BioMol Concepts, Vol. 1 (2010), pp. 325–334 • Copyright © by Walter de Gruyter • Berlin • New York. DOI 10.1515/BMC.2010.031

Review

The epigenetic regulator Cfp1

David G. Skalnik

Wells Center for Pediatric Research, Section of Pediatric Hematology/Oncology, Departments of Pediatrics and Biochemistry and Molecular Biology, Indiana University School of Medicine, 1044 W. Walnut St., Indianapolis, IN 46202, USA

e-mail: dskalnik@iupui.edu

Abstract

Numerous epigenetic modifications have been identified and correlated with transcriptionally active euchromatin or repressed heterochromatin and many enzymes responsible for the addition and removal of these marks have been characterized. However, less is known regarding how these enzymes are regulated and targeted to appropriate genomic locations. Mammalian CXXC finger protein 1 is an epigenetic regulator that was originally identified as a protein that binds specifically to any DNA sequence containing an unmethylated CpG dinucleotide. Mouse embryos lacking CXXC finger protein 1 die prior to gastrulation, and embryonic stem cells lacking CXXC finger protein 1 are viable but are unable to achieve cellular differentiation and lineage commitment. CXXC finger protein 1 is a regulator of both cytosine and histone methylation. It physically interacts with DNA methyltransferase 1 and facilitates maintenance cytosine methylation. Rescue studies reveal that CXXC finger protein 1 contains redundant functional domains that are sufficient to support cellular differentiation and proper levels of cytosine methylation. CXXC finger protein 1 is also a component of the Setd1 histone H3-Lys4 methyltransferase complexes and functions to target these enzymes to unmethylated CpG islands. Depletion of CXXC finger protein 1 leads to loss of histone H3-Lys4 tri-methylation at CpG islands and inappropriate drifting of this euchromatin mark into areas of heterochromatin. Thus, one function of CXXC finger protein 1 is to serve as an effector protein that interprets cytosine methylation patterns and facilitates crosstalk with histone-modifying enzymes.

Keywords: chromatin; cytosine methylation; epigenetics; gene regulation; histone methylation.

Introduction

Epigenetics refers to heritable patterns of gene expression that occur in the absence of altered DNA sequence. This concept is illustrated by a variety of phenomena, including X-chromosome inactivation, in which one X chromosome in each cell of a developing female blastocyst becomes irreversibly inactivated; genomic imprinting, in which maternally and paternally derived alleles of a gene are differentially expressed; and the observation that diverse tissues express distinct sets of genes to permit unique functional properties, yet each (with rare exceptions) carries identical genetic information (1–4).

Epigenetic information is largely encoded within chromatin structure. A major class of epigenetic modifications is post-translational modification of histones. Dozens of distinct covalent modifications at specific amino acid residues have been identified, including acetylation, methylation, phosphorylation, and sumoylation (2, 5, 6). Many of these modifications are tightly correlated with either transcriptionally active euchromatin or transcriptionally silenced heterochromatin. Relatively subtle changes of covalent modifications can result in major changes in chromatin structure. For example, trimethylation at the lysine 4 position of histone H3 is associated with transcriptionally active euchromatin (peaking at promoter sequences), whereas methylation of the lysine 9 position of histone H3 is correlated with transcriptionally repressed heterochromatin. The histone code hypothesis predicts that specific covalent modifications of chromatin (or combinations of modifications) serve as binding sites for effector proteins (also referred to as 'reader' proteins) that modulate gene expression (7).

A second major class of epigenetic modification is cytosine methylation, which usually occurs in the context of CpG dinucleotides, and is strongly associated with gene repression (8). DNA methyltransferase (Dnmt) enzymes are responsible for adding a methyl group to form 5-methylcytosine. Dnmt1 is the major maintenance methyltransferase, which preferentially acts on hemimethylated DNA, the immediate product of DNA replication (9). Dnmt3A and Dnmt3B are *de novo* methyltransferases that are responsible for establishing cytosine methylation patterns during early development (10). Approximately 75% of CpG dinucleotides in the human genome are methylated, most of which reside within repetitive DNA elements. The CpG dinucleotide is under-represented in mammalian genomes ($\sim$10% of the expected frequency), with the exception of CpG islands, in which clusters of unmethylated CpG dinucleotides are found at the expected frequency near the promoters of $\sim$50% of genes (11–13). This unusual dinucleotide distribution appears to be a consequence of DNA repair efficiency following spontaneous cytosine deamination. Cytosine deamination produces uracil, which is efficiently recognized as a site of DNA dam-

age by the uracil glycosidase repair enzyme, leading to the subsequent restoration of the cytosine nucleotide. However, deamination of 5-methylcytosine produces thymine, which is repaired at a lower efficiency. Thus, inefficient repair represents a mutagenic pressure that resulted in the loss of most methylated CpG dinucleotides over evolutionary time. The exception to this is CpG islands, which are generally unmethylated and therefore not subject to this mutagenic pressure.

Lower eukaryotes, such as yeast and worms, do not utilize cytosine methylation as an epigenetic regulatory mechanism. However, this machinery is critical for the development of mammals, because ablation of any of the *DNMT* genes in mice is lethal (10, 14). Furthermore, mutations in Dnmt3B or the methyl-CpG binding protein MeCP2 lead to immunodeficiency, centromere instability, and facial anomaly syndrome, or the progressive neurogenerative disorder Rett syndrome, respectively (15–19). Alterations in cytosine methylation patterns are also frequently observed in cancer (20–23). Tumor cells typically carry a globally hypomethylation genome. Paradoxically, these cells also often exhibit hypermethylation of tumor suppressor gene promoters, leading to gene repression. Re-expression of these genes following treatment with Dnmt inhibitors can slow tumor growth, and pharmacological modulators of epigenetic enzymes are being tested as therapeutic agents (24–30).

The orderly restriction of lineage potential during development is associated with chromatin remodeling and a progressive accumulation of heterochromatin and restriction of gene expression (31–36). Stem cells carry a relatively open chromatin structure, including a novel bivalent chromatin signature found at many critical developmental control genes characterized by the co-existence of euchromatic (histone H3-Lys4 trimethylation) and heterochromatic (histone H3-Lys 27 tri-methylation) epigenetic marks (37). Bivalent chromatin is abundant in stem cells but is lost upon cellular differentiation (38). Another feature of stem cell epigenetics is the prevalence of non-CpG cytosine methylation, which is lost upon cellular differentiation and is reacquired upon production of induced pluripotent stem (iPS) cells (39, 40). It is now clear that resetting to a stem cell chromatin state is necessary for successful nuclear reprogramming during iPS cell production (41–43), and the efficiency of iPS cell generation can be increased by modulating the epigenetic machinery (44–47).

