

SFMBT1 functions with LSD1 to regulate expression of canonical histone genes and chromatin-related factors

Jin Zhang,¹ Roberto Bonasio,¹ Francesco Strino,² Yuval Kluger,² J. Kim Holloway,³ Andrew J. Modzelewski,³ Paula E. Cohen,³ and Danny Reinberg^{1,4}

¹Howard Hughes Medical Institute, Department of Biochemistry, New York University School of Medicine, New York, New York 10016, USA; ²Department of Pathology, Yale Cancer Center, Yale University School of Medicine, New Haven, Connecticut 06520, USA; ³Department of Biomedical Sciences, College of Veterinary Medicine, Cornell University, Ithaca, New York 14853, USA

SFMBT1 (Scm [Sex comb on midleg] with four MBT [malignant brain tumor] domains 1) is a poorly characterized mammalian MBT domain-containing protein homologous to *Drosophila* SFMBT, a *Polycomb* group protein involved in epigenetic regulation of gene expression. Here, we show that SFMBT1 regulates transcription in somatic cells and during spermatogenesis through the formation of a stable complex with LSD1 and CoREST. When bound to its gene targets, SFMBT1 recruits its associated proteins and causes chromatin compaction and transcriptional repression. SFMBT1, LSD1, and CoREST share a large fraction of target genes, including those encoding replication-dependent histones. Simultaneous occupancy of histone genes by SFMBT1, LSD1, and CoREST is regulated during the cell cycle and correlates with the loss of RNA polymerase II at these promoters during G2, M, and G1. The interplay between the repressive SFMBT1–LSD1–CoREST complex and RNA polymerase II contributes to the timely transcriptional regulation of histone genes in human cells. SFMBT1, LSD1, and CoREST also form a stable complex in germ cells, and their chromatin binding activity is regulated during spermatogenesis.

[Keywords: LSD1; SFMBT; polycomb]

Supplemental material is available for this article.

Received November 27, 2012; revised version accepted March 15, 2013.

Transcriptional competency of chromatin is a function of its structure, which in turn is dictated in part by the distinctive properties of the histone species comprising the nucleosomal unit. These determinant nucleosomal features arise from the regulated expression of histone genes that encode canonical and variant histone proteins (Campos and Reinberg 2009; Banaszynski et al. 2010) and the regulated catalysis of post-translational modifications (PTM) to which they are subjected (Kouzarides 2007).

Coincident with DNA replication, the genes encoding canonical histones initiate transcription rapidly at the start of S phase, and the resultant mRNAs are quickly degraded at the end of S phase (Osley 1991; Stein et al. 2006; Marzluff et al. 2008). The specific and timely execution of this process is enforced by proteins that bind to stem-loops in the 3' untranslated region (UTR) of histone mRNAs and regulate their metabolism at the post-transcriptional level (Marzluff et al. 2008). Multiple copies of the genes

for canonical histones are organized as clusters in the genome. The multigene family encoding the five canonical histones H1, H2A, H2B, H3, and H4 is up-regulated at the transcriptional level during DNA replication (S phase) and maps to a large *HIST1* locus and the smaller *HIST2*, *HIST3*, and *HIST4* clusters in both human and mouse genomes (Marzluff et al. 2002). This genomic organization likely facilitates the efficient and coordinated histone gene expression associated with S phase.

Besides canonical histones, multiple histone variants convey important information during chromatin-based processes, including transcriptional regulation, DNA damage and repair, and centromere and kinetochore assembly (Sarma and Reinberg 2005; Banaszynski et al. 2010). In contrast to canonical histone mRNAs, histone variant mRNAs are mostly polyadenylated, and their expression is not regulated as tightly through the cell cycle (Marzluff et al. 2008). Some testis-specific histone variants are also encoded within the replication-dependent histone clusters but are only expressed during spermatogenesis. These testis-specific histones replace their canonical counterparts during meiosis and are in turn replaced by transition proteins and protamines after meiosis (Kimmins and Sassone-Corsi 2005; Banaszynski et al. 2010).

⁴Corresponding author

E-mail danny.reinberg@nyumc.org

Article is online at <http://www.genesdev.org/cgi/doi/10.1101/gad.210963.112>.

Freely available online through the *Genes & Development* Open Access option.

Many histone PTMs and the processes involved in establishing, removing, recognizing, and propagating these marks exert profound effects on chromatin structure, gene transcription, and epigenetic inheritance (Berger 2007; B Li et al. 2007; Campos and Reinberg 2009). Histone PTMs (or their absence) exert their functions by creating binding surfaces that are recognized by specific protein domains that are present, often in modular fashion, in several chromatin-associated proteins and orchestrate the recruitment of multisubunit complexes that further affect chromatin function and transcription (Maurer-Stroh et al. 2003; Ruthenburg et al. 2007; Taverna et al. 2007). For example, the malignant brain tumor (MBT) domain is a binding module that recognizes mono- and dimethylated lysines on histone tails (Bonasio et al. 2010) through a pocket lined with aromatic residues (Sathyamurthy et al. 2003; Wang et al. 2003; H Li et al. 2007; Min et al. 2007; Taverna et al. 2007).

Three such MBT domain-containing proteins (Supplemental Table S1) have been identified in *Drosophila*: L(3)mbt, Sex comb on midleg (Scm), and Scm with four MBT domains (Sfmbt). In flies, *Sfmbt* and *Scm* belong to the *Polycomb* group (PcG) of genes, which are critical for the epigenetic control of gene expression and the maintenance of cellular identity (Simon and Kingston 2009; Beisel and Paro 2011). PcG genes typically encode proteins that assemble into multisubunit protein complexes (Supplemental Table S1) that associate with chromatin and alter its structure to enforce transcriptional repression at the epigenetic level (Simon and Kingston 2009; Beisel and Paro 2011). Among these, the best studied are *Polycomb*-repressive complex-1 (PRC1) and PRC2, which play key roles during development and exhibit conserved functions in most animals and plants (Simon and Kingston 2009; Beisel and Paro 2011). *Drosophila* Sfmbt (dSfmbt) is a *Polycomb* protein that forms a less well-studied *Polycomb* complex named Pho-repressive complex (PhoRC) (Klymenko et al. 2006). Unlike PRC1 and PRC2, PhoRC lacks a mammalian counterpart. In fact, the other subunit of *Drosophila* PhoRC, Pho, is only poorly conserved in mammals, and its ortholog, YY1, does not stably associate with any MBT domain-containing proteins (Cai et al. 2007; Wu et al. 2007a). L3MBTL2, a mammalian homolog of dSfmbt that lacks the C-terminal SPM domain [named after the three *Drosophila* proteins in which it was discovered: Scm, Ph, and L(3)mbt], forms a complex with E2F6 and several *Polycomb* proteins, such as RING1A, RING1B, and MBLR (Ogawa et al. 2002; Trojer et al. 2011) but not YY1.

SFMBT1 and SFMBT2 are additional mammalian homologs of dSfmbt that contain four MBT domains at the N terminus and an SPM domain at the C terminus, sharing the same domain architecture as dSfmbt (Bonasio et al. 2010). Although overexpressed SFMBT2 and YY1 interact in 293 cells, they do not form a stable complex (Kuzmin et al. 2008). Thus, the questions of which protein complexes comprise mammalian SFMBT proteins and what their functions are remain unanswered.

Previous investigations have revealed the cellular functions of other MBT domain-containing proteins. L3MBTL1,

a mammalian homolog of dL(3)mbt, compacts chromatin *in vitro* in a histone PTM-dependent manner (Trojer et al. 2007). Under the same experimental setting, the chromatin compaction caused by L3MBTL2 binding does not require histone PTMs (Trojer et al. 2011). Both L3MBTL2 and SFMBT1 repress gene transcription when recruited to an integrated luciferase reporter (Wu et al. 2007b; Trojer et al. 2011), but the natural genomic targets of SFMBT1 and its physiological role in regulating gene expression are not known.

Here, we demonstrate that SFMBT1 and SFMBT2 are functionally divergent, as they interact with distinct groups of polypeptides, and show that a ternary SFMBT1–LSD1–CoREST complex targets genes enriched for functions related to chromatin and nucleosome assembly, including the replication-dependent histone gene clusters. Consistent with this, some histone genes are derepressed after knockdown of SFMBT1–LSD1–CoREST. SFMBT1 is highly expressed in mouse germ cells, where it associates with the synaptic regions of meiotic chromosomes in pachytene stage spermatocytes. SFMBT1 and LSD1 simultaneously occupy the promoters of many genes during pachytene stage spermatocyte but not in spermatids, suggesting that they function in a stage-specific fashion during spermatogenesis.

