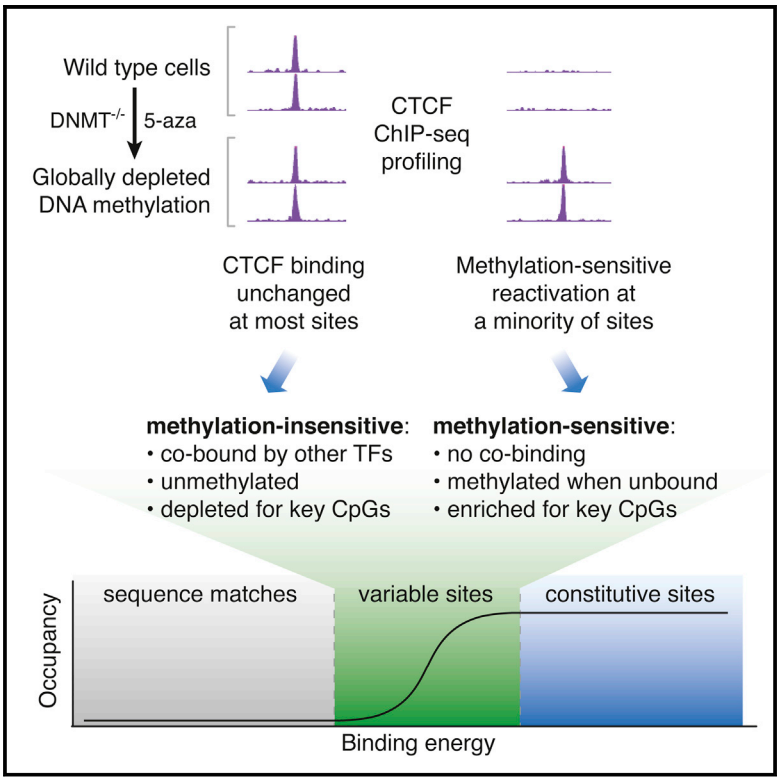
Article

Cell Reports

Role of DNA Methylation in Modulating Transcription Factor Occupancy

Graphical Abstract



Authors

Matthew T. Maurano, Hao Wang, Sam John, ..., Theresa Canfield, Kristen Lee, John A. Stamatoyannopoulos

Correspondence

matthew.maurano@nyumc.org (M.T.M.), jstam@uw.edu (J.A.S.)

In Brief

Alterations of DNA methylation in malignancy and development are frequently interpreted as affecting transcriptional activity. Maurano et al. find that, upon genomic abrogation of DNA methylation, binding of the canonically methylation-sensitive transcriptional regulator CTCF is largely unaffected. Their results suggest that a limited set of methylation-sensitive CTCF sites are variable across cell types and that key sequence and chromatin features predict methylation sensitivity of CTCF binding.

Highlights

- Most CTCF binding in vivo is unaltered by genomic abrogation of DNA methylation
- A limited set of CTCF sites from other cell lineages is methylation sensitive
- Methylation-sensitive CTCF sites are highly variable across cell types
- Key sequence and chromatin features predict methylation sensitivity of CTCF binding

Accession Numbers

GSE30263
GSE50610

Maurano et al., 2015, Cell Reports *12*, 1184–1195
August 18, 2015 ©2015 The Authors
http://dx.doi.org/10.1016/j.celrep.2015.07.024

OPEN ACCESS
CellPress

Cell Reports
Article

Role of DNA Methylation in Modulating Transcription Factor Occupancy

Matthew T. Maurano,[1,3,4,*] Hao Wang,[1,3] Sam John,[1,5] Anthony Shafer,[1] Theresa Canfield,[1] Kristen Lee,[1] and John A. Stamatoyannopoulos[1,2,*]

[1]Department of Genome Sciences, University of Washington, Seattle, WA 98195, USA
[2]Division of Oncology, Department of Medicine, University of Washington, Seattle, WA 98195, USA
[3]Co-first author
[4]Present address: Institute for Systems Genetics, New York University Langone Medical Center, New York, NY 10016, USA
[5]Present address: NIH, National Cancer Institute, Laboratory of Genome Integrity, Bethesda, MD 20892, USA
*Correspondence: matthew.maurano@nyumc.org (M.T.M.), jstam@uw.edu (J.A.S.)
http://dx.doi.org/10.1016/j.celrep.2015.07.024

SUMMARY

Although DNA methylation is commonly invoked as a mechanism for transcriptional repression, the extent to which it actively silences transcription factor (TF) occupancy sites in vivo is unknown. To study the role of DNA methylation in the active modulation of TF binding, we quantified the effect of DNA methylation depletion on the genomic occupancy patterns of CTCF, an abundant TF with known methylation sensitivity that is capable of autonomous binding to its target sites in chromatin. Here, we show that the vast majority (>98.5%) of the tens of thousands of unoccupied, methylated CTCF recognition sequences remain unbound upon abrogation of DNA methylation. The small fraction of sites that show methylation-dependent binding in vivo are in turn characterized by highly variable CTCF occupancy across cell types. Our results suggest that DNA methylation is not a primary groundskeeper of genomic TF landscapes, but rather a specialized mechanism for stabilizing intrinsically labile sites.

INTRODUCTION

DNA methylation is required for mammalian development and plays a central role in imprinting (Jones, 2012; Li et al., 1992). Cytosine methylation in the context of CpG dinucleotides has been widely invoked as a causal mechanism for transcriptional repression at promoter regions, and a correlation between DNA methylation and gene expression has long been recognized (Jones and Taylor, 1980). Recent findings indicate that dynamic demethylation during development is largely restricted to promoter-distal regulatory elements marked by DNase I-hypersensitive sites (DHSs) (Hon et al., 2013). However, the mechanism by which DNA methylation perturbs chromatin state and regulatory elements in a site-specific fashion remains obscure (Deaton and Bird, 2011). Various possibilities have been suggested including direct potentiation of repressive chromatin features (Collings et al., 2013; Ramirez-Carrozzi et al., 2009), recruitment of methylcytosine-specific repressive factors (Baubec et al., 2013; Lewis et al., 1992; Tate and Bird, 1993), and physical obstruction of the transcription factor (TF) DNA interface (Hu et al., 2013; Tate and Bird, 1993). Consistent with the latter hypothesis, DNA methylation is specifically depleted at occupied TF-binding sites in vivo (Groudine and Conkin, 1985; Lister et al., 2009; Neph et al., 2012b; Thurman et al., 2012). Although DNA methylation is widely assumed to inhibit TF occupancy in vivo, mechanistic studies suggest that methylation of TF recognition sites may follow TF evacuation (Brandeis et al., 1994; Feldmann et al., 2013; Lienert et al., 2011; Lin et al., 2000; Macleod et al., 1994; Matsuo et al., 1998; Stadler et al., 2011; Thurman et al., 2012). It thus remains unclear whether altered methylation patterns themselves invoke transcriptional repression or are instead downstream consequences of other regulatory factors.

Here we investigate the causal relationship between genome-wide DNA methylation and TF occupancy using the model TF CCCTC-binding factor (zinc finger protein) (CTCF), an abundant TF with known methylation sensitivity that is capable of autonomous binding to its target sites in chromatin (Phillips and Corces, 2009). CTCF occupancy is tightly anticorrelated with DNA methylation at its binding sites in vivo (Phillips and Corces, 2009; Wang et al., 2012), and its binding to DNA is abrogated by methylation in vitro (Bell and Felsenfeld, 2000; Filippova et al., 2001; Hark et al., 2000; Renda et al., 2007). Using a combination of genetic and chemical DNA methylation depletion experiments coupled to genome-wide occupancy analysis, we found that the CTCF-binding landscape remains largely unchanged in response to removal of DNA methylation. However, we observed a small minority of sites that reproducibly exhibited methylation-dependent occupancy. These sites were distinguished by high CpG content, the presence of specific CTCF recognition sequences that incorporated CpGs at critical positions in the binding interface, and an absence of preexisting regulatory activity within the assayed cell type. Despite the potential for CTCF occupancy at tens of thousands of potentially competent recognition elements genome-wide, reactivation was restricted specifically to elements that displayed highly variable CTCF occupancy when assayed across 40 diverse cell types. Our results suggest that DNA methylation does not play

