Published in final edited form as:
Curr Opin Genet Dev. 2013 April ; 23(2): 116–123. doi:10.1016/j.gde.2012.11.008.

Mechanisms by which transcription factors gain access to target sequence elements in chromatin

Michael J. Guertin and **John T. Lis**
Dept. of Molecular Biology and Genetics Cornell University

Abstract

Transcription factors (TF) bind DNA sequence motifs, but the presence of a consensus DNA element is not sufficient to direct TF binding to chromatin. Recent genomic data have revealed that accessibility, as measured by DNase sensitivity and the presence of active histone marks, is necessary for TF binding. DNA sequence provides the initial specification of the accessibility of DNA elements within chromatin that permits TF binding. In yeast, it is known that poly(dA-dT) tracts directly encode low-nucleosome occupancy at promoters. Recent evidence suggests that CpG islands in mammals are inherently refractory to higher-order chromatin structure and remain accessible, despite favoring nucleosome formation in vitro. Taken together, these studies support a model for how accessibility originates and then propagates throughout regulatory cascades and development.

Introduction

Specific DNA sequence elements are sufficient to direct transcription factor (TF) binding in prokaryotes; however, in higher organisms, chromatin often occludes TF binding [1–4]. In eukaryotes, DNA is wrapped around nucleosomes and forms higher order chromatin structures that restrict TF access. The first high resolution in vivo measure of the accessibility of chromatin structure came from the study of candidate heat shock genes [5]. DNase I footprinting revealed that the 5′ end of *Hsp70* and *Hsp83* were highly sensitive to digestion prior to heat shock induction [5]. Post-translational modifications of histones provide an independent measure of chromatin structure. For instance, histone acetylation is associated with actively transcribed genes [6,7], but can also be a precursor to transcription and permit subsequent activation [8]. DNase I signals overlap with histone acetylation marks [9], suggesting that histone acetylation contributes to the molecular basis for DNase I sensitivity. Multiple types of histone acetylation and H3K4 methylation often co-occur and are associated with transcriptionally active or potentiated chromatin [10–13]. Recently, André Martins developed a probabilistic model that infers DNase I sensitivity from histone modification data, reinforcing the qualitative link between DNase sensitivity and active histone marks with a quantitative model [14]. (Fig. 1)

Correspondence should be addressed to both MJG @ GuertinMJ@gmail.com and JTL @ johnlis@cornell.edu.

Present address of Michael J. Guertin: Laboratory of Receptor Biology and Gene Expression, Center for Cancer Research, NCI, National Institutes of Health, Bethesda, Maryland 20892

Early studies revealed that TF bound DNA and DNase I sensitivity co-occurred at candidate promoters [5,15]; however, these measurements did not resolve the causality of either event. Did pre-existing DNase I hypersensitivity permit TF binding or did TF binding affect the local chromatin environment? Candidate gene analyses have shown that accessible chromatin correlated with inducibly-bound TF binding sites for several factors [16–18]. To test whether accessible chromatin was necessary and sufficient to direct TF binding to consensus elements, genomic assays were needed to examine the comprehensive set of bound target elements and unbound consensus elements. The recent advent of molecular genomics approaches that measure both the accessibility of DNA (DNase-seq, FAIRE-seq, and ChIP-seq) [19–21] and the inducible binding of transcription factors (ChIP-seq) [10,12,22,23] allows this causality to be addressed in a comprehensive manner.

Genomic assays reveal that active chromatin is required for TF binding

The first genome-wide study looking at histone marks and inducible TF binding indicated that active chromatin marks preceded TF binding [24]. The authors found that STAT1 binding sites were marked by H3K4 methylation prior to interferon-induced STAT1 binding [24]. These data suggested that an active chromatin state is necessary for inducible binding. However, it remained unclear if potential, but unoccupied, STAT1 binding sites also harbored active marks, but remained unbound after stimulation.

To test whether or not potential binding sites in active chromatin remain untargeted, we directly compared the chromatin landscape (histone modification and chromatin associated factors) at inducibly-bound, heat shock factor (HSF) DNA elements (HSEs) to unoccupied HSEs [10]. We found that inducibly-bound HSEs resided in chromatin characterized by histone acetylation and H3K4 methylation and unbound HSEs lacked these histone marks (Fig. 2). A similar study showed that the glucocorticoid receptor (GR) specifically binds to target elements that are pre-marked by DNase I hypersensitive signal prior to ligand treatment [12]. Taken together, these data indicate that TFs are specifically targeted to consensus elements within a region of active chromatin.

