

# Cross-talk between Site-specific Transcription Factors and DNA Methylation States\*

Published, JBC Papers in Press, October 22, 2013, DOI 10.1074/jbc.R113.512517

Adam Blattler<sup>†§1</sup> and Peggy J. Farnham<sup>‡2</sup>

From the <sup>†</sup>Department of Biochemistry and Molecular Biology, Keck School of Medicine, University of Southern California, Los Angeles, California 90089 and the <sup>‡</sup>Genetics Graduate Group, University of California, Davis, California 95616

DNA methylation, which occurs predominantly at CpG dinucleotides, is a potent epigenetic repressor of transcription. Because DNA methylation is reversible, there is much interest in understanding the mechanisms by which it can be regulated by DNA-binding transcription factors. We discuss several models that, by incorporating sequence motifs, CpG density, and methylation levels, attempt to link the binding of a transcription factor with the acquisition or loss of DNA methylation at promoters and distal regulatory elements. Additional *in vivo* genome-wide characterization of transcription factor binding patterns and high-resolution DNA methylation analyses are clearly required for stronger support of each model.

DNA methylation is a potent epigenetic repressor of transcription; promoters and enhancers that display high levels of methylation are essentially inactive (1, 2). Patterns of DNA methylation are tightly regulated in a tissue-specific manner, resulting in epigenetic specification of gene expression. Additionally, widespread genomic DNA hypomethylation in conjunction with local promoter-specific hypermethylation (which leads to inappropriate gene silencing) is centrally implicated in a myriad of human diseases from immunodeficiency-centromeric instability-facial anomalies syndrome to cancer (3–7). DNA methylation is potentially reversible (8), and there is much interest in understanding the mechanisms by which DNA methylation patterns are established. In this review, we focus on regulation of methylation patterns at mammalian genomic elements such as promoters and enhancers; for a summary of the regulation of large scale changes in methylation during development, we suggest a recent review by Smith and Meissner (9). It has been proposed that site-specific regulation of DNA methylation is mediated by DNA-binding transcription factors (TFs)<sup>3</sup> that interact with specific promoter or

enhancer regions (10–14). The functional relationship between DNA methylation and TF binding has been a subject of much interest for decades. Several models have been put forth that attempt to integrate the ability of a TF to bind to hyper- or hypo-methylated DNA with the acquisition and/or loss of DNA methylation at regulatory elements (Fig. 1). These include (a) protection from acquisition of DNA methylation upon binding of a TF to unmethylated DNA, (b) promotion of DNA methylation upon binding of a TF to unmethylated DNA, (c) reversal of DNA methylation upon binding of a TF to a region containing methylated DNA, and (d) reinforcement of repression upon binding of a TF to methylated DNA.

In mammalian genomes, DNA methylation occurs predominantly at CpG dinucleotides. Regions of the genome are generally classified as those that have high CpG density (*e.g.* promoters) *versus* low CpG density (most of the genome). Additionally, the genome can be divided into categories based on CpG methylation levels: low methylation (promoters), intermediate methylation (distal regulatory elements), and high CpG methylation (most of the genome) (15). Both the CpG density and the methylation status of a region can have important implications with respect to the influence of DNA methylation on TF binding. For example, the interaction of TFs that contain a CpG dinucleotide in their recognition motif with DNA will be influenced by methylation status in regions having both high and low CpG density, whereas TFs that do not have a CpG in their motif may interact differently with genomic regions that have sparse or dense methylation levels. In addition, some factors specifically recognize methylated *versus* unmethylated CpGs in the absence of an extended DNA motif. These issues, DNA recognition sequence, CpG density, and methylation levels, are considered in each of the models described below.

## Model A: Protection from Acquisition of Methylation upon Transcription Factor Binding to Unmethylated DNA

Approximately 70% of promoters in the human genome are classified as CpG island promoters, having a high density of CpG dinucleotides. However, although CpG dinucleotides are the substrate for DNA methylation in mammalian genomes, they are largely protected from methylation events at CpG island promoters, which are characterized by active, unmethylated chromatin in most cell types. It has been proposed that CpG island promoters can be protected from acquisition of DNA methylation by binding of TFs. For example, proteins having a CXXC zinc finger domain have been identified that bind to unmethylated CpGs in a sequence-independent manner. The first factor discovered with these properties was CXXC1,<sup>4</sup> which binds to clustered, unmethylated CpGs (16–18). CXXC1 (which is also sometimes called CFP1) can recruit

