nature genetics

Nucleosome dynamics define transcriptional enhancers

Housheng Hansen He[1,2,6], Clifford A Meyer[1,6], Hyunjin Shin[1], Shannon T Bailey[2], Gang Wei[3], Qianben Wang[2,5], Yong Zhang[1,5], Kexin Xu[2], Min Ni[2], Mathieu Lupien[2,5], Piotr Mieczkowski[4], Jason D Lieb[4], Keji Zhao[3], Myles Brown[2] & X Shirley Liu[1]

Chromatin plays a central role in eukaryotic gene regulation. We performed genome-wide mapping of epigenetically marked nucleosomes to determine their position both near transcription start sites and at distal regulatory elements, including enhancers. In prostate cancer cells, where androgen receptor binds primarily to enhancers, we found that androgen treatment dismisses a central nucleosome present at androgen receptor binding sites that is flanked by a pair of marked nucleosomes. A new quantitative model built on the behavior of such nucleosome pairs correctly identified regions bound by the regulators of the immediate androgen response, including androgen receptor and FOXA1. More importantly, this model also correctly predicted previously unidentified binding sites for other transcription factors present after prolonged androgen stimulation, including OCT1 and NKX3-1. Therefore, quantitative modeling of enhancer structure provides a powerful predictive method to infer the identity of transcription factors involved in cellular responses to specific stimuli.

© 2010 Nature America, Inc. All rights reserved.

Transcription in eukaryotes is regulated by transcription factors that associate with the genome in a cell-type- and condition-specific manner. Chromatin organization forms part of the basis for this cell-type specificity by allowing or denying transcription factor access to DNA. The basic units of chromatin structure are the nucleosomes, which are known to restrict the *in vivo* access of certain classes of transcription factors[1]. Intensive work has been done to reveal the correlation between nucleosome position, histone modification and gene expression[2–4]. Genome-wide nucleosome occupancy maps have been generated for *Saccharomyces cerevisiae*[5–7], *Drosophila melanogaster*[8] and *Caenorhabditis elegans*[9], but high-quality human nucleosome occupancy data is more difficult to acquire because of the large size of the human genome[10,11]. Although the nucleosome positioning patterns are well established at transcription start sites (TSS), they are less well known at enhancers. Functional enhancers are *cis*-regulatory DNA elements that are independent of orientation and position and can act at variable distances from the TSS of the genes they regulate[12,13]. Monomethylated H3K4 (H3K4me) has been shown to be associated with transcription factor binding at enhancers, trimethylated H3K4 (H3K4me3) with the TSS, and dimethylated H3K4 (H3K4me2) with both the TSS and enhancers[14,15]. To characterize the pattern of nucleosome positioning at enhancers, we used nucleosome-resolution chromatin immunoprecipitation deep sequencing (ChIP-Seq) of H3K4me, H3K4me2 and H3k4me3 in the prostate cancer cell line LNCaP in response to a time-course of stimulation by the androgen receptor agonist 5-α-dihydrotestosterone (DHT).

By comparing regions of enriched histone modification (**Supplementary Table 1**) to androgen receptor and FOXA1 binding sites (after 4 h of treatment with DHT) and the promoter regions of genes from the RefSeq database, we found H3K4me2 ChIP-Seq to be the most efficient in identifying both promoters and putative enhancers (**Supplementary Fig. 1**). To differentiate intergenic transcription factor–binding sites from promoters, we removed H3K4me2 regions with strong H3K4me3 ChIP-Seq signals from subsequent analyses.

We next examined whether H3K4me2 showed positional trends relative to the known androgen receptor and FOXA1 binding sites. The H3K4me2 signal (based on average tag count) was highest near androgen receptor binding sites in the absence of DHT (**Fig. 1a**). Upon DHT stimulation and concomitant androgen receptor binding, the H3K4me2 signal decreased at the binding sites and increased in the flanking regions (**Fig. 1b**). The same analysis relative to the binding sites of FOXA1 showed a bimodal tag-count distribution both before (**Fig. 1c**) and after (**Fig. 1d**) DHT treatment, consistent with the role of FOXA1 as a pioneer factor in facilitating androgen receptor binding[16].

To investigate whether nucleosome positioning could explain the pattern observed in the previous analysis, we determined the likely positions of H3K4me2-marked nucleosomes using the nucleosome positioning from sequencing (NPS) algorithm (**Supplementary Table 1**)[17]. The distance from the androgen receptor motif in the binding site to the nearest detected nucleosome increased to ~200 bp

[1]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute and Harvard School of Public Health, Boston, Massachusetts, USA. [2]Department of Medical Oncology, Dana-Farber Cancer Institute and Harvard Medical School, Boston, Massachusetts, USA. [3]Laboratory of Molecular Immunology, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, Maryland, USA. [4]Department of Biology, Carolina Center for the Genome Sciences and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina, USA. [5]Present addresses: Department of Molecular and Cellular Biochemistry and the Comprehensive Cancer Center, Ohio State University College of Medicine, Columbus, Ohio, USA (Q.W.); School of Life Science and Technology, Tongji University, Shanghai, China (Y.Z.); and Department of Genetics, Norris Cotton Cancer Center, Dartmouth Medical School, Lebanon, New Hampshire, USA (M.L.). [6]These authors contributed equally to this work. Correspondence should be addressed to X.S.L. (xsliu@jimmy.harvard.edu) or M.B. (myles_brown@dfci.harvard.edu).

