

Widespread plasticity in CTCF occupancy linked to DNA methylation

Hao Wang,^{1,5} Matthew T. Maurano,^{1,5} Hongzhu Qu,^{1,2,5} Katherine E. Varley,³ Jason Gertz,³ Florencia Pauli,³ Kristen Lee,¹ Theresa Canfield,¹ Molly Weaver,¹ Richard Sandstrom,¹ Robert E. Thurman,¹ Rajinder Kaul,¹ Richard M. Myers,³ and John A. Stamatoyannopoulos^{1,4,6}

¹Department of Genome Sciences, University of Washington, Seattle, Washington 98195, USA; ²Laboratory of Disease Genomics and Individualized Medicine, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing 100029, China; ³HudsonAlpha Institute for Biotechnology, Huntsville, Alabama 35806, USA; ⁴Department of Medicine, University of Washington, Seattle, Washington 98195, USA

CTCF is a ubiquitously expressed regulator of fundamental genomic processes including transcription, intra- and inter-chromosomal interactions, and chromatin structure. Because of its critical role in genome function, CTCF binding patterns have long been assumed to be largely invariant across different cellular environments. Here we analyze genome-wide occupancy patterns of CTCF by ChIP-seq in 19 diverse human cell types, including normal primary cells and immortal lines. We observed highly reproducible yet surprisingly plastic genomic binding landscapes, indicative of strong cell-selective regulation of CTCF occupancy. Comparison with massively parallel bisulfite sequencing data indicates that 41% of variable CTCF binding is linked to differential DNA methylation, concentrated at two critical positions within the CTCF recognition sequence. Unexpectedly, CTCF binding patterns were markedly different in normal versus immortal cells, with the latter showing widespread disruption of CTCF binding associated with increased methylation. Strikingly, this disruption is accompanied by up-regulation of CTCF expression, with the result that both normal and immortal cells maintain the same average number of CTCF occupancy sites genome-wide. These results reveal a tight linkage between DNA methylation and the global occupancy patterns of a major sequence-specific regulatory factor.

[Supplemental material is available for this article.]

The polyfunctional regulator CTCF plays a central role in multiple complex genomic processes, including transcription (Baniahmad et al. 1990; Filippova et al. 1996; Vostrov and Quitschke 1997), imprinting (Bell and Felsenfeld 2000; Hark et al. 2000), and long-range chromatin interactions and subnuclear localization (Yusufzai et al. 2004; Splinter 2006; Hou et al. 2008). Cohesin, a major mediator of chromosomal contacts during mitosis (Seitan et al. 2011), is tightly co-localized with CTCF, indicating a key function for CTCF in chromosome pairing (Parelho et al. 2008; Rubio et al. 2008; Wendt et al. 2008). CTCF has also been connected with multiple malignancies, including by the association of mutations in its gene locus (Filippova et al. 1998), through its anti-proliferative effect (Rasko et al. 2001), and through regulatory interactions with tumor suppressor genes (Butcher et al. 2004; Witcher and Emerson 2009; Soto-Reyes and Recillas-Targa 2010; Dávalos-Salas et al. 2011).

CTCF is ubiquitously expressed, and it is widely believed that CTCF binding patterns are largely invariant between cell types (Kim et al. 2007; Cuddapah et al. 2008; Heintzman et al. 2009), though diverse regulatory mechanisms at individual loci have been described (Lefevre et al. 2008; Sekimata et al. 2009; Witcher and Emerson 2009; Lai et al. 2010; Shukla et al. 2011). In addition,

at a small number of loci, variable CTCF occupancy has been linked with DNA methylation *in vivo* (Kanduri et al. 2000; Pant et al. 2003), and *in vitro* studies suggest that methylation may hinder CTCF binding at certain sequence elements (Bell and Felsenfeld 2000; Hark et al. 2000; Filippova et al. 2001; Renda et al. 2007). However, neither the degree to which CTCF binding patterns vary between different cell types nor the relationship of such variability with DNA methylation is currently known.

We therefore sought to establish the cellular selectivity of CTCF binding and to define its relationship with methylation on a global scale. By using genome-wide occupancy profiling and reduced representation bisulfite sequencing (RRBS), we establish that a majority of CTCF sites are cell-selective, and link 41% of this variable CTCF occupancy to differential DNA methylation. We further observe markedly different CTCF binding patterns distinguishing normal and immortal cells, which are associated with increased methylation and up-regulation of CTCF expression. These results indicate a global linkage between DNA methylation and the occupancy patterns of an important genome regulator.

Results

Widespread plasticity of CTCF occupancy patterns

To assess CTCF binding variation genome-wide, we localized and quantified CTCF occupancy by ChIP-seq in 19 diverse cell types, including seven immortal cell lines and 12 normal cell types. We generated two biological replicates for each cell type. Both repli-

⁵These authors contributed equally to this work.

⁶Corresponding author

E-mail jstam@uw.edu

Article and supplemental material are at <http://www.genome.org/cgi/doi/10.1101/gr.136101.111>. Freely available online through the *Genome Research* Open Access option.

cates were of high enrichment and exhibited high concordance (average correlation of 0.93) (Supplemental Fig. S1). We found that CTCF binds an average of about 55,000 sites in each tested cell type (Supplemental Fig. S1A). In total, we identified 77,811 distinct binding sites across all 19 cell types.

To survey binding variability genome-wide, we conservatively assessed how many cell types demonstrated binding at each site using a dual-threshold strategy to prevent bias toward variable sites (see Methods). In all 19 cell types, 27,662 binding sites were present. However, 50,149 binding sites were found to be unbound in at least one cell type (Supplemental Table S2). Thus, 64% of CTCF sites are found to vary in at least one cell type, demonstrating the existence of a widespread variability in CTCF occupancy. These variable sites exhibited clear occupancy differences between bound and unbound cell types, including at the well-known *H19/IGF2* imprinted locus (Fig. 1A–C). Variable binding sites were occupied in an average of 10 of 19 cell lines, implying a high degree of shared regulation between cell types (Fig. 1D). Indeed, between any two cell types, an average of 72% of bound sites were in common (Supplemental Fig. S3). Variable sites had a similar genomic localization (Fig. 1E) compared with constitutive sites.