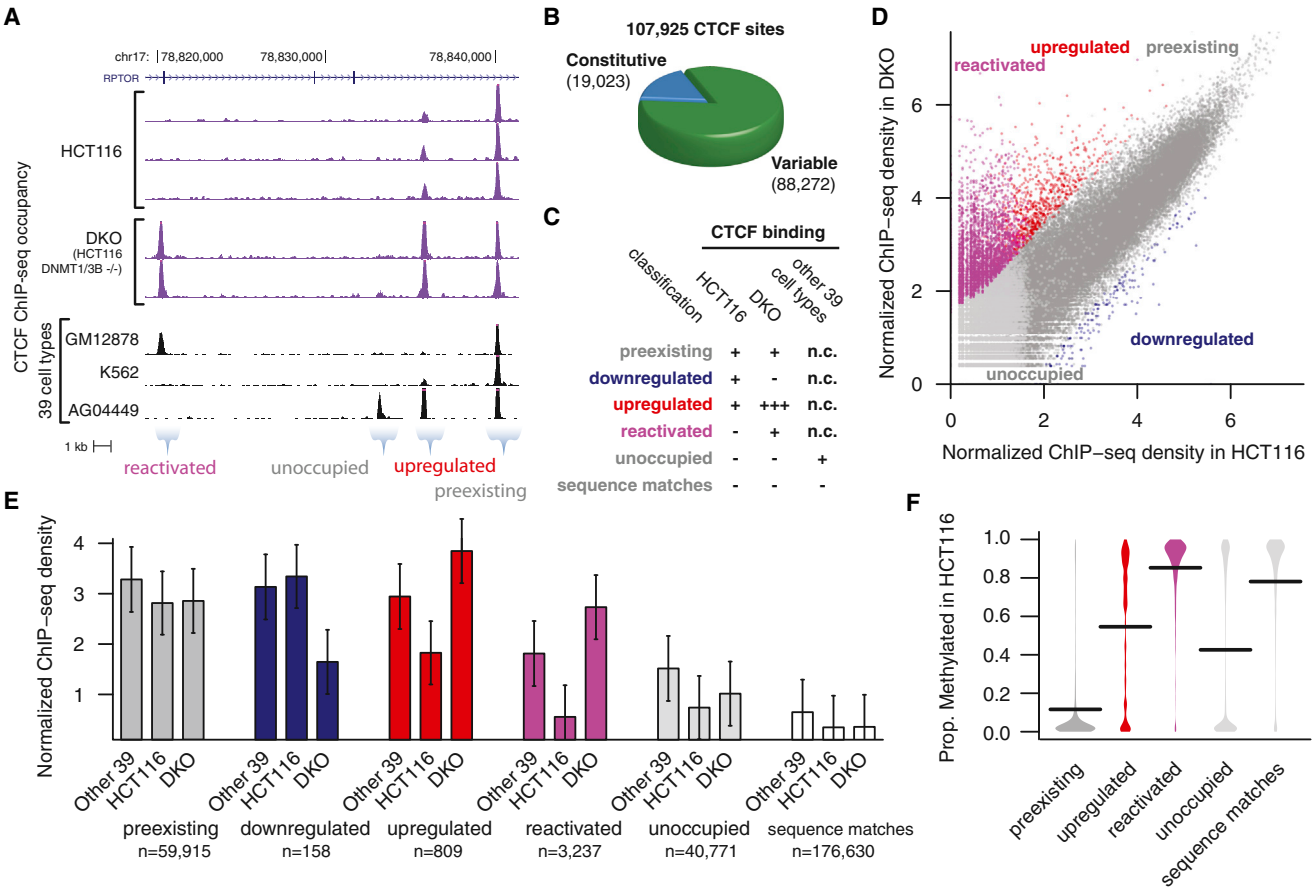


Figure 1. Profiling of TF Binding in Stably Demethylated Cells

(A) Normalized ChIP-seq profiles of CTCF occupancy in colorectal cancer cells depleted for methylation (HCT116, wild-type; DKO, double knockout of DNMT1 and DNMT3B) shows reactivated sites at RPTOR locus present in DKO, but not HCT116.

(B) Identification of CTCF-binding sites across 40 cell types. To conservatively identify variable CTCF sites, peaks were required to overlap a 0.5% FDR hotspot in at least one cell type, but presence of a 1% FDR hotspot was sufficient to establish activity in subsequent cell types. The majority of CTCF-binding sites varies across cell types.

(C–F) CTCF sites in HCT116 and DKO were classified based on ChIP-seq peak presence/absence (+/−) in HCT116; significant difference in occupancy upon demethylation in DKO (+ or +++, increased occupancy at 1% FDR using DESeq); and peak presence/absence (n.c., not considered) in other 39 cell types. (C) Summary of classification scheme is shown. (D) Global view shows occupancy in HCT116 and DKO cells. A minority of sites exhibited increased occupancy in DKO (red and magenta sites above diagonal). (E) Quantitative comparison shows mean occupancy at each class of sites in HCT116, DKO, or other cell types (measured as 90th percentile occupancy across 39 other cell types per site). Error bars represent SD. (F) Average methylation across 9,208 CTCF sites in HCT116 is shown. Reactivated sites are characterized by high preexisting methylation and unoccupied sites exhibit both unmethylated and unmethylated populations. Downregulated sites (n = 5 with methylation data) are not shown. Horizontal bars represent mean.

a significant primary role in repressing CTCF occupancy, but rather serves a more specialized function targeted at labile occupancy sites.

RESULTS

Most CTCF Binding Is Unaltered by Genomic Abrogation of DNA Methylation

To comprehensively assess the degree of CTCF cell-type-specific occupancy, we performed dense profiling of the normal CTCF occupancy landscape across 40 cell types in replicate (Tables S1 and S2). We identified an average of 61,944 CTCF sites in each cell type (range, 38,703–75,854) and a total of 107,295 CTCF sites across all cell types, which increased the previously documented CTCF-binding landscape (Wang et al., 2012) by nearly 40% (Table S3). These data confirm that the CTCF-binding landscape comprises a core minority of constitutive sites (n = 19,023), but that the majority of its binding is cell type specific (Figure 1; Figure S1A).

Partial disruption of the DNA methyltransferases DNMT1 and DNMT3B in colorectal carcinoma cells (HCT116) through gene targeting stably reduces global DNA methylation by 83%–95% (Akalin et al., 2012; Rhee et al., 2002) and results in genome-wide changes in gene expression and chromatin structure (Lay et al., 2015; Pandiyan et al., 2013; Reddington et al., 2013). However, the extent to which the observed alterations in chromatin structure and gene expression are directly caused by demethylation of specific sites versus secondary effects

is undetermined. To determine the degree to which CTCF occupancy was actively inhibited by DNA methylation, we performed chromatin immunoprecipitation sequencing (ChIP-seq) on HCT116 cells harboring homozygous knockouts of the major DNA methyltransferases DNMT1 and DNMT3B. Figure 1A shows the results of multiple replicates of CTCF ChIP-seq in double knockout (DKO) HCT116 cells in comparison with wild-type HCT116 cells. To gauge precisely the alteration to CTCF binding in the context of reduced genomic methylation, we partitioned CTCF sites along two axes as follows: (1) significant differences in occupancy (false discovery rate [FDR] 1%) between wild-type HCT116 and DKO cells, and (2) whether or not each site was occupied in wild-type HCT116 cells (Figure 1C; Experimental Procedures). CTCF-binding sites from the 39 other cell types that were not occupied either in HCT116 or DKO cells (unoccupied) are shown for comparison. This analysis revealed that the majority of sites in DKO cells were preexisting in wild-type HCT116 cells and showed unchanged CTCF occupancy (by sequencing read density) in DKO cells (Figure 1D). This suggested that DNMT knockout had at best a minor effect on global CTCF occupancy patterns.

To confirm a global lack of activation of CTCF occupancy upon DNMT ablation, we quantified CTCF occupancy at 176,630 stringent matches genome-wide for the CTCF-binding motif (Kim et al., 2007) that were completely null for CTCF occupancy (i.e., did not exhibit CTCF occupancy in vivo in any cell type with intact DNMTs; Figure 1E; Figures S1B–S1D; Experimental Procedures). These elements showed complete lack of occupancy, similar to that of a set of random sequences without the CTCF recognition sequence. We thus concluded that the CTCF-binding landscape is not significantly influenced by a drastic reduction in genomic methylation.

Reactivation at a Limited Set of Methylation-Sensitive Sites

Although the vast majority of preexisting CTCF occupancy sites, as well as the vast majority of unoccupied CTCF recognition sites, were unaffected by the reduction of DNA methylation, we noticed that a small compartment of CTCF sites (n = 4,204) showed significant (FDR 1%) methylation sensitivity. These elements comprised 158 CTCF sites with reduced CTCF occupancy (downregulated sites); 809 sites with increased CTCF occupancy (upregulated sites); and 3,237 sites found in DKO cells, but not wild-type HCT116 cells (reactivated sites). Reactivated elements were strongly occupied, showing similar occupancy in DKO cells to preexisting sites in either cell type (Figure 1E), suggesting that occupancy at a minority of CTCF sites is uniquely susceptible to methylation.

To test whether binding at reactivated elements relied directly on DNA methyltransferase function, we profiled CTCF occupancy in single knockouts of DNMT1 and DNMT3B. In contrast to DKO cells, each single knockout line is known to exhibit little difference in global DNA methylation (Rhee et al., 2002). Congruent with this, perturbation of single DNMT knockouts did not show significant changes in CTCF occupancy (Figures S1E and S1F). We thus conclude that novel CTCF occupancy at a limited subset of CTCF recognition elements is a direct consequence of the abrogation of methylation.

To link reactivation at these elements to relief of methylation-dependent repression, we examined methylation levels in HCT116 cells (Varley et al., 2013) at a subset of CTCF sites. Occupied sites demonstrated low methylation and unbound sites showed high methylation (Figure 1F; Table S4), consistent with older observations (Lister et al., 2009; Thurman et al., 2012; Wang et al., 2012). Although reactivation occurred almost exclusively (96%) at methylated sites, only 50% of methylated sites were reactivated. Similarly, 90% of unbound CTCF elements were methylated but exhibited virtually no reactivation. Given the low global level of remaining methylation in DKO cells (Akalin et al., 2012; Rhee et al., 2002), it is unlikely the lack of reactivation could have resulted from localized persistence of methylation at unoccupied sites. Nevertheless, we verified using published Methyl-Cap profiling of HCT116 and DKO cells (Simmer et al., 2012) that both reactivated and unoccupied sites showed essentially complete depletion of methylation in DKO cells (Figure S1G). Thus, while methylation is not the sole barrier to in vivo occupancy at the majority of sites genome-wide, loss of DNA methylation clearly potentiates CTCF binding at a subset of recognition sites.