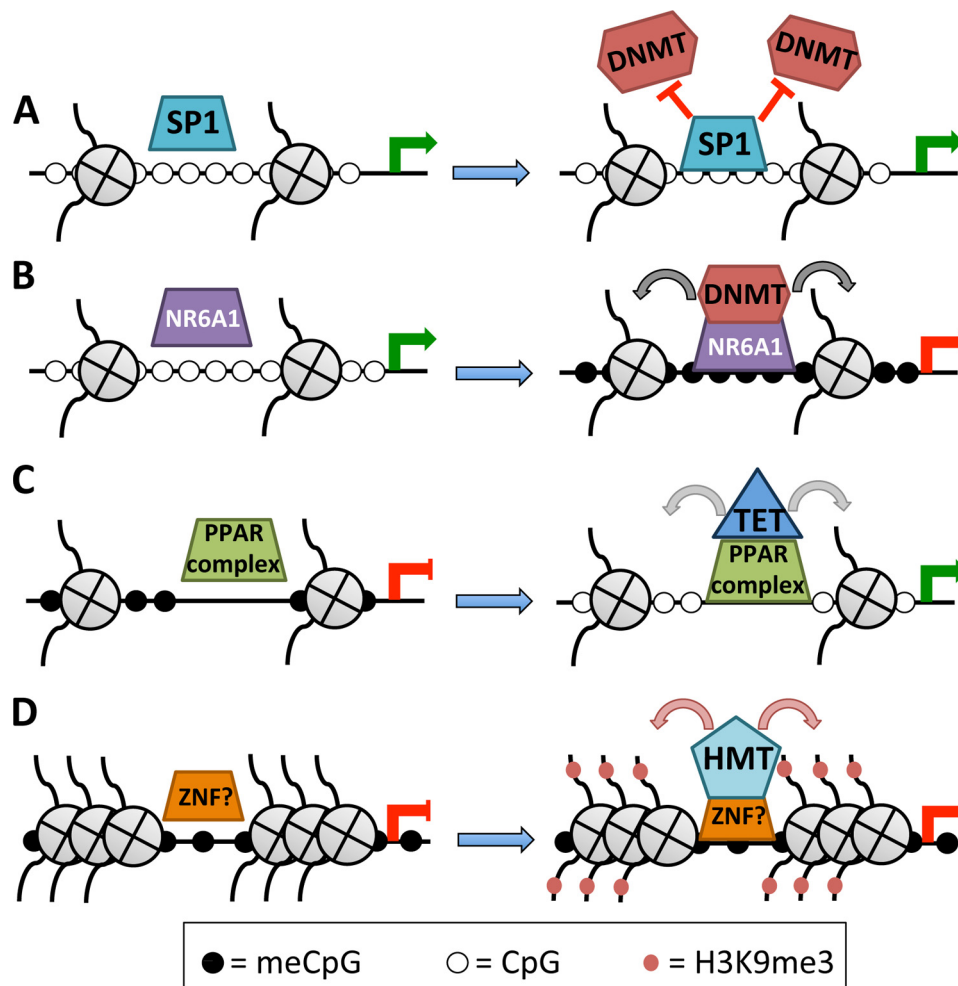
\* This work was supported, in whole or in part, by National Institutes of Health Grant HG006996 through the United States Public Health Service.

<sup>1</sup> Supported in part by National Institutes of Health Predoctoral Training Fellowship F3100HG6114.

<sup>2</sup> To whom correspondence should be addressed: Norris Comprehensive Cancer Center, 1450 Biggy St., NRT 6503, University of Southern California, Los Angeles, CA 90089-9601. Tel.: 323-442-8015; E-mail: peggy.farnham@med.usc.edu.

<sup>3</sup> The abbreviations used are: TF, transcription factor; MBD, methyl CpG-binding domain; DNMT, DNA methyl transferase; TET, Ten-eleven-translocation; 5hmC, 5-hydroxymethylcytosine; 5mC, 5-methylcytosine; ZNF, zinc finger; KRAB, Krüppel-associated box; ChIP-seq, ChIP sequencing.

<sup>4</sup> The official HUGO gene symbol has been used to refer to all transcription factors and other genes; the full name of each gene can be found on the HUGO Gene Nomenclature Committee web site.



**FIGURE 1. The effect of site-specific factors on DNA methylation.** A, site-specific factors, such as SP1, bind unmethylated DNA, preventing DNA methyltransferases from accessing the promoter or enhancer. B, site-specific factors, such as NR6A1, can recruit DNA methyltransferases to unmethylated DNA, resulting in CpG methylation (meCpG). C, site-specific factors such as PPARG, which do not have a CpG in their motifs, can bind between methylated CpGs within a methylated region, recruiting TET proteins that oxidize 5-methylcytosine and reverse DNA methylation. D, site-specific factors, perhaps ZNF proteins, bind methylated motifs, recruiting histone methyltransferases and other repressive machinery.

a histone H3 lysine 4 (H3K4) methyltransferase and protect bound regions from DNA methylation (18). Other proteins that have a CXXC domain, such as mixed lineage leukemia family members, can also recognize unmethylated CpGs (17). Interestingly, the methyl CpG-binding domain (MBD) protein MBD1 (see Model D) has a CXXC domain that is responsible for low levels of methylation-independent DNA binding, suggesting that it may have multiple roles in genome regulation (19). DNA methyltransferase (DNMT) 1 is also a CXXC domain-containing protein. Binding to unmethylated CpGs may seem contradictory for a factor whose role is to maintain DNA methylation through DNA replication. However, the targeting specificity of DNMT1 is largely dictated by its interaction with UHRF1, which recognizes hemi-methylated DNA. Studies suggest that when DNMT1 binds to unmethylated DNA via the CXXC domain, it is catalytically inactive (20). Therefore, the CXXC domain may help to limit the activity of DNMT1 to hemi-methylated, as opposed to unmethylated, DNA. CXXC domain-containing proteins require access to both the major and the minor grooves and therefore must either bind to nucleosome-free regions or bind to the linker

DNA between nucleosomes because the physical association of DNA with histone octamers can prevent simultaneous access to the major and minor groove (21). This suggests that CXXC domain-containing proteins may help to block *de novo* DNA methylation at unmethylated, nucleosome-free regions. Reversal of new methylation events at an unmethylated region may be another mechanism by which CXXC proteins function. Ten-eleven-translocation (TET)1 and TET3 are CXXC proteins that can generate 5-hydroxymethylcytosine (5hmC) by oxidation of 5-methylcytosine (5mC); TET proteins can further oxidize 5hmC to 5-formylcytosine and 5-carboxylcytosine, but no enzyme has been found that can completely convert 5mC to C (22). TET1 has been shown to localize to unmethylated promoters, suggesting that it may reverse inappropriate *de novo* methylation at a nearby CpG by converting newly formed 5mC to 5hmC. Because 5hmC is not a good binding substrate for MBD proteins, this could help to keep promoter regions unmethylated. However, due to lack of sequence specificity and the need to access both the major and the minor grooves, it is likely that interaction with other DNA-binding factors is required for TET proteins to access highly methylated regions (see Model C).

