Efficient differentiation of murine embryonic stem cells requires the binding of CXXC finger protein 1 to DNA or methylated histone H3-Lys4

Jyothi Mahadevan [a] and David G. Skalnik [a,b *]

[a] Department of Biochemistry and Molecular Biology, Indiana University School of Medicine, Indianapolis, IN 46202

[b] Biology Department, School of Science, Indiana University-Purdue University Indianapolis, Indianapolis, IN 46202

*Corresponding author: David Skalnik, 402 North Blackford St., LD 222, Indianapolis, IN 46202; phone, (317) 274-8510; FAX, (317) 274-0628; email, dskalnik@iupui.edu

Running title: Cfp1 is a histone-binding protein

Keywords: Chromatin, epigenetics, histones, cellular differentiation

Abbreviations: aa, amino acid; AP, alkaline phosphatase; Cfp1, CXXC finger protein 1; DAPI, 4',6-diamidino-2-phenylindone; EB, embryoid bodies; ES, embryonic stem; GST, glutathione S-transferase; LIF, leukemia inhibitory factor; PHD, plant homeodomain; WT, wild type;

Abstract

Mammalian CXXC finger protein 1 (Cfp1) is a DNA-binding protein that is a component of the Setd1 histone methyltransferase complexes and is a critical epigenetic regulator of both histone and cytosine methylation. Murine embryonic stem (ES) cells lacking Cfp1 exhibit a loss of histone H3-Lys4 tri-methylation (H3K4me3) at many CpG islands, and a mis-localization of this epigenetic mark to heterochromatic sub-nuclear domains. Furthermore, these cells fail to undergo cellular differentiation *in vitro*. These defects are rescued upon introduction of a Cfp1-expression vector. Cfp1 contains an N-terminal plant homeodomain (PHD), a motif frequently observed in chromatin associated proteins that functions as a reader module of histone marks. Here, we report that the Cfp1 PHD domain directly and specifically binds to histone H3K4me1/me2/me3 marks. Introduction of individual mutations at key Cfp1 PHD residues (Y28, D44, or W49) ablates this histone interaction both *in vitro* and *in vivo*. The W49A point mutation does not affect the ability of Cfp1 to rescue appropriate restriction of histone H3K4me3 to euchromatic sub-nuclear domains or *in vitro* cellular differentiation in Cfp1-null ES cells. Similarly, a mutated form of Cfp1 that lacks DNA-binding activity (C169A) rescues *in vitro* cellular differentiation. However, rescue of Cfp1-null ES cells with a double mutant form of Cfp1 (W49A, C169A) results in partially defective *in vitro* differentiation. These data define the Cfp1 PHD domain as a reader of histone H3K4me marks and provide evidence that this activity is involved in the regulation of lineage commitment in ES cells.

1. Introduction

Appropriate control of chromatin structure is critical for regulation of cell fate decisions in embryonic stem (ES) cells during the process of lineage commitment and cellular differentiation. Pluripotent ES cells display a genome-wide expansion of euchromatic domains, including tri-methylation of histone H3 at the lysine 4 position (H3K4me3) near gene promoters [1, 2]. Conversely, ES cells undergoing differentiation exhibit dynamic changes in chromatin structure from an open, accessible state to a more compact and silent form as evidenced by the accumulation of highly condensed heterochromatic domains in lineage-committed cells [3-5].

Mammalian Setd1 histone lysine methyltransferase complexes (Setd1a/Setd1b) are euchromatin-specific enzyme complexes that, in addition to the MLL complexes (MLL 1-4), catalyze the methylation of histone H3K4 [6, 7]. These complexes contain a catalytic SET [Su(var)3-9, Enhancer of Zeste, trithorax] domain and are analogous to the yeast Set1/COMPASS complex [8, 9]. The mammalian Setd1a and Setd1b complexes are identical with the exception of the catalytic components (Setd1a/Setd1b proteins). In addition to a unique catalytic subunit, the Setd1 and MLL histone H3K4 methyltransferase complexes all contain four common components; Ash2l, Wdr5, Dpy30, and Rbbp5 [10]. In addition, Setd1a/Setd1b complexes also contain Wdr82 and CXXC finger protein 1 (Cfp1) [7, 11].

Cfp1 is encoded by the *CXXC1* gene [12, 13]. This protein contains a 60 amino acid (aa) cysteine-rich CXXC domain that confers DNA-binding activity of Cfp1 specifically to sequences containing unmethylated CpG dinucleotides [13, 14]. Cfp1 also contains a Set1 interaction domain (SID) that is required for interaction with the Setd1 complexes, acidic, basic, and coiled-coil domains, and a HEAT domain that facilitates protein-protein interactions in other transcription factors [13, 15, 16]. Cfp1 also contains an N-terminal plant homeodomain (PHD). Plant homeodomains are typically found in proteins that function as chromatin modifiers and mediate transcriptional control [17-20]. The Cfp1 PHD domain (27-73 aa) conforms to a conserved Cys_4-His-Cys_3 motif [13].

Cfp1 is an important epigenetic regulator that is indispensable for vertebrate development. Deletion of the murine *CXXC1* gene results in failure of gastrulation and peri-implantation death at 4.5 to 6.5 days post coitus [21]. Moreover, post-gastrulation depletion of Cfp1 in zebrafish embryos abolishes primitive hematopoiesis [22]. Cfp1 is

also critical for murine hematopoiesis, as ablation of a conditional *CXXC1* allele in adult bone marrow cells leads to death within two weeks as a consequence of hematopoietic failure [23]. Murine *CXXC1*[-/-] ES cells are viable, but lack the ability to differentiate *in vitro* following removal of leukemia inhibitory factor (LIF) [24].

CXXC1[-/-] ES cells additionally exhibit epigenomic defects, including a marked reduction in global 5-methylcytosine levels due to impairment of maintenance DNA methylation, and an increase in global levels of histone H3K4me [11, 24]. In the absence of Cfp1, the Setd1a complex and its product, histone H3K4me3, are partially mis-localized to DAPI-bright heterochromatin as observed by confocal microscopy. This suggests that Cfp1 restricts the Setd1a complex to euchromatin [25]. Furthermore, chromatin immunoprecipitation studies revealed that Cfp1 is localized at ~80% of CpG islands (CGIs), and is required for deposition of histone H3K4me3 marks at CGIs [26] [27]. Moreover, absence of Cfp1 results in the appearance of ectopic histone H3K4me3 peaks at genomic regions devoid of CGIs or transcription start sites [27]. Together, these findings identify Cfp1 as a critical regulator of epigenetic modifications required to restrict Setd1/histone H3K4me3 to unmethylated CGIs located at gene promoters.