Reactivation Occurs at In Vivo-Verified Sites

We queried whether reactivated elements represented activation of CTCF sites with documented capacity for occupancy in some cell type versus de novo activation of novel occupancy sites. Strikingly, 91% of sites found in DKO cells, but not HCT116 cells, also were detected in at least one other cell type, suggesting reactivation occurs at a subset of methylation-sensitive sites with an intrinsic capacity for in vivo CTCF occupancy (Figure 2A). Notably, 40,771 potentially reactivatable sites (i.e., found in other cell types, but not HCT116) were not affected, suggesting that methylation-sensitive sites are distinguished from inert CTCF recognition sequence matches by specific characteristics.

Preferential Reactivation at CTCF Sites Silenced in Transformed Cells

We next computed the frequency of CTCF occupancy at various classes of CTCF recognition sites. We observed that binding at both upregulated and reactivated sites appeared to be highly variable across cell types (Figure 2B). Methylation-sensitive occupancy was observed frequently at two classes of sites as follows: (1) those found in almost all cell types (excluding HCT116), and (2) sites found exclusively in malignancy-derived or immortalized cell lines (Figure 2C). We observed CpG island-associated hypermethylation in immortal, but not normal, lines (Figure 2D), overlapping a subset of methylation-associated cell-type-selective CTCF sites (Wang et al., 2012). This finding is consistent with the regulation of several tumor suppressors and oncogenes by methylation-dependent CTCF sites (Butcher et al., 2004; Dávalos-Salas et al., 2011; Soto-Reyes and Recillas-Targa, 2010; Witcher and Emerson, 2009). Overall, the vast majority (93%) of reactivated CTCF sites had at least one CpG within the 44-bp region of protein-DNA interaction (compared with only 54% of unoccupied sites), with reactivated sites more frequently harboring CpGs within the recognition sequence (Figure 2E). Likewise, whereas 29% of the reactivated sites were in CpG islands, this was the case for only 10% of unoccupied

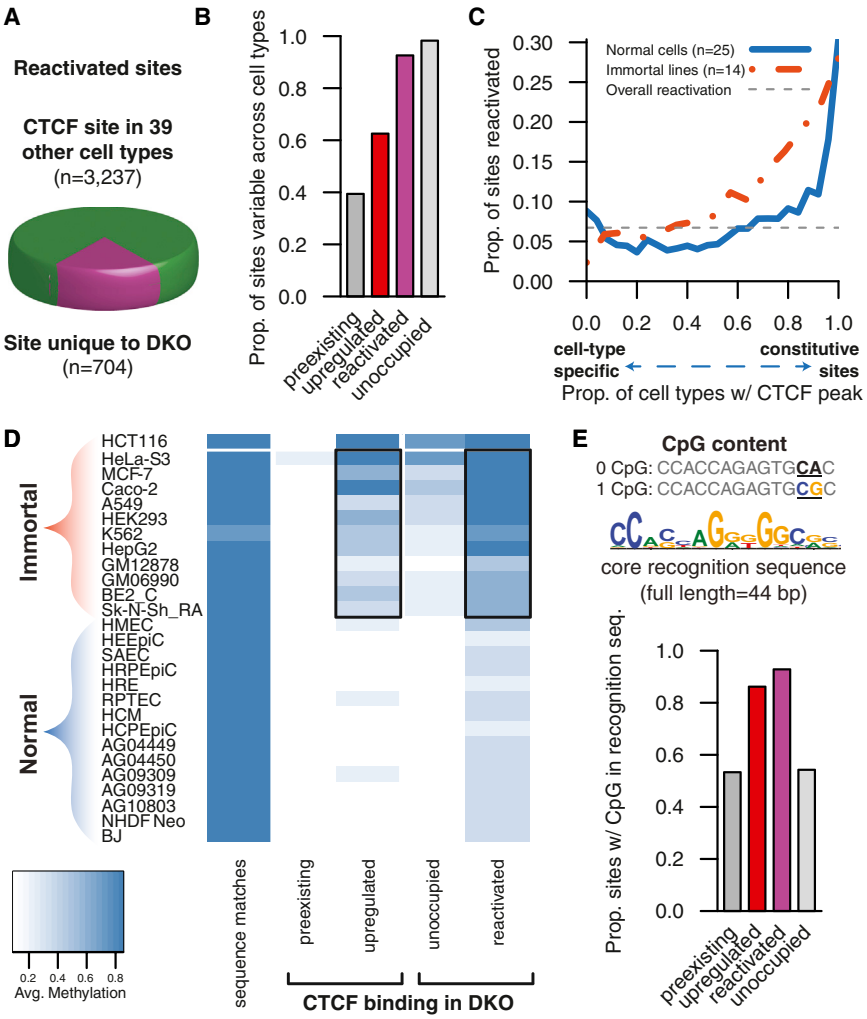


Figure 2. Reactivation of Predetermined Lineage-Specific Sites

(A) Overlap of reactivated sites with CTCF-binding sites known from other cell types. The minority of sites unique to DKO represents novel binding sites not known from our catalog of 40 cell types.

(B) Overlap of methylation sensitivity with sites variable across cell types (y axis) is shown.

(C) Reactivation frequency (y axis, computed as the number of reactivated sites divided by the number of reactivated and reactivated sites) at sites found in all normal cell types but silenced in HCT116 (solid line, right), as well as immortal-only sites not found in normal lines (solid line, left), is shown.

(D) Methylation at CTCF sites across 29 cell types (Varley et al., 2013). Note increased methylation in immortalized cell lines in reactivated and upregulated sites (boxes).

(E) Proportion of sites with a CpG in their recognition sequence. Note the increased frequency in upregulated and reactivated sites.

sites. We thus conclude that many of the methylation alterations observed at high-CpG regions in the context of malignant transformation (Varley et al., 2013; Wang et al., 2012) are accompanied by disrupted CTCF occupancy at a highly specific set of labile sites.

Cooperative Binding at CTCF Sites Impedes Methylation Sensitivity

Despite the very large potential genomic occupancy space defined by stringent matches to the CTCF recognition sequence, reactivation is not observed outside of experimentally verified binding sites (by ChIP-seq). We thus asked what features distinguish reactivated CTCF sites from potential recognition sites that were not responsive to demethylation. Chromatin context and DNA accessibility are obvious candidate factors in discriminating such sites (John et al., 2011). To probe the role of DNA accessibility in potentiating CTCF occupancy upon demethylation, we profiled HCT116 and DKO cells using DNase sequencing (DNase-seq). We also quantified chromatin modification state using ChIP-seq for trimethylation of histone 3 lysine 4 (H3K4me3) and acetylation of histone 3 lysine 27 (H3K27ac) (Figure 3A; Table S5). At the vast majority of sites, we found CTCF occupancy to be tightly associated with the DHSs; however, only 17% of reactivated sites overlapped a DHS in HCT116 cells (Figure 3B; Table S6). We thus concluded that de novo recruitment of CTCF does not require targeting to accessible chromatin sites prior to demethylation.

Indeed, reactivated CTCF sites overlapped DHSs in HCT116 cells significantly less frequently than unoccupied sites (Figure 3B; Figure S2A). Furthermore, the association of reactivation with CpG content was almost entirely specific to sites without preexisting accessible chromatin (Figure 3C). The presence of H3K4me3 was similarly antagonistic to reactivation (Figure S2B), and the presence of both DNase I hypersensitivity and H3K4me3 was associated with an absence of methylation specifically at unoccupied sites, despite a lack of CTCF occupancy (Figure 1F; Figure S2C). Finally, unoccupied sites were enriched for the occupancy of 17 TFs (Figure S2D) studied by the ENCODE project (ENCODE Project Consortium, 2012). Thus, pre-established cell-type-specific binding by other TFs actively impedes CTCF recruitment, regardless of DNA methylation.

