Transcription factor binding predicts histone modifications in human cell lines

Dan Benveniste[a,1], Hans-Joachim Sonntag[b,1], Guido Sanguinetti[a,c,2], and Duncan Sproul[b,2]

[a]School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, United Kingdom; [b]Medical Research Council Human Genetics Unit, Medical Research Council Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, United Kingdom; and [c]Synthetic and Systems Biology, University of Edinburgh, Edinburgh EH9 3JD, United Kingdom

Gene expression in higher organisms is thought to be regulated by a complex network of transcription factor binding and chromatin modifications, yet the relative importance of these two factors remains a matter of debate. Here, we show that a computational approach allows surprisingly accurate prediction of histone modifications solely from knowledge of transcription factor binding both at promoters and at potential distal regulatory elements. This accuracy significantly and substantially exceeds what could be achieved by using DNA sequence as an input feature. Remarkably, we show that transcription factor binding enables strikingly accurate predictions across different cell lines. Analysis of the relative importance of specific transcription factors as predictors of specific histone marks recapitulated known interactions between transcription factors and histone modifiers. Our results demonstrate that reported associations between histone marks and gene expression may be indirect effects caused by interactions between transcription factors and histone-modifying complexes.

epigenetics | gene regulation

Gene expression is the fundamental process through which genetic information is dynamically and specifically deployed within cells. It is, therefore, of vital importance to all organisms and tightly controlled at both the transcriptional and posttranscriptional levels. Consequently, the elucidation of gene-regulatory mechanisms has been a central focus of biological research, with the area of transcriptional regulation having attracted intense attention over the last four decades.

The canonical players in transcriptional regulation are sequence-specific DNA-binding transcription factors (TFs) that modulate gene expression by facilitating or inhibiting the recruitment of RNA polymerase to gene promoters (1). This paradigm has provided a powerful unifying mechanism for transcription, validated by a large amount of experimental evidence over the last five decades (see, e.g., ref. 2). Further evidence of the power of TFs to act as master regulators of gene expression and cell identity is illustrated by their ability to reprogram differentiated fibroblasts into embryonic stem (ES) cells (3).

Research in the field of epigenetics has, however, suggested an alternative view that places posttranslational modifications of the histone subunits of nucleosomes in a central role of transcriptional regulation. The finding that particular combinations of histone modifications are associated with active and repressed gene promoters (4) has led to suggestions that a histone code controls gene expression (5, 6). Support for this hypothesis has come from the recent application of bioinformatic approaches to whole-genome measurements of both histone modifications and gene expression, which have demonstrated that gene expression can be predicted from histone modifications (7, 8). This model has generated intense interest and is part of the stimulation behind the search for epigenetic causes of human disease (9).

However, although histone modifications likely play key roles in gene expression, significant uncertainty remains as to the relative importance of chromatin-based and TF-based mechanisms of regulation. Histone modifications are themselves tightly

regulated and exhibit dynamic behavior during cellular processes (10). Several studies have also delineated direct interactions between histone-modifying enzymes and TFs (11). Most importantly, currently known mechanisms of histone modification deposition are not sequence specific, leading some to caution against overinterpreting correlative evidence for a regulatory role of a histone code (12, 13).

In this work, we exploit the richness of the recently released Encyclopedia of DNA Elements (ENCODE) datasets (14) to interrogate the relationship between TF binding and histone modifications in a large-scale computational experiment. We find that DNA sequence is remarkably predictive of the presence of histone modifications at promoters in three different cell lines. By comparative analysis using TF chromatin immunoprecipitation followed by sequencing (ChIP-Seq) data as input, we find that histone modifications can be predicted significantly more accurately from TF-binding patterns than from DNA sequence. We also show that the predictive power of TF-binding data extends to predict histone modifications genome-wide on a large dataset of putative functional loci. Furthermore, TF-based predictors trained on data from one cell line accurately predict histone modifications in a different cell line. Our use of statistical modeling affords insights into the relative predictive power of each TF, recapitulating known interactions between TFs and histone-modifying enzymes. Our results show that the correlative evidence for a regulatory role of chromatin is equally well

Significance

The regulation of gene expression is fundamental to biology and is classically predicated on binding of transcription factor proteins to DNA. This view is challenged by large-scale studies correlating gene expression with posttranslational modifications of the histone proteins with which DNA is complexed in cells. Here, we show through a large-scale computational study that histone modifications can be predicted with remarkable accuracy from transcription factor-binding profiles, recapitulating known interactions between transcription factors and chromatin-modifying enzymes. Our results demonstrate that associations between gene expression and histone modifications do not necessarily imply a direct regulatory role for these modifications, but can be explained equally well as an indirect effect of interactions between transcription factors and chromatin-modifying enzymes.

explained as an indirect effect of TF binding and suggest that interactions between TFs and histone-modifying enzymes might be important in driving the deposition of histone modifications.