Distinct CTCF binding landscapes in normal vs. immortal cells

To understand whether binding variability follows a similar pattern in related cell types, we performed an unsupervised hierarchical clustering of variable CTCF binding sites (see Methods). We found that the variable CTCF binding landscape distinguished three groups (Fig. 2A). The first group of immortal cells consists of malignancy-derived and EBV-immortalized cell lines, including several carcinomas (colorectal, Caco-2; cervical, HeLa-S3; hepatocellular, HepG2), neuroblastoma (SK-N-SH_RA), and retinoblastoma (WERI-Rb-1) and EBV-transformed lymphoplastoid (GM06990). The remaining two groups consist of normal cell types of limited proliferative potential: The second group consists of three epithelial cell types, including renal cortical (HRE), small airway (SAEC), and esophageal (HEEpiC) mucosal epithelia, and the third group consists of fibroblasts, including abdominal (AG10803), toe skin (AG09309), gum (AG09319), aortic adventitial (AoAF), foreskin (BJ), mammary (HMF), pulmonary artery (HPAF), and pulmonary (HPF) and brain microvascular endothelium (HBMEC). Principal component analysis and bootstrap assessment of the uncertainty in the hierarchical clustering confirmed a separation between the normal cell types and remaining cell lines, although the epithelial line HRE

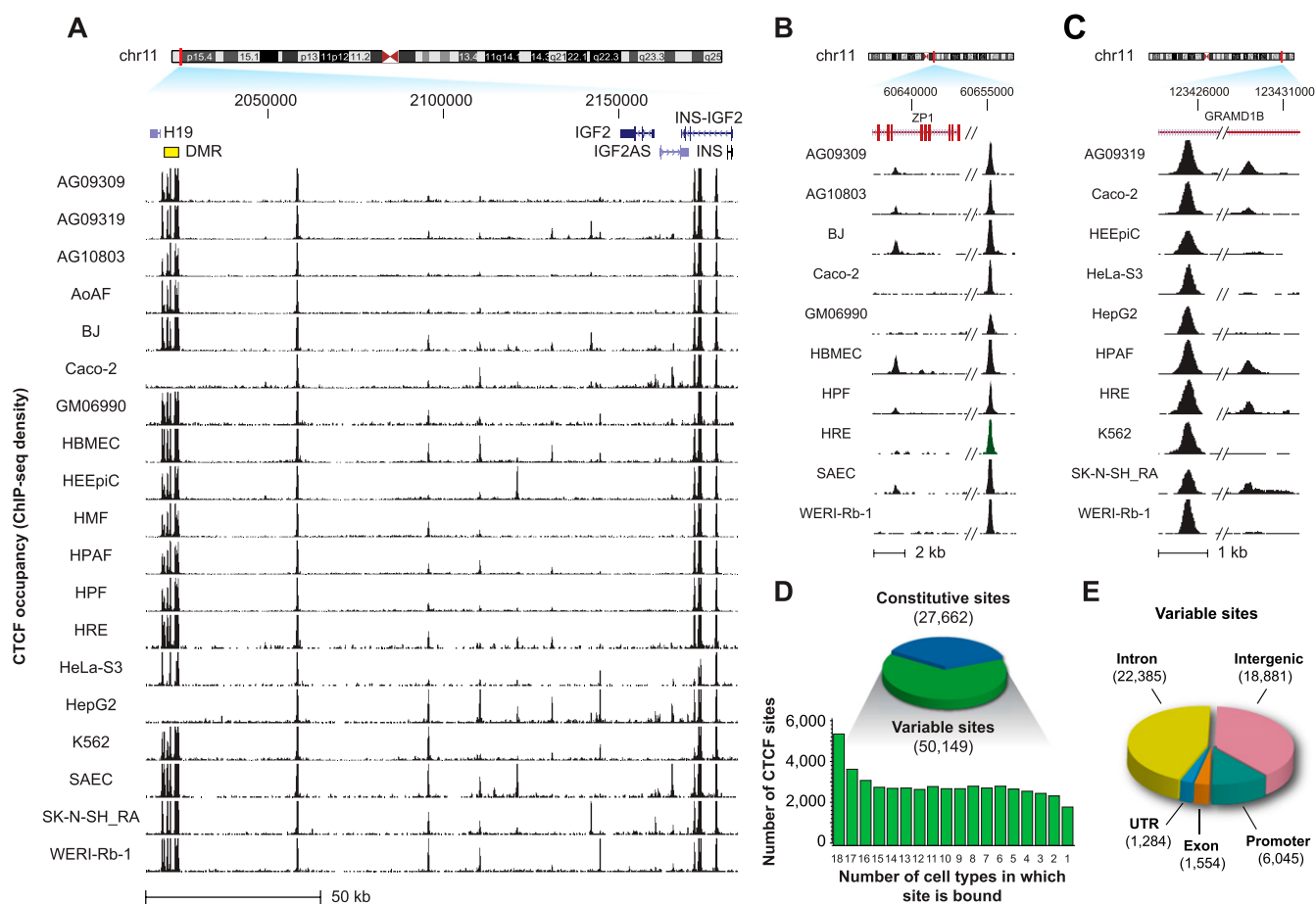


Figure 1. CTCF in vivo binding exhibits widespread plasticity. (A–C) Constitutive and variable CTCF sites. (A) The *H19/IGF2* imprinted locus in multiple human cell types. Note the total silencing in two cell lines of the seven CTCF sites in the differentially methylated region (DMR; yellow box at left), and the complex pattern of cell-selective CTCF binding flanked by constitutive sites. Location (hg19), chr11:2,015,000–2,184,000. (B,C) Additional examples of variable sites. (D) Genome-wide analysis of CTCF binding in 19 cell types reveals 77,811 distinct binding sites; 27,662 sites are constitutively present in all cell types; 50,149 variable sites exhibiting a wide range of selectivity are present in a subset of one to 18 cell types (below). (E) Genomic distribution of variable sites is similar to constitutive sites (Supplemental Fig. S2A).

