Published in final edited form as:

Epigenomics. 2013 April ; 5(2): 113–116. doi:10.2217/epi.13.13.

The correlation between histone modifications and gene expression

Xianjun Dong and

Program in Bioinformatics and integrative Biology, University of Massachusetts Medical School,
364 Plantation Street, Worcester MA 01605, USA

Zhiping Weng

Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School,
364 Plantation Street, Worcester, MA 01605, USA

Keywords

chromatin state; epigenetics; gene expression; histone modification

“...transcriptional regulation is a complex and dynamic process of which histone modifications are a key component.”

In the nuclei of eukaryotic cells, DNA wraps around the octamer of histone proteins to form the nucleosome, in a structure like ‘beads on a string’, which makes up the basic unit of chromatin. Chromatin further folds into higher-level structures, loosely or tightly, which helps to determine the accessibility of the DNA. For instance, actively transcribed regions tend to be in looser chromatin structures so that transcription factors and RNA polymerases can access the genes. Chromatin structure can be altered by various post-translational modifications of the N-terminal tail residues of histone proteins. For example, acetylation of a lysine residue can neutralize its positive charge and weaken the binding between the histone and the negatively charged DNA, which exposes the DNA to regulatory proteins. Methylation is another common type of histone modification; for example, the lysine at the fourth position of the H3 histone can be mono-, di- or tri-methylated (denoted as H3K4me1, H3K4me2 and H3K4me3, respectively).

By examining histone modification patterns at highly conserved noncoding regions in mouse embryonic stem cells, Bernstein *et al.* found ‘bivalent domains’ of histone modifications (i.e., harboring both the repressive mark H3K27me3 and the active mark H3K4me3) near genes with poised transcription [1]. When embryonic stem cells differentiate into more specialized cells (e.g., neural precursor cells), a subset of the bivalent domains are resolved (i.e., H3K27me3 becomes weaker, while H3K4me3 becomes stronger, and these loci

© 2013 Future Medicine Ltd

Author for Correspondence: zhiping.weng@umassmed.edu.

Financial & competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

coincide with genes that are actively transcribed in neural precursor cells). Thus, combinations of histone marks are indicative of transcriptional states.

Barski *et al.* mapped 20 histone methylations of lysine and arginine residues in human CD4⁺ T cells using chromatin immunoprecipitation followed by sequencing (ChIP-seq) [2]. They found that monomethylated H3K27, H3K9, H4K20, H3K79 and H2BK5 were linked to gene activation, while trimethylated H3K27, H3K9 and H3K79 were linked to gene repression. In a later study, the group profiled 39 additional histone modifications in human CD4⁺ T cells [3]. They identified more than 3000 genes that were highly expressed in these cells and the promoters of these genes showed high levels of 17 histone modifications (called a his-tone modification module). Other studies also investigated the correlation between individual histone marks and gene expression, although not in a quantitative way [4,5].

“Some histone marks are established as consequences of active transcription; however, these histone marks provide memory of recent transcriptional activity and signal for additional transcriptional regulation.”

Recent studies derived quantitative models for correlating various chromatin marks with gene expression. A genome-wide analysis of H3K4me3 and H3K27me3 showed that H3K4me3 correlated with gene expression in mouse naïve T cells, and two types of cytotoxic T cells, while H3K27me3 anti-correlated with expression [6]. Karli *et al.* [7] built a quantitative model to predict gene expression using histone modification data from the above human T-cell studies [2,3]. They concluded that a small number of histone marks could predict gene expression almost as accurately as using all 38 marks and one histone variant (e.g., correlation coefficient $r = 0.75$ for the H3K27ac+H3K4me1+H3K20me1 three-modification model vs $r = 0.77$ for the full model). They also showed that different sets of chromatin marks performed best for predicting the expressions of genes with high- and low-CpG promoters – H3K27ac and H4K20me1 were the most predictive marks for HCP genes, while H3K4me3 and H3K79me1 were the most predictive for LCP genes. They also showed that the model trained on CD4⁺ T cells could predict expression in CD36⁺ cells and CD133⁺ cells, indicating that the correlation between histone modifications and gene expression is general, rather than cell-type-specific.

“Many histone marks that are not coupled active transcription function in transcriptional regulation processes.”