Results

SFMBT1, LSD1, and CoREST form a stable complex

To identify polypeptides associated with SFMBT1 and SFMBT2, we performed Flag-HA tandem affinity purifications (TAP) in 293T-REx cells that expressed C-terminal HA/Flag-tagged (–CHF) SFMBT1 and SFMBT2 at endogenous levels upon induction with doxycycline (dox). As a control, extracts from 293T-REx cells transfected with the empty vector were subjected to the same purification scheme. Affinity-purified polypeptides were resolved by SDS-PAGE, visualized by silver staining (Fig. 1A, left), and identified by mass spectrometry. Although SFMBT1 and SFMBT2 share the same domain architecture, the sets of associated polypeptides were largely distinct (Fig. 1A, right), suggesting a remarkable functional diversification. SFMBT1 associated with an LSD1- and CoREST-containing complex that included KDM1A/LSD1, RCOR1/CoREST, ZNF516, ZNF217, HDAC2, and PHF21A (Ballas et al. 2001; Hakimi et al. 2002; Humphrey et al. 2001; You et al. 2001). None or only a small number (one to six) of peptides belonging to these proteins were detected in the SFMBT2-CHF TAP, likely an indication that trace amounts of SFMBT1-containing complexes were present in the SFMBT2 TAP, as evidenced by the presence of seven peptides for SFMBT1 (Fig. 1A). On the other hand, SFMBT2 associated with castor zinc finger 1 (CASZ1), ubiquitin-conjugating enzyme E2O (UBE2O), and PHC1B, none of which were detectable in the SFMBT1-CFH TAP. Both SFMBT1 and SFMBT2 interacted with components of the PRC1.4 complex (Gao et al. 2012) as well as SAM/SPM domain-containing proteins such as SAMD1, L3MBTL3, SCML2, L3MBTL4, and SCMHI, consistent with the known

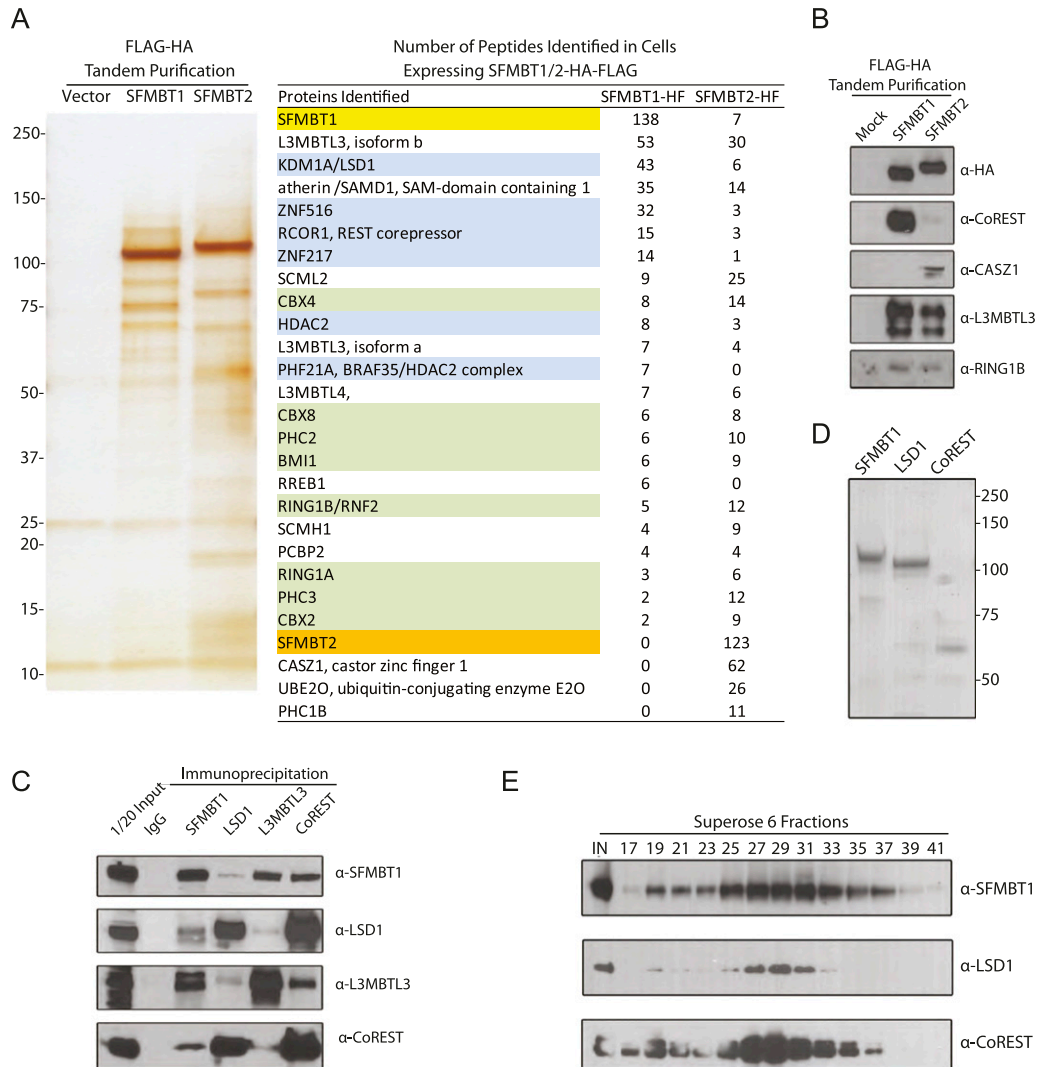


Figure 1. SFMBT1, LSD1, and CoREST form a stable complex. (A) Polypeptides identified in SFMBT1/2 TAPs. (Left) Silver staining of elutions from TAP resolved on a 4%–15% gradient SDS-PAGE. The most prominent bands in lanes 2 and 3 migrate at the predicted molecular weight of SFMBT1-FH and SFMBT2-FH, respectively. (Right) Associated polypeptides and total number of peptides identified by mass spectrometry in a representative experiment. Bait proteins SFMBT1 and SFMBT2 are highlighted in yellow and orange, respectively, components of the LSD1–CoREST complex are highlighted in blue, and those of the PRC1 complex are in green. (B) Western blot analysis of specific associated polypeptides using elutions from TAPs. (C) Immunoprecipitation from HeLa S3 nuclear extract using antibodies against endogenous SFMBT1, LSD1, L3MBTL3, and CoREST. (D) Coomassie blue staining of purified recombinant proteins derived from baculovirus-infected Sf9 cells. (E) Profile of Superose 6 gel filtration column fractions assayed by Western blot. A mix of equivalent amounts of SFMBT1, LSD1, and CoREST proteins was used as input, and the odd-numbered fractions after the void volume were analyzed.

ability of the SPM domain to mediate homo- and heterotypic interactions (Peterson et al. 1997).

We validated the specificity of these interactions by Western blot using TAP elutions (Fig. 1B) and endogenous immunoprecipitations from HeLa S3 nuclear extracts (Fig. 1C). SFMBT1 associated with LSD1, CoREST, and L3MBTL3 in HeLa S3 nuclear extracts. Immunoprecipitations using the SFMBT2 antibody were not performed, as it also recognized SFMBT1, likely because of the sequence homology between the two paralogs (see below; data not shown). To test whether the interactions between SFMBT1, LSD1, and CoREST led to the formation

of a stable complex, recombinant versions of these proteins were separately expressed in insect cells (Fig. 1D), purified and mixed *in vitro*, and subjected to size exclusion chromatography. The elution profile of the three proteins coincided and peaked at an apparent mass of ~300 kDa (Fig. 1E, fractions 27–31), which corresponds to the sum of their expected molecular weight (~250 kDa), suggesting that they form a stable complex. The sedimentation profile of endogenous SFMBT1, LSD1, and CoREST from mouse lysates also coincided and peaked between 150 kDa and 443 kDa (Fig. 6D; see below), indicating that these three polypeptides form the stable

core of a complex and that the additional proteins detected by mass spectrometry likely reflect more transient or sporadic interactions. We conclude that SFMBT1 directly associates with LSD1 and CoREST in a stable complex, which we call SLC, from the initials of the three main components.

Genomic targets of the SLC complex

Having found that SFMBT1 forms a stable complex with LSD1 and CoREST (the SLC complex), we sought to identify its target genes using chromatin immunoprecipitation (ChIP) followed by deep sequencing (ChIP-seq). ChIP-seq experiments were performed with antibodies against endogenous SFMBT1, LSD1, and CoREST in HeLa S3 cells, where the interactions between these proteins had been validated (Fig. 1C). Our genome-wide analyses identified thousands of enriched regions (ERs) for these three proteins (see the Materials and Methods), the large majority of which were located within 2 kb from transcriptional start sites (TSSs) (Supplemental Fig. S1A), further implicating the role of the SLC complex in the regulation of transcription.

In keeping with the biochemical observations reported above, 70% of SFMBT1 targets were also occupied by LSD1 or CoREST, and overall, 1554 genes were bound by all three components (Fig. 2A). Within these targets, 465 promoters containing the highest levels of SFMBT1, LSD1, and CoREST (Supplemental Fig. S1B, red box) were enriched for genes with gene ontology (GO) terms related to nucleosome and chromatin assembly functions (Fig. 2B). Consistent with this, the SLC complex occupied the TSS and/or proximal promoter of many histone gene loci, the most prominent being the replication-dependent, histone genes within the *HIST1* (Fig. 2C, left) and *HIST2* (Supplemental Fig. S1C) clusters. High occupancy of the SLC components was also evident at other loci, including *PWWP2A* (Fig. 2C, right), a gene encoding a PWWP domain-containing protein, and *SIX2* (data not shown), a gene encoding a homeobox-containing transcription factor that has important functions during development (Christensen et al. 2008).

