Supporting Information

Ali et al. 10.1073/pnas.1310156110

SI Materials and Methods

Cell Culture and Fly Stocks. C2C12 were grown in DMEM media supplemented with 10% (vol/vol) inactive FBS plus penicillin/streptomycin (Invitrogen) at 37 °C and diluted 1:20 every 2 d. S2 cells were maintained at 25 °C in CCM3 media (HyClone) supplemented with 10% (vol/vol) inactive FBS and penicillin/streptomycin and diluted 1:10 every 3 d. Oregon R flies were used in this study.

Lentiviral-Based DNA Adenine Methyltransferase Identification Chromatin Profiling. Murine mixed-lineage leukemia 5 (MLL5; 1868 aa) was cloned from RNA extracted from mouse erythroid progenitors (CD71+, low Ter119+) (1). The cDNA was sequenced to confirm its integrity. An entry clone was obtained by amplifying MLL5 cDNA with attR site-containing primers and transferred to the destination vector pLgw EcoDam-V5-RFC1 using the gateway system (2). Generation of viral particles was performed at the viral core facility (Fred Hutchinson Cancer Research Center) using a combination of the DNA adenine methyltransferase (Dam) fusion plasmid and the GFP expression lentiviral vector pFG12 in a 9:1 ratio. Virus was titered in C2C12 cells based on GFP expression until reaching 10% of GFP+ cells. The 1×10^6 cells were infected in the presence of polybrene (4 mL/mL), and genomic DNA was extracted 48 h postinfection. Genomic DNA was treated and processed according to ref. 2. Dam-UpSET and Dam DNA were labeled by random priming with Cy5 and Cy3 hexamers, respectively, and hybridized to a NimbleGen promoter tiling array (MM8_RefSeq_promoter design covering ~32,000 Refseq genes plus CpG islands) according to the manufacturer's protocol (Roche NimbleGen Inc.).

Data Analysis. Data analysis was performed using R/Bioconductor (www.R-project.org/; www.bioconductor.org/). Raw signals of corresponding experimental replicates were loess-normalized within arrays and quantile-normalized between arrays. Enrichment statistics (test vs. control signals) were computed using the SAM algorithm within Bioconductor (3). FDR values of the SAM statistic were determined using the package locfdr. Region summarization was performed using the HMM algorithm of TileMap (4). We applied this HMM model to reliably establish the regions in the genome bound by MLL5 (5). Probes were considered to be bound significantly if the posterior probability of the HMM was greater than 0.9 and larger than 500 bp for promoter arrays. Statistical tests were performed using R defaults. Track visualization was carried out using the University of California, Santa Cruz, browser as the ENCODE RNA-sequencing (RNA-seq) and ChIP-sequencing (ChIP-seq) data are already available in this interface. *P* values were obtained with Welch two-sample *t* test to evaluate significance to unsorted promoters. The 4,000 random sequences ranging between 1 and 3 kb were generated in XLSTAT considering a similar number of binding sites for each quintile set of genes. Data used the *Mus musculus*-built mm8. End analyses (meta-analysis) to transcriptional start sites (TSSs) and CpG islands were performed as described (6). To evaluate overlaps between MLL5-bound promoters and histone modifications in C2C12 cells, we lift over the MLL5-bound regions from the built mm8 to mm9. Overlaps were analyzed using Galaxy (7–9).

RNA-seq data for C2C12 cells was obtained from the mouse ENCODE consortium (Wold Lab, California Institute of Technology) as a bam file. Using htseq-count, we generated counts from the bam files, and reads mapping multiple locations were discarded. The resulting genes and their reads were divided into quintiles from low-expressing (Q1) to high-expressing (Q5) genes for Fig. 1 *F* and *G* and Fig. S1 *C–F*. A total of 121 genes were removed that had zero counts in both samples.

Only data generated for C2C12 cells (myoblast) was used in this study: RNA-seq (GSM929774), RNA polymerase, H3K4me3, H3K4me2, H3K4me1, and H3K27me3 (GSE25308).