Certain site-specific DNA-binding TFs have been shown to bind to unmethylated regulatory regions and protect them from methylation by blocking *de novo* methylation events (Fig. 1A). One such factor is SP1, which binds the consensus motif CCGCCC, a sequence that is overrepresented in CpG islands (23–25). Lienert *et al.* (26) showed that mutating SP1-binding sites within the *Gtf2a11* promoter resulted in increased DNA methylation of the region and silencing of the *Gtf2a11* gene. CTCF has also been shown to block methylation of a bound region (27, 28). CTCF binds the *Igf2*/H19 imprinting control region and acts as a boundary element for the control of imprinted expression of maternal and paternal copies of *Igf2* and *H19* genes (29, 30). Schoenherr *et al.* (27) showed that introducing mutations at the four CTCF-binding sites within the imprinting control region resulted in loss of CTCF binding and a substantial increase in methylation. Thus, binding of factors such as SP1 and CTCF can protect regulatory regions from gaining methylation. Binding of transcription factors to promoter regions can also influence DNA methylation by causing active transcription. Active transcription of GC-rich promoter regions results in the formation of DNA-RNA hybrids (R-loops), which protect the promoter-proximal transcribed region from the action of DNMTs (31).

#### Model B: Promotion of Methylation upon Transcription Factor Binding to Unmethylated DNA

Several DNA-binding factors have been shown to associate with DNMTs and promote the methylation of an unmethylated genomic region (32, 33). For example, *in vitro* experiments by Brenner *et al.* (33) showed that MYC can interact with DNMT3B, forming a ternary complex with MIZ1, which can bind to the *CDKN1A* promoter, and Velasco *et al.* (34) showed that E2F6 can recruit DNMT3B to a set of promoters, resulting in their methylation and subsequent repression. Similarly, Suzuki *et al.* (35) demonstrated an interaction of DNMT3B with SPI1 (commonly known as PU.1), and ChIP experiments confirmed that DNMT3A and DNMT3B bind a reporter promoter only in the presence of SPI1; repression of the reporter gene occurred when SPI1 was co-transfected with DNMT3A or DNMT3B. Experiments by de la Rica *et al.* (36) confirm that SPI1 interacts with DNMT3B and show that the factors co-localize at promoters that gain DNA methylation during osteoclastogenesis. Sato *et al.* (37) showed that NR6A1 (commonly known as germ cell nuclear factor (GCNF)) interacts with DNMT3A and DNMT3B and that co-expression of NR6A1 with DNMT3A is sufficient to drive DNA methylation of an engineered promoter containing an NR6A1-binding site. In addition to site-specific factors that can directly interact with DNMTs, there are several examples of indirect recruitment of DNMTs to the genome by TFs involved in the deposition of repressive histone modifications. Recent studies have demonstrated a spatial relationship between DNA methylation and the trimethylation of histone H3 lysine 9 or 27 (H3K9me3, H3K27me3) (38, 39). The H3K9 methyltransferase SETDB1 is a known component of heterochromatin maintenance machinery and works in conjunction with MBD1, HP1, and histone deacetylase proteins to silence gene expression (40–42). Li (43) showed that the N-terminal domain of SETDB1 can directly

interact with both DNMT3A and DNMT3B and that SETDB1 and DNMT3A co-localize at the *RASSF1* promoter. Although SETDB1 is not a DNA-binding protein, it does associate with the TRIM28 (commonly known as KAP1) repression complex (44) and is recruited along with TRIM28 to specific genomic sites by KRAB domain-containing zinc finger proteins (KRAB-ZNFs) such as ZNF274 (44–47). There are over 300 KRAB-ZNFs, representing the largest class of TFs encoded in the human genome. Cell type-specific expression of KRAB-ZNFs could provide a mechanism by which DNMTs are targeted to promoters in a tissue-specific manner. However, very few KRAB-ZNFs have been functionally characterized due to their relatively low expression in most tissue types. Therefore, the importance of these factors in specifying DNA methylation patterns remains to be seen. Finally, Viré *et al.* (48) showed that the H3K27 methyltransferase EZH2, a component of the polycomb repressive complex PRC2, can interact with DNMT1, DNMT3A, and DNMT3B. EZH2 does not directly interact with DNA, but rather is recruited to its binding sites by factors such as JARID2 (49) and by long noncoding RNAs (50). EZH2 binding to the *MYT1* promoter is required for the recruitment of DNMTs, suggesting that EZH2 may be involved in establishing DNA methylation at promoter regions (48). However, other studies have shown that DNA methylation is retained at promoters upon the depletion of EZH2 (51). Although it is possible that EZH2 plays a role in the recruitment of DNMTs at some genomic loci, more information is necessary to elucidate its role in regulating DNA methylation. In summary, although recruitment of a DNMT to an unmethylated regulatory element by interaction with a site-specific DNA-binding factor is an attractive model to explain methylation of specific promoters and enhancers, most of the studies to date have focused on a single promoter. ChIP-seq experiments have not yet identified a TF whose genome-wide binding sites show a large degree of overlap with the binding of a DNMT.

#### Model C: Reversal of Methylation upon Transcription Factor Binding to Methylated DNA

A small number of studies have shown that TFs can bind to a methylated region and mediate the reversal of DNA methylation. Stadler *et al.* (15) used homologous recombination to insert an *in vitro* methylated DNA fragment into mouse cells and measured its methylation at a later time point. They found that a fragment containing a wild-type, but not mutated, CTCF motif showed loss of methylation. This seems to be contradictory to *in vitro* studies showing that methylation of the CTCF motif abrogates binding (52). However, the DNA region used by Stadler *et al.* (15) had low CpG density and only partial methylation; perhaps in that study, CTCF binding occurred between methylated CpGs. In a similar experiment, genomic regions corresponding to REST-binding sites were analyzed in wild-type and REST knock-out embryonic stem cells. Regions surrounding the REST-binding sites were unmethylated in wild-type cells, but those same locations were methylated in the knock-out cells. Rescue of REST expression in the knock-out cells resulted in loss of methylation at the binding sites, suggesting that REST could bind to the methylated region and mediate a loss of methylation. Again, this study analyzed low CpG den-