Received 2 October 2009; accepted 9 February 2010; published online 7 March 2010; doi:10.1038/ng.545

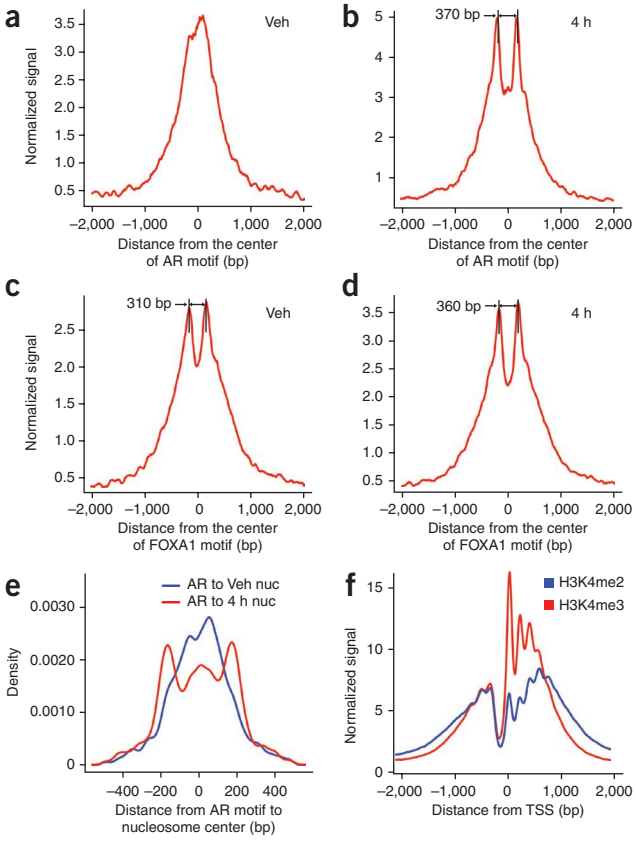

Figure 1 Signal distribution and nucleosome position analysis in the androgen receptor and FOXA1 binding regions identified by ChIP-chip experiments and the TSS. (**a**–**d**) H3K4me2 signal distribution relative to the center of the androgen receptor (AR) motif (**a**,**b**) and FOXA1 motif (**c**,**d**) in the binding regions. The *x* axis represents the distance to the center of the best AR or FOXA1 motif match in a given binding site. The *y* axis represents normalized ChIP-Seq tag count numbers. Veh, unstimulated condition; 4 h, stimulated conditions with treatment of DHT for 4 h. (**e**) Distance from the AR motif to the center of the nearest nucleosome in the AR binding sites under vehicle (red) and 4 h after DHT stimulation (blue). (**f**) H3K4me2 and H3K4me3 signal distribution relative to the TSS.

after DHT stimulation (**Fig. 1e** and **Supplementary Fig. 2a**), indicating that nucleosomes tend to be less occupied (destabilized) at the binding site itself and more occupied (stabilized) at adjacent nucleosomes. Notably, the locations of the most positioned nucleosomes are concordant between DHT treatment and vehicle control (**Supplementary Fig. 2b**). This suggests that before androgen receptor activation, androgen receptor binding loci are already marked with two well-positioned H3K4me2-containing nucleosomes that are 250–450 bp apart and flank the precise binding sites, along with a well-positioned nucleosome occluding the binding site itself. After androgen receptor activation, H3K4me2-modified nucleosomes are destabilized at the androgen receptor binding sites and are better positioned at the two flanking loci. Although the chromatin structure relative to TSS is characterized by a nucleosome-free region immediately upstream and a series of well positioned nucleosomes downstream (**Fig. 1f**), we found that in general, only two well-positioned nucleosomes are present at transcription factor–bound enhancers (**Fig. 1b–d**).

To validate our observations and rule out the possibility of ChIP-Seq artifacts or loss of the H3K4me2 mark rather than loss of the nucleosome, we conducted H3K4me2 ChIP–quantitative PCR (qPCR) and input DNA–qPCR on five androgen receptor binding sites that showed these H3K4me2 patterns near *TMPRSS2*, *STK39*, *KLK3*, *TMC6* and *TRIM35*. In all cases, the qPCR results showed that the overall nucleosome occupancy decreased over the transcription factor binding sites and increased in the flanking regions to the same extent as for the H3K4me2-marked nucleosomes. We examined two of these cases, *TMPRSS2* and *STK39*, in greater detail using primers that tile regions of interest at a finer resolution (**Fig. 2b**) and obtained similar results.