Antibodies. MLL5s antibody was obtained from Orbigen and shown to be specific to this isoform (10). Antibodies used in this study include rabbit anti-H3K4me3 (Active Motif), H3K9me3, and H3T6ph (Abcam). Antibodies for UpSET are described in ref. 11. For generating a MLL5 antibody that recognizes both long and short isoforms, a DNA fragment corresponding to the 1- to 100-aa region of MLL5 was cloned in-frame into pGEX-5X. The GST–MLL5 fusion protein was produced and purified from *Escherichia coli* (Rossetta; Novagen/EMD/Millipore). MLL5 antibody was produced and protein A purified by Thermo Scientific.

Indirect Immunofluorescence. For S2 cell staining, 2×10^6 cells were resuspended in 100 µL of growing media and allowed to attach to poly-L-lysine–coated slides for 30 min in a humid chamber. For C2C12 cell staining, 5×10^4 cells were grown on four-chamber slides for 24 h. Cells were washed in PBS and fixed in 2% EM-grade formaldehyde (Polysciences, Inc.) diluted in PBS for 10 min at room temperature. Cells were washed twice with PBS and permeabilized with 0.2% Triton X-100 in PBS. Slides were blocked with 1.5% BSA in PBS. Primary antibody was incubated for 1.5 h and subsequently washed twice with PBS + 0.05 Tween-20 (Fisher Scientific). Secondary antibodies were incubated for 1 h at room temperature. Alexa 488 and 568 (1:1,000 dilution; Invitrogen Life Technologies) were used for fluorescence visualization. A total of 0.1 µg/mL of DAPI (Invitrogen Life Technologies) was used for DNA. The antibodies were diluted as follows: anti-H3T3ph (1:300), anti-H3T6ph (1:300), anti-UpSET N (P2E7 and P5E7, 1:10), and anti-UpSET C (P1G11 and P2D11, 1:10).

For costaining with MLL5 antibodies and histone modifications, we directly labeled the antibodies using a Zenon Alexa 488 and 568 Rabbit IgG Labeling Kit following the manufacturer's instructions (Invitrogen Life Technologies). Direct-labeled antibodies were incubated for 1.5 h at room temperature in the presence of 1.5% BSA in PBS. After three consecutive washes with PBS, slides were fixed with 4% EM-grade formaldehyde (Polysciences, Inc.) for 10 min at room temperature. Slides were washed twice with PBS and mounted in SlowFade Gold with DAPI (Invitrogen Life Technologies). A slight increase in background staining was observed with this technique, probably due to the passive diffusion of the reagents from the kit. Cells were imaged on a DeltaVision microscope. The 1-µm average projections of 0.2-µm stacks (5) are shown with adjustments in brightness and contrast only.

Ovary Immunostaining. Immunofluorescence and confocal microscopy were performed as described previously (12, 13) using a Zeiss LSM 780 confocal microscope. The 2-d-old well-fed females were dissected in CCM3 media, washed in PBS, and fixed in 6% EM-grade formaldehyde (Polysciences, Inc.) diluted in PBS, with 3× vol of heptane. After washing, tissues were blocked in 1.5% BSA, hand-dissected, then incubated with primary antibodies at 4 °C overnight. The following antibodies were used: anti-H3T6ph (1:300), a combination of five monoclonal antibodies for UpSET mix (1:10). Alexa 488 and 568 (1:1,000 dilution; Invitrogen Life Technologies) were used for fluorescence visualization, and 0.1 µg/mL of DAPI (Invitrogen Life Technologies) was used for

DNA. Ovaries were further mounted in SlowFade Gold with DAPI (Invitrogen Life Technologies).

Polytene Chromosome Staining. Salivary gland polytene chromosomes from wild-type third instar larvae were fixed and prepared as previously described (14). Mouse monoclonal antibodies for UpSET (PE57 and P1G11) were used in a 1:6 dilution. Other antibodies used at 1:500 dilution were rabbit anti-H3T3ph (Upstate-Millipore) and anti-H3T6ph (Abcam). Polytene chromosomes were imaged on a DeltaVision microscope. Average projections of 0.2-mm stacks are shown with adjustments in brightness and contrast only.