The pleiotropic defects observed in *CXXC1*[-/-] ES cells are rescued by introduction of a Cfp1 expression vector, thus providing a useful model system with which to probe structure/function relationships of Cfp1. Using this approach, it was found that Cfp1 possesses redundant functional domains [28]. For example, the C169A mutation in the Cfp1 DNA-binding domain abolishes interaction with DNA [14], whereas the C375A mutation in the Cfp1 SID domain ablates interaction with the Setd1 complex [29]. Introduction of either of these point mutations individually within Cfp1 does not affect the rescue of ES cell differentiation. However, introduction of both of these mutations into Cfp1 leads to a loss of rescue activity towards cellular differentiation. Therefore, the retention of either the DNA-binding activity or association with the Setd1 complexes is necessary for Cfp1 to support ES cell differentiation, suggesting that these domains have redundant function [28]. In contrast, rescue of appropriate sub-nuclear localization of H3K4me3 to euchromatin, as assessed by confocal microscopy, requires both the CXXC and the SID domains of Cfp1.

Surprisingly, the DNA-binding deficient form of Cfp1 (C169A) rescues appropriate histone H3K4me3 levels at most CGIs in *CXXC1*[-/-] ES cells, but does not rescue the

inappropriate occurrence of ectopic histone H3K4me3 peaks outside of CGIs [27]. Thus, Cfp1 DNA-binding activity is not necessary for the targeting of histone H3K4me3 to CGIs, and other Cfp1 domain(s) may perform this function [27]. Importantly, Spp1 (the yeast analog of Cfp1) contains a PHD domain that binds to histone H3K4me2/3 [30]. Studies on other PHD-containing proteins such as Ing2 and Bptf have revealed that PHD domains function as effectors of H3K4me2/3 and play a role in modulating gene expression [31, 32]. However, the specific histone-binding properties of Cfp1 and their underlying function(s) are largely unknown. Experiments were conducted to determine the histone-binding properties of the mammalian Cfp1 PHD domain and analyze the functional significance of this activity.

2. Results

2.1 The Cfp1 PHD domain exhibits histone-binding activity

Sequence comparison of the N-terminal PHD domain of Cfp1 to similar domains in other known histone H3K4me2/3 binders such as Phf2, Ing2, and Bptf [31-33] revealed the conservation of residues Y28, D44, and W49 previously found to be critical for binding to histone H3K4me3 (indicated by boxes in Fig. 1A). GST pull-down assays were performed to assess the histone-binding properties of the Cfp1 PHD domain. When incubated with a mixture of calf thymus histones, a bacterially expressed GST-Cfp1 PHD domain (1-102 aa) fusion protein interacts with histone H3K4me1, H3K4me2, and H3K4me3, but not with H3K9me3 (a modification associated with heterochromatin) (Fig. 1B). This finding was further validated by performing histone peptide pull-down assays using biotin-labelled histone H3 N-terminal tail peptides (1-21 aa) that were variably methylated. The GST-Cfp1 PHD fusion protein was specifically retained by biotinylated histone tail peptides H3K4me1, H3K4me2, and H3K4me3, but not by H3K9me3 (Fig. 1C). These results reveal a direct interaction between the amino terminal domain of Cfp1 and histone H3 methylated at the K4 position.

2.2 Mutations in the Cfp1 PHD domain ablate its binding to histone H3K4me3

Site-directed mutagenesis was performed on the Cfp1 PHD domain to assess the structural requirements for *in vitro* binding activity to histone H3K4me3. Individual mutation of Y28, D44, or W49 to alanine led to a dramatic reduction in the binding of the GST-Cfp1 (1-102 aa) domain to biotin-tagged histone H3K4me3 peptide (Fig. 2A).

In order to assess the histone-binding properties of Cfp1 *in vivo*, wild type (WT) or mutated forms of Flag-tagged full length Cfp1 (1-656 aa) were transiently expressed in HEK-293 cells, followed by immunoprecipitation using Flag-M2 conjugated agarose beads and immunoblotting for the histone H3K4me3 mark. WT Cfp1 co-immunoprecipitated endogenous histone H3K4me3, but this *in vivo* interaction was abolished for the Y28A, D44A, and W49A forms of full-length Cfp1 (Fig. 2B). Together, these data indicate that the Cfp1 PHD domain is necessary and sufficient for the binding of Cfp1 to histone H3K4me3 both *in vitro* and *in vivo*.

2.3 The histone-binding activity of Cfp1 is dispensable for appropriate restriction of histone H3K4me3 to euchromatin sub-nuclear domains in ES cells

ES cells that lack Cfp1 ($CXXC1^{-/-}$) exhibit an inappropriate drifting of histone H3K4me3 to 4', 6-diamidino-2-phenylindole (DAPI)-bright heterochromatin, suggesting that Cfp1 is required to restrict the Set1 histone methyltransferase complexes to euchromatic CpG islands. This defect in sub-nuclear targeting of the histone H3K4me3 mark is rescued by expression of Cfp1 in $CXXC1^{-/-}$ ES cells, but not by mutated forms of Cfp1 that lack either DNA-binding activity (C169A) or the ability to associate with the Set1 complex (C375A) [25].

To determine if the interaction of Cfp1 with methylated histone H3K4 is necessary for appropriate sub-nuclear localization of histone H3K4me3, $CXXC1^{-/-}$ ES cells were stably transfected with an expression vector encoding either WT Flag-tagged Cfp1 or Cfp1 containing W49A mutation (or the empty expression vector) (Fig. 3A). Individual clones were screened for transgene expression by immunoblotting using Cfp1 antiserum (Fig. 3B). Previous studies indicated that ES cells heterozygous for the disrupted CXXC allele ($CXXC1^{+/-}$) do not show a phenotype despite expressing ~50% of the WT level of Cfp1 protein [24]. Hence, stably transfected ES clones expressing at least 50% of the WT level of Cfp1 were analyzed by confocal microscopy (Fig. 3B).

Similar to previous observations [25], the percent co-localization of histone H3K4me3 with DAPI-bright heterochromatin was significantly increased in $CXXC1^{-/-}$ ES cells (22.1%) and $CXXC1^{-/-}$ ES cells carrying the empty expression vector (21.9%) compared to that found in $CXXC1^{+/+}$ ES cells (3.5%) or $CXXC1^{-/-}$ ES cells rescued with the WT Cfp1 expression vector (3.3%). Importantly, a rescue of proper restriction of histone H3K4me3 to euchromatic sub-nuclear domains was observed in $CXXC1^{-/-}$ ES cells expressing the W49A form of Cfp1 (4.6%) (Fig. 3C). These results demonstrate that the PHD domain-mediated interaction of Cfp1 with histone H3K4me3 is not required for appropriate restriction of histone H3K4me3 to DAPI-dim euchromatic sub-nuclear domains.

