MOLECULAR AND CELLULAR BIOLOGY, Nov. 2011, p. 4310–4318 Vol. 31, No. 21
0270-7306/11/$12.00 doi:10.1128/MCB.06092-11

The COMPASS Family of H3K4 Methylases in *Drosophila*▽

Man Mohan,[1] Hans-Martin Herz,[1] Edwin R. Smith,[1] Ying Zhang,[1] Jessica Jackson,[3] Michael P. Washburn,[1,2] Laurence Florens,[1] Joel C. Eissenberg,[3] and Ali Shilatifard[1]*

Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, Missouri 64110[1]; Department of Pathology and Laboratory Medicine, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160[2]; and Edward A. Doisy Department of Biochemistry and Molecular Biology, Saint Louis University School of Medicine, 1100 South Grand Boulevard, St. Louis, Missouri 63104[3]

Received 9 August 2011/Returned for modification 11 August 2011/Accepted 22 August 2011

Methylation of histone H3 lysine 4 (H3K4) in *Saccharomyces cerevisiae* is implemented by Set1/COMPASS, which was originally purified based on the similarity of yeast Set1 to human MLL1 and *Drosophila melanogaster* Trithorax (Trx). While humans have six COMPASS family members, *Drosophila* possesses a representative of the three subclasses within COMPASS-like complexes: dSet1 (human SET1A/SET1B), Trx (human MLL1/2), and Trr (human MLL3/4). Here, we report the biochemical purification and molecular characterization of the *Drosophila* COMPASS family. We observed a one-to-one similarity in subunit composition with their mammalian counterparts, with the exception of LPT (lost plant homeodomains [PHDs] of Trr), which copurifies with the Trr complex. LPT is a previously uncharacterized protein that is homologous to the multiple PHD fingers found in the N-terminal regions of mammalian MLL3/4 but not *Drosophila* Trr, indicating that *Trr* and *LPT* constitute a split gene of an MLL3/4 ancestor. Our study demonstrates that all three complexes in *Drosophila* are H3K4 methyltransferases; however, dSet1/COMPASS is the major monoubiquitination-dependent H3K4 di- and trimethylase in *Drosophila*. Taken together, this study provides a springboard for the functional dissection of the COMPASS family members and their role in the regulation of histone H3K4 methylation throughout development in *Drosophila*.

Histone H3 lysine 4 methylation (H3K4me) is associated with the transcriptionally active regions of the genome in yeast, flies, and mammals (3, 23, 35). Set1 was identified as a component of a macromolecular protein complex named COMPASS (complex of proteins associated with Set 1), as the first H3K4 methylase, and it is responsible for all mono-, di-, and trimethylation of H3K4 in yeast (22, 31, 40, 52). In *Drosophila melanogaster*, four SET domain-containing proteins, namely, Trithorax (Trx), Trithorax-related (Trr), dSet1, and Ash1, have been reported to implement H3K4 methylation (10). All but Ash1, which has subsequently been demonstrated to be an H3K36 methyltransferase (49, 59), are related to subunits of the six COMPASS and COMPASS-like complexes in mammals. *trx* was originally characterized as a gene that when mutated caused homeotic transformations (6, 18). Detailed genetic and molecular analyses showed that Trx is required to maintain activation states of its target genes throughout development and counteracts the repressive effects of the Polycomb group proteins (PcG) (39, 41). Trr was identified based on sequence similarity to Trx but was shown to function in the regulation of hormone-responsive gene expression (42). dSet1 was identified based on sequence homology to the *Saccharomyces cerevisiae* and mammalian Set1 proteins (53, 58).

In mammals, there are at least six SET1-related proteins that form COMPASS-like complexes, namely, SET1A, SET1B, and MLL1 to MLL4. SET1A and SET1B are orthologous to dSet1; MLL1 and MLL2 are orthologous to *Drosophila* Trx; MLL3 and MLL4 (also known as ALR) are orthologous to *Drosophila* Trr (33, 43, 45). All of the mammalian COMPASS family of H3K4 methylases share ASH2L, RBBP5, DPY30, and WDR5 as common components. Analysis of the mammalian complexes allows classification into three classes based on unique components within each class: COMPASS, represented by SET1A and SET1B, contains WDR82 and CXXC1 (27, 53), proteins implicated in regulating trimethylation by yeast COMPASS; the MLL1/2 complexes contain Menin, implicated in targeting MLL1 to the Hox genes (55, 56); the MLL3/4 complexes contain PTIP, PA-1, and NCOA6 (8), which are important for the gene-specific targeting of these complexes, and UTX, a histone H3K27 demethylase thought to be involved in counteracting PcG-mediated gene silencing (10, 17, 29).

In this study, we purified and characterized the dSet1, Trx, and Trr complexes. In contrast to a previous report that Trx formed a heterotrimeric complex with CBP and SBF1 (38), we found instead that Trx forms a COMPASS-like complex containing orthologs of all known components of the MLL1 complex in mammals. Our studies also demonstrate that *Drosophila* Set1 is the major contributor to the bulk *in vivo* dimethylation and trimethylation of H3K4 and that this depends on a conserved form of histone cross talk, where monoubiquitinated H2B is required for H3K4 trimethylation by dSet1. We also find that mammalian MLL3/4 are represented in flies by two genes, *Trr* and *LPT*, and that the encoded proteins exist together in a COMPASS-like Trr complex. Taken together, our evidence for the existence of one representative complex in *Drosophila* for each of the three classes of the six COMPASS family proteins in mammals provides a

* Corresponding author. Mailing address: Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110. Phone: (816) 926-4465. Fax: (816) 926-2080. E-mail: ASH@Stowers.org.
▽ Published ahead of print on 29 August 2011.

unique opportunity to discover the differences in the targeting and function of H3K4 methylation by these complexes.

MATERIALS AND METHODS

Antibodies, plasmids, cell lines, and transgenics. Antibodies recognizing the FLAG tag were obtained from Sigma, H3, H3K4me1, H3K4me2, and H3K4me3 were obtained from Abcam. The phosphorylated polymerase II (Pol II) C-terminal domain rat monoclonal antibodies were obtained from Millipore. Antibodies to dSet1 (amino acids [aa] 1308 to 1641), Trx (aa 2785 to 3131), Trr (aa 541 to 827), and LPT (aa 85 to 242) were expressed as fusion proteins in *Escherichia coli* and sent to Pocono Rabbit Farm and Laboratories for antibody production in rabbits. The dUTX antibody has been described previously (14). Polytene chromosome squashes were essentially performed as described in reference 46. Clonal S2 cells expressing Ash2 were generated by cotransfecting pRmHa3-FLAG-HA-Ash2 (full length) (48) and pCoBlast (Invitrogen), followed by selection in soft agar (54). Upstream activating sequence (UAS)-dUTX-hemagglutinin (HA)-FLAG transgenic flies were generated by standard procedures (BestGene, Chino Hills, CA).