In order to determine whether the pattern of nucleosome positioning we observed at the androgen receptor binding sites could be applied more broadly to other transcription factors, we used the change in H3K4me2-marked nucleosomes at androgen receptor binding sites after acute androgen stimulation to develop a general model of nucleosome positioning at enhancers. We identified ~65,000 well-positioned nucleosome pairs (after 4 h of DHT treatment) separated by the characteristic distance of 250–450 bp in which promoter-proximal pairs were removed on the basis of having greater levels of H3K4me3 than H3K4me2 (**Fig. 3a**). Based on the analysis of the androgen receptor binding sites, we developed a quantitative model reflecting the changes in the H3K4me2 signal in the flanking nucleosomes and in the region between them (**Fig. 3a**). Running the model resulted in a nucleosome stabilization-destabilization (NSD) score *S* (defined in **Fig. 3a** and having the distribution shown in **Supplementary Fig. 3**) for each pair of appropriately spaced nucleosomes. When we ranked all the nucleosome pairs by NSD score and grouped them into bins of 500, we found that the top-scoring bins showed the highest enrichment at androgen receptor binding sites (**Fig. 3b**).

To further test the functional relevance of the regions identified by our model, we examined evolutionary conservation across the 5,000 highest-scoring paired nucleosomes. We saw three phastCons conservation peaks—one major peak at the nucleosome-depleted regions between the paired nucleosomes and peaks at the regions flanking each of these nucleosomes (**Fig. 3c**). This suggests evolutionary pressure not only on the transcription factor–binding sites between the paired nucleosomes but also on the regions immediately outside the paired nucleosomes.

To investigate the nature of nucleosome depletion in the regions between the paired nucleosomes, we studied the DNA sequence features in these regions. We observe that, consistent with previous models[18,19], simple A/T content and AA/TT/TA/AT dinucleotides were depleted in nucleosome-enriched regions and enriched in nucleosome-depleted regions, whereas GC dinucleotides showed the opposite trend (**Fig. 3d**). In addition, the stabilization of nucleosomes flanking the transcription factor–binding sites supports a model in which binding of non-nucleosomal proteins such as transcription factors forms boundaries that direct the *in vivo* positioning of nearby nucleosomes[20]. A recent study also suggests that nucleosomes containing both H2A.Z and H3.3 are intrinsically labile, therefore facilitating the access of transcription factors at regulatory sites *in vivo*[21]. We performed H2A.Z ChIP-qPCR at five representative androgen receptor binding sites (**Supplementary Fig. 4**). The results showed that H2A.Z is enriched at the central nucleosome as compared to the two flanking nucleosomes for all the five sites tested, suggesting that the central nucleosome may be more labile. Notably, the mean positions of the paired nucleosomes at androgen receptor binding loci appear to be the same in both bound (DHT) and unbound (vehicle) conditions (**Fig. 2** and **Supplementary Fig. 2b**), suggesting the existence of a common enhancer architecture that enables access of transcription factors to DNA.

