous fashion as to how histone PTMs work together to form a functional code. Also worthy of noting is that many of the enzymes responsible for writing, erasing, and reading histone methylation and acetylation on histone proteins are also responsible for modifying the C-terminus of p53 and certainly other non-histone proteins,^{75,76} echoing Jacob's visionary sentiment that the same elements are often used to create new products of increasing complexity.

Our piece in 2000 was framed as a hypothesis with the hope that it would stimulate discussion and lead to subsequent tests of its central tenets. Much of this has happened, and we look forward to much more along these lines. While contention over use of the word "code" may eventually lead to an alternative designation in future years, we are confident that debates over diction will not hinder the elegant work that the chromatin community has collectively produced at a remarkable pace. We close with a prediction: We will indeed witness a period of further enlightenment with regard to how cellular enzymes tinker with both histone and non-histone proteins alike to create increasingly complex patterns of regulatory mechanisms in the years to come. Coloring the chromatin code with even more shades will be part of the fun.^{77,78}

Acknowledgements

We thank the many researchers whose studies have helped to expand our understanding of both the *lac* operon and the histone code, and we apologize to those whose work could not be cited here due to space constraints. We also thank Nara Lee, Scott Rothbart, and the members of the Allis laboratory for insightful conversations and comments on the manuscript, and Nara Lee and Stephen Fuchs for assistance with the illustrations contained in this piece.

References

1. Jacob, F. (1977). Evolution and tinkering. *Science*, **196**, 1161–1166.
2. Jacob, F. & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J Mol Biol*, **3**, 318–356.
3. Jacob, F., Perrin, D., Sanchez, C. & Monod, J. (1960). [Operon: a group of genes with the expression

- coordinated by an operator.]. *C R Hebd Seances Acad. Sci.* **250**, 1727–1729.
4. Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell*, **98**, 1–4.
 5. Kornberg, R. D. & Lorch, Y. (1999). Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell*, **98**, 285–294.
 6. Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature*, **389**, 251–260.
 7. Kornberg, R. D. (1974). Chromatin structure: a repeating unit of histones and DNA. *Science*, **184**, 868–871.
 8. Oudet, P., Gross-Bellard, M. & Chambon, P. (1975). Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. *Cell*, **4**, 281–300.
 9. Wolffe, A. P. & Hayes, J. J. (1999). Chromatin disruption and modification. *Nucleic Acids Res.* **27**, 711–720.
 10. Berger, S. L. (2007). The complex language of chromatin regulation during transcription. *Nature*, **447**, 407–412.
 11. Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, **128**, 693–705.
 12. Clapier, C. R. & Cairns, B. R. (2009). The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* **78**, 273–304.
 13. Ho, L. & Crabtree, G. R. (2010). Chromatin remodeling during development. *Nature*, **463**, 474–484.
 14. Talbert, P. B. & Henikoff, S. (2010). Histone variants—ancient wrap artists of the epigenome. *Nat. Rev. Mol. Cell Biol.* **11**, 264–275.
 15. Strahl, B. D. & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, **403**, 41–45.
 16. Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S. C. *et al.* (2007). Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell Proteomics*, **6**, 812–819.
 17. Sakabe, K., Wang, Z. & Hart, G. W. (2010). β-N-Acetylglucosamine (O-GlcNAc) is part of the histone code. *Proc. Natl Acad. Sci. USA*, **107**, 19915–19920.
