Molecular Cell
Article

# An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting

Ling Cai,[1,2,9] Scott B. Rothbart,[1,2,9] Rui Lu,[1,2,9] Bowen Xu,[1,2,3] Wei-Yi Chen,[4] Ashutosh Tripathy,[2] Shira Rockowitz,[6] Deyou Zheng,[6] Dinshaw J. Patel,[7] C. David Allis,[5] Brian D. Strahl,[1,2,3] Jikui Song,[8,*] and Gang Greg Wang[1,2,3,*]
[1]UNC Lineberger Comprehensive Cancer Center
[2]Department of Biochemistry and Biophysics
[3]Biological and Biomedical Sciences Program
University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA
[4]Laboratory of Biochemistry and Molecular Biology
[5]Laboratory of Chromatin Biology and Epigenetics
The Rockefeller University, New York, NY 10065, USA
[6]Department of Neurology, Albert Einstein College of Medicine, Bronx, NY 10461, USA
[7]Structural Biology Program, Memorial Sloan-Kettering Cancer Center, New York, NY 10065, USA
[8]Department of Biochemistry, University of California, Riverside, CA 92521, USA
[9]These authors contributed equally to this work
*Correspondence: jikui.song@ucr.edu (J.S.), greg_wang@med.unc.edu (G.G.W.)
http://dx.doi.org/10.1016/j.molcel.2012.11.026

## SUMMARY

**Polycomb repressive complex 2 (PRC2) regulates pluripotency, differentiation, and tumorigenesis through catalysis of histone H3 lysine 27 trimethylation (H3K27me3) on chromatin. However, the mechanisms that underlie PRC2 recruitment and spreading on chromatin remain unclear. Here we report that histone H3 lysine 36 trimethylation (H3K36me3) binding activity is harbored in the Tudor motifs of PRC2-associated polycomb-like (PCL) proteins PHF1/PCL1 and PHF19/PCL3. Ectopically expressed PHF1 induced Tudor-dependent stabilization of PRC2 complexes on bulk chromatin and mediated spreading of PRC2 and H3K27me3 into H3K36me3-containing chromatin regions. In murine pluripotent stem cells, we identified coexistence of H3K36me3, H3K27me3, and PHF19/PCL3 at a subset of poised developmental genes and demonstrated that PHF19/PCL3 Tudor function is required for optimal H3K27me3 and repression of these loci. Collectively, our data suggest that PCL recognition of H3K36me3 promotes intrusion of PRC2 complexes into active chromatin regions to promote gene silencing and modulate the chromatin landscape during development.**

## INTRODUCTION

Modulation of the chromatin landscape by covalent histone posttranslational modifications (PTMs) represents a fundamental way of regulating DNA-templated processes such as gene transcription (Chi et al., 2010; Kouzarides, 2007). Histone H3 lysine 27 trimethylation (H3K27me3) promotes gene silencing, whereas promoter-associated histone H3 lysine 4 trimethylation (H3K4me3), together with gene body-associated H3 lysine 36 trimethylation (H3K36me3), demarcates active genes (Mikkelsen et al., 2007). PRC2 complexes are the major enzymatic machineries responsible for writing H3K27me3, and PRC2-mediated gene silencing is involved in various biological processes including stem cell pluripotency, differentiation, and cancer progression (Bracken and Helin, 2009; Margueron and Reinberg, 2011). While a reconstituted tetrameric PRC2 core complex (comprising EZH2, EED, SUZ12, and NURF55/RbAp48) is sufficient to induce H3K27me3 (Cao and Zhang, 2004; Margueron and Reinberg, 2011), a number of PRC2-associated factors, including EED (Margueron et al., 2009), JARID2 (Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009), and noncoding RNAs (Bracken and Helin, 2009; Margueron and Reinberg, 2011), that either modulate H3K27me3 catalysis or help to stabilize and tether PRC2 to appropriate genomic regions have recently been identified. Despite these recent advances, mechanisms responsible for PRC2 targeting and spreading on active chromatin to promote de novo repressed states remain poorly defined.

In this study, we show that two PCL family proteins (PHF1/PCL1 and PHF19/PCL3), accessory components of the PRC2 core complex (Cao et al., 2008; Hunkapiller et al., 2012; Sarma et al., 2008), harbor H3K36me3-reading activity within their N-terminal Tudor motifs. Biochemical, biophysical, and structural analyses reveal tight binding to H3K36me3 through a conserved aromatic cage formed by PCL Tudor domains. Our gene regulation and genomics analysis, using both overexpression and knockdown systems, define the PCL Tudor-H3K36me3 interaction as critical for both the targeting and spreading of PRC2 into active chromatin regions and for the maintenance of optimal repression of poised developmental

genes where PCL, H3K36me3, and H3K27me3 coexist. Our studies shed important light on the regulation of PRC2 by PCL proteins, implicating H3K36me3 reading in this process, and contribute to our understanding of the dynamics of gene expression and chromatin remodeling associated with development and disease (Chi et al., 2010).

## RESULTS

### PHF1 Directly Binds to H3K36me3 through a Conserved N-Terminal Tudor Motif

We began this study by attempting to identify, by mass spectrometry, H3K36me3-reader proteins using synthetic histone tail peptides as bait for nuclear extracts. In addition to the previously identified putative H3K36me3-reading effectors NSD2 and MSH6 (Vermeulen et al., 2010), we identified PHF1 (Figures 1A and S1A), a known accessory component of H3K27me3-promoting PRC2 complexes (Cao et al., 2008; Sarma et al., 2008). PHF1 and its homologs MTF2/PCL2 and PHF19/PCL3 have several putative chromatin-interacting domains (Cao et al., 2008) including a highly conserved N-terminal Tudor domain and two plant homeodomain (PHD) fingers (Figure S1B). In-solution peptide pull-down assays with recombinant GST-fusion proteins of each of these individual domains of PHF1 indicated that the Tudor domain facilitated H3K36me3 recognition (Figures 1B and S1C). To confirm this interaction, we established stable cell lines expressing wild-type (WT) or N-terminal truncated forms of PHF1 lacking either the Tudor domain or PHD fingers. Indeed, WT PHF1, but not that lacking the Tudor domain, pulled down with H3K36me3 peptides (Figures 1C and S1D). Furthermore, PHF1 coimmunoprecipitated (coIP) with endogenous histones that contained H3K36me3 in a Tudor-dependent manner (Figure 1D). Collectively, our data implicated the PHF1 Tudor domain ($PHF1_{Tudor}$) as an H3K36me3-reading effector module.

### $PHF1_{Tudor}$ Binds Specifically and with High Affinity to H3K36me3

We next sought to determine the histone-binding specificity of $PHF1_{Tudor}$ in an unbiased manner by using a recently developed high-density histone peptide microarray platform (Fuchs et al., 2011; Rothbart et al., 2012b). Arrays were printed with a library of unmodified and modified peptides (Table S1) containing known single and combinatorial histone PTMs (methylation, acetylation, and phosphorylation) on all four core histones (H2A, H2B, H3, and H4). Interrogation of the GST-tagged $PHF1_{Tudor}$ by peptide arrays (Figure 1E, top panel) revealed a strong preference for various H3K36me3-containing peptides (Figures 1E and S1E). In addition, $PHF1_{Tudor}$ displayed weak interactions with other methylated histone peptides such as H3 lysine 9 trimethylation (H3K9me3) and H3K27me3 (Figures 1F, S1F and S1G). Similar results were obtained using another peptide array platform (Figures S1H and S1I). Using isothermal titration calorimetry (ITC), we next quantified these observed interactions. Consistent with array results and peptide pull-downs, ITC measurements of the $PHF1_{Tudor}$ demonstrated high affinity and specificity for H3K36me3, with a dissociation constant ($K_D$) of ~4.2 μM, ~19 μM, and ~38 μM for tri-, di-, and monomethylated H3K36 states, respectively (Figure 1G, Table 1, and Figures S1L–S1O). Compared to H3K36me3, interactions with H3K9me3 and H3K27me3 were significantly weaker, with a $K_D$ of ~38 μM and ~168 μM, respectively (Figures S1P and S1Q and Table 1). We further determined the histone-binding specificity toward nucleosomal substrates in a GST pull-down assay using GST-$PHF1_{Tudor}$ and mononucleosomes purified from 293 cells. Again, $PHF1_{Tudor}$ preferentially bound to nucleosomes containing H3K36me3—but not other examined methylations, including those marked with H3K27me3, H3K9me3, H3K4me3, trimethylation of H4 lysine 20, or asymmetric dimethylation of H3 arginine 2 and arginine 17 (Figure 1H, lane of WT). Collectively, these data showed that $PHF1_{Tudor}$ motif demonstrates tight affinity and high selectivity toward H3K36me3.

Full-length PHF1 localizes exclusively in the nucleus (Figure 3D). We therefore asked whether association of full-length PHF1 with chromatin is dependent on H3K36me3. To do this, we diminished global levels of H3K36me3/H3K36me2 on chromatin by overexpressing an H3K36me3/H3K36me2 demethylase JMJD2A/KDM4A in 293 cells stably expressing a FLAG-tagged form of PHF1. Consistent with previous findings (Klose and Zhang, 2007), we observed a dramatic reduction in H3K36me3/H3K36me2 and H3K9me3 after biochemical separation of JMJD2A/KDM4A-overexpressing cells into chromatin-bound and soluble cytoplasmic/nucleoplasm fractions with no change detected in H3K4me3 or H3K27me3 (Figure 1I). Importantly, compared to control, a reduction was observed in the chromatin-bound fraction of PHF1, concurrent with an increase in its soluble pool (Figure 1I), thus demonstrating that PHF1 is less stable on chromatin in the absence of global H3K36me3/H3K36me2. Collectively, our in vitro and cellular biochemical data comprehensively showed that $PHF1_{Tudor}$ specifically and strongly binds to H3K36me3 and that this interaction is necessary to target PHF1 to chromatin. In addition, the $PHF1_{Tudor}$-H3K36me3 interaction represents the first high-affinity effector interaction to H3K36me3, which is several orders of magnitude tighter than all of the previously reported H3K36me3-reading motifs (Sun et al., 2008; Vezzoli et al., 2010; Xu et al., 2008).

### Structural Analysis of $PHF1_{Tudor}$ Revealed a Unique H3K36me3-Reading Pocket

To gain insight into the molecular basis for recognition of H3K36me3 by $PHF1_{Tudor}$, we performed NMR spectroscopy and obtained the solution structure of $PHF1_{Tudor}$ (residues 7–83) in complex with an $H3_{31-41}$K36me3 peptide. Initial comparison of $^{1}H,^{15}N$-HSQC spectra of the unbound $PHF1_{Tudor}$ with those in the presence of H3K36me3 indicated that, while the majority of $PHF1_{Tudor}$ residues exhibited no detectable change in chemical shifts, several residues showed a large perturbation, indicating a role for these residues in histone tail engagement (Figure S2A). The NMR structures of the $PHF1_{Tudor}$-H3K36me3 complex revealed that residues 31–81 of $PHF1_{Tudor}$ adopted a classic Tudor domain fold (Figure 2A and Table 2) as initially observed in the Survival of Neuron Tudor domain (Selenko et al., 2001), whereas flanking residues were disordered, evidenced by poor chemical shift dispersion and lack of NOE connectivity. We also found that the structure of $PHF1_{Tudor}$ is

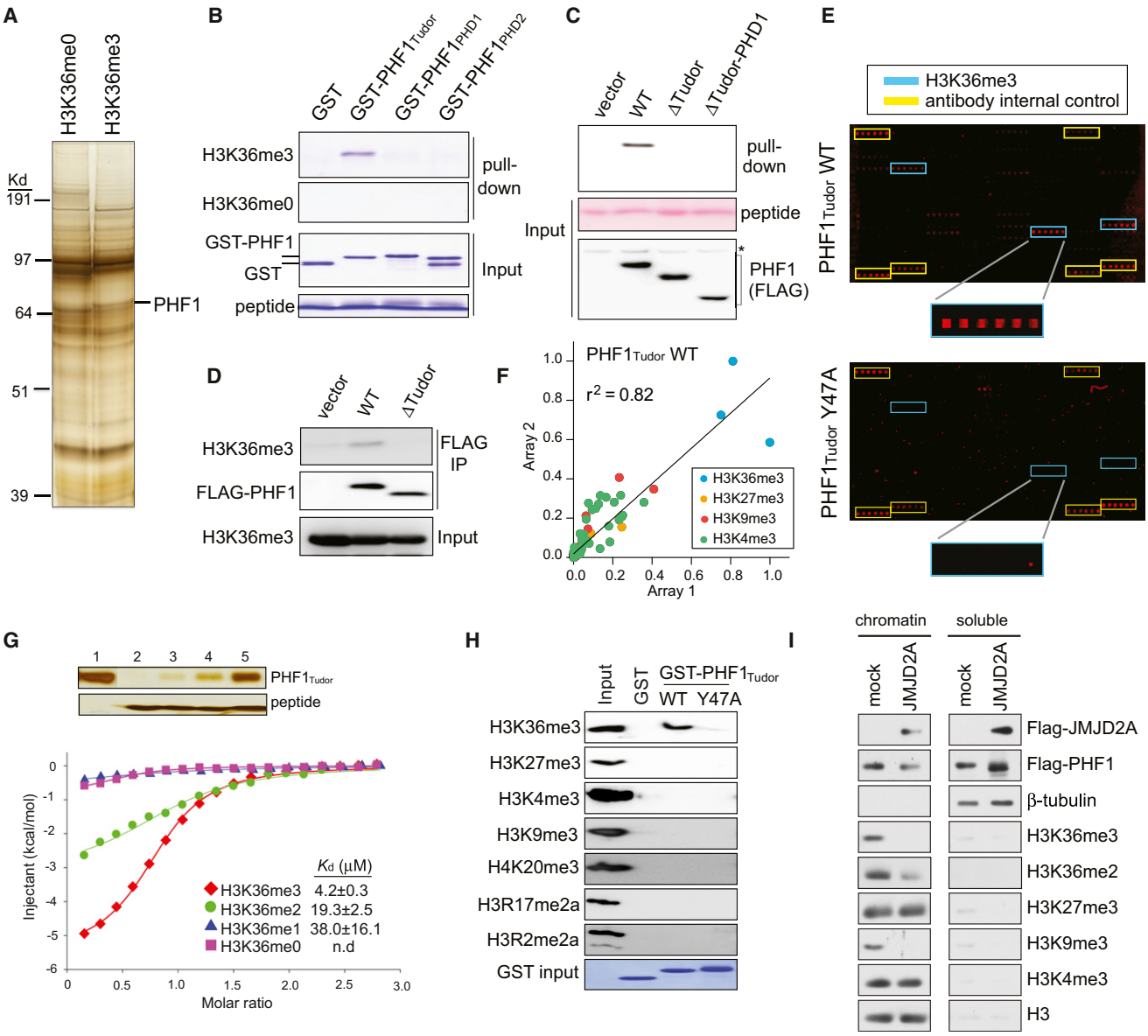



**Figure 1. The N-Terminal Tudor Motif of PHF1 Specifically Recognizes H3K36me3**

(A) Enrichment of PHF1 in pull-downs using biotinylated histone tail peptides that contain H3K36me3 as compared to that using unmodified H3K36. See also Figure S1.

(B) Pull-down of GST recombinant proteins fused to different PHF1 domains using H3K36me3 or unmodified H3K36 peptides (top two panels). Input is shown in bottom panels.

(C) Immunoblot following H3K36me3 peptide pull-down (top panel) from extracts of cells transduced with vector or that encoding the full-length PHF1 or deletion forms. Input of peptides and PHF1 is shown in middle and bottom panels.

(D) coIP using FLAG antibodies detecting a Tudor-dependent association of PHF1 (middle) to H3K36me3-containing histones (top) in 293 cells.

(E) Representative scan images of histone peptide microarrays probed with either WT or mutant (Y47A) GST-PHF1 Tudor proteins. Highlighted are positions of H3K36me3-containing peptides (blue) and IgG control (yellow). See also Table S1.

(F) Scatter plot of two peptide arrays probed with the PHF1 Tudor motif. Peptides are colored according to the legend. All other peptides are shown in black. Correlation coefficient was derived from linear regression analysis using GraphPad Prism.

(G) ITC measurements of binding affinities of PHF1 Tudor to H3 peptides containing different H3K36 methylation states. Insert represents silver staining of 1% of input (lane 1) and pull-down samples of PHF1 Tudor (top panel) using peptides harboring un-, mono-, di-, or trimethylated H3K36 (lanes 2–5, bottom panel).

(H) Immunoblot using various histone methylation antibodies on mononucleosomes pulled down by either WT or mutant (Y47A) GST-PHF1 Tudor proteins.

(I) Immunoblot of chromatin-bound and soluble cell fractions following 48 hr transient expression of JMJD2A/KDM4A. Mock indicates vector control.