The previously mentioned studies were performed in cell lines with an inducing agent (interferon, hormone, and heat stress), but defining the temporal order of chromatin structural change and TF binding is more difficult within a developing organism. Multiple cell types can confound the quantitative analysis of both TF binding and chromatin structure. Additionally, the analysis requires capturing cells at a stage just prior to a TF binding, which is difficult to accomplish in a manner that does not perturb the developmental process [25,26]. To overcome these limitations and study the context-dependent manner of transcription factor binding in erythroid differentiation, Wu et al. generated an inducible GATA1 cell line [27]. They found that erythroid progenitors lacking GATA1 retain the chromatin state (H3K4 methylation and DNase I sensitivity) that is permissive for binding [27]. Others have captured the early events in the reprogramming of differentiated cells to induced pluripotent stem cells (iPSC) using an assay that allows for cells to be distinguished by their number of cell divisions [28]. Reprogramming is initiated by ectopic expression of Oct4, Sox2, Klf4, and c-Myc, and the authors showed that the primary targets of these factors pre-exist in an accessible state [28].

A special class of TFs termed “pioneer factors” are often the first detectable transcription factors binding a region of chromatin in vivo and can access nucleosomal DNA in vitro [29]. Recent genomic studies have shown that H3K4 methylation and DNase I sensitivity precedes binding of the pioneer factors FoxA1 and GATA1 [18,27,30]. FoxA1's high affinity for nucleosomal DNA [29] likely allows FoxA1 to bind to transiently accessible chromatin that may be inaccessible to other factors (reviewed in [31]), but the in vivo evidence suggests that these sites are not heterochromatic. Here we use “heterochromatic” to

refer to the annotation of chromatin states defined by the presence and absence of specific chromatin marks that likely result in higher order chromatin structure [11,13]. FoxA1 and other TFs are defined as pioneers, because they precede binding of other TFs. These studies indicate that the mechanism and function of pioneer factor binding is not appreciably different from TFs that are not considered pioneers: both classes bind to relatively decondensed chromatin marked by active histone marks and subsequently reinforce and expand the accessible region [10,12,32].

We propose that many factors have the potential to pioneer a region. For instance, AP1 binding precedes GR binding and maintains accessible chromatin at over 70% of GR binding sites, but it seems that GR is acting as a classical pioneer at 15% of the binding sites [12,32]. We have found that the presence of GAGA Associated Factor (GAF) at HSEs positively influences HSF binding at a subset of sites [33].

Accessibility increases concomitantly upon Transcription Factor binding

Just as chromatin landscape influences TF binding, so does TF binding influence chromatin state. The histone marks and DNase I signal that are predictive of TF binding also increase in intensity upon inducible binding [5,10,12,19–23,34,35]. This observation is consistent with the role of TFs in recruiting coactivator complexes that are capable of modifying histone tails and remodeling nucleosomes [24,36–43]. Concomitant increases in locus accessibility may reinforce and expand the boundaries of accessible chromatin to allow access for other TFs and large molecular complexes. This cascade can lead to accumulation of paused RNA polymerase II near promoters, which can exclude nucleosomes from promoters and help maintain an active and factor-accessible chromatin state [44,45].

Sequence composition encodes accessible chromatin

The mechanism by which accessibility originates to allow TF binding is an unresolved but active subject of investigation. Ultimately, DNA sequence must direct the origin of accessible DNA. In yeast, the low nucleosome density at promoters is largely specified by DNA sequence, mainly poly(dA-dT) tracts, producing promoters that are generally devoid of nucleosomes permitting TF access [17,46,47].

CpG islands, which are hallmarks of approximately 60% of human and mouse promoters [25,48], favor nucleosome formation in vitro [49], but are nucleosome-depleted in vivo [50]. It seems counterintuitive that the DNA sequence of CpG island regulatory elements would favor nucleosome formation. We hypothesize that CpG island sequence is inherently refractory to higher order compaction by linker histones and retains the chromatin in a transiently accessible state amenable to TF binding.