To analyze the genomic distribution of the SLC complex in the context of chromatin structure and gene activity, we computed the read densities of SFMBT1, LSD1, and CoREST as well as those of well-characterized histone PTMs and RNA polymerase II (Pol II) in the top 1000 ERs for SFMBT1. The binding profiles of SLC components were highly correlated, and unsupervised hierarchical clustering assigned them to the same branch, separated from the other chromatin features analyzed (Fig. 2D). Among histone PTMs, the distribution of the SLC complex more closely resembled that of H3K4me2 (Fig. 2D), a known target for LSD1 demethylation activity (Shi et al. 2004). The distribution of LSD1 and CoREST, but not that of SFMBT1, also correlated with that of H3K4me1 (Fig. 2D), a known marker for enhancer function (Spitz and Furlong 2012); however, LSD1–CoREST-binding sites that also contained SFMBT1 (i.e., the targets of the SLC complex) exhibited comparably lower levels of

H3K4me1 and higher levels of H3K4me2 (Fig. 2E). This is consistent with the observation that the peak of SLC complex occupancy always corresponds with a depression (“valley”) in H3K4me1 density (Fig. 2C). Overall, our genome-wide analysis suggests that the SLC complex occupies a structurally and functionally distinct subset of LSD1–CoREST targets and that SFMBT1 might modulate the function of LSD1 and its associated subunits in these regions.

SFMBT1 recruits its associated proteins to gene targets

To determine the downstream consequence of SFMBT1 binding to chromatin as well as that of the SLC complex, we tethered SFMBT1 to an integrated luciferase reporter gene using a dox-inducible GAL4-upstream activation sequence (UAS) system previously described (Vaquero et al. 2004). We induced the expression of stably transfected GAL4-SFMBT1 in 293T-REx cells and analyzed the transcription of the stably integrated luciferase reporter as well as its chromatin structure (Fig. 3A). Immunoprecipitated chromatin using antibodies specific to the endogenous candidates was analyzed using primers spanning the TK promoter and the start of the luciferase gene (*TK-LUC*) (Fig. 3A, red line). When bound to *TK-LUC*, GAL4-SFMBT1 recruited LSD1 and also CBX8, a component of PRC1.4 (Fig. 3B; Gao et al. 2012), consistent with the biochemical composition of SFMBT1-containing complexes determined by affinity purification (Fig. 1A). Upon SFMBT1 binding, we detected increased nucleosome density, decreased levels of H3K4me3 (a mark that correlates with active transcription), and increased levels of H3K27me2/3 (repressive marks) (Fig. 3B), indicating that the presence of SFMBT1 induces the formation of an overall repressive chromatin configuration. GAL4 alone and GAL4-SFMBT2 failed to recruit LSD1 (Fig. 3C), indicating that SLC assembly on and recruitment to chromatin are specific functions of SFMBT1. Genome-wide analyses revealed that, in addition to being recruited to the UAS-containing reporter, GAL4-tagged SFMBT1 and SFMBT2 retained the ability to bind to the endogenous SFMBT targets (data not shown); thus, we were able to confirm the specificity of LSD1 recruitment by GAL4-SFMBT1 and not GAL4-SFMBT2 upon induction (Fig. 3D). Moreover, when cells were stably transfected with shRNA against SFMBT1, a reduction of SFMBT1, LSD1, and CoREST was observed at 10 SLC ERs tested, including the *SIX2* promoter (Fig. 3E; data not shown), indicating that SFMBT1 is not only sufficient but also necessary for SLC assembly at these endogenous target genes.

SFMBT1 compacts chromatin and represses transcription

To better understand the molecular mechanism responsible for the SFMBT1-induced formation of repressive chromatin (see Fig. 3B), we analyzed the effect of SFMBT1 truncations and single-point mutations (Fig. 4A) on the expression of the luciferase reporter located downstream from the artificial SFMBT1-tethering site (Fig. 3A). In the case of L3MBTL1, a protein with domain architecture

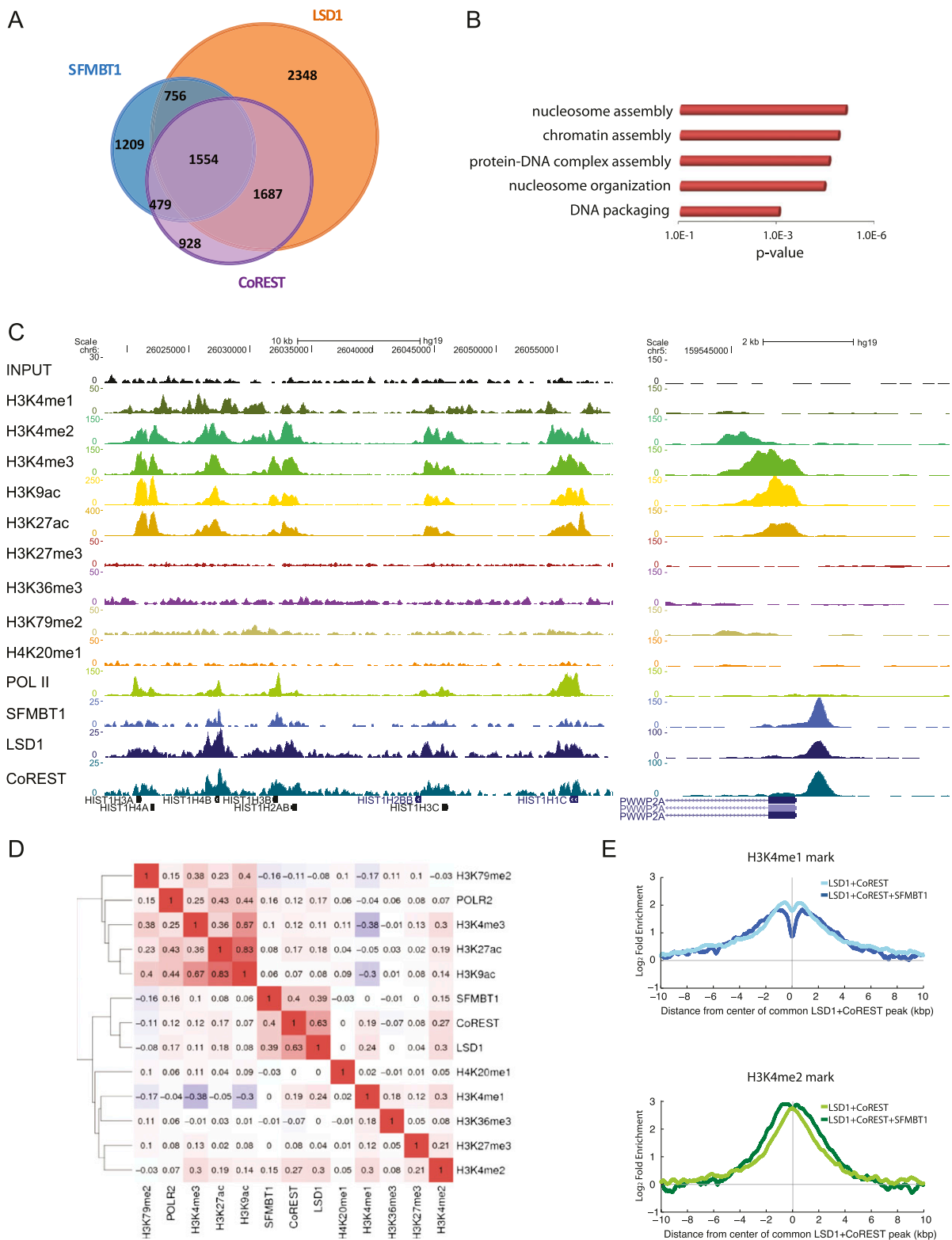


Figure 2. SFMBT1, LSD1, and CoREST share high percentages of target genes. (A) Venn diagram showing the overlap of gene targets for SFMBT1 (blue), CoREST (violet), and LSD1 (orange), as determined by ChIP-seq. Gene targets were determined by Genomic Regions Enrichment of Annotation Tool (GREAT) with the “basal plus extension” setting. (B) GO analysis of the ERs targeted by the SLC complex using the Database for Annotation, Visualization, and Integrated Discovery (DAVID). The X-axis corresponds to the P-value (in logarithmic scale). (C) Read density profile for input, H3K4me1, H3K4me2, H3K4me3, H3K9ac, H3K27ac, H3K27me3, H3K36me3, H3K79me2, H4K20me1, Pol II, SFMBT1, LSD1, and CoREST at two of the most enriched loci. The X-axis corresponds to genomic location, and the Y-axis corresponds to normalized ChIP-seq signal density. (D) Clustering of the SLC complex components with Pol II (POLR2) and multiple histone PTMs according to pairwise Pearson correlation scores of their ChIP-seq signal in the top 1000 SFMBT1 ERs. Red represents positive correlation, and blue represents negative correlation. Darker color indicates stronger positive/negative correlation. (E) H3K4me1 (top panel) and H3K4me2 (bottom panel) ChIP-seq signal around ERs that are co-occupied by LSD1 and CoREST with no or low SFMBT1 (lighter color) or high level of SFMBT1 (LSD1, CoREST, and SFMBT1) (darker color). The X-axis represents the distance from the center of ERs bound by LSD1 and CoREST, and the Y-axis represents log₂ (fold enrichment).