Chromatin Dynamics at Reactivated CTCF Sites

Given that the removal of DNA methylation might indirectly enable CTCF occupancy subsequent to a broader relaxation of chromatin state (Ooi et al., 2007; Thomson et al., 2010), we then asked whether relaxed chromatin structure required the reactivation of CTCF binding. DNase I accessibility increased in tandem with CTCF occupancy at reactivated and upregulated sites and decreased at downregulated sites, but only a subset of CTCF reactivation was accompanied by concomitant reactivation of

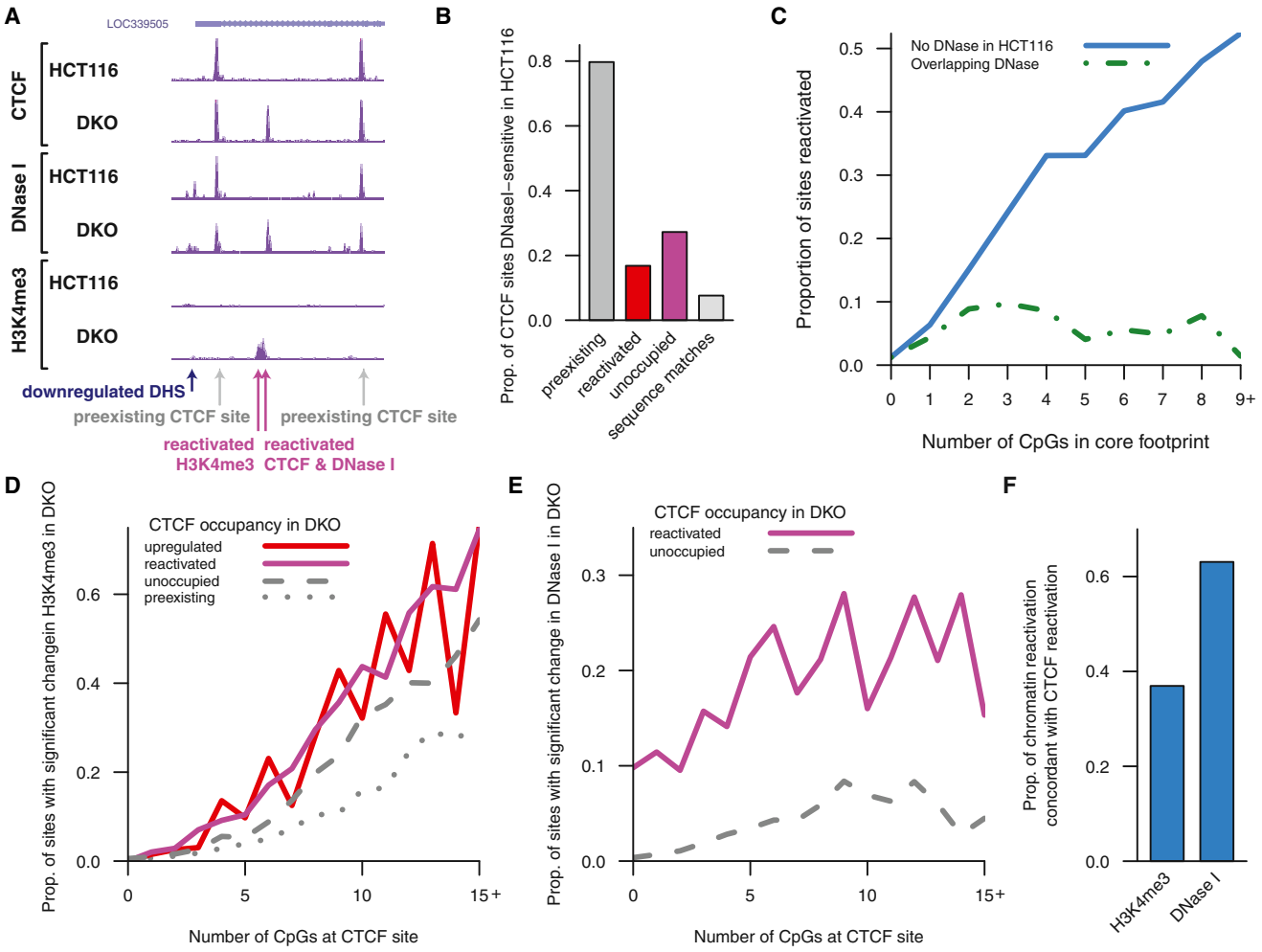


Figure 3. Chromatin Dynamics at Reactivated CTCF Sites

(A) Profiling of CTCF, DNase I, and H3K4me3 in HCT116 and DKO cells is shown.

(B) Proportion of CTCF sites overlapping DNase I-hypersensitive sites (DHSs). Although preexisting CTCF sites are stereotypically hypersensitive to DNase I cleavage, reactivation upon demethylation occurs at sites inaccessible in HCT116.

(C) Reactivation frequency parallels local CpG content (solid blue line), while the presence of a preexisting DHS is associated with a near-total lack of reactivation regardless of CpG content (dashed green line).

(D and E) Increased CpG content (x axis) is strongly associated with reactivation of H3K4me3 (D), but not DNase I (E). Shown are sites without H3K4me3 peak (D) or DHS and CTCF peak (E) in HCT116. Note strong reactivation of H3K4me3 at unoccupied CTCF sites in contrast to a lack of increase in DNase I at these sites.

(F) While H3K4me3 increases irrespective of CTCF reactivation, DNase I at CTCF sites coincides frequently with reactivation of CTCF occupancy.

H3K4me3 (Figures S2E–S2I). In fact, H3K4me3 reactivation occurred regardless of CTCF recruitment and depended strongly upon CpG content (Figure 3D), consistent with the recognition of unmethylated CpG islands by Cfp1 (part of the Set1 methyltransferase complex) (Thomson et al., 2010). In contrast, the relationship between DNase I reactivation and CpG content was conditional on CTCF reactivation (Figures 3E and 3F). Thus, we conclude that, while H3K4me3 is deposited in unmethylated regions of high CpG content regardless of CTCF occupancy, DNase I accessibility specifically marks regulatory factor occupancy.

We investigated whether other TFs might exhibit global sensitivity to DNA methylation by examining genome-wide profiles of H3K4me3, H3K27ac, and DNase I in HCT116 and DKO cells. The vast majority of H3K4me3 (84%) and H3K27ac (88%) peaks in DKO cells overlapped with DHSs. DKO cells manifested an expanded accessible chromatin landscape, comprising a 61% increase in the number of DHSs (Figures S2J–S2L). This represents a surprisingly small increase given the expanse of potentially reactivated DHSs known from other cell types. Detailed statistical analysis revealed 12,279 DHSs with differential accessibility, 96% of which were not present in HCT116 cells. Fully 70% of these DHSs were present in a survey of 124 cell types (Maurano et al., 2012a), confirming preferential reactivation of silence sites from other lineages. This set of reactivated DHSs included a majority of the reactivated CTCF sites (between 48% and 77%, depending on FDR cutoff); but, despite the large number of genomic CTCF occupancy sites and its pivotal role at

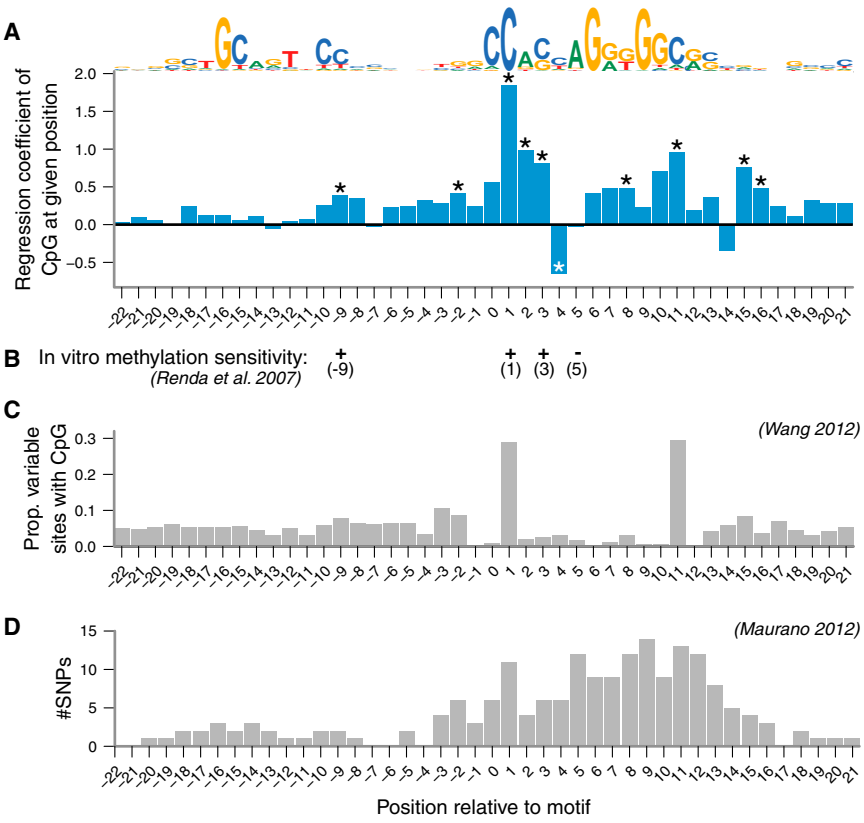


Figure 4. CpGs at Key Positions in the Protein-DNA Binding Interface Drive Reactivation

(A) Linear regression coefficients (y axis) estimating the contribution of the presence of a CpG dinucleotide in the CTCF recognition sequence (consensus binding motif shown at top) to reactivation upon abrogation of methylation. Window is defined as the 44-bp extent of protein-DNA contact demarcated by DNase I footprint. *Position is significant in regression ($p < 0.05$, Bonferroni corrected).

(B) Methylation-sensitive positions identified in vitro through electrophoretic mobility shift assay (EMSA) are shown (Renda et al., 2007).

(C) Association of CpG presence with methylation-associated cell-type variability across 19 cell types is shown (Wang et al., 2012).

(D) SNPs associated with significant alteration in occupancy across individuals are shown (from Figure 3B; Maurano et al., 2012b).

imprinted loci, fully 90% of reactivated DHSs were not CTCF sites (Figure S2M).