**Table 1. Summary of Thermodynamic and Curve-Fitting Parameters for ITC Assays Using Recombinant PHF1 or PHF19 Tudor Domain Proteins and Various Histone Peptides**

| Protein | Peptide | ΔH (cal/mol) | ΔS (cal/mol/deg) | Ka ($M^{-1}$) | N | Chi2/DoF | $K_D$ (μM) |
|---|---|---|---|---|---|---|---|
| $PHF1^{Tudor}$, WT | $H3_{(28–43)}$K36me3 | −5,626 ± 108 | 4.45 | $2.38 \times 10^5 \pm 1.91 \times 10^4$ | 0.79 ± 0.01 | 5,415 | 4.2 ± 0.3 |
| $PHF1^{Tudor}$, WT | $H3_{(28–43)}$K36me2 | −3,994 ± 261 | 7.26 | $5.18 \times 10^4 \pm 6.79 \times 10^3$ | 0.94 ± 0.04 | 4,338 | 19.3 ± 2.5 |
| $PHF1^{Tudor}$, WT | $H3_{(28–43)}$K36me1 | −817 ± 342 | 17.3 | $2.63 \times 10^4 \pm 1.12 \times 10^4$ | 0.70 ± 0.23 | 658.3 | 38.0 ± 16.1 |
| $PHF1^{Tudor}$, WT | $H3_{(28–43)}$K36me0 | – | – | – | – | – | N.D. |
| $PHF1^{Tudor}$, WT | $H3_{(19–35)}$K27me3 | −5,141 ± 469 | −1.21 | $5.93 \times 10^3 \pm 486$ | 1.03 ± 0.07 | 347.7 | 168 ± 13.8 |
| $PHF1^{Tudor}$, WT | $H3_{(1–15)}$K9me3 | −817 ± 343 | 17.3 | $2.63 \times 10^4 \pm 1.12 \times 10^4$ | 0.70 ± 0.23 | 658.3 | 38.0 ± 16.2 |
| $PHF1^{Tudor}$, W41A | $H3_{(28–43)}$K36me3 | −1,355 ± 457 | 15.8 | $3.32 \times 10^4 \pm 1.20 \times 10^4$ | 0.68 ± 0.18 | 1,456 | 30.1 ± 10.9 |
| $PHF1^{Tudor}$, Y47A | $H3_{(28–43)}$K36me3 | – | – | – | – | – | N.D. |
| $PHF1^{Tudor}$, F65A | $H3_{(28–43)}$K36me3 | – | – | – | – | – | N.D. |
| $PHF1^{Tudor}$, F71A | $H3_{(28–43)}$K36me3 | – | – | – | – | – | N.D. |
| $PHF1^{Tudor}$, E66K | $H3_{(28–43)}$K36me3 | −6,471 ± 229 | −1.94 | $4.39 \times 10^4 \pm 3.76 \times 10^3$ | 0.60 ± 0.02 | 5,055 | 22.8 ± 2.0 |
| $PHF19^{Tudor}$, WT | $H3_{(27–46)}$K36me3 | −6,610 ± 267 | 0.466 | $1.60 \times 10^5 \pm 1.65 \times 10^4$ | 0.58 ± 0.02 | 7,436 | 6.2 ± 0.6 |
| $PHF19^{Tudor}$, WT | $H3_{(27–46)}$K36me1 | – | – | – | – | – | N.D. |
| $PHF19^{Tudor}$, Y56A | $H3_{(27–46)}$K36me3 | – | – | – | – | – | N.D. |

N.D., not determined. See also Figures S1L–S1S and S2G.

dominated by a β-barrel formed by five antiparallel β strands and that association of $PHF1_{Tudor}$ with H3K36me3 is mediated by direct contacts to both H3K36me3 and several surrounding histone tail residues (Figures 2B and 2C). The trimethylammonium side chain of H3K36me3 fits into an aromatic cage at one end of the β-barrel formed by residues Trp41, Tyr47, Phe65, Ser69, Asp67, and Phe71 (Figure 2D). Notably, Trp41 and Phe71 are not conserved in the Tudor domain of *Drosophila* polycomb-like (Pcl) protein (Figure S2B), which was reported to lack the ability to bind to histone methylation (Friberg et al., 2010), suggesting the role of these aromatic residues in formation of a closed cage composition. The overall structure of the H3K36me3-binding cage found in $PHF1_{Tudor}$ is reminiscent of those previously reported for the BRPF1 PWWP domain (Vezzoli et al., 2010) and the Eaf3 chromodomain (for comparison, see Figures S2C–S2E) (Sun et al., 2008; Xu et al., 2008). Several distinct features of $PHF1_{Tudor}$ were also revealed by our analysis, mainly occurring at contacts to histone residues surrounding H3K36me3. First, a salt bridge formed between the H3K37 side chain and the carboxylate of E66 of $PHF1_{Tudor}$ (Figure 2C, a dashed line). Second, the H3P38 side chain interacted with a hydrophobic patch generated by L38 and L46 of $PHF1_{Tudor}$ (Figures 2B and 2C). Third, hydrophobic contacts were observed between the H3H39 side chain and L38 and L48 of $PHF1_{Tudor}$ (Figures 2B and 2C).

To examine the functional relevance of observed contacts to H3K36me3 recognition, we introduced individual alanine substitutions to each of the H3K36me3-caging residues. We found that these single point mutations abrogated the binding of $PHF1_{Tudor}$ to H3K36me3 peptides on arrays (Figure 1E, Y47A, bottom panel) and by ITC (Figure 2E, Table 1, and Figure S1R) and peptide pull-downs (Figures 2F and S2F, left panel). Point mutations also abrogated binding of full-length PHF1 to H3K36me3 peptides (Figure 2G) and to H3K36me3-containing histones as assayed by GST-pull-downs (Figure 1H; Y47A) and coIPs (Figure 2H). Furthermore, disruption of the salt bridge to H3K37 by mutating E66 of $PHF1_{Tudor}$ to K66 also caused a significant reduction in H3K36me3 binding ($K_D$ = ~22.8 μM; Figures S1S and S2F and Table 1). Collectively, our structural and mutational analysis demonstrated that unlike other known H3K36me3 effector proteins, numerous van der Waals contacts and electrostatic interactions contribute to a high-affinity association of $PHF1_{Tudor}$ with H3K36me3.

## $PHF1_{Tudor}$ Alters Localization of PHF1-PRC2 Complexes, but Not PRC2 Core Structure

PHF1 was reported to interact with PRC2 core complexes where PHF1 modulates the enzymatic activity of associated PRC2 core complexes in vitro (Cao et al., 2008; Sarma et al., 2008). To determine whether PHF1 primarily associates with PRC2, we used mass spectrometry analysis to identify PHF1-associated factors from 293 stable cell lines. The top unique hits of identified proteins were the four subunits of tetrameric PRC2 core complexes (SUZ12, EED, NURF55/RbAp48, and EZH2 or EZH1) (Figure 3A and Table S2), consistent with comigration of PHF1 and PRC2 core components observed in gel filtration chromatography analyses (Cao et al., 2008; Sarma et al., 2008). We next sought to determine whether the $PHF1_{Tudor}$ motif was needed for PRC2 complex formation. coIPs with EZH2 and SUZ12 in cells stably expressing WT or Tudor-mutated forms of PHF1 showed that membership in the PRC2-PHF1 complex was not dependent on $PHF1_{Tudor}$ (Figures 3B, S3A, and S3B). Notably, while JARID2 was previously identified as associating with the PRC2 core complex (Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009), we did not detect JARID2 in either mass spectrometry analysis (Table S2) or coIP experiments with PHF1 (Figure S3C). These results are consistent with previous proteomics analysis of JARID2-associated complexes that failed to identify PCL proteins such as PHF1 or MTF2 (Shen et al., 2009). These studies collectively indicate that PHF1 primarily assembles in protein complexes with PRC2, potentially to modulate PRC2 targeting via its

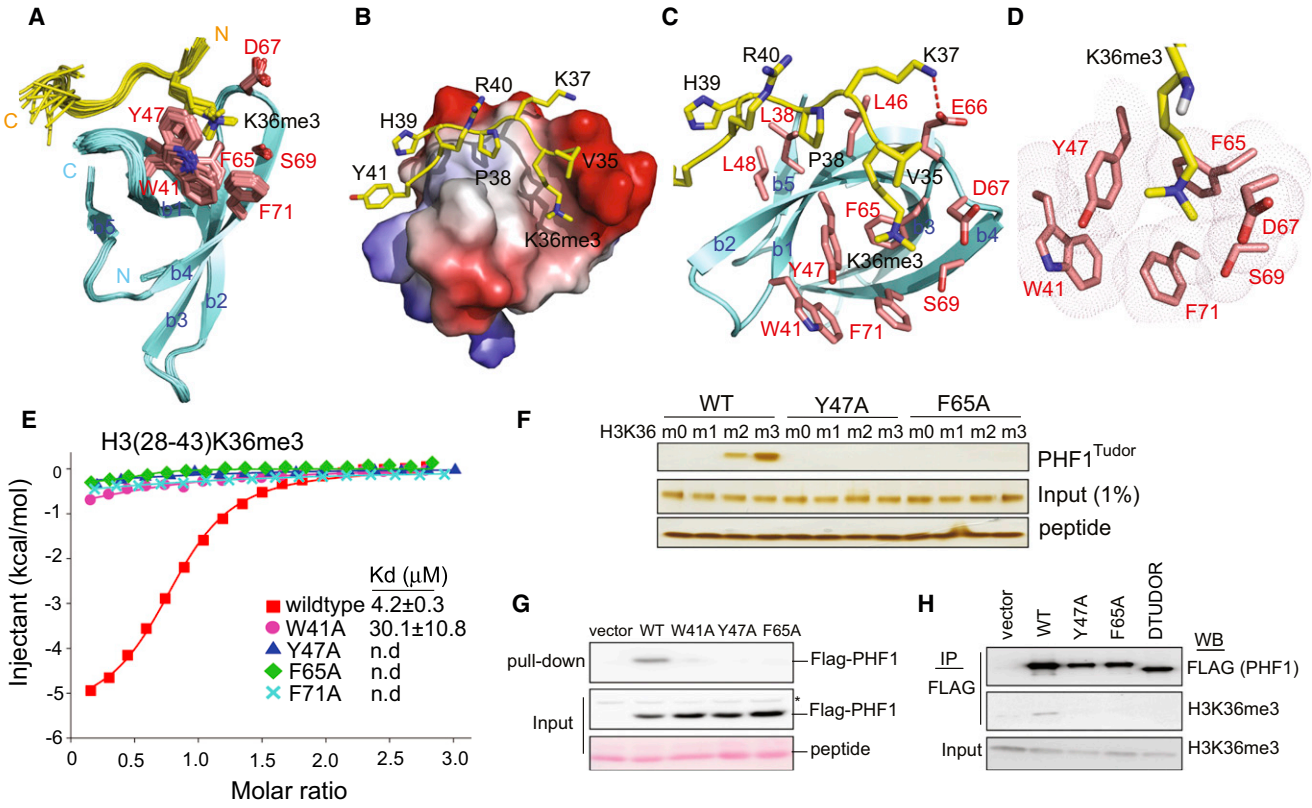



**Figure 2. Structural and Biophysical Analysis Revealed Intermolecular Interaction between PHF1$^{TUDOR}$ and H3K36me3-Containing Histones**

(A) Backbone superposition of 20 energy-minimized conformers representing the NMR structure of PHF1 Tudor in complex with an H3K36me3 peptide. For clarity, residues 1–30 and 82–83 of PHF1 Tudor and 31–34 of H3K36me3 peptide are not shown due to disorder. The PHF1 Tudor (cyan) and bound H3K36me3 peptide are shown in ribbon representation with side chains of H3K36me3-binding cage (red) and H3K36me3 (yellow) in stick representation. See also Figure S2.

(B) Surface electrostatic representation of PHF1 Tudor bound to H3K36me3 peptide.

(C) Representative structure of PHF1 Tudor-H3K36me3 complex with side chains involved in intermolecular contacts. The salt bridge formed between PHF1 Tudor E66 and H3K37 is depicted as a red dashed line.

(D) Positioning of the H3K36me3 side chain within an aromatic cage of the indicated residues (pink) on the surface of PHF1 Tudor.

(E) ITC measurements of binding affinities of WT or mutant PHF1 Tudor to H3K36me3 peptides.

(F) Silver staining of pull-down samples (top panel) of WT or mutant (Y47A and F65A) forms of PHF1 Tudor using histone tail peptides that harbor either un-, mono-, di-, or trimethylated H3K36. Inputs of protein and peptide used are shown in middle and bottom panels.

(G) Anti-FLAG immunoblot following peptide pull-down (top panel) from extracts of cells transduced with either empty vector or vector encoding a WT or mutant form of FLAG-tagged PHF1.

(H) Western blots examining association of PHF1 (top) to endogenous H3K36me3-containing histones (middle) in HeLa cells that expressed a WT or mutant form of FLAG-tagged PHF1.

histone-reading ability, and that PCL proteins and JARID2 reside in distinct PRC2 subcomplexes.

We next sought to determine whether the PHF1$_{Tudor}$ domain was required for PRC2 recruitment to chromatin. To this end, we fractionated PHF1 stable expression cells into chromatin and soluble fractions and found that, consistent with our observation following JMJD2A/KDM4A overexpression (Figure 1I), chromatin association of PHF1 was perturbed in the absence of Tudor recognition of H3K36 methylation (Figure 3C, lane 1 versus lanes 2 and 3). Notably, PRC2 association with bulk chromatin was also hindered in the absence of a functional PHF1$_{Tudor}$ in our PHF1 overexpression system, evidenced by a loss of SUZ12 and EZH2 from the chromatin fraction and a concurrent increased soluble pool of these PRC2-associated factors (Figure 3C). Consistent with these results, confocal immunofluorescence microscopy of WT PHF1 showed punctate nuclear foci overlapping with DAPI-dense regions of chromatin, while Tudor mutant forms of PHF1 showed diffuse distribution throughout the nucleus with accumulations in the chromatin-free nucleoplasm (Figures 3D and S3D). Collectively, our data suggest that PHF1 alters chromatin localization or stabilization of PHF1 and associated PRC2 proteins via an H3K36me3/H3K36me2-recognizing PHF1$_{Tudor}$, a function separated from assembly of a higher-order PHF1-PRC2 complex.

## PHF1 Induces Spreading of PRC2 and H3K27me3 into H3K36me3-Marked Loci

Our data presented thus far suggest that PCL family proteins may link PRC2 core complexes to active chromatin via H3K36me3 recognition and potentially modulate PRC2

**Table 2. Statistics for the 20 Energy-Minimized Conformers of $PHF1_{Tudor}$ in Complex with $H3_{31–41}$K36Me3**

| Distance constraints | |
|---|---|
| Long ([i−j] > 5) (intramolecular) | 250 |
| Medium (1 < [i−j] ≤ 5) | 106 |
| Sequential ([i−j] = 1) | 234 |
| Intraresidue (i = j) | 462 |
| Intermolecular | 38 |
| Dihedral angle constraints (φ and ψ) | 96 |
| Hydrogen bond constraints | 16 |
| Average pairwise rmsd to the mean structure (Å)[a] | |
| Backbone ( $C^{\alpha}$, C′, N, O) | 0.26 ± 0.05 |
| Heavy atoms | 0.72 ± 0.10 |
| Deviations from idealized covalent geometry | |
| Bond (Å) | 0.008 ± 0.002 |
| Angles (°) | 0.744 ± 0.012 |
| Impropers (°) | 0.479 ± 0.010 |
| Rmsd from experimental distance restraints (Å) | 0.032 ± 0.001 |
| Rmsd from experimental dihedral restraints (°) | 0.578 ± 0.027 |
| Ramachandran statistics (% of all residues)[a] | |
| Most favored | 91.3 |
| Additionally allowed | 8.7 |
| Generously allowed | 0.0 |
| Disallowed | 0.0 |

[a]Statistics are given for residues 31–81 of $PHF1_{Tudor}$ and residues 35–40 of the $H3_{31–41}$K36Me3 peptide.

stabilization and spreading during conversion of transcriptionally active/permissive states to repressed ones—a process often observed in cell differentiation and lineage specification (Mikkelsen et al., 2007). We next tested this hypothesis in HeLa cells, whose endogenous PHF1 expression is low (Boulay et al., 2011), and asked whether ectopic expression of PHF1 could shift chromatin boundaries that separate H3K27me3 from established H3K36me3 domains. Profiling of HeLa cells by chromatin immunoprecipitation coupled with deep sequencing (ChIP-seq) revealed that the *HOX-B* gene clusters, known to be regulated by PRC2 (Cao et al., 2008), exhibited a bimodal distribution of H3K27me3 and H3K36me3 (Figure 4A). H3K27me3 covered the distal HOX-B genes (from *HOXB13* to *HOXB7*), and H3K36me3 decorated the proximal ones (from *HOXB6* to *HOXB1*). We first confirmed this bimodal pattern using directed ChIPs (Figures S4A and S4B). In examination of changes in *HOX-B* gene expression before and after ectopic PHF1 expression, we found that WT and Tudor mutant forms of PHF1 demonstrated a chromatin environment-dependent effect on *HOX* repression. First, both WT and Tudor mutant PHF1 repressed *HOXB9*, *HOXB8*, and *HOXB7*, three distal *HOX*s marked by high H3K27me3 and low H3K36me3 (Figures 4A, 4B, and S4C). Using ChIP assays, we detected exogenous FLAG-tagged PHF1 as well as an increase in PRC2 residence and H3K27me3 at *HOXB9* and *HOXB8* (Figures 4C–4E and S4D), suggesting that independent of the Tudor motif, preexisting PRC2 may recruit PHF1 to H3K27me3-positive regions where PHF1 either enhances PRC2 enzymatic activities as previously reported (Cao et al., 2008; Sarma et al., 2008) or further stabilizes the chromatin binding of PRC2. While we detected no change in expression of proximal *HOX*s positioned deep within the H3K36me3 domain including *HOXB4*, *HOXB3*, and *HOXB1* (Figures 4A and 4B), WT—but not Tudor mutant PHF1—was able to repress *HOXB6* and *HOXB5* (Figure 4B, red line versus green, blue or purple), two *HOX* genes situated exactly on the junction of H3K36me3 and H3K27me3 domains (Figure 4A). These data suggested a Tudor-dependent mechanism for PRC2 targeting to regions of *HOXB6* and *HOXB5*. Indeed, PHF1 residence at *HOXB6* and *HOXB5* was completely dependent on Tudor function (Figure 4C), as was the residence of the PRC2 core complex components EZH2 and SUZ12 and the spread of H3K27me3 into these regions (Figures 4D–4F). No dramatic change in H3K36me3 was detected at examined *HOX-B* loci (Figure S4E). A similar Tudor-dependent H3K27me3 spreading was also observed at *PBX1* and *TADA1/ADA1*, two other developmentally critical genes that also exhibit bimodal distribution of H3K36me3 (at gene bodies) and H3K27me3 (at promoters) (Figures S4F and S4G). Taken together, our data demonstrate that through Tudor-mediated recognition of H3K36me3, ectopically expressed PHF1 is able to mediate the targeting and spreading of associated PRC2 core complexes to H3K36me3-demarcated active chromatin regions to modulate gene expression.