Vector pBacEGFP/MT-FLAG-HA was constructed by blunt cloning a PCR fragment containing the *Metallothionein* (MTN) promoter, multiple-cloning site, and poly(A) tail from the pMT-FLAG-HA vector with primers 5′-CGTTGCA GGACAGGATGTGGTGCCCGAT and 5′-CAAAAGGTTTTGCCATTGTTC TTCTTGGAGCCGTTGTT at a unique EcoRI site in pBac-Polyhedrin-EGFP. cDNAs of CG17293 (dWdr82), CG5585 (dRbbp5), and CG11750 (dPA1) were cloned in frame to FLAG-HA downstream of the *MTN* promoter. These constructs were cotransfected with Bsu36I-digested viral DNA with the BacPAK8 kit from Clontech (Mountain View, CA).

FLAG purification and MudPIT analysis. Nuclear extracts were prepared essentially as described previously for mammalian cells (32). Briefly, affinity purifications were carried out using nuclear extracts prepared from 3 L of cells grown in SFX medium in Spinner flasks. Cells were induced with 200 μM $CuSO_4$ for 2 days at a density of 2×10^6 cells/ml. Nuclear extracts were incubated with FLAG-coated beads overnight and washed 3 times with buffer containing 300 mM NaCl. Complexes were eluted with FLAG peptide and precipitated with trichloroacetic acid. Precipitated protein mixtures were digested with endoproteinase Lys-C and trypsin (Roche) and subjected to multidimensional protein identification technology (MudPIT) analysis as previously described (53).

Immunoprecipitations. Volumes of 1.5 ml of nuclear extracts were incubated with 5 μl serum or 5 μg of rabbit IgG for 3 h and washed 3 times for 5 min with 5 ml of 300 mM NaCl-containing buffer. One-sixth of the immunoprecipitation (IP) mixture was used per blot assay.

***In vitro* methyltransferase assay.** The methyltransferase activity of purified complexes was tested toward recombinant histone H3 as described previously (28). FLAG eluates (5 μl) were incubated overnight with recombinant histone H3 (1 μg) in histone methyltransferase (HMTase) reaction buffer with cold *S*-adenosyl methionine (SAM). Methylation of histone H3 was tested by the application of the reaction mixture to SDS-PAGE and Western analysis with antibodies from Abcam (catalog numbers Ab1791 and Ab8895) to histone H3 and methylated H3K4.

Fly lines. The *Drosophila* UAS-*dSet1*-RNAi (VDRC 40682), UAS-*trx*-RNAi (VDRC 108122), UAS-*trr*-RNAi (Bloomington 29563), UAS-*LPT*-RNAi (Bloomington 25994), UAS-*dBre1*-RNAi (32), and UAS-*dWdr82*-RNAi (VDRC 25246) lines were used.

RESULTS

dSet1 is responsible for the bulk of H3K4 trimethylation and dimethylation in flies. To understand the contribution of all the known H3K4 methyltransferases, dSet1, Trx, and Trr, in flies, we performed UAS-*GAL4*-mediated RNA interference (RNAi) knockdown of each enzyme in the posterior compartment of third-instar wing imaginal discs. In these discs, the posterior compartment is marked by green fluorescent protein (GFP), and the anterior compartment with no knockdown serves as an internal control. We generated antibodies specific to dSet1, Trx, and Trr and tested their specificities on wing imaginal discs. Significant loss of signal in the posterior half demonstrated the specificity of these antibodies (Fig. 1a, a′, d, d′, i, and i′; compare the posterior GFP compartment with the anterior compartment). We then stained the wing imaginal discs with antibodies specific to H3K4 di- and trimethylation and observed that dSet1 knockdown led to near total loss of both of the aforementioned marks (compare H3K4me2 and H3K4me3 signals in the anterior versus the posterior domain). However, Trx and Trr knockdown had only marginal effects on the levels of either of these marks.

To gain insight into the function of dSet1, we performed polytene chromosome immunolocalizations with dSet1-specific antibodies and observed that, similar to a recently published report (1), dSet1 is widely distributed throughout euchromatin (Fig. 2A, panel a), and in some cases, low levels of staining were evident at puffs. We also observed a significant colocalization of dSet1 with the Ser5 and Set2 phosphorylated forms of RNA Pol II (5) in polytene chromosome squashes (Fig. 2C and D), supporting the fact that dSet1 is an H3K4 methyltransferase, a mark associated with active transcription. We also performed polytene chromosome staining with our Trx antibody (Trx-CT, specific for the C-terminal half of the protein) (Fig. 2A, panel c) and observed a number of strongly stained loci, which is similar to previously published reports of Trx localization (7). We also observed that a large number of weaker but significant signals of Trx localized to interbands. This was also the case with our Trr-specific antibody (Fig. 2A, panel b), an observation similar to previously published reports (42). The broadly dispersed distributions of dSet1, Trx, and Trr proteins genome-wide suggest that these proteins have the potential to regulate a large number of genes. Detailed comparisons of the distal X chromosome (Fig. 2B) indicated that these proteins can overlap but also show distinct distribution patterns.

Affinity purification of the COMPASS family from *Drosophila*. A previous report suggested that Trx is found in a heterotrimeric complex with CBP and SBF1 (38). This complex of Trx is different than the yeast Set1/COMPASS and human MLL/COMPASS-like complex in that it lacks subunits of the COMPASS family shown to be critical for H3K4 methylase function within the complex (9, 17, 22, 53). To understand the composition of each of the known H3K4 methyltransferases, dSet1, Trx, and Trr of *Drosophila*, and their relative contributions to H3K4 methylation in flies, we set out to biochemically isolate their complexes. We used two methods to make FLAG-HA-tagged proteins in *Drosophila* S2 cells. The first is a widely used method of stable transfection in cells, expressing the protein under the $CuSO_4$-inducible *Metallothionein* (*Mtn*) promoter (20). The second method uses a baculovirus superinfection system (24). Baculovirus, which infects and replicates in lepidopteran cells, can also infect *Drosophila* S2 cells but is unable to replicate (24). We modified the strategy presented by Lee et al. (24) and cloned FLAG-HA versions of dWdr82, dPA1, and dRbbp5 under the regulation of the *Mtn* promoter. We generated complete viruses harboring these constructs and infected S2 cells, followed by induction with $CuSO_4$. Immunofluorescence and western analyses with the HA antibody demonstrated that we could infect and express the desired proteins (Fig. 3A and B). For either expression strategy, we isolated complexes by FLAG affinity purification from nuclear extracts from induced cells. Immunopre-