Methylation Modulates the CTCF-DNA Binding Interface at Labile Sites

To examine whether the association of CpG content with reactivation (Figure 3C) localized to specific positions in the protein-DNA interface (Renda et al., 2007; Wang et al., 2012; Wiench et al., 2011), we used a regression model to quantify the contributions of CpG dinucleotides at each position in the recognition sequence. We found that critical CpGs were concentrated in the core binding region, consistent with in vitro sensitivity to methylation (Renda et al., 2007), the positions of CpGs at sites of methylation-associated cell-type-selective binding (Wang et al., 2012), and sensitivity to single-nucleotide variants (Maurano et al., 2012b; Figure 4). Reactivation was associated with the strength of the match to the CTCF consensus sequence, also consistent with a direct effect on CTCF binding (Figure S3A). By classifying CTCF-binding sites by sequence similarity into three known binding modes (Bowers et al., 2009; Filippova et al., 1996; Nakahashi et al., 2013; Ohlsson et al., 2001; Rhee and Pugh, 2011), including engagement of zinc fingers 3–7 (core sites) and additional specific interactions by zinc fingers 8–10 (upstream and extended spacing sites), we found that core sites were reactivated 42% more frequently than upstream sites (Figure S3B), supporting speculation that different modes of engagement of its 11 zinc fingers may confer functional selectivity (Ohlsson et al., 2010).

Cell-Type-Specific Reactivation Highlights Methylation-Independent Repression

We next asked whether reactivation of methylation-sensitive CTCF sites could be reproduced in an independent system combining both a different cellular context and a different methodology for depleting DNA methylation. Chemical inhibition of DNA methyltransferases by 5-aza-2′-deoxycytidine (5-aza-CdR) transiently reduces global methylation levels and has been reported to reactivate TF binding and increase gene expression (Hagemann et al., 2011; Komashko and Farnham, 2010; Qiu et al., 2010; Rodriguez et al., 2010; Witcher and Emerson, 2009). We thus profiled the effect of 5-aza-CdR treatment on CTCF binding in K562 cells, which have an erythroid phenotype that differs markedly from the colonic epithelial phenotype of HCT116 cells (Thurman et al., 2012). In K562 cells, 5-aza-CdR treatment resulted in weaker reactivation than in DKO cells, including 767 reactivated, 191 upregulated, and 4 downregulated sites (Figures 5A and 5B). We used a targeted sodium bisulfite approach to confirm a more limited reduction of methylation than in DKO cells (Experimental Procedures; Table S7; Figure S4), consistent with previous reports (Hagemann et al., 2011; Pandiyan et al., 2013; Rhee et al., 2002). Reactivation was again largely specific to sites of preexisting methylation: 79% of reactivated sites were methylated in K562 cells. However, we found a complex relationship between methylation and occupancy (Figure S4D) that suggested widespread secondary effects, consistent with previous observations that 20% of variable CTCF sites are unmethylated in all cell types (Wang et al., 2012) and that most 5-aza-CdR-induced alterations in gene expression or chromatin structure occur at previously unmethylated sites (Komashko and Farnham, 2010).

Despite the lesser extent of reactivation with 5-aza-CdR in K562 cells, we found that fully 69% of sites reactivated with 5-aza-CdR overlapped sites that were concordantly reactivated

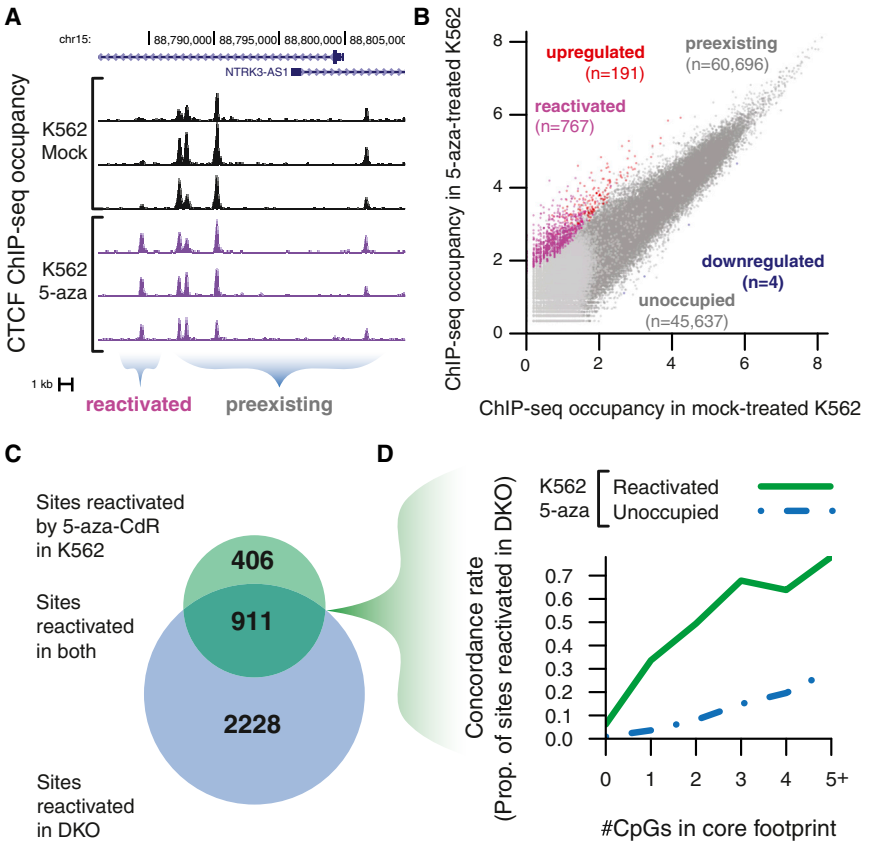


Figure 5. Cell-Type-Specific Reactivation Highlights Methylation-Independent Repression

(A) CTCF ChIP-seq at NTRK3 locus shows reactivated and preexisting sites in a mock- and 5-aza-CdR-treated K562 cells.

(B) Quantitative comparison of CTCF occupancy in mock and treated cells shows limited but reproducible reactivation (FDR 1%). Sites were classified (labels) based on the presence of peak in mock-treated K562 and significant differential occupancy in treated cells.

(C) Overlap with reactivation in DKO at 46,726 potential sites not occupied in both K562 and HCT116; FDR 5% hotspots. The majority of 5-aza-CdR-reactivated sites are concordantly reactivated in DKO.

(D) Sites reactivated in K562 (solid line) are more likely to be reactivated in DKO (y axis) than unoccupied sites (dashed line), regardless of CpG content.

(E) Reduced preexisting methylation (y axis) in both K562 (left) and HCT116 (right) distinguishes sites reactivated specifically by genetic or chemical means, suggesting that non-concordant sites are attributable to methylation-independent silencing. p values by Wilcoxon signed-rank test. Error bars represent SD.

(F) Model of cell-type-specific repression of TF occupancy. Concordant reactivation implies methylation-dependent silencing across cell types, while discordant reactivation implies methylation-independent silencing in the unoccupied cell type.

in HCT116/DKO cells (Figure 5C). Sites reactivated in both cell types demonstrated CpG-dependent reactivation (Figure 5D) and high preexisting methylation in both K562 and HCT116 cells, further supporting the existence of a predetermined class of methylation-dependent sites. Sites reactivated only in a single cell type demonstrated significantly less methylation (Figure 5E), revealing the presence of methylation-independent silencing mechanisms in the unoccupied cell type (Figure 5F).

Sequence Recognition of Labile Methylation-Sensitive Sites

We examined the relative predictive power of genomic characteristics to recognize the minority of sites that were reactivated (Figures 6A and 6B). The single most predictive factor was the number of CpGs at critical positions in the recognition sequence. The number of CpGs in the flanking regions was almost as predictive, probably due to the regionally high CpG density of CTCF sites. The second most predictive factor was lack of chromatin accessibility in HCT116 cells.