Results

To computationally explore the mechanisms responsible for the deposition of histone marks in mammalian genomes, we made use of the recently released ENCODE datasets (14). To focus on a biologically homogenous sample population of genomic elements, we initially examined which factors might determine histone modification patterns at gene promoters. Briefly, we defined 29,828 unique protein-coding transcription start sites (TSSs) from the ENSEMBL database and assigned histone modifications to each of these based on ChIP-Seq data for the three ENCODE tier 1 cell lines (H1 ES cells, K562 erythroleukemia cells, and GM12878 lymphoblastoid cells). Each TSS was assigned a positive label for a given histone mark if a ChIP-Seq peak was detected within 100 bp; if not, it was assigned a negative label. The size of the window was chosen to capture promoter proximal histone marks, which are most likely to interact with transcription. We tested promoter-associated marks that could be assigned to ≥5% of TSSs in at least one cell line. Out of the 10 histone marks assayed in all tier I ENCODE cell lines, this resulted in the following set of marks: H3K4me3, H3K4me2, H3K4me1, H3K9ac, H3K27ac, and H3K27me3. H3K4me1 was present above threshold only in the K562 cell line (6.1% of TSSs). We then further discarded H3K4me2 as it is primarily found flanking H3K4me3 at promoters (4) (Table S1). We used logistic regression (LR) to investigate which factors might determine histone modification patterns at these promoters by testing their ability to predict the presence of each histone modification independently.

DNA Sequence Predicts Presence of Histone Marks. Initially, we examined whether histone modifications might be predictable from DNA sequence alone. We therefore extracted sequence features (k-mers) from genomic regions of ±2 kb from the TSS by counting the frequency of all possible k letter words in the A, C, G, T alphabet and merging reverse complement pairs to prevent strand biases; we used $k = 6$ in our analysis as in ref. 15.

We then trained LR classifiers on these 6-mer counts using a random subset of 70% of the TSSs for each histone modification and cell line. Each classifier was then tested on its ability to predict the status of the remaining 30% of TSSs and assessed

by examination of receiver operator characteristic (ROC) curves, which plot the true-positive rate vs. the false-positive rate. To test the stability of these predictions, we repeated this procedure 10 times for each mark and cell line, computing the mean and SE in the area under the curve (AUC) for these 10 iterations. The AUC scores for predictions of all histone marks were very high in H1 cells, ranging from 0.806 for H3K27me3 to 0.918 for H3K4me3 (Fig. 1 and Table 1). Sequence-based predictions of histone marks in the other two cell lines also gave high AUCs (Fig. S1 and Table S2).

The ability of DNA sequence to predict histone marks was independent of the window size used to define TSSs as marked by a particular modification, as high AUCs were observed at multiple window sizes (Table S3). We obtained similar levels of accuracy for a selection of marks in H1 cells using a 6-mer support vector machine in place of LR (Fig. S2), suggesting that the predictability of histone modification patterns from genomic sequence is robust and independent of the method used.

TF ChIP-Seq Significantly Improves Prediction of Histone Marks over DNA Sequence. A plausible explanation for the predictability of histone modifications from sequence is that this may be a by-product of histone modifications being predictable from sequence-specific TFs. To test this, we constructed histone modification classifiers based upon data from TF ChIP-Seq experiments. TF ChIP-Seq data were downloaded for the three cell lines and filtered to remove proteins that lacked sequence-specific DNA-binding TF activity or that possessed histone-modifying activity (*SI Materials and Methods*). This resulted in data on 30 TFs assayed in H1 cells, 45 in K562, and 51 in GM12878. Of these, 17 TFs were assayed in all three cell lines. A complete list of all TFs used is given in Dataset S1. We tested the ability of TF-binding locations to predict the histone modification status of promoters by calculating input-normalized read count values for a window ±2 kb from each TSS. These read counts were then used as input features to train LR classifiers on 10 samples of 70% of the TSSs and tested on the remaining 30% of TSSs as above.