It has been known for over 40 years that linker histones, like H1, preferentially bind AT-rich linker DNA (reviewed in [51]). Recently, the molecular interactions that are responsible for this selectivity have been characterized [52]. In short, the nonpolar region of the linker histone's globular domain preferentially interacts with the thymine methyl groups through hydrophobic interactions [52]. These data support the hypothesis that CpG islands are refractory to higher-order compaction in vivo. Nonetheless, this does not completely explain the fact that CpG island promoters are generally highly accessible in vivo [50], despite favoring nucleosome assembly in vitro [49]. TFs are highly bound to CpG islands in vivo, so it is clear that their elements are initially accessible and that TF binding propagates a more accessible chromatin landscape. Consistent with this idea, CpG islands were found to be sufficient to direct the *de novo* recruitment of the CxxC finger protein 1 (Cfp1) and subsequent H3K4 methylation [53]. Cfp1 is a component of the Set1 H3K4 methylation complex [54], and has been shown to interact with unmethylated CpG-rich DNA in vitro

[55]. Methylation of H3K4 is inhibitory to DNA methylation, so H3K4 methylation may also reinforce and maintain the unmethylated DNA state of CpG islands [56]. KDM2A, a H3K36 demethylase, also targets linker DNA between nucleosomes in CpG islands by a CxxC zinc finger domain protein [57,58], consistent with the idea that CpG islands do not efficiently incorporate linker histones for compaction.

Non-CxxC zinc finger proteins must also target the linker DNA between nucleosomes at CpG islands, and in fact, mammalian sequence-specific TFs, as a class, have a GC-bias in their cognate binding sites [59] and TFs bind to GC-rich regions in vivo [50]. Thus, CpG islands, which constitute a majority of mammalian promoters, have inherent properties amenable for TF binding. Although these GC-rich sequences can interact with nucleosomes, they appear to be dissuaded from higher-order packing and are enriched for TF binding elements. Importantly, CpG island promoters can be found in a repressive state in vivo, so repressive TFs and cofactors can also bind to the region and precipitate repressive chromatin and further condensation.

This model of CpG island structure and function is highlighted by the example of embryonic stem cell differentiation. CpG island promoters that are uncommitted to the repressive or active state, so-called bivalent promoters, are highly depleted of DNA elements that are the targets of TFs that are expressed in the embryonic stem cell [60,61]. These same promoters in differentiated cells are highly acetylated and more accessible, presumably because these cells express TFs that are competent to bind the region and tilt the balance from the bivalent to the active state [60,62].

The DNA sequence of *Drosophila*, which is not enriched for CpG-island and poly(dA-dT) tracts in its promoters, can also influence TF binding. A recent study has shown that the high GC content in the 10kb flanking Male-specific lethal (MSL) binding sites can influence MSL binding [63]. We speculate that the sequence in flanking regions of *Drosophila* also may occlude higher order chromatin compaction.

Our model: DNA sequence directs chromatin structure to allow TF binding

We propose that DNA sequences that inherently promote unstable nucleosomes (AT-rich stretches) or disfavor higher order compaction (CpG islands) are sufficient to keep chromatin in a state that is at least transiently accessible to TF binding and uncommitted to the repressive or active transcription state (Figure 3A). Throughout development, or in response to regulatory cascades, previously unexpressed or inactive TFs (often pioneer TFs) and nucleosome modifying enzymes are targeted to their cognate binding sites (Figure 3B) through their interaction with accessible DNA (often CG-rich, unmethylated, linker DNA between nucleosomes). Targeting of the TFs cause subsequent hyper-modification of histones in the region and tilt the balance toward the repressive or active state, as directed by the TFs (Figure 3B). Subsequent recruitment of cofactors or RNA polymerase itself is sufficient to enhance and expand the boundaries of the accessible or inaccessible chromatin (Figure 3C). In some cases, TF binding will directly or indirectly activate or repress the transcription of other sequence-specific TFs or cofactors (Figure 3C). The cognate elements of transcriptionally activated TFs will be targeted if they reside within active chromatin (Figure 3D). The *trans* targets of transcriptionally repressed TFs will return to the default accessibility state, in the absence of factors with redundant functions that reinforce the chromatin state. This cascade can be subject to further perturbation by activating inactive TFs through environmental factors (e.g. heat shock, hormone treatment, stress, etc.). Genomic tracking of chromatin and transcription changes throughout a regulatory cascade will test the validity of the proposed model.

Inference and prediction of Transcription Factor binding sites

A complete understanding of the DNA and chromatin features that influence TFBS selection would permit accurate prediction of TFBS intensities using only DNA sequence and chromatin landscape. Many TFBS models infer TF binding, rather than predict inducible binding patterns *de novo* [26,64–66]. These models are developed with data for the chromatin state concurrent with TF binding, so the effect of preexisting chromatin states upon TF binding is confounded by the influence of the TF on the local chromatin environment. These models also use position-specific weight matrices to predict potential binding sites, which can be problematic if the assumption of independence between positions in the binding site is not met. Recently, many protein/DNA binding assays have been developed to directly measure the TF/DNA binding affinity [14,67–69]. These assays can be combined with in vivo chromatin landscape data to predict *de novo* TFBS intensity from chromatin features. For instance, we used PB-seq [14] to measure the binding energy landscape of all potential binding sites in a genome in a chromatin-free context and chromatin landscape data from the modENCODE consortium [13] to develop a probabilistic TF binding model. Our model revealed that the intensity of each modification could be used to predict more than just whether or not HSF would bind, but also the intensity of binding. As more chromatin data becomes available, these models will be expanded upon and refined.