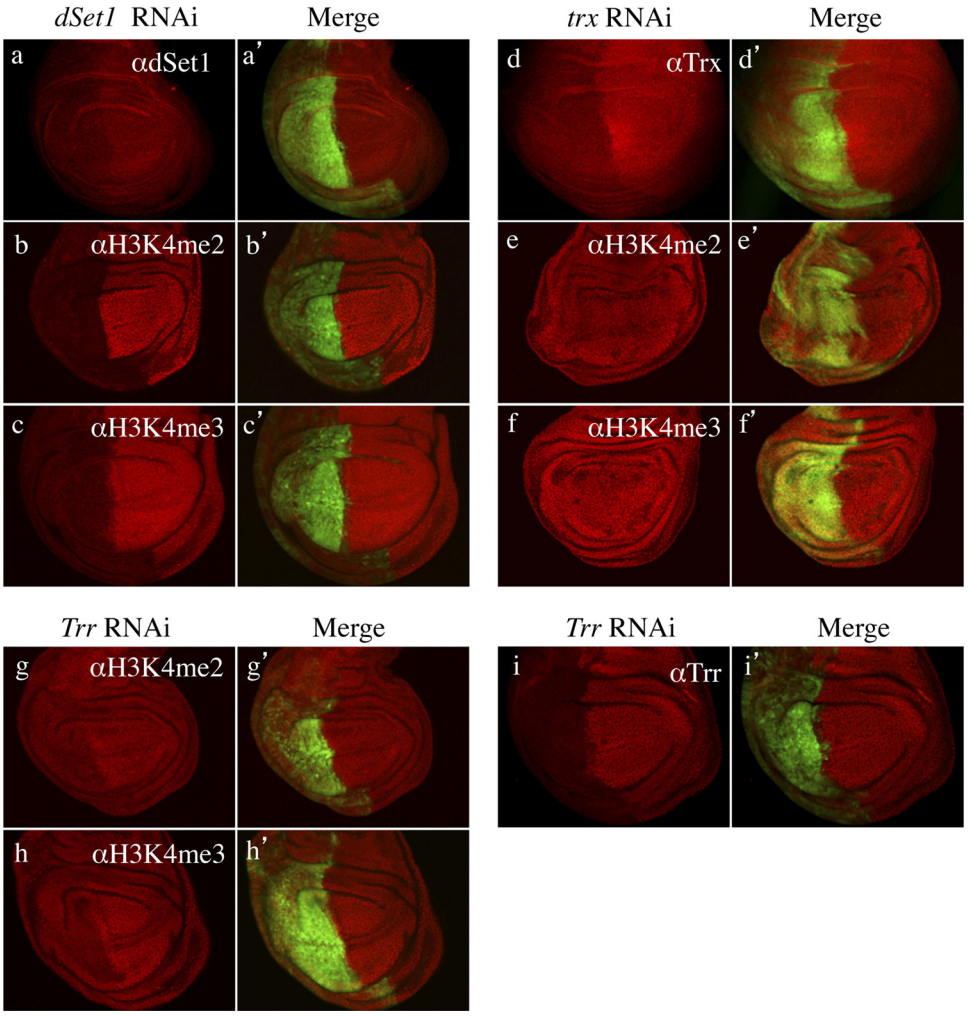

FIG. 1. dSet1 is the major H3K4 methyltransferase, while Trx and Trr are minor contributors of global H3K4 methylation. Wing imaginal discs from *Drosophila* third-instar larvae were stained with antibodies specific to H3K4 methyltransferases. With the UAS-GAL4 system, *enGAL4* was used to drive the expression of UAS-hairpin constructs targeting dSet1, Trx, and Trr; the domain of knockdown was marked by GFP. Loss of signal in the posterior half demonstrates the knockdown of dSet1 (a and a′), Trx (d and d′), and Trr (i and i′) and the specificity of the antibodies to their respective proteins. dSet1 knockdown results in near total loss of both H3K4me2 (b and b′) and H3K4me3 (c and c′) from the discs, whereas Trx (e and f′) and Trr (g and h′) knockdown does not result in any significant reduction of either H3K4me2 or H3K4me3.

cipitated complexes were subjected to protein identification using MudPIT.

We first chose to probe the interaction partners of Ash2, as its homolog in mammals, Ash2L, is one of the common components of all H3K4 methylase complexes (40, 47, 53). We reasoned that purification of a common component would give us an overall understanding of the conservation of the *Drosophila* complexes. We found that our FLAG-Ash2 preparations contained significant amounts of dSet1 (CG40351 [59]), Trx, and Trr, suggesting that all three proteins could associate in COMPASS-like complexes (Fig. 4). In support of these findings, we also found significant enrichment of Wds (will die slowly [16]), the *Drosophila* ortholog of WDR5, a shared component of all the H3K4 methylase complexes (47). In mammals, Menin is unique to the MLL1/2 complexes, while PTIP, PA-1, UTX, and NCOA6 are unique to the MLL3/4 complexes (8, 17, 19, 37). Their *Drosophila* orthologs, Menin 1 (13), PTIP (11), and dUTX (14, 46), were also identified in the Ash2 preparations. We also found high enrichment for previously uncharacterized *Drosophila* proteins transcribed by the CG5585, CG17293, CG6444, and CG17446 genes. Sequence searches and alignments of these proteins with human and yeast proteins revealed that CG17293 is very closely related to Wdr82, whereas CG5585, CG6444, CG17446, and CG14023 are homologous to RBBP5 (15), DPY30, CXXC1, and NCOA6, respectively (Fig. 5A). RBBP5 and DPY30 are common components of all the major H3K4 methylase complexes, and CXXC1 and WDR82 are unique components of the SET1A/B (hCOMPASS) complexes (10, 25, 26, 53).

The presence of three different SET domain-containing proteins in Ash2 purifications was consistent with the presence of three distinct HMT complexes with unique and shared subunits, similar to the six COMPASS-like complexes in mammals. To further define the composition of these complexes, we chose to tag and purify at least one unique component of each of the dSet1 (dWdr82) and Trr (dPA1 and dUTX) complexes.

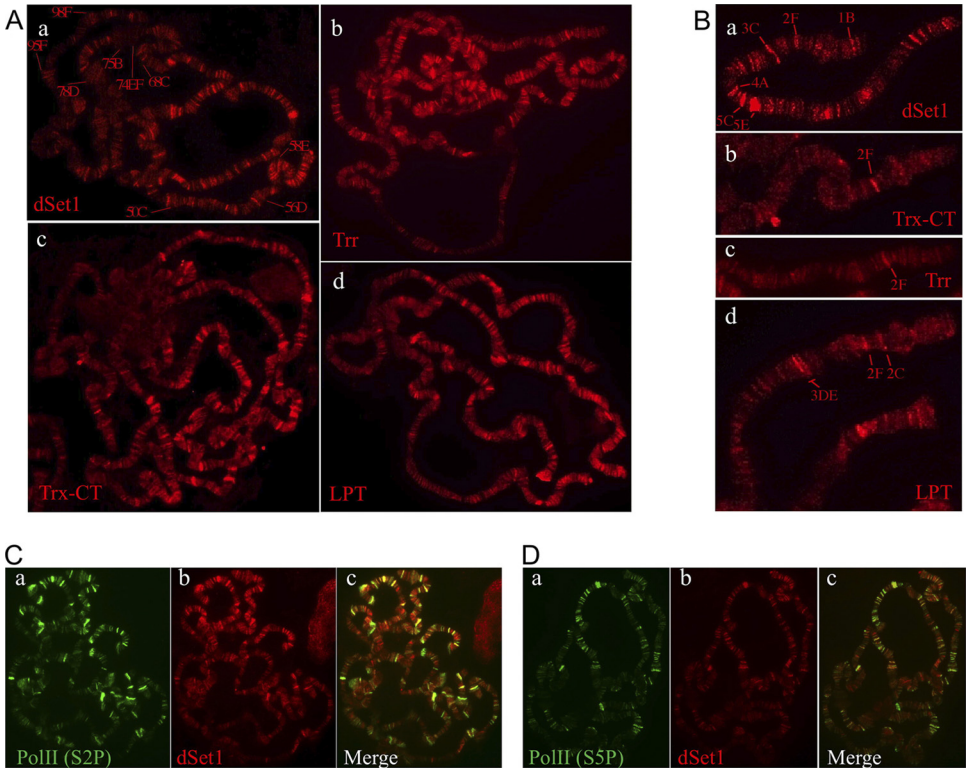

FIG. 2. Distributions of dSet1, Trx, Trr, and LPT on polytene chromosomes. (A) Immunostaining of dSet1, Trx, Trr, and LPT on polytene chromosome squashes from third-instar larval salivary glands. Labels in panel a denote the cytological positions of landmark puffs. (B) Details of the distal X chromosome stained for of dSet1, Trx, Trr, and LPT. Labels denote the cytological positions of the indicated stained sites. Trx-CT, antibody raised against the C-terminal domain of Trithorax protein. (C and D) Colocalization of dSet1 (red) with RNA Pol II (S2P; green) (C) and with RNA Pol II (S5P; green) (D). The significant amounts of yellow signals in the merge channels (C, panel c) and (D, panel c) indicate colocalization of dSet1 with active regions of the genome.