2.4 Interaction of Cfp1 with DNA or methylated histone H3K4 is necessary for efficient differentiation of ES cells

$CXXC1^{-/-}$ ES cells fail to differentiate *in vitro* when cultured in growth medium lacking leukemia inhibitory factor (LIF), as evidenced by colony morphology and persistent expression of stem cell markers such as alkaline phosphatase (AP) and Oct4. This defect can be reversed by introduction of a Cfp1 expression vector [24] (Fig. 4 and 5). To determine if the interaction of Cfp1 with methylated histone H3K4 is necessary for the rescue of cellular differentiation, $CXXC1^{-/-}$ ES cells expressing the W49A mutated form of Cfp1 were cultured in growth medium lacking LIF for a period of 10 days. These cells produced robust embryoid bodies (EBs) and subsequently formed extensive outgrowths characteristic of endoderm formation, indistinguishable from similarly cultured $CXXC1^{+/+}$ ES cells and $CXXC1^{-/-}$ ES cells expressing WT Cfp1 (Fig. 4). In each case, greater than 95% of cells down-regulated AP activity (Fig. 5). These results indicate that a mutated form of Cfp1 that is deficient in binding to methylated histone H3K4 supports cellular differentiation of $CXXC1^{-/-}$ ES cells.

Previous studies revealed that a mutated form of Cfp1 deficient in DNA-binding activity (C169A) and a mutated form of Cfp1 that fails to interact with the Set1 complexes (C375A) are both able to rescue cellular differentiation when expressed in $CXXC1^{-/-}$ ES cells. However, a mutated form of Cfp1 containing both of these point mutations lacks rescue activity. Thus, while neither of these domains are required for rescue activity, retention of one of these domains is required, suggesting that these domains exhibit redundant activity in support of rescue activity [28]. To determine if a similar functional redundancy exists between the Cfp1 PHD and DNA-binding domains, the W49A and C169A mutations were introduced into full-length Cfp1. $CXXC1^{-/-}$ ES cells stably expressing doubly mutated Cfp1 (W49A, C169A) were induced to differentiate by removal of LIF. Ten days after the withdrawal of LIF, examination of morphology revealed that only a few EBs could adhere and form outgrowths that remained small over time (Fig. 4). This finding was strikingly different from the well-differentiated morphology of $CXXC1^{+/+}$ ES cells and $CXXC1^{-/-}$ ES cells expressing WT Cfp1 or the W49A or C169A mutated forms of Cfp1. This indicates that $CXXC1^{-/-}$ ES cells expressing doubly mutated Cfp1 (W49A, C169A) are unable to undergo efficient differentiation. Consistent with the EB morphology results, 41% of these ES cells

continued to express AP activity 10 days following removal of LIF, indicating a partial failure of cellular differentiation (Fig. 5).

3. Discussion

The results reported here demonstrate that the mammalian Cfp1 PHD domain interacts with methylated histone H3K4 (H3K4me1/me2/me3). The Cfp1 PHD domain, expressed as a fusion with GST in bacterial cells, binds to methylated H3K4 histone tail peptides, indicating that this interaction is direct. Moreover, the Cfp1 PHD domain fails to recognize histone H3K9me3, a heterochromatin mark, illustrating the specificity of the observed interaction. This is consistent with a recent report that the PHD domain of murine Cfp1 interacts specifically with methylated histone H3K4 *in vitro* [34]. Transiently expressed Flag-tagged Cfp1 co-immunoprecipitates endogenous histone H3K4me3 from nuclear extracts, indicating that this interaction occurs *in vivo* within the nuclei of cells. Thus, Cfp1, a component of the euchromatin-specific Setd1 histone methyltransferase complexes that catalyze the deposition of histone H3K4me marks, also interacts with this epigenetic mark. Absence of Cfp1 results in a loss of histone H3K4me3 at CGIs and this defect is rescued by expression of Cfp1 [26, 27]. These data suggest that Cfp1 may participate in a feed-forward mechanism that facilitates the maintenance of these marks at CGIs to reinforce an active chromatin status. Interestingly, other epigenetic regulatory complexes adopt similar mechanisms for mediating activation or repression of specific genomic loci. For instance, Wdr5, a common component of the MLL and Setd1 complexes, functions as an effector protein by binding to nucleosomes containing histone H3K4me2 and facilitating global and gene specific deposition of histone H3K4me3 [35]. Another example is HP1, which not only associates with the histone H3K9me3 methyltransferase Suv39H1, but also recognizes the methylated H3K9 mark through its amino-terminal chromodomain, thus promoting the spread of heterochromatin domains [36, 37]. Additionally, Pygo2 contains a PHD domain that recognizes histone H3K4me2/3, and this interaction is critical to mammary progenitor cell expansion. Pygo2 also facilitates the generation of these marks by physically interacting with Wdr5 at Wnt-specific loci and in bulk chromatin, thereby creating a positive feedback loop for the establishment of active chromatin domains [38].

The Y28, D44 and W49 residues of the Cfp1 PHD domain are conserved in many histone H3K4me3-binding proteins such as Bptf, Phf2, and Ing2 [31-33]. Of these, Y28 and W49 of the Cfp1 PHD correspond to Y10 and W32 of the Bptf PHD and are involved in formation of the aromatic amino acid cage that securely positions the tri-methyl group of histone H3K4me3 [32]. Corresponding amino acids Y7/W29 present in Phf2, and

Y215/W238 present in Ing2 establish contacts with the tri-methylammonium group of lysine 4 [33, 39]. The point mutations Y28A, D44A, and W49A in Cfp1 PHD and full-length Cfp1 resulted in ablation of their association with histone H3K4me3. This suggests that only the N-terminal PHD domain of Cfp1 is sufficient for binding to methylated H3K4. Moreover, it is also likely that the tri-methyl group of histone H3K4me3 is stabilized by aromatic residues Y28 and W49, similar to that in other known H3K4me3-binding proteins.

Cfp1 is required to restrict the sub-nuclear localization of Setd1 and histone H3K4me3 to euchromatin domains [25]. Moreover, Cfp1 also regulates the levels of the H3K4me3 mark at many CGI-rich gene promoters, and the DNA-binding function of Cfp1 is surprisingly not necessary for this function [27]. The study reported here used confocal microscopy to demonstrate that the histone-binding activity of Cfp1 is not necessary for the restriction of histone H3K4me3 to DAPI-dim euchromatin in ES cells. However, it will be important to determine if the binding of Cfp1 to methylated H3K4 is required for the deposition and maintenance of histone H3K4me3 at CGI-rich promoters.