How the myriad of identified epigenetic modifications influence one another to coordinate appropriate chromatin structure and gene expression is a major area of investigation. It is clear that crosstalk occurs between the two major epigenetic arms (cytosine methylation and the histone code), as well as between distinct histone modifications (48–50). For example, sites of cytosine methylation serve as binding sites for methyl-CpG binding proteins such as MeCP2 and Mbd1, which then recruit histone modifying enzymes such as histone deacetylase enzymes to remodel the local histone code and facilitate heterochromatin formation. Furthermore, Dnmt proteins associate with histone deacetylase complexes (51–53); cytosine methylation in *Neurospora* is dependent on methylation of histone H3; inhibition of histone deacetylase activity by trichostatin A results in a loss of cytosine methylation (54–56); disruption of the Suv39h1 histone H3-Lys9 methyltransferase gene leads to altered localization of Dnmt3b and decreased cytosine methylation at pericentric satellite repeats (57); loss of Dnmt1 leads to perturbations in the histone code consistent with reduced heterochromatin (58); methylation of histone H4-Arg3 recruits Dnmt3A (59); and monoubiquitination of histone H2B is required for histone H34-Lys4 methylation (60–62). Thus, DNA methylation and chromatin condensation are highly integrated and mutually reinforcing mechanisms that establish and maintain heterochromatin, thus providing a unifying framework for the control of chromatin structure and gene regulation (63).

Mammals utilize an exquisitely intricate complement of epigenetic regulatory mechanisms to control chromatin structure and gene expression during development. For example, yeast cells contain a single histone H3-Lys4 methyltransferase complex (Set1/COMPASS). However, human cells express numerous distinct histone H3-Lys4 methyltransferases, including Setd1A, Setd1B, Smyd3, Mll1, Mll2, Mll3, Mll4, and Set9 (64–71). Although generally widely expressed, these mammalian methyltransferases provide non-redundant functions, as loss of a single member of the family can lead to disease or death. For example, chromosomal translocations involving the gene encoding the Mll1 histone H3-Lys4 methyltransferase are frequently found in leukemia (72–77); genetic disruption of the *MLL1* or *MLL2* genes leads to embryonic lethality in mice (78, 79); and depletion of the Smyd3 histone H3-Lys4 methyltransferase by short interfering RNA (siRNA) treatment leads to suppression of cell growth (65). It is likely that non-redundant function of each histone H3-Lys4 methyltransferase is a result of distinct target gene specificity, but the nature of these gene targets and the mechanisms utilized to achieve unique subnuclear targeting of each methyltransferase are largely unknown.

Over the past 15 years, a multitude of epigenetic marks have been identified and correlated with active or inactive chromatin, and numerous enzymes that add or remove these marks have been identified. However, relatively little is currently known regarding the regulation and genomic targeting of these enzymes.

Cloning and characterization of Cfp1

Given the precedence of methyl-CpG binding proteins that serve as readers or effectors of cytosine methylation patterns, studies were undertaken to determine whether a complementary set of effectors exist that are specific for unmethylated CpG dinucleotides, such as would be found at typical CpG islands. Ligand screening of an expression library with a double-stranded oligonucleotide probe containing an unmethylated CpG dinucleotide resulted in the isolation of cDNAs encoding human CXXC finger protein 1 (Cfp1) (80). This widely expressed protein is encoded by the *CXXC1* gene and exhibits a binding specificity for unmethylated DNA, as methylation of CpG dinucleotide(s) present within a DNA-binding site ablates Cfp1-binding affinity (80–82). Serial

selection and amplification of Cfp1-binding sites from a pool of random double-strand oligonucleotide sequences reveals that Cfp1 binds to any DNA sequence containing an unmethylated CpG dinucleotide, although greater affinity is exhibited for sequences that contain adenine or cytosine at the two flanking positions (82). The DNA-binding activity of Cfp1 is contributed by the CXXC domain (Figure 1), which is also found in several other proteins involved in the recognition or establishment of cytosine methylation, including Mbd1, Dnmt1, and Mll1. Cfp1 also contains two plant homeodomains, which are found in several dozen chromatin-associated proteins and are thought to serve as protein/protein interaction modules (83).

Consistent with binding affinity for unmethylated DNA, Cfp1 exhibits transcriptional activation activity in cotransfection assays, which depends on the presence of a CpG dinucleotide within the promoter of the reporter gene construct (80). Confocal microscopy reveals that Cfp1 forms euchromatin nuclear speckles, and serial extraction of nuclear extracts indicates that Cfp1 associates strongly with the nuclear matrix (84). Examination of various truncated forms of Cfp1 indicates that the central portion of this protein (including the acidic, basic, and coiled-coil domains) is required to direct speckling and matrix association, and that this subnuclear distribution is required for transcriptional transactivation activity (84).

Function of Cfp1 in vertebrate development

Homologous recombination in murine embryonic stem (ES) cells was performed to ablate the *CXXC1* gene. ES cells heterozygous for the disrupted *CXXC1* allele were injected into blastocysts to generate chimeric mice. Following germline transmission of the disrupted allele, heterozygous mice were bred to assess Cfp1 function during mammalian development. Heterozygous mice express approximately 50% of wild-type levels of Cfp1 and appear normal, but no viable mice were recovered that lack Cfp1 expression (85). Examination of embryos at various stages of gestation revealed that loss of Cfp1 leads to a peri-implantation death (E4.5–5.5). Blastocysts lacking a *CXXC1* gene appear normal and implant in the uterine wall, but these embryos fail to gastrulate. Thus, Cfp1 is required for early mammalian embryogenesis.

Additional experimental approaches demonstrate a role for Cfp1 during later developmental stages. Injection of antisense morpholino oligonucleotides into two-cell zebrafish embryos to deplete Cfp1 leads to a variety of developmental defects, including cardiac edema, runting, incomplete vasculature formation, a failure of primitive hematopoiesis, elevated levels of apoptosis, and death (86). The specificity of this phenotype was established by repeating the study with a second independent antisense oligonucleotide sequence. Furthermore, the phenotype was rescued by co-injection of mRNA encoding murine Cfp1. Similarly, depletion of Cfp1 using short hairpin RNA leads to cell death in human leukemia cell lines (87).

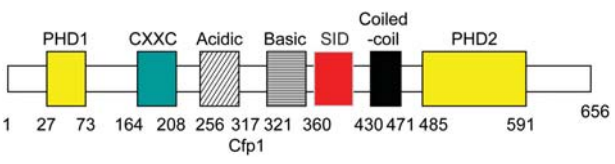

Figure 1 Schematic representation of the domain structure of Cfp1.
Numbers refer to amino acid residue positions. PHD, plant homeodomain; CXXC, DNA-binding domain; SID, Setd1 interaction domain.

Cfp1 is an epigenetic regulator: cytosine methylation

Given the binding affinity of Cfp1 for unmethylated CpG dinucleotides, additional studies were conducted to assess the epigenome in cells lacking Cp1. ES cell lines lacking Cfp1 were derived from blastocysts resulting from breeding between mice heterozygous for the deleted *CXXC1* allele. The successful isolation and propagation of Cfp1-deficient ES cell lines demonstrates that Cfp1 is not essential for stem cell viability (88). This finding also suggests that the timing of death for murine embryos lacking a *CXXC1* gene is due to a requirement for Cfp1 during specific developmental stages (e.g., gastrulation), rather than the time at which maternally derived Cfp1 protein is depleted. Importantly, the isolation of ES cells lacking Cfp1 provides a convenient reagent with which to probe the molecular function(s) of Cfp1.