## PHF19/PCL3 Genetic Complementation in Murine Pluripotent Stem Cells Demonstrates a Tudor-Dependent Repression of Differentiation-Associated Genes

We next sought to examine the role of the PCL Tudor motif in PRC2-mediated gene silencing under physiological conditions. Endogenous expression of *PHF1* in HeLa and HEK293 cells is low (Boulay et al., 2011), and its knockdown only induced subtle changes in global H3K27me3 levels and expression of several tested PRC2-regulated genes (data not shown). Since endogenous *PHF1* expression was only found highly expressed in germ cells (Kawakami et al., 1998), we focused on PHF19 (also known as PCL3) (Figure S1B), a murine *PHF1* homolog that is highly expressed in murine ESCs and ESC-like embryonic carcinoma cell line F9 (Figure S5A). Importantly, the Tudor motif of PHF19 showed similar preference and affinity toward H3K36me3 and demonstrated a similar requirement of aromatic caging residues for binding to H3K36me3 peptides (Table 1; Figures S1J, S1K, S2F, and S2G), indicating a functional redundancy among these proteins in recognition of H3K36 methylation. We also found that, unlike a recent study showing that *Phf19* knockdown in ESCs also led to downregulation of *Suz12* (Hunkapiller et al., 2012), *Phf19* knockdown in F9 cells did not cause any detectable change in expression of all examined PRC2 complex genes including *Ezh2*, *Suz12*, and *Mtf2/Pcl2*, as examined by RT-PCR (Figure S5A) and immunoblot (Figure S5B). Knocking down *Phf19* in F9 embryonic carcinoma cells led to derepression of numerous PRC2 direct target genes, as revealed by gene array analysis (Tables S3 and S4) and RT-qPCR (Figure S5A). Consistent with a *Phf19* knockdown

A

| Protein name | Symbol | coverage% | # of unique peptides | total peptide hits |
|---|---|---|---|---|
| PHD finger protein 1, isoform 2 | PHF1, aka PCL1 | 39.3% | 22 | 119 |
| polycomb protein SUZ12 | SUZ12 | 34.9% | 27 | 121 |
| polycomb protein EED | EED | 23.2% | 11 | 51 |
| retinoblastoma binding protein 4 | NURF55, aka RbBp4/RbAp48 | 16.7% | 8 | 16 |
| enhancer of zeste 2 | EZH2 | 16.4% | 13 | 36 |
| enhancer of zeste 1 | EZH1 | 7.8% | 7 | 31 |
| histone deacetylase 1 | HDAC1 | 6.8% | 3 | 3 |
| retinoblastoma binding protein 7 | RbBp7, aka RbAp46 | 5.2% | 2 | 5 |

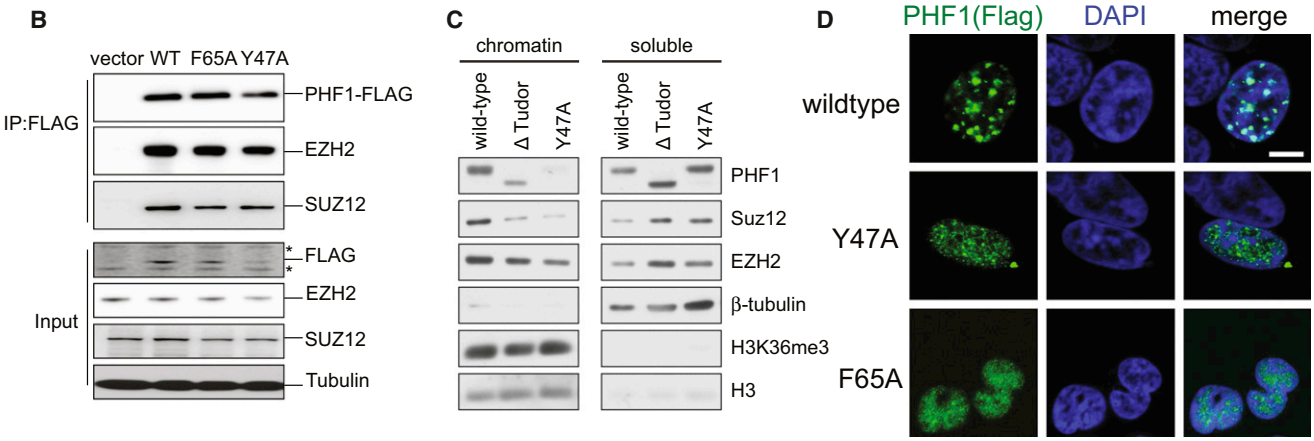



**Figure 3. PHF1$_{Tudor}$ Alters Subnuclear Localization of PHF1-PRC2 Complexes, but not PRC2 Core Structure Formation**

(A) Summary of protein peptides identified by mass spectrometry that specifically associate with FLAG-tagged PHF1 complexes purified from 293 stable cell lines. Numbers shown in the table are those of coverage percentage of amino acid sequences, unique peptides, and total peptide hits. See also Table S2.

(B) coIP of FLAG-tagged WT or Tudor-mutated forms of PHF1 with PRC2 core complex components EZH2 and SUZ12. Bottom panels are immunoblots of input. Asterisk represents nonspecific bands. See also Figure S3.

(C) Immunoblot of the indicated FLAG-PHF1 stable expression cell lines separated into chromatin and soluble fractions.

(D) Representative confocal immunofluorescence images revealing localization of WT versus Tudor mutant form of PHF1 proteins (FLAG-tagged, green) and chromatin (stained by DAPI, blue) in transiently transfected 293 cells. Scale bar, 5 μm.

experiment performed in E14 murine ESCs (Hunkapiller et al., 2012), the upregulated, but not downregulated, genes found in *Phf19*-knockdown cells were enriched with pathways associated with organismal development and cell differentiation (Tables S3 and S4). Among the top 100 upregulated genes, more than 40% are H3K27me3 positive at their promoters, demonstrating a significant overlap between H3K27me3 and Phf19-repressed genes (p value $< 2.2 \times 10^{-16}$, data not shown). Such observed premature derepression of development programs in pluripotent cells is similar to phenotypes observed in *Jarid2* null ESCs (Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009). Thus, F9 embryonic carcinoma cells serve as a suitable model to study functions of the Phf19 Tudor motif.

Previously, it has been shown that Phf19 binding sites strongly correlated to those of PRC2 complexes and H3K27me3 peaks (Hunkapiller et al., 2012). Our analysis of ChIP-seq profiling data obtained from murine E14 ESCs (Hunkapiller et al., 2012; Xiao et al., 2012) further revealed a moderate genome-wide association between Phf19 and H3K36me3, where higher levels of Phf19 binding were found at promoters of genes with higher levels of H3K36me3 among genes (n = 19,387) containing intermediate to low levels of H3K27me3 at promoters (Figure S5C). Analyses of these ESC ChIP-seq data also revealed a subset of PRC2-targeted genes marked by H3K36me3, which coexisted with H3K27me3 and Phf19, as exemplified by *Otx2* (see boxed region, Figure S5D). Using directed ChIP (Figures 5A and S5E–S5H) and sequential re-ChIP analyses (Figure 5B), we confirmed such coexistence of H3K36me3 and H3K27me3 at PRC2 target genes, including *Otx2*, *Meis1*, *Hoxa5*, and *Fgf15*, in both ESCs and F9 cells. At these bivalent PRC-targeted genes, H3K36me3 tends to increase toward gene bodies at the 3′ end, and H3K27me3 is enriched at promoter regions in the 5′ end. In addition, H3K36me3 was typically detected to a lesser extent when compared to that found at a locus of the housekeeping gene *Rps19* (Figure S5E). Detection of such H3K27me3-H3K36me3 bivalent genes is consistent with recent genomic studies showing coexistence of H3K36me3 and H3K27me3 at multivalent development genes in zebrafish sperm cells (Wu et al., 2011) as well as a proteomic study detecting a small percentage of H3 containing both H3K27 and H3K36 methylation in mammalian cells (Voigt et al., 2012).

Next, we used Phf19-expressing F9 cells to examine the function of the H3K36me3-recognizing Tudor motif in regulating the bivalent genes described above. Concurrent with derepression of these development genes following Phf19 knockdown (Figure S5A), we detected a significant reduction in levels of

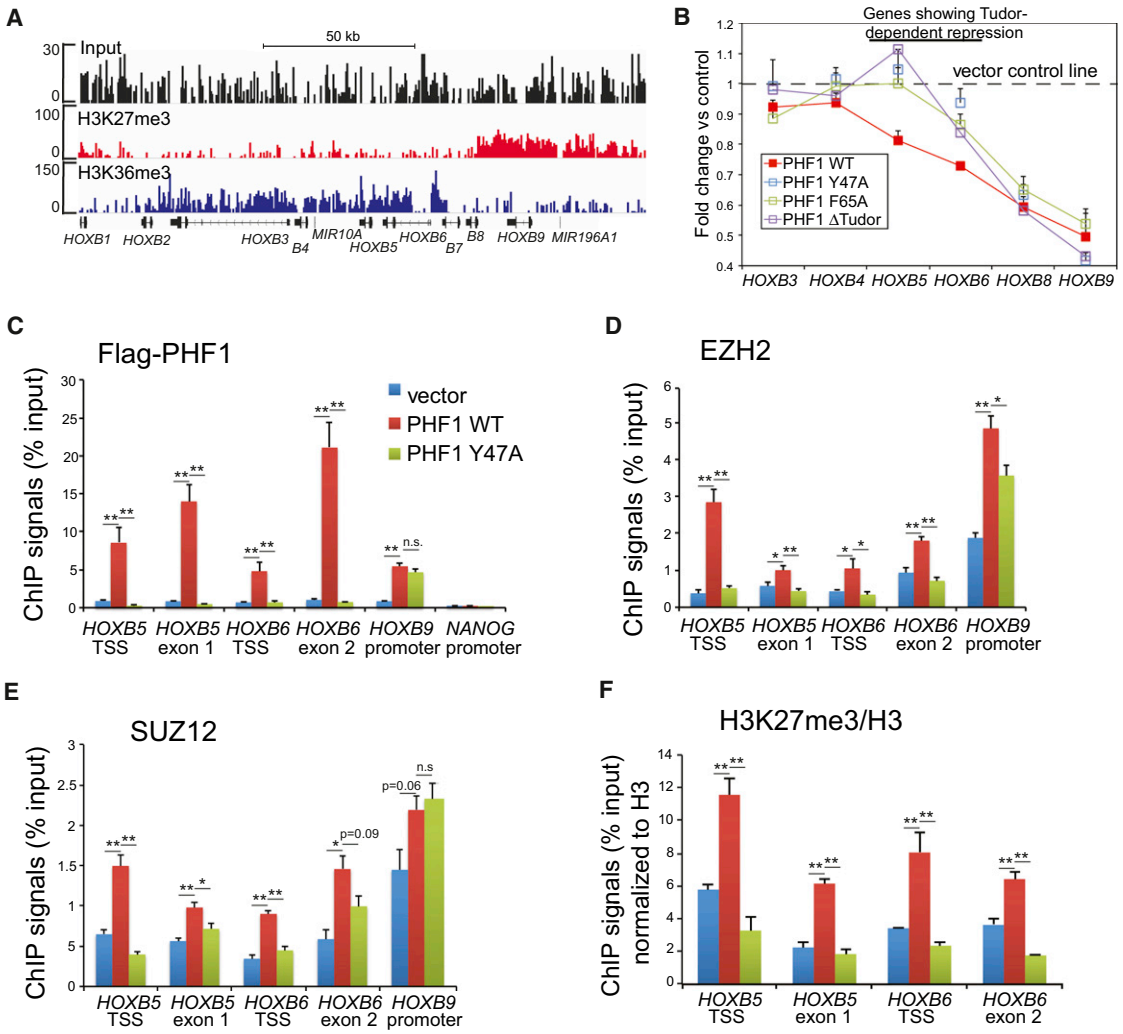



**Figure 4. Ectopic Expression of PHF1 Induced a Tudor-Dependent Spreading of PRC2 Complexes and H3K27me3 into an Adjacent H3K36me3 Region within *HOX-B* Gene Clusters**

(A) ChIP-seq revealing a bimodal distribution of H3K27me3 and H3K36me3 at *HOX-B* gene clusters. Position of each *HOX* is indicated at the bottom. See also Figure S4.

(B) Change in expression of *HOX-B* genes in HeLa stable cells with ectopic expression of either WT or Tudor mutant form of PHF1. Overexpression of the WT, but not Tudor mutant form of PHF1, led to repression of *HOXB5* and *HOXB6* (highlighted by a line on the top), two genes situated at the H3K36me3-H3K27me3 conjunction (A). Data of relative mRNA levels (y axis) from three independent experiments were normalized against vector control (shown as a dash line) and then presented as mean ± SD.

(C–F) ChIP for FLAG-tagged PHF1 (C), EZH2 (D), SUZ12 (E), and H3K27me3 (F) across different *HOX-B* genes in HeLa cells overexpressing either WT PHF1 or its Tudor mutant form. Data of ChIP signals (y axis) from three independent experiments were normalized to 1% of input used, data of histone modification (F) further normalized to those of total histone H3, and then presented as mean ± SD. Statistics shown are t test comparisons of ChIP in cells transduced with WT PHF1 to that with control or the mutant (Y47A) form. *$p < 0.05$; **$p < 0.005$; n.s., not significant.

H3K27me3 at all examined target genes (Figures 5C and 5D). No measurable change in H3K36me3 was found at tested loci (Figure S5I). Similar to PHF1 (Figure 3B) and findings of a previous report (Hunkapiller et al., 2012), both WT and Tudor mutant forms of PHF19 efficiently associated with PRC2 complexes (Figure S5J). Introduction of a WT hairpin-resistant PHF19 into F9 cells repressed *Otx2* and other tested target genes (*Meis1*, *Hoxa5*, *Fgf15*) in either parental F9 cells (*Otx2* as tested in Figure 5E) or cells with endogenous Phf19 knocked down (Figure 5F); the mutants that failed to engage H3K36me3 binding either led to *Otx2* derepression in a dominant-negative fashion in parental F9 cells (Figure 5E) or failed to restore gene repression (Figure 5F). As examined by ChIP assays at all tested gene loci, repression of gene expression was found to be correlated to the binding of WT, but not Tudor mutant PHF19 (Figure 5G), the recruitment of PRC2 core components (Figure 5H), and the level of H3K27me3 (Figure S5K). Collectively, our data suggest that PHF19, another member of PCL family

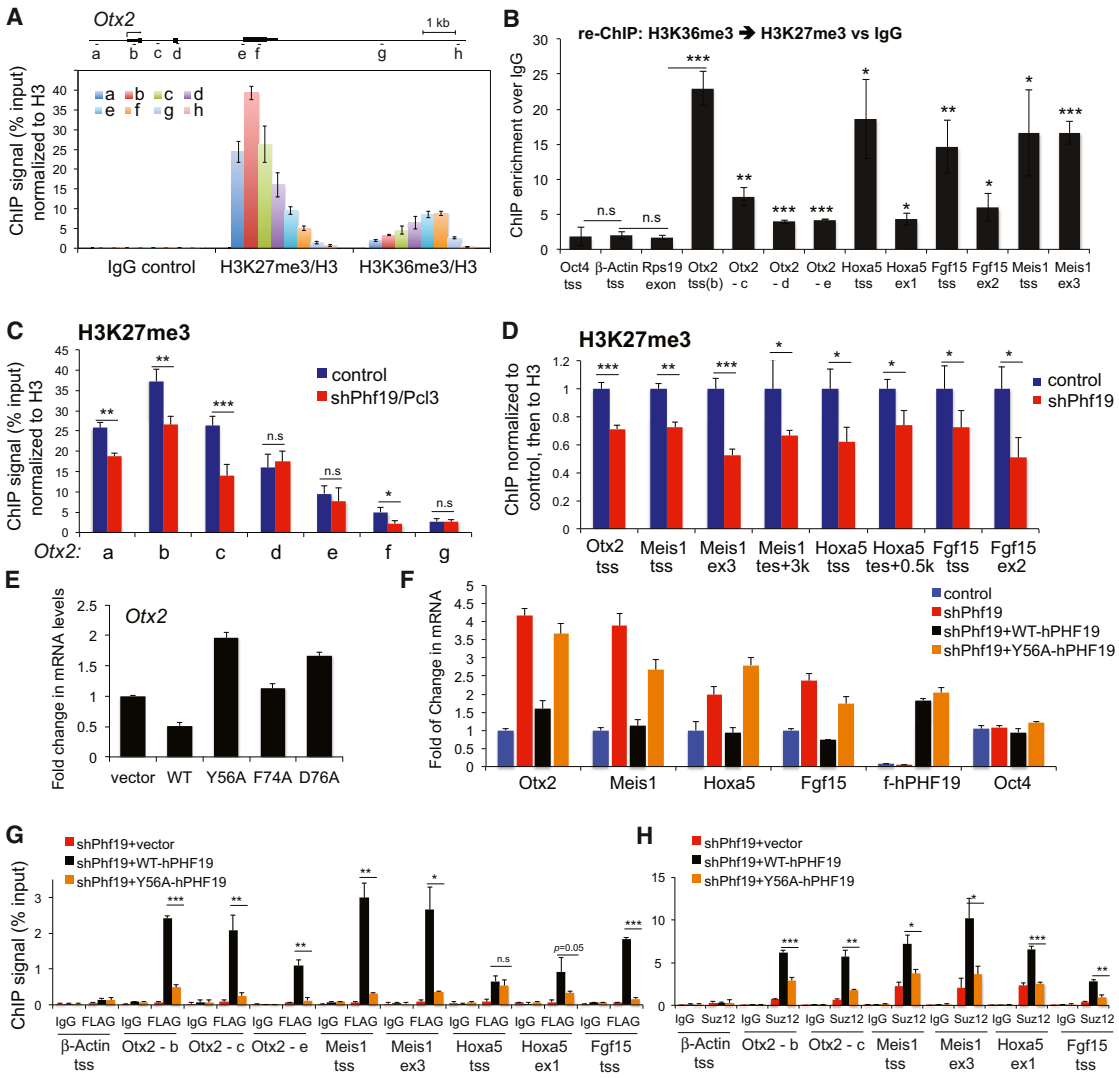



**Figure 5. Phf19/Pcl3 Demonstrates a Tudor-Dependent Repression of Differentiation-Associated Genes**

(A) ChIP for H3K36me3 and H3K27me3 across the *Otx2* gene in F9 embryonic carcinoma cells. The diagram on top depicts the genomic organization of *Otx2* with positions of each ChIP PCR amplicon marked as (a–h) (not drawn to scale). Data of ChIP signals (y axis) from three independent experiments were normalized to 1% of input and then to histone H3 and presented as mean ± SD. See also Figure S5.