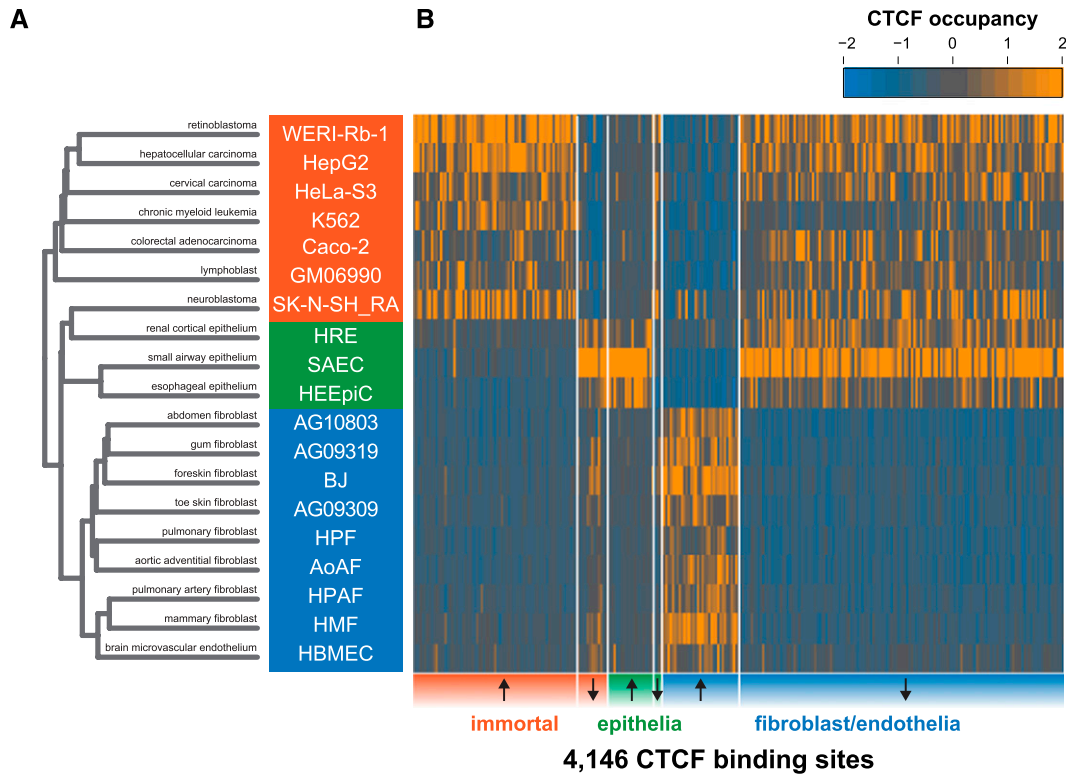


Figure 2. CTCF occupancy distinguishes similar cell types. (A) Unsupervised hierarchical clustering of binding at all CTCF sites. (B) CTCF occupancy at 4146 variable binding sites that distinguish immortal cell lines, epithelia, fibroblasts and endothelia (Methods). *x*-axis, CTCF binding sites in chromosomal order, separated into sites that are up-regulated and down-regulated (arrows) in each of the three groups (immortal, epithelial, fibroblast, and endothelial). Color corresponds to Z-score of normalized ChIP-seq density.

was less clearly distinguished (Supplemental Fig. S4). We then sought to identify the specific binding differences characterizing these three groups. We identified 4146 specific binding sites whose occupancy was significantly different between these groups at a false-discovery rate (FDR) of 1% (Methods) (Fig. 2B; Supplemental Table S3). These results suggest that CTCF occupancy exhibits major regulatory differences distinguishing immortal cell lines from normal epithelium, endothelium, and fibroblasts.

Variable CTCF occupancy linked to CpG methylation

Pre-existing methylation can antagonize CTCF binding *in vitro* (Bell and Felsenfeld 2000; Hark et al. 2000; Kanduri et al. 2000). Therefore we asked whether differential methylation was associated with variable sites *in vivo*. To study this, we compared CTCF occupancy and RRBS data (Fig. 3A). We studied a subset of CTCF sites in 13 cell types ($n = 6,707$) for which RRBS data were available from the ENCODE project (KE Varley, J Gertz, KM Bowling, SL Parker, TE Reddy, F Pauli, MK Cross, BA Williams, JA Stamatoyannopoulos, GE Crawford, et al., *in prep.*). We obtained methylation status of 44,048 CpGs dinucleotides in the region centered on these sites (see Methods), with each CpG monitored in an average of 12 out of 13 cell types (Supplemental Fig. S6).

First, we assessed the overall methylation status at the 6707 CTCF sites with RRBS data. We found that methylation was substantially more variable at variable CTCF sites than at constitutive ones (Supplemental Fig. S5). Only 10% of these sites tested showed intermediate methylation status (between 25% and 75% methylation) (Supplemental Fig. S6). Overall, 98% of CTCF sites were un-

methyated (defined as <50% methylation) in at least one of the cell types tested, confirming an inverse relationship between methylation and CTCF occupancy. However, 47% of CTCF sites were methylated (>50% methylation) in at least one cell type, suggesting a widespread potential link between methylation and CTCF occupancy.

To quantify the global association of differential methylation status with variable CTCF occupancy, we performed a linear regression analysis at the 6707 sites for which we had RRBS data (Fig. 3B; see Methods). Four thousand ninety-nine (61%) of these sites exhibited variable CTCF binding in the 13 cell types tested. Of the 4099 variable sites with RRBS data, 1677 (41%) showed a significant association (5% FDR) between methylation and occupancy (Fig. 3C). At significant sites, increased methylation was negatively associated with occupancy in 98% of cases. The magnitude of the association between methylation and occupancy was strong: Occupancy was on average 87% lower at significant sites in the methylated cell types relative to the unmethylated cell types (Fig. 3D). Further supporting a strong link to methylation, 67% of variable methylation was associated with a concomitant affect on occupancy. The remaining 36% of sites with variable methylation that was not associated with occupancy nevertheless demonstrated an aggregate reduction in occupancy in methylated cell types (Supplemental Fig. S7), confirming the overall inverse association of methylation with CTCF occupancy but suggesting that this relationship may be complicated by additional factors at this subset of sites.

Next we asked if the inverse relationship between methylation and CTCF occupancy is characterized by regional hyper-