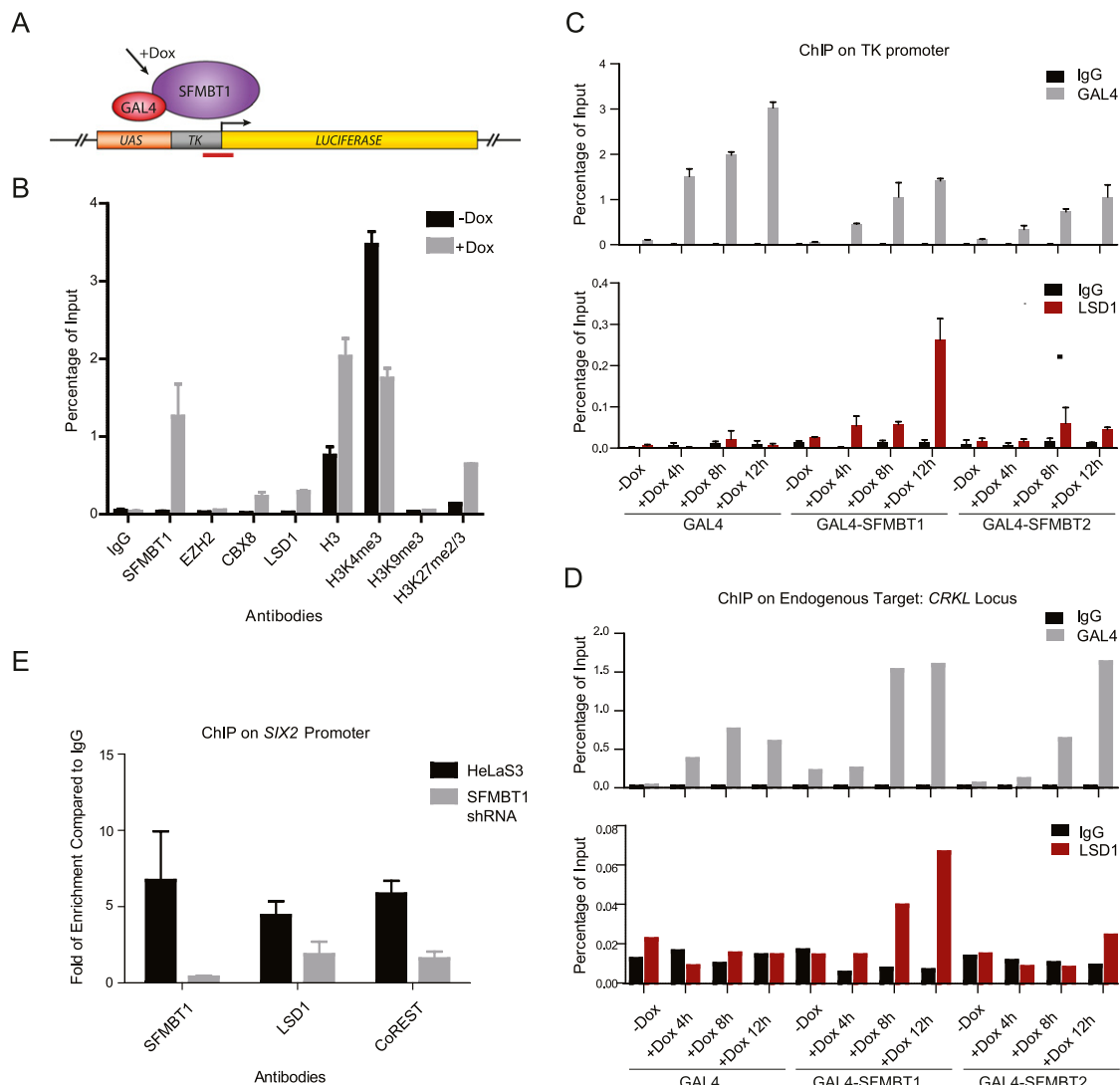


Figure 3. SFMBT1 recruits its associated polypeptides to gene targets. (A) Schematic representation of the GAL4-UAS targeting system. GAL4-tagged SFMBT1 was expressed upon dox treatment and bound to the UAS. The UAS lies upstream of the TK promoter and the luciferase gene. The red line indicates the position of the ChIP primer used in the following figures. (B) ChIP enrichment at the TK-luciferase locus without (–dox) and with (+dox) GAL4-SFMBT1 expression. Antibodies used in this assay are shown next to the X-axis. ChIP enrichment expressed as percentage of input is plotted on the Y-axis. Bars indicate the mean of three independent experiments + SEM. (C) GAL4-SFMBT1 specifically recruits LSD1 to the TK promoter. ChIP enrichments are shown without dox induction (–dox) or after 4-, 8-, and 12-h inductions of GAL4/GAL4-SFMBT1/GAL4-SFMBT2 expression for an IgG control (black), ChIP with a GAL4 antibody (gray), or LSD1 antibody (red). Bars represent the mean of two biological replicates each with two qPCRs + SEM. (D) ChIP enrichment at the *CRKL* locus. ChIP enrichments are shown without dox induction (–dox) or after 4-, 8-, and 12-h induction of GAL4/GAL4-SFMBT1/GAL4-SFMBT2 expression for an IgG control (black), ChIP with a GAL4 antibody (gray), or LSD1 antibody (red). Bars represent the mean of two experiments. (E) ChIP enrichment of SFMBT1, LSD1, and CoREST at the *SIX2* promoter without (HeLa S3) and with SFMBT1 shRNA transfected. Bars represent the mean of two biological replicates each with two qPCRs + SEM.

similar to SFMBT1, the MBT domains bind to methylated lysines, resulting in chromatin compaction, and are required for transcriptional repression (Trojer et al. 2007). However, in contrast to several other MBT domain-containing proteins (Bonasio et al. 2010), SFMBT1 did not associate with specific histone lysine modifications *in vitro* (Fig. 4B). Nevertheless, SFMBT1 did bind to histone tails (Fig. 4B), and mutations of the aromatic residues within

the MBT domains reduced (F173A and Y196A) or abolished (W180A) its histone tail-binding activity (Fig. 4C).

Next, we determined the contributions of wild-type and mutant SFMBT1 to the regulation of transcription. When GAL4-SFMBT1 was expressed, transcription from the integrated luciferase reporter was repressed by ~10-fold, and the four MBT domains at the N terminus of SFMBT1 were sufficient for transcriptional repression