(B) Sequential re-ChIP showing fold of enrichment by comparing ChIP signals of H3K27me3 over IgG control in a sequential IP following the first IP using H3K36me3 antibodies. Signal of ChIP (y axis) from three independent experiments were normalized to 1% of input and then to nonspecific IgG control and presented as mean ± SD. Statistics shown are t test comparisons of each tested loci to three negative control loci, the transcriptional start site (TSSs) of *Oct4* and β-actin, and the gene body of *Rps19*. *$p < 0.05$; **$p < 0.01$; ***$p < 0.005$; n.s., not significant.

(C and D) ChIP examining H3K27me3 levels associated with *Otx2* (C) and other examined development genes (D) in F9 cells transduced with control or *Phf19*-specific hairpins. Data of ChIP signals (y axis) from three independent experiments were normalized to 1% of input and then to histone H3 and presented as mean ± SD. ex, exon;TES, transcriptional ending site.

(E) Real-time PCR detecting *Otx2* expression levels in F9 cells following overexpression of WT and Tudor mutant forms of *PHF19/PCL3*. Data of relative mRNA levels (y axis) from three independent experiments were normalized to vector transduced cells and presented as mean ± SD.

(F) Real-time PCR of *Otx2*, *Meis1*, *Hoxa5*, *Fgf15*, and *Oct4* expression in F9 cells following *Phf19* knockdown (sh_Phf19) and reintroduction of hairpin-resistant WT and Tudor mutant forms of FLAG-tagged human PHF19/PCL3 (F-PHF19). y axis represents fold change after normalization of data from three independent experiments to control cells (presented as mean ± SD). See also Tables S3 and S4.

(G and H) ChIP examining the residence of FLAG-tagged PHF19/PCL3 (G) and SUZ12 (H) at *Otx2*, *Meis1*, *Hoxa5*, and *Fgf15* genes after transduction of hairpin-resistant WT or Tudor mutant PHF19/PCL3 into knockdown cells. Signal of ChIP (y axis) from three independent experiments were normalized to 1% of input and presented as mean ± SD. IgG and TSS of β-actin were used as antibody and locus control, respectively. Statistics shown are t test comparisons of ChIP in cells transduced with WT PHF19 to that with Tudor mutant forms. *$p < 0.01$; **$p < 0.001$; ***$p < 0.0001$.

proteins, also regulates repression of PRC2 targets via a functional Tudor motif that recognizes H3K36me3.

## DISCUSSION

### PCL Family Identified as High-Affinity H3K36me3 Effector Proteins

In this study, we have identified the PCL family of Tudor motifs as H3K36me3 effector modules. Our structural, biochemical, and biophysical studies of these Tudor motifs revealed an aromatic H3K36me3-caging pocket that forms direct contacts to H3K36me3 as well as to the surrounding H3 residues, which explained our observed high affinity and selectivity. In comparison to the previously identified H3K36me3-reading effectors BRPF1 and EAF3, whose affinities to H3K36me3 were measured in the high μM to low mM range ($K_D$ = ~0.3–3 mM), both the PHF1 and PHF19 Tudor motifs bind to H3K36me3 with affinities in the low μM range ($K_D$ = ~5–7 μM). These $K_D$ measurements are similar to affinities measured for effectors of other common histone methylation marks, including H3K4me3-recognizing PHD fingers ($K_D$ = ~1–10 μM) (Li et al., 2006; Wang et al., 2009), H3K27me3- and H3K9me3-recognizing chromo domains ($K_D$ = ~5–50 μM) (Bernstein et al., 2006; Kaustov et al., 2011), and an H3K9me3-recognizing Tudor domain (Rothbart et al., 2012a). Notably, PCL family proteins also contain two PHD fingers, one of which is in close proximity to the Tudor motif (Figure S1B); the histone-binding activities of these domains, as well as their influence on PCL functions, remain to be determined.

### PCL Proteins Serve as a Bridging and/or Targeting Mechanism for PRC2 to Intrude into Active Chromatin Regions

Our collective biochemical and genetic studies demonstrate a PRC2-targeting mechanism wherein the cofactor PCL proteins either stabilize, tether, or establish intrusion of PRC2 core components into the H3K36me3-containing chromatin via Tudor-mediated recognition of H3K36me3 and histone H3. coIP studies of PHF1 showed that this protein associates with PRC2 core components and assembles a higher-order complex efficiently, a process independent of the Tudor motif. We further show that PCL proteins promote PRC2 chromatin association and catalysis of H3K27me3, consistent with previous observations (Cao et al., 2008; Hunkapiller et al., 2012; Sarma et al., 2008). Our studies of ectopic PHF1 expression, modulation of global H3K36me3, as well as knockdown of endogenous Phf19 (without changing PRC2 core complexes per se) demonstrated a role of PCL proteins in the bridging of PRC2 to H3K36me3-containing nucleosomes, a mechanism that may act together with those provided by other PRC2-targeting mechanisms such as EED (Margueron et al., 2009), JARID2 (Li et al., 2010; Pasini et al., 2010; Peng et al., 2009; Shen et al., 2009), and noncoding RNAs (Bracken and Helin, 2009; Margueron and Reinberg, 2011). As we did not detect association of PHF1 or PHF19 to non-PRC2 target genes that are highly H3K36me3 positive (exemplified by *CD44* in Figure S4H and β-actin in Figure 5G), an intriguing model would be that noncoding RNA, DNA-binding, and chromatin-binding activities associated with PRC2 complexes and cofactors (for example EED and JARID2) collectively mediate initial PRC2 recruitment to H3K27me3-positive promoters/nucleosomes, while PCL proteins further modulate PRC2 stabilization and tether PRC2 to adjacent H3K36me3/H3K36me2-positive nucleosomes toward gene bodies. Although it has been previously shown that active histone methylation marks, including H3K36me3, inhibit PRC2-mediated H3K27me3 (Schmitges et al., 2011; Yuan et al., 2011), our studies demonstrate that PRC2 spreads and intrudes into active chromatin during gene silencing and chromatin remodeling, a critical process concurrent with development and differentiation (Mikkelsen et al., 2007).

### Coexisting Together: How Do Seemingly Biologically Opposing Chromatin Marks Work Together in the Chromatin Landscape?

Our genomics and ChIP analysis revealed that relatively lower levels of H3K36me3 exist in previously described poised bi- or multivalent chromatin state genes found in ESCs (Mikkelsen et al., 2007; Xiao et al., 2012). Such detected H3K36me3 may represent residual species of H3K36me3 docked by Phf19 in processes of PRC2-mediated chromatin remodeling associated with lineage differentiation and development. Indeed, a recent mass spectrometry-based analysis identified asymmetrically methylated, bivalent mononucleosomes that carry H3K36me3/H3K36me2 along with H3K27me3 (Voigt et al., 2012). We suspect that these asymmetrically methylated mononucleosomes are more likely to be found at conjunction regions of H3K27me3 and H3K36me3 in those bivalent or bimodal genes or loci as examined in our study (exemplified by *HOX-B* gene cluster in Figure 4A, *PBX1* in Figure S4F, and *Otx2* in Figure 5A) where PCL proteins serve as a bridging and stabilizing mechanism for PRC2 complex encroachment and spreading.

We also observed overall separation of high peaks/domains of H3K27me3 and H3K36me3 in ChIP-seq as previously reported (Mikkelsen et al., 2007; Xiao et al., 2012) (data not shown), and we suspect that PCL proteins may be involved in a dynamic process aimed at conversion of active chromatin state to the poised/repressed state by recruiting additional factors such as H3K36 demethylases to achieve completion of remodeling. We also hypothesize that residual H3K36me3 marks retained at these poised development genes might serve as a memory of cellular and epigenetic states, a mechanism that separates them from those silenced permanently. In accordance, expression of PCL family genes is dynamically regulated during organismal development, with *PHF1* specifically expressed in murine germ cells (Kawakami et al., 1998) and *Mtf2*/*Pcl2* preferentially expressed in undifferentiated ESCs (Walker et al., 2010). Both *PHF1* and *PHF19* were reported to be translocated or overexpressed in various human tumor types (Micci et al., 2006; Wang et al., 2004). Collectively, our studies strongly indicate that the chromatin- and H3K36me3-recognizing activities harbored in mammalian PCL proteins are critical for PRC2 targeting and regulation of developmentally critical genes. They further shed light on the modulation of the chromatin landscape and gene expression dynamics during processes of development and oncogenesis.

## EXPERIMENTAL PROCEDURES

### Plasmid Construction

Plasmids containing cDNA of PHF1 (also known as polycomb-like 1 or PCL1, NCBI accession NM_024165) and PHF19 (also known as polycomb-like 3 or PCL3, NCBI accession BC125076) were purchased from Open Biosystems. cDNAs were fused to either HA-FLAG or 2xFLAG tag by PCR and cloned into MSCV retroviral expression vectors (Clontech) and pGEX GST-fusion plasmids (GE Life Science). Deletion and mutation were introduced by PCR and site-directed mutagenesis. All plasmids used were confirmed by sequencing.

### Recombinant Glutathione S-Transferase Protein Production

Glutathione S-transferase (GST)-fusion proteins were generated and purified as previously described (Ruthenburg et al., 2011; Wang et al., 2009).

### Histone Peptide Microarrays

Methods for fabrication of the histone peptide array platform (with peptide identities listed in Table S1) and analysis of effector protein binding were previously described (Fuchs et al., 2011; Rothbart et al., 2012b). A second histone peptide array platform was also obtained and used according to the manufacturer's specifications (Active Motif).

### Chromatin Fractionation

Chromatin and associated proteins were isolated from asynchronously growing 293 cells as described (Rothbart et al., 2012a).

### ChIP and Sequential Re-ChIP Assays

ChIP was performed using a previously described protocol (Wang et al., 2009). ChIP signals were represented as a percentage of input chromatin, and fold of enrichment was calculated by normalizing against signals of nonspecific IgG. Sequential re-ChIP was performed using a previously described protocol (Voigt et al., 2012). Information regarding primers used is described in Table S5.

### NMR Spectroscopy

Collection of NMR spectra and structural calculations are detailed in Supplemental Experimental Procedures.

## ACCESSION NUMBERS

The structural coordinates of PHF1 Tudor domain protein in the complex with $H3_{31–41}$K36me3 peptide have been deposited in the Protein Data Base under accession number 2M0O. The microarray data of Phf19 shRNA knockdown have been deposited in GEO database under accession number GSE42463.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, five tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2012.11.026.

## ACKNOWLEDGMENTS

We thank K. Krajewski for synthesizing peptides, M. Vernon for gene microarray analysis, Y. Zhang for providing JMJD2 plasmids, and T. Magnuson for anti-EED antibodies used in this study. L.C. and S.B.R. are supported by postdoctoral fellowships awarded by the US Army-DOD Prostate Cancer Research Program and the UNC Lineberger Comprehensive Cancer Center Basic Sciences Training Program (T32CA09156), respectively. This research was also supported by grants from the Leukemia & Lymphoma Society and STARR Foundation to both D.J.P. and C.D.A., and National Institutes of Health (NIH) grants to B.D.S. (GM068088). J.S. is supported by funds from UC-Riverside, and G.G.W. is supported by the NIH/NCI "Pathway to Independence" Award in Cancer Research (CA151683) and University Cancer Research Fund (UCRF) of the NC State. G.G.W. is a Martin D. Abeloff, M.D. V Scholar of the V Foundation for Cancer Research.

Received: September 20, 2012
Revised: October 31, 2012
Accepted: November 21, 2012
Published: December 27, 2012

## REFERENCES

Bernstein, E., Duncan, E.M., Masui, O., Gil, J., Heard, E., and Allis, C.D. (2006). Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. Mol. Cell. Biol. *26*, 2560–2569.

Boulay, G., Rosnoblet, C., Guérardel, C., Angrand, P.O., and Leprince, D. (2011). Functional characterization of human Polycomb-like 3 isoforms identifies them as components of distinct EZH2 protein complexes. Biochem. J. *434*, 333–342.

Bracken, A.P., and Helin, K. (2009). Polycomb group proteins: navigators of lineage pathways led astray in cancer. Nat. Rev. Cancer *9*, 773–784.

Cao, R., and Zhang, Y. (2004). SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol. Cell *15*, 57–67.

Cao, R., Wang, H., He, J., Erdjument-Bromage, H., Tempst, P., and Zhang, Y. (2008). Role of hPHF1 in H3K27 methylation and Hox gene silencing. Mol. Cell. Biol. *28*, 1862–1872.

Chi, P., Allis, C.D., and Wang, G.G. (2010). Covalent histone modifications—miswritten, misinterpreted and mis-erased in human cancers. Nat. Rev. Cancer *10*, 457–469.

Friberg, A., Oddone, A., Klymenko, T., Müller, J., and Sattler, M. (2010). Structure of an atypical Tudor domain in the Drosophila Polycomblike protein. Protein Sci. *19*, 1906–1916.

Fuchs, S.M., Krajewski, K., Baker, R.W., Miller, V.L., and Strahl, B.D. (2011). Influence of combinatorial histone modifications on antibody and effector protein recognition. Curr. Biol. *21*, 53–58.

Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., McCleary, D., Ramalho-Santos, M., Krogan, N., Ren, B., Song, J.S., and Reiter, J.F. (2012). Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. PLoS Genet. *8*, e1002576.

Kaustov, L., Ouyang, H., Amaya, M., Lemak, A., Nady, N., Duan, S., Wasney, G.A., Li, Z., Vedadi, M., Schapira, M., et al. (2011). Recognition and specificity determinants of the human cbx chromodomains. J. Biol. Chem. *286*, 521–529.

Kawakami, S., Mitsunaga, K., Kikuti, Y.Y., Ando, A., Inoko, H., Yamamura, K., and Abe, K. (1998). Tctex3, related to Drosophila polycomblike, is expressed in male germ cells and mapped to the mouse t-complex. Mamm. Genome *9*, 874–880.

Klose, R.J., and Zhang, Y. (2007). Regulation of histone methylation by demethylimination and demethylation. Nat. Rev. Mol. Cell Biol. *8*, 307–318.

Kouzarides, T. (2007). Chromatin modifications and their function. Cell *128*, 693–705.

Li, H., Ilin, S., Wang, W., Duncan, E.M., Wysocka, J., Allis, C.D., and Patel, D.J. (2006). Molecular basis for site-specific read-out of histone H3K4me3 by the BPTF PHD finger of NURF. Nature *442*, 91–95.

Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B.E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. Genes Dev. *24*, 368–380.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., 3rd, Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., et al. (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Micci, F., Panagopoulos, I., Bjerkehagen, B., and Heim, S. (2006). Consistent rearrangement of chromosomal band 6p21 with generation of fusion genes JAZF1/PHF1 and EPC1/PHF1 in endometrial stromal sarcoma. Cancer Res. *66*, 107–112.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553–560.

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature *464*, 306–310.

Peng, J.C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., and Wysocka, J. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell *139*, 1290–1302.

Rothbart, S.B., Krajewski, K., Nady, N., Tempel, W., Xue, S., Badeaux, A.I., Barsyte-Lovejoy, D., Martinez, J.Y., Bedford, M.T., Fuchs, S.M., et al. (2012a). Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. Nat. Struct. Mol. Biol. *19*, 1155–1160.

Rothbart, S.B., Krajewski, K., Strahl, B.D., and Fuchs, S.M. (2012b). Peptide microarrays to interrogate the "histone code". Methods Enzymol. *512*, 107–135.

Ruthenburg, A.J., Li, H., Milne, T.A., Dewell, S., McGinty, R.K., Yuen, M., Ueberheide, B., Dou, Y., Muir, T.W., Patel, D.J., and Allis, C.D. (2011). Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. Cell *145*, 692–706.

Sarma, K., Margueron, R., Ivanov, A., Pirrotta, V., and Reinberg, D. (2008). Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Mol. Cell. Biol. *28*, 2718–2731.

Schmitges, F.W., Prusty, A.B., Faty, M., Stützer, A., Lingaraju, G.M., Aiwazian, J., Sack, R., Hess, D., Li, L., Zhou, S., et al. (2011). Histone methylation by PRC2 is inhibited by active chromatin marks. Mol. Cell *42*, 330–341.

Selenko, P., Sprangers, R., Stier, G., Bühler, D., Fischer, U., and Sattler, M. (2001). SMN tudor domain structure and its interaction with the Sm proteins. Nat. Struct. Biol. *8*, 27–31.

Shen, X., Kim, W., Fujiwara, Y., Simon, M.D., Liu, Y., Mysliwiec, M.R., Yuan, G.C., Lee, Y., and Orkin, S.H. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell *139*, 1303–1314.

Sun, B., Hong, J., Zhang, P., Dong, X., Shen, X., Lin, D., and Ding, J. (2008). Molecular basis of the interaction of Saccharomyces cerevisiae Eaf3 chromo domain with methylated H3K36. J. Biol. Chem. *283*, 36504–36512.

Vermeulen, M., Eberl, H.C., Matarese, F., Marks, H., Denissov, S., Butter, F., Lee, K.K., Olsen, J.V., Hyman, A.A., Stunnenberg, H.G., and Mann, M. (2010). Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. Cell *142*, 967–980.