Conclusions

Ultimately, the goal is to understand TF binding and function within a developmental or regulatory framework. This understanding of networks would allow one to predict the transcriptional output of binding events and the cascade of regulatory binding and transcription cycles that follows in normal and disease states. Carefully designed kinetic experiments that monitor changes in chromatin structure, TF binding, and transcription throughout stages of development and regulatory networks will be necessary to unravel the complex cascade of molecular events precipitated by differentiation or response to stimuli.

References

1. Lin S, Riggs AD. The general affinity of lac repressor for E. coli DNA: implications for gene regulation in procaryotes and eucaryotes. Cell. 1975; 4:107–111. [PubMed: 1092468]
2. Hippel von PH, Revzin A, Gross CA, Wang AC. Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects. Proc Natl Acad Sci. 1974; 71:4808–4812. [PubMed: 4612528]
3. Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, Struhl K. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. Mol Cell. 2006; 24:593–602. [PubMed: 17188034]
4. Wade JT, Reppas NB, Church GM, Struhl K. Genomic analysis of LexA binding reveals the permissive nature of the Escherichia coli genome and identifies unconventional target sites. Genes Dev. 2005; 19:2619–2630. [PubMed: 16264194]
5. Wu C. The 5′ ends of Drosophila heat shock genes in chromatin are hypersensitive to DNase I. Nature. 1980; 286:854–860. [PubMed: 6774262]
6. Allfrey VG, Faulkner R, Mirsky AE. Acetylation and methylation of histones and their possible role in hte regulation of RNA synthesis. Proc Natl Acad Sci. 1964; 51:786. [PubMed: 14172992]
7. Pogo BG, Allfrey VG, Mirsky AE. RNA synthesis and histone acetylation during the course of gene activation in lymphocytes. Proc Natl Acad Sci. 1966; 55:805–812. [PubMed: 5219687]
8. Hebbes TR, Thorne AW, Clayton AL, Crane-Robinson C. Histone acetylation and globin gene switching. Nucleic Acids Res. 1992; 20:1017–1022. [PubMed: 1549462]

9. Hebbes TR, Clayton AL, Thorne AW, Crane-Robinson C. Core histone hyperacetylation co-maps with generalized DNase I sensitivity in the chicken beta-globin chromosomal domain. EMBO J. 1994; 13:1823–1830. [PubMed: 8168481]
10. Guertin MJ, Lis JT. Chromatin landscape dictates HSF binding to target DNA elements. PLoS Genet. 2010; 6:e1001114. [PubMed: 20844575] ••This is the first genome-wide study to directly compare the chromatin landscape at targeted binding sites to potential binding sites for an inducibly bound TF. Heat Shock Factor was specifically targeted to cognate elements that were enriched for active chromatin marks prior to heat shock induction.
11. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol. 2010; 28:817–825. [PubMed: 20657582]
12. John S, Sabo PJ, Thurman RE, Sung M-H, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat Genet. 2011; 43:264–268. [PubMed: 21258342] ••This study found that Glucocorticoid Receptor (GR) is specifically targeted to GR elements marked by DNase I hypersensitivity prior to induction with glucocorticoid. The authors were the first to show that genome-wide, binding of a TF results in a concomitant increase in accessibility at binding sites.
13. Kharchenko PV, Alekseyenko AA, Schwartz YB, Minoda A, Riddle NC, Ernst J, Sabo PJ, Larschan E, Gorchakov AA, Gu T, et al. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature. 2011; 471:480–485. [PubMed: 21179089]
14. Guertin MJ, Martins AL, Siepel A, Lis JT. Accurate prediction of inducible transcription factor binding intensities in vivo. PLoS Genet. 2012; 8:e1002610. [PubMed: 22479205] •This study was the first use genome-wide TF/DNA binding affinity measurements (PB-seq) and pre-existing chromatin landscape to predict inducible TF binding intensities in vivo.
15. Carr A, Biggin MD. Accessibility of transcriptionally inactive genes is specifically reduced at homeoprotein-DNA binding sites in Drosophila. Nucleic Acids Res. 2000; 28:2839–2846. [PubMed: 10908343]
16. Guccione E, Martinato F, Finocchiaro G, Luzi L, Tizzoni L, Dall' Olio V, Zardo G, Nervi C, Bernard L, Amati B. Myc-binding-site recognition in the human genome is determined by chromatin context. Nat. Cell Biol. 2006; 8:764–770. [PubMed: 16767079]
17. Sekinger EA, Moqtaderi Z, Struhl K. Intrinsic histone-DNA interactions and low nucleosome density are important for preferential accessibility of promoter regions in yeast. Mol Cell. 2005; 18:735–748. [PubMed: 15949447]
18. Lupien M, Eeckhoute J, Meyer CA, Wang Q, Zhang Y, Li W, Carroll JS, Liu XS, Brown M. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. Cell. 2008; 132:958–970. [PubMed: 18358809] ••These authors found that FoxA1 specifically targets elements that are marked by H3K4 dimethylation in vivo. Despite FoxA1's ability to bind nucleosomal DNA in vitro, these data showed that the targeted FoxA1 motif was marked by active chromatin in vivo.
19. Hesselberth JR, Chen X, Zhang Z, Sabo PJ, Sandstrom R, Reynolds AP, Thurman RE, Neph S, Kuehn MS, Noble WS, et al. Global mapping of protein-DNA interactions in vivo by digital genomic footprinting. Nat Methods. 2009; 6:283–289. [PubMed: 19305407]
20. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, Furey TS, Crawford GE. High-resolution mapping and characterization of open chromatin across the genome. Cell. 2008; 132:311–322. [PubMed: 18243105]
21. Giresi PG, Lieb JD. Isolation of active regulatory elements from eukaryotic chromatin using FAIRE (Formaldehyde Assisted Isolation of Regulatory Elements). Methods. 2009; 48:233–239. [PubMed: 19303047]
22. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, Lee B-K, Sheffield NC, Gräf S, Huss M, Keefe D, et al. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res. 2011; 21:1757–1767. [PubMed: 21750106]
23. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, et al. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods. 2007; 4:651–657. [PubMed: 17558387]