sity and partially methylated genomic regions, suggesting that REST may have bound to an unmethylated motif within a partially methylated domain (15). In fact, the REST consensus-binding motif does not contain a highly conserved CpG, indicating that DNA methylation likely does not play a role in its physical interaction with DNA, allowing the protein to bind in regions containing methylated CpGs (15). The mechanism by which the binding of factors such as CTCF and REST could lead to demethylation of once methylated regions is not yet clear. A possible mechanism could be via interaction with TET proteins. However, a direct interaction of CTCF or REST with a TET protein has not yet been demonstrated. Recent studies by de la Rica *et al.* (36) have shown that TET2 can interact with SPI1 and that both factors bind promoters that lose DNA methylation during osteoclastogenesis. This link between SPI1 and low DNA methylation levels is in agreement with data showing that SPI1-binding sites are unmethylated in acute myelogenous leukemia cells that highly express SPI1.<sup>5</sup> As noted above, SPI1 can also interact with DNMT3B. Thus, SPI1 may play multiple roles in regulating DNA methylation. The DNA-binding motifs for the activator protein 1 (AP-1) complex and NFκB1 were also found to be enriched within regions that become demethylated during osteoclastogenesis, suggesting that these proteins may also play a role in TET protein recruitment. Another recent study has identified a TF that can indirectly recruit TET proteins to a methylated enhancer region (53). PPARG is a nuclear hormone receptor that interacts with co-activators to regulate the expression of adipocyte-specific genes. The PPARG co-activator complex is poly(ADPribose)ylated when recruited to the DNA. The TET proteins can bind to the poly(ADPribose)ylated complex and catalyze the conversion of 5mC to 5hmC, thereby inducing region-specific demethylation (53). The bZIP protein CEBPA has been shown to bind methylated DNA *in vivo* and *in vitro*, and Rishi *et al.* (54) found that 15–25% of CEBPA-binding sites are methylated in keratinocytes and adipocytes. Binding of CEBPA resulted in enhanced expression of target promoters. However, contrary to Model C, the CEBPA-bound promoters, which are enriched for the sequence TGACGTCA, remained methylated. It is not clear how the promoters of CEBPA-bound genes are expressed when methylated, but perhaps these regions have low CpG density and therefore cannot effectively recruit repressive machinery. Another possibility is that at a subset of CEBPA-bound promoters, one allele is unmethylated and expressed whereas the other allele is methylated and repressed. This phenomenon has been observed for a small number of human promoters in a recent study (55).

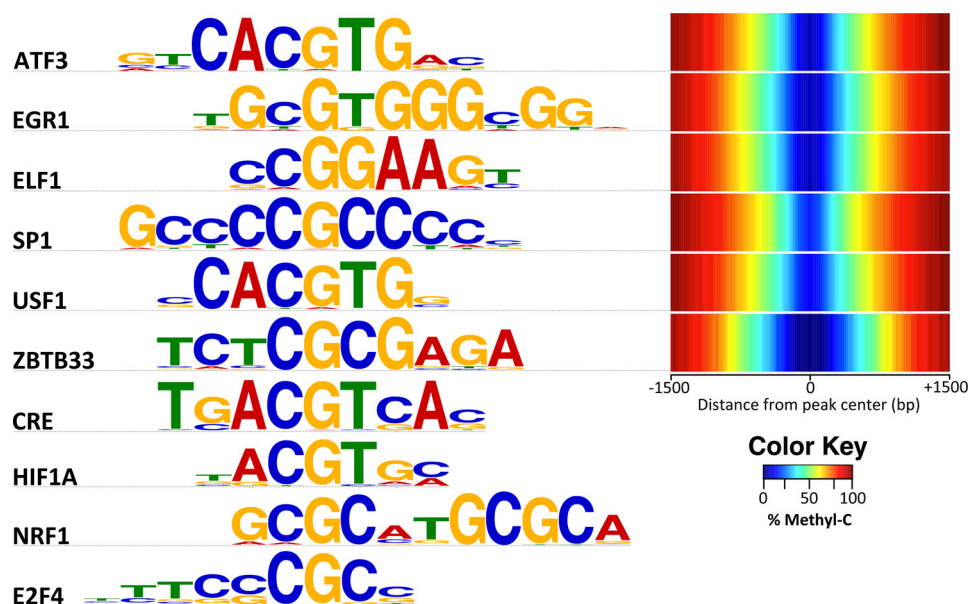
#### **Model D: Reinforcement of Repression upon Transcription Factor Binding to Methylated Binding Motifs**

In Model C, TFs bind to methylated regions but *not* to methylated CpGs. However, proteins have been identified that bind methylated CpGs with varying degrees of specificity, with or without use of an extended recognition motif (56, 57). Although it is unlikely that factors that bind to methylated CpGs in the

absence of an extended motif are involved in site-specific regulation of DNA methylation, they may be important for maintaining global patterns of methylated and unmethylated domains. Members of the MBD family, including MBD1, MBD2, MBD4, and MECP2, contain specific domains responsible for recognition of methylated sequences (58–60). MBD3, another member of the MBD family, does not have the same affinity for methylated DNA. A study using biotin-tagged MBD proteins demonstrated that this family binds methylated DNA *in vivo* with the highest affinity in regions with high concentrations of methylated CpGs, whereas methylated regions with low CpG content dictate lower levels of MBD binding (19). This affinity for binding to regions densely populated with methylated CpGs, along with the fact that MBDs can recruit repressive histone modifying complexes to their binding sites (61–63), could reinforce silencing of methylated CpG island promoters, perhaps after methylation has been initiated by the binding of site-specific factors.