Vezzoli, A., Bonadies, N., Allen, M.D., Freund, S.M., Santiveri, C.M., Kvinlaug, B.T., Huntly, B.J., Göttgens, B., and Bycroft, M. (2010). Molecular basis of histone H3K36me3 recognition by the PWWP domain of Brpf1. Nat. Struct. Mol. Biol. *17*, 617–619.

Voigt, P., LeRoy, G., Drury, W.J., 3rd, Zee, B.M., Son, J., Beck, D.B., Young, N.L., Garcia, B.A., and Reinberg, D. (2012). Asymmetrically modified nucleosomes. Cell *151*, 181–193.

Walker, E., Chang, W.Y., Hunkapiller, J., Cagney, G., Garcha, K., Torchia, J., Krogan, N.J., Reiter, J.F., and Stanford, W.L. (2010). Polycomb-like 2 associates with PRC2 and regulates transcriptional networks during mouse embryonic stem cell self-renewal and differentiation. Cell Stem Cell *6*, 153–166.

Wang, S., Robertson, G.P., and Zhu, J. (2004). A novel human homologue of Drosophila polycomblike gene is up-regulated in multiple cancers. Gene *343*, 69–78.

Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J., and Allis, C.D. (2009). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature *459*, 847–851.

Wu, S.F., Zhang, H., and Cairns, B.R. (2011). Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. Genome Res. *21*, 578–589.

Xiao, S., Xie, D., Cao, X., Yu, P., Xing, X., Chen, C.C., Musselman, M., Xie, M., West, F.D., Lewin, H.A., et al. (2012). Comparative epigenomic annotation of regulatory DNA. Cell *149*, 1381–1392.

Xu, C., Cui, G., Botuyan, M.V., and Mer, G. (2008). Structural basis for the recognition of methylated histone H3K36 by the Eaf3 subunit of histone deacetylase complex Rpd3S. Structure *16*, 1740–1750.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J. Biol. Chem. *286*, 7983–7989.

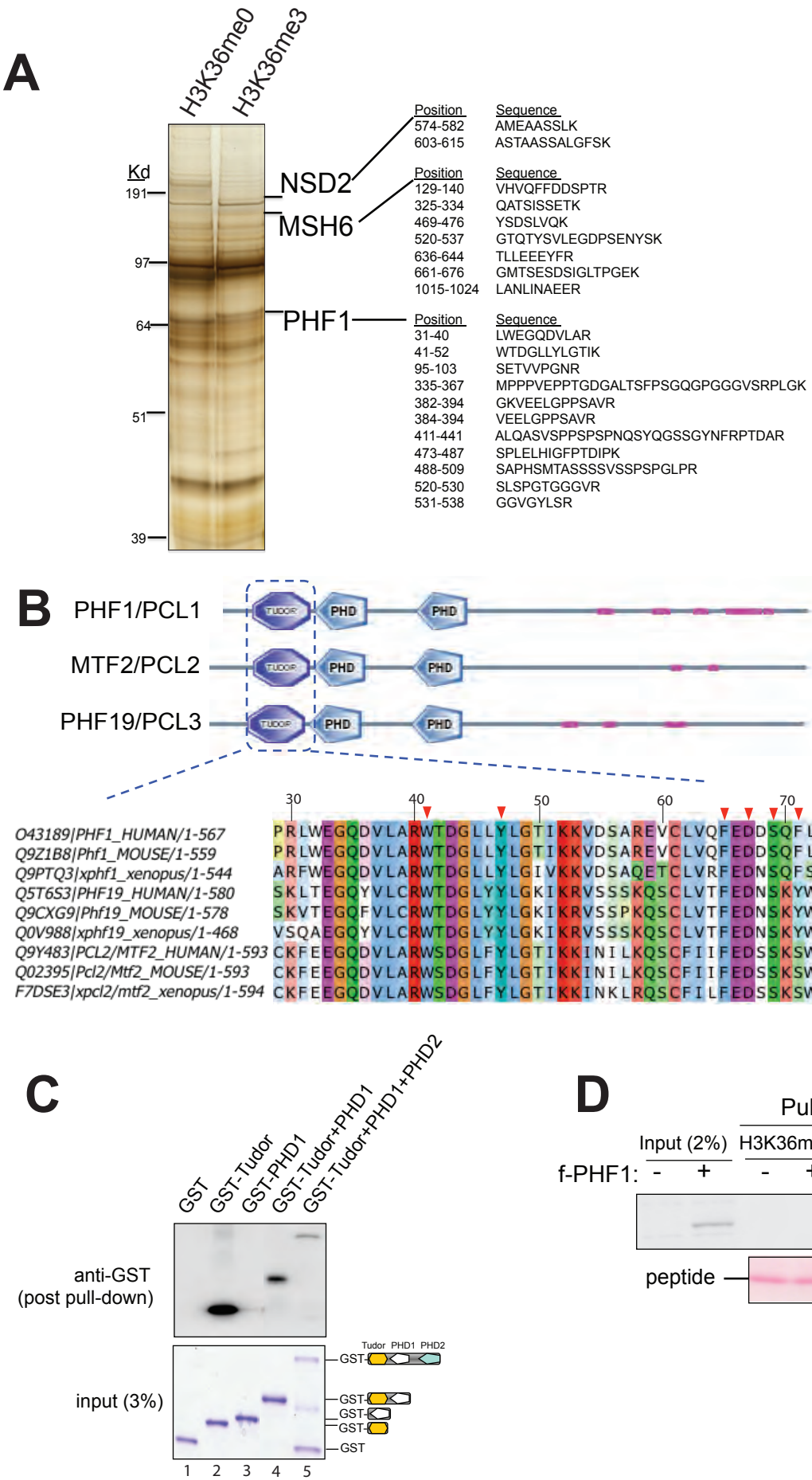
**Molecular Cell, Volume *49***

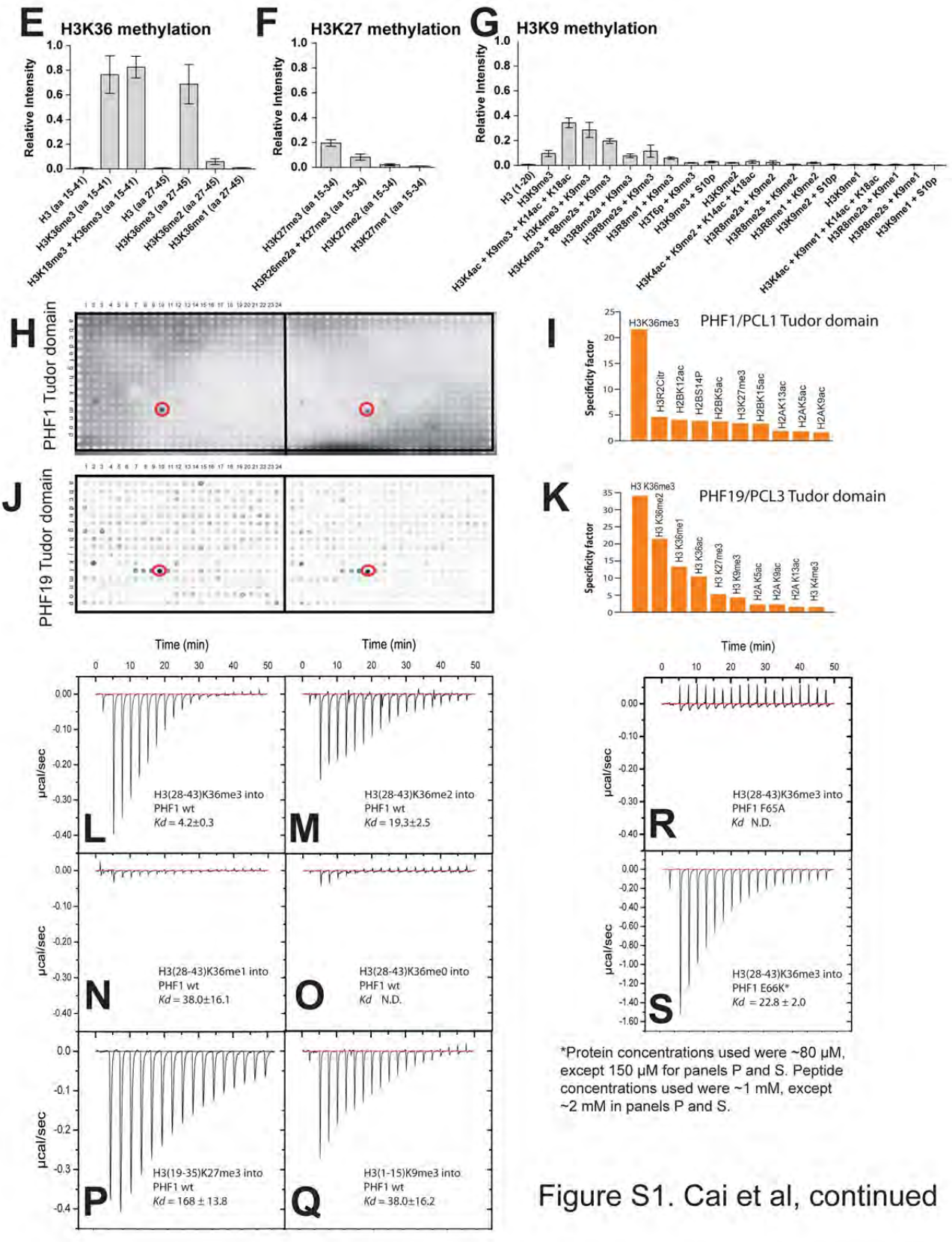
## Supplemental Information

# An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting

Ling Cai, Scott B. Rothbart, Rui Lu, Bowen Xu, Wei-Yi Chen, Ashutosh Tripathy, Shira Rockowitz, Deyou Zheng, Dinshaw J Patel, C. David Allis, Brian D. Strahl, Jikui Song, and Gang Greg Wang

Figure S1, Cai et al.






*Protein concentrations used were ~80 µM, except 150 µM for panels P and S. Peptide concentrations used were ~1 mM, except ~2 mM in panels P and S.

Figure S1. Cai et al, continued

**Figure S1. $PHF1_{Tudor}$ and $PHF19_{Tudor}$ Exhibiting High Affinity and Selectivity for H3K36me3, Related to Figure 1 and Table 1**

(A) Enrichment of NSD2, MSH6 and PHF1 in pull-downs using biotinylated histone H3 peptides (residue 27-46) that contain H3K36me3, as compared to that using the unmodified-H3K36 version (H3K36me0). Shown on the right are sequences of peptides identified by mass spectrometry in gel slides cut from the SDS-PAGE lane loaded with H3K36me3 pull-down samples.

(B) Protein domain structure of three PCL family proteins, PHF1/PCL1, MTF2/PCL2 and PHF19/PCL3, with their Tudor motifs highlighted in a dashed box. Lower panel, alignment of the Tudor motif of PCL family protein homologues in human, mouse, and frog using JalView alignment tools (sequences acquired from UniProt database). Highlighted with red arrowhead are residues involved in the formation of an H3K36me3-recognizing 'cage' (see Figure 2).

(C) Anti-GST immunoblot (top) of pull-down samples using biotinylated H3K36me3 peptides and the indicated GST fused to different PHF1 domains (lanes 2-5). Lower panel represents coomassie staining of 3% of protein input (with domain structures shown on the right).

(D) Anti-FLAG immunoblot (top panel) of input (lane 1-2) of extracts from cells expressing vector control (-) or a full-length FLAG-tagged PHF1 (+), as well as samples of pull-down using unmodified-H3K36 or H3K36me3 peptides. Bottom panel shows ponceau-S staining of peptides.

(E-G) Quantification of binding of GST-$PHF1_{Tudor}$ motif to histone tail peptides that contain methylation of H3K36 (E), H3K27 (F), or H3K9 (G) as measured by peptide arrays. Histone tail sequence coverage and modification(s) contained are shown below. Relative intensities are

calculated by averaging signal intensities from two replicate arrays (with a total of 48 individual spots per peptide) ± SEM.

(H-K) Anti-GST immunoblot images (panels H and J) and graphical representation of highest binding events detected (panels I and K) showing the binding specificity of GST-PHF1$_{Tudor}$ (H-I) or GST-PHF19$_{Tudor}$ (J-K) motifs using Active Motif Histone Peptide Arrays. Peptides were spotted in duplicate as shown in two boxes on the same array. Immunoblot signals and the position of H3K36me3 peptides are highlighted with red circles.

(L-S) ITC titration experimental curves of wildtype (wt) or mutant forms of PHF1$_{Tudor}$ bound to different H3 peptides as indicated. Concentrations of protein used were approximately 80 μM, except 150 μM in panels P and S, and those of peptides used are approximately 1 mM, except 2 mM in panels P and S. N.D., not determined.

Figure S2, Cai et al.

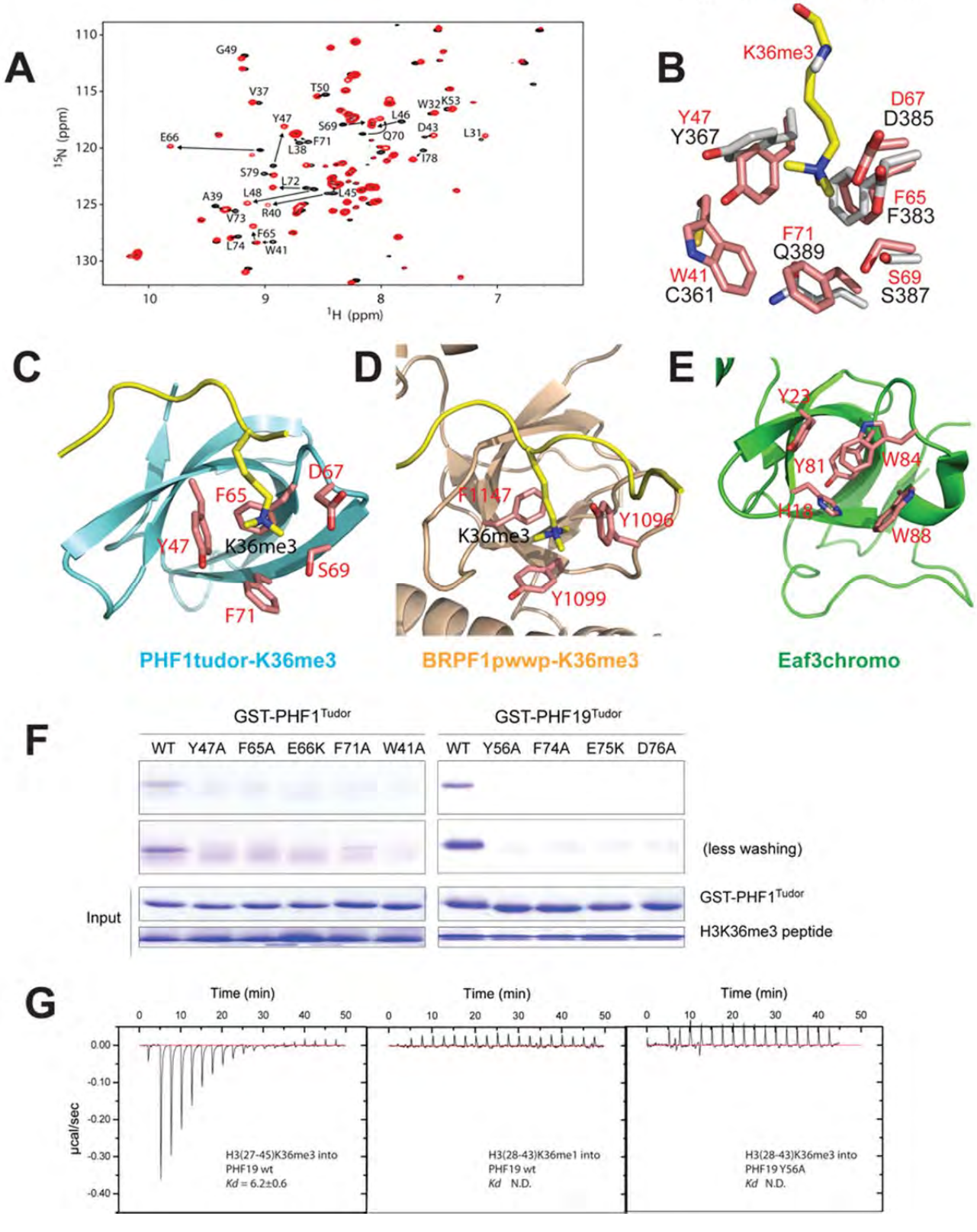


**Figure S2. Structural and Mutational Analysis Revealing Intermolecular Interaction between $PHF1_{Tudor}$ or $PHF19_{Tudor}$ and H3K36me3 Peptides, Related to Figure 2 and Tables 1-2**

(A) Spectral overlap of $PHF1_{Tudor}$ free (black) and in complex with an $H3_{31-41}$K36me3 peptide (red). Residues with substantial chemical shift changes upon complex formation are labeled with arrows.

(B) Structural superposition of the $PHF1_{Tudor}$ aromatic cage (in pink) with the equivalent region of *Drosophila* Pcl Tudor domain (in grey, PDB accession number 2XK0).

(C-E) Comparison of the H3K36me3-binding 'cage' in $PHF1_{Tudor}$ (C) and $BRPF1_{PWWP}$ (D, PDB accession number 2X4Y), and a proposed H3K36me3-binding cage in $Eaf3_{chromo}$ (E, PDB accession number 3E9G).

(F) Biotinylated H3K36me3 peptide pull-down using recombinant $PHF1_{Tudor}$ (left panel) or $PHF19_{Tudor}$ (right) proteins, either wildtype (WT) or containing a single point mutation. Top two panels are commassie blue stainings of pull-down samples, and bottom two panels are those showing amounts of proteins and peptides used.

(G) ITC titration experimental curves of wild-type (wt) or mutant forms of or $PHF19_{Tudor}$ bound to H3K36me3 peptides as indicated.