Cloning and Protein Purification. The plant homeodomain (PHD) fingers of MLL5 (117–181), Set3 (115–183), and UpSET (852–917) were cloned into a pCool (modified pGEX2T) expression vector with ampicillin resistance. The proteins were expressed in *E. coli* BL21 Rosetta-2 (DE3) pLysS cells grown in LB or in $^{15}NH_4Cl$ or $^{15}NH_4Cl$ and ^{13}C-glucose minimal media supplemented with 60 μM $ZnCl_2$. After induction with isopropyl β-D-1-thiogalactopyranoside (0.5 mM) for 16 h at 18 °C, cells were harvested by centrifugation and lysed by sonication. The unlabeled and ^{15}N-labeled GST fusion proteins were purified on glutathione Sepharose 4B beads (GE Healthcare). The GST tag was cleaved with thrombin protease (GE Healthcare). The proteins were concentrated into PBS buffer (pH 7.4) supplemented with 5 mM DTT.

X-Ray Crystallography. The MLL5 PHD domain was crystallized in bound state with H3K4me3 (1–12) peptide. The purified PHD domain (9 mg/mL) was incubated overnight with the histone peptide in a 1:1.5 molar ratio in PBS buffer (pH 7.4) and 5 mM DTT before setting up crystallization. Initial crystallization was set up in 96-well sitting-drop plates at 18 °C and 25 °C, mixing 1.0 μL each of protein–peptide solution and the precipitant conditions. The condition was optimized to 0.14 mM potassium acetate and 35% PEG 3350 at 18 °C. Good diffraction-quality crystals were obtained by seeding and growing by hanging-drop vapor diffusion method at 18 °C. Data sets were collected at the Advanced Light Source 4.2.2 beamline. The structure of the complex was solved using single-wavelength anomalous dispersion with Zn anomalous signal at 1.0 Å. The crystal diffracted up to 1.48 Å and was determined to belong to space group $P2_12_12_1$. HKL2000 was used to process the data sets. Location of each Zn atom was determined using SHELX C/D. Initially, a partial model was build using Phenix AutoBuild. Manual model-building was performed using Coot, and the structure was refined using Phenix refine. The final structure was verified by PROCHECK.

PCR Mutagenesis. Point mutants were prepared using the Stratagene QuikChange XL Site-Directed Mutagenesis Kit according to the manufacturer's instructions.

Peptide Microarray. Peptide synthesis and validation, microarray fabrication, effector protein hybridization and detection, and data analysis were performed as previously described (15) with the following modification. Each peptide listed in Dataset S1 was spotted in triplicate eight times per array. Triplicate spots were averaged and treated as a single value for subsequent statistical analysis.

NMR Spectroscopy and Sequence-Specific Assignments. NMR experiments were performed at 298 K on a Varian INOVA 600-MHz spectrometer using pulse-field gradients to suppress artifacts and eliminate water signal. The NMR samples contained 0.95-mM uniformly $^{15}N/^{13}C$-labeled MLL5 PHD in PBS buffer (pH 6.0) and 10% D_2O. The $^1H,^{15}N$ heteronuclear single-quantum coherence (HSQC), HNCACB (16), CBCA(CO)NH (17), C(CO)NH (18), and HC(CO)NH (18) were recorded and analyzed for sequential and spin system assignments. The NMR data were processed with nmrPipe (19) and analyzed with CCPNMR Analysis v1.6 (20) and nmrDraw.

NMR Titrations of Histone Peptides. The $^1H,^{15}N$ HSQC spectra of 0.1-mM uniformly ^{15}N-labeled wild-type or mutated PHD fingers of MLL5, Set3, and UpSET were collected on a Varian INOVA 600 MHz spectrometer. The spectra were recorded at 298K using 1,024 × 160 increments, and a spectral width of 8,804.8 × 1,944.3 Hz in the ^1H and ^{15}N dimensions, respectively. The binding was characterized by monitoring chemical shift changes in $^1H,^{15}N$ HSQC spectra of the PHD finger as differently modified histone tail peptides (synthesized by the University of Colorado Denver Peptide Core Facility) were added stepwise. The dissociation constants (K_ds) were determined by a nonlinear least-squares analysis in KaleidaGraph using the equation

$$\Delta\delta = \Delta\delta_{\max}\left(([L]+[P]+K_d) - \sqrt{([L]+[P]+K_d)^2 - 4[P][L]}\right)\Big/2[P],$$

where [L] is concentration of the peptide, [P] is concentration of the protein, Δδ is the observed chemical shift change, and $\Delta\delta_{max}$ is the normalized chemical shift change at saturation. Normalized (21) chemical shift changes were calculated using the equation $\Delta\delta = \sqrt{(\Delta\delta H)^2 + (\Delta\delta N/5)^2}$, where Δδ is the change in chemical shift in parts per million (ppm).