24. Robertson AG, Bilenky M, Tam A, Zhao Y, Zeng T, Thiessen N, Cezard T, Fejes AP, Wederell ED, Cullum R, et al. Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. Genome Res. 2008; 18:1906–1917. [PubMed: 18787082]
25. Li X-Y, Thomas S, Sabo PJ, Eisen MB, Stamatoyannopoulos JA, Biggin MD. The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. Genome Biology. 2011; 12:R34. [PubMed: 21473766]
26. Kaplan T, Li X-Y, Sabo PJ, Thomas S, Stamatoyannopoulos JA, Biggin MD, Eisen MB. Quantitative models of the mechanisms that control genome-wide patterns of transcription factor binding during early Drosophila development. PLoS Genet. 2011; 7:e1001290. [PubMed: 21304941]
27. Wu W, Cheng Y, Keller CA, Ernst J, Kumar SA, Mishra T, Morrissey C, Dorman CM, Chen K-B, Drautz D, et al. Dynamics of the epigenetic landscape during erythroid differentiation after GATA1 restoration. Genome Res. 2011; 21:1659–1671. [PubMed: 21795386] •The authors found that during erythroid differentiation, the chromatin states that permit GATA1 binding are established at the stage of lineage commitment, independent of the master regulator GATA1. Moreover, this study reinforced the idea that pioneer factors access regions that are pre-marked with active histone marks and DNase I hypersensitivity.
28. Koche RP, Smith ZD, Adli M, Gu H, Ku M, Gnirke A, Bernstein BE, Meissner A. Reprogramming factor expression initiates widespread targeted chromatin remodeling. Cell Stem Cell. 2011; 8:96–105. [PubMed: 21211784]
29. Cirillo LA, McPherson CE, Bossard P, Stevens K, Cherian S, Shim EY, Clark KL, Burley SK, Zaret KS. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. EMBO J. 1998; 17:244–254. [PubMed: 9427758]
30. Sérandour AA, Avner S, Percevault F, Demay F, Bizot M, Lucchetti-Miganeh C, Barloy-Hubler F, Brown M, Lupien M, Métivier R, et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. Genome Res. 2011; 21:555–565. [PubMed: 21233399]
31. Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. Genes Dev. 2011; 25:2227–2241. [PubMed: 22056668]
32. Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schiltz RL, Miranda TB, Sung M-H, Trump S, Lightman SL, et al. Transcription Factor AP1 Potentiates Chromatin Accessibility and Glucocorticoid Receptor Binding. Mol Cell. 2011; 43:145–155. [PubMed: 21726817]
33. Sharma, S.; Guertin, MJ.; Lis, JT. Undergraduate Thesis. Cornell University; 2012. The Genomic Binding Profile of GAGA Element Associated Factor (GAF) in Drosophila S2 cells.
34. Voss TC, Schiltz RL, Sung M-H, Yen PM, Stamatoyannopoulos JA, Biddie SC, Johnson TA, Miranda TB, John S, Hager GL. Dynamic Exchange at Regulatory Elements during Chromatin Remodeling Underlies Assisted Loading Mechanism. Cell. 2011; 146:544–554. [PubMed: 21835447]
35. Guertin MJ, Petesch SJ, Zobeck KL, Min IM, Lis JT. Drosophila Heat Shock System as a General Model to Investigate Transcriptional Regulation. Cold Spring Harbor Symposia on Quantitative Biology. 2011; 75:1–9.
36. John S, Sabo PJ, Johnson TA, Sung M-H, Biddie SC, Lightman SL, Voss TC, Davis SR, Meltzer PS, Stamatoyannopoulos JA, et al. Interaction of the Glucocorticoid Receptor with the Chromatin Landscape. Mol Cell. 2008; 29:611–624. [PubMed: 18342607]
37. Johnson TA, Elbi C, Parekh BS, Hager GL, John S. Chromatin remodeling complexes interact dynamically with a glucocorticoid receptor-regulated promoter. Mol. Biol. Cell. 2008; 19:3308–3322. [PubMed: 18508913]
38. Vo N, Goodman RH. CREB-binding Protein and p300 in Transcriptional Regulation. J Biol Chem. 2001; 276:13505–13508. [PubMed: 11279224]
39. Kim MY, Hsiao SJ, Kraus WL. A role for coactivators and histone acetylation in estrogen receptor alpha-mediated transcription initiation. EMBO J. 2001; 20:6084–6094. [PubMed: 11689448]
40. Peterson CL, Workman JL. Promoter targeting and chromatin remodeling by the SWI/SNF complex. Current Opinion in Genetics & Development. 2000; 10:187–192. [PubMed: 10753786]