 18. Collins, R. E., Northrop, J. P., Horton, J. R., Lee, D. Y., Zhang, X., Stallcup, M. R. & Cheng, X. (2008). The ankyrin repeats of G9a and GLP histone methyltransferases are mono- and dimethyllysine binding modules. *Nat. Struct. Mol. Biol.* **15**, 245–250.
 19. Taverna, S. D., Li, H., Ruthenburg, A. J., Allis, C. D. & Patel, D. J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **14**, 1025–1040.
 20. Vezzoli, A., Bonadies, N., Allen, M. D., Freund, S. M., Santiveri, C. M., Kvinlaug, B. T. *et al.* (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. *Nat. Struct. Mol. Biol.* **17**, 617–619.
 21. Wang, Y., Reddy, B., Thompson, J., Wang, H., Noma, K., Yates, J. R., III & Jia, S. (2009). Regulation of Set9-mediated H4K20 methylation by a PWWP domain protein. *Mol. Cell*, **33**, 428–437.
 22. Liu, W., Tanasa, B., Tyurina, O. V., Zhou, T. Y., Gassmann, R., Liu, W. T. *et al.* (2010). PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature*, **466**, 508–512.
 23. Kelly, A. E., Ghenoïu, C., Xue, J. Z., Zierhut, C., Kimura, H. & Funabiki, H. (2010). Survivin reads phosphorylated histone H3 threonine 3 to activate the mitotic kinase Aurora B. *Science*, **330**, 235–239.
 24. Lim, W. A. & Pawson, T. (2010). Phosphotyrosine signaling: evolving a new cellular communication system. *Cell*, **142**, 661–667.
 25. Ruthenburg, A. J., Allis, C. D. & Wysocka, J. (2007). Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell*, **25**, 15–30.
 26. Sims, R. J., III & Reinberg, D. (2006). Histone H3 Lys 4 methylation: caught in a bind? *Genes Dev.* **20**, 2779–2786.
 27. Smith, E. & Shilatifard, A. (2010). The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. *Mol. Cell*, **40**.
 28. Lee, J. S., Smith, E. & Shilatifard, A. (2010). The language of histone crosstalk. *Cell*, **142**, 682–685.
 29. Turner, B. M. (2007). Defining an epigenetic code. *Nat. Cell Biol.* **9**, 2–6.
 30. Ptashne, M. (2007). On the use of the word ‘epigenetic.’ *Curr. Biol.* **17**, R233–236.
 31. Allfrey, V. G., Faulkner, R. & Mirsky, A. E. (1964). Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl Acad. Sci. USA*, **51**, 786–794.
 32. Struhl, K. (1998). Histone acetylation and transcriptional regulatory mechanisms. *Genes Dev.* **12**, 599–606.
 33. Pazin, M. J. & Kadonaga, J. T. (1997). What’s up and down with histone deacetylation and transcription? *Cell*, **89**, 325–328.
 34. Katan-Khaykovich, Y. & Struhl, K. (2002). Dynamics of global histone acetylation and deacetylation *in vivo*: rapid restoration of normal histone acetylation status upon removal of activators and repressors. *Genes Dev.* **16**, 743–752.
 35. Jacobson, R. H., Ladurner, A. G., King, D. S. & Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. *Science*, **288**, 1422–1425.
 36. Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A. *et al.* (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*, **131**, 58–69.
 37. Ruthenburg, A. J., Li, H., Patel, D. J. & Allis, C. D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **8**, 983–994.
 38. Fischle, W., Tseng, B. S., Dormann, H. L., Ueberheide, B. M., Garcia, B. A., Shabanowitz, J. *et al.* (2005). Regulation of HP1–chromatin binding by histone H3 methylation and phosphorylation. *Nature*, **438**, 1116–1122.
 39. Zippo, A., Serafini, R., Rocchigiani, M., Pennacchini, S., Krepelova, A. & Oliviero, S. (2009). Histone crosstalk between H3S10ph and H4K16ac generates a histone code that mediates transcription elongation. *Cell*, **138**, 1122–1136.
 40. Huang, Y., Fang, J., Bedford, M. T., Zhang, Y. & Xu, R. M. (2006). Recognition of histone H3 lysine-4