Previous data indicated that retention of either the DNA-binding activity of Cfp1 or interaction with the Setd1 complexes is necessary for rescue of *in vitro* differentiation of *CXXC1*[-/-] ES cells, suggesting that the CXXC and SID domains of Cfp1 are redundant functional domains [28]. Similarly, the data reported in this study reveals that ablation of the interaction between Cfp1 and methylated histone H3K4 does not affect the ability to rescue cellular differentiation of ES cells. However, disruption of both the histone H3K4me-binding and DNA-binding activities of Cfp1 results in compromised rescue of cellular differentiation. Following induction of differentiation, *CXXC1*[-/-] ES cells expressing Cfp1 containing point mutations in both PHD and CXXC domains fail to form extensive outgrowths and ~40% of these cells are unable to down-regulate AP activity, indicating a failure to undergo lineage commitment. Furthermore, RT-PCR analysis of this cell population following induction of differentiation detects up-regulation of lineage commitment markers including Gata-4 (endoderm), Gp-IIb (megakaryocyte), c-fms (myeloid), MHC-β (cardiac), and β-H1 (primitive erythroid) as well as persistent expression of the pluripotency marker Oct4 (unpublished observation, J.M.). Presumably the AP-negative cells in the population express these markers of cellular differentiation, while the AP-positive cells continue to express Oct4. These data indicate that although neither the PHD nor the CXXC domains of Cfp1 is required to support *in vitro*

differentiation of ES cells, these domains exhibit redundant function, and provide functional significance to the histone-binding property of the Cfp1 PHD domain.

Previous data revealed that the C-terminal half of Cfp1 (361-656 aa) is sufficient for the rescue of *in vitro* differentiation in *CXXC1*[-/-] ES cells [28]. However, the data reported here suggests that either the PHD domain or the CXXC domain is required to rescue defective ES cell differentiation in the context of full-length Cfp1. Both the PHD and CXXC domains are present in the N-terminal half of Cfp1. The fact that the C-terminal half of Cfp1, which lacks the PHD and CXXC domains, effectively rescues cellular differentiation could possibly be due to folding differences in truncated proteins compared to the full-length protein. Another possibility could be that the C-terminal half of Cfp1 lacks an uncharacterized 'inhibitory' domain that inhibits rescue activity.

Appropriate genomic targeting of the Setd1 histone methyltransferases involves intricate contributions from multiple chromatin-binding activities embedded within the subunits of each methyltransferase complex. For example, the Setd1 complexes contain Wdr82 that is lacking in the other COMPASS-like complexes (MLL1-4) [7, 11, 40]. Wdr82 interacts with serine 5-phosphorylated C-terminal domain repeats of RNA polymerase II and recruits the Setd1a complex to the transcriptional start sites of active genes [41]. Further, mono-ubiquitination of histone H2B is required for association of the yeast analog of Wdr82 (Cps35) to chromatin and the subsequent tri-methylation of histone H3K4 [42]. Cfp1 is another component that is restricted to the Setd1 members of the histone H3K4 methyltransferase family, and the Cfp1 DNA-binding domain may contribute to the targeting of these complexes to unmethylated CGIs and deposition of the histone H3K4me3 mark at these sites. In addition, Ash2l, a core component present in all the mammalian histone H3K4 methyltransferase complexes, binds DNA via its winged helix motif and this binding promotes histone H3K4 methylation [43, 44]. Another shared core component, Wdr5, associates with chromatin by interacting with methylated histone H3K4 to catalyze tri-methylation of K4 [35]. Finally, despite sharing these numerous chromatin-binding components, the Setd1a and Setd1b complexes exhibit non-overlapping sub-nuclear distributions within euchromatin regions as observed by confocal immunofluorescence [7]. The Setd1a and Setd1b enzymes contain highly divergent central domains that may confer distinct genomic targeting behavior. Indeed, co-immunoprecipitation studies revealed distinct interactions of these central domains with putative targeting molecules [45]. The finding that Cfp1 exhibits histone-binding

activity in addition to DNA-binding activity adds another layer of complexity to the regulation of the Setd1 methyltransferases.

4. Experimental Procedures

4.1 Cell culture

Human Embryonic Kidney-293 (HEK-293) cell line was grown in Dulbecco's Modified Eagle's medium (Gibco BRL Life Technologies, Grand Island, NY) enriched with 10% Hyclone Fetal Clone III Serum (GE Healthcare, United Kingdom). ES cells were cultured on tissue culture dishes coated with 0.1% gelatin in growth media as described [28]. Generation of the murine $CXXC1^{-/-}$ ES cell line has been previously described [24]. Analysis of *in vitro* differentiation was performed as described [24] in the absence of LIF-conditioned media. Histochemical staining for alkaline phosphatase activity was performed using leukocyte alkaline phosphatase kit (Sigma Aldrich, St. Louis, MO) as described [24].

4.2 Sub-cloning of plasmids

A cDNA fragment encoding Cfp1 (1-102 aa) was sub-cloned into pGEX-4T1, a bacterial vector system for expression of GST-tagged fusion proteins. A cDNA encoding full-length Cfp1 (1-656 aa) was sub-cloned into pcDNA3.1/Hygro vector system (Invitrogen, Carlsbad, CA) for expression in mammalian cell line HEK-293 and $CXXC1^{-/-}$ ES cells. Conserved residues within Cfp1 (Y28, D44, and W49) thought to be important for histone H3K4me3 binding were replaced by alanines using a QuickChange II Site Directed Mutagenesis kit (Agilent Technologies, Santa Clara, CA) as per the manufacturer's instructions using the following primers:

Y28A F: 5' GAGAATGCGCCCATCGCATGCATCTGCCGCAAAC 3'
Y28A R: 5' GTTTGCGGCAGATGCATGCGATGGGCGCATTCTC 3'
D44A F: 5' CTTCATGATCGGGTGTGCCAACTGCAATGAGTGG 3'
D44A R: 5' CCACTCATTGCAGTTGGCACACCCGATCATGAAG 3'
W49A F: 5' GTGTGACAACTGCAATGAGGCCTTCCATGGGGACTGCATCC 3'
W49A R: 5' GGATGCAGTCCCCATGGAAGGCCTCATTGCAGTTGTCACAC 3'

4.3 Transient and stable transfection

Transient transfection of pcDNA 3.1/ Hygro vector expressing Flag-Cfp1 (1-656 aa) was performed in HEK-293 cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol. In order to stably express WT or mutated Cfp1 in $CXXC1^{-/-}$ ES cells, 25 µg of each plasmid was linearized using the restriction enzyme Nru1 (New England Biolabs, Ipswich, MA) and electroporated into these cells. ES growth medium supplemented with 200 µg/ml hygromycin B was replaced every other day for fourteen days until individual colonies were isolated and expanded. These clones were assayed for expression of Cfp1 by western blotting using an antiserum against Cfp1 [7]. Only clones expressing Cfp1 protein > 50% of WT Cfp1 levels were chosen for further study.