© 2010 Nature America, Inc. All rights reserved.
npg

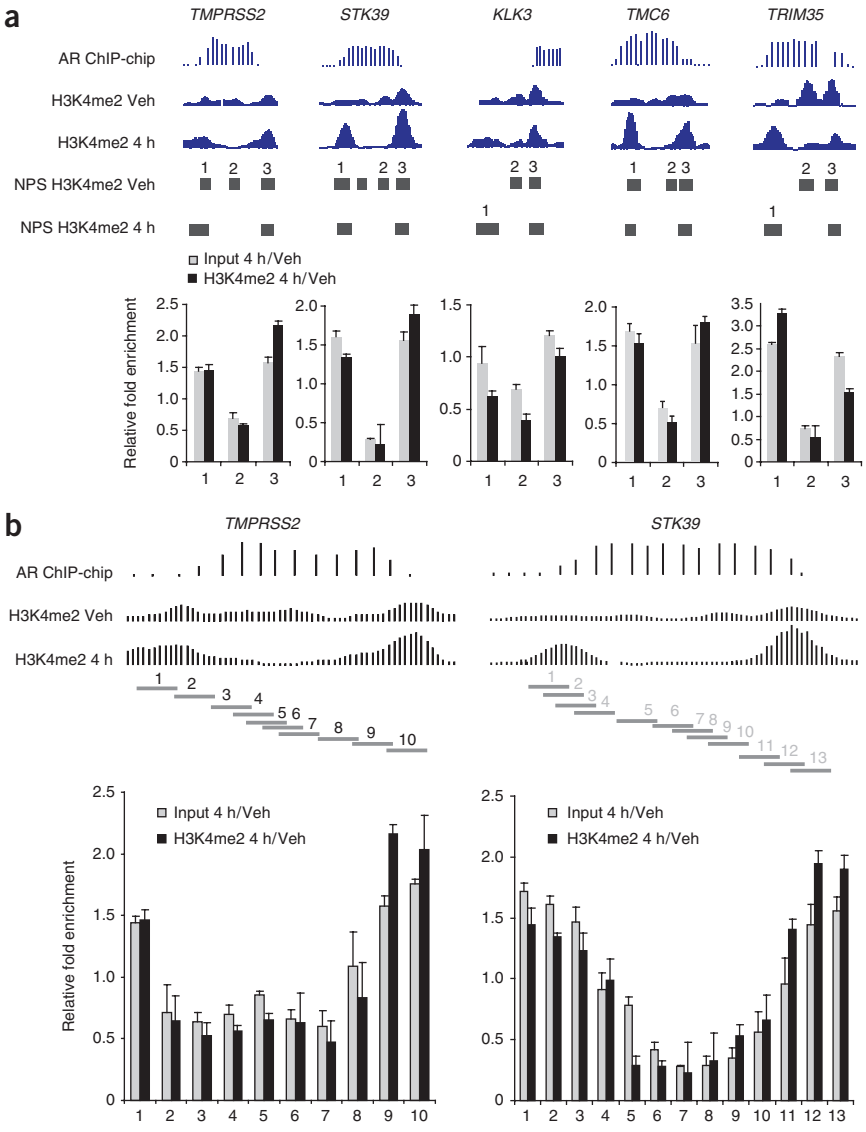

Figure 2 qPCR validation of the nucleosomes stabilized-destabilized around androgen receptor (AR) binding sites. (**a**) Five AR binding sites near the genes *TMPRSS2*, *STK39*, *KLK3*, *TMC6* and *TRIM35*. AR ChIP-chip, AR ChIP-chip signals; H3K4me2 Veh and H3K4me2 4h, H3K4me2 ChIP-Seq signals before and after 4 h of DHT treatment. Input 4 h/Veh, qPCR assay of nucleosome fold change for DHT treatment relative to vehicle; H3K4me2 4 h/Veh, qPCR assay of fold change for H3K4me2 signal for DHT treatment relative to vehicle. Error bar, s.d. Each horizontal bar represents a NPS peak region. (**b**) Detailed qPCR analysis of the AR binding sites near *TMPRSS2* and *STK39*. Each horizontal bar represents a qPCR amplification region.

To determine whether our model could be used to impute the identity of transcription factors that bind between nucleosome pairs, we searched for motifs in the 1,000 top NSD-scoring nucleosome-paired regions. The top motifs identified were from the forkhead and steroid-receptor families (**Fig. 3e**, **Supplementary Fig. 5a**,**b** and **Supplementary Table 2**), which were previously shown to be responsible for the androgen response in prostate cancer[16]. Notably, this approach predicted androgen receptor binding at several sites that were not previously identified by ChIP-chip, and all sites in a representative sample were validated by ChIP-qPCR (**Fig. 4a** and **Supplementary Fig. 6a**).

If the model is valid more generally, it should identify key transcription factors regulating the response of a cell population to a stimulus using the H3K4me2 and H3K4me3 nucleosome-resolution ChIP-Seq data alone. Because the response to 4 h of androgen exposure to cells in the LNCap line was dominated by androgen receptor and FOXA1, we generated nucleosome-resolution H3K4me2 ChIP-Seq data at 16 h after DHT treatment; we predicted that secondary transcriptional responses would dominate at this point. After applying our prediction model, we found NKX3-1 and OCT1 motifs to be highly significantly associated with high NSD-scoring regions ($P < 1 \times 10^{-7}$; **Supplementary Fig. 5c**,**d** and **Supplementary Table 2**), whereas the androgen receptor motif was not (**Fig. 3f** and **Supplementary Table 2**). *NKX3-1* is a homeobox gene involved in the normal prostate development that marks the prostate luminal epithelial stem cell and is a putative tumor-suppressor gene in the prostate[22,23]. OCT1 has been shown previously to collaborate with androgen receptor at a subset of androgen receptor binding loci in LNCaP cells[24]. Although OCT1 is constitutively expressed in LNCaP cells, NKX3-1 was induced fourfold by androgens at both the 4 h and 16 h time points[25].

To test whether these factors were truly differentially bound, we selected sites with high NSD scores (at 16 h compared to 4 h) and central sequences closely resembling the NKX3-1 or OCT1 motifs defined in the TRANSFAC database (**Fig. 3e**). ChIP-qPCR was performed to compare binding under vehicle conditions to that after DHT stimulation (at 4 h and 16 h). DHT-dependent NKX3-1 recruitment was validated for 18 out of 22 selected regions to a degree comparable to that of two known[26] NKX3-1 binding sites (**Fig. 4b** and **Supplementary Fig. 6b**). Although previously identified OCT1 binding sites have been located in close proximity to androgen receptor binding sites[24], our model predicted a set of putative DHT-responsive OCT1 binding sites that were independent of androgen receptor binding. ChIP-qPCR of these sites showed a strong response to DHT stimulation, with all nine sites having greater enrichment of OCT1 binding at 16 h compared with 4 h (**Fig. 4c** and **Supplementary Fig. 6c**).