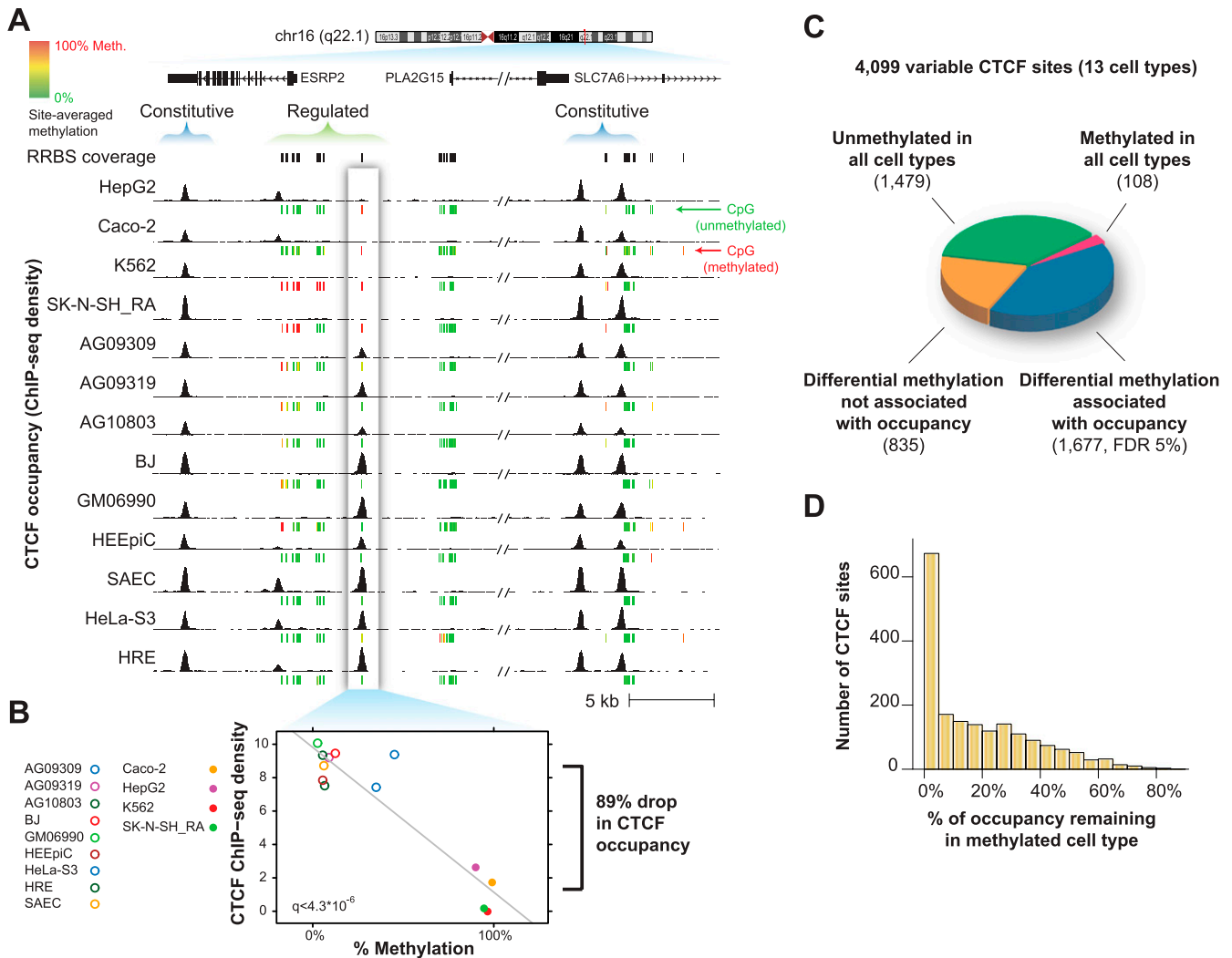


Figure 3. Impact of DNA methylation on cell-selective CTCF binding. (A) Example CTCF binding sites, where occupancy (*above*) quantitatively increases as local CpG methylation decreases (*below*). Green indicates CpG is 0% methylated; yellow, 50%; and red, 100%. (B) Quantitative analysis of methylation at the boxed CTCF binding site in A. (C) Global impact of methylation at variable CTCF sites monitored by RRBS. Sixty-five percent of sites with cell-type selective patterns of methylation also exhibited differences in occupancy. (D) At methylated binding sites, occupancy was reduced on average by 87% compared with cell lines without methylation at the same site. Shown are sites where increased methylation was associated with decreased occupancy (98% of all significant sites).

methylation or if instead methylation is concentrated specifically at the region of protein–DNA interaction. We examined the location of all CpG dinucleotides relative to the CTCF motif at sites with variable methylation. Indeed, sites of differential methylation associated with occupancy differences showed an enrichment of CpG dinucleotides at two positions in the CTCF recognition sequence (Fig. 4). This finding is consistent with previous reports showing methylation outside the recognition sequence does not affect CTCF binding *in vitro* (Engel et al. 2004; Chadwick 2008). Within the recognition sequence, methylation at one of these CpGs (position 1) has been shown to inhibit binding of CTCF *in vitro* (Renda et al. 2007). The second (position 11) is the predominant CpG in the motif, which has been shown to have a higher rate of C–T transitions at vertebrate-conserved binding sites (Kim et al. 2007), consistent with germline methylation. Interestingly, constitutively unmethylated CTCF sites also showed an enrichment of CpGs at these two positions compared with

differentially methylated sites without an association to occupancy (Supplemental Fig. S8). Given that the latter sites nevertheless exhibit substantial methylation variability, this suggests that the absence of CpGs at these positions may decouple CTCF occupancy from differential methylation at these sites. Overall, 29% of CTCF recognition sequences genome-wide contain a CpG at positions 1 and/or 11, and 52% of recognition sequences contain a CpG anywhere in the sequence. The genome-wide prevalence of “susceptible” CTCF sites suggests a widespread potential for interaction between CTCF and methylation.

Methylated-associated remodeling of CTCF binding in immortal cell lines

Paralleling prior reports of widespread hypermethylation in cancer (Jones and Baylin 2007; KE Varley, J Gertz, KM Bowling, SL Parker, TE Reddy, F Pauli, MK Cross, BA Williams, JA Stamatoyannopoulos,

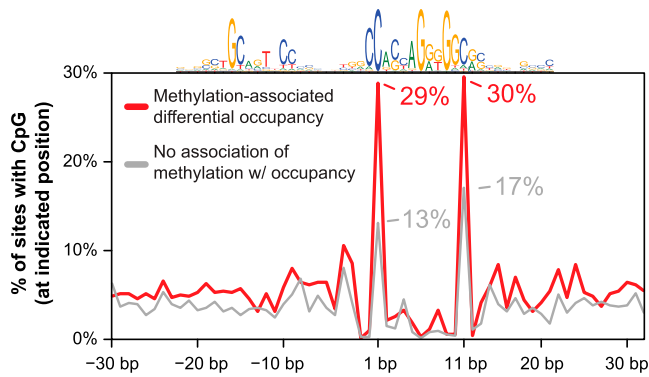


Figure 4. Sites significantly affected by methylation are enriched for CpGs at two positions. Frequency of a CpG (y-axis) at positions relative to the CTCF motif (x-axis) is shown for sites with variable methylation that is associated (red) and is not associated (gray) with occupancy changes. Note that at positions 1 and 11, there is a 2.2- and 1.8-fold enrichment, respectively, for the presence of a CpG at sites where the variable methylation was not associated with occupancy. Twenty-nine percent of CTCF motifs genome-wide contain a CpG at one or both of these positions.

GE Crawford, et al., in prep.), we observed a bimodal pattern of methylation at CTCF sites distinguishing normal and immortal cell types (Fig. 5A). At 31% of the sites where differential methylation was associated with CTCF occupancy, methylation was observed throughout the 13 normal and immortal cell types (average number of methylated cell types, 7.3). In contrast, the remaining 69% of sites were characterized by cell-specific hypermethylation constrained to the six immortal lines (average number of methylated cell lines, 2.1) (Fig. 5A, strip at right). Notably, although the neuroblastoma line SK-N-SH_RA clusters with epithelial cell types based purely on CTCF binding (Fig. 2A), it exhibits the hypermethylation characteristic of the other immortal lines. Surprisingly, the increased methylation in immortal lines does not correspond to a decrease in the total number of bound CTCF sites (Fig. 5B). Strikingly, we also observed that CTCF transcript levels are significantly higher in the immortal cell lines (Fig. 5C). This disruption of CTCF binding in immortal cell lines is further distinguished by a unique association between CTCF occupancy and methylation at promoter sites. Of the promoter CTCF sites where methylation was significantly associated with occupancy, 98% (281 of 288) of these sites were characterized by hypermethylation in the immortal lines (Fig. 5D). These results suggest a widespread methylation-associated remodeling of the CTCF binding landscape in immortal cell lines.

Discussion

Surprising plasticity of the CTCF occupancy landscape

This study exposes a previously unappreciated degree of plasticity within the binding landscape of the master genomic regulator CTCF. Previous studies in a small number of cell types had uncovered only limited cell-type specificity (Kim et al. 2007; Cuddapah et al. 2008; Heintzman et al. 2009). We further associate differential methylation with 41% of this variable binding at a subset of sites overlapping existing RRBS data in 13 cell types. We specifically linked this variable methylation to the presence of a CpG at two key positions relative to the consensus motif. Finally, we observe the maintenance of a stable total amount of CTCF genomic binding sites in immortal cell lines despite their altered localization as-

sociated with increased methylation. Our results show that methylation is indeed a global feature of the regulatory diversity of CTCF, and our approach is readily extensible to the repertoire of vertebrate transcription factors.