4.4 GST protein production

GST-tagged WT and mutated Cfp1 PHD (1-102 aa) were prepared by inducing 250 ml bacterial cultures (final OD_{600}: 0.6-1.0) with 0.1mM IPTG for 3 hours. Lysis buffer containing 25 mM Tris pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100, and cOMPLETE protease inhibitor tablets (Roche, Indianapolis, IN) was used to lyse pelleted bacterial cells. This lysate was sonicated three times for 30 s each to shear bacterial DNA. After high speed centrifugation (15000 rpm for 10 min), the supernatant was incubated with GSH-Agarose beads (Sigma Aldrich). The GST proteins were eluted using reduced glutathione (Sigma Aldrich) and dialyzed in 1X PBS three times before use in binding assays.

4.5 Pull down assays/ Immunoprecipitations

4.5.1 GST pulldown

Two micrograms of purified GST-Cfp1 PHD was incubated with 1 µg calf thymus total histones (Worthington, Lakewood, NJ) in binding buffer (50 mM Tris-HCl 7.5, 1 M NaCl, 1% NP-40, 0.5 mM EDTA, 1 mM PMSF, and 1X Roche cOMPLETE protease inhibitor) at 4ºC overnight followed by 1 hour incubation with glutathione agarose beads. Proteins bound to beads were denatured by 1X Laemmli sample buffer and subjected to western

blotting using antibodies against total histone H3, histone H3K4me1/me2/me3, and histone H3K9me3 (Abcam, Cambridge, MA).

4.5.2 Histone peptide pulldown

One-to-two micrograms of biotinylated, N-terminal tail peptides (1-21 aa) of histone H3K4me1/me2/me3 or histone H3K9me3 (Anaspec EGT, Fremont, CA) was incubated overnight with 2 µg purified GST-Cfp1 PHD in a binding buffer composed of 50 mM Tris pH 7.5, 300 mM NaCl and 0.05% NP-40. Streptavidin sepharose beads (GE Healthcare Lifesciences) were then added for 1 hour to immobilize the biotinylated histone peptides and pulldown the interacting Cfp1 PHD domain. After extensive washing, western blotting was performed on the eluates using an antibody directed against the GST tag (Thermo Scientific, Waltham, MA).

4.5.3 Flag Immunoprecipitation

Nuclear extracts were prepared from HEK-293 cells transiently expressing WT or mutated forms of Flag-Cfp1 (1-656 aa) using benzonase nuclease (Millipore, Billerica, MA). The soluble nuclear extract was incubated with Flag M2-conjugated agarose (Sigma Aldrich). Further, these beads were washed in cytoskeletal buffer [46], boiled in 1X Laemmli sample buffer and analyzed by western blot for the endogenous histone H3K4me3 mark.

4.6 Confocal microscopy

ES cells (2×10^4) were seeded onto sterile coverslips placed on each well of a 24 well tissue culture dish and grown for 2 days. ES medium was removed and cells were washed twice with cold 1X PBS. Cells were fixed using 4% paraformaldehyde for 20 min at room temperature followed by three washes with cold 1X PBS. Cells were then permeabilized in PBS-T (0.5% Triton X-100 in PBS) for 10 min followed by a single wash with 1X PBS. Blocking buffer (1% BSA and 5% Normal Donkey Serum (Santa Cruz Biotechnologies, Santa Cruz, CA) in PBS-T was prepared fresh and added for 1 hour at room temperature with gentle rocking. After a wash with 1X PBS, the primary antibody for histone H3K4me3 (Abcam) was diluted in blocking buffer (1:100) and added to the cells for overnight incubation at 4ºC. The next day, cells were washed three times in 1X

PBS and incubated with FITC-conjugated secondary antibody (Santa Cruz) diluted 1:200 in blocking buffer for 2 hours at room temperature. After 3 washes with 1X PBS, cells were stained with 0.1 μg/ml of 4', 6-diamidino-2-phenylindole (DAPI) solution for 20 min at room temperature. Coverslips were washed once in H_2O and twice in 1X PBS before mounting on slides with Vectashield mounting medium (Vector Laboratories, Burlingame, CA). A Zeiss LSM 700 confocal imaging system (Department of Ophthalmology, Eugene and Marilyn Glick Eye Institute) was used to observe staining patterns of nuclei. This system utilized excitation wavelengths of 405 nm and 488 nm for visualizing DAPI staining and FITC staining respectively. Images obtained were analyzed using MetaMorph v7 (Molecular Devices, Sunnyvale, CA) with similar thresholding values for each image. Thresholding for DAPI was done so as to include DAPI bright regions whereas for FITC, thresholding was performed to exclude background and non-specific staining. Percent co-localization of histone H3K4me3 and DAPI was determined for 7-37 nuclei for every cell line using the MetaMorph co-localization module.

4.7 Statistical analysis

Two-tailed Student's t-tests assuming equal variance were used for significance testing. A difference of p value < 0.05 was considered as statistically significant.

FIGURE LEGENDS

Figure 1. The PHD domain of Cfp1 binds histone H3K4me1/me2/me3.

A) Sequence comparison of the Cfp1 PHD domain and similar domains in other known histone H3K4me3-binders. Conserved residues important for binding to histone H3K4me3 are enclosed in boxes. Phf2, PHD finger protein 2; Ing2, Inhibitor of growth protein 2; and Bptf, Bromodomain PHD finger transcription factor. Sequence alignment was performed using the Clustal O program.

B) Purified GST-Cfp1 PHD (1-102 aa) or GST alone was incubated with a mixture of calf thymus histones followed by addition of GSH agarose beads to immobilize the formed complex. Bound histone proteins were analyzed by immunoblotting (IB) using antisera against mono-, di-, or tri-methylated histone H3K4, tri-methylated histone H3K9, or total histone H3. Representative immunoblots are shown from three independent experiments.

C) Histone peptide pull-downs were performed by incubating purified GST-Cfp1 PHD (1-102 aa) with N-terminal, biotin-labeled histone H3 peptides (1-21 aa) carrying the indicated modifications, followed by addition of streptavidin sepharose beads. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum. Representative immunoblots are shown from three independent experiments.

Figure 2. Mutations in the Cfp1 PHD domain ablate histone H3K4me3-binding activity.

A) WT or the indicated mutated forms of the Cfp1 PHD domain (1-102 aa) were expressed as GST fusion proteins and incubated with biotin-labeled histone H3K4me3 peptide (1-21 aa), followed by pull-down (PD) with streptavidin sepharose. Bound proteins were analyzed by immunoblotting (IB) with GST antiserum. Representative immunoblots are shown from three independent experiments.

B) Nuclear extracts prepared from HEK-293 cells transiently expressing wild-type or mutated Flag-tagged Cfp1 (1-656 aa) were subjected to Flag immunoprecipitation. Binding of Flag-Cfp1 to histone H3K4me3 was analyzed by immunoblotting with histone H3K4me3 antiserum. Representative immunoblots are shown from two independent experiments.

Figure 3. Binding of Cfp1 to methylated histone H3K4 is dispensable for appropriate restriction of histone H3K4me3 to euchromatic sub-nuclear domains in ES cells.