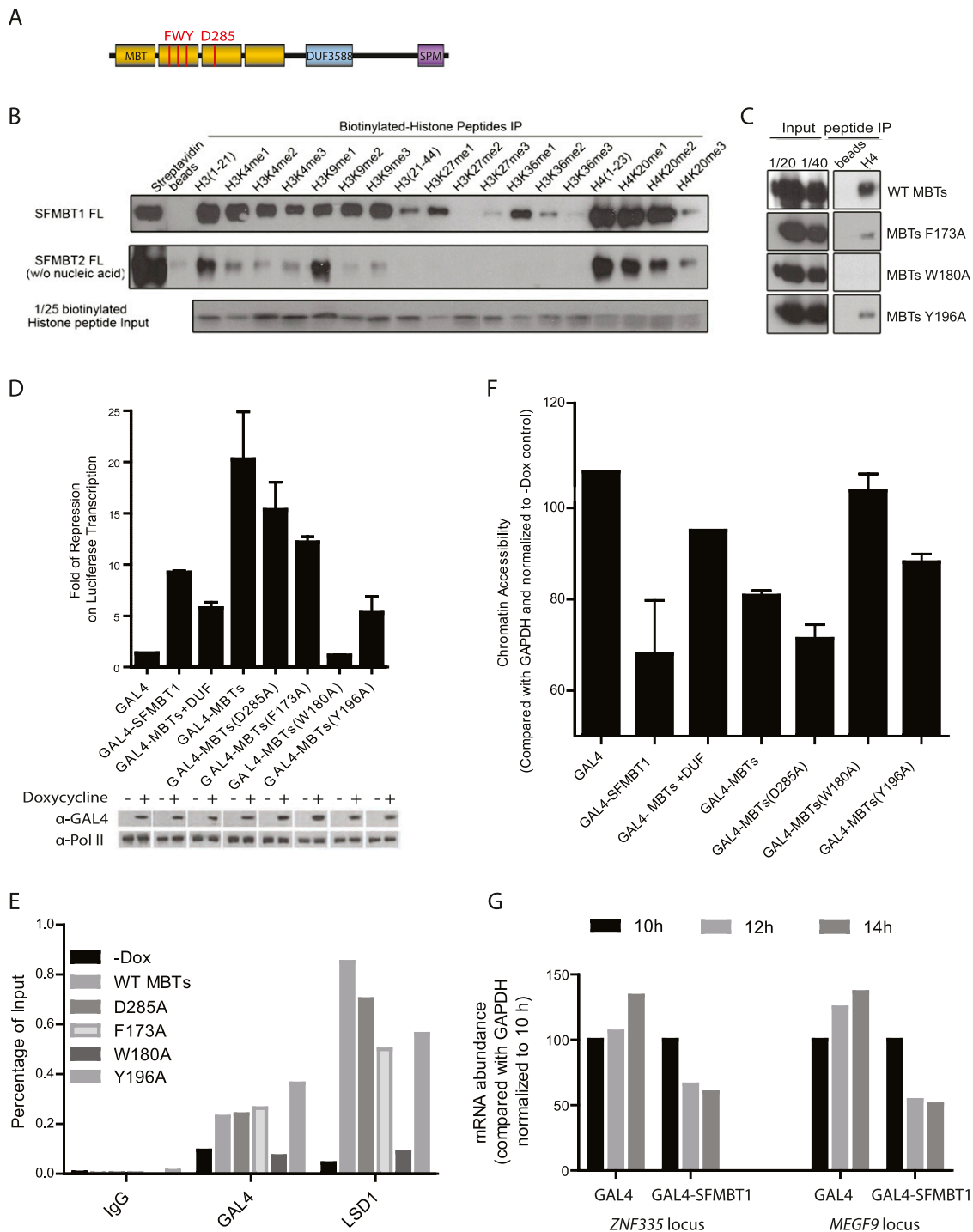


Figure 4. SFMBT1 represses transcription. (A) Domain architecture of the SFMBT1 protein. Positions of mutations discussed in the text are represented at the *top* by red lines with their corresponding amino acids. (B) Histone-binding activity of SFMBT1 proteins. Pull-down assay of full-length SFMBT1 or SFMBT2 recombinant proteins using biotinylated histone peptides. After pull-down, proteins bound to the histone peptides were loaded on a SDS-PAGE for Western blotting. The SFMBT2 proteins used in this assay were purified with a DE52 column to remove contaminating nucleic acid. One-twenty-fifth of the peptide input was loaded for Western blotting analysis and detected with a biotin antibody. (C) Pull-down assay of MBT domains of SFMBT1 (wild-type MBTs) or mutant domains (F173A, W180A, or Y196A) using biotinylated histone H4 peptides. (D) Effects of SFMBT1—either full-length, truncated, or mutant in the MBT domain—on luciferase transcription. The extent of repression compared with GAL4 control is indicated on the Y-axis (the higher the bar, the more repression of the reporter). Bars indicate the mean of triplicates + SEM. Expression levels of all GAL4-tagged proteins are shown at the *bottom* before (–) and after (+) induction. (E) The MBT domains of SFMBT1 recruit LSD1 to the TK promoter-luciferase locus upon dox induction. MBT domain mutants with no (W180A) or decreased (F173A and Y196A) binding affinity toward histone tails exhibit no or decreased recruitment of LSD1 to this locus. ChIP enrichment as percentage of input is plotted on the Y-axis; bars represent the mean of two experiments. (F) Effects of full-length SFMBT1, truncated SFMBT1, and MBT domain mutations on chromatin accessibility. The Y-axis corresponds to the percentage of chromatin accessibility relative to that of the *GADPH* gene and normalized by nonprotein control (–dox). Bars represent the mean of triplicates + SEM. (G) Target gene expression without (–dox) and with (+dox) GAL4 or GAL4-SFMBT1 binding. mRNA abundance (as compared with *GADPH* and normalized to 10 h after induction) of two endogenous targets (*ZNF335* and *MEGF9*) is shown after 10-, 12-, and 14-h induction of GAL4/GAL4-SFMBT1 expression. Bars represent the mean of two experiments.

(Fig. 4D). Mutations in SFMBT1 that affect its histone tail-binding activity showed decreased (F173A and Y196A) or abolished (W180A) transcriptional repression (Fig. 4D), consistent with the severity of the histone tail-binding defect. In addition, the MBT domain alone was sufficient to recruit LSD1 to the luciferase locus, and the W180A mutant that did not bind histone tail also failed to recruit LSD1 to chromatin (Fig. 4E).

SFMBT1 appeared to carry out its repressive function at least partially through changes in chromatin accessibility, which were tested by nuclease digestion (see the Materials and Methods for a detailed description of this assay) and confirmed that the chromatin at the integrated *TK-LUC* locus adopted a more compacted state upon SFMBT1 binding (Fig. 4F). The same mutations in the MBT domains that interfered with transcriptional repression also interfered with chromatin compaction.

As the genome-wide distribution of GAL4-SFMBT1 recapitulated that of endogenous SFMBT1 (see above), we used this inducible overexpression system to determine the functional consequences of increased SFMBT1 binding on the transcription of its endogenous target genes. Increased occupancy of GAL4-SFMBT1 at endogenous loci was detected 8 h after dox induction (Fig. 3D). To observe functional consequence after GAL4-SFMBT1 binding, changes in mRNA levels of several target genes were analyzed at 10–14 h after dox induction. GAL4-SFMBT1 led to decreased expression of these genes, whereas GAL4 alone was ineffectual (Fig. 4G).

In summary, our findings demonstrate that SFMBT1 binding to chromatin leads to a repressive chromatin state, which includes compacted chromatin brought about by interactions of the MBT domains with histone tails and by changes in histone PTMs, likely as a consequence of the recruitment of SFMBT1-associated histone-modifying complexes such as LSD1/CoREST and PRC1. This repressive chromatin state resulted in decreased transcription of both a luciferase reporter and endogenous target genes. Importantly, the chromatin-binding activity of the MBT domain of SFMBT1 was necessary for these functions.

Dynamic regulation of histone loci

To analyze the function of the SLC complex with respect to transcription from its endogenous gene targets, HeLa S3 cells were treated with either random negative control siRNAs (NC siRNA) or siRNAs against SFMBT1, LSD1, and CoREST (triple siRNA). Despite >60% knockdown for each of these mRNAs, cell proliferation was not significantly affected on the time scale of this experiment (data not shown). We harvested total RNA from these cells and sequenced the polyA⁺ fraction (PA⁺) as well as a fraction depleted of polyA⁺ and ribosomal RNA (Ribo⁻) that contained the bulk of the canonical, nonpolyadenylated histone transcripts. The analysis of the Ribo⁻ fraction revealed that several histone genes within the canonical histone clusters (*HIST1*, *HIST2*, *HIST3*, and *HIST4*) were derepressed upon combined knockdown of the three SLC core components (Fig. 5A), suggesting that the SLC complex repressed histone gene transcription. We noted

that the genes most affected were expressed at lower levels and were therefore most likely to suffer the consequences of a partial depletion of a repressive complex by siRNAs. *HIST1H1T* and *HIST3H3*, two testis-specific histone variant genes located at these clusters, were also derepressed in cells treated with triple siRNA (Fig. 5A).

The occupancy profiles obtained through ChIP-seq indicated that the replication-dependent histone loci targeted by the SLC complex were also occupied by Pol II and contained various active histone PTMs, including H3K4me2/3, H3K9ac, and H3K27ac (Fig. 2C), which are normally associated with active transcription, in apparent opposition to our findings regarding the repressive role of the SLC complex. However, the Pol II enrichment at nonhistone targets of the SLC complex was much lower than at these histone targets (Fig. 2C, right panel; data not shown). Given that ChIP-seq data reveal the chromatin-binding profile of an entire cell population, proteins that appear to occupy the same targets may not be bound at the same time or in the same cell. Therefore, we hypothesized that the SLC complex and the transcription machinery at these loci might be dynamically regulated to ensure timely histone gene expression, which must be tightly coupled to the different phases of the cell cycle. To test this hypothesis, we arrested HeLa S3 cells at the G1/S border with double thymidine treatment and released them to obtain a synchronized population (Supplemental Fig. S2A). We harvested the synchronized cells at different times after release and hence at different stages of the cell cycle. ChIP-seq analysis of the SLC components in these cell populations revealed that the SLC complex bound to histone loci during G1 but not S phase (Fig. 5B), consistent with the fact that replication-dependent histones are repressed during G1 phase (SLC bound) and expressed during S phase (SLC not bound). We then performed a “meta-gene” analysis, combining all histone genes in the *HIST1* and *HIST2* clusters, and observed a clear preference for the SLC complex to bind during S phase compared with G1 phase (Supplemental Fig. S2C). However, the most visually inspected non-histone targets did not exhibit these changes in SLC binding during the cell cycle. These differences were not due to changes in the cellular levels of SLC during the cell cycle (Supplemental Fig. S2B).