The status of cytosine methylation in ES cells lacking Cp1 was assessed by several methods, including: (i) isoschizomer Southern blot analysis, in which the presence of cytosine methylation at a specific gene locus can be determined using a methyl-sensitive restriction enzyme; (ii) thin layer chromatographic measurement of global 5-methylcytosine levels; and (iii) methyl acceptance assay, in which the ability of a genomic DNA sample to accept methyl groups in an *in vitro* assay is inversely proportional to the degree of global genomic cytosine methylation present in the sample. All of these approaches indicate that ES cells lacking Cfp1 exhibit a 60–70% decline in genomic cytosine methylation levels (88). Cytosine methylation deficiency was observed at all genomic loci analyzed, including repetitive DNA elements, single copy genes, and imprinted genes. A similar deficit of global cytosine methylation was observed in zebrafish embryos following treatment with Cfp1 antisense oligonucleotides (86). Remarkably, the cytosine methylation deficiency observed in *CXXC1*-null murine ES cells is rescued upon introduction of a Cfp1 expression vector into these cells. Even differential methylation at the paternally imprinted H19 locus was rescued by Cfp1 (88). Failure to rescue was only observed at the insulin-like growth factor 2 receptor locus, presumably because male ES cells are unable to regenerate a maternal imprint mark. The ability of these cells to restore proper patterns of cytosine methylation following re-expression of Cfp1 illustrates the robust plasticity of this epigenetic modification and suggests that residual epigenetic

modifications that persist following loss of Cfp1 provide sufficient information for re-establishment of appropriate cytosine methylation patterns upon restoration of the appropriate regulatory apparatus.

Analysis of cellular extracts reveals that cells lacking Cfp1 exhibit reduced Dnmt activity (88). These cells exhibit approximately a 60% decline of methyltransferase activity when a hemimethylated oligonucleotide substrate is utilized, which corresponds to the preferred target for the maintenance Dnmt1 enzyme. In contrast, methyltransferase activity towards an unmethylated DNA substrate is normal, suggesting that *de novo* methyltransferase activity is unaffected by loss of Cfp1. This conclusion is further supported by the finding that ES cells lacking Cfp1 exhibit normal kinetics for the acquisition of *de novo* cytosine methylation of a transduced provirus (88). However, the provirus fails to become progressively more heavily methylated over time, which is consistent with a deficiency in maintenance Dnmt activity. Indeed, an indistinguishable pattern of provirus methylation was previously observed when a similar study was performed using ES cells lacking the *DNMT1* gene (14). Cfp1 physically interacts with Dnmt1 and *CXXC1*-null ES cells contain reduced levels of Dnmt1 protein (89, 90). However, Cfp1-deficient ES cells express elevated levels of Dnmt1 transcript. Instead, reduced steady-state levels of Dnmt1 in these cells are a consequence of reduced efficiency of Dnmt1 translation and reduced Dnmt1 protein half-life (90).

ES cells lacking Cfp1, although viable, are unable to achieve cellular differentiation *in vitro* (88). Upon induction of differentiation by removal of leukemia inhibitory factor from the growth medium, these cells fail to downregulate stem cell markers such as alkaline phosphatase and Oct4 and fail to induce a variety of lineage-restricted markers such as brachyury (mesoderm), Gata4 (endoderm), c-fms (myeloid), or β-MHC (cardiac). This is consistent with the behavior of Cfp1-null blastocysts, which are unable to gastrulate, and suggests a model in which Cfp1 is required for the remodeling of chromatin structure that is necessary for stem cell differentiation. Importantly, *in vitro* differentiation capacity is restored to these cells following introduction of a Cfp1 expression vector (88).

CXXC1-null ES cells also exhibit hypersensitivity to DNA-damaging agents such as ionizing radiation, cisplatin, etoposide, hydrogen peroxide, methylmethanesulfonate, and temozolomide (91). However, these cells exhibit normal sensitivity to non-genotoxic agents such as methotrexate and paclitaxel. Cfp1-null ES cells also exhibit approximately a 50% decline in the level and enzymatic activity of the apurunic/apyrimidinic endonuclease (Ape) DNA repair enzyme. Whether the observed hypersensitivity to DNA-damaging agents is due to increased DNA accessibility as a consequence of relaxed chromatin structure (i.e., reduced levels of cytosine methylation) or reduced efficiency of the repair of such lesions, or both, remains to be determined. Normal levels of Ape expression and sensitivity to DNA-damaging agents are restored following rescue of *CXXC1*-null ES cells with the Cfp1 expression vector (91).

Taken together, these findings establish Cfp1 as a critical regulator of maintenance DNA methylation, chromatin structure, stem cell function, and embryonic development.

Cfp1 is an epigenetic regulator: histone methylation

Although the decline in cytosine methylation in the absence of Cfp1 is dramatic, several lines of evidence suggest that Cfp1 serves additional functions. For example, mouse embryos lacking Dnmt1 die later in gestation (around E9.5–10.5) compared to *CXXC1*-null embryos (E4.5–5.5), despite suffering from a more severe loss of cytosine methylation (14, 85). In addition, Cfp1 homologs are found in lower eukaryotes such as yeast and worms that lack cytosine methylation. Interestingly, the Cfp1 homologs found in these organisms lack the CXXC DNA-binding domain, suggesting that Cfp1 performs an ancestral function in these organisms, and acquired DNA-binding activity in higher eukaryotes that utilize cytosine methylation as an epigenetic regulatory mechanism.

Sucrose gradient equilibrium centrifugation revealed that endogenous Cfp1 associates with a complex of ~450 kDa (67). Co-immunoprecipitation studies were performed to gain additional insight into the molecular basis of Cfp1 function. Cfp1 was found to interact with several proteins that represent the mammalian homologs of the yeast Set1/COMPASS histone H3-Lys4 methyltransferase complex (67). This includes Setd1A, Ash2L, Rbbp5, Wdr5, and Wdr82. As mentioned above, mammalian cells express numerous distinct histone H3-Lys4 methyltransferase complexes. Further studies revealed that Cfp1 additionally interacts with Setd1B, but not with the Mll members of the histone H3-Lys4 methyltransferase family (68). It is the Setd1 enzymes that appear to be responsible for the bulk of histone H3-Lys4 methylation in mammalian cells, because siRNA-mediated depletion of either Setd1A or Cfp1 leads to dramatic global reduction in histone H3-Lys4 tri-methylation (92, 93).