Some site-specific TFs recognize a sequence that contains a CpG within an extended DNA recognition motif. Because methylation of the cytosine causes a major structural change in the nucleotide, it is likely that DNA-protein interactions will be influenced (either positively or negatively) by methylation of the CpG. An analysis of the JASPAR motif database revealed that 25% of all characterized motifs contain a CpG within the recognition sequence. Of course, not all CpG-containing motifs have the CpG dinucleotide sequence at a critical position that would influence DNA-protein interactions. However, there are a few motifs for which the CpG is located in a critical position (Fig. 2). One would predict that changes in the methylation state of these motifs would impact protein binding. To date, the influence of DNA methylation on protein-DNA interactions has mainly been investigated *in vitro* using binding methods such as gel mobility shift assays and structural methods such as x-ray crystallography. Using such techniques, proteins have been characterized as belonging to three classes: those that prefer to bind to unmethylated DNA (27, 64), those that prefer to bind to methylated DNA (12, 14), and those that are agnostic to the presence of DNA methylation (65, 66). For example, Campanero *et al.* (67) showed that CpG methylation differentially regulates the response of certain E2F elements to different E2F family members. The E2F consensus motif can contain two CpGs (TTTSSCGC can be TTTCGCGC); when unmethylated, these are the strongest E2F recognition motifs. However, motifs having two CpGs cannot be bound by E2F1–5 when the sites are methylated. In contrast, methylation of an E2F-binding site that contained only one CpG did not affect binding of E2F2–5, but abrogated E2F1 binding (67). The helix-loop-helix DNA-binding proteins MYC, USF1, and TFE3 can all bind to a CACGTG motif. Methylation of the central CpG strongly affected MYC, but not USF1 or TFE3, binding *in vitro* (65). bZIP proteins bind to palindromic CRE motifs that have a central CG dinucleotide; *in vitro* studies show that methylation enhances CEBPA and CEBPB binding but inhibits binding of CREB1, ATF4, JUN, JUND, CEBPD, and CEBPG (11, 54, 68). Harrington *et al.* (66) found that methylation of the CpG in the motif CCGCCC did not affect DNA binding by SP1, whereas other studies have shown that methylation of the first C of the

<sup>5</sup> B. P. Berman, personal communication.



**FIGURE 2. Site-specific transcription factors containing CpG dinucleotides within their recognition motif.** An analysis of HOMER v4.3 and FactorBook motif databases (78, 79) identified a small number of motifs having at least one CpG dinucleotide at a critical position. These included motifs bound by members of the ATF family (ATF1, ATF3), EGR1, members of the ETS family (ETS1, ELF1, ELK4, GABPA), SP1, the MYC family (MYC, MAX, NMYC, USF1, BHLHE40), ZBTB33, CRE-binding factors, HIF1A, NRF1, and members of the E2F family. Shown are examples of the motif for a member of each family that has a critical CpG in its recognition sequence. ChIP-seq data for ATF3, EGR1, ELF1, SP1, USF1, and ZBTB33 produced by the ENCODE consortium (77) was compared with whole genome bisulfite sequencing data (A. Blattler and P. J. Farnham, unpublished data); all data are from HCT116 colorectal cancer cells. On the right, the degree of DNA methylation of a region  $\pm 1500$  from the center of each ChIP-seq peak is plotted for those TFs. In all cases, DNA methylation is absent from the center of the TF-binding sites. To determine the *in vivo* relationship between TF binding and DNA methylation, experiments such as this must be performed comparing ChIP-seq data with whole genome DNA methylation data in matched cell types.

motif in combination with the C of the CpG abrogates SP1 binding (64).

Whether methylation of a CpG within a motif has the same effect on DNA binding *in vivo* has not yet been investigated for most TFs. However, several TFs with CpG-containing motifs have been studied by ChIP-seq in the same cells for which genome-wide DNA methylation patterns are known. ZBTB33 (commonly known as Kaiso), ZBTB38, and ZBTB4 are highly related proteins that have three tandem C2H2 zinc finger (ZNF) domains responsible for methyl-DNA recognition *in vitro*. A recent study showed that ZBTB33 binds unmethylated promoters *in vivo* (69), indicating that *in vitro* analyses may not reflect the binding properties of proteins in the context of a chromatin environment. Methylation of CpG-dense regions creates a highly condensed heterochromatin structure that can prevent factors from accessing a methylated motif. Thus, although a factor such as ZBTB33 may prefer binding to a methylated motif, it is not found at methylated promoters *in vivo*. A similar analysis of the binding patterns of Kaiso and five other CpG motif-binding factors reveals an absence of DNA methylation at the center of the binding sites, suggesting that in general methylation is inhibitory to *in vivo* binding of these TFs (Fig. 2).

Several proteins, such as KLF4, ZFP57, and CEBPB, have been implicated in binding to methylated DNA *in vivo*. Using ChIP-seq and whole genome bisulfite sequencing data, Spruijt *et al.* (71) showed that  $\sim 18\%$  of the genomic regions bound by KLF4, one of the four pluripotency factors identified by Takahashi and Yamanaka (70), are highly methylated in mouse embryonic stem cells. However, when the analysis was restricted to the binding motif, the levels of DNA methylation

dropped significantly. Using a protein microarray, Hu *et al.* (57) also identified KLF4 in a set of 47 proteins (approximately half of which were ZNFs) that can bind to methylated DNA *in vitro*. Interestingly, the methylated sequence bound by KLF4 was different from the preferred unmethylated motif. The authors tested a small number of genomic sites and showed that KLF4 could bind to methylated sequences *in vivo* using ChIP followed by bisulfite sequencing (57). However, the functional significance of these binding sites is not yet clear. Quenneville *et al.* (12) showed that the zinc finger protein ZFP57 can bind to a methylated TGCCGC motif *in vitro* and used bisulfite treatment of ChIP DNA to show that ZFP57 can bind to the methylated allele of three imprinted mouse genes. However, the DNA methylation status of the other 11,000 ZFP57 ChIP-seq binding sites was not analyzed. CEBPB is a bZIP protein that can bind to the 8-mer TTGCGCAA, and it has been shown that methylation enhances *in vitro* binding to this motif (11). However, this motif is not a preferred binding motif *in vivo*, and the few sites that do contain this motif are not in the top ranked peak list as determined by ChIP-seq. The authors (12) suggest that perhaps, similar to ZBTB33 (69), access of CEBPB to highly methylated regions is prevented at condensed, heterochromatic regions. However, the study did identify  $\sim 200$  places in the genome where CEBPB bound that had greater than 50% methylation frequency. It is likely that CpG density, percentage of methylation, and nucleosome density of a genomic region greatly affect the ability of factors such as ZBTB33, KLF4, ZFP57, or CEBPB to access a methylated motif in the genome (69, 72, 73).