Although we cannot rule out that similar oscillations may also occur at some nonhistone genes, we did not observe major changes in SLC occupancy in the vicinity of the top nonhistone target genes (for example, see *PWWP2A* in Supplemental Fig. S2D). To confirm our genome-wide observation and obtain more quantitative results, we performed ChIP-quantitative real-time PCR (qPCR) on a subset of histone loci (Fig. 5C). The repetitive nature of these clusters prevented a more extensive qPCR-based validation, as only a few primer sets yielded unique products (data not shown). In addition to confirming the genome-wide observation that the SLC complex is preferentially bound to histone loci during S phase, these data also showed an inverse trend for Pol II, which was mostly enriched during early S phase and gradually decreased thereafter (Fig. 5C), suggesting that the simultaneous

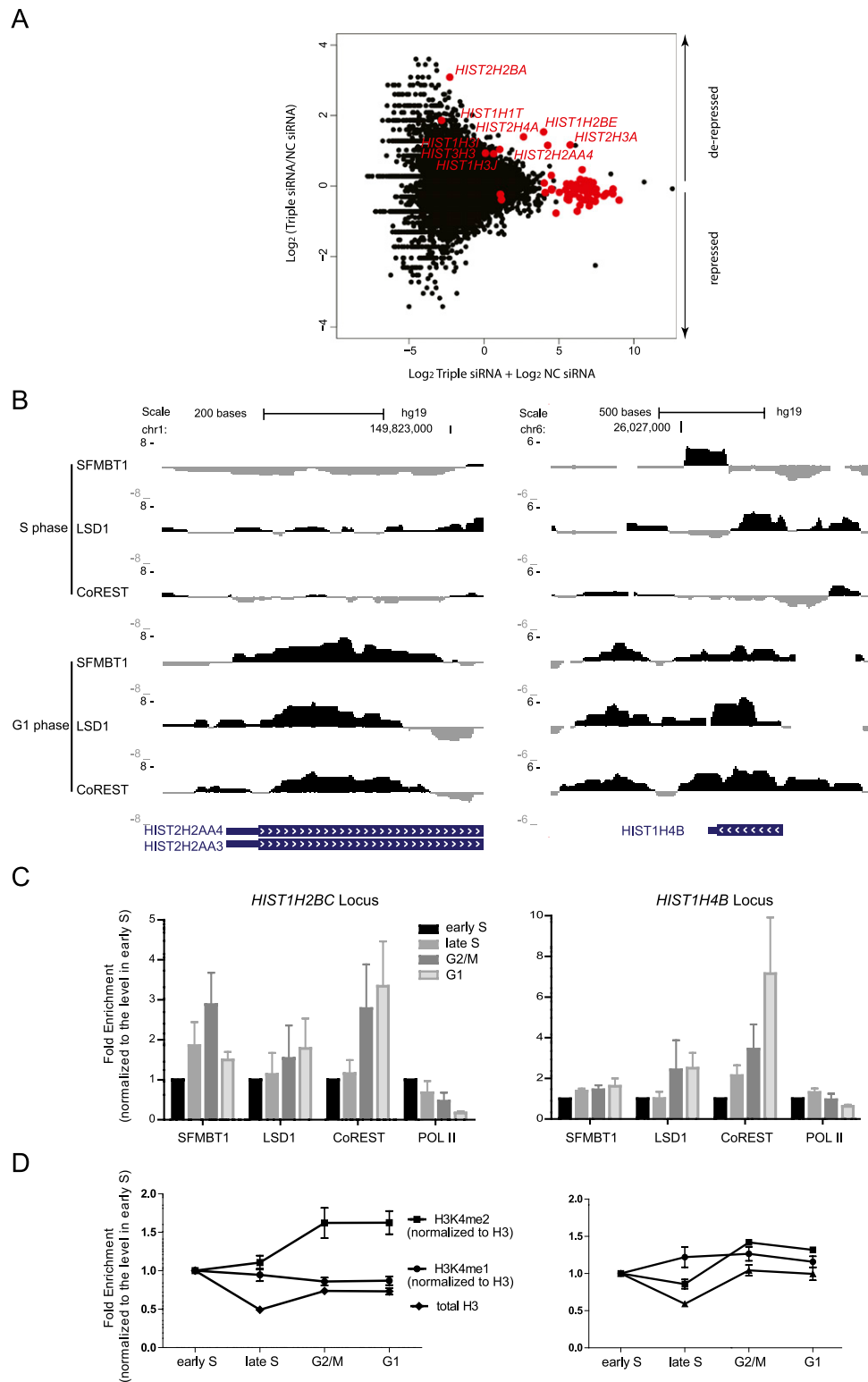


Figure 5. The SLC complex dynamically regulates histone genes. *(A)* M-A plot of changes in protein-coding gene expression comparing SFMBT1–LSD1–CoREST knockdown (Triple siRNA) and negative control siRNA knockdown (NC siRNA). Histone genes located in the replication-dependent loci are colored red. The Y-axis corresponds to \log_2 (expression level in Triple siRNA/NC siRNA). The X-axis corresponds to the \log_2 of the mean expression in both samples (see the Materials and Methods). All expression values were calculated in reads per kilobase per million (RPKM) by mapping RNA-seq reads to the gene models in ENSEMBL version 65. *(B)* Read density profile for SFMBT1, LSD1, and CoREST at *HIST1H2AA4/3* and *HIST1H4B* in two cell cycle stages: S and G1. The X-axis corresponds to genomic location, and the Y-axis corresponds to normalized ChIP-seq signal density (reads per 10 million) of each antibody after subtracting IgG signal. *(C)* Dynamic regulation of the enrichment of Pol II and the SLC complex components on histone genes during different stages of the cell cycle. The Y-axis corresponds to fold enrichment of ChIP for the indicated factors normalized to IgG and then normalized again to the enrichment level measured during early S phase. Bars represent the mean of four experiments + SEM. *(D)* ChIP enrichment of total H3, H3K4me1, and H3K4me2 at the *HIST1H2BC* locus (*left*) and the *HIST1H4B* locus (*right*) during different stages of the cell cycle. The Y-axis corresponds to fold enrichment of ChIP for the indicated factors (histone marks are normalized to total H3 levels first) as normalized to the enrichment level measured during early S phase. Bars represent the mean of four experiments + SEM.

occupancy observed in the asynchronous population is the result of superimposed profiles from cells in S phase, when these genes are active and Pol II predominates, and cells in other phases of the cell cycle, when these genes are repressed and bound by the SLC complex.

We also analyzed the histone marks on these histone loci through the cell cycle. Total H3 levels decreased through the progression of S phase and then increased in G2/M and G1 phase to reach a level comparable with that measured in early S phase (Fig. 5D). After normalization to histone H3 levels, we observed an increased density of H3K4me2 during G2/M and G1 phase, when the SLC complex occupied these histone loci; however, no increase or a slight decrease of H3K4me1 was found in G2/M and G1 compared with late S phase (Fig. 5D). This is consistent with our findings that the presence of SFMBT1, LSD1, and CoREST correlates with the appearance of H3K4me2 (Fig. 2D) and that SLC complex targets contain more H3K4me2 and less H3K4me1 than LSD1–CoREST targets that are not bound by SFMBT1 (Fig. 2E).

The SLC complex is abundant in mouse testes and associates with meiotic chromatin at the pachytene stage

To better understand the physiological role of SFMBT proteins in the context of the whole organism, we analyzed a panel of mouse tissues and found that both *Sfmbt1* and *Sfmbt2* mRNA exhibited the highest level of expression in mouse testes (Fig. 6A). Consistent with this observation, the SFMBT1 protein was enriched in whole lysates of mouse testes (Fig. 6A, bottom) and was detected in isolated and fractionated germ cells, with a higher expression in spermatocytes than in spermatids (Fig. 6B). Compared with several other MBT domain-containing proteins, SCMH1, a mouse homolog of the MBT and *Polycomb* protein SCM, also appeared to be elevated in germ cells. This finding is consistent with a previous study demonstrating that *Scmh1*^{-/-} mice display defects in gametogenesis (Takada et al. 2007).

We next examined whether SFMBT1 in mouse testes lysate was stably associated with the same set of polypeptides that we identified in human cell lines. Two independent SFMBT1 antibody immunoprecipitations were performed and subjected to mass spectrometry for polypeptide identification. Although the presence of CoREST was likely masked by the abundant band of the immunoglobulin heavy chain (which has a similar molecular weight), we detected peptides from LSD1 in both immunoprecipitations (Fig. 6C). To confirm that the SLC is a stable complex in primary tissues, we fractionated mouse testis lysate on a glycerol gradient and found that SFMBT1, LSD1, and CoREST sedimented with very similar profiles (Fig. 6D).