41. Martinato F, Cesaroni M, Amati B, Guccione E. Analysis of Myc-induced histone modifications on target chromatin. PLoS ONE. 2008; 3:e3650. [PubMed: 18985155]

42. Zhang X, Bolt M, Guertin MJ, Chen W, Zhang S, Cherrington BD, Slade DJ, Dreyton CJ, Subramanian V, Bicker KL, et al. Peptidylarginine deiminase 2-catalyzed histone H3 arginine 26 citrullination facilitates estrogen receptor α target gene activation. Proc Natl Acad Sci. 2012 currently published online before print.

43. Hu G, Schones DE, Cui K, Ybarra R, Northrup D, Tang Q, Gattinoni L, Restifo NP, Huang S, Zhao K. Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1. Genome Res. 2011; 21:1650–1658. [PubMed: 21795385]

44. Gilchrist DA, Santos Dos G, Fargo DC, Xie B, Gao Y, Li L, Adelman K. Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. Cell. 2010; 143:540–551. [PubMed: 21074046]

45. Shopland LS, Hirayoshi K, Fernandes M, Lis JT. HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. Genes Dev. 1995; 9:2756–2769. [PubMed: 7590251]

46. Mai X, Chou S, Struhl K. Preferential Accessibility of the Yeast his3 Promoter Is Determined by a General Property of the DNA Sequence, Not by Specific Elements. Molecular and cellular biology. 2000; 20:6668–6676. [PubMed: 10958664]

47. Yuan G-C, Liu Y-J, Dion MF, Slack MD, Wu LF, Altschuler SJ, Rando OJ. Genome-scale identification of nucleosome positions in S. cerevisiae. Science. 2005; 309:626–630. [PubMed: 15961632]

48. Antequera F. Structure, function and evolution of CpG island promoters. Cell. Mol. Life Sci. 2003; 60:1647–1658. [PubMed: 14504655]

49. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, Sidow A. Determinants of nucleosome organization in primary human cells. Nature. 2011; 474:516–520. [PubMed: 21602827]

50. Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, et al. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. Genome Res. 2012 currently published online before print.

51. Zlatanova J, Yaneva J. Histone H1-DNA interactions and their relation to chromatin structure and function. DNA Cell Biol. 1991; 10:239–248. [PubMed: 2029335]

52. Cui F, Zhurkin VB. Distinctive sequence patterns in metazoan and yeast nucleosomes: implications for linker histone binding to AT-rich and methylated DNA. Nucleic Acids Res. 2009; 37:2818–2829. [PubMed: 19282449]

53. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD, et al. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature. 2010; 464:1082–1086. [PubMed: 20393567] •This study found that the CxxC zinc finger protein Cfp1 is targeted to CpG island chromatin in vivo and that artificial CpG clusters was sufficient to recruit Cfp1.