These TF-based classifiers predicted the histone modification status of TSSs with a high degree of accuracy, irrespective of the window size chosen for defining positive regions (Fig. 2A, Fig. S3, Table 1, and Tables S3 and S4). A quantitative comparison of sequence and TF-based predictions demonstrated that TF LR always significantly outperformed sequence-based LR models (Fig. 2B, $P \leq 10^{-5}$ rank-sum test). All points fall significantly
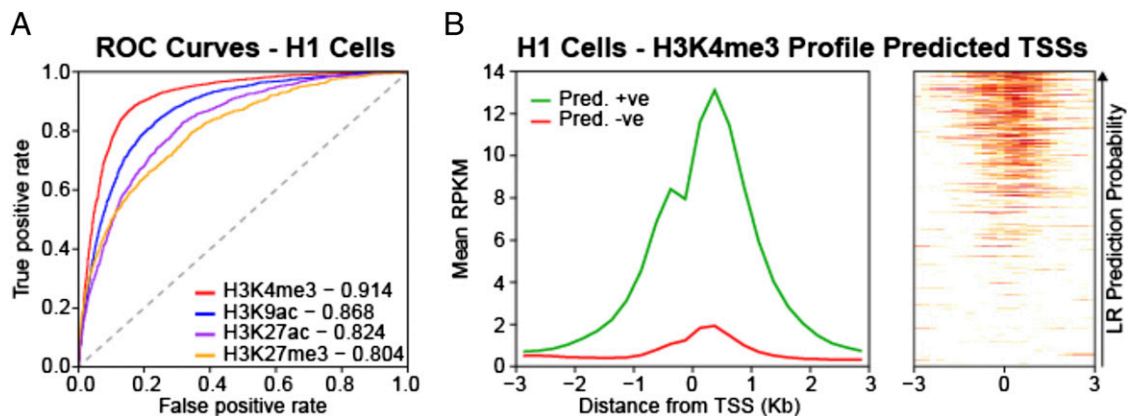

Fig. 1. Histone modifications can be predicted from DNA sequence. (A) Representative ROC curves of the performance of k-mer LR-based classifiers for histone modifications at gene promoters in H1 cells. The AUC for each task is indicated in the legend. The ROC curves shown are for a single iteration of a 70–30 split of the data. (B) H3K4me3 profile at test set promoters in H1 cells. Shown on the *Left* is the mean H3K4me3 profile at promoters predicted to be positive (green) and negative (red) for H3K4me3 in a single iteration of the analysis. The cutoff used was $P = 0.5$. The panel on the *Right* shows the profile at all of the promoters in the test set ordered by their predicted probability of being marked by H3K4me3 (white, low; red, high).

Table 1. Predictions of histone modification presence in H1 cells (mean AUC ± SE)

Mark	Sequence promoters	TF promoters	TF DNase loci	TF Enhancers
H3K4me1	N.D.	N.D.	0.854 (±0.001)	0.842 (±0.003)
H3K4me3	0.918 (±0.001)	0.950 (±0.001)	0.974 (±0.001)	0.962 (±0.001)
H3K9ac	0.867 (±0.001)	0.921 (±0.001)	0.976 (±0.001)	0.961 (±0.001)
H3K27ac	0.828 (±0.002)	0.909 (±0.001)	0.968 (±0.001)	0.950 (±0.001)
H3K27me3	0.808 (±0.002)	0.877 (±0.002)	0.916 (±0.001)	0.918 (±0.002)

above the diagonal line (equal predictive power), indicating that TFs are considerably more predictive of histone modifications than sequence. (Notice that prediction of H3K27me3 was carried out only in H1 cells, as this mark was not present at a sufficiently high number of TSSs in the other cell lines to meet our inclusion criteria.) TF-based models also resulted in a greater separation between the observed histone modification profiles at predicted positive and negative TSSs when compared to sequence-based models (Fig. 2C). Our results, therefore, demonstrate that the binding patterns of TFs accurately predict the histone modification status of mammalian gene promoters.

TF ChIP-Seq Enables Predictions Across Different Cell Lines. Histone modifications exhibit dynamic and cell-specific patterns that clearly cannot be explained on the basis of DNA sequence content. TFs do exhibit cell type specificity and, given the excellent performance of TF-based classifiers in single-cell lines, we tested to what extent the TF-based classifiers trained on data from one cell line (e.g., H1) can predict histone markings on another cell line (e.g., K562).