We were unable to detect expression of FLAG-HA-Menin 1, but Trx-specific components could be deduced by comparing Ash2-associated proteins that were not found in dSet1 or Trr complexes (Fig. 4A).

MudPIT analysis of dWdr82-FLAG eluate showed the presence of all the homologous components of yeast COMPASS and human COMPASS, namely, dSet1, dCXXC1, dDpy30, dRbbp5, Ash2, and Wds (Fig. 4). The dPA1-FLAG eluate was enriched for Trr, Ash2, dRbbp5, dDpy30, Wds, dPTIP, dNCoA6, and dUTX, suggesting the conservation between the *MLL3*/4 and Trr complexes. As dUTX is an H3K27 demethylase, its presence in an H3K4 methylation complex suggests that the complex first derepresses the genes by removing H3K27 methylation and then methylates at H3K4, a histone mark associated with active transcription. To further support our observation that dUTX purifies with Trr, we generated UAS-*dUTX-HA-FLAG* flies and FLAG-immunopurified complexes from embryos ubiquitously expressing dUTX-HA-FLAG under the control of *Act5C-Gal4*. We found significant enrichment of Trr, Ash2, dRbbp5, dDpy30, Wds, PTIP, and NCoA6, confirming that dUTX is a component of the Trr complex. Altogether, we found evidence for representatives of all three classes of mammalian COMPASS-like complexes (Fig. 4A). Importantly, all three complexes, like their mammalian counterparts, were capable of implementing methylation of H3K4 (Fig. 4B).

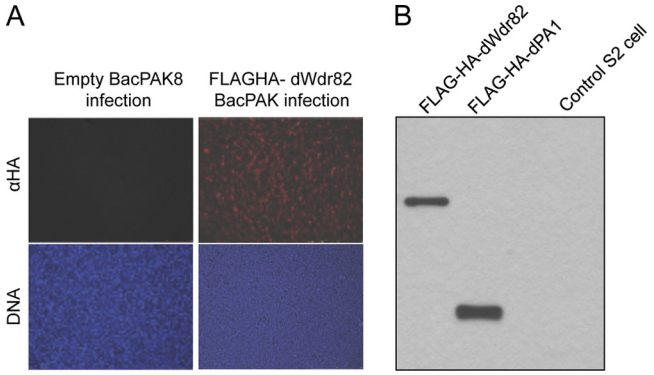

FIG. 3. Baculovirus superinfection for the expression of epitope-tagged proteins in *Drosophila* S2 cells. (A) Immunohistochemistry against the HA epitope in S2 cells infected with either pBacEGFP/MT-FLAG-HA (empty BacPAK8) or pBacEGFP/MT-FLAG-HA-dWdr82, showing the expression of tagged protein. (B) Western blot analysis with HA antibody demonstrated the expression of dWdr82 and dPA1 in S2 cells infected with baculovirus (pBacEGFP-MT-FLAG-HA-dWdr82 and pBacEGFP-MT-FLAG-HA-dPA1).

MudPIT analyses of the FLAG eluates of Ash2 (a shared component of dCOMPASS, Trx, and Trr complexes) and dPA1 and dUTX (unique components of the Trr complex) repeatedly revealed large amounts of another protein, CG5591. A protein BLAST search revealed a significant sim-

A

	Yeast SET1/ COMPASS	Human SET1/ COMPASS	Drosophila SET1/ COMPASS	Human MLL1/2	Human MLL3/4	Drosophila Trr		Shared components
	Ref. 22, 31, 41, 53	Ref. 27, 54	FLAG-HA:dWdr82	Ref. 17, 56-58	Ref. 8, 29	FLAG-HA: CG11750 (dPA1)	FLAG-HA-dUTX	FLAG-HA-Ash2
CG17293 dWdr82			1208/ 0.234			0	8/ 0.00046	33/ 0.0032
CG40351 dSet1			94/ 0.0035			0	9/ 0.00010	121/ 0.0023
CG17446 dCXXC1			10/ 0.00092			0	0	19/ 0.00089
CG6677 Ash2			7/ 0.00077			5/ 0.00037	377/ 0.012	2048/ 0.114
CG5585 Rbbp5			7/ 0.00088			0	241/ 0.009	110/ 0.007
CG6444 Dpy30			2/ 0.00091			0	72/ 0.0099	311/ 0.072
CG17437 Wds			24/ 0.0041			9/ 0.001	207/ 0.0105	34/ 0.0029
CG1710 Hcf1			62/ 0.0025			17/ 0.00047	46/ 0.00056	119/ 0.0025
CG13778 Menin1			1/ 0.00008			0	0	3/ 0.00012
CG8651Trithorax /MLL 1/2			0			0	0	8/ 0.000067
CG3848Trithorax related/MLL3/4			0			28/ 0.00048	430/ 0.0033	74/ 0.00096
CG32133 PTIP			0			43/ 0.00078	474/ 0.0038	24/ 0.00033
CG11750 dPTIP Assoc. 1			0			90/ 0.031	0	5/ 0.0013
CG5640 dUTX			0			22/ 0.0008	1664/ 0.027	38/ 0.0010
CG5591 LPT			0			10/ 0.00028	143/ 0.0018	29/ 0.00061
CG14023 dNCOA6			0			9/ 0.00015	108/ 0.00081	17/ 0.00022

B

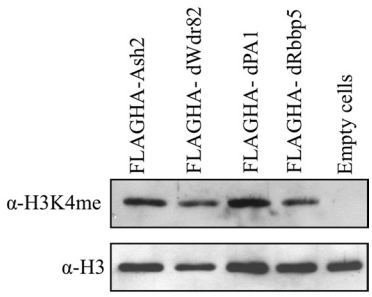

FIG. 4. Composition of COMPASS and COMPASS-like complexes in *Drosophila.* (A) Proteins purified and identified by MudPIT with FLAG-HA-dWdr82, Ash2, dPA1, and dUTX are shown in the table. For Ash2, a clonal S2 cell line expressing FLAG-HA-Ash2 was used. For dWdr82 and dPA1, baculovirus infection systems were used (see Fig. 3). Ubiquitously driven (*Act5C-Gal4*) UAS-*dUTX-HA-2×FLAG* fly embryos were used to purify dUTX complexes, and *yw* embryos were used as a control. Each box contains the number of peptides identified followed by the normalized spectral abundance factor (dNSAF) (60). Red boxes indicate components identified in yeast and mammalian complexes; blue boxes indicate subunits of *Drosophila* complexes identified in this study; green shows components expected due to homology. (B) HMTase assays with Ash2, dWdr82, dPA-1 (CG11750), dRbbp5 (CG5585), and wild-type S2 FLAG eluates on free recombinant histones were performed overnight in the presence of SAM and analyzed by Western blotting using antibodies specific to H3K4me1. The H3K4me1 signal demonstrated the histone methyltransferase activity of purified complexes. Total H3 levels were assayed as a loading control.