We next attempted to ascertain the regulation and biological function of SFMBT1 and LSD1 during spermatogenesis. Because spermatogenesis is carried out in the mouse testes after birth, postnatal development (PND) experiments were performed and showed that the expression levels of SFMBT1, SFMBT2, LSD1, and CoREST peaked

during days 14–20 (Supplemental Fig. S3A), when mid-prophase I stage spermatocytes appear (Cohen et al. 2006). A considerable increase in SFMBT1 protein expression was also detected within the same time frame (Supplemental Fig. S3B).

Consistent with this regulated expression, immunofluorescence staining revealed that SFMBT1 and LSD1 were enriched at meiotic chromosomes in pachytene stage spermatocytes (Fig. 6E). Specifically, SFMBT1 started to accumulate at synaptic regions at the zygotene stage (Fig. 6E, left panel, red arrows). At the pachytene stage, SFMBT1 was bound to all synapsed autosomes and the pseudoautosomal region (PAR) of the X and Y chromosomes (sex body) (Fig. 6E, left panel). Interestingly, SYCP1, a commonly used spermatocyte marker identified in the SFMBT1 affinity-purified fraction from testes lysate (Fig. 6C), marks synapses between homologous chromosomes and is located at the synaptic regions as well (Cohen and Pollard 2001), suggesting that the colocalization between these two proteins observed by microscopy may be due to their interaction in vivo. LSD1 staining showed similar meiotic chromosomal association, although it strongly bound both synapsed and unsynapsed regions of the X and Y chromosomes at the pachytene stage (Fig. 6E, middle panel), which correlated with a depletion of the H3K4me2 mark at this stage (Fig. 6E, right panel). The differential staining of SFMBT1 and LSD1 at the sex chromosomes suggests that although LSD1 is stably associated with SFMBT1 in mouse testes (Fig. 6C), it likely engages in additional roles that are independent of SFMBT1, which is consistent with the existence of LSD1–CoREST targets that do not contain SFMBT1 in human cell lines (see above).

The association of SLC components to meiotic chromosomes prompted us to identify genomic targets of SFMBT1 and LSD1 at different stages of spermatogenesis. SFMBT1, LSD1, and Pol II ChIP-seq were performed in pachytene stage spermatocytes and spermatids. Consistent with their expression levels, both SFMBT1 and LSD1 were highly enriched at ERs identified from spermatocytes but barely detectable in spermatids (Fig. 6F). Significant Pol II enrichment at many loci in both germ cell populations (Fig. 6F; data not shown) confirmed that this difference in chromatin binding of SFMBT1 was not a consequence of spermatid chromatin being less accessible. This result confirmed that our ChIP-seq data faithfully recapitulated the dynamic binding and targeting of SFMBT1 during spermatogenesis and that the regulation of SFMBT1 expression in germ cells correlates with its binding to chromatin.

Seventy-five percent of all SFMBT1 ERs identified in pachytene spermatocytes (1596 out of 2134) were also occupied by LSD1 (Supplemental Fig. S3C). The target genes shared by SFMBT1 and LSD1 in pachytene spermatocytes were enriched for GO terms related to RNA binding and processing, such as “ncRNA metabolic processes” (Supplemental Fig. S3D). Finally, we found that some histone genes were bound by both SFMBT1 and LSD1 in pachytene spermatocytes (Fig. 6F, right panel), which suggests that transcriptional regulation of histone

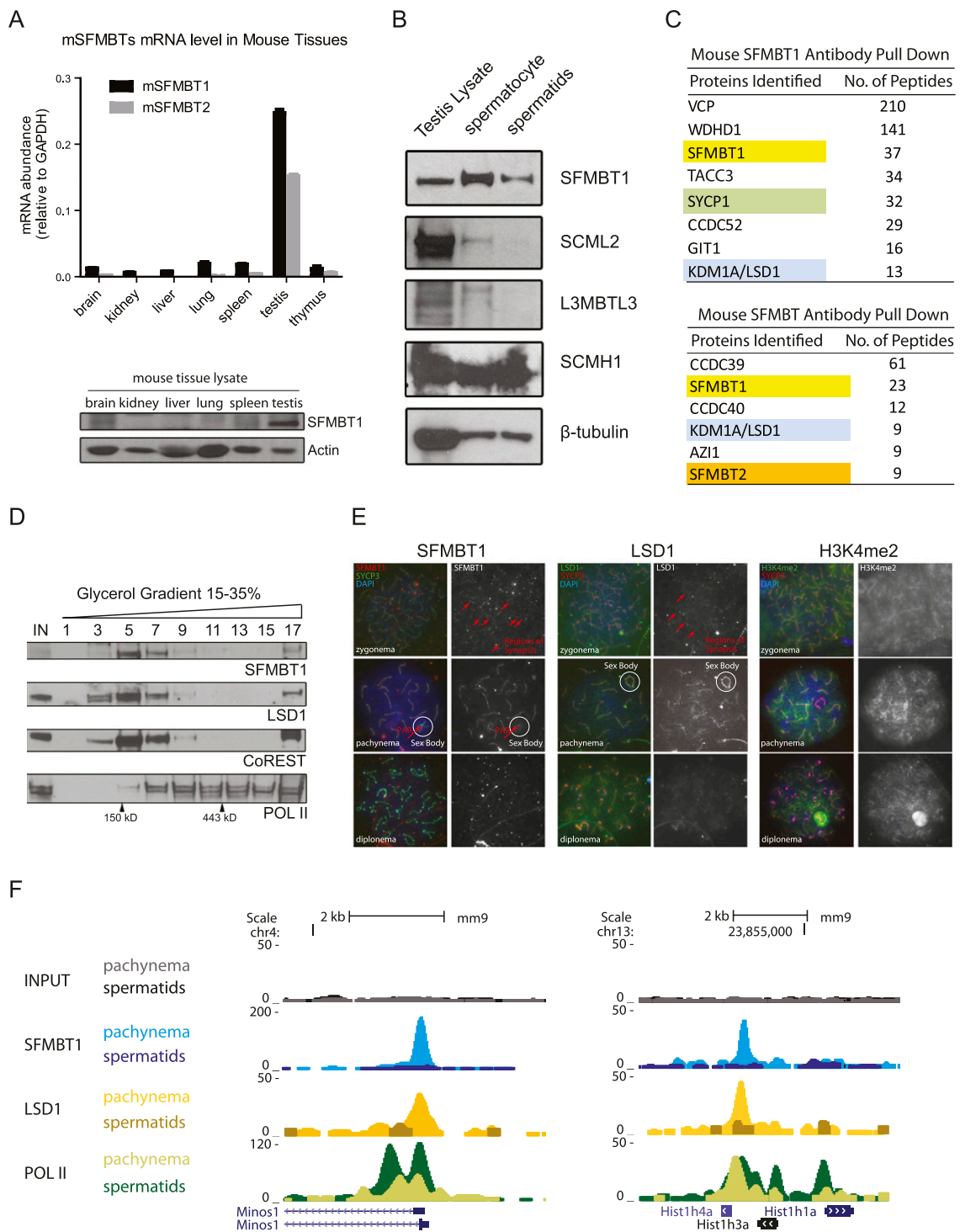


Figure 6. The SLC complex is abundant in mouse testes and associates with meiotic chromatin at the pachyene stage. (A) SFMBT1 is highly expressed in mouse testis compared with several other mature tissues, as analyzed in RT-qPCR (*top* panel) and Western blotting (*bottom* panel). Bars represent the percentage of mRNA abundance compared with *GAPDH* + SEM. The same amount of lysate for each tissue was loaded for Western blotting. (B) Western blotting analysis of protein expression levels in testis whole-cell lysate, spermatocyte, and spermatid lysate. (C) Mass spectrometric analysis of affinity-purified proteins from testis lysate. Mouse SFMBT1 antibody specifically recognizes SFMBT1; mouse SFMBT2 antibody recognizes both SFMBT1 and SFMBT2. The identified polypeptides and total peptide number are listed. Bait proteins SFMBT1 and SFMBT2 are colored yellow and orange, respectively. Two associated proteins—LSD1 and SYCP1—are highlighted in blue and green, respectively. (D) SFMBT1, LSD1, and CoREST form a stable complex in mouse testes, as they peak in the same fractions when mouse testes lysate is subjected to a 15%–35% glycerol gradient. Two reference proteins with masses of 150 kDa and 443 kDa were also analyzed using the same gradient, and their peak fractions are marked with black arrowheads. (E) Immunofluorescence staining of meiotic prophase I chromosome spreads. SFMBT1, LSD1, and H3K4me2 staining are arranged on the *left*, *middle*, and *right* panels, respectively. The *left* side of each shows overlaid staining of SFMBT1/LSD1/H3K4me2, SYCP3, and DAPI (fluorescent dye color indicated as the color in the legend). The *right* side shows single-channel staining of SFMBT1/LSD1/H3K4me2. (F) Superimposed read density profile for input (gray), SFMBT1 (blue), LSD1 (gold), and Pol II (green) in pachyene stage spermatocytes (lighter color) and in spermatids (darker color). The X-axis corresponds to genomic location, and the Y-axis corresponds to normalized ChIP-seq signal density. The position of RefSeq genes is shown at the *bottom*.

genes is a widespread and conserved function of the SLC complex and not a peculiarity of transformed cell lines.