We investigated whether the genomic regions identified with our model might be of regulatory importance. We defined '4 h' and '16 h' sites as the 5,000 regions with the highest NSD scores after 4 h of DHT treatment compared to the vehicle and at 16 h compared to 4 h of DHT treatment, respectively. We then examined gene expression microarray assays obtained for vehicle treatment and the 4 h and 16 h DHT treatments and compared imputed differential binding with differential gene expression. At both 4 h and 16 h of DHT treatment, the differentially expressed genes were more highly associated with imputed binding sites than with genes that were not differentially expressed (**Fig. 4d**). Further analysis showed that the likelihood of a gene being upregulated increased with the number and score of paired nucleosome sites in the vicinity of the TSS (**Fig. 4**,**f**). In contrast, when we examined the

© 2010 Nature America, Inc. All rights reserved.

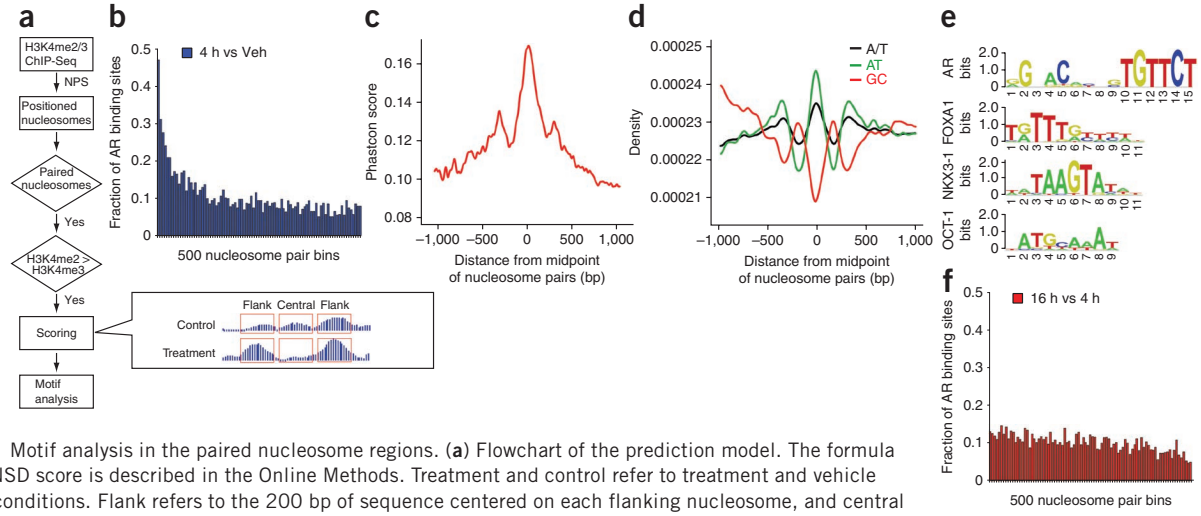

Figure 3 Motif analysis in the paired nucleosome regions. (**a**) Flowchart of the prediction model. The formula for the NSD score is described in the Online Methods. Treatment and control refer to treatment and vehicle control conditions. Flank refers to the 200 bp of sequence centered on each flanking nucleosome, and central refers to the sequence between these regions. (**b**) The fraction of androgen receptor binding sites in NSD score ranked paired nucleosome bins with decreasing score (at 4 h compared to vehicle). Paired nucleosome regions are ranked by scores representing the differences in H3K4me2 tag counts before and after DHT treatment. These ranked regions are grouped into bins of 500. Shown here is the number of regions in each bin that overlap with androgen receptor ChIP-chip–enriched regions. (**c**) Evolutionary conservation in the vicinity of the 5,000 highest-scoring nucleosome pairs. Mean phastCons scores representing DNA sequence conservation over 17 species is plotted as a function of the distance from the midpoint between paired nucleosomes. (**d**) DNA sequence content associated with nucleosome positioning. The 5,000 highest-scoring paired nucleosome regions, aligned at the midpoint, were analyzed for simple DNA sequence features: the distribution of A/T mononucleotides (black), GC dinucleotides (red) or AT dinucleotides (green). (**e**) Logos of androgen receptor (AR), FOXA1, NKX3-1 and OCT1 motifs from TRANSFAC library. (**f**) The fraction of AR binding sites in score-ranked paired nucleosome bins with decreasing score (at 16 h compared to 4 h).

relationship between number and score of paired nucleosome sites for downregulated genes, we found no correlation (**Supplementary Fig. 7**). These results suggest that the sites with high NSD scores are functional enhancers that have a functional role in gene regulation.