A) A schematic representation of WT and mutated forms of Cfp1 that were stably expressed in $CXXC1^{-/-}$ ES cells. Each "X" indicates the approximate location of the indicated point mutation within Cfp1. The black circle at the N-terminus of Cfp1 represents the Flag tag. The nomenclature of each of the stably expressed constructs is represented on the right of the respective schematic.

B) Immunoblot analysis was performed on whole cell lysates obtained from stably expressing ES clones using Cfp1 antiserum to determine the protein expression level of Cfp1, which was compared to Cfp1 expression in WT ES cells. The membrane was re-probed with β-actin antiserum as a loading control. A representative immunoblot from three independent experiments is shown.

C) Sub-nuclear localization of endogenous histone H3K4me3 was detected in $CXXC1^{+/+}$, $CXXC1^{-/-}$, and $CXXC1^{-/-}$ ES cells expressing WT or W49A Cfp1 (or carrying the empty expression vector), using a primary rabbit antibody (anti-histone H3K4me3) and bovine anti-rabbit IgG-fluorescein isothiocyanate (FITC)-conjugated secondary antibody. After counterstaining with DAPI, the nuclei were examined by confocal microscopy. Yellow regions in the merge and co-localized images indicate co-localization. The numbers within brackets indicate the percent co-localization for the illustrated nucleus, calculated using the MetaMorph co-localization module. The numbers shown below indicate the average percent co-localization (+/- standard error) of histone H3K4me3 with DAPI-bright heterochromatin. Between 7 and 37 nuclei were examined for each ES cell clone. Asterisks represent statistically significant difference ($p<0.05$) compared to $CXXC1^{+/+}$ ES cells. Similar results were obtained using two independent ES clones carrying each construct.

Figure 4. Interaction of Cfp1 with histone H3K4me or DNA is necessary for formation of extensive EB outgrowths.

$CXXC1^{+/+}$, $CXXC1^{-/-}$, and $CXXC1^{-/-}$ ES cells expressing WT Cfp1 or the indicated Cfp1

mutations (or carrying the empty expression vector) were grown in ES media lacking LIF. Morphology of ES cell colonies was studied by microscopy ten days after induction of differentiation (10X magnification). Similar results were obtained using three independent ES clones for each construct.

Figure 5. Interaction of Cfp1 with histone H3K4me or DNA is necessary for efficient down-regulation of alkaline phosphatase activity during *in vitro* differentiation

$CXXC1^{+/+}$, $CXXC1^{-/-}$, and $CXXC1^{-/-}$ ES cells expressing WT Cfp1 or the indicated mutated forms of Cfp1 (or carrying the empty expression vector) were grown in ES media lacking LIF. After a period of ten days, cells were collected from bacterial culture plates and trypsinized into a single cell suspension. These cells were re-seeded into tissue culture dishes coated with gelatin and were stained for alkaline phosphatase activity using a leukocyte alkaline phosphatase kit. Images were captured at 10X magnification. Cells containing reddish deposits were scored positive for alkaline phosphatase activity. A minimum of 300 cells were scored for each cell line. The numbers within the parentheses indicate the percentage of alkaline phosphatase-positive cells for the indicated cell line. Similar results were obtained using three independent ES clones for each construct.

Acknowledgements

REFERENCES

1.	Guenther, M.G., S.S. Levine, L.A. Boyer, R. Jaenisch, and R.A. Young, *A chromatin landmark and transcription initiation at most promoters in human cells.* Cell, 2007. **130**(1): p. 77-88.
2.	Meissner, A., T.S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B.E. Bernstein, C. Nusbaum, D.B. Jaffe, A. Gnirke, R. Jaenisch, and E.S. Lander, *Genome-scale DNA methylation maps of pluripotent and differentiated cells.* Nature, 2008. **454**(7205): p. 766-70.
3.	Meshorer, E., D. Yellajoshula, E. George, P.J. Scambler, D.T. Brown, and T. Misteli, *Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells.* Dev Cell, 2006. **10**(1): p. 105-16.
4.	Chen, T. and S.Y. Dent, *Chromatin modifiers and remodellers: regulators of cellular differentiation.* Nat Rev Genet, 2014. **15**(2): p. 93-106.
5.	Meshorer, E. and T. Misteli, *Chromatin in pluripotent embryonic stem cells and differentiation.* Nat Rev Mol Cell Biol, 2006. **7**(7): p. 540-6.
6.	Jenuwein, T., G. Laible, R. Dorn, and G. Reuter, *SET domain proteins modulate chromatin domains in eu- and heterochromatin.* Cell Mol Life Sci, 1998. **54**(1): p. 80-93.
7.	Lee, J.H., C.M. Tate, J.S. You, and D.G. Skalnik, *Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex.* J Biol Chem, 2007. **282**(18): p. 13419-28.
8.	Dillon, S.C., X. Zhang, R.C. Trievel, and X. Cheng, *The SET-domain protein superfamily: protein lysine methyltransferases.* Genome Biol, 2005. **6**(8): p. 227.
9.	Shilatifard, A., *The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis.* Annu Rev Biochem, 2012. **81**: p. 65-95.
10.	Herz, H.M., A. Garruss, and A. Shilatifard, *SET for life: biochemical activities and biological functions of SET domain-containing proteins.* Trends Biochem Sci, 2013. **38**(12): p. 621-39.
11.	Lee, J.H. and D.G. Skalnik, *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**(50): p. 41725-31.
12.	Carlone, D.L., S.R. Hart, P.D. Ladd, and D.G. Skalnik, *Cloning and characterization of the gene encoding the mouse homologue of CpG binding protein.* Gene, 2002. **295**(1): p. 71-7.
13.	Voo, K.S., D.L. Carlone, B.M. Jacobsen, A. Flodin, and D.G. Skalnik, *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**(6): p. 2108-21.
14.	Lee, J.H., K.S. Voo, and D.G. Skalnik, *Identification and characterization of the DNA binding domain of CpG-binding protein.* J Biol Chem, 2001. **276**(48): p. 44669-76.
15.	Groves, M.R., N. Hanlon, P. Turowski, B.A. Hemmings, and D. Barford, *The structure of the protein phosphatase 2A PR65/A subunit reveals the conformation of its 15 tandemly repeated HEAT motifs.* Cell, 1999. **96**(1): p. 99-110.
16.	Perry, J. and N. Kleckner, *The ATRs, ATMs, and TORs are giant HEAT repeat proteins.* Cell, 2003. **112**(2): p. 151-5.
17.	Aasland, R., T.J. Gibson, and A.F. Stewart, *The PHD finger: implications for chromatin-mediated transcriptional regulation.* Trends Biochem Sci, 1995. **20**(2): p. 56-9.