ilarity between CG5591 and the N terminus of MLL3 and MLL4 (ALR) (Fig. 5B). A subsequent protein domain search revealed the presence of 7 plant homeodomain (PHD) fingers and an HMG box, arranged in the same manner as the PHDs and HMG box found on MLL3/4 of mammals (Fig. 5B). PHD fingers, represented by the $Cys_4HisCys_3$ zinc finger motif, are involved in protein-protein interactions, some of which bind to H3K4-methylated histones (12). In contrast, Trr shares only a single PHD finger, FYRN/FYRC, and SET/Post-Set domains with its mammalian homologs MLL3 and MLL4. This suggests that in *Drosophila*, CG5591 and Trr together represent the ortholog of mammalian MLL3/4. BLAST searches and sequence analyses indicated that only the orders of *Diptera* and *Hymenoptera* have MLL3/4 N and C termini expressed from distinct genes, consistent with the split having occurred just once in a common ancestor (Fig. 5C). Therefore, we have renamed CG5591 as LPT, for lost PHDs of Trr. We generated an antibody specific to LPT (Fig. 5D) and performed polytene chromosome squashes from third-instar larval salivary glands. We observed an extensive distribution of LPT similar to the Trr distribution on chromosomes (Fig. 2A, panel d, and B, panel d). Since all of the available antibodies against Trr, dUTX, and LPT were generated in rabbits, we were unable to perform colocalization studies. However, immunoprecipitation of S2 cell extracts with LPT antibody enriched for Trr and dUTX (Fig. 5E). Our reproducible finding of LPT with shared components of the COMPASS family in *Drosophila*, and with unique components of the Trr complex, suggests that LPT is an integral component of the functional Trr/COMPASS-like complex.

Having established the one-to-one relationship between the H3K4 complexes of flies and mammals, we wanted to know whether their functional mechanism is conserved. Therefore, we decided to further our understanding by probing the conservation of the histone cross talk mechanism used to target H3K4 trimethylation by dSet1. In yeast and in humans, it has been shown that yBre1/human Bre1(RNF20/40)-mediated monoubiquitination of H2B is required for the generation of H3K4 trimethylation (51, 53), and to a lesser extent H3K4 dimethylation by SET1. To test this histone cross talk, we performed knockdown of Bre1 using UAS-*GAL4*-mediated RNAi in the posterior half of wing imaginal discs and found a

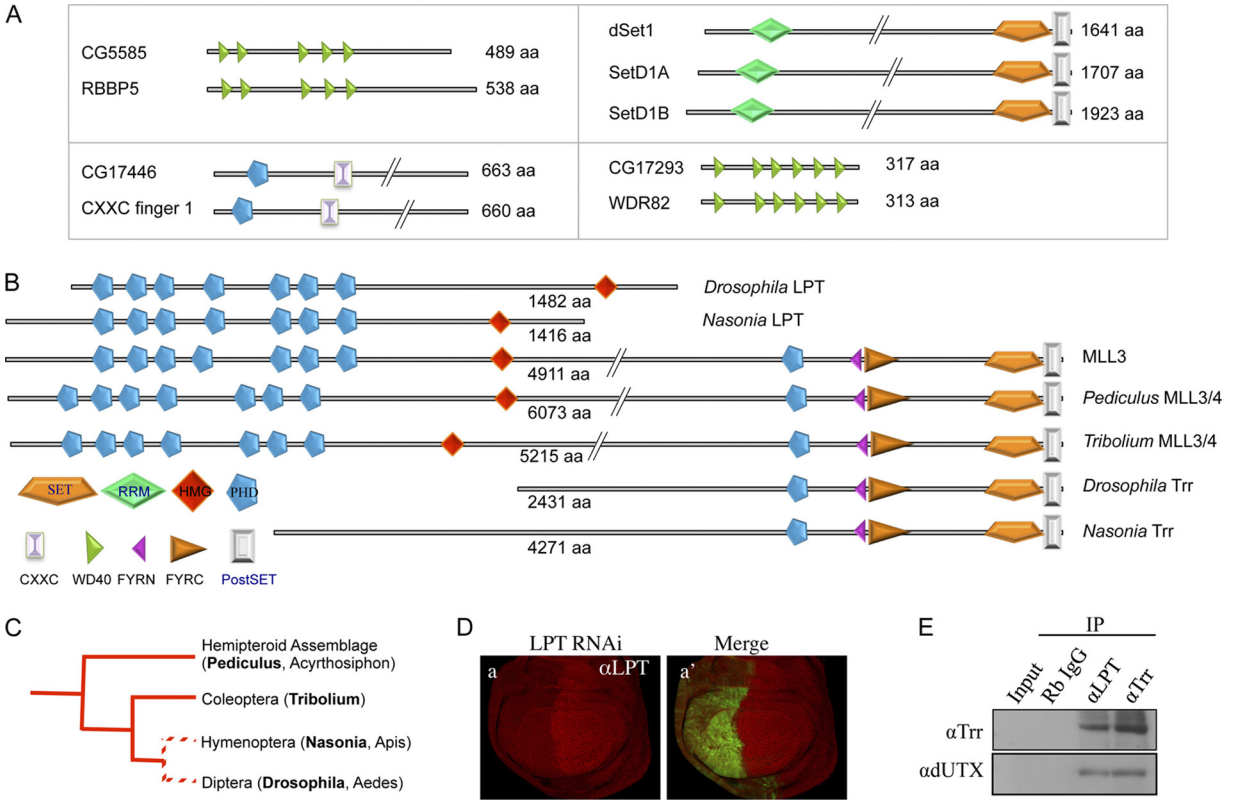

FIG. 5. Ortholog analysis of COMPASS and COMPASS-like subunits between humans and *Drosophila.* (A) Proteins identified with purifications of shared and unique components of dCOMPASS and COMPASS-like complexes were domain aligned with human orthologs using SMART (http://smart.embl-heidelberg.de). WD40 repeat-containing proteins, such as dWdr82, and dRbbp5, were found to have similar domain organization with their human counterparts. (B) LPT is the missing link between MLL3/MLL4 and Trithorax-related (Trr) in *Drosophila*. The schematic alignment indicates that two proteins in *Drosophila*, LPT and Trr, likely resulted from a split into two genes of an ancestral *MLL3/4* gene. Shown are LPT from *Drosophila melanogaster* (GI:24762433) and parasitic wasp (*Nasonia vitripennis*; GI:156551804), *MLL3*-related proteins from human, body louse (*Pediculus humanus*; GI:242016925), flour beetle (*Tribolium castaneum*; GI:270001730), and Trithorax-related protein (Trr) from *Drosophila* (GI:24639197) and *Nasonia* (GI:156551806). LPT together with the Trithorax-related proteins contain the overall domain architecture of MLL3/4 proteins from other insects and other animals, including vertebrates. (C) Schematic view of the phylogenetic relationship (30) between the insect orders represented in panel B. Bold text highlights the genus name for the protein represented in panel B, with a second representative of the order listed if we were able to determine that it also had a split or intact *MLL3/4* gene. (D) RNAi-mediated knockdown of LPT using *enGal4* resulted in significant loss of signal in the posterior half of the wing disc, demonstrating the knockdown of LPT and the specificity of the antibody. (E) Whole-cell lysates from S2 cells were immunoprecipitated with LPT or Trr antibody or rabbit IgG, and the blots were probed with either Trr or dUTX antibody. Input was 1% of the immunoprecipitated nuclear extracts.

strong reduction in H3K4 trimethylation and a significant loss of dimethylation (Fig. 6A, panels a and b).