Examination of histone methylation patterns reveals that ES cells lacking Cfp1 contain elevated levels of histone H3-Lys4 methylation (the product of the Set1-catalyzed reaction) and decreased levels of histone H3-Lys9 methylation (67). This is consistent with reduced levels of cytosine methylation in these cells and indicates reduced levels of heterochromatin. This result also suggests that one function of Cfp1 is to inhibit or restrict the activity of the Setd1 histone H3-Lys4 methyltransferase complexes. Data in support of this idea is provided by confocal microscopy examination of the subnuclear distribution of Setd1A (94). In wild-type ES cells, the Setd1A protein is nearly exclusively localized to 4′,6-diaminidino-2-phenylindone (DAPI)-dim areas of euchromatin. However, in ES cells lacking Cfp1 approximately 25% of the Setd1A protein inappropriately colocalizes with DAPI-bright areas of heterochromatin (Figure 2). Restriction of Setd1A to euchromatin domains is restored following introduction of a Cfp1 expression vector into the knockout ES cell line. Similar results were obtained when the subnu-

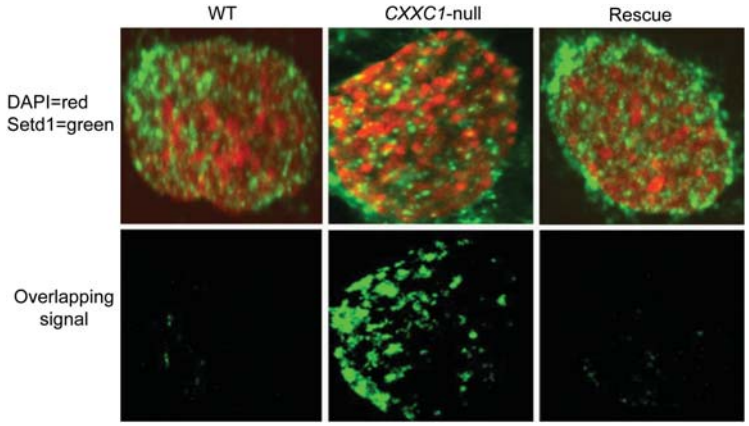

Figure 2 Cfp1 is required to restrict Setd1A protein to euchromatin.
Confocal microscopy was utilized to assess the subnuclear distribution of the Setd1A protein in wild-type (WT), *CXXC1*-null, or rescued (*CXXC1*-null cells transfected with a Cfp1 expression vector) murine ES cell lines. The top row illustrates the relative subnuclear distribution of Setd1A (green) and DAPI-bright heterochromatin (red). Colocalization is indicated by a yellow or orange color. The bottom row reveals the extent of overlapping signal (displayed in green).

clear distribution of histone H3-Lys4 tri-methylation, the product of the Setd1A-catalyzed reaction, was analyzed (94).

In further support for this model, a recent report from the Bird group demonstrated that Cfp1 is found at ~80% of CpG islands and that siRNA-mediated depletion of Cfp1 results in the loss of Set1A occupancy and a consequent reduction of histone H3-Lys4 trimethylation at these sites (92). Taken together, these data suggest a model in which Cfp1 functions as a reader of cytosine methylation patterns and functions to tether the Setd1 histone H3-Lys4 methyltransferase complexes to unmethylated CpG islands, thus providing crosstalk between patterns of cytosine methylation and the histone code (Figure 3). Because of the predominance of heterochromatin in cells, this model explains the apparent paradox of reduced histone H3-Lys4 methylation at CpG islands following Cfp1 depletion, while at the same time global levels of this modification are elevated.

Structure/function studies of Cfp1

The ability to rescue various aspects of the *CXXC1*-null ES cell phenotype by expression of full-length Cfp1 provides an attractive system with which to probe structure/function relationships of Cfp1. Functional domains of Cfp1 that are necessary and/or sufficient to support rescue activity were identified following introduction of various truncated and/or mutated versions of Cfp1 cDNA into knockout ES cells. Surprisingly, these studies revealed that Cfp1 contains redundant functional domains (91, 95). Expression in *CXXC1*-null ES cells of either the amino terminal half (aa 1–367) or carboxyl terminal half (aa 361–656) of Cfp1 is sufficient to rescue the defects observed in cytosine methylation, hypersensitivity to DNA-damaging agents, and *in vitro* differentiation capacity. Point mutations within Cfp1 functional domains were analyzed to further probe the requirements for Cfp1 rescue activity. A point mutation within the CXXC domain (C169A) ablates Cfp1 DNA-binding activity (82), and a point mutation within the SID domain (C375A) ablates the interaction between Cfp1 and the Set1 enzymes (89). Introduction of either of these mutations into the full-length Cfp1 cDNA does not affect rescue activity, demonstrating that neither DNA-binding activity nor interaction with the Set1 complexes is required for rescue activity. However, ablation of DNA-binding activity within the 1–367 aa Cfp1 fragment or ablation of Set1 interaction within the 361–656 aa Cfp1 fragment eliminates rescue activity. Similarly, introduction of

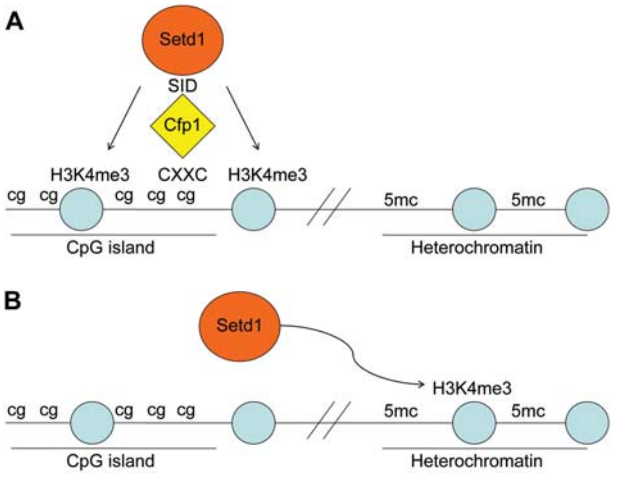

Figure 3 Model of Cfp1 function in tethering the Setd1 histone H3-Lys4 methyltransferase complexes to unmethylated CpG islands.
(A) The CXXC (DNA-binding) and SID (Setd1-interaction) domains function to tether the Setd1 complex at unmethylated CpG islands, resulting in the acquisition of histone H3-Lys4 trimethylation (H3K4me3) at adjacent nucleosomes (blue circles). CG indicates CpG dinucleotides. (B) In the absence of Cfp1, the Set1d1 complexes are no longer restricted to CpG islands, leading to the loss of histone H3-Lys4 trimethylation at these sites and to the inappropriate drifting of this mark into areas of heterochromatin that contain 5-methylcytosine (5mc).

both point mutations into the full-length Cfp1 cDNA also ablates rescue activity. Thus, retention of either DNA-binding activity or interaction with the Set1 complexes is required for Cfp1 function in these rescue assays.

A prediction of the model that Cfp1 tethers Sed1 complexes to unmethylated CpG islands is that both the CXXC and SID domains of Cfp1 should be required to permit appropriate targeting of the Setd1 protein to euchromatic CpG islands. This was confirmed using confocal microscopy to determine what Cfp1 domains are required to rescue normal subnuclear distribution following transfection into *CXXC1*-null ES cells (94). In contrast to the rescue of cytosine methylation, *in vitro* differentiation, and hypersensitivity to DNA-damaging agents, proper subnuclear distribution of Setd1A and histone H3-Lys4 tri-methylation requires both the DNA-binding and Setd1-interaction domains of Cfp1. This finding reveals distinct functional requirements for rescue of distinct aspects of the *CXXC1*-null ES cell phenotypes and reveals that proper euchromatic targeting of Setd1A is not required to support proper patterns of cytosine methylation and *in vitro* differentiation.