As many of the features surveyed are not independent, we used a logistic regression model to assess their combined predictive power (Experimental Procedures). A model considering just chromatin accessibility and number of CpGs was more predictive (area under the curve [AUC] of 0.82) than either factor alone and almost as predictive as the full model (AUC 0.84), indicating that the remaining factors provide only minor additional discriminative power. Although we observed increased reactivation at exons, the increased G+C content of protein-coding sequence is presumably responsible for a markedly increased reactivation frequency relative to other genomic sequence (Figure S5A), which suggests an altered regime of methylation-regulatory factor interaction at these sites (Stergachis et al., 2013). At sites for which methylation data were available, the degree of methylation alone had strong predictive power; but, this predictive power was largely redundant when DNase I accessibility and number

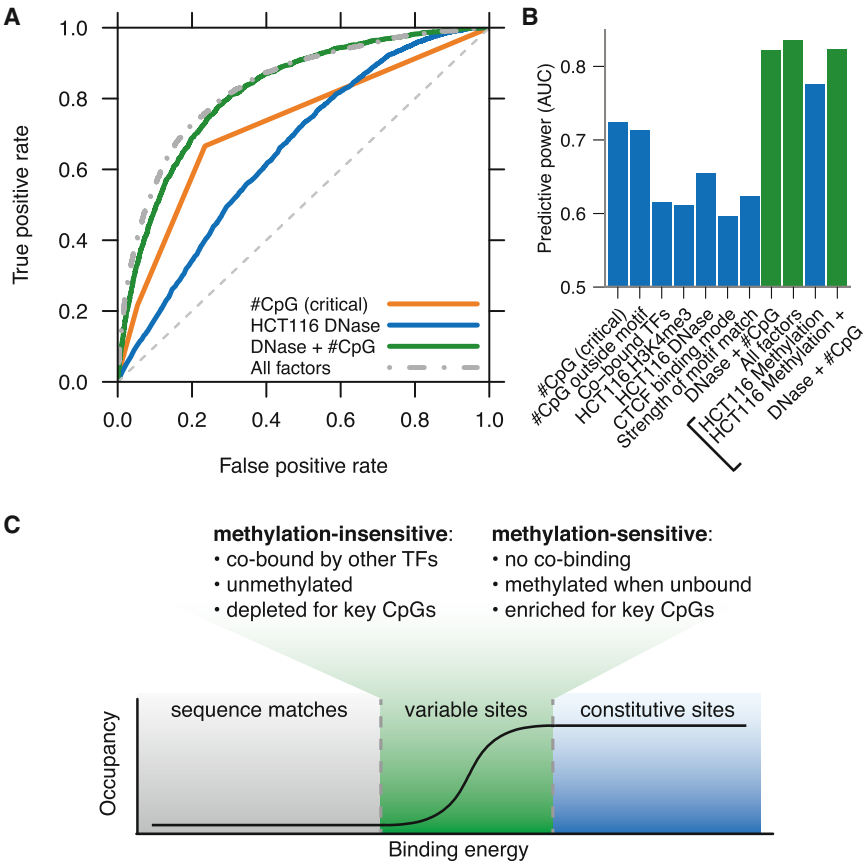


Figure 6. Recognition of Labile Methylation-Sensitive Sites

(A and B) Predictive power of genomic features to distinguish reactivated from unoccupied sites, illustrated by receiver-operator curves (ROCs). Area under the curve (AUC) summarizes overall predictive power. (A) Full ROCs highlight for models considering (1) just CpGs at specific positions, (2) DNase I in HCT116, (3) both together, or (4) all features together (Experimental Procedures). Dashed gray line indicates a random predictor and has an AUC of 0.5. A perfect predictor would be plotted as a right angle and have an AUC of 1. (B) Detailed predictive power (y axis) of selected genomic features is shown. Models include methylation (bracketed) measured only at sites with RRBS data. Green bars indicate linear models combining multiple factors. Other TF occupancy refers to the sum of HCT116 ENCODE TF ChIP-seq track densities.

(C) Model shows methylation sensitivity at labile CTCF sites.

of CpGs additionally were considered, suggesting that the latter features primarily provide indication of methylation status (Figure S2C). Thus, just as sequence features are associated with CTCF-binding sites conserved across six mammals (Schmidt et al., 2012) and SNPs affecting CTCF occupancy (Ding et al., 2014; Maurano et al., 2012b), a simple sequence model considering CpG content can reliably distinguish the minority of reactivated sites (Figure 6C).

DISCUSSION

The potential for DNA methylation to shape TF-binding landscapes, with consequent effects on gene expression patterns, has been invoked frequently but not studied systematically. Although methylation-sensitive CTCF binding to the Igf2/H19-imprinted locus represents the paradigmatic model of the relationship of DNA methylation to gene regulation (Bell and Felsenfeld, 2000; Hark et al., 2000), the results reported herein reveal a limited and highly specific response of CTCF occupancy to DNA methylation depletion. The observed effects are consistent across cellular contexts and under both transient and stable inhibition of DNA methyltransferases, suggesting their generality. Given the global lack of correlation between methylation changes and altered binding, these results place a limit on the extent of CTCF-mediated coupling between DNA methylation and genome organization, cautioning against a facile interpretation of alterations of DNA methylation in oncogenesis.

Our results contrast with the well-established relationship between DNA methylation and reduced CTCF binding (Bell and Felsenfeld, 2000; Wang et al., 2012), and they qualify that this relationship is only causal at a small subset of sites in vivo. Yet although DNA methylation does not play a significant primary role in repressing TF occupancy, the presence of methylation sensitivity at labile sites suggests that it may nevertheless modulate their epigenetic stability. This interpretation is consistent with a model whereby DNA methylation is passively deposited in the abrogation of TF binding and acts as a cooperative switch to prevent the return of binding after a reprogramming event (Stadler et al., 2011; Wang et al., 2012).

Strikingly, reactivation is strictly limited to in vivo CTCF sites from other cell types, highlighting that these silent sites are globally distinguished from other matches to the CTCF recognition sequence. Although CTCF preferentially recognizes a DNA sequence with higher information content than most other human TFs—fully 24% of CTCF recognition sequences are within 500 bp of an in vivo CTCF-binding site—our work confirms that its recognition sequence alone is insufficient for recruitment even at unmethylated sites. But it remains unclear how bona fide CTCF sites are distinguished from inert sequence matches: these silent sites are not marked by DNase I accessibility and, in fact, reactivation is less frequent in the presence of cofactors. However, approximately half of reactivated and unoccupied sites are accessible to DNase I in human embryonic stem cells, compared to only 11% of sequence matches, suggesting that the marking of bona fide CTCF sites occurs early in development. It is possible that topological considerations (Rao et al., 2014) or cooperative binding with as yet undetermined factors play a role in the determination of its in vivo binding sites.

Despite the secondary role of DNA methylation at most CTCF sites, the small set of reproducibly reactivated sites is distinguished by solitary CTCF binding and clear sequence characteristics including the presence of CpGs at key positions in the binding interface. The strongest factor distinguishing reactivated sites is the presence of CpG dinucleotides at key positions of the protein-DNA recognition interface, consistent with previous results showing that the methylation is specifically depleted at sites of TF occupancy marked by DNase I footprints (Groudine and Conkin, 1985; Neph et al., 2012b; Stadler et al., 2011). This selective effect of sequence features is reminiscent of single nucleotide genetic variants within the CTCF recognition sequence (Maurano et al., 2012b), and in turn suggests that TF-DNA recognition interfaces are intrinsically buffered against both genetic and epigenetic perturbations. The conferral of methylation-sensitive binding by the presence of CpG dinucleotides at key positions in the recognition interface at elements showing labile occupancy across cell types provides a potentially unifying link among sequence features, DNA methylation, and epigenetic regulatory state (Gaidatzis et al., 2014; Neph et al., 2012b).

EXPERIMENTAL PROCEDURES

Cell Culture and ChIP-Seq/DNaseI-Seq

HCT116 and DKO cells have been described previously (Rhee et al., 2002). Cells were cultured in appropriate medium and maintained in a humidified incubator at 37°C in the presence of 5% CO_2. (Table S1). K562 cells were treated with 5-aza-CdR (decitabine; Sigma, A3656) dissolved in DMSO to 10 mM. The drug was administered at 1 μM daily for 3 days. Control K562 cells were mock-treated with DMSO. Conditions for other cell types were as described in Table S1.

Chromatin immunoprecipitations were performed as described previously for CTCF (Wang et al., 2012) and H3K4me3 and H3K27ac (Thurman et al., 2012). Briefly, cells were crosslinked for 10 min in 1% formaldehyde and quenched in 125 mM glycine. Chromatin was sheared by Bioruptor (Diagenode) and incubated with antibody conjugated to Dynabeads (M-280, Invitrogen). After reversing crosslinks, immunoprecipitated DNA was purified by phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation. DNase was performed as described previously (John et al., 2011; Thurman et al., 2012), whereby small (<500-bp) fragments are isolated from lysed nuclei following DNase treatment. Libraries generated from immunoprecipitated or DNase-treated DNA were sequenced on an Illumina Genome Analyzer IIx or HiSeq 2000 by the High-Throughput Genomics Center (University of Washington) according to a standard protocol. Most experiments were performed on two independent biological replicates (Tables S2 and S5).

Data Processing

ChIP-seq and DNase data were mapped to the human genome (GRCh37/hg19) using bowtie (Langmead et al., 2009) with the options "bowtie -n 2 -m 1 -e 70–best" allowing up to two mismatches. Reads mapping to multiple locations were then excluded, and reads with identical 5′ ends and strand were presumed to be PCR duplicates and were excluded using Picard MarkDuplicates. Smoothed density tracks were generated using bedmap (https://bedops.readthedocs.org/) to count the number of reads overlapping a sliding 150-bp window, with a resolution of 20 bp (Neph et al., 2012a). Density tracks for display were normalized for sequencing depth by a global linear scaling to 10 million reads.