Cheng *et al.* developed support vector regression (SVR) models that used histone modifications to predict gene expression [8]. They showed a strong correlation ($r = 0.75$) between predicted and measured expression values for modEN-CODE data in worms. They also showed that the model learned from one species performed well in another species. Recent data from the ENCODE project [9] enabled our laboratory to further explore this relationship in a comprehensive manner. In addition to working on many cell lines, ENCODE production groups measured transcriptional activity with multiple techniques (e.g., cap analysis of gene expression [CAGE], RNA-Seq and RNA-paired-end tags [PET]) by extracting RNA using different protocols (e.g., polyA⁺, polyA⁻) from different cellular compartments (e.g., whole cell, nuclear, cytosolic etc.). Taking advantage of the rich

ENCODE data, we studied the relationships between chromatin features and transcriptional activities in various cellular contexts [10]. We developed a novel two-step model combining classification with regression, through which we could not only verify the general relationship found in previous studies, but also better reveal the chromatin marks differing in predicting the 'on/off' status or the dynamic range of expression. By comparing the top chromatin features in predicting expression from CAGE and RNA-Seq, we confirmed that transcription initiation and transcription elongation were characterized by different groups of chromatin features. We also showed that polyA⁺ RNAs were significantly better predicted than polyA⁻ RNAs, which suggests they might be regulated by different mechanisms.

Although the predictive relationship between chromatin features and gene expression can be accurately modeled, there are several outstanding questions: is gene expression the cause or consequence of histone modification? What are the protein factors that connect histone modifications to gene expression? In a recent article, Henikoff and Shilatifard argued that the correlation between histone modification and gene expression did not equate causation and, to the contrary, that histone modifications were more likely the consequences than the causes of transcription [11]. The best examples of histone modifications that result from active transcription are H3K4me3 and H3K36me3, which are methylated by two histone lysine methyltransferases, Set1 (as a component of the COMPASS complex) and Set2, respectively, in yeast. Set1 binds to the heptapeptide repeat in the C-terminal domain of the largest subunit of RNA polymerase II during the early elongation phase, when the serine residue at position 5 (ser5) of the heptapeptide is phosphorylated, but ser2 is unphosphorylated [12]. In contrast, Set2 binds to the same heptapeptide repeat of RNA polymerase II during the late elongation phase when ser2 is phosphorylated, but ser5 is unphosphorylated [13]. Both Set1 and Set2 are recruited by polymerase-associated factors (e.g., the Paf1 protein complex in yeast), while these factors travel with RNA polymerase II during elongation [14,15]. Accordingly, the level of H3K4me3 peaks at the 5'-ends of actively transcribed genes and tapers off toward the 3'-ends of the genes, while the H3K36me3 level starts low at the 5'-ends of expressed genes and gradually increases toward the 3'-ends [16]. H3K79me3 is another histone mark coupled with active transcription in a manner highly similar to H3K4me3. H3K79me3 is catalyzed by another histone methyltransferase Dot1p; the process is dependent on the Paf1 protein complex [14] and, accordingly, H3K79me3 level peaks at the 5'-end and decreases toward the 3'-end. Moreover, both H3K79me3 and H3K4me3 depend on ubiquitination of H2B on lysine 123 [14,17].

But do histone modifications have any regulatory impact on transcription? To answer this question, one approach is to mutate, delete or knock down the genes that encode the histone-modifying enzymes or to inhibit these enzymes with small molecules. The results of this approach can be difficult to interpret, because these enzymes often modify additional non-histone proteins and multiple enzymes can modify the same histone residue. A more direct approach is to mutate the histone residue that bears the modification; however, this approach is technically difficult in higher eukaryotes that have many copies of each histone gene. Carrozza *et al.* combined both approaches to study the function of H3K36me3 [18]. They found that the H3K36me3 that is catalyzed by Set2 inside gene bodies is recognized by

the Rpd3 histone deacetylase in the Rpd3S complex in yeast. Consequently, Rpd3 erases histone acetylation associated with transcriptional elongation. Both deletion of the *set2* gene and mutation of H3K36 to alanine caused increased acetylation at actively transcribed open reading frames and led to spurious initiation from cryptic start sites within these regions [18].

Among the histone modifications whose levels anti-correlate with gene expression, H3K9me3 is catalyzed by the Su(var)3–9 family of histonespecific methyltransferase (SUV39H in human and Suv39h in mouse) [19], and H3K27me3 is catalyzed by Polycomb repressive complex 2 (PRC2) [20,21]. Suv39h catalyzes H3K9me3 at pericentric repeats and then heterochromatin protein 1 (HP1) binds to the trimethylated H3K9 residues and recruits DNA methyltransferase 3b (Dnmt3b) to establish heterochromatin in these regions. *Suv39h* null mouse embryonic stem cells exhibit an altered DNA methylation profile in pericentric regions, which gives rise to transcripts across major satellite repeats [22]. E(z) is the catalytic subunit of PRC2 that specifically methylates H3K27 in chromatin and represses gene expression. In *Drosophila*, all 23 copies of canonical histone genes are clustered in the *HisC* locus in the genome. Mutant flies without *HisC* but complemented with a transgene cassette that contains 12 copies of histone genes are phenotypically normal [23]. Pengelly *et al.* modified the transgene cassette so that all H3K 27 residues were mutated to arginine. The *H3K27R* mutant showed very similar phenotypes as the *E(z)* mutant; in particular, several Polycomb group genes (*Ubx*, *Abd-B*, *Scr* and *en*) were mis-expressed and the differentiated mutant cells show homeotic transformations [24].