Genomic targeting of Setd1 histone H3-Lys4 methyltransferase complexes

The complexity of histone H3-Lys4 methyltransferases present in mammalian cells presumably reflects the requirement for elaborate control of chromatin structure needed for execution of complex developmental programs. The fact that each methyltransferase contributes non-redundant function is illustrated by the severe developmental aberrations observed upon misregulation of a single family member. For example, ablation of the *Mll1* gene leads to embryonic death in mice, and chromosomal translocations of this gene are commonly found in human leukemia. Relatively little is known about how each member of the histone H3-Lys4 methyltransferase family is targeted to appropriate genomic sites for action. The data described above demonstrate that one function of Cfp1 is to tether the Setd1A complex to unmethylated CpG islands. However, as described below, genomic targeting of Setd1 complexes is complex, and Cfp1 cannot be the primary determinant for the genomic targeting of the Setd1 complexes.

Several components of the mammalian histone H3-Lys4 methyltransferase complexes have been implicated in genomic targeting. In addition to the previously described role of Cfp1, the yeast homolog (Spp1) has been demonstrated to bind to sites of histone H3-Lys4 methylation (96) and to the carboxyl-tail of histone H2B (97). Whether these histone-binding properties have been conserved in mammalian Cfp1 remains to be determined. In addition, Wdr5 is a Wd40 domain-containing subunit that is a member of all complexes of this methyltransferase family and binds to the histone H3-Lys4 residue (98–101). Wdr82, another Wd40-containing protein, directly interacts with the amino terminal RNA recognition domain within the Setd1A and Setd1B enzymes, and also interacts with RNA polymerase II (pol II) molecules containing Ser5-phosphorylated C-terminal repeats (102). This form of pol II is characteristic of initiating pol II found near promoters (103). Thus, the action of Wdr82 tethers the Setd1 methyltransferases to transcription start sites, consistent with a peak of histone H3-Lys4 trimethylation that is observed at these genomic sites. Hence, multiple protein/protein interactions appear to contribute to the proper genomic targeting of mammalian histone H3-Lys4 methyltransferases.

Perhaps the most dramatic example of the complexity of histone methyltransferase targeting is revealed by examination of the Setd1A and Setd1B complexes. The composition of these complexes is identical, except for the identity of the enzymatically active (Setd1) component. Despite this similarity, confocal microscopy analysis of the subnuclear distribution of these factors reveals that they exhibit nearly non-overlapping localization to euchromatic domains (Figure 4) (68). Given that all of the putative targeting molecules described above are shared by the Setd1A and Setd1B complexes, this observation indicates that yet additional targeting signals are contributed by other molecules that interact differentially between the Setd1A and Setd1B proteins. One candidate that could contribute this activity is Hcf1, which has been implicated in targeting of the Setd1A complex to the herpesvirus immediate early promoters (104).

Outlook

Research to date has revealed Cfp1 to be an essential regulator of epigenetic marks such as cytosine methylation and histone methylation, and to be essential for embryonic vertebrate development. Given its direct intersection with and regulation of Dnmt1 and the Setd1 histone H3-Lys4 methyltransferases, a better understanding of Cfp1 function could provide important information regarding crosstalk and integration between these two epigenetic marks. Recent developments in chromatin immunoprecipitation and high-throughput

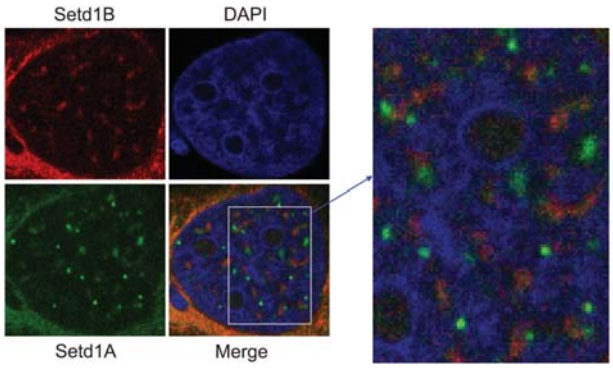

Figure 4 The Setd1A and Setd1B proteins exhibit a nearly non-overlapping subnuclear distribution.
The subnuclear distribution of endogenous Setd1A (green) and Setd1B (red) proteins were detected by indirect immunofluorescence and compared with the localization of DAPI-bright heterochromatin domains (blue). An expanded view of the merged image is shown on the right, revealing that Setd1A and Setd1B reside in distinct DAPI-dim (euchromatic) domains.

sequencing makes feasible a full analysis of the altered epigenome in ES cells lacking Cfp1 (and following rescue), which should reveal new insights into the genomic targeting of epigenetic modifying enzymes. Additional studies are needed to develop a comprehensive understanding of how the redundant functional domains of Cfp1 support its pleiotropic functions. Furthermore, development of conditional knockout mice will permit analysis of Cfp1 function in adult stem cells (e.g., hematopoietic stem cells) and homeostasis of mature tissues.

Highlights

- Cfp1 binds specifically to DNA containing unmethylated CpG dinucleotides.
- Cfp1 is required for early mammalian embryogenesis.
- Murine ES cells lacking Cfp1 are viable, but carry a perturbed epigenome and are unable to achieve cellular differentiation.
- Murine ES cells lacking Cfp1 are hypersensitive to DNA-damaging agents.
- Cfp1 facilitates maintenance cytosine methylation.
- Cfp1 is a component of the Setd1 histone H3-Lys4 methyltransferase complexes.
- Cfp1 functions to restrict Setd1A to unmethylated CpG islands and is required to target histone H3-Lys4 tri-methylation to these sites.
- Cfp1 contains redundant functional domains that are sufficient to support cytosine methylation and cellular differentiation.
- Cfp1 appears to be necessary for chromatin remodeling that is required for stem cell differentiation.

Acknowledgments

The author thanks Jeong-Heon Lee and Erika Dobrota, who performed the experiments presented in Figures 3 and 4, respectively. This research was supported by the National Science Foundation, the Riley Children's Foundation, and the Lilly Endowment. The author declares no competing financial interests.