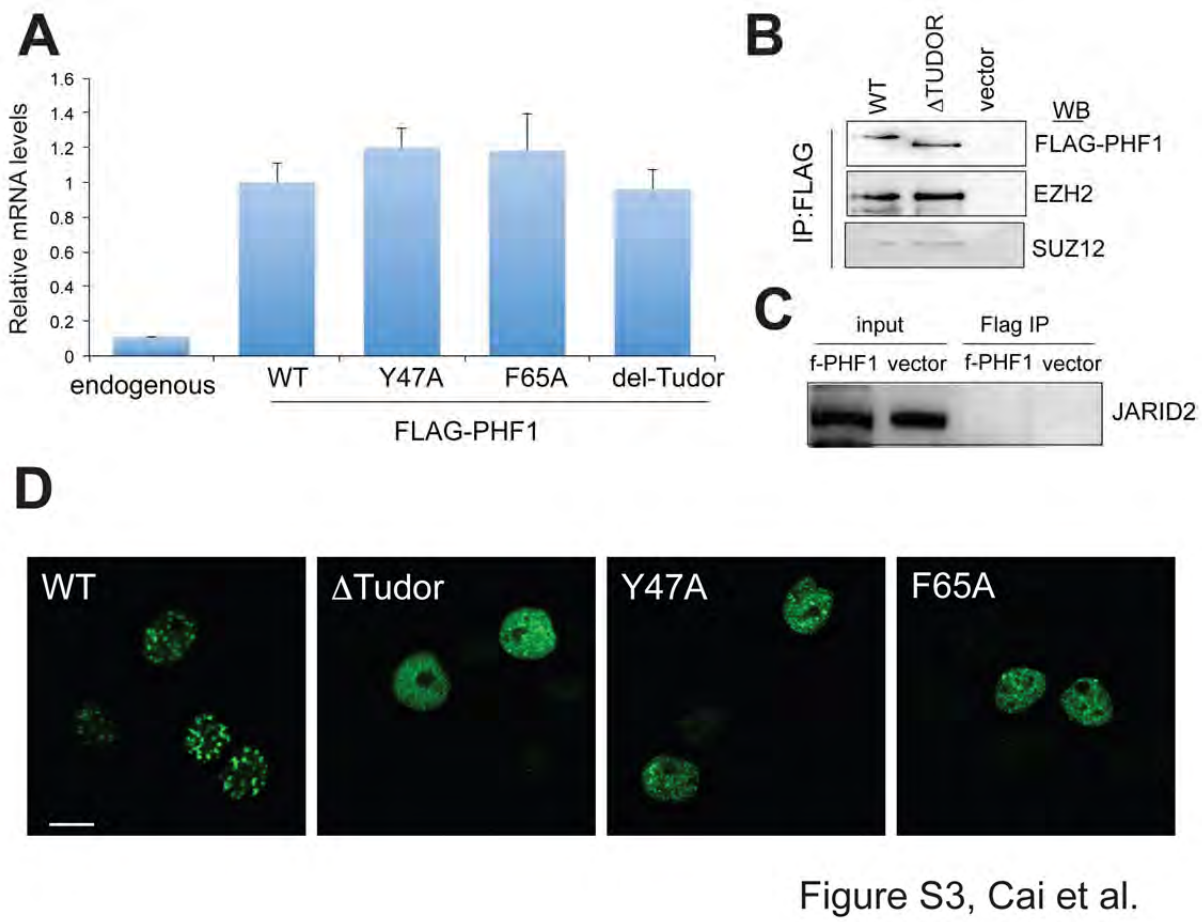



**Figure S3. PHF1 Facilitates Chromatin Localization of PRC2 Complexes through a TUDOR-Dependent Mechanism, Related to Figure 3**

(A) RT-qPCR analysis of PHF1 mRNA levels in stable expression cell lines used. Data are presented as mean ± SD.

(B-C) Western blots (WB) of FLAG IP samples to probe the association of wildtype (WT) versus Tudor-deleted PHF1 with the PRC2 complex (panel B), as well as association of wildtype PHF1 with JARID2 (panel C). Samples loaded are nuclear extracts prepared from 293 cells after transient transfection of empty vector, Flag-tagged wildtype or Tudor-deleted PHF1.

(D) Representative confocal immunofluorescence images revealing different nuclear distribution patterns of wildtype versus Tudor mutant forms of Flag-tagged PHF1 in transiently transfected 293 cells. Scale bar, 10 μm.

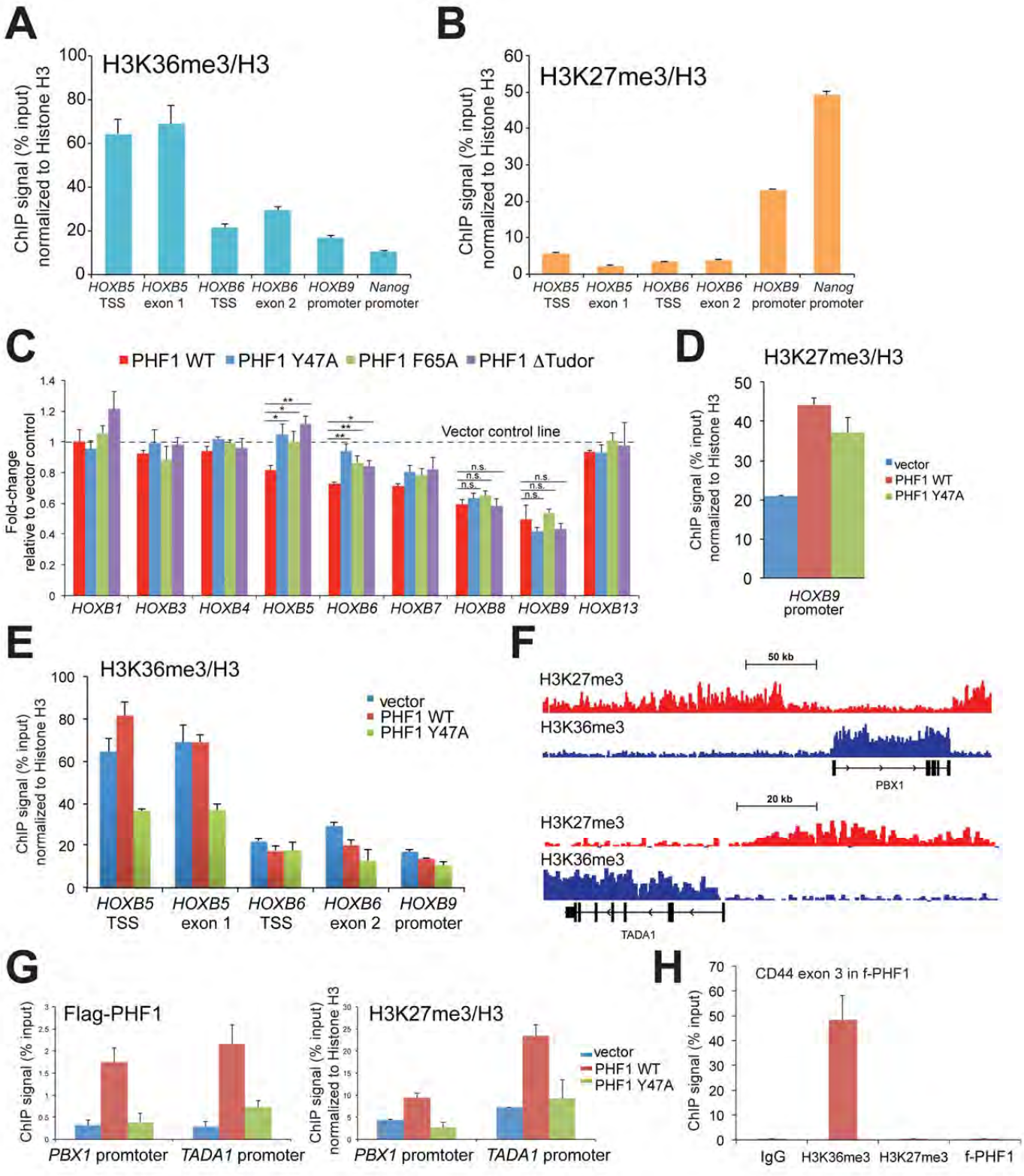



Figure S4. Cai et al

**Figure S4. PHF1-Mediated Spreading of PRC2 Complexes and H3K27me3 into H3K36me3 Regions through a TUDOR-Dependent Mechanism, Related to Figure 4**

(A-B) ChIP-qPCR verification of distribution of H3K36me3 (A) and H3K27me3 (B) at *HOX-B* gene clusters revealed by ChIP-Seq. Primers targeting promoter, transcription starting site (TSS) or exon were designed based on the ChIP-Seq results. Data of ChIP signals from three independent experiments were normalized to 1% of input used, data of histone modification ChIPs further normalized to those of total histone H3, and then presented as mean ± SD.

(C) RT-qPCR analysis of *HOX-B* mRNA levels in HeLa cells with ectopic expression of wildtype (WT) or mutant PHF1. Data of relative gene expression levels were normalized to those in HeLa transduced with empty vector (dash line), and presented as mean ± SD. *, $p<0.05$; **, $p<0.005$.

(D) ChIP for H3K27me3 at *HOXB9* promoter region in HeLa cells with ectopic expression of wildtype or mutant PHF1. ChIP signals were normalized to those of 1% of input, and levels of histone modification were normalized to total histone H3. Data presented as mean ± SD.

(E) ChIP for H3K36me3 at *HOX-B* gene clusters in HeLa cells with ectopic expression of wildtype or mutant PHF1. Data of ChIP signals were normalized to those of 1% of input, then normalized to those of total histone H3, and presented as mean ± SD.

(F) *PBX1* and *TADA1/ADA1* genes show a bimodal distribution of H3K27me3 and H3K36me3 as revealed by ChIP-Seq.

(G) ChIP for Flag-PHF1 and H3K27me3 at *PBX1* and *TADA1* promoter regions in HeLa cells with ectopic expression of wildtype or mutant PHF1. Data of ChIP signals from three independent experiments were normalized to 1% of input used, data of histone modification ChIPs further normalized to those of total histone H3, and then presented as mean ± SD.

(H) ChIP for H3K36me3, H3K27me3 and Flag-PHF1 at *CD44* exon 4 locus in Hela cells with ectopic expression of wildtype PHF1. ChIP signals were normalized to those of 1% of input. Data presented as mean ± SD.

Figure S5. Cai et al.

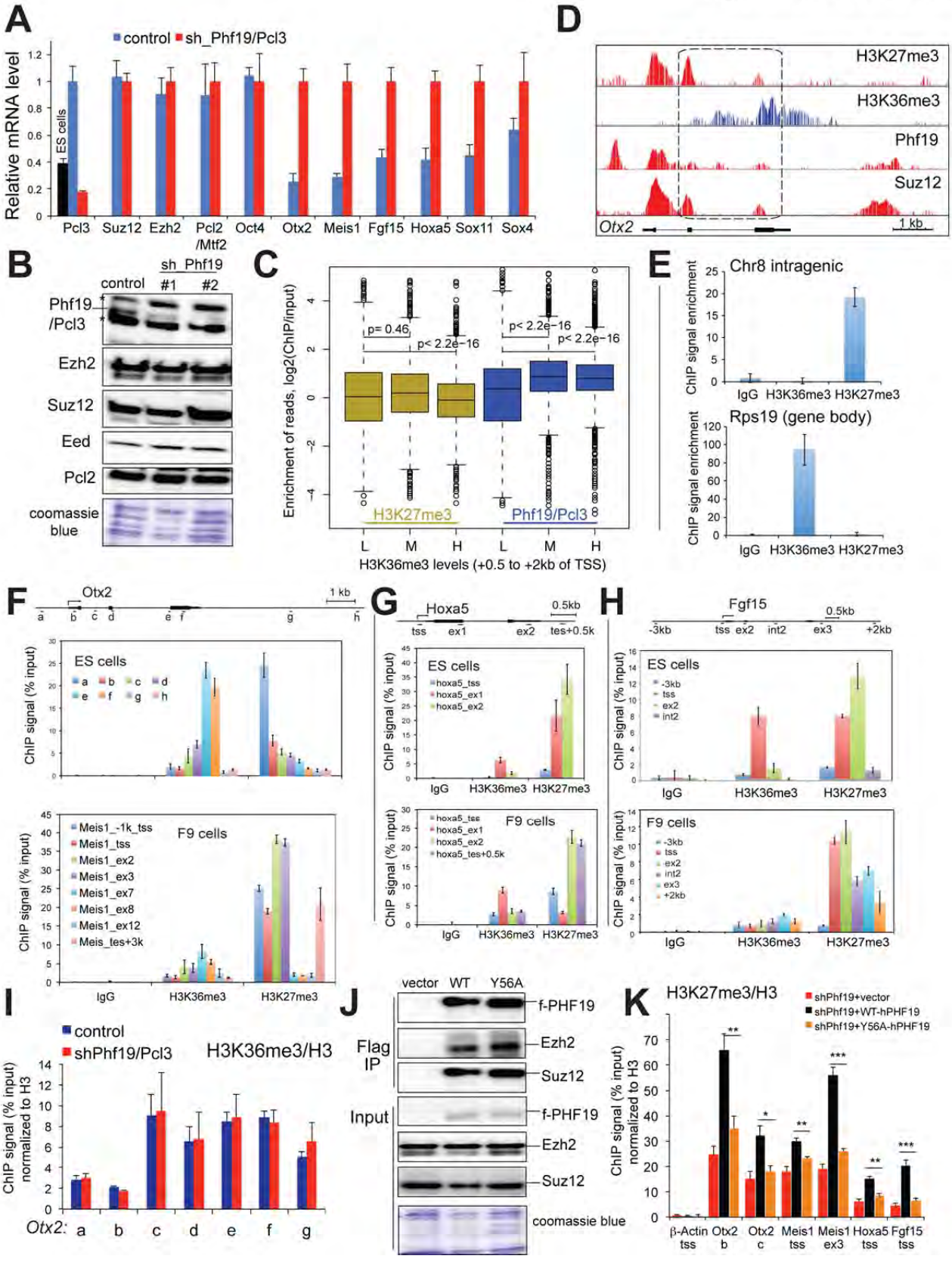


**Figure S5. PHF19/PCL3 Demonstrating a Tudor-Dependent Repression of Differentiation-Associated Genes, Related to Figure 5**

(A) RT-qPCR analysis for mRNA levels of genes as indicated on top of each panel in F9 embryonic carcinoma cells stably transduced with control (blue) or shRNA (red) against Phf19/Pcl3 (shPhf19). Relative mRNA levels (y axis) from three independent experiments were normalized to shRNA transduced cells (except those of Phf19 which were normalized to F9 control transduced cells), and presented as mean ± SD.

(B) Western blots using indicated antibodies and nuclear extracts prepared from F9 cells transduced with control or Phf19-specific shRNAs (two independent line #1-2). Stars indicate nonspecific bands. Coomassie blue staining was used for input control.

(C) ChIP-Seq revealing a moderate genome-wide correlation of higher Phf19/Pcl3 binding at promoters of genes with higher H3K36me3 levels. Genes with intermediate to low levels of H3K27me3 were separated into three groups based on their H3K36me3 modification: low (L), medium (M) and high (H). The H3K27me3 (gold) and Phf19/Pcl3 (blue) ChIP-Seq read counts (normalized by input) at promoters of these genes were compared and plotted as boxplots. For detailed description of analysis, see Supplemental Method.

(D) ChIP-Seq of murine ES cells revealing coexistence of H3K36me3, H3K27me3, and binding of Phf19 and Suz12 at the *Otx2* gene body as highlighted in the boxed region.

(E-H) ChIP for H3K36me3 and H3K27me3 at two control loci (E), *Otx2 (*F*), Meis1* (F), *Hoxa5* (G), and *Fgf15* (H) genes in either murine ES cells or F9 cells. Gene structure and ChIP primers targeting regions of putative transcription starting sites (tss), exon (ex), or transcription ending sites (tes) are shown at the top of each panel. Specificity of ChIP and re-ChIP (Figure 5B) was evaluated by two control loci, an H3K27me3-positive intragenic locus on chromosome

8 (E, top) and an H3K36me3-high locus at gene-body of a housekeeping gene, *Rps19* (E, bottom). Data of ChIP signals (y axis) from three independent experiments were normalized to 1% of input used and to histone H3, and then presented as mean ± SD.

(I) ChIP of H3K36me3 across the *Otx2* gene in F9 cells transduced with control (blue) or Phf19-specific shRNA (red). Data of ChIP (y axis) from three independent experiments were normalized to 1% of input and to histone H3, and presented as mean ± SD.

(J) Western blots using indicated antibodies of FLAG IP or nuclear extract input samples prepared from cells transduced with empty vector, a human wild-type or Tudor-mutant (Y56A) human PHF19 (Flag-tagged, f-PHF19).

(K) ChIP of H3K27me3 (normalized to total H3) at *Otx2, Meis1, Hoxa5* and *Fgf15* after transduction of hairpin-resistant wildtype or Tudor mutant (Y47A) human PHF19/PCL3 into knockdown cells. β-Actin tss were used as locus control. Data of ChIP (y axis) were normalized to 1% of input and to histone H3, and presented as mean ± SD. Statistics shown are t-test comparisons of ChIP in cells transduced with wild-type PHF19 to that with the Y47A form. *, $p<0.05$; **, $p<0.005$, ***, $p<0.005$.

