Cell Cycle 11:24, 4487–4488; December 15, 2012; © 2012 Landes Bioscience

Uncovering the interplay between DNA sequence preferences of transcription factors and nucleosomes

Sarah A. Teichmann,[1,*] Philip A. Wigge[2] and Varodom Charoensawan[3,4,*]

[1]MRC Laboratory of Molecular Biology; Cambridge, UK; [2]Sainsbury Laboratory; University of Cambridge; Cambridge, UK; [3]Department of Biochemistry; Faculty of Science; Mahidol University; Bangkok, Thailand; [4]Integrative Computational BioScience (ICBS) Center; Mahidol University; Nakhon Pathom, Thailand

The genome of a cell contains two major types of information. The coding regions specify amino acid sequences of the cognate proteins, whereas the promoter and enhancer elements control when and where these genes will be expressed. Unlike the genetic code, our understanding of transcriptional regulation is far from complete: knowledge of a genome sequence alone is insufficient to predict its transcriptome. Major determinants of the transcriptional program of the cell are transcription factors (TFs), sequence-specific DNA-binding proteins and histone octamers, which are necessary to condense DNA so it can be packaged within the nucleus. Since TFs and histones are the most abundant classes of DNA-binding proteins in the cell, it has been suggested for many years that their interplay may be important in determining the transcriptome. Indeed, detailed studies of individual TFs, notably heat shock and temperature-responsive proteins,[1,2] have shown that TFs bind to nucleosome-occluded DNA sequences and induce nucleosome repositioning. However, global studies indicate that TFs use multiple mechanisms to regulate transcription.[3] This raises the question of whether there are general principles describing how TFs interact with nucleosomes to regulate the transcriptome.

A recent study by two of the authors (V.C. and S.A.T.) has addressed this question by integrating in vitro and in vivo DNA-binding preferences of histones, and 137 budding yeast *Saccharomyces cerevisiae* TFs and quantitatively assessed the extent to which the intrinsic DNA-binding specificities of yeast TFs and histones influence their actual binding positions in a population of cells.[4] Fortunately, there are extensive data, determined in vitro, for the intrinsic binding sequences for many yeast TFs. The majority of these TFs show a high correlation of their preferred DNA sequences with histone footprints (termed "histone-correlated" group, HC), indicating that for these HC TFs to function in the cell, they must compete with and displace nucleosomes to bind their cognate sites.

A second class of TFs are the "histone-anti-correlated" group, HA, which intrinsically prefer non-overlapping binding sites to histones. Since TFs of the HA class do not need to displace nucleosomes, it was predicted that in vitro determined nucleosome positions around the binding sites of the HA class should match actual nucleosome positions in the cell closely, as compared with the HC class. This is indeed the case, since the difference in proportions of nucleosome-enriched TF-binding sites, based on nucleosome occupancy determined in vitro compared with in vivo, is significantly larger for the HC TFs. HC TFs therefore change the chromatin landscape by displacing nucleosomes when they bind their target sites, causing a difference between in vitro determined histone occupancy and in vivo data (**Fig. 1**).

This finding raised the question of whether there is a functional significance associated with TFs having these opposite effects on chromatin structure. Strikingly, analysis of high-throughput experimental data revealed that TFs with preferred recognition sequences similar to histones are frequently associated with gene activation. Our global analysis supports earlier detailed studies for individual TFs, suggesting that activators compete with histones, and induce nucleosome disruption/repositioning upon their binding to DNA in living cells.[5,6] Opening of chromatin structure is predicted to facilitate transcription by allowing access of RNA PolII. It is worth noting, however, that we did not find a statistically significant direct association between the HC group and activators per se. This could be because the functional regulatory modes (activator or repressor) have only been verified for about two-thirds of the TFs in our data set.

Apart from the distinct modes of interplay between different types of TFs and nucleosomes,[4] earlier studies have documented the influences of other promoter features on transcriptional regulation, including different histone architecture[3] and cooperativity of multiple TF-binding sites within close proximity.[7] Our recent analysis confirms previous findings that co-binding of multiple TFs within close proximity can yield combinatorial effects in lowering the nucleosome occupancy around the co-occurring TF-binding sites.

We observed a higher level of histone depletion around co-occurring TF-binding sites, especially within clusters of multiple HC TF-binding sites, or TF-binding sites that match strongly to the consensus motif, as compared with isolated binding sites (**Fig. 1**). Consequently, this might promote accessibility of binding sites in proximity and help fine-tuning

*Correspondence to: Sarah A. Teichmann and Varodom Charoensawan; Email: sat@mrc-lmb.cam.ac.uk and varodom.cha@mahidol.ac.th
Submitted: 09/24/12; Accepted: 10/16/12
http://dx.doi.org/10.4161/cc.22666
Comment on: Charoensawan V, et al. Mol Cell 2012; 47:183-92; PMID:22841002; http://dx.doi.org/10.1016/j.molcel.2012.06.028.