References

1. Bartolomei MS, Tilghman SM. Genomic imprinting in mammals. Annu Rev Genet 1997; 31: 493–525.
2. Felsenfeld G, Groudine M. Controlling the double helix. Nature 2003; 421: 448–53.
3. Heard E, Clerc P, Avner P. X-chromosome in activation in mammals. Annu Rev Genet 1997; 31: 571–610.
4. Reik W, Walter J. Imprinting mechanisms in mammals. Curr Opin Genet Dev 1998; 8: 154–64.
5. Grewal SI, Moazed D. Heterochromatin and epigenetic control of gene expression. Science 2003; 301: 798–802.
6. Jaenisch R, Bird A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet (Suppl) 2003; 33: 245–54.
7. Chi P, Allis CD, Wang GG. Covalent histone modifications – miswritten, misinterpreted and mis-erased in human cancers. Nat Rev Cancer 2010; 10: 457–69.
8. Bird A. DNA methylation patterns and epigenetic memory. Genes Dev 2002; 16: 6–21.
9. Yoder JA, Soman NS, Verdine GL, Bestor T. DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanisms-based probe. J Mol Biol 1997; 270: 385–95.
10. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for *de novo* methylation and mammalian development. Cell 1999; 99: 247–57.
11. Antequera F, Bird A. Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci USA 1993; 90: 11995–9.
12. Gardiner-Garden M, Frommer M. CpG islands in vertebrate genomes. J Mol Biol 1987; 196: 261–82.
13. Larsen F, Gundersen G, Lopez R, Prydz H. CpG islands as gene markers in the human genome. Genomics 1992; 13: 1095–107.
14. Lei H, Oh SP, Okano M, Juttermann R, Goss KA, Jaenisch R, Li E. *De novo* DNA cytosine methyltransferase activities in mouse embryonic stem cells. Development 1996; 122: 3195–205.
15. Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U, Zoghbi HY. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet 1999; 23: 185–8.
16. Chen R, Akbarian S, Tudor M, Jaenisch R. Deficiency of methyl-CpG binding protein-2 in CNS neurons results in Rett-like phenotype in mice. Nat Genet 2001; 27: 327–31.
17. Guy J, Hendrich B, Holmes M, Martin JE, Bird A. A mouse MeCP2-null mutation causes neurological symptoms that mimic Rett syndrome. Nat Genet 2001; 27: 322–6.
18. Hansen RC, Wijmenga C, Luo P, Stanek AM, Canfield TK, Weemaes CM. The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc Natl Acad Sci USA 1999; 96: 14412–7.
19. Xu G-L, Bestor TH, Bourc'his D, Hsieh C-L, Tommerup N, Bugge M, Hulten M, Qu X, Russo JJ, Viegas-Pequignot E. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. Nature 1999; 402: 187–91.
20. Baur AS, Shaw P, Burri N, Delacretaz F, Bosman FT, Chaubert P. Frequent methylation silencing of $p15^{INK4b}$ (MTS2) and $p16^{INK4a}$ (MTS1) in B-cell and T-cell lymphomas. Blood 1999; 94: 1773–81.
21. Guo SX, Taki T, Ohnishi H, Piao HY, Tabuchi K, Bessho F, Hanada R, Ynagisawa M, Hayashi Y. Hypermethylation of p16 and p15 genes and RB protein expression in acute leukemia. Leukemia Res 2000; 24: 39–46.
22. Issa JP, Zehnbauer BA, Kaufmann SH, Biel MA, Baylin SB. HIC1 hypermethylation is a late event in hematopoietic neoplasms. Cancer Res 1997; 57: 1678–81.
23. Katzenellenbogen RA, Baylin SB, Herman JG. Hypermethylaion of the DAP-kinase CpG island is a common alteration in B-cell malignancies. Blood 1999; 93: 4347–53.
24. Braiteh F, Soriano AO, Garcia-Manero G, Hong D, Johnson MM, De Padua Silva L, Yang H, Alexander S, Wolff J, Kurzrock R. Phase I study of epigenetic modulation with 5-azacytidine and valproic acid in patients with advanced cancers. Clin Cancer Res 2008; 14: 6296–301.
25. Fenaux P, Mufti GJ, Hellstorn-Lindberg E, Santini V, Finelli C, Giagounidis A, Schoch R, Gatternmann N, Sanz G, List A, Gore SD, Seymour JF, Bennett JM, Byrd J, Backstorm J, Zim-

merman L, McKenzie D, Beach CL, Silverman LR. Efficacy of azacitidine compared with that of conventional care regimens in the treatment of higher-risk myelodysplastic syndromes: a randomised, open-label, phase III study. Lancet Oncol 2009; 10: 223–32.
26. Hennessy BT, Garcia-Manero G, Kantarjian HM, Giles FJ. DNA methylation in haematological malignancies: the role of decitabine. Expert Opin Investig Drugs 2003; 12: 1985–93.
27. Kelly TK, De Carvalho DD, Jones PA. Epigenetic modifications as therapeutic targets. Nat Biotechnol 2010; 28: 1069–78.
28. Lin J, Gilbert J, Rfudek MA, Zwiebel JA, Gore S, Jiemjit A, Zhao M, Baker SD, Smbinder RF, Herman JG, Donehower RC, Carducci MA. A phase I dose-finding study of 5-azacytidineincombination with sodium phenylbutyrate in patients with refractory solid tumors. Clin Cancer Res 2009; 15: 6241–9.
29. Lyons J, Bayar E, Fine G, McCullar M, Rolens R, Rubinfeld J, Rosenfeld C. Decitabine: development of a DNA methyltransferase inhibitor for hematological malignancies. Curr Opin Invest Drugs 2003; 4: 1442–50.
30. Wijermans P, Lubbert M, Verhoef G, Bosly A, Ravoet C, Andre M, Ferrant A. Low-dose 5-aza-2′-deoxycytidine, a DNA hypomethylating agent, for the treatment of high-risk myelodysplastic syndrome: a multicenter phase II study in elderly patients. J Clin Oncol 2000; 18: 956–62.
31. Akashi K, He X, Chen J, Iwasaki H, Niu C, Steenhard B, Zhang J, Haug J, Li L. Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. Blood 2003; 101: 383–90.
32. Attema JL, Papathanasiou P, Forsberg EC, Xu J, Smale ST, Weissman IL. Epigenetic characterization of hematopoietic stem cell differentiation using miniChIP and bisulfite sequencing analysis. Proc Natl Acad Sci USA 2007; 104: 12371–6.
33. Efroni S, Duttagupta R, Cheng JC, Dehghani H, Hoeppner DJ, Dash C, Bazett-Jones DP, Le Grice S, McKay RD, Buetow KH, Gineras TR, Misteli T, Meshorer E. Global transcription in pluripotent embryonic stem cells. Cell Stem Cell 2008; 2: 437–47.
34. Meissner A. Epigenetic modifications in pluripotent and differentiated cells. Nat Biotechnol 2010; 28: 1079–88.
35. Mohammad HP, Baylin SB. Linking cell signaling and the epigenetic machinery. Nat Biotechnol 2010; 28: 1033–8.
36. Surani MA, Hayashi K, Hajkova P. Genetic and epigenetic regulators of pluripotency. Cell 2007; 128: 747–62.
37. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Reil R, Schreiber SL, Lander ES. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 2006; 125: 315–26.
38. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007; 448: 553–60.
39. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, Low HM, Wing K, Sung K, Rigoutsos I, Loring J, Wei CL. Dynamic changes in the human methylome during differentiation. Genome Res 2010; 20: 320–31.
40. Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G, Tonti-Filippini J, Nery JR, Lee L, Ye Z, Ngo QM, Edsell L, Antosiewicz-Bourget J, Stewart R, Ruotti V, Millar AH, Thomson JA, Ren B, Ecker JR. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 2009; 462: 315–22.
41. Hajkova P. Epigenetic reprogramming – taking a lesson from the embryo. Curr Opin Cell Biol 2010; 22: 342–50.
42. Maherali N, Sridharan R, Xie W, Utikal J, Eminli S, Arnold K, Stadtfeld M, Tachechko R, Tchieu J, Jaenisch R, Plath K, Hochedlinger K. Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell 2007; 1: 55–70.
43. Mikkelsen TS, Hanna J, Zhang X, Ku M, Wernig M, Schorderet P, Bernstein BE, Jaenisch R, Lander ES, Meissner A. Dissecting direct reprogramming through integrative genomic analysis. Nature 2008; 454: 49–55.
44. Huangfu D, Osafune K, Maehr R, Guo W, Eijkelenboom A, Chen S, Muhlestein W, Melton DA. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol 2008; 26: 1269–75.
45. Mali P, Chou BK, Yen JY, Zou J, Dowey S, Brodsky RA, Oh JE, Yu W, Baylin SB, Yusa K, Bradley A, Meyers DJ, Mukherjee C, Cole PA, Cheng L. Butyrate greatly enhances derivation of human induced pluripotent stem cells by promoting epigenetic remodeling and the expression of pluripotency-associated genes. Stem Cells 2010; 28: 713–20.
46. Shi Y, Do J, Desponts C, Hahm H, Scholer HR, Ding S. A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell Stem Cell 2008; 2: 525–8.
47. Stadtfeld M, Hochedlinger K. Induced pluripotency: history, mechanisms, and applications. Genes Dev 2010; 24: 2239–63.
48. Cedar H, Bergman Y. Linking DNA methylation and histone modifications: patterns and paradigms. Nat Rev Genet 2009; 10: 295–304.
49. Cheng X, Blumenthal RM. Coordinated chromatin control: structural and functional linkage of DNA and histone methylation. Biochemistry 2010; 49: 2999–3008.
50. Lee J-S, Smith E, Shilatifard A. The language of histone crosstalk. Cell 2010; 142: 682–5.
51. Robertson KD, Ait-Si-Ali S, Yokochi T, Wade PA, Jones PL, Wolffe AP. DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. Nat Genet 2000; 25: 338–42.
52. Fuks F, Burgers WA, Brehm A, Hughes-Davies L, Kouzarides T. DNA methyltransferase Dnmt1 associates with histone deacetylase activity. Nat Genet 2000; 24: 88–91.
53. Fuks F, Burgers WA, Godin N, Kasai M, Kouzarides T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. EMBO J 2001; 20: 2536–44.
54. Tamaru H, Selker EU. A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. Nature 2001; 414: 277–83.
55. Selker EU. Trichostatin A causes selective loss of DNA methylation in *Neurospora*. Proc Natl Acad Sci USA 1998; 95: 9430–5.
56. Tamaru H, Zhang X, McMillen D, Singh PB, Nakayama JI, Grewal SI, Allis CD, Cheng X, Selker EU. Trimethylated lysine 9 of histone H3 is a mark for DNA methylation in *Neurospora crassa*. Nat Genet 2003; 34: 75–9.
57. Lehnertz B, Ueda Y, Derijck AAHA, Braunschweig U, Perez-Burgos L, Kubicek S, Chen T, Li E, Jenuwein T, Peters A. Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr Biol 2003; 13: 1192–200.
58. Espada J, Ballestar E, Fraga MF, Villar-Garea A, Juarranz A, Stockert JC, Robertson KD, Fuks F, Esteller M. Human DNA