In yeast and in humans, CPS35/WDR82 has been shown to interact with chromatin in an H2B ubiquitination-dependent manner (28, 53), and it was proposed that the interaction of SET1 with CPS35/WDR82 on chromatin could result in the assembly of trimethylation-competent COMPASS. To assess this phenomenon in flies, we conducted *UAS-GAL4*-mediated RNAi against dWdr82, and we observed a significant reduction only in the levels of H3K4 trimethylation, but not of dimethylation (Fig. 6B, panels c and d), indicating that the regulation of methylation states implemented by COMPASS in yeast, flies, and mammals is functionally conserved.

DISCUSSION

Modifications of histones and the protein machinery for the generation and removal of such modifications are highly conserved and are associated with processes such as transcription, replication, recombination, repair, and RNA processing (21, 43). Histone H3K4 methylation, particularly trimethylation, has been mapped to transcription start sites in all eukaryotes tested (4) and is generally believed to be a hallmark of active transcription. The H3K4 methylation machinery was first identified in yeast and named Set1/COMPASS (31). More recently, six H3K4 methyltransferase complexes have been identified in humans, including SET1A/B, which are subunits of human COMPASS, and MLL1 to MLL4, which are found in COMPASS-like complexes (17, 33, 57).

Although Trx and Trr were identified quite some time ago, their relative contributions to different states of overall H3K4 methylation were not known. Studies of human cells and *Drosophila* cells showed that SET1 is the major contributor of H3K4 trimethylation levels in cells (53, 58). During the preparation of the manuscript, a study of *Drosophila* also showed that dSet1, as a part of COMPASS, is responsible for the

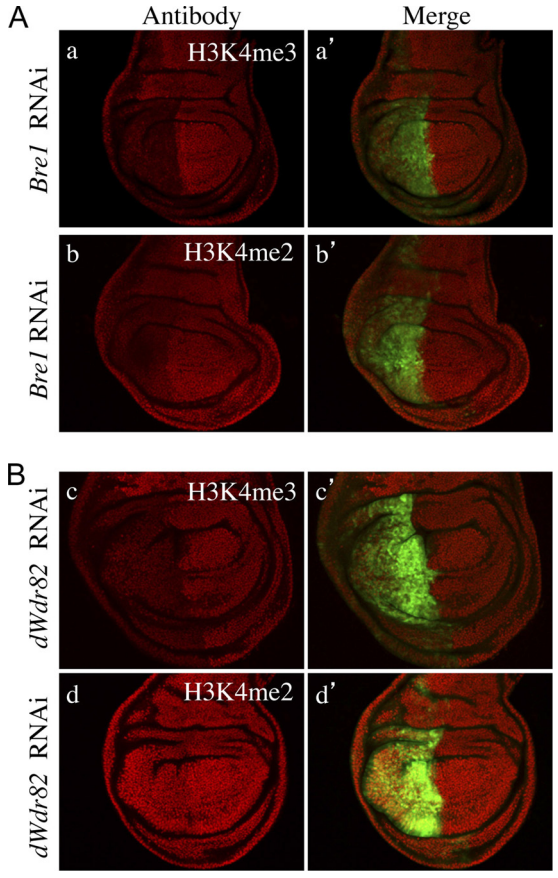

FIG. 6. Requirement of monoubiquitinated H2B by COMPASS for H3K4 trimethylation is conserved in flies. RNAi-mediated knockdown of Bre1, the homolog of yeast BRE1 and human RNF20/40 in wing imaginal discs, led to significantly reduced levels of H3K4 trimethylation (a and a′) and H3K4 dimethylation (b and b′). RNAi-mediated knockdown of dWdr82 in wing imaginal discs led to a similar reduction in H3K4 trimethylation (c and c′), but not H3K4 dimethylation (d and d′), demonstrating the mechanistic conservation between yeast, *Drosophila*, and mammalian SET1 complexes in generating H3K4 trimethylation.

majority of H3K4 di- and trimethylation (1), which is in line with our findings presented here. These findings suggest that dSet1 could be responsible for the deposition of H3K4 trimethylation at the transcription start sites of the most actively transcribed genes as a consequence of postinitiation recruitment via the PAF complex (45). Trx and Trr both show extensive distribution along polytene chromosomes, although neither protein is required for bulk levels of H3K4me3. Perhaps Trx and Trr implement H3K4 methylation in a more gene-specific manner, at distinct stages of transcriptional regulation, or alternatively, have other substrates or functions.

Our biochemical studies have demonstrated that the *Drosophila* complexes are very similar to their mammalian counterparts in subunit composition. Our studies have also demonstrated the utility of a baculovirus superinfection system for expressing proteins in *Drosophila* cells. Large-scale transient transfections offer several potential advantages over generating clonal stable cell lines, one of which is that the overexpression of some proteins could be toxic to cells. This can be a problem even when using inducible promoters, such as the *Mtn* promoter, due to leaky expression under uninduced conditions. Moreover, the baculovirus infection and expression strategy took about 3 weeks from the cloning of the cDNA into the viral vector, generating the virus, infection of S2 cells, and purification of the complexes from nuclear extracts. In contrast, conventional cloning took 4 months from cloning the cDNA into the vector to generating and characterizing the clonal cell lines. We also purified FLAG-HA-dWDR82 from both stably transfected S2 cells and from the superinfection system and found that both strategies yielded a strikingly similar enrichment of target proteins (data not shown).

We found that all of the COMPASS family members in *Drosophila* have several common subunits, namely, Ash2, Rbbp5, Wdr5, and Dpy30, which are homologs of CPS60, CPS50, CPS30, and CPS25, respectively, as well as each having complex-specific subunits (Fig. 7) (10). Many of these subunits have established, conserved roles in both the yeast and mammalian complexes: ASH2L is required for proper H3K4 trimethylation, as is CPS60 in yeast; both WDR5 in humans and CPS30 in yeast are required for the mono-, di-, and trimethylation of H3K4, and each is required for proper formation of the COMPASS and MLL complexes (9, 47). Conservation of this degree in the H3K4 methylation machinery suggests that *Drosophila* might have similar machinery. However, it had previously been reported that Trx forms a complex with CBP and SBF (38), but no corresponding complexes have been found in mammals (53).