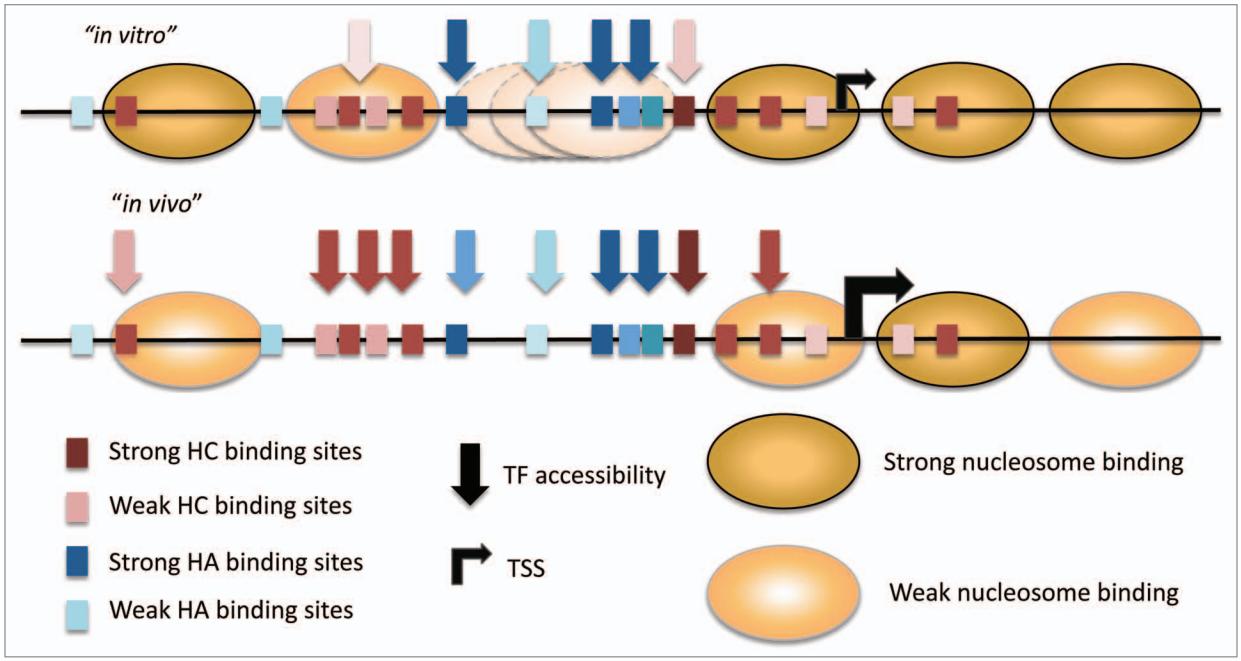

Figure 1. Summary of in vitro and in vivo TF and nucleosome-binding landscapes. Based on TF and nucleosome intrinsic (in vitro) preferences, only the binding sites of HA and only very strong HC should be accessible. In the YPD growth medium (in vivo), however, nucleosomes can be disrupted/repositioned from their "preferred" DNA sequences upon the binding of HC TFs. Lower nucleosome occupancy is observed over clusters of multiple TF-binding sites, especially the sites bound by HC TFs and those that match strongly with consensus motifs. This results in more binding sites becoming accessible and a concomitant increase in gene expression.

transcriptional rate of genes in response to changes in environmental conditions.[8] Combinatorial binding of eukaryotic TFs might be employed to increase affinity and specificity and, thus, help minimize non-functional binding.[9] It may also be used to compensate the loss in binding strength of the sites, which are less similar to consensus motifs.[7]

In addition to genome-wide principles for the interplay between DNA sequence, histone occupancy and TF binding in the regulation of gene expression, our computational analysis generates a direct testable prediction that HC/activators should compete with nucleosomal histones and, thus, cause greater changes in nucleosome organization at increased concentrations than HA/repressors. Alternatively, this hypothesis can also be comprehensively assessed by comparing the reduction of nucleosome occupancy in wild-type and histone-depleted yeast strains.[10] The greater difference in nucleosome occupancy between the two strains is expected at the binding sites of HC/activators, as compared with HA/repressors. Transcriptional regulation occurs in a chromatin context, and so understanding the underlying rules of how TFs and histones interact is of key importance. Recent advances in cataloging the functional elements of the human genome[11] will help extend these studies to multicellular eukaryotes.

References

1. Shivaswamy S, et al. PLoS Biol 2008; 6:e65; PMID:18351804; http://dx.doi.org/10.1371/journal.pbio.0060065.
2. Kumar SV, et al. Nature 2012; 484:242-5; PMID:22437497; http://dx.doi.org/10.1038/nature10928.
3. Zaugg JB, et al. Genome Res 2012; 22:84-94; PMID:21930892; http://dx.doi.org/10.1101/gr.124099.111.
4. Charoensawan V, et al. Mol Cell 2012; 47:183-92; PMID:22841002; http://dx.doi.org/10.1016/j.molcel.2012.06.028.
5. Bai L, et al. Mol Cell 2011; 42:465-76; PMID:21596311; http://dx.doi.org/10.1016/j.molcel.2011.03.028.
6. Ganapathi M, et al. Nucleic Acids Res 2011; 39:2032-44; PMID:21081559; http://dx.doi.org/10.1093/nar/gkq1161.
7. Bilu Y, et al. Genome Biol 2005; 6:R103; PMID:16356266; http://dx.doi.org/10.1186/gb-2005-6-12-r103.
8. Zhou X, et al. Mol Cell 2011; 42:826-36; PMID:21700227; http://dx.doi.org/10.1016/j.molcel.2011.05.025.
9. Mirny LA. Proc Natl Acad Sci USA 2010; 107:22534-9; PMID:21149679; http://dx.doi.org/10.1073/pnas.0913805107.
10. Gossett AJ, et al. PLoS Genet 2012; 8:e1002771; PMID:22737086; http://dx.doi.org/10.1371/journal.pgen.1002771.
11. Raney BJ, et al. Nucleic Acids Res 2011; 39(Database issue):D871-5; PMID:21037257; http://dx.doi.org/10.1093/nar/gkq1017.