54. Lee J-H, Skalnik DG. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J. Biol. Chem. 2005; 280:41725–41731. [PubMed: 16253997]

55. Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. Cloning of a Mammalian Transcriptional Activator That Binds Unmethylated CpG Motifs and Shares a CXXC Domain with DNA Methyltransferase, Human Trithorax, and Methyl-CpG Binding Domain Protein 1. Molecular and cellular biology. 2000; 20:2108–2121. [PubMed: 10688657]

56. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin S-P, Allis CD, et al. DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature. 2007; 448:714–717. [PubMed: 17687327]

57. Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ. CpG islands recruit a histone H3 lysine 36 demethylase. Mol Cell. 2010; 38:179–190. [PubMed: 20417597] •The authors found that the CxxC zinc finger protein KDM2A is targeted to CpG island chromatin in vivo to help define the chromatin landscape observed at CpG islands.

58. Zhou JC, Blackledge NP, Farcas AM, Klose RJ. Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA. Mol Cell Biol. 2012; 32:479–489.

[PubMed: 22083960] ••This study used chromatin templates in vitro to show that CpG islands are targeted by the CxxC domain through specific interaction with linker DNA between nucleosomes.

59. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes Dev. 2011; 25:1010–1022. [PubMed: 21576262]
60. Ku M, Koche RP, Rheinbay E, Mendenhall EM, Endoh M, Mikkelsen TS, Presser A, Nusbaum C, Xie X, Chi AS, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. 2008; 4:e1000242. [PubMed: 18974828]
61. Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE. GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. 2010; 6:e1001244. [PubMed: 21170310]
62. Roh T-Y, Cuddapah S, Zhao K. Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev. 2005; 19:542–552. [PubMed: 15706033]
63. Alekseyenko AA, Ho JWK, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E, et al. Sequence-Specific Targeting of Dosage Compensation in Drosophila Favors an Active Chromatin Context. PLoS Genet. 2012; 8:e1002646. [PubMed: 22570616]
64. Narlikar L, Gordân R, Hartemink AJ. A nucleosome-guided map of transcription factor binding sites in yeast. PLoS Comput. Biol. 2007; 3:e215. [PubMed: 17997593]
65. Pique-Regi R, Degner JF, Pai AA, Gaffney DJ, Gilad Y, Pritchard JK. Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. 2011; 21:447–455. [PubMed: 21106904]
66. Boyle AP, Song L, Lee B-K, London D, Keefe D, Birney E, Iyer VR, Crawford GE, Furey TS. High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. Genome Res. 2011; 21:456–464. [PubMed: 21106903]
67. Berger MF, Bulyk ML. Universal protein-binding microarrays for the comprehensive characterization of the DNA-binding specificities of transcription factors. Nat Protoc. 2009; 4:393–411. [PubMed: 19265799]
68. Liu J, Stormo GD. Combining SELEX with quantitative assays to rapidly obtain accurate models of protein-DNA interactions. Nucleic Acids Res. 2005; 33:e141. [PubMed: 16186128]
69. Liu X, Noll DM, Lieb JD, Clarke ND. DIP-chip: rapid and accurate determination of DNA-binding specificity. Genome Res. 2005; 15:421–427. [PubMed: 15710749]

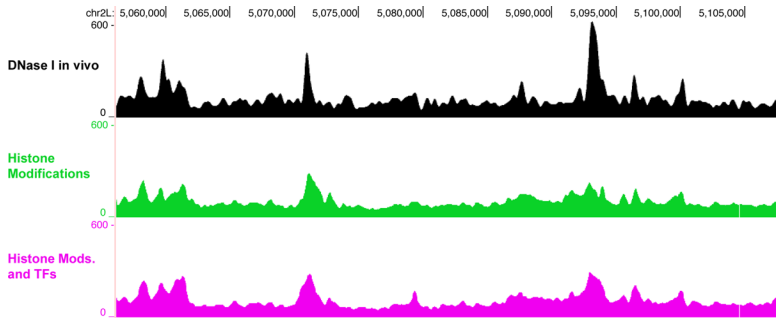

Figure 1. DNase I intensity can be modeled using histone marks and TF binding data

DNase I hypersensitivity landscape is inferred by models that use histone modification profiles and TF profiles. Incorporating non-histone chromatin-bound factors into the model increases accuracy, which is consistent with the role of TFs having an additive effect upon DNase I hypersensitivity [5,12].