By performing ChIP-Seq for H3K4me2 and H3K4me3, we profiled at high resolution the changes in nucleosome occupancy that occur at enhancers in a human prostate cancer cell line in response to androgen stimulation. Analysis of nucleosome occupancy near

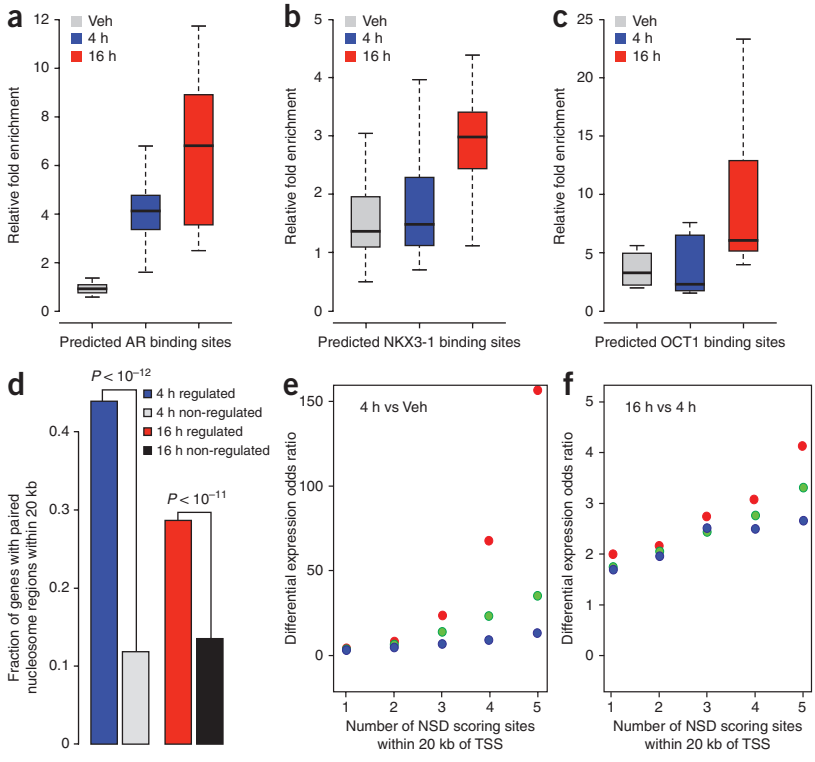

Figure 4 ChIP-qPCR and gene expression analysis of NSD scoring sites. (**a**–**c**) ChIP-qPCR validation of predicted androgen receptor (AR) (**a**), NKX3-1 (**b**) and OCT1 (**c**) binding sites. Box plots were generated from ChIP-qPCR data obtained from three independent experiments testing 10 sites for AR, 22 sites for NKX3-1 and 9 sites for OCT1. The individual ChIP-qPCR assays are shown in **Supplementary Figure 5**. (**d**) Correlation of paired nucleosome regions with gene expression. The fraction of differentially regulated genes with paired nucleosome regions within 20 kb is shown. The top 5,000 paired nucleosome regions were selected under the conditions of DHT treatment for 4 h versus vehicle and DHT treatment for 16 h versus DHT treatment for 4 h. Differentially regulated genes were identified as described in the Online Methods. 4 h regulated, fraction of DHT 4 h versus vehicle-treated differentially regulated genes having at least one DHT 4 h versus vehicle-paired nucleosome region within 20 kb of the transcription start site. 4 h non-regulated, fraction of non-regulated genes under the same condition. 16 h regulated and 16 h non-regulated, fractions under DHT treatment for 16 h versus 4 h. (**e**,**f**) Correlation of score and number of NSD scoring sites and upregulated gene expression, 4 h versus vehicle treatment (**e**) and 16 h versus 4 h treatment (**f**). The *x* axis represents the lower bound *n* of the number of sites within 20 kb of the TSS of a gene, and the *y* axis represents the odds ratio calculated by the formula (upregulated genes with at least *n* sites/non-regulated genes with at least *n* sites)/(all upregulated genes/all non-regulated genes). Red, green and blue dots represent the top 5,000, 10,000 and 20,000 NSD score sites, respectively.

© 2010 Nature America, Inc. All rights reserved.

androgen receptor and FOXA1 binding sites revealed a striking pattern of nucleosome stabilization in a pair of nucleosomes flanking the binding site together with elimination of a nucleosome at the binding site itself. We found several intrinsic characteristics of the central nucleosome that distinguish it from the flanking ones: the central nucleosome's sequence is more evolutionarily conserved and has a higher A/T content than those flanking it, and its histone octamer is more likely to contain the H2A.Z variant. This suggests that it may be intrinsically less stable than the flanking nucleosomes. Thus the apparent differences in nucleosome stability may be the result of the combination of DNA sequence, histone octamer composition and transcription factor binding. We developed a new quantitative model and scoring function, the NSD score, that not only correctly identified the sites of androgen receptor binding but also made it possible to predict the binding of other factors, including NKX3-1 and OCT1, that mediate secondary transcriptional responses. Thus, this model defines the characteristic pattern of nucleosome occupancy changes associated with enhancers and can be used to infer dynamic transcription factor binding events that occur when a cell population transitions between states.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturegenetics/.