We therefore generated LR-based classifiers on the data for the 17 TFs that were assayed in all three cell lines. These classifiers were each able to accurately predict the histone modification status of gene promoters in the cell type they were trained upon; however, in each case, this was to a lower degree than classifiers trained on the full set of TFs available for that cell line (Fig. 3, diagonal figures). Therefore, the inclusion of more complete information on TF-binding patterns in predictive models improves the performance of these models, as would be expected if TF-binding patterns were important in determining the histone modification status of promoters.

The classifiers based upon the common TFs were then tested for the ability to predict the histone modification status across cell types. Although the cross-cell predictions produced reduced performance, we observed striking accuracy in the prediction of histone marks based on classifiers trained in a separate cell type (Fig. 3 for H3K4me3 and Fig. S4 for other marks). The reduced performance of cross-cell predictors might suggest that cell type-specific TFs play a strong role in determining histone modification profiles. This fact may be particularly prominent in the prediction of GM12878 marks from K562-trained classifiers where the greatest drop in performance was noted. We were unable to perform this analysis for H3K27me3, a repressive mark associated with Polycomb complexes, as it occupies large diffuse chromatin domains in nonstem cells (16); this resulted in too few peaks being called in either GM12878 or K562 to meet our inclusion criteria.

Overall, the results of our cross-cell analysis support our hypothesis by demonstrating the ability of data on TF-binding profiles to predict the histone modification status of mammalian gene promoters.

TFs Predictive Power Recapitulates Known TF–Histone Interactions. LR modeling produces a vector of weights determining the sensitivity of the histone mark prediction to changes in the TF input. Analysis of these weights revealed several features that corroborated reported interactions between TFs and histone-modifying enzymes (Fig. 4 and Fig. S5). In H1 ES cells, SP4 was the highest weighted predictor of H3K4me3 marking and the presence of its paralogue specificity protein 1 (SP1) was also positively predictive of H3K4me3. These TFs were also positively predictive of the other two active marks analyzed (H3K9ac and H3K27ac), and SP1 received positive weights in predicting all three active marks in K562 and GM12878 cells (Fig. S5). SP1 binding on human chromosomes 21 and 22 has previously been associated with CpG island promoters (17), which are generally enriched for H3K4me3 (18). The top predictive TF for H3K27me3 in H1 cells, transcription factor 12 (TCF12), has
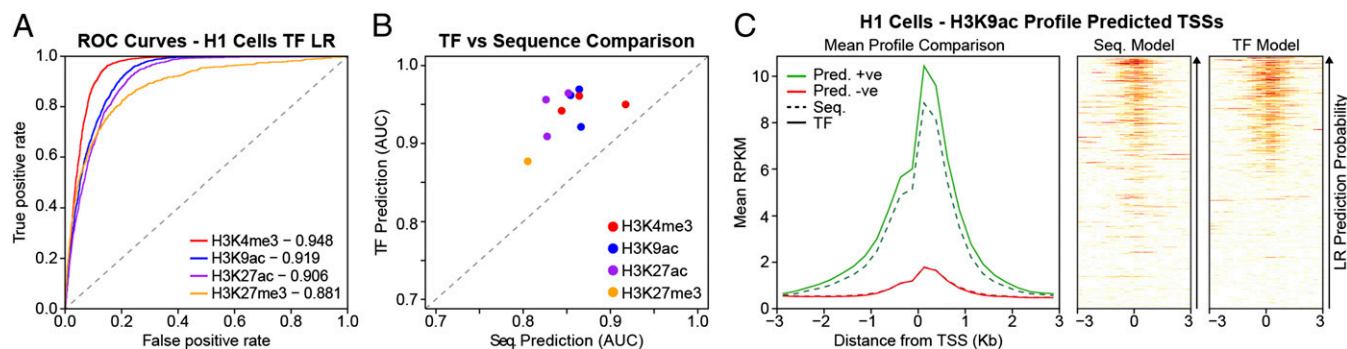