As described above, to date there are very few documented cases of a site-specific factor binding robustly to a methylated

motif within a CpG-dense, highly methylated region. However, Liu *et al.* (74) made the observation that both ZBTB33 and ZFP57 contain zinc finger domains with a conserved arginine preceding the first zinc-binding histidine residue (termed the RH motif) and that this arginine interacts with the methyl group of the 5mC. They postulate that this RH motif may be a feature of zinc finger proteins that bind methylated DNA. If this is true, 224 of the 330 human KRAB ZNFs may have the ability to recognize methylated DNA, at least *in vitro* (74). Importantly, KRAB-ZNFs interact with repressive transcription complexes. Perhaps a subset of these factors will have the ability to bind to methylated DNA and recruit DNMTs or repressive histone-modifying complexes to the genome.

## CONCLUSIONS

The models described above present four mechanisms by which binding of TFs might promote or inhibit DNA methylation and influence transcription. Unfortunately, it is not yet possible to predict whether a TF will promote or inhibit methylation when bound to a specific regulatory element. TFs have multiple protein interaction domains and can interact with both co-activators and co-repressors; for example, SPI1 can interact with both DNMT3B and TET2 (35, 36). It is likely that the effect a TF has on DNA methylation at a given regulatory element will be influenced by other proteins recruited to that site. Therefore, it is extremely important that the relationship between TFs and DNA methylation be examined in a relevant physiological context. Much of the experimental evidence in support of the four models has been collected using reporter assays or single endogenous elements as a model system. Fortunately, with the advent of new technologies that allow investigation of the genome-wide *in vivo* binding patterns of TFs (75–77), along with comprehensive gene expression and DNA methylation analyses, it is now possible to investigate the relationship between DNA methylation, TF binding, and gene expression on a global scale in a variety of cell types under diverse physiological and developmental conditions. Future studies that intersect TF-binding sites with binding sites of DNMTs or repressive histone-modifying complexes may identify factors that help establish or reinforce DNA methylation. Conversely, intersection of TF-binding sites with sites bound by the TET proteins may identify site-specific factors that are important in blocking or reversing DNA methylation. The recent introduction of experiments that combine traditional chromatin immunoprecipitation with bisulfite treatment and sequencing of the ChIP DNA may also shed light on the methylation state of TF-bound DNA (38, 39). In conclusion, we hope that the models described within this review will provide a useful framework with which to interpret the expanding amounts of genome-scale data and contribute to a more complete understanding of the transcriptional dysregulation that results in a wide array of human diseases.

*Acknowledgments*—We thank Seth Fietze and Kwangho Lee for critical reading of the manuscript, Dennis Hazelett for help with the motif analyses, and Lijing Yao for assistance with DNA methylation analyses.

## REFERENCES

1. Jones, P. A. (2012) Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492
2. Deaton, A. M., and Bird, A. (2011) CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022
3. Pawlak, M., and Jaenisch, R. (2011) *De novo* DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state. *Genes Dev.* **25**, 1035–1040
4. Moss, T. J., and Wallrath, L. L. (2007) Connections between epigenetic gene silencing and human disease. *Mutat. Res.* **618**, 163–174
5. Okano, M., Xie, S., and Li, E. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* **19**, 219–220
6. Hansen, R. S., Wijmenga, C., Luo, P., Stanek, A. M., Canfield, T. K., Weemaes, C. M., and Gartler, S. M. (1999) The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 14412–14417
7. Jin, B., Tao, Q., Peng, J., Soo, H. M., Wu, W., Ying, J., Fields, C. R., Delmas, A. L., Liu, X., Qiu, J., and Robertson, K. D. (2008) DNA methyltransferase 3B (DNMT3B) mutations in ICF syndrome lead to altered epigenetic modifications and aberrant expression of genes regulating development, neurogenesis and immune function. *Hum. Mol. Genet.* **17**, 690–709
8. Franchini, D.-M., Schmitz, K.-M., and Petersen-Mahrt, S. K. (2012) 5-Methylcytosine DNA DEMETHYLATION: More than losing a methyl group. *Annu. Rev. Genet.* **46**, 419–441
9. Smith, Z. D., and Meissner, A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220
10. Prokhorchouk, A., Hendrich, B., Jørgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhorchouk, E. (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* **15**, 1613–1618
11. Mann, I. K., Chatterjee, R., Zhao, J., He, X., Weirauch, M. T., Hughes, T. R., and Vinson, C. (2013) CG methylated microarrays identify a novel methylated sequence bound by the CEBPB ATF4 heterodimer that is active *in vivo*. *Genome Res.* **23**, 988–997
12. Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, L., Pedone, P. V., Grimaldi, G., Riccio, A., and Trono, D. (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* **44**, 361–372
13. Bartels, S. J. J., Spruijt, C. G., Brinkman, A. B., Jansen, P. W. T. C., Vermeulen, M., and Stunnenberg, H. G. (2011) A SILAC-based screen for methyl-CpG binding proteins identifies RBP-J as a DNA methylation and sequence-specific binding protein. *PLoS One* **6**, e25884
14. Filion, G. J. P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhorchouk, E., and Defossez, P.-A. (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol. Cell Biol.* **26**, 169–181
15. Stadler, M. B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E. J., Gaidatzis, D., Tiwari, V. K., and Schübeler, D. (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495
16. Xu, C., Bian, C., Lam, R., Dong, A., and Min, J. (2011) The structural basis for selective binding of non-methylated CpG islands by the CFP1 CXXC domain. *Nat. Commun.* **2**, 227–228
17. Long, H. K., Blackledge, N. P., and Klose, R. J. (2013) ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Trans.* **41**, 727–740
18. 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., Turner, D. J., Illingworth, R., and Bird, A. (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. *Nature* **464**, 1082–1086
19. 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
20. Song, J., Rechko, O., Bestor, T. H., and Patel, D. J. (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036–1040