CTCF data from 39 cell types were processed as described previously (Wang et al., 2012). Briefly, a master peak list (Table S3) was established from ENCODE project 2% irreproducible discovery rate (IDR)-processed SPP peak calls (Kharchenko et al., 2008; Li et al., 2011; ftp://ftp.ebi.ac.uk/pub/databases/ensembl/encode/supplementary/integration_data_jan2011/byDataType/peaks/june2012/spp/optimal/), with locations adjusted to center on matches to the nearest CTCF motif ($p < 10^{-4}$, FIMO) if the motif was within 50 bp. An IDR-thresholded peak was additionally required to overlap a 0.5% FDR hotspot in both replicates. To conservatively identify variable sites, peak presence in a given cell type was established by the presence of a 1% FDR hotspot (John et al., 2011; also http://www.uwencode.org/proj/hotspot/) in any replicate. Peaks only found in DKO and/or 5-aza-CdR-treated K562 cells were not included in total peak counts. We identified possible further CTCF-binding sequences using the program FIMO ($p < 10^{-5}$), as well as random sequences without a CTCF-binding sequence ($p > 10^{-4}$). Both sets were required to be further than 500 bp from any known in vivo CTCF site, to consist of base pairs 90% of which are uniquely mappable by 36-bp reads, and not to be on the ENCODE blacklist (ENCODE Project Consortium, 2012). We identified DHSs by the presence of a hotspot (FDR 1%) in any HCT116 replicate or a hotspot (FDR 0.5%) in DKO cells; novel sites in DKO cells were included in the analysis if they overlapped no unthresholded hotspot in any HCT116 replicate. H3K4me3 and H3K27ac regions were called using hotspot.

At both CTCF sites and DHSs separately, we measured CTCF occupancy or DNase I accessibility by the number of reads overlapping the 150-bp region and H3K4me3 and H3K27ac density by reads overlapping the 2150-bp region. We then used the package DESeq (Anders and Huber, 2010) to identify significant differences. CTCF ChIP-seq peaks or DHSs discordant between the two previously published replicates and the new replicate for HCT116 (FDR 10%) were removed. Pairwise differences were called between samples for each ChIP-seq and DNase I dataset using estimateDispersions (method = pooled, sharingMode = maximum, and fitType = local) and nbinomTest. After observing a near-complete lack of reactivation at sequence matches and random sites (less than ten sites), we subsequently excluded these sites and recomputed pairwise differences. We applied the Benjamini-Hochberg procedure to control for multiple testing. To obtain normalized data at all sites, we used estimateDispersions as before but with fitType = parametric, and then we applied a variance-stabilizing transformation and averaged the occupancy of all replicates. These values were scaled to [0, 10] by subtracting the global minimum and dividing by the global maximum * 10.

We obtained the location of CTCF sites relative to genes using RefSeq and relative to CpG islands and repetitive regions in RepeatMasker from the University of California, Santa Cruz (UCSC) Genome Browser.

Classification of CTCF Motif Models

CTCF binds in a multivalent fashion, whereby three modes of binding are distinguished by the presence and position of an upstream motif (Nakahashi et al., 2013). At each site we chose the best matching (by FIMO p value) of three motif models.

Per-nt Regression Model

We used the lm() function in R to perform a logistic regression considering all preexisting, upregulated, unoccupied, and reactivated CTCF sites with a recognizable motif as follows:

$$\text{reactivation} \sim c + x_{-22} + \ldots + x_{21} + \text{upstream} + (\text{CpGs in flanking region}),$$

where upstream represents an indicator variable for the presence of either upstream motif. Multiple testing correction was performed using the Bonferroni method.

ROC and PR curves were computed using the R package ROCR. Logistic models for multiple factors were constructed using the lm() function in R.

Monitoring Methylation

Purified DNA from mock- and 5-aza-CdR-treated K562 cells was fragmented following the Agilent SureSelect Methyl-Seq protocol with slight modifications (Table S7). DNA (4 μg) from each of the samples was fragmented in a Covaris S2 under the following conditions: 10% duty cycle at intensity 5 for 60 s, with 200 cycles per burst for 6 cycles with mode set to sweeping. The 250-bp fragmented DNA was then end-repaired and adenylated, followed by ligation to adapters synthesized with 5′-methylcytosine in place of of cytosine. The

adaptor-ligated library was purified using AMPure XP beads (Beckman Coulter Genomics). Next, 500 ng of each library was hybridized to Agilent SureSelect Methyl-Seq biotinylated RNA baits (84 Mbp) for 24 hr at 65°C. The biotinylated probe/target hybrids were captured on Dynal MyOne Streptavidin T1 (Invitrogen), washed, eluted, and desalted following purification on a MinElute PCR column (QIAGEN), as described in the SureSelect protocol.

Bisulfite conversion of the purified captured library was performed using the EZ DNA Methylation Gold Kit (Zymo Research) as per the manufacturer's instructions. The bisulfite-converted captured library was amplified by PCR with a minimal amount of PCR cycles, then purified using AMPure XP beads and quantified by Qubit dsDNA assay (Invitrogen) following the SureSelect protocol.

Samples were diluted to a working concentration of 10 nm. Cluster generation was performed for each sample and loaded onto a single lane of an Illumina HiSeq flowcell. Single-end sequencing was performed for 36 cycles according to the manufacturer's instructions.

Capture bisulfite reads were mapped to the human genome (hg19) using Bismark (Krueger and Andrews, 2011) and bowtie2 beta 6 (Langmead et al., 2009) with the options "-n 1" and excluding duplicate reads and combining reads from the top and bottom strands. Methylation was monitored at CpGs with eight or more read coverage and averaged for all CpGs within the 150-bp window at each CTCF site. We obtained eight or more times coverage in both samples for 658,010 CpGs.

Reduced representation bisulfite sequencing (RRBS) data for 29 cell types are available from the UCSC Genome Browser (Varley et al., 2013), processed as described previously (Wang et al., 2012).

ACCESSION NUMBERS

The accession numbers for the ChIP-seq, DNase-seq, and bisulfite data reported in this paper are GEO: GSE30263 and GSE50610.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures and seven tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2015.07.024.

AUTHOR CONTRIBUTIONS

H.W., M.T.M., S.J., A.S., T.C., and K.L. performed research. M.T.M. analyzed data. M.T.M. and J.A.S. designed research and wrote the paper.

ACKNOWLEDGMENTS

We thank Bert Vogelstein for graciously providing HCT116 DNMT1/3B$^{-/-}$ and DKO cells. We thank Bob Thurman for informatics support. This work was supported by NIH fellowship F31MH094073 to M.T.M. and grants U54HG004592, U01ES017156, and S10OD017999 to J.A.S.

Received: November 3, 2014
Revised: June 14, 2015
Accepted: July 10, 2015
Published: August 6, 2015

REFERENCES

Akalin, A., Garrett-Bakelman, F.E., Kormaksson, M., Busuttil, J., Zhang, L., Khrebtukova, I., Milne, T.A., Huang, Y., Biswas, D., Hess, J.L., et al. (2012). Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet. *8*, e1002781.

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol. *11*, R106.

Baubec, T., Ivánek, R., Lienert, F., and Schübeler, D. (2013). Methylation-dependent and -independent genomic targeting principles of the MBD protein family. Cell *153*, 480–492.

Bell, A.C., and Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature *405*, 482–485.

Bowers, S.R., Mirabella, F., Calero-Nieto, F.J., Valeaux, S., Hadjur, S., Baxter, E.W., Merkenschlager, M., and Cockerill, P.N. (2009). A conserved insulator that recruits CTCF and cohesin exists between the closely related but divergently regulated interleukin-3 and granulocyte-macrophage colony-stimulating factor genes. Mol. Cell. Biol. *29*, 1682–1693.

Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994). Sp1 elements protect a CpG island from de novo methylation. Nature *371*, 435–438.

Butcher, D.T., Mancini-DiNardo, D.N., Archer, T.K., and Rodenhiser, D.I. (2004). DNA binding sites for putative methylation boundaries in the unmethylated region of the BRCA1 promoter. Int. J. Cancer *111*, 669–678.

Collings, C.K., Waddell, P.J., and Anderson, J.N. (2013). Effects of DNA methylation on nucleosome stability. Nucleic Acids Res. *41*, 2918–2931.

Dávalos-Salas, M., Furlan-Magaril, M., González-Buendía, E., Valdes-Quezada, C., Ayala-Ortega, E., and Recillas-Targa, F. (2011). Gain of DNA methylation is enhanced in the absence of CTCF at the human retinoblastoma gene promoter. BMC Cancer *11*, 232.

Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. Genes Dev. *25*, 1010–1022.

Ding, Z., Ni, Y., Timmer, S.W., Lee, B.-K., Battenhouse, A., Louzada, S., Yang, F., Dunham, I., Crawford, G.E., Lieb, J.D., et al. (2014). Quantitative genetics of CTCF binding reveal local sequence effects and different modes of X-chromosome association. PLoS Genet. *10*, e1004798.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Feldmann, A., Ivanek, R., Murr, R., Gaidatzis, D., Burger, L., and Schübeler, D. (2013). Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. PLoS Genet. *9*, e1003994.

Filippova, G.N., Fagerlie, S., Klenova, E.M., Myers, C., Dehner, Y., Goodwin, G., Neiman, P.E., Collins, S.J., and Lobanenkov, V.V. (1996). An exceptionally conserved transcriptional repressor, CTCF, employs different combinations of zinc fingers to bind diverged promoter sequences of avian and mammalian c-myc oncogenes. Mol. Cell. Biol. *16*, 2802–2813.

Filippova, G.N., Thienes, C.P., Penn, B.H., Cho, D.H., Hu, Y.J., Moore, J.M., Klesert, T.R., Lobanenkov, V.V., and Tapscott, S.J. (2001). CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the DM1 locus. Nat. Genet. *28*, 335–343.