Methylation-associated disruption of CTCF binding in immortal lines

Although CTCF binding varied across all 13 cell types, we observed unique patterns of CTCF occupancy specific to the immortal cell lines. Interestingly, CTCF overexpression has previously been associated with resistance to apoptosis in breast cancer cell lines (Docquier et al. 2005) and with DNMT3B overexpression (Butcher et al. 2004). Further, the unique occurrence of hypermethylation-associated abrogation of CTCF occupancy at promoters in immortal lines is notable, given the involvement of CTCF in the methylation-associated silencing of known tumor suppressors and oncogenes (Witcher and Emerson 2009; Lai et al. 2010; Soto-Reyes and Recillas-Targa 2010). We found that the immortal cell lines we profiled have the same overall amount of genomic CTCF binding sites despite a redistribution of CTCF occupancy from binding sites subject to hypermethylation. The concomitant up-regulation of CTCF expression may therefore represent a cancer-associated compensatory mechanism. This inverse correlation is compatible with the existence of a stabilizing mechanism acting through increased CTCF expression to maintain a constant level of genomic binding despite increased methylation at its target sites, although further study in an expanded set of cell types will be necessary.

The role of DNA methylation in regulation of transcription factor occupancy

Although DNA methylation is widely invoked as a causal mechanism for transcriptional repression, surprisingly little *in vivo* evidence is available. While experimentally directed methylation can prevent binding of CTCF and other factors *in vitro* (Tate and Bird 1993; Renda et al. 2007), the mechanisms establishing methylation patterns *in vivo* remain unknown, and its precise relationship with gene expression remains unclear (Enver et al. 1988; Selker 1990; Walsh and Bestor 1999). Likewise, our results do not distinguish whether demethylation facilitates subsequent CTCF binding or whether bound CTCF maintains an unmethylated domain.

An alternative model has DNA methylation deposited passively in the wake of independent abrogation of transcription factor binding. This model is equally consistent with evidence that transcription factor binding sites appear to be generally depleted for DNA methylation (Mukhopadhyay et al. 2004; Lister et al. 2009; Thurman et al. 2012) and that binding sites recognized by certain sequence-specific factors have been associated with lack of methylation (Straussman et al. 2009; Dickson et al. 2010; Gebhard et al. 2010; Lienert et al. 2011). Indeed, there is evidence that the binding of some transcription factors, including CTCF, is sufficient to effect a local demethylated state (Matsuo et al. 1998; Lin et al. 2000; Stadler et al. 2011). But if *in vivo* methylation was deposited generally at unoccupied binding sites, then how would this process interact with the *in vitro* methylation sensitivity of common transcription factors?

The well-investigated *H19/Igf2* imprinted locus offers an appropriate example: CTCF binding there has been shown necessary to maintain an existing unmethylated state (Schoenherr et al. 2002; Pant et al. 2004). However, CTCF is not the originator of the unmethylated state (Matsuzaki et al. 2010), implying a limited capacity to directly affect methylation. Perhaps methylation instead

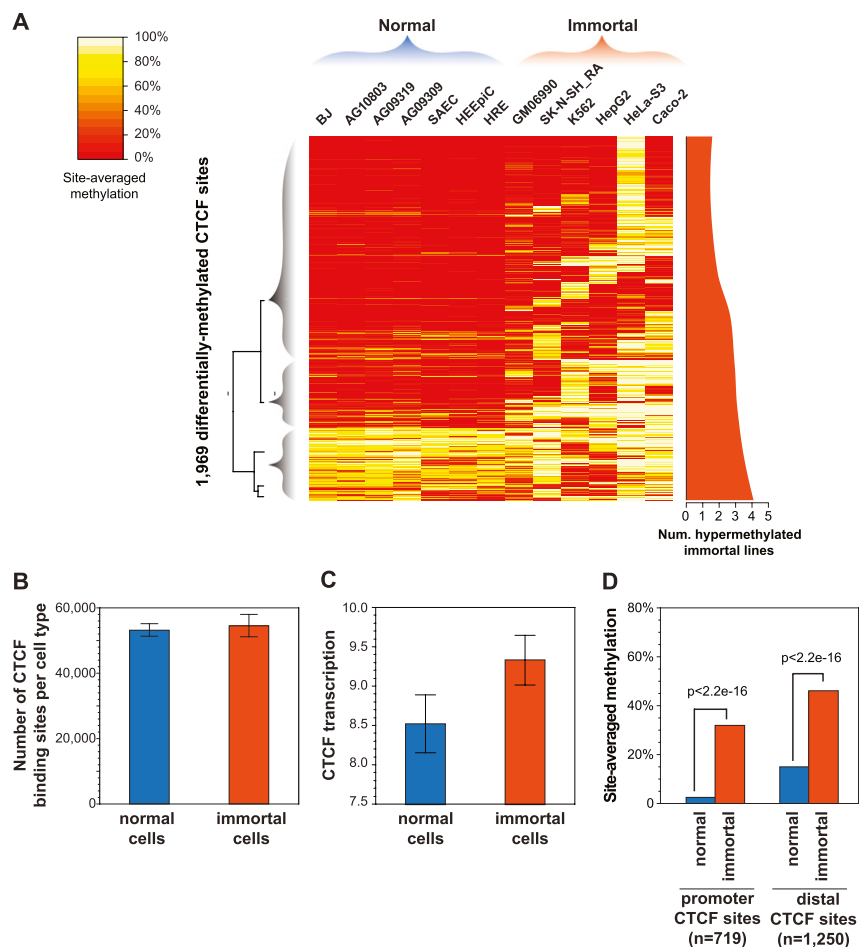


Figure 5. Cell-selective patterns of methylation associated with occupancy differences. (A) Methylation status at 1969 CTCF sites where differential methylation is significantly associated with occupancy differences. Color corresponds to the percentage of bisulfite sequencing tags at each site overlapping methylated CpG positions. Dendrogram (left) highlights pattern of hypermethylation in immortal cell lines. (Right) Smoothed plot of number of immortal lines exhibiting hypermethylation at each site. (B) Immortal lines show no significant difference in number of occupied CTCF sites (y -axis, mean). Error bars, SD. (C) immortal lines demonstrate increased CTCF transcript levels (y -axis, mean). Error bars, SD. (D) Immortal lines exhibit increased methylation relative to the other cell types, though significant promoter methylation is rarely observed in normal lines. y -axis, genome-wide median of per-site methylation. P -values, Wilcoxon. Promoter, ± 2.5 kb of RefSeq transcription start site.

acts as a cooperative switch to prevent the return of CTCF after a reprogramming event. In this model, rather than guiding binding localization, methylation is a general amplifier of perturbations to transcription factor occupancy.