21. Zhou, J. C., Blackledge, N. P., Farcas, A. M., and Klose, R. J. (2012) Recognition of CpG island chromatin by KDM2A requires direct and specific interaction with linker DNA. *Mol. Cell Biol.* **32**, 479–489
22. Pastor, W. A., Aravind, L., and Rao, A. (2013) TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**, 341–356
23. 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
24. 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
25. Kriwacki, R. W., Schultz, S. C., Steitz, T. A., and Caradonna, J. P. (1992) Sequence-specific recognition of DNA by zinc-finger peptides derived from the transcription factor Sp1. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 9759–9763
26. 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
27. Schoenherr, C. J., Levorse, J. M., and Tilghman, S. M. (2003) CTCF maintains differential methylation at the *Igf2/H19* locus. *Nat. Genet.* **33**, 66–69
28. Wang, H., Maurano, M. T., Qu, H., Varley, K. E., Gertz, J., Pauli, F., Lee, K., Canfield, T., Weaver, M., Sandstrom, R., Thurman, R. E., Kaul, R., Myers, R. M., and Stamatoyannopoulos, J. A. (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome Res.* **22**, 1680–1688
29. Lee, D.-H., Singh, P., Tsai, S. Y., Oates, N., Spalla, A., Spalla, C., Brown, L., Rivas, G., Larson, G., Rauch, T. A., Pfeifer, G. P., and Szabó, P. E. (2010) CTCF-dependent chromatin bias constitutes transient epigenetic memory of the mother at the *H19-Igf2* imprinting control region in prospermatogonia. *PLoS Genet.* **6**, e1001224
30. 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
31. Ginno, P. A., Lott, P. L., Christensen, H. C., Korf, I., and Chédin, F. (2012) R-loop formation is a distinctive characteristic of unmethylated human CpG island promoters. *Mol. Cell* **45**, 814–825
32. Di Croce, L. (2002) Methyltransferase Recruitment and DNA Hypermethylation of target promoters by an oncogenic transcription factor. *Science* **295**, 1079–1082
33. Brenner, C., Deplus, R., Didelot, C., Lorient, A., Viré, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., Pelicci, P. G., Amati, B., Kouzarides, T., de Launoit, Y., Di Croce, L., and Fuks, F. (2005) Myc represses transcription through recruitment of DNA methyltransferase corepressor. *EMBO J.* **24**, 336–346
34. Velasco, G., Hubé, F., Rollin, J., Neuillet, D., Philippe, C., Bouzinba-Segard, H., Galvani, A., Viegas-Péquignot, E., and Francastel, C. (2010) Dnmt3b recruitment through E2F6 transcriptional repressor mediates germ-line gene silencing in murine somatic tissues. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9281–9286
35. Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T., Hara, E., Tenen, D. G., Hozumi, N., and Oikawa, T. (2006) Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. *Oncogene* **25**, 2477–2488
36. de la Rica, L., Rodríguez-Ubrea, J., García, M., Islam, A. B., Urquiza, J. M., Hernando, H., Christensen, J., Helin, K., Gómez-Vaquero, C., and Ball-est, E. (2013) PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. *Genome Biol.* **14**, R99
37. Sato, N., Kondo, M., and Arai, K.-I. (2006) The orphan nuclear receptor GCNRF recruits DNA methyltransferase for *Oct-3/4* silencing. *Biochem. Biophys. Res. Commun.* **344**, 845–851
38. Brinkman, A. B., Gu, H., Bartels, S. J. J., Zhang, Y., Matarese, F., Simmer, F., Marks, H., Bock, C., Gnirke, A., Meissner, A., and Stunnenberg, H. G. (2012) Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138
39. Statham, A. L., Robinson, M. D., Song, J. Z., Coolen, M. W., Stirzaker, C., and Clark, S. J. (2012) Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. *Genome Res.* **22**, 1120–1127
40. Verschuer, P. J., van der Kraan, I., de Leeuw, W., van der Vlag, J., Carpenter, A. E., Belmont, A. S., and van Driel, R. (2005) *In vivo* HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Mol. Cell Biol.* **25**, 4552–4564
41. Sarraf, S. A., and Stancheva, I. (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* **15**, 595–605
42. Fuks, F., Burgers, W. A., Godin, N., Kasai, M., and Kouzarides, T. (2001) Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J.* **20**, 2536–2544
43. Li, H. (2006) The histone methyltransferase SETDB1 and the DNA methyltransferase DNMT3A interact directly and localize to promoters silenced in cancer cells. *J. Biol. Chem.* **281**, 19489–19500
44. Schultz, D. C. (2002) SETDB1: a novel KAP-1-associated histone H3, lysine 9-specific methyltransferase that contributes to HP1-mediated silencing of euchromatic genes by KRAB zinc-finger proteins. *Genes Dev.* **16**, 919–932
45. Agata, Y., Matsuda, E., and Shimizu, A. (1999) Two novel Krüppel-associated box-containing zinc-finger proteins, KRAZ1 and KRAZ2, repress transcription through functional interaction with the corepressor KAP-1 (TIF1 $\beta$ /KRIP-1). *J. Biol. Chem.* **274**, 16412–16422
46. O'Geen, H., Squazzo, S. L., Iyengar, S., Blahnik, K., Rinn, J. L., Chang, H. Y., Green, R., and Farnham, P. J. (2007) Genome-wide analysis of KAP1 binding suggests autoregulation of KRAB-ZNFs. *PLoS Genet.* **3**, e89
47. Fietze, S., O'Geen, H., Blahnik, K. R., Jin, V. X., and Farnham, P. J. (2010) ZNF274 recruits the histone methyltransferase SETDB1 to the 3' ends of ZNF genes. *PLoS One* **5**, e15082
48. Viré, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morrey, L., Van Eynde, A., Bernard, D., Vanderwinden, J.-M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y., and Fuks, F. (2006) The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439**, 871–874
49. Pasini, D., Cloos, P. A. C., Walfridsson, J., Olsson, L., Bukowski, J.-P., Johansen, J. V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* **464**, 306–310
50. Margueron, R., and Reinberg, D. (2011) The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349
51. McGarvey, K. M., Greene, E., Fahrner, J. A., Jenuwein, T., and Baylin, S. B. (2007) DNA methylation and complete transcriptional silencing of cancer genes persist after depletion of EZH2. *Cancer Res.* **67**, 5097–5102
52. Gao, J., Wang, J., Wang, Y., Dai, W., and Lu, L. (2011) Regulation of *Pax6* by CTCF during induction of mouse ES cell differentiation. *PLoS One* **6**, e20954
53. Fujiki, K., Shinoda, A., Kano, F., Sato, R., Shirahige, K., and Murata, M. (2013) PPAR $\gamma$ -induced PARylation promotes local DNA demethylation by production of 5-hydroxymethylcytosine. *Nat. Commun.* **4**, 2262
54. Rishi, V., Bhattacharya, P., Chatterjee, R., Rozenberg, J., Zhao, J., Glass, K., Fitzgerald, P., and Vinson, C. (2010) CpG methylation of half-CRE sequences creates C/EBP $\alpha$  binding sites that activate some tissue-specific genes. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20311–20316
55. Kelly, T. K., Liu, Y., Lay, F. D., Liang, G., Berman, B. P., and Jones, P. A. (2012) Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. *Genome Res.* **22**, 2497–2506
56. Sasai, N., and Defossez, P.-A. (2009) Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *Int. J. Dev. Biol.* **53**, 323–334
57. Hu, S., Wan, J., Su, Y., Song, Q., Zeng, Y., Nguyen, H. N., Shin, J., Cox, E., Rho, H. S., Woodard, C., Xia, S., Liu, S., Lyu, H., Ming, G. L., Wade, H., Song, H., Qian, J., and Zhu, H. (2013) DNA methylation presents distinct binding sites for human transcription factors. *eLife* **2**, e00726–e00726
58. Hendrich, B., and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell Biol.* **18**, 6538–6547
59. Avvakumov, G. V., Walker, J. R., Xue, S., Li, Y., Duan, S., Bronner, C.,