methyltransferase 1 is required for maintenance of the histone H3 modification pattern. J Biol Chem 2004; 279: 37175–84.
59. Zhao Q, Rank G, Tan YT, Li H, Moritz RL, Simpson RJ, Cerruti L, Curtis DJ, Patel DJ, Allis CD, Cunningham JM, Jane SM. PRMT5-mediated methylation of histone H4R3 recruits DNMT3A,
coupling histone and DNA methylation in gene silencing. Nat Struct Mol Biol 2009; 16: 304–11.
60. Kim J, Guermah M, McGinty RK, Lee J-S, Tang Z, Milne TA, Shilatifard A, Muir TW, Roeder RG. RAD6-mediated transcription-coupled H2B ubiquitylation directly stimulates H3K4 methylation in human cells. Cell 2009; 137: 459–71.
61. Dover J, Schneider J, Tawiah-Boateng MA, Wood A, Dean K, Johnston M, Shilatifard A. Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. J Biol Chem 2002; 277: 28368–71.
62. Lee J-S, Shukla A, Schneider J, Swanson SK, Washburn MP, Florens L, Bhaummi SR, Shilatifard A. Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell 2007; 131: 1084–96.
63. Burgers WA, Fuks F, Kouzarides T. DNA methyltransferases get connected to chromatin. Trends Genet 2002; 18: 275–7.
64. Goo Y-H, Sohn YC, Kim D-H, Kim S-W, Kang M-J, Jung D-J, Kwak E, Barlev NA, Berger SL, Chow VT, Roeder RG, Azorsa DO, Meltzer PS, Suh PG, Song EJ, Lee KJ, Lee YC, Lee JW. Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Mol Cell Biol 2003; 23: 140–9.
65. Hamamoto R, Furukawa Y, Morita M, Iimura Y, Silva FP, Li M, Yagyu R, Nakamura Y. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. Nat Cell Biol 2004; 6: 731–40.
66. Hughes CM, Rozenblatt-Rosen O, Milne TA, Copeland TD, Levine SS, Lee JC, Hayes DN, Shanmugam KS, Bhattacharjee A, Biondi CA, Kay GF, Hayward NK, Hess JL, Meyerson M. Menin associates with a trithorax family histone methyltransferase complex and with the Hoxc8 locus. Mol Cell 2004; 13: 587–97.
67. Lee J-H, Skalnik DG. CpG binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J Biol Chem 2005; 280: 41725–31.
68. Lee J-H, Tate CM, You J-S, Skalnik DG. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. J Biol Chem 2007; 282: 13419–28.
69. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, Hess JL. MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell 2002; 10: 1107–17.
70. Nakamura T, Mori T, Tada S, Krajewski W, Rozovskaia T, Wassell R, Dubois G, Mazo A, Corcie CM, Canaani E. ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell 2002; 10: 1119–28.
71. Nishioka K, Chuikov S, Sarma K, Erdjument-Bromage H, Allis CD, Tempst P, Reinberg D. Set9, a novel histone H3 methyltransferase that facilitates transcription by precluding histone tail modifications required for heterochromatin formation. Genes Dev 2002; 16: 479–89.
72. Domer PH, Fakharzadeh SS, Chen CS, Jockel J, Johansen L, Silverman GA, Kersey JH, Korsmeyer SJ. Acute mixed-lineage leukemia t(4; 11)(q21; q23) generates an MLL-AF4 fusion product. Proc Natl Acad Sci USA 1993; 90: 7884–8.
73. Gu Y, Nakamura T, Alder H, Prasad R, Canaani O, Cimino G, Croce CM, Canaani E. The t(4; 11) chromosome translocation of human acute leukemias fuses the ALL-1 gene, related to *Drosophila* trithorax, to the AF-4 gene. Cell 1992; 71: 701–8.
74. Ma Q, Alder H, Nelson KK, Chatterjee D, Gu Y, Nakamura T, Canaani E, Croce CM, Sircussa LD, Buchberg AM. Analysis of the murine ALL-1 gene reveals conserved domains with human ALL-1 and identifies a motif shared with DNA methyltransferases. Proc Natl Acad Sci USA 1993; 90: 6350–4.
75. Prasad R, Yano T, Sorio C, Nakamura T, Rallapalli R, Gu Y, Leshkowitz D, Croce CM, Canaani E. Domains with transcriptional regulatory activity within the ALL1 and AF4 proteins involved in acute leukemia. Proc Natl Acad Sci USA 1995; 92: 12160–4.
76. Tkachuk DC, Kohler S, Cleary ML. Involvement of a homolog of *Drosophila* Trithorax by 11q23 chromosomal translocations in acute leukemias. Cell 1992; 71: 691–700.
77. Zeleznik-Le N, Harden AM, Rowley JD. 11q23 translocations split the ''AT-hook'' cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the mixed-lineage leukemia (MLL) gene. Proc Natl Acad Sci USA 1994; 91: 10610–4.
78. Ayton P, Sneddon SF, Palmer DB, Rosewell IR, Owen MJ, Young B, Presley R, Subramanian V. Truncation of the MLL gene in exon 5 by gene targeting leads to early preimplantation lethality of homozygous embryos. Genesis 2001; 30: 201–12.
79. Glaser S, Schaft J, Lubitz S, Vintersten K, van der Hoeven F, Tufteland KR, Aasland R, Anastassiadis K, Ang SL, Stewart AF. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. Development 2006; 133: 1423–32.
80. Voo KS, Carlone DL, Jacobsen BM, Flodin A, Skalnik DG. Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. Mol Cell Biol 2000; 20: 2108–21.
81. Carlone DL, Hart SRL, Ladd PD, Skalnik DG. Cloning and characterization of the gene encoding the mouse homologue of CpG binding protein. Gene 2002; 295: 71–7.
82. Lee J-H, Voo KS, Skalnik DG. Identification and characterization of the DNA binding domain of CpG-binding protein. J Biol Chem 2001; 276: 44669–76.
83. Aasland R, Gibson TJ, Stewart F. The PHD finger: implications for chromatin-mediated transcriptional regulation. Trends Biochem Sci 1995; 20: 56–9.
84. Lee J-H, Skalnik DG. CpG binding protein is a nuclear matrix- and euchromatin-associated protein localized to nuclear speckles containing human trithorax: identification of nuclear matrix targeting signals. J Biol Chem 2002; 277: 42259–67.
85. Carlone DL, Skalnik DG. CpG binding protein is crucial for early embryonic development. Mol Cell Biol 2001; 21: 7601–6.
86. Young SRL, Mumaw C, Marrs JA, Skalnik DG. Antisense targeting of CXXC finger protein 1 inhibits genomic cytosine methylation and primitive hematopoiesis in zebrafish. J Biol Chem 2006; 281: 37034–44.
87. Young SRL, Skalnik DG. CXXC-finger protein 1 is required for normal proliferation and differentiation of the PLB-985 myeloid cell line. DNA Cell Biol 2007; 26: 80–90.
88. Carlone DL, Lee J-H, Young SRL, Dobrota E, Butler JS, Ruiz J, Skalnik DG. Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. Mol Cell Biol 2005; 25: 4881–91.
89. Butler JS, Lee J-H, Skalnik DG. CFP1 interacts with DNMT1 independently of association with the Setd1 histone H3K4