Fluorescence Spectroscopy. Spectra were recorded at 25 °C on a Fluoromax-3 spectrofluorometer (HORIBA). The samples containing 1 μM wild-type or mutated MLL5 PHD or UpSET PHD and progressively increasing concentrations of the histone peptide were excited at 295 nm. Emission spectra were recorded over a range of wavelengths between 315 and 405 nm with a 0.5-nm step size and a 1-s integration time and averaged over three scans. The K_d values were determined using a nonlinear least-squares analysis and the equation

$$\Delta I = \Delta I_{\max}\left(([L]+[P]+K_d) - \sqrt{([L]+[P]+K_d)^2 - 4[P][L]}\right)\Big/2[P],$$

where [L] is the concentration of the histone peptide, [P] is the concentration of MLL5 PHD, ΔI is the observed change of signal intensity, and ΔI_{max} is the difference in signal intensity of the free and bound states of the PHD finger. The K_d value was averaged over three separate experiments, with error calculated as the SD between the runs.

Proximity Ligation Assays. HEK 293T cells grown on polylysine-coated coverslips were transfected with plasmids expressing FLAG-MLL5 (wild type, W141A, D128K, F125A). Cells were fixed, blocked, and incubated with anti-FLAG M2 antibody (catalog no. F1804; Sigma) and anti-H3K4me3 antibody (catalog no. ab8580; Abcam). After washing, coverslips were incubated with secondary antibodies conjugated with Duolink proximity ligation assay (PLA) probes (catalog nos. DUO92003 and DUO92004; OLINK Bioscience), followed by hybridization, ligation, and amplification using Duolink PLA Detection Reagents (catalog no. DUO92008; OLINK Bioscience) according to manufacturer instructions. Images were acquired using a LSM710 confocal microscope (Carl Zeiss) and analyzed using ZEN 2009 software. For quantification analysis, 40 figures were randomly captured, and the numbers of red spots in nucleus of each cell were counted manually (400–500 cells per sample). Statistical significance was assessed by nonparametric analysis using Mann–Whitney *U* test. Differences were considered statistically significant at $P < 0.05$.