18. Sanchez, R. and M.M. Zhou, *The PHD finger: a versatile epigenome reader.* Trends Biochem Sci, 2011. **36**(7): p. 364-72.

19. Wysocka, J., T. Swigut, H. Xiao, T.A. Milne, S.Y. Kwon, J. Landry, M. Kauer, A.J. Tackett, B.T. Chait, P. Badenhorst, C. Wu, and C.D. Allis, *A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling.* Nature, 2006. **442**(7098): p. 86-90.

20. O'Connell, S., L. Wang, S. Robert, C.A. Jones, R. Saint, and R.S. Jones, *Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein.* J Biol Chem, 2001. **276**(46): p. 43065-73.

21. Carlone, D.L. and D.G. Skalnik, *CpG binding protein is crucial for early embryonic development.* Mol Cell Biol, 2001. **21**(22): p. 7601-6.

22. Young, S.R., C. Mumaw, J.A. Marrs, and D.G. Skalnik, *Antisense targeting of CXXC finger protein 1 inhibits genomic cytosine methylation and primitive hematopoiesis in zebrafish.* J Biol Chem, 2006. **281**(48): p. 37034-44.

23. Chun, K.T., B. Li, E. Dobrota, C. Tate, J.H. Lee, S. Khan, L. Haneline, H. HogenEsch, and D.G. Skalnik, *The epigenetic regulator CXXC finger protein 1 is essential for murine hematopoiesis.* PLoS One, 2014. **9**(12): p. e113745.

24. Carlone, D.L., J.H. Lee, S.R. Young, E. Dobrota, J.S. Butler, J. Ruiz, and D.G. Skalnik, *Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein.* Mol Cell Biol, 2005. **25**(12): p. 4881-91.

25. Tate, C.M., J.H. Lee, and D.G. Skalnik, *CXXC finger protein 1 restricts the Setd1A histone H3K4 methyltransferase complex to euchromatin.* FEBS J, 2010. **277**(1): p. 210-23.

26. Thomson, J.P., P.J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A.R. Kerr, A. Deaton, R. Andrews, K.D. James, D.J. Turner, R. Illingworth, and A. Bird, *CpG islands influence chromatin structure via the CpG-binding protein Cfp1.* Nature, 2010. **464**(7291): p. 1082-6.

27. Clouaire, T., S. Webb, P. Skene, R. Illingworth, A. Kerr, R. Andrews, J.H. Lee, D. Skalnik, and A. Bird, *Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells.* Genes Dev, 2012. **26**(15): p. 1714-28.

28. Tate, C.M., J.H. Lee, and D.G. Skalnik, *CXXC finger protein 1 contains redundant functional domains that support embryonic stem cell cytosine methylation, histone methylation, and differentiation.* Mol Cell Biol, 2009. **29**(14): p. 3817-31.

29. Butler, J.S., J.H. Lee, and D.G. Skalnik, *CFP1 interacts with DNMT1 independently of association with the Setd1 Histone H3K4 methyltransferase complexes.* DNA Cell Biol, 2008. **27**(10): p. 533-43.

30. Shi, X., I. Kachirskaia, K.L. Walter, J.H. Kuo, A. Lake, F. Davrazou, S.M. Chan, D.G. Martin, I.M. Fingerman, S.D. Briggs, L. Howe, P.J. Utz, T.G. Kutateladze, A.A. Lugovskoy, M.T. Bedford, and O. Gozani, *Proteome-wide analysis in Saccharomyces 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**(4): p. 2450-5.

31. Shi, X., T. Hong, K.L. Walter, M. Ewalt, E. Michishita, T. Hung, D. Carney, P. Pena, F. Lan, M.R. Kaadige, N. Lacoste, C. Cayrou, F. Davrazou, A. Saha, B.R. Cairns, D.E. Ayer, T.G. Kutateladze, Y. Shi, J. Cote, K.F. Chua, and O. Gozani, *ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression.* Nature, 2006. **442**(7098): p. 96-9.

32. Li, H., S. Ilin, W. Wang, E.M. Duncan, J. Wysocka, C.D. Allis, and D.J. Patel, *Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF.* Nature, 2006. **442**(7098): p. 91-5.

33.   Wen, H., J. Li, T. Song, M. Lu, P.Y. Kan, M.G. Lee, B. Sha, and X. Shi, *Recognition of histone H3K4 trimethylation by the plant homeodomain of PHF2 modulates histone demethylation.* J Biol Chem, 2010. **285**(13): p. 9322-6.

34.   Eberl, H.C., C.G. Spruijt, C.D. Kelstrup, M. Vermeulen, and M. Mann, *A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics.* Mol Cell, 2013. **49**(2): p. 368-78.

35.   Wysocka, J., T. Swigut, T.A. Milne, Y. Dou, X. Zhang, A.L. Burlingame, R.G. Roeder, A.H. Brivanlou, and C.D. Allis, *WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development.* Cell, 2005. **121**(6): p. 859-72.

36.   Elgin, S.C. and S.I. Grewal, *Heterochromatin: silence is golden.* Curr Biol, 2003. **13**(23): p. R895-8.

37.   Aagaard, L., G. Laible, P. Selenko, M. Schmid, R. Dorn, G. Schotta, S. Kuhfittig, A. Wolf, A. Lebersorger, P.B. Singh, G. Reuter, and T. Jenuwein, *Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31.* EMBO J, 1999. **18**(7): p. 1923-38.

38.   Gu, B., P. Sun, Y. Yuan, R.C. Moraes, A. Li, A. Teng, A. Agrawal, C. Rheaume, V. Bilanchone, J.M. Veltmaat, K. Takemaru, S. Millar, E.Y. Lee, M.T. Lewis, B. Li, and X. Dai, *Pygo2 expands mammary progenitor cells by facilitating histone H3 K4 methylation.* J Cell Biol, 2009. **185**(5): p. 811-26.

39.   Pena, P.V., F. Davrazou, X. Shi, K.L. Walter, V.V. Verkhusha, O. Gozani, R. Zhao, and T.G. Kutateladze, *Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2.* Nature, 2006. **442**(7098): p. 100-3.

40.   Wu, M., P.F. Wang, J.S. Lee, S. Martin-Brown, L. Florens, M. Washburn, and A. Shilatifard, *Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS.* Mol Cell Biol, 2008. **28**(24): p. 7337-44.

41.   Lee, J.H. and D.G. Skalnik, *Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes.* Mol Cell Biol, 2008. **28**(2): p. 609-18.

42.   Lee, J.S., A. Shukla, J. Schneider, S.K. Swanson, M.P. Washburn, L. Florens, S.R. Bhaumik, and A. Shilatifard, *Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS.* Cell, 2007. **131**(6): p. 1084-96.