Fig. 2. Histone modifications can be predicted from TF-binding data. (A) Representative ROC curves of the performance of TF ChIP-Seq LR-based classifiers for histone modifications at gene promoters in H1 cells. The AUC for each task is indicated in the legend. The ROC curves shown are for a single iteration of a 70–30 split of the data. (B) TF-binding prediction outperforms DNA sequence. Shown is a scatter plot comparing the AUCs achieved from TF-binding LR classifiers (y axis) and DNA sequence LR classifiers (x axis). Each point represents the mean of 10 computational experiments for one histone mark in one cell line. (C) H3K9ac profile at test set promoters. Shown on the *Left* is the mean H3K9ac profile at test set promoters predicted to be positive (green) and negative (red) for H3K9ac. Both predictions were performed on the same set of promoters. The dashed lines are predictions from sequence, and the solid lines are predictions from TFs. Notice the higher average of TF predicted positive marks. On the *Right* are heat maps of H3K9ac levels (white, low; red, high) at the individual promoters ordered by predicted positive probability (increasing along the y axis) as provided by sequence LR (*Center*) and by TF LR (*Right*).
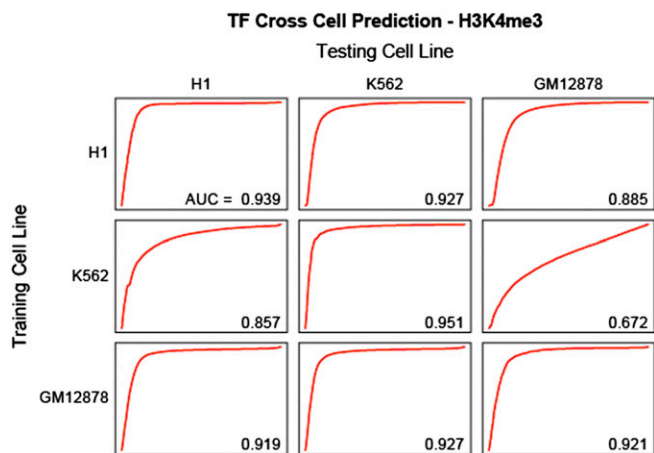
TF Cross Cell Prediction - H3K4me3

Fig. 3. TF-binding data can predict histone marks across cell lines. Shown are ROC curves for TF-binding LR classifiers for H3K4me3. Classifiers were trained on the cell line indicated by the row and tested on each of the three cell lines (indicated by the column). The AUC is given on the plot in each case.

recently been reported as a repressor of E-cadherin in association with the H3K27me3 methylase enhancer of zeste homologue 2 (EZH2) (19).

Interestingly, the core pluripotency factor NANOG was predictive of the activation-associated marks in H1 cells and POU5F1 (OCT4) was predictive of H3K9ac and H3K27ac, consistent with their roles in specifying ES cell identity (3). In K562 cells, the oncogene CMYC was the second strongest predictor of activation-associated histone marks consistent with this cancer cell line being derived from an erythroleukemia (20). Two lymphocyte-specific TFs, nuclear factor of activated T-cells, calcineurin-dependent 1 (NFATC1) and interferon regulatory factor 4 (IRF4), were the strongest predictors of activation-associated marked promoters in GM12878 lymphoblastoid cells. IRF4 in particular has been shown to be critical for the development of B cells from which lymphoblastoid cell lines are derived (21). Although NFATC1 is generally regarded as a T-cell–specific factor, its ablation in B cells affects both their proliferation and differentiation (22). We also note that these key predictors of histone modifications in GM12878 cells were not assayed in K562 cells, potentially explaining the drop in cross-cell predictive performance of classifiers trained in K562 cells and tested on GM12878 cells (Fig. 3). Taken together, these results suggest that cell type-specific TFs are strong predictors of histone modifications in the analyzed cell lines, consistent with the cell type-specific distribution of histone marks observed in epigenomic profiling studies.

TF-Binding Patterns Predict Histone Modifications Genome-Wide. To comprehensively test the capacity of TF-binding patterns to predict histone modification states, we tested our predictive approach on a diverse set of 1.2 million functional elements defined by the ENCODE consortium using DNase hypersensitivity assays (14). This set includes the total promoters and putative distal regulatory elements defined from 125 different cell lines. H3K4me1 was also included in this analysis as it is regarded as a defining mark of enhancers. The strong predictive power of TFs for histone modifications was confirmed in this large-scale analysis, with TF LR achieving AUCs ranging from 0.862 to 0.970 in H1 cells (Table 1 and Tables S5 and S6).