1. Sawado T, et al. (2008) H3 K79 dimethylation marks developmental activation of the beta-globin gene but is reduced upon LCR-mediated high-level transcription. *Blood* 112(2):406–414.
2. Vogel MJ, Peric-Hupkes D, van Steensel B (2007) Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc* 2(6):1467–1478.
3. Tusher VG, Tibshirani R, Chu G (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA* 98(9):5116–5121.
4. Ji H, Wong WH (2005) TileMap: Create chromosomal map of tiling array hybridizations. *Bioinformatics* 21(18):3629–3636.
5. Straub T, Grimaud C, Gilfillan GD, Mitterweger A, Becker PB (2008) The chromosomal high-affinity binding sites for the Drosophila dosage compensation complex. *PLoS Genet* 4(12):e1000302.
6. Deal RB, Henikoff JG, Henikoff S (2010) Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* 328(5982):1161–1164.
7. Blankenberg D, et al. (2010) Galaxy: A web-based genome analysis tool for experimentalists. *Curr Protoc Mol Biol* 89:19.10.1–19.10.21.
8. Giardine B, et al. (2005) Galaxy: A platform for interactive large-scale genome analysis. *Genome Res* 15(10):1451–1455.
9. Goecks J, Nekrutenko A, Taylor J, Galaxy T; Galaxy Team (2010) Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol* 11(8):R86.
10. Sebastian S, et al. (2009) MLL5, a trithorax homolog, indirectly regulates H3K4 methylation, represses cyclin A2 expression, and promotes myogenic differentiation. *Proc Natl Acad Sci USA* 106(12):4719–4724.
11. Rincon-Arano H, Halow J, Delrow JJ, Parkhurst SM, Groudine M (2012) UpSET recruits HDAC complexes and restricts chromatin accessibility and acetylation at promoter regions. *Cell* 151(6):1214–1228.
12. Rosales-Nieves AE, et al. (2006) Coordination of microtubule and microfilament dynamics by *Drosophila* Rho1, Spire and Cappuccino. *Nat Cell Biol* 8(4):367–376.
13. Liu R, et al. (2009) Wash functions downstream of Rho and links linear and branched actin nucleation factors. *Development* 136(16):2849–2860.
14. Bianchi-Frias D, et al. (2004) Hairy transcriptional repression targets and cofactor recruitment in *Drosophila*. *PLoS Biol* 2(7):E178.
15. Rothbart SB, Krajewski K, Strahl BD, Fuchs SM (2012) Peptide microarrays to interrogate the "histone code". *Methods Enzymol* 512:107–135.
16. Wittekind M, Mueller L (1993) HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha-carbon and beta-carbon resonances in proteins. *J Magn Reson B* 101(2):201–205.
17. Grzesiek S, Bax A (1992) Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein. *J Magn Reson* 96(2):432–440.
18. Grzesiek S, Anglister J, Bax A (1993) Correlation of backbone amide and aliphatic side-chain resonances in C-13/N-15-enriched proteins by isotropic mixing of C-13 magnetization. *J Magn Reson* 101(1):114–119.
19. Delaglio F, et al. (1995) NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3):277–293.
20. Vranken WF, et al. (2005) The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins* 59(4):687–696.
21. Grzesiek S, Stahl SJ, Wingfield PT, Bax A (1996) The CD4 determinant for downregulation by HIV-1 Nef directly binds to Nef. Mapping of the Nef binding surface by NMR. *Biochemistry* 35(32):10256–10261.

Fig. S1. MLL5 associates with transcribed genes. (*A*) Meta-analysis of MLL5 binding around transcriptional start sites of two biological replicates. (*B–F*) MLL5 chromatin profiles over specific genes belonging to quintiles Q4–Q1. A track with RNA-seq reads is shown for each example. The mean signal was smoothed + whiskers. Coding regions are shown in black.

Fig. S2. MLL5 binds H3K4me3. (*A*) Pie graphs showing the overlap between bound MLL5 and indicated histone PTMs. See also Table S1. (*B*) MLL5 and MLL5s distribution correlates with H3K4me3. MLL5s distribution does not correlate with H3K9me3. C2C12 cells were stained with MLL5 and MLL5s antibodies (green) as well as with anti-H3K4me3 or anti-H3K9me3 antibodies (red). Blue indicates DNA stained with DAPI. (Scale bar, 10 μm.) (*C*) Superimposed ^{1}H,^{15}N HSQC spectra of the PHD fingers of MLL5 collected upon titration with H3K4me3 (residues 1–12 of H3).

Fig. S3. A unique H3K4me3 binding mechanism of MLL5. Superimposed H3K4me3 binding pockets of PHD fingers of MLL5 (red), MLL1 (green; PDB ID code 3LQJ), bromodomain PHD finger transcription factor (BPTF, blue; PDB ID code 2F6J) and inhibitor of growth protein 2 (ING2, wheat; PDB ID code 2G6Q). H3K4me3 peptide in the MLL5 PHD–H3K4me3 complex is yellow, and H3K4me3 peptides in all other complexes are gray.