Accession codes. Raw sequence tags and processed peak files have been deposited with accession code GSE20042 in the NCBI short-read archive.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

This work was supported by grants from US National Institutes of Health (1R01 HG004069-02 to X.S.L., and 2P50 CA090381-06 to X.S.L. and M.B.), the Department of Defense (W81XWH-07-1-0037 to X.S.L.) and the Prostate Cancer Foundation (to M.B.).

AUTHOR CONTRIBUTIONS

H.H.H., C.A.M., K.Z., J.D.L., X.S.L. and M.B. designed the experiments. H.H.H., S.T.B., G.W., Q.W., K.X., M.N., M.L. and P.M. performed the experiments. C.A.M., H.H.H., H.S. and Y.Z. performed data analysis. C.A.M., H.H.H., X.S.L., M.B., J.D.L. and M.L. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics/.
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/.

1. Beato, M. & Eisfeld, K. Transcription factor access to chromatin. *Nucleic Acids Res.* **25**, 3559–3563 (1997).
2. Narlikar, L., Gordan, R. & Hartemink, A.J. A nucleosome-guided map of transcription factor binding sites in yeast. *PLOS Comput. Biol.* **3**, e215 (2007).
3. Ozsolak, F. *et al.* Chromatin structure analyses identify miRNA promoters. *Genes Dev.* **22**, 3172–3183 (2008).
4. Guttman, M. *et al.* Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227 (2009).
5. Yuan, G.C. *et al.* Genome-scale identification of nucleosome positions in *S. cerevisiae*. *Science* **309**, 626–630 (2005).
6. Lee, W. *et al.* A high-resolution atlas of nucleosome occupancy in yeast. *Nat. Genet.* **39**, 1235–1244 (2007).
7. Mavrich, T.N. *et al.* A barrier nucleosome model for statistical positioning of nucleosomes throughout the yeast genome. *Genome Res.* **18**, 1073–1083 (2008).
8. Mavrich, T.N. *et al.* Nucleosome organization in the *Drosophila* genome. *Nature* **453**, 358–362 (2008).
9. Valouev, A. *et al.* A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. *Genome Res.* **18**, 1051–1063 (2008).
10. Ozsolak, F., Song, J.S., Liu, X.S. & Fisher, D.E. High-throughput mapping of the chromatin structure of human promoters. *Nat. Biotechnol.* **25**, 244–248 (2007).
11. Schones, D.E. *et al.* Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887–898 (2008).
12. Blackwood, E.M. & Kadonaga, J.T. Going the distance: a current view of enhancer action. *Science* **281**, 60–63 (1998).
13. Bulger, M. & Groudine, M. Looping versus linking: toward a model for long-distance gene activation. *Genes Dev.* **13**, 2465–2477 (1999).
14. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
15. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318 (2007).
16. Lupien, M. *et al.* FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970 (2008).
17. Zhang, Y., Shin, H., Song, J.S., Lei, Y. & Liu, X.S. Identifying positioned nucleosomes with epigenetic marks in human from ChIP-Seq. *BMC Genomics* **9**, 537 (2008).
18. Peckham, H.E. *et al.* Nucleosome positioning signals in genomic DNA. *Genome Res.* **17**, 1170–1177 (2007).
19. Yuan, G.C. & Liu, J.S. Genomic sequence is highly predictive of local nucleosome depletion. *PLOS Comput. Biol.* **4**, e13 (2008).
20. Kornberg, R.D. & Stryer, L. Statistical distributions of nucleosomes: nonrandom locations by a stochastic mechanism. *Nucleic Acids Res.* **16**, 6677–6690 (1988).
21. Jin, C. *et al.* H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.* **41**, 941–945 (2009).
22. Korkmaz, C.G. *et al.* Analysis of androgen regulated homeobox gene NKX3.1 during prostate carcinogenesis. *J. Urol.* **172**, 1134–1139 (2004).
23. Asatiani, E. *et al.* Deletion, methylation, and expression of the NKX3.1 suppressor gene in primary human prostate cancer. *Cancer Res.* **65**, 1164–1173 (2005).
24. Wang, Q. *et al.* A hierarchical network of transcription factors governs androgen receptor-dependent prostate cancer growth. *Mol. Cell* **27**, 380–392 (2007).
25. Wang, Q. *et al.* Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**, 245–256 (2009).
26. Liu, W. *et al.* Characterization of two functional NKX3.1 binding sites upstream of the *PCAN1* gene that are involved in the positive regulation of *PCAN1* gene transcription. *BMC Mol. Biol.* **9**, 45 (2008).