Our demonstration of the presence of shared components between COMPASS and COMPASS-like complexes in *Drosophila* supports the findings that these proteins are required for the proper functional architecture critical for the methylation of H3K4. The complex-specific components we found in association with the dSet1, Trx, and Trr complexes further demonstrate a one-to-one correspondence of subunits between the *Drosophila* and human COMPASS family members that will allow the use of *Drosophila* as a model system for understanding the function of the human complexes. For example, while Set1/COMPASS is conserved from yeast to humans, it is possible that the metazoan complexes have additional functions needed for development. As the subunit compositions of both the SET1A and SET1B complexes are identical (26, 53), it is likely that their functional analysis would be hindered by redundancy between the two complexes. The presence of a single dSet1 complex in flies may serve as an excellent starting point to dissect the metazoan-specific functions of the SET1 complexes.

MLL-related proteins are multidomain proteins with the capacity to bind to many other proteins that may modulate their function. For example, Menin binds to the extreme N terminus of MLL1/2 and is required for proper targeting of the MLL1/2 complex to chromatin (17, 55, 56). Owing to its conserved components and interactions, but nonredundant nature, investigation of the *Drosophila* Trx complex promises to aid in our understanding of the MLL1 and MLL2 complexes, specifically in their role in development.

We currently have a very limited understanding of the functions of the various domains within the MLL3/4 proteins. Our identification of LPT, which is homologous to the N terminus of MLL3/4, as a component of the Trr complex indicates the

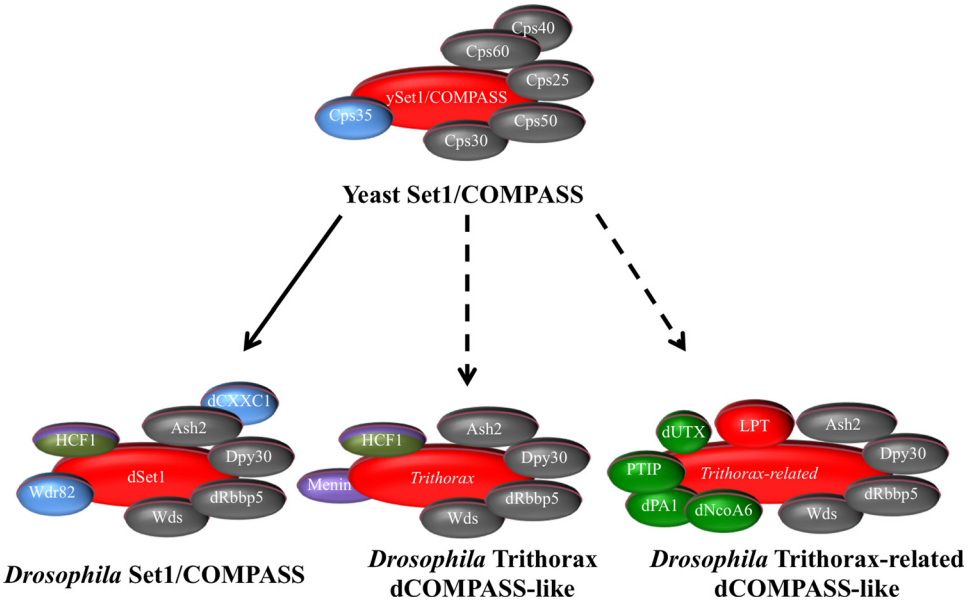

FIG. 7. Subunit composition for the histone H3K4 methyltransferase COMPASS and COMPASS-like complexes in yeast and *Drosophila*. Yeast Set1/COMPASS, the founding member of the H3K4 methylases, is represented as a direct descendant, reflecting the similarity of its protein composition to that of *Drosophila* Set1/dCOMPASS. Two other COMPASS-like complexes are the Trithorax-containing complex and the LPT/Trr complex. LPT and Trr together constitute the ortholog of the MLL3/4 proteins in mammals.

importance of PHD fingers residing in the LPT protein for the proper functioning and/or targeting of the Trr complex to chromatin. This separation of the MLL3/4 protein in *Drosophila* as Trr and LPT could allow us to more easily dissect the functions of N and C termini. Various studies have identified mutations in MLL3, MLL4, and UTX in a variety of cancers (2, 34, 36, 50). Therefore, studies of the LPT-Trr complex could improve our understanding of the targeting and regulation of these complexes with relevance to human disease.

Importantly, *Drosophila* has a single representative of each class of COMPASS family members found in mammals, in which two representatives of each complex exist. In contrast, nematodes, such as the genetically tractable *Caenorhabditis elegans*, contain only a Set1 and MLL3/4-related protein, but no MLL1/2 representative (44). Given the power of genetic manipulation, the identification of the COMPASS, Trx, and Trr complexes in *Drosophila* that share similar subunits with their mammalian counterparts will greatly facilitate our understanding of the biological functions of the H3K4 methylation machinery in development and differentiation.

ACKNOWLEDGMENTS

We are grateful to the Workman and Conaway laboratories for providing pRmHA3 and BacPAK8-EGFP vectors and to Laura Shilatifard for editorial assistance.

This work was supported by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund to H.M.H. and a grant, R01CA150265, from the National Institutes of Health to A.S.

REFERENCES

1. **Ardehali, M. B., et al.** 2011. Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J. **30:**2817–2828.
2. **Ashktorab, H., et al.** 2010. Distinct genetic alterations in colorectal cancer. PLoS One **5:**e8879.
3. **Bernstein, B. E., et al.** 2002. Methylation of histone H3 Lys 4 in coding regions of active genes. Proc. Natl. Acad. Sci. U. S. A. **99:**8695–8700.
4. **Bernstein, B. E., et al.** 2005. Genomic maps and comparative analysis of histone modifications in human and mouse. Cell **120:**169–181.
5. **Boehm, A. K., A. Saunders, J. Werner, and J. T. Lis.** 2003. Transcription factor and polymerase recruitment, modification, and movement on dhsp70 in vivo in the minutes following heat shock. Mol. Cell. Biol. **23:**7628–7637.
6. **Breen, T. R., and P. J. Harte.** 1991. Molecular characterization of the Trithorax gene, a positive regulator of homeotic gene expression in Drosophila. Mech. Dev. **35:**113–127.
7. **Chinwalla, V., E. P. Jane, and P. J. Harte.** 1995. The Drosophila trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. EMBO J. **14:**2056–2065.
8. **Cho, Y. W., et al.** 2007. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. J. Biol. Chem. **282:**20395–20406.
9. **Dou, Y., et al.** 2006. Regulation of MLL1 H3K4 methyltransferase activity by its core components. Nat. Struct. Mol. Biol. **13:**713–719.
10. **Eissenberg, J. C., and A. Shilatifard.** 2010. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev. Biol. **339:**240–249.
11. **Fang, M., et al.** 2009. Drosophila ptip is essential for anterior/posterior patterning in development and interacts with the PcG and trxG pathways. Development **136:**1929–1938.
12. **Fortschegger, K., and R. Shiekhattar.** 2011. Plant homeodomain fingers form a helping hand for transcription. Epigenetics **6:**4–8.
13. **Guru, S. C., et al.** 2001. Characterization of a MEN1 ortholog from Drosophila melanogaster. Gene **263:**31–38.
14. **Herz, H. M., et al.** 2010. The H3K27me3 demethylase dUTX is a suppressor of Notch- and Rb-dependent tumors in Drosophila. Mol. Cell. Biol. **30:**2485–2497.
15. **Higa, L. A., et al.** 2006. CUL4-DDB1 ubiquitin ligase interacts with multiple WD40-repeat proteins and regulates histone methylation. Nat. Cell Biol. **8:**1277–1283.
16. **Hollmann, M., E. Simmerl, U. Schafer, and M. A. Schafer.** 2002. The essential Drosophila melanogaster gene wds (will die slowly) codes for a WD-repeat protein with seven repeats. Mol. Genet. Genomics **268:**425–433.
17. **Hughes, C. M., et al.** 2004. Menin associates with a trithorax family histone methyltransferase complex and with the hoxc8 locus. Mol. Cell **13:**587–597.
18. **Ingham, P., and R. Whittle.** 1980. Trithorax: a new homoeotic mutation of *Drosophila melanogaster* causing transformations of abdominal and thoracic imaginal segments. Mol. Gen. Genet. **179:**607–614.
19. **Issaeva, I., et al.** 2007. Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. Mol. Cell. Biol. **27:**1889–1903.
20. **Koelle, M. R., et al.** 1991. The Drosophila EcR gene encodes an ecdysone receptor, a new member of the steroid receptor superfamily. Cell **67:**59–77.
21. **Kouzarides, T.** 2007. Chromatin modifications and their function. Cell **128:**693–705.