PNAS PNAS PNAS

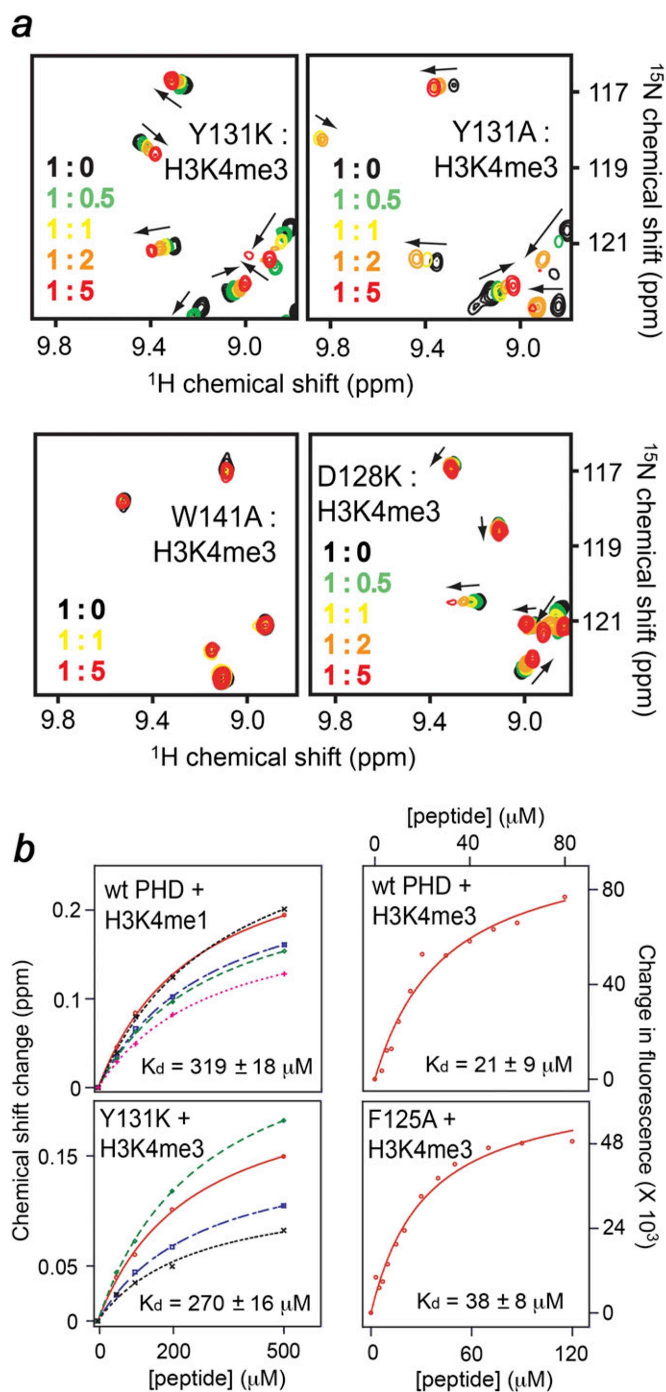

Fig. S4. Mutations in the active site disrupt binding. (*A*) Superimposed $^{1}H,^{15}N$ HSQC spectra of the indicated mutants of MLL5 PHD collected upon titration with H3K4me3 peptide (residues 1–12 of H3). Spectra are color-coded according to the protein:peptide molar ratio (*Lower Left*). (*B*) Representative binding curves used to determine the K_d values by tryptophan fluorescence and NMR.