To test whether this performance reflected the ability of TFs to predict histone marks at distal regulatory elements as well as TSSs, we sought to perform an analysis on a defined set of enhancers. Determining a suitable dataset of enhancers is nontrivial as enhancers are frequently defined in terms of the

presence of histone marks (23). We used a dataset of ~40,000 loci defined by the FANTOM5 consortium based on the presence of bidirectional capped transcripts in a set of 567 tissue and cell samples (24). TF-based LR classifiers were able to predict the histone modification status of these enhancers with an accuracy similar to that for TSSs and DNase sites (Table 1 and Tables S5 and S6). Surprisingly, we detected the promoter-associated mark H3K4me3 at a proportion of these elements in H1 cells, suggesting that some of them may function as novel promoters rather than enhancers. Our results therefore demonstrate that the binding locations of TFs accurately predict the modification status of histones genome-wide.

Discussion

The regulation of transcription is fundamental to the control of the genetic information encoded in cellular DNA. Recent attention has focused on a potential direct role for histone modifications in regulating gene expression based on their correlation with active or inactive promoters. By demonstrating that histone modifications can be predicted from the binding patterns of TFs, our results suggest that such correlations might be equally well explained as indirect effects of interactions between TFs and chromatin-modifying enzymes.

Descriptive analyses of ENCODE data have revealed significant overlaps between TF-binding loci and histone marks (14). However, to our knowledge, a quantification of the power of TFs as predictors of histone modifications was not attempted. The difference between descriptive and predictive analyses is important: for example, it would be impossible to assess whether cross-cell line predictions are possible within the limits of a descriptive analysis. Predictive analyses were performed as part of ENCODE and other studies but have focused on predicting TF binding (15) or gene expression levels (7, 8) from histone modification data. Although these predictive analyses are valuable, such models do not provide mechanistic explanations for the specificity of gene expression, in the way that TF-based regulation does (1). Intriguingly, one analysis of the ENCODE data suggested that the addition of histone modification data resulted in only minor improvements to the prediction of gene expression levels based on TF-binding data (25). Furthermore, a recent study demonstrated that depletion of the canonical activating mark H3K4me3 has only a modest effect on transcription,
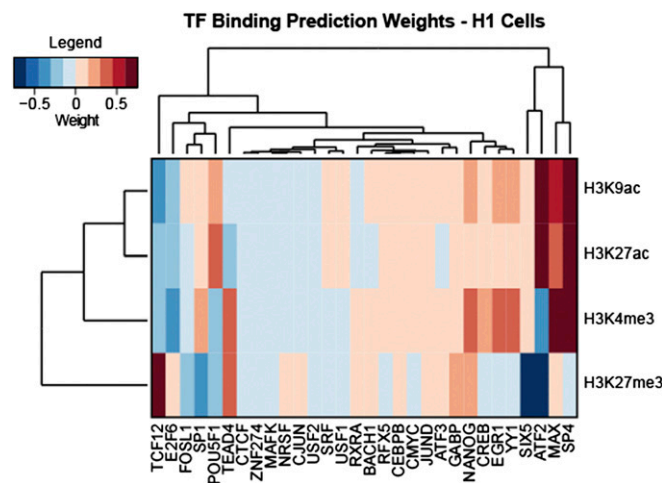

Fig. 4. TF weights uncover histone modifier–TF interactions. Heat map of LR weights from the prediction of histone modifications from TF-binding data in H1 cells. Each cell represents the weight assigned to a particular TF in predicting the occurrence of a particular histone mark. Both the rows and columns were subject to hierarchical clustering.

weakening the case for the causative role of this mark in transcription (26). It is therefore important to establish whether alternative explanatory variables may underlie the association between gene expression and histone modifications. Our results provide such an alternative explanation by demonstrating that histone modifications can themselves be predicted with high accuracy from TF binding genome-wide. We also note that DNA sequence has been shown to be predictive of histone modifications in previous studies (27, 28), but the comparative predictive power of TF-binding data were not tested.