Other sources of variable CTCF binding

Although we have shown that 41% of overall CTCF occupancy variation is significantly linked to methylation at tested sites, 36% of variable CTCF sites overlap no variable methylation at all. It is unlikely that much of this variability is associated with genetic variability in CTCF recognition sequences (Maurano et al. 2012), though some sites may associate with modified forms of CTCF (Klenova et al. 2001; Yu et al. 2004; MacPherson et al. 2009). One likely possibility is that the constantly unmethylated variable CTCF sites may represent instances of cooperative regulation that complicate a direct relationship between methylation and CTCF occupancy. Accordingly, CTCF has been known to interact with

a number of cofactors that could potentially govern its selectivity at these sites or, alternatively, maintain demethylation in the absence of CTCF binding (Chernukhin et al. 2007; Donohoe et al. 2007, 2009; Parelho et al. 2008; Rubio et al. 2008; Wendt et al. 2008; Ohlsson et al. 2010; Liu et al. 2011). Interestingly, we found that of the 36% of variable sites despite constant methylation, 76% were within 2.5 kb of a RefSeq transcription start site, compared with 38% of the variable sites associated with methylation differences. Recent work has further observed an enrichment of tethered CTCF peaks at promoters (Neph et al. 2012b), suggesting that the remaining variation in CTCF occupancy may derive from complex regulation of co-factors or variation in its specific interaction partners. Given the breadth of CTCF's regulatory functionality, our observation of global binding variation implies a widespread potential role in the translation of epigenetic marks to genome organization at thousands of sites.

Methods

Cell culture

Cells were cultured in an appropriate growth medium, with the addition of growth factors and supplements according to the suppliers' instructions (Supplemental Table S1). Cell lines were maintained in a humidified incubator at 37°C in the presence of 5% CO₂.

ChIP-seq

Suspension cells were cross-linked with formaldehyde (Sigma) at a final concentration of 1% for 10 min at room temperature. Adherent cells were first detached from the plates by 0.05% Trypsin-EDTA and Trypsin neutralizer solution (Invitrogen) and then cross-linked by 1% formaldehyde. Glycine was added to a final concentration of 0.125 M for 5 min. Cells were rinsed twice with phosphate buffered saline, lysed in lysis buffer (50 mM Tris-HCl at pH 8.0, 10 mM EDTA, 1% SDS) containing protease inhibitor cocktail (Roche), and sheared by Bioruptor (Diagenode). The chromatin was incubated with Dynabeads (M-280, sheep anti-rabbit IgG, Invitrogen)-conjugated anti-CTCF polyclonal antibody (Cell Signaling no. 2899).

The CTCF-DNA complexes were washed, eluted, and reverse cross-linked. The DNA was RNase A-, Proteinase K-treated, and purified by phenol-chloroform-isoamyl alcohol extraction and ethanol precipitation. DNA was end-repaired (End-it DNA End-repair kit, Epicentre), followed by the addition of adenine to the 3' ends (Taq DNA polymerase, NEB), and ligated to an adapter (Illumina). Purified ligation product was PCR amplified and run on a 2% agarose gel. The size-selected libraries were sequenced on an Illumina Genome Analyzer (Illumina) by the High-Throughput

Genomics Center (University of Washington) according to a standard protocol.

For each cell type, experiments were conducted on two independent biological replicates.

Identification and quantification of CTCF binding sites

We obtained Uniform Element Calls from the ENCODE project for each cell line. Briefly, peaks were called using SPP (Kharchenko et al. 2008). The set of peaks reproducible in both replicates were identified based on an irreproducible discovery rate (IDR) of 0.25% (Li et al. 2011). We then combined peak calls from 19 cell types to generate a master list of all distinct CTCF binding sites. We adjusted the peak locations to center on matches to the nearest CTCF motif ($P < 10^{-5}$, fimo) if the motif was within 50 bp.

To distinguish between variable and constitutive binding sites, for each site we examined the presence of a peak in each of 19 cell types. We used the peak calling program Hotspot (John et al. 2011) to enable a conservative procedure for the identification of variable binding sites. To reduce the misclassification of sites near the peak-calling threshold as variable, we employed separate cut-offs for calling peak presence and absence. First, for each CTCF binding site called above, we additionally required that it overlap a 0.5% FDR hotspot in both replicates of at least one cell line. Then, a binding site was counted as occupied in subsequent cell lines if a looser 1% FDR hotspot was present in one or both replicates for that cell line. Employing this looser criteria for binding in subsequent cell types results in conservative identification of variable sites. We confirmed that binding sites in cell types considered absent were substantially closer to background than sites in cell types considered active (Supplemental Fig. S2B).

ChIP-seq data were mapped to the human genome (GRCh37/hg19) using bowtie (Langmead et al. 2009) with the options “bowtie-mm -n 3 -v 3 -k 2 -phred64-quals,” allowing up to three 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. Smoothed density tracks were generated using bedmap (<http://code.google.com/p/bedops/>) to count the number of tags overlapping a sliding 150-bp window, with a step width of 20 bp (Neph et al. 2012a). Density tracks were normalized for sequencing depth by a global linear scaling to 10 million tags. We measured occupancy by the maximum normalized ChIP-seq tag density over the 134-bp region.

Reproducibility of ChIP-seq experiments was tested using Pearson correlation on normalized density tracks of chromosome 19 between each replicate.

Clustering of cell-selective CTCF binding sites

We converted the presence and absence of a given peak to 1 and 0, respectively, in 19 cell lines. We then performed hierarchical clustering with the `hclust` function in R, using the “average” method and Euclidean distance metric. We cut the dendrogram (Fig. 2A) into three groups, of immortal cell lines, epithelia, and fibroblasts. To assess the significance of these three groups, we used the R package `pclus` (Suzuki and Shimodaira 2006) and principal components analysis (Supplemental Fig. S4). We then used the package `DESeq` (Anders and Huber 2010) on the tag count at each peak to identify differentially occupied sites between each of these three groups (FDR 1%).

RRBS genome-wide methylation profiling

We downloaded RRBS methylation data for 13 cell lines from the “HAIB Methyl RRBS” track (KE Varley, J Gertz, KM Bowling, SL Parker, TE Reddy, F Pauli, MK Cross, BA Williams, JA Stamatoyannopoulos,

GE Crawford, et al., in prep.) of the UCSC Genome Browser. To measure methylation in each cell line, we combined counts for both strands in both replicates and removed data for samples with less than 8× coverage. We retained only CpGs monitored in at least six samples (Supplemental Fig. S6B).

We applied a linear regression to measure whether methylation status is associated with occupancy. We normalized CTCF occupancies using the `getVarianceStabilizedData` function of `DESeq` and then averaged replicate signals. We regressed CTCF occupancy onto the average proportion methylated of all monitored CpGs in a 134-bp region centered around the CTCF peak. We excluded 1806 sites missing RRBS data and ChIP-seq data for seven or more cell types or having too great a difference in the number of CpGs monitored between any two cell types (more than six CpGs monitored). We averaged the methylation level of all CpGs within a 134-bp window to increase sensitivity and reliability. We excluded sites where the number of monitored CpGs differed by more than four among any two cell lines. We used the R package `qvalue` to estimate an FDR (Storey and Tibshirani 2003).