In summary, transcriptional regulation is a complex and dynamic process of which histone modifications are a key component. Some histone marks are established as consequences of active transcription; however, these histone marks provide memory of recent transcriptional activity and signal for additional transcriptional regulation. Many histone marks that are not coupled with active transcription function in transcriptional regulation processes. Rapid progress in genome-wide experiments, bioinformatics analysis and new technologies that allow epigenetic enzymes to be targeted to specified DNA sequences [25] will help elucidate the genesis and biological functions of the myriad of histone modifications.

Acknowledgments

We thank B Pierce for editing the manuscript.

This work was funded by NIH grant U41 HG007000.

Biography

Xianjun Dong

Zhiping Weng

References

1. Bernstein BE, Mikkelsen TS, Xie X, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006; 125:315–326. [PubMed: 16630819]
2. Barski A, Cuddapah S, Cui K, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007; 129:823–837. [PubMed: 17512414]
3. Wang Z, Zang C, Rosenfeld JA, et al. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* 2008; 40:897–903. [PubMed: 18552846]
4. Lim PS, Hardy K, Bunting KL, et al. Defining the chromatin signature of inducible genes in T cells. *Genome Biol.* 2009; 10:R107. [PubMed: 19807913]
5. Cain CE, Blekhman R, Marioni JC, Gilad Y. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics*. 2011; 187(4):1225–1234. [PubMed: 21321133]
6. Araki Y, Wang Z, Zang C, et al. Genome-wide analysis of histone methylation reveals chromatin state-based regulation of gene transcription and function of memory CD8+ T cells. *Immunity*. 2009; 30:912–925. [PubMed: 19523850]
7. Karli R, Chung H-R, Lasserre J, Vlahovicek K, Vingron M. Histone modification levels are predictive for gene expression. *Proc. Natl Acad. Sci. USA*. 2010; 107:2926–2931. [PubMed: 20133639]
8. Cheng C, Yan K-K, Yip KY, et al. A statistical framework for modeling gene expression using chromatin features and application to modENCODE datasets. *Genome Biol.* 2011; 12:R15. [PubMed: 21324173]
9. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012; 489:57–74. [PubMed: 22955616]
10. Dong X, Greven MC, Kundaje A, et al. Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.* 2012; 13:R53. [PubMed: 22950368]

11. Henikoff S, Shilatifard A. Histone modification: cause or cog? *Trends Genet.* 2011; 27:389–396. [PubMed: 21764166]
12. Ng H-H, Robert F, Young RA, Struhl K. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. *Mol. Cell.* 2003; 11:709–719. [PubMed: 12667453]
13. Li J, Moazed D, Gygi SP. Association of the histone methyltransferase Set2 with RNA polymerase II plays a role in transcription elongation. *J. Biol. Chem.* 2002; 277:49383–49388. [PubMed: 12381723]
14. Krogan NJ, Dover J, Wood A, et al. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell.* 2003; 11:721–729. [PubMed: 12667454]
15. Krogan NJ, Kim M, Tong A, et al. Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell Biol.* 2003; 23:4207–4218. [PubMed: 12773564]
16. Rando OJ, Chang HY. Genome-wide views of chromatin structure. *Annu. Rev. Biochem.* 2009; 78:245–271. [PubMed: 19317649]
17. Sun Z-W, Allis CD. Ubiquitination of histone H2B regulates H3 methylation and gene silencing in yeast. *Nature.* 2002; 418:104–108. [PubMed: 12077605]
18. Carrozza MJ, Li B, Florens L, et al. Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell.* 2005; 123:581–592. [PubMed: 16286007]
19. Rea S, Eisenhaber F, O'Carroll D, et al. Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature.* 2000; 406:593–599. [PubMed: 10949293]
20. Kuzmichev A, Nishioka K, Erdjument-Bromage H, Tempst P, Reinberg D. Histone methyltransferase activity associated with a human multiprotein complex containing the enhancer of Zeste protein. *Genes Dev.* 2002; 16:2893–2905. [PubMed: 12435631]
21. Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature.* 2011; 469:343–349. [PubMed: 21248841]
22. Lehnertz B, Ueda Y, Derijck AA, et al. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. *Current Biol.* 2003; 13:1192–1200.
23. Günesdogan U, Jäckle H, Herzig A. A genetic system to assess in vivo the functions of histones and histone modifications in higher eukaryotes. *EMBO Rep.* 2010; 11:772–776. [PubMed: 20814422]
24. Pengelly AR, Copur Ö, Jäckle H, Herzig A, Müller J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science.* 2013; 339:698–699. [PubMed: 23393264]
25. de Groote ML, Verschure PJ, Rots MG. Epigenetic editing: targeted rewriting of epigenetic marks to modulate expression of selected target genes. *Nucleic Acids Res.* 2012; 40:10596–10613. [PubMed: 23002135]