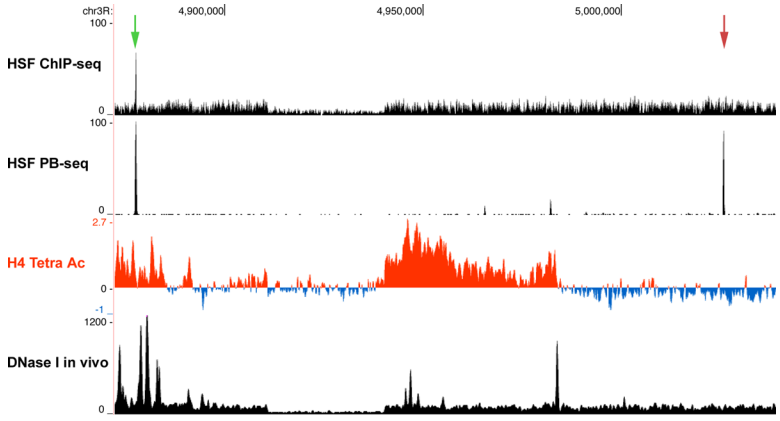

Figure 2. HSF discriminates between potential binding sites based on the pre-existing chromatin state

This region of chromosome 3R contains two strong potential HSF-binding sites (green and red arrows), measured by an in vitro protein/DNA-binding assay (PB-seq) [14]. Although the HSF-free motif (red arrow) binds with comparable affinity in vitro (PB-seq), chromatin structure restricts HSF occupancy in vivo (ChIP-seq). HSEs that are enriched for H4 acetylation and DNase I hypersensitivity during non-HS are preferentially bound by HSF in vivo (green arrow).

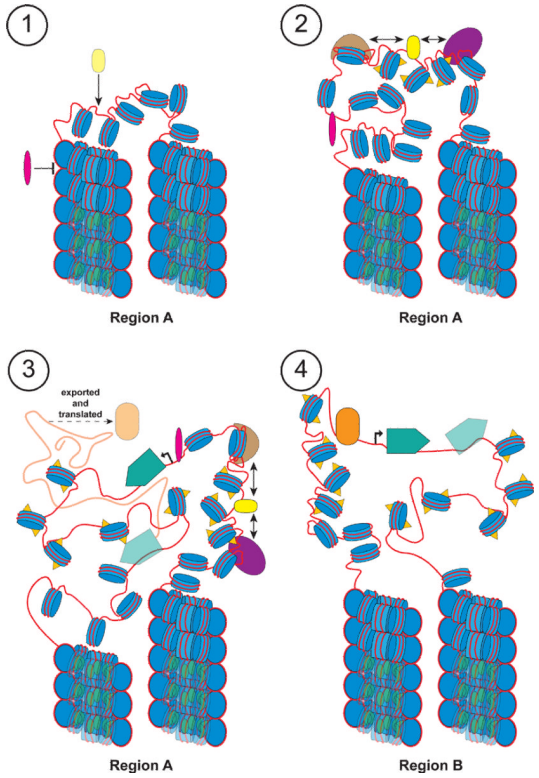

Figure 3. DNA sequence directs chromatin structure and allows transcription factor binding to precipitate regulatory cascades

(A) A CpG island is refractory to higher-order chromatin compaction (unraveled chromatin). A regulatory or developmental cascade is precipitated by the activation or expression of a transcription factor (yellow rounded rectangle) targeted to elements within the CpG island. Another TF remains (pink ellipse) unable to access elements with the H1 linker histone (green, interior crescents) condensed chromatin. (B) Targeting of the TF to the CpG island directly or indirectly results in the recruitment of nucleosome remodeling factors (brown crescent) and histone acetyltransferases (purple ellipse), which further decondenses the region and allows the other TF (pink ellipse) access to its cognate element. (C) Binding of the second TF (pink ellipse) directly or indirectly causes the recruitment and productive elongation of RNA Polymerase II (green pentagon) to a gene encoding a third TF, which is subsequently transcribed (transparent orange line) and translated (orange rounded rectangle). This causes further nucleosome loss, histone modifications (orange triangles), and decondensation of the Region A locus. (D) The translated protein (orange rounded rectangle), encoded by region A, is targeted in *trans* to Region B, which is highly acetylated and contains a transcriptionally engaged paused RNA Polymerase II. This binding event leads to the release of the paused RNA Polymerase II and activation of the gene.