RNA expression analysis

For each cell line, total RNA was extracted in two replicates from 5×10^6 cells using Ribopure (Ambion) according to the manufacturer's instructions. RNA quality was ascertained using RNA 6000 Nano Chips on a bioanalyzer (Agilent). Approximately 3 µg of total RNA for each sample was used for labeling and hybridization (University of Washington Center for Array Technology) to Affymetrix Human Exon 1.0 ST arrays (Affymetrix) using a standard protocol. Exon expression data were analyzed through Affymetrix Expression Console using gene-level RMA summarization and sketch-quantile normalization method. Measurements from both replicates were then averaged.

Data access

CTCF ChIP-seq data have been submitted to the NCBI Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) under accession no. GSE30263. Affymetrix exon array data are available under accession no. GSE19090. RRBS methylation data are under accession no. GSE27584. All three sets are available for viewing in the UCSC Genome Browser (<http://genome.ucsc.edu/>).

Acknowledgments

We thank Jeff Vierstra, Andrew Stergachis, and Sam John for critical reading of the manuscript and many helpful suggestions. We also thank Daniel Bates, Morgan Diegel, and Doug Dunn at the University of Washington High-Throughput Genomics Center for technical assistance. This work was supported by National Institutes of Health grants U54HG004592 (J.A.S.) and U54HG004576 (R.M.M.).

Author contributions: H.W., M.T.M., and J.A.S. conceived the study. H.W. and T.C. cultured cells. H.W. and K.L. produced ChIP-seq data. H.W. and M.W. generated Illumina libraries. K.E.V., J.G., and F.P. generated RRBS data under the supervision of R.M.M. M.T.M., R.S., and R.E.T. processed data. M.T.M. and H.Q. analyzed data. R.K. oversaw production data collection and aspects of primary analysis. H.W. and M.T.M. wrote the manuscript, with contributions from J.A.S.

References

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**: R106. doi: 10.1186/gb-2010-11-10-r106.

- Baniahmad A, Steiner C, Köhne AC, Renkawitz R. 1990. Modular structure of a chicken lysozyme silencer: Involvement of an unusual thyroid hormone receptor binding site. *Cell* **61**: 505–514.
- Bell AC, Felsenfeld G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**: 482–485.
- Butcher DT, Mancini-DiNardo DN, Archer TK, Rodenhiser DI. 2004. DNA binding sites for putative methylation boundaries in the unmethylated region of the *BRCA1* promoter. *Int J Cancer* **111**: 669–678.
- Chadwick BP. 2008. DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res* **18**: 1259–1269.
- Chernukhin I, Shamsuddin S, Kang SY, Bergstrom R, Kwon YW, Yu W, Whitehead J, Mukhopadhyay R, Docquier F, Farrar D, et al. 2007. CTCF interacts with and recruits the largest subunit of RNA polymerase II to CTCF target sites genome-wide. *Mol Cell Biol* **27**: 1631–1648.
- Cuddapah S, Jothi R, Schones DE, Roh TY, Cui K, Zhao K. 2008. Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res* **19**: 24–32.
- Dávalos-Salas M, Furlan-Magaril M, González-Buendía E, Valdes-Quezada C, Ayala-Ortega E, 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. doi: 10.1186/1471-2407-11-232.
- Dickson J, Gowher H, Strogantsev R, Gaszner M, Hair A, Felsenfeld G, West AG. 2010. VEZF1 elements mediate protection from DNA methylation. *PLoS Genet* **6**: e1000804. doi: 10.1371/journal.pgen.1000804.
- Docquier F, Farrar D, D'Arcy V, Chernukhin I, Robinson AF, Loukinov D, Vatolin S, Pack S, Mackay A, Harris RA, et al. 2005. Heightened expression of CTCF in breast cancer cells is associated with resistance to apoptosis. *Cancer Res* **65**: 5112–5122.
- Donohoe ME, Zhang L-F, Xu N, Shi Y, Lee JT. 2007. Identification of a Ctf cofactor, Yy1, for the X chromosome binary switch. *Mol Cell* **25**: 43–56.
- Donohoe ME, Silva SS, Pinter SF, Xu N, Lee JT. 2009. The pluripotency factor Oct4 interacts with Ctf and also controls X-chromosome pairing and counting. *Nature* **460**: 128–132.
- Engel N, West AG, Felsenfeld G, Bartolomei MS. 2004. Antagonism between DNA hypermethylation and enhancer-blocking activity at the *H19* DMD is uncovered by CpG mutations. *Nat Genet* **36**: 883–888.
- Enver T, Zhang JW, Papayannopoulou T, Stamatoyannopoulos G. 1988. DNA methylation: A secondary event in globin gene switching? *Genes Dev* **2**: 698–706.
- Filippova GN, Fagerlie S, Klenova EM, Myers C, Dehner Y, Goodwin G, Neiman PE, Collins SJ, Lobanenkov VV. 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 GN, Lindblom A, Meincke LJ, Klenova EM, Neiman PE, Collins SJ, Doggett NA, Lobanenkov VV. 1998. A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. *Genes Chromosomes Cancer* **22**: 26–36.
- Filippova GN, Thienes CP, Penn BH, Cho DH, Hu YJ, Moore JM, Klesert TR, Lobanenkov VV, Tapscott SJ. 2001. CTCF-binding sites flank CTG/CAG repeats and form a methylation-sensitive insulator at the *DM1* locus. *Nat Genet* **28**: 335–343.
- Gebhard C, Benner C, Ehrlich M, Schwarzfischer L, Schilling E, Klug M, Dietmaier W, Thiede C, Holler E, Andreesen R, et al. 2010. General transcription factor binding at CpG islands in normal cells correlates with resistance to *de novo* DNA methylation in cancer cells. *Cancer Res* **70**: 1398–1407.
- Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**: 486–489.
- Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**: 108–112.
- Hou C, Zhao H, Tanimoto K, Dean A. 2008. CTCF-dependent enhancer-blocking by alternative chromatin loop formation. *Proc Natl Acad Sci* **105**: 20398–20403.
- John S, Sabo PJ, Thurman RE, Sung M-H, Biddie SC, Johnson TA, Hager GL, Stamatoyannopoulos JA. 2011. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat Genet* **43**: 264–268.
- Jones PA, Baylin SB. 2007. The epigenomics of cancer. *Cell* **128**: 683–692.
- Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV. 2000. Functional association of CTCF with the insulator upstream of the *H19* gene is parent of origin-specific and methylation-sensitive. *Curr Biol* **10**: 853–856.
- Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat Biotechnol* **26**: 1351–1359.
- Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B. 2007. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. *Cell* **128**: 1231–1245.
- Klenova EM, Chernukhin IV, El-Kady A, Lee RE, Pugacheva EM, Loukinov DI, Goodwin GH, Delgado D, Filippova GN, Leon J, et al. 2001. Functional phosphorylation sites in the C-terminal region of the multivalent multifunctional transcriptional factor CTCF. *Mol Cell Biol* **21**: 2221–2234.
- Lai AY, Fatemi M, Dhasarathy A, Malone C, Sobol SE, Geigerman C, Jaye DL, Mav D, Shah R, Li L, et al. 2010. DNA methylation prevents CTCF-mediated silencing of the oncogene *BCL6* in B cell lymphomas. *J Exp Med* **207**: 1939–1950.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25. doi: 10.1186/gb-2009-10-3-r25.
- Lefevre P, Witham J, Lacroix CE, Cockerill PN, Bonifer C. 2008. The LPS-induced transcriptional upregulation of the chicken lysozyme locus involves CTCF eviction and noncoding RNA transcription. *Mol Cell* **32**: 129–139.
- Li Q, Brown JB, Huang H, Bickel PJ. 2011. Measuring reproducibility of high-throughput experiments. *Ann Appl Stat* **5**: 1752–1779.
- Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schübeler D. 2011. Identification of genetic elements that autonomously determine DNA methylation states. *Nat Genet* **43**: 1091–1097.
- Lin IG, Tomzynski TJ, Ou Q, Hsieh CL. 2000. Modulation of DNA binding protein affinity directly affects target site demethylation. *Mol Cell Biol* **20**: 2343–2349.
- Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo Q-M, et al. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**: 315–322.
- Liu Z, Scannell DR, Eisen MB, Tjian R. 2011. Control of embryonic stem cell lineage commitment by core promoter factor, TAF3. *Cell* **146**: 720–731.
- MacPherson MJ, Beatty LG, Zhou W, Du M, Sadowski PD. 2009. The CTCF insulator protein is posttranslationally modified by SUMO. *Mol Cell Biol* **29**: 714–725.
- Matsuo K, Silke J, Georgiev O, Marti P, Giovannini N, Rungger D. 1998. An embryonic demethylation mechanism involving binding of transcription factors to replicating DNA. *EMBO J* **17**: 1446–1453.
- Matsuzaki H, Okamura E, Fukamizu A, Tanimoto K. 2010. CTCF binding is not the epigenetic mark that establishes post-fertilization methylation imprinting in the transgenic *H19* ICR. *Hum Mol Genet* **19**: 1190–1198.
- Maurano MT, Wang H, Kutayin T, Stamatoyannopoulos JA. 2012. Widespread site-dependent buffering of human regulatory polymorphism. *PLoS Genet* **8**: e1002599. doi: 10.1371/journal.pgen.1002599.
- Mukhopadhyay R, Yu W, Whitehead J, Xu J, Lezcano M, Pack S, Kanduri C, Kanduri M, Ginja V, Vostrov A, et al. 2004. The binding sites for the chromatin insulator protein CTCF map to DNA methylation-free domains genome-wide. *Genome Res* **14**: 1594–1602.
- Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. 2012a. BEDOPS: High performance genomic feature operations. *Bioinformatics* doi: 10.1093/bioinformatics/bts277.
- Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, Sandstrom R, Johnson AK, Maurano MT, et al. 2012b. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* (in press).
- Ohlsson R, Lobanenkov V, Klenova E. 2010. Does CTCF mediate between nuclear organization and gene expression? *Bioessays* **32**: 37–50.
- Pant V, Mariano P, Kanduri C, Mattsson A, Lobanenkov V, Heuchel R, Ohlsson R. 2003. The nucleotides responsible for the direct physical contact between the chromatin insulator protein CTCF and the *H19* imprinting control region manifest parent of origin-specific long-distance insulation and methylation-free domains. *Genes Dev* **17**: 586–590.
- Pant V, Kurukuti S, Pugacheva E, Shamsuddin S, Mariano P, Renkawitz R, Klenova E, Lobanenkov V, Ohlsson R. 2004. Mutation of a single CTCF target site within the *H19* imprinting control region leads to loss of *Igf2* imprinting and complex patterns of *de novo* methylation upon maternal inheritance. *Mol Cell Biol* **24**: 3497–3504.
- Parelho V, Hadjir S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, et al. 2008. Cohesins functionally