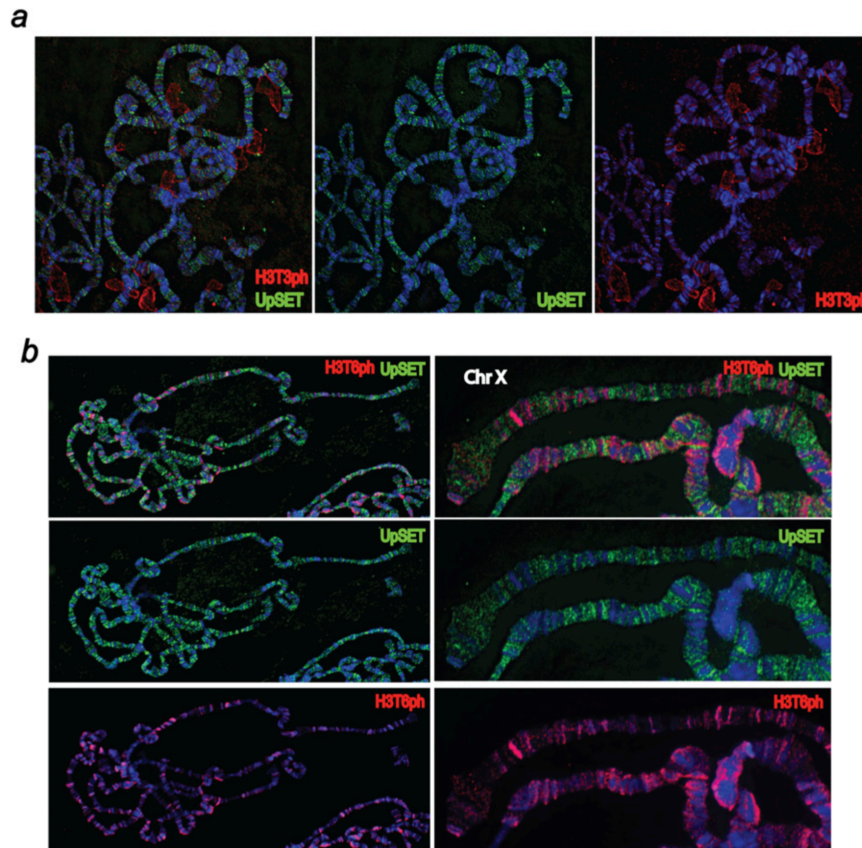

Fig. S5. Chromatin targeting by the *Drosophila* MLL5 ortholog UpSET is regulated through phosphorylation. (*A* and *B*) Display of individual staining of Fig. 5*E*. Polytene chromosomes were stained with a mix of antibodies against UpSET, N-, and C-terminus (green), and H3T3ph or H3T6ph (red). DNA was counterstained with DAPI (blue). A higher magnification view of the X chromosome lacking colocalization of UpSET and H3T6ph (*B*, *Right*).

Table S1. MLL5 overlaps with transcription-associated histone modification

	All peaks*	TSS-associated peaks†	MLL5 bound	MLL5 unbound	Bound, %	Unbound, %
H3K4me1	330,000	13,139	7,169	5,970	54.56	45.44
H3M4me2	72,244	14,114	8,974	5,140	63.58	36.42
H3K4me3	75,425	14,976	9,441	5,535	63.04	36.96
H3K27me3	210,000	4,042	1,671	2,371	41.34	58.66
H3K9Ac	64,971	12,265	8,105	4,160	66.08	33.92
RNA Pol II	46,759	10,834	7,195	3,639	66.41	33.59

This analysis does not include CpG islands not associated with protein-coding gene promoters.
*ChIP-seq peaks generated in Asp et al. (1).
†Histone modification peaks associated with TSSs within −1,500 bp to 500 bp flanking the +1 position of ∼32,000 annotated mouse genes.

1. Asp P, et al. (2011) Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc Natl Acad Sci USA* 108:E149–E158.

Table S2. Data collection and refinement statistics

Data collection and Refinment	MLL5 PHD–H3K4me3
Data collection	
Wavelength (Å)	1.000
Space group	$P2_12_12_1$
Resolution, Å	33.65–1.48
Cell dimensions, Å	a = 31.29, b = 43.64, c = 52.86, α = β = γ = 90°
No. of measured reflections	85,330
No. of unique reflections	12,590
Completeness (%)	97.0 (67.5)
Anomalous completeness (%)	96.7 (64.6)
Redundancy	6.9 (4.9)
Anomalous redundancy	3.7 (2.7)
I/σ (I)	28.4 (6.9)
R_{merge} (%)	6.1 (18.5)
Refinement	
Resolution (Å)	33.65–1.48 (1.55–1.48)
No. of reflections	22,776
No. of nonanomalous reflections	12,338
R_{factor} (%)	12.5
R_{free} (%)	14.0
No. of protein atoms	1,156
No. of heterogen atoms	127
No. of water molecules	125
Rmsd from ideal values	
Bond lengths, Å	0.004
Bond angles, °	1.054
Average B-values, Å2	
Protein chain A	15.8
Peptide chain U	22.4
Water	29.3
Zinc	8.7
Ramachandran plot analysis	
Residues in most favored regions	87.3%
Residues in additional allowed regions	12.7%
Residues in generously allowed regions	0%
Residues in disallowed regions	0%

Values in parentheses refer to data in the highest-resolution shell.

Other Supporting Information Files

Dataset S1 (XLSX)