## Table S2. Top Unique Hits of Mass Spec Sequencing of Flag-PHF1 Complexes Purified from 293 Cells, Related to Figure 3A

### PHF1

| Position | MH+ | Sequence |
|---|---|---|
| 531-538 | 808.9126 | GGVGYLSR |
| 95-103 | 959.0471 | SETVVPGNR |
| 305-313 | 983.1556 | LLSALNSHK |
| 520-530 | 988.0885 | SLSPGTGGGVR |
| 370-377 | 994.1387 | RPEPEPLR |
| 510-519 | 1084.2817 | RSAPPSPLCR |
| 395-403 | 1126.1710 | NQPEPQEQR |
| 369-377 | 1150.3262 | RRPEPEPLR |
| 370-378 | 1150.3262 | RPEPEPLRR |
| 384-394 | 1154.3086 | VEELGPPSAVR |
| 31-40 | 1187.3408 | LWEGQDVLAR |
| 305-315 | 1254.4317 | LLSALNSHKDR |
| 382-394 | 1339.5346 | GKVEELGPPSAVR |
| 173-184 | 1341.4256 | GLDWDAGHLSNR |
| 41-52 | 1380.6267 | WTDGLLYLGTIK |
| 41-53 | 1508.8008 | WTDGLLYLGTIKK |
| 380-394 | 1595.8394 | QKGKVEELGPPSAVR |
| 473-487 | 1664.9420 | SPLELHIGFPTDIPK |
| 10-30 | 2095.2792 | SGASSLWDPASPAPTSGPRPR |
| 488-509 | 2128.3216 | SAPHSMTASSSSVSSPSPGLPR |
| 450-472 | 2281.4182 | MFASFHPSASTAGTSGDSGPPDR |
| 335-367 | 3146.5452 | MPPPVEPPTGDGALTSFPSGQGPGGGVSRPLGK |

### SUZ12

| Position | MH+ | Sequence |
|---|---|---|
| 134-140 | 807.9373 | VDDMLSK |
| 247-253 | 886.0385 | SYSLLFR |
| 361-368 | 950.9805 | WTGETNDK |
| 114-121 | 1009.1682 | TLTYMSHR |
| 332-341 | 1134.2743 | ATWETILDGK |
| 404-413 | 1164.2578 | ESLTTDLQTR |
| 134-143 | 1164.3595 | VDDMLSKVEK |
| 131-140 | 1184.3931 | TFKVDDMLSK |
| 721-732 | 1193.2964 | ALETDSVSGVSK |
| 104-113 | 1194.4653 | NLIAPIFLHR |
| 369-380 | 1226.4618 | STAPIAKPLATR |
| 294-304 | 1287.5166 | TFVAQMTVFDK |
| 332-342 | 1290.4618 | ATWETILDGKR |
| 404-414 | 1292.4319 | ESLTTDLQTRK |
| 681-690 | 1303.5627 | LREMQQKLEK |
| 719-732 | 1450.5860 | EKALETDSVSGVSK |
| 102-113 | 1451.7579 | TRNLIAPIFLHR |
| 428-439 | 1593.7819 | IFYQFLYNNNTR |
| 596-611 | 1839.9507 | TITQIEEFSDVNEGEK |
| 538-555 | 2072.1647 | ASMSEFLESEDGEVEQQR |
| 287-304 | 2116.3535 | NREDGEKTFVAQMTVFDK |
| 361-380 | 2158.4191 | WTGETNDKSTAPIAKPLATR |
| 536-555 | 2301.4439 | TKASMSEFLESEDGEVEQQR |
| 381-403 | 2493.7795 | NSESLHQENKPGSVKPTQTIAVK |
| 6-40 | 2631.6943 | HGGGGGGGSGPSAGSGGGGFGGSAAVAAATASGGK |
| 691-718 | 2911.9634 | GESASPANEEITEEQNGTANGFSEINSK |
| 688-718 | 3282.4124 | LEKGESASPANEEITEEQNGTANGFSEINSK |

### EZH2

| Position | MH+ | Sequence |
|---|---|---|
| 40-46 | 828.9183 | SMFSSNR |
| 294-303 | 1079.1980 | EFAAALTAER |
| 53-61 | 1161.2999 | TEILNQEWK |
| 697-707 | 1336.5458 | YVGIEREMEIP |
| 53-63 | 1445.6181 | TEILNQEWKQR |
| 34-46 | 1527.6954 | RADEVKSMFSSNR |
| 65-78 | 1550.8417 | IQPVHILTSVSSLR |

| | | |
|---|---|---|
| 676-688 | 1589.7448 | AIQTGEELFFDYR |
| 49-61 | 1672.9217 | ILERTEILNQEWK |
| 434-451 | 1866.0350 | ESSIIAPAPAEDVDTPPR |
| 319-337 | 2055.2956 | LPNNSSRPSTPTINVLESK |
| 432-451 | 2093.3417 | VKESSIIAPAPAEDVDTPPR |
| 432-452 | 2221.5158 | VKESSIIAPAPAEDVDTPPRK |

**EZH1**

| Position | MH+ | Sequence |
|---|---|---|
| 32-39 | 832.9933 | LQANMGAK |
| 729-736 | 895.9878 | YSQADALK |
| 31-39 | 989.1808 | RLQANMGAK |
| 40-48 | 997.1821 | ALYVANFAK |
| 53-61 | 1161.2999 | TQILNEEWK |
| 737-747 | 1294.4473 | YVGIERETDVL |
| 481-492 | 1373.5640 | LPTDELMNPSQK |

**EED**

| Position | MH+ | Sequence |
|---|---|---|
| 376-385 | 1031.2584 | MLALGNQVGK |
| 314-322 | 1045.2668 | WLGDLILSK |
| 202-211 | 1086.2736 | DPNLLLSVSK |
| 368-375 | 1089.2538 | FSMDFWQK |
| 285-293 | 1091.2553 | TNRPFISQK |
| 294-302 | 1120.2530 | IHFPDFSTR |
| 276-284 | 1130.1559 | ESYDYNPNK |
| 343-355 | 1414.6450 | IKPSESNVTILGR |
| 5-19 | 1446.6155 | EVSTAPAGTDMPAAK |
| 107-120 | 1462.6024 | EGDPLVFATVGSNR |
| 5-20 | 1574.7896 | EVSTAPAGTDMPAAKK |

**RBBP4**

| Position | MH+ | Sequence |
|---|---|---|
| 16-22 | 895.0001 | VINEEYK |
| 297-304 | 974.1478 | TVALWDLR |
| 342-349 | 975.1325 | LNVWDLSK |
| 121-129 | 1068.1343 | INHEGEVNR |
| 5-15 | 1252.2800 | EAAFDDAVEER |
| 16-25 | 1322.5468 | VINEEYKIWK |
| 132-143 | 1379.6753 | YMPQNPCIIATK |
| 144-156 | 1472.6350 | TPSSDVLVFDYTK |

**RBBP7**

| Position | MH+ | Sequence |
|---|---|---|
| 5-14 | 1285.3682 | EMFEDTVEER |
| 131-142 | 1413.6776 | YMPQNPHIIATK |

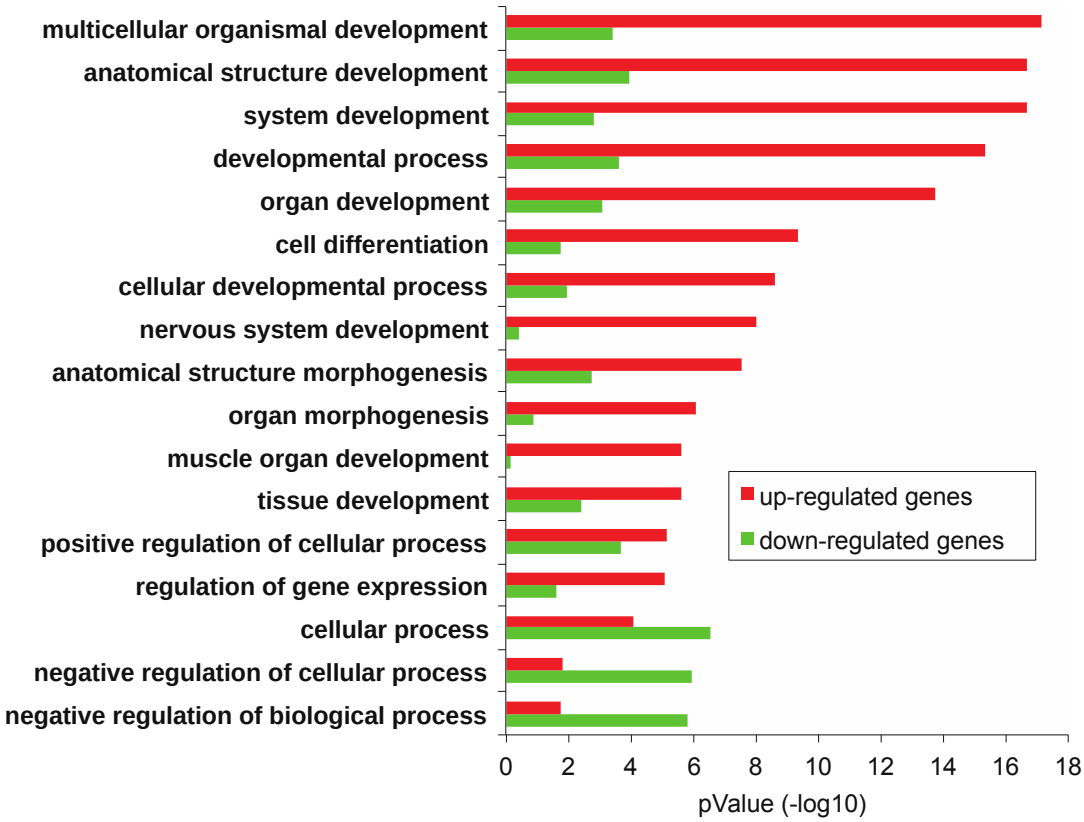
**HDAC1**

| Position | MH+ | Sequence |
|---|---|---|
| 474-482 | 973.1577 | GVKEEVKLA |
| 201-212 | 1375.4803 | YGEYFPGTGDLR |
| 78-89 | 1427.5720 | SIRPDNMSEYSK |

**SIN3A**

| Position | MH+ | Sequence |
|---|---|---|
| 677-685 | 944.0756 | AADIIDGLR |
| 1098-1104 | 955.0148 | WSDYVER |
| 866-875 | 1093.2681 | LLFSNTAAQK |
| 1044-1052 | 1097.2132 | SLLESTYQR |
| 156-167 | 1260.3897 | SQSIDTPGVISR |

**Table S3. GO Analysis of Genes Up- and Down Regulated in F9 Cells Transduced with Phf19-Specific Hairpin when Compared to Control, Related to Figures 5F and S5A**




**See Table S4 for the complete gene list.**

## Table S5. Information of primers used in this study, Related to Figures 4-5 and S4-S5

| Gene ID | Forward primer | Reverse primer | Note/Reference |
|---|---|---|---|
| ***RT-qPCR of human genes*** | | | |
| ***PHF1*** | agaaggggaaagtggaggag | ggtggaaggaagcaaacatc | |
| ***PHF1*** | gGCACTGCAGGCCTCAGTGT | tgtgaagttccaggggtgac | targeting 3'-UTR |
| ***HOXB1*** | 5'-CTCCTCTCCGAGGACAAGGAA-3' | 5'-CTGTCTTGGGTGGGTTTCTCTTAA-3' | |
| ***HOXB3*** | 5'-GCACCAACTCCACCCTCACC-3' | 5'-GCCACCACAGCCCTCTGC-3' | |
| ***HOXB4*** | 5'-CTACCCCTGGATGCGCAAAG-3' | 5'-TCCAGCTCCAAGACCTGCTG-3' | |
| ***HOXB5*** | 5'-TCCGCAAATATTCCCCTGGA-3' | 5'-CGGGTCAGGTAGCGGTTGAA-3' | |
| ***HOXB6*** | 5'-CTCCGGTCTACCCGTGGATG-3' | 5'-CCGCGTCAGGTAGCGATTGT-3' | |
| ***HOXB7*** | 5'-GACTTGGCGGCGGAGAGTAA-3' | 5'-CAGGGTCTGGTAGCGGGTGT-3' | |
| ***HOXB8*** | 5'-AACTCACTGTTCTCCAAATACAAAACC-3' | 5'-GACGGCCCGTGGTAGAACT-3' | |
| ***HOXB9*** | 5'-TTTGCGAAGGAAGCGAGGAC-3' | 5'-AGCTCCAGCGTCTGGTATTTGG-3' | |
| ***HOXB13*** | 5'-CCACTGGCTGCTGGACTGTT-3' | 5'-TATGACTGGGCCAGGTTCTTTG-3' | |
| ***GAPDH*** | gaaggtgaaggtcggagtc | gaagatggtgatgggatttc | |
| ***18s rRNA (18S5)*** | GGCGCCCCCTCGATGCTCTTAG | GCTCGGGCCTGCTTTGAACACTCT | Boulay et al 2011 |
| ***ChIP-qPCR of human genes*** | | | |
| ***HOXB5 TSS*** | TCCACAAATCAAGCCCTCCA | GATAACGCCCCGAGAAGGAG | |
| ***HOXB5 exon1*** | CCGGCTCTTACGGCTACAAT | CCAAAGTGGCTGGAGGAGG | |
| ***HOXB6 TSS*** | GGTGGAGCAGCCATGAAGAA | GCCAATCGCTGGATTCAACC | |
| ***HOXB6 exon2*** | GGCTCTCACGTTGGGACAAT | AATGGGTGGGGAGGAAGGAA | |
| ***HOXB9 promoter*** | CCACCGACTGGCTTCCTCGC | CCAGGGGTCACACCTCCCCA | |
| ***NANOG promoter*** | GAGGGGTGGGTCTAAGGTGA | ATGAGGCAACCAGCTCAGTC | |
| ***PBX1 promoter*** | CCTGTTGGAGGCTCTTCCTG | GGTTTTGAAGCAGTGCTGGG | |
| ***TADA1 promoter*** | CCCACACTTGGACTCAGCAT | ATGCGTAAAGGGGAAGAGCC | |
| ***CD44 exon 4*** | CCAACCTGAAGCATTGAAGC | CCAACCTGAAGCATTGAAGC | |
| ***RT-qPCR of murine genes*** | | | |
| ***Phf1/Pcl1*** | CAGCAGCGGCTACAACTTC | TGTGAAGTCCCAGAGGTGAC | |
| ***Phf19/Pcl3*** | ATATTGAGAGGCTGCCCTTG | AGCTCCCAGTGATGGTTGAC | |
| ***Mtf2/Pcl2*** | AGGGAATTGCACATTCATCC | CACAATGCCTGGAAATGCTA | Walker et al 2010 |
| ***Ezh2*** | AGACGTCCAGCTCCTCTGAA | ATCCTCAGTGGGAACAGGTG | Hunkapiller et al 2012 |
| ***Suz12*** | GTGCACTCTGAACTGCCGTA | CCGGTCCATTTCGACTAAAA | Hunkapiller et al 2012 |
| ***Eed*** | CTGGCAAAATGGAGGATGAT | TGGGTCAGTGTTGTGCATTT | Hunkapiller et al 2012 |
| ***Pou5f1/Oct4*** | atggggaaagaagctcagtg | caaaatgatgagtgacagacagg | gift of Laura Banaszynski |
| ***Otx2*** | AAGACCCGGTACCCAGACATC | TTGGCGGCACTTAGCTCTTC | Pasini et al., 2008 |
| ***Hoxa5*** | GCAAGCTGCACATTAGTCAC | GCATGAGCTATTTCGATCCT | |
| ***Meis1*** | AAGGTGATGGCTTGGACAAC | TGTGCCAACTGCTTTTTCTG | |
| ***Fgf15*** | GAGGACCAAAACGAACGAAATT | ACGTCCTTGATGGCAATCG | |
| ***Sox11*** | GTTGAATTCATACACTCCAATGT | GGAGATTGATCACACGATTT | Ling et al., 2009 |
| ***Sox4*** | GTTGGGGATGCAGAAGGA | TTTGCACAGACCCCAGGCGGAG | Ling et al., 2009 |
| ***Gapdh*** | ATGACATCAAGAAGGTGGTGAAG | TCCTTGGAGGCCATGTAGG | |
| ***18s RNA*** | AAGTCCCTGCCCTTTGTACACA | GCCTCACTAAACCATCCAATCG | gift of Ying Liu |
| ***ChIP-qPCR of murine genes*** | | | |
| ***Chr8 intragenic*** | AAGGGGCCTCTGCTTAAAAA | AGAGCTCCATGGCAGGTAGA | H3K27me3+; Wang et al 2009 |
| ***Actb tss*** | TTGATAGTTCGCCATGGATGACGA | ATCGATCCCCAAGAAAACCCCA | H3K27me3-; Wang et al 2009 |
| ***Rps19 gene body*** | TCTGCAGAGTGAGTGCCAGGACTATACA | CCATCCCAAGCTGTGTAAACCTGG | H3K36me3+; gift of Aaron Goldberg |
| ***Otx2 (a)*** | tatccgccattttgaggaag | ggcacttaaagcgctctctc | 1.5kb upstream of tss |
| ***Otx2 (b)*** | TTTTTAGTTAGTGCTGGAACGTGG | TGGGTAGATTTGGAGTGACGG | near tss; Pasini et al., 2008 |
| ***Otx2 (c)*** | atgccctgtccctttaatcc | ggggagggaggaatctgac | intron 1 |
| ***Otx2 (d)*** | AAGACCCGGTACCCAGACATC | ttctttggtctgggacaagg | exon 2 |
| ***Otx2 (e)*** | tagccatgtgatctgggatg | TTGGCGGCACTTAGCTCTTC | junction of intron2-exon3 |
| ***Otx2 (f)*** | CCAGCCATCTCAATCAGTCC | GGCAGTTTGGTCCTTATAATCC | exon3 |
| ***Otx2 (g)*** | cctatgttggctccaagagg | agacccttcctgtccttgc | 4kb downstream of tes |
| ***Otx2 (h)*** | gagaggaagaggaggcctaaa | tgaacttgaagggctgaaca | 7kb downstream of tes |
| ***Hoxa5 tss*** | atggaactgcgagggaaatg | cttccgacctcgggcttc | near tss/promoter |
| ***Hoxa5 ex1*** | CCCACATCAGCAGCAGAGAG | GGGTAGATCTGGGGCTGAG | exon 1 |
| ***Hoxa5 ex2*** | CACCTCGTTTAGTGCCAATG | GCTTGGAGCTATTGAGACAGG | exon 2 |
| ***Hoxa5 tes+0.5k*** | ccactgggaaactcctcaga | aaagacggcatccgtgtaag | 0.5kb downstream of tes |
| ***Fgf15 -3kb*** | CTGGCCTTCTGACTCTTTTCCT | CTGTGATCTCAAGGCAATCCTC | ~3kb upstream of tss |
| ***Fgf15 tss*** | GAGGAGAAATGCTCTGATGC | CACAACCCACAAAGCAACTG | near tss/promoter |
| ***Fgf15 ex2*** | cgctgtactcacgtgttgct | ATGGCAATCGTCTTCAGAGC | junction of intron1-exon2 |
| ***Fgf15 int2*** | caagggaatatggcaccagt | aaatgaatgcagtgggaagc | intron 2 |
| ***Fgf15 ex3*** | TCCTGGCCAGAACTAACTGG | TAAGTGGAAAAGGGGGAAGC | exon 3 |
| ***Fgf15 +2kb*** | AAGGATTCCTGCTGCAAATGT | AATGGTCATCCCAGGAGACTG | 2kb downstream of tes |
| ***Meis1 -1k_tss*** | cgctcggctcattgttcc | gtccccacctcctctcttg | ~1kb upstream of tss |
| ***Meis1 tss*** | gcagttgcaaagagggagag | gcccgctttccttgaatc | near tss/promoter |
| ***Meis1 ex2*** | atctaagctcggtgccctaac | ATGCCTACTCCATCCATACCC | exon 2 |

| | | | |
|---|---|---|---|
| ***Meis1 ex3*** | ACCCCCTCTTCCCTCTCTTAG | AATGACTCTGACGAGCAGACG | exon 3 |
| ***Meis1 ex7*** | GCCCTCTTGGAATAGAGACCA | CACTGCTGTTATCCCCACTGT | exon 7 |
| ***Meis1 ex8*** | GCACAGGTGACGATGATGAC | ACGCCCTCATGATATTGGTG | exon 8 |
| ***Meis1 ex12*** | agagcaatccatgcagaagac | gaggggaagaaaggcttcac | intron 11- exon 12 |
| ***Meis1 tes+3k*** | ctagcgggcctagagaagatg | ggaattccagcgtaaaagagc | 3kb downstream of tes |