- associate with CTCF on mammalian chromosome arms. *Cell* **132**: 422–433.
- Rasko JE, Klenova EM, Leon J, Filippova GN, Loukinov DI, Vatolin S, Robinson AF, Hu YJ, Ulmer J, Ward MD, et al. 2001. Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. *Cancer Res* **61**: 6002–6007.
- Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, Pedone PV. 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.
- Rubio ED, Reiss DJ, Welch PL, Disteché CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A. 2008. CTCF physically links cohesin to chromatin. *Proc Natl Acad Sci* **105**: 8309–8314.
- Schoenherr CJ, Levorse JM, Tilghman SM. 2002. CTCF maintains differential methylation at the *Igf2/H19* locus. *Nat Genet* **33**: 66–69.
- Seitan VC, Hao B, Tachibana-Konwalski K, Lavagnoli T, Mira-Bontenbal H, Brown KE, Teng G, Carroll T, Terry A, Horan K, et al. 2011. A role for cohesin in T-cell-receptor rearrangement and thymocyte differentiation. *Nature* **476**: 467–471.
- Sekimata M, Pérez-Melgosa M, Miller SA, Weinmann AS, Sabo PJ, Sandstrom R, Dorschner MO, Stamatoyannopoulos JA, Wilson CB. 2009. CCCTC-binding factor and the transcription factor T-bet orchestrate T helper 1 cell-specific structure and function at the interferon- γ locus. *Immunity* **31**: 551–564.
- Selker EU. 1990. DNA methylation and chromatin structure: A view from below. *Trends Biochem Sci* **15**: 103–107.
- Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479**: 74–79.
- Soto-Reyes E, 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.
- Splinter E. 2006. CTCF mediates long-range chromatin looping and local histone modification in the β -globin locus. *Genes Dev* **20**: 2349–2354.
- Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Schöler A, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, et al. 2011. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**: 490–495.
- Storey JD, Tibshirani R. 2003. Statistical significance for genomewide studies. *Proc Natl Acad Sci* **100**: 9440–9445.
- Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, Simon I, Yakhini Z, Cedar H. 2009. Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol* **16**: 564–571.
- Suzuki R, Shimodaira H. 2006. Pvcust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**: 1540–1542.
- Tate PH, Bird AP. 1993. Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr Opin Genet Dev* **3**: 226–231.
- Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* (in press).
- Vostrov AA, Quitschke WW. 1997. The zinc finger protein CTCF binds to the APB β domain of the amyloid β -protein precursor promoter. Evidence for a role in transcriptional activation. *J Biol Chem* **272**: 33353–33359.
- Walsh CP, Bestor TH. 1999. Cytosine methylation and mammalian development. *Genes Dev* **13**: 26–34.
- Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishihiro T, et al. 2008. Cohesin mediates transcriptional insulation by CCCTC-binding factor. *Nature* **451**: 796–801.
- Witcher M, Emerson BM. 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.
- Yu W, Ginja V, Pant V, Chernukhin I, Whitehead J, Docquier F, Farrar D, Tavosoidana G, Mukhopadhyay R, Kanduri C, et al. 2004. Poly(ADP-ribosylation) regulates CTCF-dependent chromatin insulation. *Nat Genet* **36**: 1105–1110.
- Yusufzai TM, Tagami H, Nakatani Y, Felsenfeld G. 2004. CTCF tethers an insulator to subnuclear sites, suggesting shared insulator mechanisms across species. *Mol Cell* **13**: 291–298.

Received December 9, 2011; accepted in revised form April 30, 2012.