43.   Chen, Y., B. Wan, K.C. Wang, F. Cao, Y. Yang, A. Protacio, Y. Dou, H.Y. Chang, and M. Lei, *Crystal structure of the N-terminal region of human Ash2L shows a winged-helix motif involved in DNA binding.* EMBO Rep, 2011. **12**(8): p. 797-803.

44.   Sarvan, S., V. Avdic, V. Tremblay, C.P. Chaturvedi, P. Zhang, S. Lanouette, A. Blais, J.S. Brunzelle, M. Brand, and J.F. Couture, *Crystal structure of the trithorax group protein ASH2L reveals a forkhead-like DNA binding domain.* Nat Struct Mol Biol, 2011. **18**(7): p. 857-9.

45.   Lee, J.H. and D.G. Skalnik, *Rbm15-Mkl1 interacts with the Setd1b histone H3-Lys4 methyltransferase via a SPOC domain that is required for cytokine-independent proliferation.* PLoS One, 2012. **7**(8): p. e42965.

46.   Lee, J.H. and D.G. Skalnik, *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**(44): p. 42259-67.
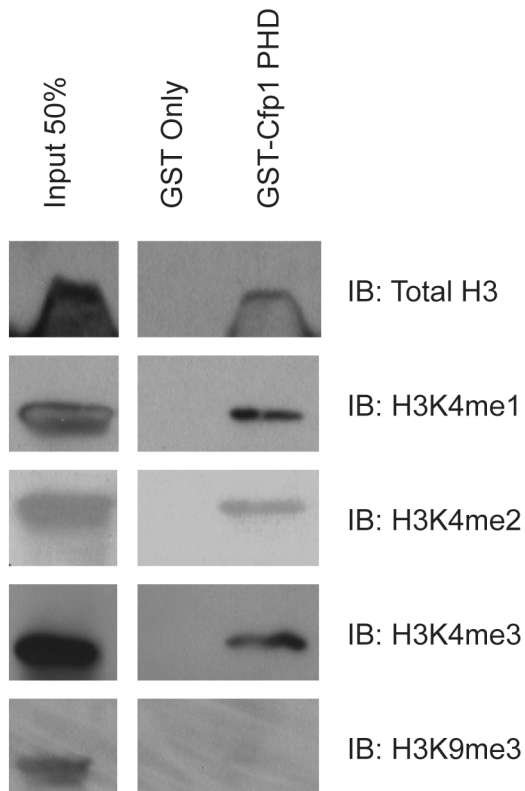
Figure 1

A

```
Cfp1   27 -IYCICRKP-DINCFMIGCDNC---NEWFHGDCIRITEKMAKAIREWYCREC 73
Phf2    6 -VYCVCRLPYDVTRFMIECDAC---KDWFHGSCVGVEEEEAPDIDIYHCPNC 53
Ing2  213 PTYCLCNQV--SYGEMIGCDNEQCPIEWFHFSCVSLTYK---PKGKWYCPKC 258
Bptf 2868 -LYCICKTPYDESKFYIGCDRC---QNWYHGRCVGILQSEAELIDEYVCPQC 2915
             **:*.        * **      :*:*  *:  :   .      :  * :*
```
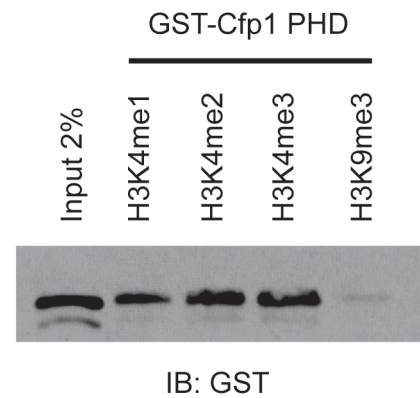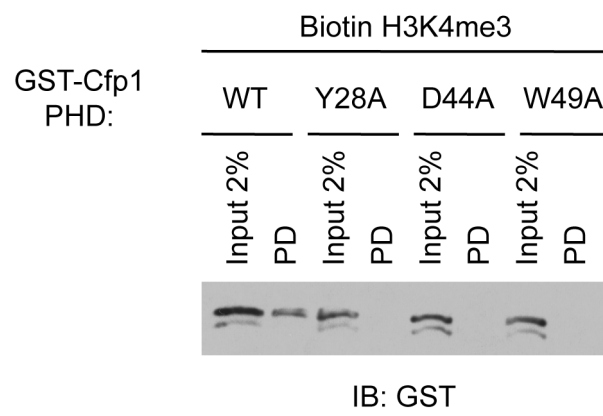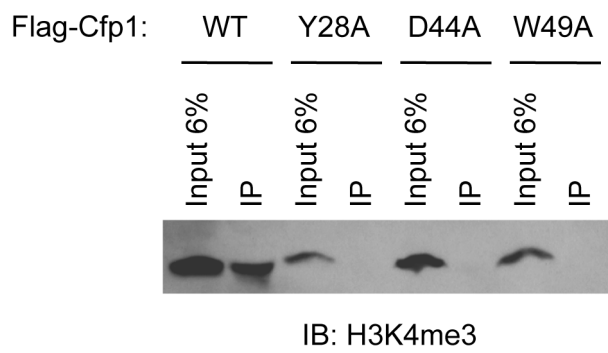
B

IB: Total H3

IB: H3K4me1

IB: H3K4me2

IB: H3K4me3

IB: H3K9me3

C

IB: GST

Figure 2

A

GST-Cfp1
PHD:

Biotin H3K4me3							
WT		Y28A		D44A		W49A	
Input 2%	PD	Input 2%	PD	Input 2%	PD	Input 2%	PD

IB: GST

B

Flag-Cfp1:

WT		Y28A		D44A		W49A	
Input 6%	IP	Input 6%	IP	Input 6%	IP	Input 6%	IP

IB: H3K4me3

Figure 3

A

PHD CXXC Acidic Basic Coiled-Coil SID HEAT

1 27 73 164 208 256 317 321 360 422 430 471 485 591 656 1-656 (WT)

C169A

W49A

W49A, C169A

● FLAG Tag

B

CXXC1 +/+ CXXC1 -/- Vector WT W49A, C169A C169A W49A

Cfp1

β-actin

C

	H3K4me3	DAPI	Merge	Co-localization	
CXXC1 +/+					[3.6%] 3.5% ± 0.5
CXXC1 -/-					[21.9%] *22.1% ± 3.4
Vector					[16.7%] *21.9% ± 1.4
WT					[3.0%] 3.3% ± 0.8
W49A					[2.0%] 4.6% ± 0.6

Figure 4

CXXC1 $^{+/+}$

CXXC1 $^{-/-}$

Vector

WT

W49A

C169A

W49A, C169A

Figure 5

CXXC1 $^{+/+}$ (< 5%)   CXXC1 $^{-/-}$ (> 95%)   Vector (> 95%)

WT (< 5%)   W49A (< 5%)   C169A (< 5%)

W49A, C169A
(41%)