22. **Krogan, N. J., et al.** 2002. COMPASS, a histone H3 (lysine 4) methyltransferase required for telomeric silencing of gene expression. J. Biol. Chem. **277:**10753–10755.
23. **Krogan, N. J., et al.** 2003. The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol. Cell **11:**721–729.
24. **Lee, D. F., C. C. Chen, T. A. Hsu, and J. L. Juang.** 2000. A baculovirus superinfection system: efficient vehicle for gene transfer into Drosophila S2 cells. J. Virol. **74:**11873–11880.
25. **Lee, J. H., and D. G. Skalnik.** 2005. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J. Biol. Chem. **280:**41725–41731.
26. **Lee, J. H., and D. G. Skalnik.** 2008. Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. Mol. Cell. Biol. **28:**609–618.
27. **Lee, J. H., C. M. Tate, J. S. You, and D. G. Skalnik.** 2007. Identification and characterization of the human Set1B histone H3-Lys4 methyltransferase complex. J. Biol. Chem. **282:**13419–13428.
28. **Lee, J. S., et al.** 2007. Histone crosstalk between H2B monoubiquitination and H3 methylation mediated by COMPASS. Cell **131:**1084–1096.
29. **Lee, M. G., et al.** 2007. Demethylation of H3K27 regulates polycomb recruitment and H2A ubiquitination. Science **318:**447–450.
30. **Maddison, D. R., and K. S. Schultz.** 2007. The Tree of Life web project. http://tolweb.org.
31. **Miller, T., et al.** 2001. COMPASS: a complex of proteins associated with a Trithorax-related SET domain protein. Proc. Natl. Acad. Sci. U. S. A. **98:** 12902–12907.
32. **Mohan, M., et al.** 2010. Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes Dev. **24:**574–589.
33. **Mohan, M., C. Lin, E. Guest, and A. Shilatifard.** 2010. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. Nat. Rev. Cancer **10:**721–728.
34. **Morin, R. D., et al.** 2011. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature **476:**298–303.
35. **Ng, H. H., F. Robert, R. A. Young, and K. Struhl.** 2003. Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell **11:**709–719.
36. **Parsons, D. W., et al.** 2011. The genetic landscape of the childhood cancer medulloblastoma. Science **331:**435–439.
37. **Patel, S. R., D. Kim, I. Levitan, and G. R. Dressler.** 2007. The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. Dev. Cell **13:**580–592.
38. **Petruk, S., et al.** 2001. Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. Science **294:**1331–1334.
39. **Ringrose, L., and R. Paro.** 2004. Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. Annu. Rev. Genet. **38:**413–443.
40. **Roguev, A., et al.** 2001. The Saccharomyces cerevisiae Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. EMBO J. **20:**7137–7148.
41. **Schuettengruber, B., D. Chourrout, M. Vervoort, B. Leblanc, and G. Cavalli.** 2007. Genome regulation by polycomb and trithorax proteins. Cell **128:**735–745.
42. **Sedkov, Y., et al.** 2003. Methylation at lysine 4 of histone H3 in ecdysone-dependent development of Drosophila. Nature **426:**78–83.
43. **Shilatifard, A.** 2008. Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. Curr. Opin. Cell Biol. **20:**341–348.
44. **Smith, E., C. Lin, and A. Shilatifard.** 2011. The super elongation complex (SEC) and MLL in development and disease. Genes Dev. **25:**661–672.
45. **Smith, E., and A. Shilatifard.** 2010. The chromatin signaling pathway: diverse mechanisms of recruitment of histone-modifying enzymes and varied biological outcomes. Mol. Cell **40:**689–701.
46. **Smith, E. R., et al.** 2008. Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. Mol. Cell. Biol. **28:**1041–1046.
47. **Steward, M. M., et al.** 2006. Molecular regulation of H3K4 trimethylation by ASH2L, a shared subunit of MLL complexes. Nat. Struct. Mol. Biol. **13:**852–854.
48. **Suganuma, T., et al.** 2008. ATAC is a double histone acetyltransferase complex that stimulates nucleosome sliding. Nat. Struct. Mol. Biol. **15:**364–372.
49. **Tanaka, Y., Z. Katagiri, K. Kawahashi, D. Kioussis, and S. Kitajima.** 2007. Trithorax-group protein ASH1 methylates histone H3 lysine 36. Gene **397:** 161–168.
50. **van Haaften, G., et al.** 2009. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat. Genet. **41:**521–523.
51. **Wood, A., et al.** 2003. Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. Mol. Cell **11:**267–274.
52. **Wood, A., et al.** 2007. Ctk complex-mediated regulation of histone methylation by COMPASS. Mol. Cell. Biol. **27:**709–720.
53. **Wu, M., et al.** 2008. Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol. Cell. Biol. **28:**7337–7344.
54. **Wyss, C.** 1979. Cloning of Drosophila cells: effect of vitamins and yeast extract components. Somatic Cell Genet. **5:**23–28.
55. **Yokoyama, A., and M. L. Cleary.** 2008. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. Cancer Cell **14:**36–46.
56. **Yokoyama, A., et al.** 2005. The Menin tumor suppressor protein is an essential oncogenic cofactor for MLL-associated leukemogenesis. Cell **123:**207–218.
57. **Yokoyama, A., et al.** 2004. Leukemia proto-oncoprotein MLL forms a SET1-like histone methyltransferase complex with Menin to regulate Hox gene expression. Mol. Cell. Biol. **24:**5639–5649.
58. **Yokoyama, R., A. Pannuti, H. Ling, E. R. Smith, and J. C. Lucchesi.** 2007. A plasmid model system shows that Drosophila dosage compensation depends on the global acetylation of histone H4 at lysine 16 and is not affected by depletion of common transcription elongation chromatin marks. Mol. Cell. Biol. **27:**7865–7870.
59. **Yuan, W., et al.** 2011. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J. Biol. Chem. **286:**7983–7989.
60. **Zhang, Y., Z. Wen, M. P. Washburn, and L. Florens.** 2010. Refinements to label free proteome quantitation: how to deal with peptides shared by multiple proteins. Anal. Chem. **82:**2272–2281.