Gaidatzis, D., Burger, L., Murr, R., Lerch, A., Dessus-Babus, S., Schübeler, D., and Stadler, M.B. (2014). DNA sequence explains seemingly disordered methylation levels in partially methylated domains of mammalian genomes. PLoS Genet. *10*, e1004143.

Groudine, M., and Conkin, K.F. (1985). Chromatin structure and de novo methylation of sperm DNA: implications for activation of the paternal genome. Science *228*, 1061–1068.

Hagemann, S., Heil, O., Lyko, F., and Brueckner, B. (2011). Azacytidine and decitabine induce gene-specific and non-random DNA demethylation in human cancer cell lines. PLoS ONE *6*, e17388.

Hark, A.T., Schoenherr, C.J., Katz, D.J., Ingram, R.S., Levorse, J.M., and Tilghman, S.M. (2000). CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature *405*, 486–489.

Hon, G.C., Rajagopal, N., Shen, Y., McCleary, D.F., Yue, F., Dang, M.D., and Ren, B. (2013). Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat. Genet. *45*, 1198–1206.

Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H.N., Shin, J., Cox, E., Rho, H.S., Woodard, C., et al. (2013). DNA methylation presents distinct binding sites for human transcription factors. eLife *2*, e00726.

John, S., Sabo, P.J., Thurman, R.E., Sung, M.-H., Biddie, S.C., Johnson, T.A., Hager, G.L., and Stamatoyannopoulos, J.A. (2011). Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. Nat. Genet. *43*, 264–268.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Jones, P.A., and Taylor, S.M. (1980). Cellular differentiation, cytidine analogs and DNA methylation. Cell *20*, 85–93.

Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. Nat. Biotechnol. *26*, 1351–1359.

Kim, T.H., Abdullaev, Z.K., Smith, A.D., Ching, K.A., Loukinov, D.I., Green, R.D., Zhang, M.Q., Lobanenkov, V.V., and Ren, B. (2007). Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell *128*, 1231–1245.

Komashko, V.M., and Farnham, P.J. (2010). 5-azacytidine treatment reorganizes genomic histone modification patterns. Epigenetics *5*, 229–240.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lay, F.D., Liu, Y., Kelly, T.K., Witt, H., Farnham, P.J., Jones, P.A., and Berman, B.P. (2015). The role of DNA methylation in directing the functional organization of the cancer epigenome. Genome Res. *25*, 467–477.

Lewis, J.D., Meehan, R.R., Henzel, W.J., Maurer-Fogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992). Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell *69*, 905–914.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915–926.

Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779.

Lienert, F., Wirbelauer, C., Som, I., Dean, A., Mohn, F., and Schübeler, D. (2011). Identification of genetic elements that autonomously determine DNA methylation states. Nat. Genet. *43*, 1091–1097.

Lin, I.G., Tomzynski, T.J., Ou, Q., and Hsieh, C.L. (2000). Modulation of DNA binding protein affinity directly affects target site demethylation. Mol. Cell. Biol. *20*, 2343–2349.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.-M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Macleod, D., Charlton, J., Mullins, J., and Bird, A.P. (1994). Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island. Genes Dev. *8*, 2282–2292.

Matsuo, K., Silke, J., Georgiev, O., Marti, P., Giovannini, N., and Rungger, D. (1998). An embryonic demethylation mechanism involving binding of transcription factors to replicating DNA. EMBO J. *17*, 1446–1453.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012a). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

Maurano, M.T., Wang, H., Kutyavin, T., and Stamatoyannopoulos, J.A. (2012b). Widespread site-dependent buffering of human regulatory polymorphism. PLoS Genet. *8*, e1002599.

Nakahashi, H., Kwon, K.-R.K., Resch, W., Vian, L., Dose, M., Stavreva, D., Hakim, O., Pruett, N., Nelson, S., Yamane, A., et al. (2013). A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Rep. *3*, 1678–1689.

Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012a). BEDOPS: high-performance genomic feature operations. Bioinformatics *28*, 1919–1920.

Neph, S., Vierstra, J., Stergachis, A.B., Reynolds, A.P., Haugen, E., Vernot, B., Thurman, R.E., John, S., Sandstrom, R., Johnson, A.K., et al. (2012b). An expansive human regulatory lexicon encoded in transcription factor footprints. Nature *489*, 83–90.

Ohlsson, R., Renkawitz, R., and Lobanenkov, V. (2001). CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. Trends Genet. *17*, 520–527.

Ohlsson, R., Lobanenkov, V., and Klenova, E. (2010). Does CTCF mediate between nuclear organization and gene expression? BioEssays *32*, 37–50.

Ooi, S.K.T., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.-P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature *448*, 714–717.

Pandiyan, K., You, J.S., Yang, X., Dai, C., Zhou, X.J., Baylin, S.B., Jones, P.A., and Liang, G. (2013). Functional DNA demethylation is accompanied by chromatin accessibility. Nucleic Acids Res. *41*, 3973–3985.

Phillips, J.E., and Corces, V.G. (2009). CTCF: master weaver of the genome. Cell *137*, 1194–1211.

Qiu, X., Hother, C., Ralfkiær, U.M., Søgaard, A., Lu, Q., Workman, C.T., Liang, G., Jones, P.A., and Grønbæk, K. (2010). Equitoxic doses of 5-azacytidine and 5-aza-2′deoxycytidine induce diverse immediate and overlapping heritable changes in the transcriptome. PLoS ONE *5*, e12994.

Ramirez-Carrozzi, V.R., Braas, D., Bhatt, D.M., Cheng, C.S., Hong, C., Doty, K.R., Black, J.C., Hoffmann, A., Carey, M., and Smale, S.T. (2009). A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. Cell *138*, 114–128.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell *159*, 1665–1680.

Reddington, J.P., Perricone, S.M., Nestor, C.E., Reichmann, J., Youngson, N.A., Suzuki, M., Reinhardt, D., Dunican, D.S., Prendergast, J.G., Mjoseng, H., et al. (2013). Redistribution of H3K27me3 upon DNA hypomethylation results in de-repression of Polycomb target genes. Genome Biol. *14*, R25.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P.V. (2007). Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. J. Biol. Chem. *282*, 33336–33345.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408–1419.

Rhee, I., Bachman, K.E., Park, B.H., Jair, K.-W., Yen, R.-W.C., Schuebel, K.E., Cui, H., Feinberg, A.P., Lengauer, C., Kinzler, K.W., et al. (2002). DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. Nature *416*, 552–556.

Rodriguez, C., Borgel, J., Court, F., Cathala, G., Forné, T., and Piette, J. (2010). CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. Biochem. Biophys. Res. Commun. *392*, 129–134.

Schmidt, D., Schwalie, P.C., Wilson, M.D., Ballester, B., Gonçalves, A., Kutter, C., Brown, G.D., Marshall, A., Flicek, P., and Odom, D.T. (2012). Waves of retrotransposon expansion remodel genome organization and CTCF binding in multiple mammalian lineages. Cell *148*, 335–348.

Simmer, F., Brinkman, A.B., Assenov, Y., Matarese, F., Kaan, A., Sabatino, L., Villanueva, A., Huertas, D., Esteller, M., Lengauer, T., et al. (2012). Comparative genome-wide DNA methylation analysis of colorectal tumor and matched normal tissues. Epigenetics *7*, 1355–1367.

Soto-Reyes, E., and Recillas-Targa, F. (2010). Epigenetic regulation of the human p53 gene promoter by the CTCF transcription factor in transformed cell lines. Oncogene *29*, 2217–2227.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature *480*, 490–495.

Stergachis, A.B., Haugen, E., Shafer, A., Fu, W., Vernot, B., Reynolds, A., Raubitschek, A., Ziegler, S., LeProust, E.M., Akey, J.M., and Stamatoyannopoulos, J.A. (2013). Exonic transcription factor binding directs codon choice and affects protein evolution. Science *342*, 1367–1372.

Tate, P.H., and Bird, A.P. (1993). Effects of DNA methylation on DNA-binding proteins and gene expression. Curr. Opin. Genet. Dev. *3*, 226–231.

Thomson, J.P., Skene, P.J., Selfridge, J., Clouaire, T., Guy, J., Webb, S., Kerr, A.R.W., Deaton, A., Andrews, R., James, K.D., et al. (2010). CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature *464*, 1082–1086.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Varley, K.E., Gertz, J., Bowling, K.M., Parker, S.L., Reddy, T.E., Pauli-Behn, F., Cross, M.K., Williams, B.A., Stamatoyannopoulos, J.A., Crawford, G.E., et al. (2013). Dynamic DNA methylation across diverse human cell lines and tissues. Genome Res. *23*, 555–567.

Wang, H., Maurano, M.T., Qu, H., Varley, K.E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., et al. (2012). Widespread plasticity in CTCF occupancy linked to DNA methylation. Genome Res. *22*, 1680–1688.

Wiench, M., John, S., Baek, S., Johnson, T.A., Sung, M.-H., Escobar, T., Simmons, C.A., Pearce, K.H., Biddie, S.C., Sabo, P.J., et al. (2011). DNA methylation status predicts cell type-specific enhancer activity. EMBO J. *30*, 3028–3039.

Witcher, M., and Emerson, B.M. (2009). Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary. Mol. Cell *34*, 271–284.