methyltransferase complexes. DNA Cell Biol 2008; 27: 533–43.
90. Butler JS, Palam LR, Tate CM, Sanford JR, Wek RC, Skalnik DG. DNMT1 protein synthesis is reduced in CFP1-deficient embryonic stem cells. DNA Cell Biol 2009; 28: 223–31.
91. Tate CM, Fishel M, Holleran J, Egorin J, Skalnik DG. Embryonic stem cells deficient in the epigenetic regulator CFP1 demonstrate hypersensitivity to DNA-damaging agents and decreased APE1 protein expression and endonuclease activity. DNA Repair 2009; 8: 1411–23.
92. Thomson JP, Skene PJ, Selfridge J, Coouaire T, Guy J, Webb S, Kerr ARW, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A. CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 2010; 464: 1082–6.
93. Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn MP, Shilatifard A. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol 2008; 28: 7337–44.
94. Tate CM, Lee J-H, Skalnik DG. CXXC finger protein 1 restricts the Setd1A histone H3-Lys4 methyltransferase complex to euchromatin. FEBS J 2010; 277: 210–23.
95. Tate CM, Lee J-H, Skalnik DG. CXXC finger protein 1 contains redundant functional domains that support embryonic stem cell cytosine methylation, histone methylation, and differentiation. Mol Cell Biol 2009; 29: 3817–31.
96. Shi X, Kachirskaia I, Walter KL, Kui J-HA, Lake A, Davrazou F, Chan SM, Martin DGE, Fingerman IM, Briggs SD, Howe L, Utz PJ, Kutateladze TG, Lugovskoy AA, Bedford MT, Gozani O. Proteome-wide analysis in *S. cerevisiae* identifies several PHD fingers as novel direct and selective binding modules of histone H3 methylated at either lysine 4 or lysine 36. J Biol Chem 2007; 282: 2450–5.
97. Chandrasekharan MB, Huang F, Chen YC, Sun ZW. Histone H2B C-terminal helix mediates trans-histone H3K4 methylation independent of H2B ubiquitination. Mol Cell Biol 2010; 30: 3216–32.
98. Couture J-F, Collazo E, Trievel RC. Molecular recognition of histone H3 by the WD40 protein WDR5. Nat Struct Mol Biol 2006; 13: 698–703.
99. Han Z, Guo L, Wang H, Shen Y, Deng XW, Chai J. Structural basis for the specific recognition of methylated histone H3 lysine 4 by the WD-40 protein WDR5. Mol Cell 2006; 22: 137–44.
100. Ruthenburg AJ, Wang W, Graybosch DM, Li H, Allis CD, Patel DJ, Verdine GL. Histone H3 recognition and presentation by the WDR5 module of the MLL1 complex. Nat Struct Mol Biol 2006; 13: 704–12.
101. Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Brurlingame AL, Roeder RG, Brivanlou AH, Allis CD. WDR5 associates with histone H3 methylated at K4 and is essential for H3K4 methylation and vertebrate development. Cell 2005; 121: 859–72.
102. Lee J-H, Skalnik DG. Wdr82 is a CTD-binding protein that recruits the Setd1A histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. Mol Cell Biol 2008; 28: 609–18.
103. Morris DP, Michelotti GA, Schwinn DA. Evidence that phosphorylation of the RNA polymerase II carboxyl-terminal repeats is similar in yeast and humans. J Biol Chem 2005; 280: 31368–77.
104. Narayanan A, Ruyechan WT, Kristie TM. The coactivator host cell factor-1 mediates Set1 and MLL1 H3K4 trimethylation at herpesvirus immediate early promoters for initiation of infection. Proc Natl Acad Sci USA 2007; 104: 10835–40.