An important benefit of our predictive analysis is the ability to quantify the relative importance of different TFs in predicting chromatin state. We find that highly predictive TFs are often cell-specific TFs that had been previously implicated in mechanisms of histone modification. Several examples of previously reported interactions between TFs and the epigenetic machinery are automatically inferred by our approach. For example, the H3K4me3 trimethylating mixed-lineage leukemia (MLL) complex has been shown to be recruited to the β-Globin locus through interaction with NFE2 (29) and is also reported to interact with E2F6 (30, 31) and the estrogen receptor (32, 33). Furthermore, examples of adaptor proteins such as SIN3A that bridge TFs to repressive complexes containing histone deacetylases (HDACs) have also been described (34). MLL interaction with E2F factors has also been suggested to be mediated through HCF-1 (35), and a computational study (36) has recently associated E2F with the SET1 methyltransferase subunit CFP1. A striking recent example of TF–histone interaction is the finding that insertion of the RE1-silencing transcription factor (REST) binding motif was capable of inducing ectopically H3K27me3 in mouse ES cells (37). In our analysis, we find REST is positively predictive of H3K27me3 localization, albeit weakly. Instead, we find that binding of TCF12, a TF that has recently been reported to interact with the H3K27me3 methyltransferase EZH2 (19), is the most predictive of H3K27me3 deposition, potentially reflecting differences between human and mouse ES cells. It is remarkable that our model was able to recapitulate such knowledge directly from binding data.

Family- and population-level genetic analyses have described the association of DNA sequence variants with altered histone modification patterns (38–40), suggesting a causal role of sequence-specific factors in determining chromatin state. The genetic inheritance of DNA sequence variants has also been suggested to be the primary cause of allele-specific variations in levels of DNA methylation (41). These analyses implicate the alteration in sequence-specific TF binding by DNA sequence variants as the primary cause of allele-specific variation in epigenetic state in mammalian genomes. An analysis of a handful of mouse promoters in ES cells has also demonstrated that DNA methylation state is primarily determined by the presence of binding sites for sequence-specific TFs (42). Taken together, these reports provide further support for the simplest interpretation of our work, that interactions between TFs and the epigenetic machinery, whether direct or indirect, play central roles in determining epigenetic state in mammalian genomes.

In summary, our results demonstrate a remarkable level of prediction of epigenetic marks from TF-binding profiles. Although such analyses do not demonstrate a causative role of TFs in determining epigenetic state in the genome, they show that previously reported associations between chromatin state and expression may be indirect effects. Our results confirm the need for caution in the mechanistic interpretation of genome-wide analyses (13), and provide useful pointers toward the complex biochemical pathways regulating gene expression.

Materials and Methods

Datasets Used. Datasets were downloaded from the Encyclopedia of DNA Elements UCSC repository at https://genome.ucsc.edu/ENCODE/. A complete list of all of the identifiers of the datasets used, as well as scripts to recreate our analysis, are available upon request. TSSs were retrieved from ENSEMBL using the human reference genome hg19. A detailed description of the pipeline used is given in *SI Materials and Methods*.

Statistical Methods. Throughout the paper, we used logistic regression as the classifier of choice due to its simplicity and interpretability. Logistic regression assigns the label 1 to an output with probability given by the logistic function of the input **x** as follows:

$$p(t = 1 | \mathbf{x}, \mathbf{w}) = \frac{1}{1 + \exp[-\mathbf{w}^T \mathbf{x}]},$$

where **w** is a vector of weights of the same dimensionality of the input data. The weights were learnt by maximum likelihood on a training set; throughout the paper, we used random splits of the data into 70% training and 30% testing, and reported only results on test data.

1. Ptashne M, Gann A (2002) *Genes and Signals* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
2. Ptashne M (2014) The chemistry of regulation of genes and other things. *J Biol Chem* 289(9):5417–5435.
3. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126(4):663–676.
4. Barski A, et al. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* 129(4):823–837.
5. Zhang Y, Reinberg D (2001) Transcription regulation by histone methylation: Interplay between different covalent modifications of the core histone tails. *Genes Dev* 15(18):2343–2360.
6. Turner BM (2007) Defining an epigenetic code. *Nat Cell Biol* 9(1):2–6.
7. Karlić R, Chung H-R, Lasserre J, Vlahovicek K, Vingron M (2010) Histone modification levels are predictive for gene expression. *Proc Natl Acad Sci USA* 107(7):2926–2931.
8. Dong X, et al. (2012) Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol* 13(9):R53.
9. Rakyan VK, Down TA, Balding DJ, Beck S (2011) Epigenome-wide association studies for common human diseases. *Nat Rev Genet* 12(8):529–541.
10. Métivier R, et al. (2003) Estrogen receptor-alpha directs ordered, cyclical, and combinatorial recruitment of cofactors on a natural target promoter. *Cell* 115(6):751–763.
11. Schuettengruber B, Martinez AM, Iovino N, Cavalli G (2011) Trithorax group proteins: Switching genes on and keeping them active. *Nat Rev Mol Cell Biol* 12(12):799–814.
12. Ptashne M (2013) Epigenetics: Core misconcept. *Proc Natl Acad Sci USA* 110(18):7101–7103.
13. Henikoff S, Shilatifard A (2011) Histone modification: Cause or cog? *Trends Genet* 27(10):389–396.