- methylation by the double tudor domain of JMJD2A. *Science*, **312**, 748–751.
41. Li, H., Ilin, S., Wang, W., Duncan, E. M., Wysocka, J., Allis, C. D. & Patel, D. J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. *Nature*, **442**, 91–95.
 42. Matthews, A. G., Kuo, A. J., Ramon-Maiques, S., Han, S., Champagne, K. S., Ivanov, D. *et al.* (2007). RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature*, **450**, 1106–1110.
 43. Pena, P. V., Davrazou, F., Shi, X., Walter, K. L., Verkhusha, V. V., Gozani, O. *et al.* (2006). Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature*, **442**, 100–103.
 44. Shi, X., Hong, T., Walter, K. L., Ewalt, M., Michishita, E., Hung, T. *et al.* (2006). ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature*, **442**, 96–99.
 45. Wysocka, J., Swigut, T., Xiao, H., Milne, T. A., Kwon, S. Y., Landry, J. *et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature*, **442**, 86–90.
 46. Tsai, W. W., Wang, Z., Yiu, T. T., Akdemir, K. C., Xia, W., Winter, S. *et al.* (2010). TRIM24 links a non-canonical histone signature to breast cancer. *Nature*, **468**, 927–932.
 47. Oliver, S. S. & Denu, J. M. (2011). Dynamic interplay between histone H3 modifications and protein interpreters: emerging evidence for a “Histone Language”. *ChemBiochem*, **12**, 299–307.
 48. Li, B., Gogol, M., Carey, M., Lee, D., Seidel, C. & Workman, J. L. (2007). Combined action of PHD and chromo domains directs the Rpd3S HDAC to transcribed chromatin. *Science*, **316**, 1050–1054.
 49. Young, N. L., Dimaggio, P. A. & Garcia, B. A. (2010). The significance, development and progress of high-throughput combinatorial histone code analysis. *Cell. Mol. Life. Sci.* **67**, 3983–4000.
 50. Voigt, P. & Reinberg, D. (2011). Histone tails: ideal motifs for probing epigenetics through chemical biology approaches. *ChemBiochem*, **12**, 236–252.
 51. Garcia, B. A., Shabanowitz, J. & Hunt, D. F. (2007). Characterization of histones and their post-translational modifications by mass spectrometry. *Curr. Opin. Chem. Biol.* **11**, 66–73.
 52. Siuti, N. & Kelleher, N. L. (2007). Decoding protein modifications using top-down mass spectrometry. *Nature Methods*, **4**, 817–821.
 53. Garcia, B. A., Pesavento, J. J., Mizzen, C. A. & Kelleher, N. L. (2007). Pervasive combinatorial modification of histone H3 in human cells. *Nature Methods*, **4**, 487–489.
 54. Thomas, C. E., Kelleher, N. L. & Mizzen, C. A. (2006). Mass spectrometric characterization of human histone H3: a bird’s eye view. *J. Proteome Res.* **5**, 240–247.
 55. Young, N. L., DiMaggio, P. A., Plazas-Mayorca, M. D., Baliban, R. C., Floudas, C. A. & Garcia, B. A. (2009). High throughput characterization of combinatorial histone codes. *Mol. Cell. Proteomics*, **8**, 2266–2284.
 56. Bock, I., Dhayalan, A., Kudithipudi, S., Brandt, O., Rathert, P. & Jeltsch, A. (2011). Detailed specificity analysis of antibodies binding to modified histone tails with peptide arrays. *Epigenetics*, **6**.
 57. Egelhofer, T. A., Minoda, A., Klugman, S., Lee, K., Kolasinska-Zwierz, P., Alekseyenko, A. A. *et al.* (2011). An assessment of histone-modification antibody quality. *Nat. Struct. Mol. Biol.* **18**, 91–93.
 58. Fuchs, S. M., Krajewski, K., Baker, R. W., Miller, V. L. & Strahl, B. D. (2011). Influence of combinatorial histone modifications on antibody and effector protein recognition. *Curr. Biol.* **21**, 53–58.
 59. Schones, D. E. & Zhao, K. (2008). Genome-wide approaches to studying chromatin modifications. *Nat. Rev. Genet.* **9**, 179–191.
 60. Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z. *et al.* (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
 61. Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G. *et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*, **448**, 553–560.
 62. Wang, Z., Zang, C., Rosenfeld, J. A., Schones, D. E., Barski, A., Cuddapah, S. *et al.* (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40**, 897–903.
 63. Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J. *et al.* (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*, **125**, 315–326.
 64. Zhou, V. W., Goren, A. & Bernstein, B. E. (2011). Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18.
 65. Gerstein, M. B., Lu, Z. J., Van Nostrand, E. L., Cheng, C., Arshinoff, B. I., Liu, T. *et al.* (2010). Integrative analysis of the *Caenorhabditis elegans* genome by the modENCODE Project. *Science*, **330**, 1775–1787.
 66. Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J. *et al.* (2010). Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature*. Dec 22. [Epub ahead of print].
 67. modENCODE Consortium, Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L. *et al.* (2010). Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science*, **330**, 1787–1797.
 68. Bua, D. J., Kuo, A. J., Cheung, P., Liu, C. L., Migliori, V., Espejo, A. *et al.* (2009). Epigenome microarray platform for proteome-wide dissection of chromatin-signaling networks. *PLoS One*, **4**, e6789.
 69. Garske, A. L., Oliver, S. S., Wagner, E. K., Musselman, C. A., LeRoy, G., Garcia, B. A. *et al.* (2010). Combinatorial profiling of chromatin binding modules reveals multisite discrimination. *Nat. Chem. Biol.* **6**, 283–290.
 70. Kleine-Kohlbrecher, D., Christensen, J., Vandamme, J., Abarrategui, I., Bak, M., Tommerup, N. *et al.* (2010). A functional link between the histone demethylase PHF8 and the transcription factor ZNF711 in X-linked mental retardation. *Mol. Cell.* **38**, 165–178.
 71. Vermeulen, M., Eberl, H. C., Matarese, F., Marks, H., Denissov, S., Butter, F. *et al.* (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell*, **142**, 967–980.
 72. Bartke, T., Vermeulen, M., Xhemalce, B., Robson, S. C., Mann, M. & Kouzarides, T. (2010). Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell*, **143**, 470–484.

-
73. Sims, R. J., III & Reinberg, D. (2008). Is there a code embedded in proteins that is based on post-translational modifications? *Nat. Rev. Mol. Cell. Biol.* **9**, 815–820.
 74. Bode, A. M. & Dong, Z. (2004). Post-translational modification of p53 in tumorigenesis. *Nat. Rev. Cancer.* **4**, 793–805.
 75. Glozak, M. A., Sengupta, N., Zhang, X. & Seto, E. (2005). Acetylation and deacetylation of non-histone proteins. *Gene*, **363**, 15–23.
 76. Huang, J. & Berger, S. L. (2008). The emerging field of dynamic lysine methylation of non-histone proteins. *Curr. Opin. Genet. Dev.* **18**, 152–158.
 77. Filion, G. J., van Bommel, J. G., Braunschweig, U., Talhout, W., Kind, J., Ward, L. D. *et al.* (2010). Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell*, **143**, 212–224.
 78. Schubeler, D. (2010). Chromatin in multicolor. *Cell*, **143**, 183–184.