- Arrowsmith, C. H., and Dhe-Paganon, S. (2008) Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455**, 822–825
60. Qian, C., Li, S., Jakoncic, J., Zeng, L., Walsh, M. J., and Zhou, M.-M. (2008) Structure and hemimethylated CpG binding of the SRA domain from human UHRF1. *J. Biol. Chem.* **283**, 34490–34494
61. Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.* **23**, 58–61
62. Nan, X., Ng, H.-H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389
63. Fuks, F. (2005) DNA methylation and histone modifications: teaming up to silence genes. *Curr. Opin. Genet. Dev.* **15**, 490–495
64. Clark, S. J., Harrison, J., and Molloy, P. L. (1997) Sp1 binding is inhibited by <sup>m</sup>Cp<sup>m</sup>CpG methylation. *Gene* **195**, 67–71
65. Prendergast, G. C., Lawe, D., and Ziff, E. B. (1991) Association of Myc, the murine homolog of Max, with c-Myc stimulates methylation-sensitive DNA binding and Ras cotransformation. *Cell* **65**, 395–407
66. Harrington, M. A., Jones, P. A., Imagawa, M., and Karin, M. (1988) Cytosine methylation does not affect binding of transcription factor Sp1. *Proc. Natl. Acad. Sci. U.S.A.* **85**, 2066–2070
67. Campanero, M. R., Armstrong, M. I., and Flemington, E. K. (2000) CpG methylation as a mechanism for the regulation of E2F activity. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6481–6486
68. Iguchi-Ariga, S. M., and Schaffner, W. (1989) CpG methylation of the cAMP-responsive enhancer/promoter sequence TGACGTCA abolishes specific factor binding as well as transcriptional activation. *Genes Dev.* **3**, 612–619
69. Blattler, A., Yao, L., Wang, Y., Ye, Z., Jin, V. X., and Farnham, P. J. (2013) ZBTB33 binds unmethylated regions of the genome associated with actively expressed genes. *Epigenetics Chromatin* **6**, 13
70. Takahashi, K., and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676
71. Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W. T. C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., Eberl, H. C., Mensinga, A., Brinkman, A. B., Lephikova, K., Müller, U., Walter, J., Boelens, R., van Ingen, H., Leonhardt, H., Carell, T., and Vermeulen, M. (2013) Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159
72. Choy, J. S., Wei, S., Lee, J. Y., Tan, S., Chu, S., and Lee, T.-H. (2010) DNA methylation increases nucleosome compaction and rigidity. *J. Am. Chem. Soc.* **132**, 1782–1783
73. 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
74. Liu, Y., Zhang, X., Blumenthal, R. M., and Cheng, X. (2013) A common mode of recognition for methylated CpG. *Trends Biochem. Sci.* **38**, 177–183
75. Yang, J.-H., Li, J.-H., Jiang, S., Zhou, H., and Qu, L.-H. (2013) ChIPBase: a database for decoding the transcriptional regulation of long non-coding RNA and microRNA genes from ChIP-Seq data. *Nucleic Acids Res.* **41**, D177–D187
76. Liu, T., Ortiz, J. A., Taing, L., Meyer, C. A., Lee, B., Zhang, Y., Shin, H., Wong, S. S., Ma, J., Lei, Y., Pape, U. J., Poidinger, M., Chen, Y., Yeung, K., Brown, M., Turpaz, Y., and Liu, X. S. (2011) Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12**, R83
77. ENCODE Project Consortium, Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C., and Snyder, M. (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74
78. Wang, J., Zhuang, J., Iyer, S., Lin, X., Whitfield, T. W., Greven, M. C., Pierce, B. G., Dong, X., Kundaje, A., Cheng, Y., Rando, O. J., Birney, E., Myers, R. M., Noble, W. S., Snyder, M., and Weng, Z. (2012) Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res.* **22**, 1798–1812
79. Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H., and Glass, C. K. (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589