14. Bernstein BE, et al.; ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414):57–74.
15. Arvey A, Agius P, Noble WS, Leslie C (2012) Sequence and chromatin determinants of cell-type-specific transcription factor binding. *Genome Res* 22(9):1723–1734.
16. Hawkins RD, et al. (2010) Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* 6(5):479–491.
17. Cawley S, et al. (2004) Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. *Cell* 116(4):499–509.
18. Thomson JP, et al. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* 464(7291):1082–1086.
19. Lee CC, et al. (2012) TCF12 protein functions as transcriptional repressor of E-cadherin, and its overexpression is correlated with metastasis of colorectal cancer. *J Biol Chem* 287(4):2798–2809.
20. Klein E, et al. (1976) Properties of the K562 cell line, derived from a patient with chronic myeloid leukemia. *Int J Cancer* 18(4):421–431.
21. Lu R (2008) Interferon regulatory factor 4 and 8 in B-cell development. *Trends Immunol* 29(10):487–492.
22. Bhattacharyya S, et al. (2011) NFATc1 affects mouse splenic B cell function by controlling the calcineurin—NFAT signaling network. *J Exp Med* 208(4):823–839.
23. Bulger M, Groudine M (2010) Enhancers: The abundance and function of regulatory sequences beyond promoters. *Dev Biol* 339(2):250–257.
24. Andersson R, et al.; FANTOM Consortium (2014) An atlas of active enhancers across human cell types and tissues. *Nature* 507(7493):455–461.
25. Cheng C, et al. (2012) Understanding transcriptional regulation by integrative analysis of transcription factor binding data. *Genome Res* 22(9):1658–1667.

BIOPHYSICS AND COMPUTATIONAL BIOLOGY

26. Clouaire T, et al. (2012) Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev* 26(15):1714–1728.
27. Yuan GC (2009) Targeted recruitment of histone modifications in humans predicted by genomic sequences. *J Comput Biol* 16(2):341–355.
28. van Heeringen SJ, et al. (2014) Principles of nucleation of H3K27 methylation during embryonic development. *Genome Res* 24(3):401–410.
29. Demers C, et al. (2007) Activator-mediated recruitment of the MLL2 methyltransferase complex to the beta-globin locus. *Mol Cell* 27(4):573–584.
30. Dou Y, et al. (2005) Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF. *Cell* 121(6):873–885.
31. Takeda S, et al. (2006) Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. *Genes Dev* 20(17):2397–2409.
32. Dreijerink KM, et al. (2006) Menin links estrogen receptor activation to histone H3K4 trimethylation. *Cancer Res* 66(9):4929–4935.
33. Mo R, Rao SM, Zhu YJ (2006) Identification of the MLL2 complex as a coactivator for estrogen receptor alpha. *J Biol Chem* 281(23):15714–15720.
34. Knoepfler PS, Eisenman RN (1999) Sin meets NuRD and other tails of repression. *Cell* 99(5):447–450.
35. Tyagi S, Chabes AL, Wysocka J, Herr W (2007) E2F activation of S phase promoters via association with HCF-1 and the MLL family of histone H3K4 methyltransferases. *Mol Cell* 27(1):107–119.
36. Schweikert G, Cseke B, Clouaire T, Bird A, Sanguinetti G (2013) MMDiff: Quantitative testing for shape changes in ChIP-Seq data sets. *BMC Genomics* 14:826.
37. Arnold P, et al. (2013) Modeling of epigenome dynamics identifies transcription factors that mediate Polycomb targeting. *Genome Res* 23(1):60–73.
38. Kasowski M, et al. (2013) Extensive variation in chromatin states across humans. *Science* 342(6159):750–752.
39. McVicker G, et al. (2013) Identification of genetic variants that affect histone modifications in human cells. *Science* 342(6159):747–749.
40. Kilpinen H, et al. (2013) Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* 342(6159):744–747.
41. Gertz J, et al. (2011) Analysis of DNA methylation in a three-generation family reveals widespread genetic influence on epigenetic regulation. *PLoS Genet* 7(8): e1002228.
42. Lienert F, et al. (2011) Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* 43(11):1091–1097.