## SUPPLEMENTAL EXPERIMENTAL PROCEDURES

**Biotinylated peptide pull-down.** Pull-down assays using biotinylated histone peptides were performed as previously described (Wang et al., 2009; Wysocka et al., 2006) with the following modifications. Biotin-conjugated histone peptides were incubated for 2-3 hours with high-capacity neutraAvidin resin (Pierce), followed by extensive washing to remove unbound peptides. To identify interacting proteins from nuclear extracts, ~50 μl of peptide-Avidin resin complexes were incubated with nuclear extracts in buffer containing 20 mM Hepes pH 7.9, 200mM KCl, 0.05% NP-40, 5% glycerol, 1mM DTT, 0.1mM PMSF and protease inhibitor cocktail (Roche) (Wysocka et al., 2006). To pull down recombinant proteins or extracts of cells with overexpressed candidate proteins, ~10 μl of peptide-Avidin resin was typically used (Wang et al., 2009; Wysocka et al., 2006). After extensive washing, proteins bound to resin were separated by SDS-PAGE, and examined by silver stain followed by mass spectrometry, by coomassie blue staining, or by western blot.

**Tissue culture.** HEK293, HeLa, mouse E14 embryonic stem (ES) cells, and F9 embryonic carcinoma cells were obtained from ATCC and maintained using recommended culture conditions. MSCV-based retrovirus encoding FLAG-tagged PHF1 was generated and used for establishing stable expression cell lines as previously described (Wang et al., 2009). After 10 days of drug selection, stable transgene expression was examined by quantitative RT-PCR and immunoblot analysis.

**Antibodies and immunoblot.** Antibodies used in this study are those against FLAG (M2, Sigma), GST (GE Life Science), EZH2 (BD bioscience 612666, Millipore 07-400), SUZ12 (Abcam, ab12073), PHF19/PCL3 (Proteintech 11895-1-AP), JARID2 (Abcam, ab48137), MTF2/PCL2 (Active Motif 61153) and Tubulin (Sigma). Histone antibodies used include those against general H3 (Abcam ab1791), H3K36me3 (Abcam ab9050), H3K27me3 (Millipore 07-449 and Abcam ab6002), H3K4me3 (Abcam ab8580), H3K9me3 (Abcam ab8898), H4K20me3 (Abcam ab9053), H3R2me2a (Millipore 07-213), H3R17me2a (Millipore 07-214), and H4R3me2a (Millipore 05-808).

**Co-immunoprecipitation (CoIP) and GST pull-down.** Samples used for CoIP were either nuclear extracts prepared according to the Dignam protocol (Dignam et al., 1983) or mononucleosomal preparation using limited micrococcal nuclease (MNase) digestion as previously described (Ruthenburg et al., 2011; Wang et al., 2009). For FLAG IP, high salt extracted or MNase-digested nuclear preparation was incubated with M2 FLAG resin (Sigma). After extensive washing, M2 FLAG resin was added with protein sample buffer, and bound proteins were subject to SDS-PAGE separation and western blot. Examination of protein-histone association was performed according to a previously described crosslink-based protocol (Ricke and Bielinsky, 2005). GST pull down of mononucleosomes was carried out as described before (Ruthenburg et al., 2011; Wang et al., 2009).

**Peptide synthesis and isothermal Titration Calorimetry (ITC**). Peptides used were synthesized by UNC High-Throughput Peptide Synthesis and Peptide Array Facility as previously described (Fuchs et al., 2011; Rothbart et al., 2012), and quality of all used peptides were examined using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry and analytical HPLC (>98~100% in purity). ITC experiments were carried out at

UNC Macromolecular Interactions Facility using a MicroCal AutoITC-200 system (GE Healthcare). Typically, peptides and proteins were subject to extensive dialysis against the same buffer, and ITC titrations were performed at 4-6°C in a buffer consisting of 100 mM NaCl, 20 mM Tris-HCl pH 7.5, and 2mM beta-mercaptoethanol using a previously described protocol (Ruthenburg et al., 2011; Wang et al., 2009). Proteins at concentrations of ~ 70-200 μM were loaded into the ITC cell and peptides at 10-15 fold higher concentrations were loaded into the ITC syringe. 19 injections of 2-μl peptide were made automatically into the cell. The resulting binding isotherms were analyzed using the Origin 7.0 software package (Origin Lab) and were fit to a one-site binding model (Wang et al., 2009).

**Purification of Flag-PHF1 complexes** Nuclear extract was prepared using the Dignam protocol (Dignam et al., 1983) from nuclei of about five liters of suspension culture of a HA-Flag-PHF1 293 stable cell line. After dialysis against low salt buffer (150 mM NaCl, 20mM Hepes pH 7.90, 0.025% NP40, 0.2 mM EDTA, 10% glyecerol, 1 mM DTT, 0.2 mM PMSF), nuclear extract was subject to high-speed centrifugation to remove precipitation and then applied to a heparin affinity purification column. Following a salt gradient elution step on ATKA chromatography system (GE life science), fractions were collected and probed with anti-Flag (M2) immunoblot. Fractions that contain Flag-PHF1 were combined, dialyzed against low salt buffer, and subject to affinity purification using M2 anti-Flag resin. After extensive washing, bound proteins were eluted with 3xFlag peptides, separated on a protein gel, and visualized by silver staining. The whole protein lane was excised into pieces and subjected to liquid chromatography-tandem mass spectrometry (LC-MS/MS) analysis (Harvard Medical School).

**shRNA knockdown and rescue construction** Lentivirus-based shRNA plasmids (pLKO.1) that target mouse *Phf19* were obtained from Open Biosystems. Lentivirus was generated in

293 cells and used according to providers' protocols. The hairpin (TRCN0000096073, with sequence: CCTCAAGTCCTCTATCACCAA) showed the best knockdown of Phf19/Pcl3 and was used to generate stable cell lines. An shRNA-resistant rescue cDNA was generated by introducing silent mutations (highlighted by underline: CCT**G**AAGTC**A**TC**A**ATCAC**G**AA) to the hairpin target region of human *PHF19*.

**Immunofluoresence** For immunofluoresence assays, 2xFLAG-tagged PHF1 transfected 293T cells were fixed with cold methanol and permeablized by 0.2% Triton X-100, followed by immunostaining with M2 anti-FLAG antibodies (Sigma) and Alexa 488 conjugated secondary antibody. The nuclei were stained with 4, 6-diamino-2-phenylindole (DAPI, 0.1 μg/ ml). Fluorescence was detected in an LSM 510 confocal microscope. Coefficient of co-localization and statistical analysis was examined as described previously (Wang et al., 2009).

**Quantitative RT–PCR analysis.** Reverse transcription of RNA was performed using random primers and cDNA Reverse Transcription kits according to the manufacturers protocols (Invitrogen and Applied Biosystems). PCR amplicons (80–150 bp) were designed to span over large intronic regions. Exon–intron information was obtained from the UCSC genome browser (http://genome.ucsc.edu/). Quantitative PCR was performed in triplicate using the SYBR Green master mix reagent (Applied Biosystems) on an ABI 7900 qPCR system. Information of primers used is described in Table S5.

**Microarray analysis.** Total RNA was extracted and the transcript expression was quantified using Affymetrix GeneChip Mouse arrays as described (Wang et al., 2009). RNA hybridization, scanning and signal quantification were performed by the UNC Functional Genomics Core. Hybridization signals were retrieved and normalized, followed by differential expression analysis and statistical analysis using GeneSpring Analysis Platform GX 11 (Agilent

Technologies). GO analysis of genes showing fold of change over 1.5 folds was performed using DAVID Functional Annotation Tool (http://david.abcc.ncifcrf.gov).

**ChIP-sequencing (ChIP-Seq) and data analysis.** ChIP-Seq data for H3K27me3 and H3K36me3 in HeLa cells were obtained from the ENCODE Consortium (NCBI GEO database, GSE29611, ENCODE project and Broad Institute). ChIP-Seq data of Phf19/Pcl3 binding sites, H3K27me3 and H3K36me3 in mouse ES cells were obtained from previous publications (Hunkapiller et al., 2012; Mikkelsen et al., 2007; Xiao et al., 2012). For each of the transcripts (n= 25,296) of mouse Refseq genes, we counted ChIP-seq reads located at its promoter (-1kb to +500 bp of TSS) and the TSS-adjacent transcribed regions (+500 bp to +2kb of TSS). Gene sets with intermediate and low H3K27me3 were generated based on histogram analysis (data not shown) and by excluding genes with high H3K27me3 at their promoters (>= 50 reads, ~25% of total), and then we divided the rest into three groups based on H3K36me3 at the TSS-adjacent regions: low (< 5 reads, n=6,200), medium (5 ~ 12 reads, n=6,554), and high (>= 12 reads, n=6633). The Phf19/Pcl3 and H3K27me3 abundance at the promoters of these three groups of genes were then analyzed and compared for significant difference using the Student's t-test. The Phf19/Pcl3 ChIP-Seq was done with two biological replicates (Hunkapiller et al., 2012), which yielded identical results in our analysis (data not shown).

**Statistics.** All results are presented as the mean ± s.d. for at least three independent experiments unless otherwise noted. Where indicated, statistical analyses were performed using a Student's t-test.

**NMR spectroscopy.** NMR spectra were collected at the New York Structural Biology Center (NYSBC) using 600, 700, and 900 MHz Bruker spectrometers equipped with $^{1}H$, $^{15}N$, $^{13}C$ triple-resonance cryogenic probes. Unless indicated otherwise, the sample temperature was

controlled at 25°C. A suite of 3D heteronuclear NMR experiments, including HNCACB, CBCA(CO)NH, HNCO, HBHA(CO)NH, and HCCH-TOCSY were acquired for sequential backbone and non-aromatic side chain assignments of $PHF1_{Tudor}$ free in solution and in the complex with $H3_{31-41}$K36me3. 2D NOESY ($\tau_{mix}$ = 100 ms), 3D $^{15}$N-edited NOESY-HSQC ($\tau_{mix}$ = 100 ms), 3D aromatic $^{13}$C-edited NOESY-HSQC ($\tau_{mix}$ = 100 ms) and 3D aliphatic $^{13}$C-edited NOESY-HSQC ($\tau_{mix}$ = 100 ms) data sets were acquired and used for additional assignments (side chain amide groups and aromatic groups) and distance constraints. In addition, a [$^{13}$C,$^{15}$N]-filtered,$^{13}$C-edited NOESY spectrum (Zwahlen, 1997) was collected for the [$^{15}$N,$^{13}$C]-labeled $PHF1_{Tudor}$ bound to unlabeled $H3_{31-41}$K36me3 to obtain intermolecular NOEs. The spectra were processed and analyzed, respectively, with the NMRPipe (Delaglio, 1995) and Sparky (http://www.cgl.ucsf.edu/home/sparky) software packages.

**Structure calculations** The structure of $PHF1_{Tudor}$ in complex with $H3_{31-41}$K36me3 was first calculated using the CYANA program (Guntert et al., 1997) and subsequently refined with the Xplor-NIH program (Schwieters et al., 2003). Interproton distance constraints were derived from 2D NOESY, 3D $^{15}$N-edited NOESY-HSQC and 3D $^{13}$C-edited NOESY-HSQC spectra. Backbone $\phi$ and $\psi$ angles were derived from TALOS-based analysis of backbone chemical shifts(Cornilescu et al., 1999). A number of hydrogen bonds derived from chemical shifts analysis and from observed NOEs characteristic for $\alpha$-helices and $\beta$-sheets, were added in the final rounds of structure refinement. The final structures were validated by Procheck-NMR (Laskowski et al., 1996), and the statistics for the 20 final structures are listed in Table 2.

**NMR protein sample preparation** The gene encoding residues 7-83 of $PHF1_{Tudor}$ was N-terminally fused to SUMO preceded by a His(x6)-tag in a modified pRSFDuet-1 vector (Novagen Inc.), with a ubiquitin-like-protease 1 (ULP1) cleavage site in the linker between the

two proteins. Uniformly $^{15}N$,$^{13}C$-labeled protein was expressed in *E. Coli* BL21(RIL) cell strain in M9 minimum medium supplemented with $^{13}C$-labeled glucose and $^{15}N$-labeled $NH_4Cl$. The fusion protein was first purified using Ni-NTA affinity column and then subject to UPL1 cleavage overnight. After UPL cleavage, $PHF1_{Tudor}$ was separated from the SUMO tag through a second Ni-NTA chromatography step followed by gel filtration. For preparation of NMR sample, 0.3-0.5 mM [$^{13}C$, $^{15}N$]-$PHF1_{Tudor}$, mixed with $H3_{31-41}$K36me3 in a molar ratio of 1:2, was dissolved in NMR buffer (20 mM sodium phosphate, 50 mM NaCl, 5 mM DTT, 90%$H_2O$/10% $D_2O$) at pH 7.0.

## SUPPLEMENTAL REFERENCES


Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. Journal of biomolecular NMR *13*, 289-302.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic acids research *11*, 1475-1489.

Fuchs, S.M., Krajewski, K., Baker, R.W., Miller, V.L., and Strahl, B.D. (2011). Influence of combinatorial histone modifications on antibody and effector protein recognition. Current biology : CB *21*, 53-58.

Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997). Torsion angle dynamics for NMR structure calculation with the new program DYANA. Journal of molecular biology *273*, 283-298.

Hunkapiller, J., Shen, Y., Diaz, A., Cagney, G., McCleary, D., Ramalho-Santos, M., Krogan, N., Ren, B., Song, J.S., and Reiter, J.F. (2012). Polycomb-like 3 promotes polycomb repressive complex 2 binding to CpG islands and embryonic stem cell self-renewal. PLoS Genet *8*, e1002576.

Laskowski, R.A., Rullmannn, J.A., MacArthur, M.W., Kaptein, R., and Thornton, J.M. (1996). AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. Journal of biomolecular NMR *8*, 477-486.

Ling, K.H., Hewitt, C.A., Beissbarth, T., Hyde, L., Banerjee, K., Cheah, P.S., Cannon, P.Z., Hahn, C.N., Thomas, P.Q., Smyth, G.K.*, et al.* (2009). Molecular networks involved in mouse cerebral corticogenesis and spatio-temporal regulation of Sox4 and Sox11 novel antisense transcripts revealed by transcriptome profiling. Genome Biol *10*, R104.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P.*, et al.* (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Pasini, D., Hansen, K.H., Christensen, J., Agger, K., Cloos, P.A., and Helin, K. (2008). Coordinated regulation of transcriptional repression by the RBP2 H3K4 demethylase and Polycomb-Repressive Complex 2. Genes & development *22*, 1345-1355.

Ricke, R.M., and Bielinsky, A.K. (2005). Easy detection of chromatin binding proteins by the Histone Association Assay. Biological procedures online *7*, 60-69.

Rothbart, S.B., Krajewski, K., Strahl, B.D., and Fuchs, S.M. (2012). Peptide microarrays to interrogate the "histone code". Methods Enzymol *512*, 107-135.

Ruthenburg, A.J., Li, H., Milne, T.A., Dewell, S., McGinty, R.K., Yuen, M., Ueberheide, B., Dou, Y., Muir, T.W., Patel, D.J., and Allis, C.D. (2011). Recognition of a mononucleosomal histone modification pattern by BPTF via multivalent interactions. Cell *145*, 692-706.

Schwieters, C.D., Kuszewski, J.J., Tjandra, N., and Clore, G.M. (2003). The Xplor-NIH NMR molecular structure determination package. J Magn Reson *160*, 65-73.

Wang, G.G., Song, J., Wang, Z., Dormann, H.L., Casadio, F., Li, H., Luo, J.L., Patel, D.J., and Allis, C.D. (2009). Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. Nature *459*, 847-851.

Wysocka, J., Swigut, T., Xiao, H., Milne, T.A., Kwon, S.Y., Landry, J., Kauer, M., Tackett, A.J., Chait, B.T., Badenhorst, P.*, et al.* (2006). A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. Nature *442*, 86-90.

Xiao, S., Xie, D., Cao, X., Yu, P., Xing, X., Chen, C.C., Musselman, M., Xie, M., West, F.D., Lewin, H.A.*, et al.* (2012). Comparative epigenomic annotation of regulatory DNA. Cell *149*, 1381-1392.