© 2010 Nature America, Inc. All rights reserved.

ONLINE METHODS

ChIP-Seq and ChIP-qPCR. The prostate cancer cell line LNCaP was obtained from the American Type Culture Collection. ChIP and ChIP-Seq library construction for histone marks H3K4me1, H3K4me2 and H3K4me3 were performed as previously described[14] and the libraries were sequenced to 35 bp with the Illumina Genome Analyzer. ChIP experiments for androgen receptor, NKX3-1 and OCT1 as well as ChIP-qPCR experiments were performed as previously described[27]. The PCR primers used are listed in **Supplementary Table 3**.

Peak calling. Significantly ($P < 1 \times 10^{-5}$) enriched regions were detected using the MACS software[28] using default parameters. Mononucleosomes were detected using the NPS analysis of nucleosome positions with default parameters[17].

Androgen receptor and FOXA1 binding sites definition. Data sets of receptor (false discovery rate (FDR) of 5%) and FOXA1 (FDR of 1%) binding sites were from our previous works[16,25].

Model for identifying differential transcription factor–binding locations. NPS was used to identify nucleosome positions based on treatment condition data. Nucleosome intervals were defined as the 200-bp region centered on the center position of the NPS-identified nucleosomes. Nucleosome pairs with center positions lying in a range between 250 bp and 450 bp from the NPS-identified nucleosomes were identified for further analysis. For each nucleosome pair, the number of tags in the nucleosomes and in the inter-nucleosomal region was counted.

A tag was considered to belong to a genomic interval if, when shifted 73 bp in a strand-directed direction, the entire tag fell within that interval. These pairs are then given a NSD score (S) by the formula

$$S = \left(\sqrt{n^{\text{treat}}_{\text{flank}} / n_{\text{treat}}} - \sqrt{n^{\text{control}}_{\text{flank}} / n_{\text{control}}} \right) - \left(\sqrt{n^{\text{treat}}_{\text{central}} / n_{\text{treat}}} - \sqrt{n^{\text{control}}_{\text{central}} / n_{\text{control}}} \right)$$

This method takes into account changes in H3K4me2 ChIP-Seq tag counts falling on the flanking nucleosomes as well as the region lying between them. Tag counts were scaled in proportion to the overall counts in the treatment and control samples. In this formula, n is the tag count and the superscripts 'treat' and 'control' refer to the DHT and vehicle conditions, respectively. The subscript 'flank' refers to the 200 bp of sequence centered on each flanking nucleosome, and 'central' refers to the sequence between these regions (**Fig. 3a**).

Motif statistics. Known DNA motifs that are enriched relative to the center of TF binding sites predicted by our model were identified using the z statistic described below. Motif analysis was conducted on 600-bp DNA segments, with each segment representing one nucleosome pair. Each segment was derived from a 1-kb sequence centered at the midpoint between a nucleosome pair from which a region of 200 bp centered on each of the nucleosomes was excluded. All subsequences within these sequences were scored by a TRANSFAC motif[29] and the genomic background sequence composition to identify hits above a probability cutoff was determined. Let x_i be a value between 0 and 1 that denotes the relative location of motif hit i out of n total motif hits (where 0 and 1 represent the center and edge of the sequence). We define a z score, $z = \sigma_{i=1-n}\,(x_i - 0.5)/\sqrt{(n/12)}$, that represents the positional bias of a motif toward the centers of the regions. Different integer cutoffs were tested for each motif, and the cutoff resulting in the highest z was selected. This statistic is based on the assumptions that insignificant DNA motifs will be uniformly distributed across the sequences and the null distribution of $\sigma_{i=1-n}\,x_i$ can be estimated as the n-fold convolution of uniform density functions.

Gene expression analysis. Affymetrix U133 Plus 2.0 microarray data were analyzed using the robust multi-array average (RMA)[30] using a custom chip definition file (CDF) (v11) mapping to RefSeq genes[31]. Differentially-expressed genes were identified using the significance analysis of microarrays algorithm at a local FDR of 10%[32]. The nearest putative TF binding sites were associated with each non-redundant RefSeq gene transcription start site. The statistical significance association between putative TF binding sites was calculated using Fisher's exact test in which genes were categorized according to whether they were differentially expressed and whether they had at least one putative TF binding site within 20 kb of the transcription start site.

27. Wang, Q., Carroll, J.S. & Brown, M. Spatial and temporal recruitment of androgen receptor and its coactivators involves chromosomal looping and polymerase tracking. *Mol. Cell* **19**, 631–642 (2005).
28. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
29. Matys, V. *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34**, D108–D110 (2006).
30. Irizarry, R.A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
31. Dai, M. *et al.* Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* **33**, e175 (2005).
32. Tusher, V.G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98**, 5116–5121 (2001).

© 2010 Nature America, Inc. All rights reserved.

doi:10.1038/ng.545