Discussion

SFMBT1- and SFMBT2-containing complexes

Here, we present a comprehensive biochemical characterization of two mammalian homologs of the *Drosophila* PcG protein dSfmbt: SFMBT1 and SFMBT2.

Drosophila Sfmbt stably associates with Pho, forming the PhoRC complex (Klymenko et al. 2006), which has no clear counterpart in mammals. The existence of this complex led to a model in which Pho provides sequence-specific DNA-binding activity, while Sfmbt provides mono- and dimethylated histone tail-binding activity, and together these functions were proposed to target *Polycomb*-specific genes, followed by the recruitment of other *Polycomb* complexes like PRC1 and PRC2 (Muller and Kassis 2006; Schwartz and Pirrotta 2008; Simon and Kingston 2009). To understand the functions of mammalian SFMBT1 and SFMBT2, we sought to identify polypeptides associated with these proteins. Our results demonstrated that SFMBT1 specifically associated with components of the LSD1/CoREST histone demethylase complex and that SFMBT2 bound different polypeptides, including CASZ1, a multiple zinc finger domain-containing protein (Fig. 1). Both SFMBT1 and SFMBT2 associated with the PRC1.4 complex, the closest homolog of *Drosophila* PRC1 (Gao et al. 2012). The SFMBT2–CASZ1 association is reminiscent of the *Drosophila* PhoRC complex in several aspects: First, both complexes contain a zinc finger protein (Pho and CASZ1) and a histone tail binder (dSfmbt and SFMBT2). Second, both complexes interact with PRC1. Third, the PhoRC–PRC1 association is important for the proposed recruitment mechanism. Based on these similarities, we speculate that CASZ1 may bind DNA in a sequence-specific fashion, similar to Pho, and that therefore the CASZ1–SFMBT2 complex may be the functional homolog of the dPhoRC complex.

Several SAM/SPM domain-containing proteins were identified in preparations of purified SFMBT1 and SFMBT2. Although SAM/SPM domains participate in protein–protein interactions (Peterson et al. 1997; Kim et al. 2005), we did not recover all SAM/SPM proteins. The high representation of MBT proteins in this group suggested that these associations were specific. It will be interesting to ascertain whether mammalian MBT proteins can function collaboratively in certain cellular processes despite their distinct binding partners and presumably nonredundant functions (Bonasio et al. 2010; Trojer et al. 2011). Most importantly, the identification of polypeptides associated with SFMBT1 provided an inroad to the functional characterization of this protein.

Genomic targets and function of the SLC complex in somatic cells

Our results determined that SFMBT1, LSD1, and CoREST formed a biochemically defined complex (SLC complex) in somatic cells and mouse testis and that these proteins

exhibited a high degree of overlap in their genome-wide distribution, consistent with their joint functioning on chromatin (Fig. 2A, Supplemental Fig. S3C). The integrity of this functional module was confirmed by its formation upon recruitment of GAL4-SFMBT1 and by the fact that depletion of SFMBT1 by RNAi caused a decrease in LSD1 and CoREST occupancy at SLC target genes (Fig. 3E).

Although genome-wide targets of LSD1 have been identified in several other cell lines (Garcia-Bassets et al. 2007; Tsai et al. 2010; Whyte et al. 2012), the biochemical identification of the SLC complex prompted us to focus on a specific feature of LSD1 function when associated with SFMBT1 and CoREST. Specifically, we found that the SLC complex bound to many replication-dependent histone genes (Fig. 2B).

We assessed the function of SFMBT1 through gain of function using an inducible GAL4-SFMBT1 expression system and through loss of function via RNAi. When targeted to the luciferase locus, GAL4-SFMBT1 induced changes on chromatin reflective of a more repressed, compacted chromatin state. Our results indicate that the biochemical function of SFMBT1 might be similar to that of other MBT proteins, including L3MBTL1 and L3MBTL2, which compact chromatin arrays in vitro, although by different mechanisms (Trojer et al. 2007, 2011). In all cases studied to date, chromatin compaction and transcriptional repression require the MBT domain, and SFMBT1 is no exception. Moreover, a truncation of SFMBT1 containing only the MBT domain was sufficient for transcriptional repression, chromatin compaction, and LSD1 recruitment (Figs. 3, 4). Finally, the chromatin-binding ability of the MBT domain is essential, since point mutations that abolish or reduce binding to histone tails were functionally defective (Fig. 4).

Our results pinpoint the role of SFMBT1 in regulating the dynamic transcription of canonical histone genes through the cell cycle. The occupancy of the SLC complex fluctuated in opposition to that of Pol II occupancy at the histone loci tested, suggesting that the SLC complex plays a repressive role and may need to assemble quickly during the cell cycle to counteract the intense transcription of these histone genes during S phase (Fig. 7A). Consistent with this observation, knockdown of SLC complex components caused the derepression of several histone genes. Previous studies showed that MBT family members associate with E2F/Rb complexes and participate in cell cycle-related cellular processes (Harrison et al. 2007; Lu et al. 2007; Trojer et al. 2007, 2011). For example, *Drosophila* *L(3)mbt* and *Sfmbt* were identified in an RNAi screen as chromatin regulators of E2F repression (Lu et al. 2007), and L3MBTL2 associates with E2F6 and binds to E2F6 gene targets (Trojer et al. 2011). Together with these observations, our results suggest a broad role for MBT proteins in orchestrating chromatin and transcriptional changes during the cell cycle.

MBT domains bind mono- and dimethylated histone tails, and several studies have analyzed their specificity toward different lysine residues (Bonasio et al. 2010). SFMBT1 and SFMBT2 lack the aspartic acid and the aromatic cage at conserved positions (comprising residues

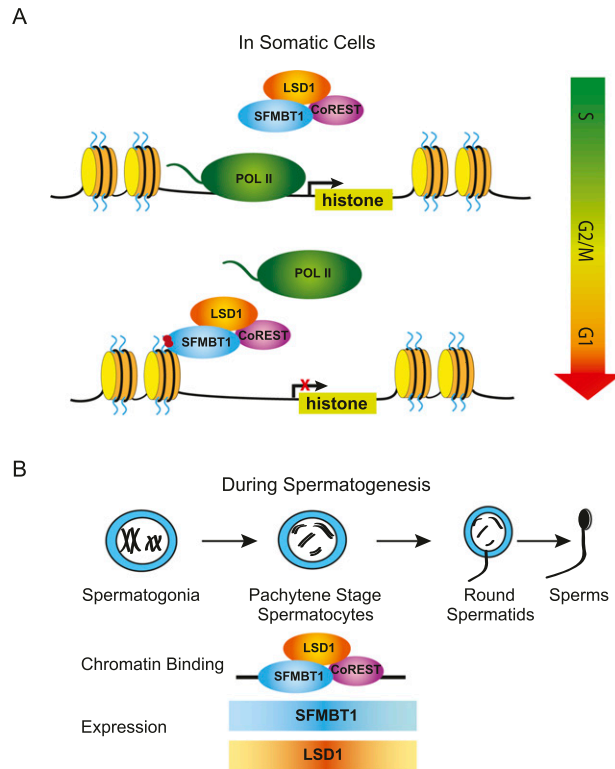


Figure 7. Dynamic regulation of the SLC complex during the cell cycle. (A) Schematic depiction of the regulation of the SLC complex through the cell cycle. During S phase, when the replication-dependent histone genes are transcribed, Pol II is highly associated with these loci, while the SLC complex does not bind (Fig. 5C). When S phase is completed and expression of histone genes is turned off, the SLC complex is recruited to the histone gene promoters, possibly through binding of the MBT domains of SFMBT1 to H3K4me2 marks, and precludes binding of Pol II, thus repressing transcription of these loci. (B) The gene expression and chromatin-binding activity of SFMBT1 and LSD1 are dynamically regulated during spermatogenesis. Four representative stages of spermatogenesis are shown at the top, with black lines representing the status of chromatin: replicated (spermatogonia), synapsed (pachytene stage spermatocytes), and segregated (spermatids). Both SFMBT1 and LSD1 associate with meiotic chromosomes in pachynema (Fig. 6E,F), suggesting critical functions of the complex at this stage of meiosis. The cyan (SFMBT1) and orange (LSD1) gradients indicate that the expression levels of these proteins are regulated during meiosis and peak at the pachytene stage (Fig. 6B; Supplemental Fig. S3A,B).

F-W-Y) that are essential for methylated lysine binding (Supplemental Fig. S4, highlighted yellow; Wang et al. 2003; H Li et al. 2007; Min et al. 2007). Instead, a few aromatic residues are scattered in the second MBT domain of SFMBT1 (Fig. 4A; Supplemental Fig. S4), suggesting the possibility of a wider "aromatic cage" or of a binding module that is somewhat distinct from that of other MBT proteins. In an initial screening of histone peptide array binding, SFMBT1 bound to H3K27me3 peptide, but this association was not confirmed using isothermal titration calorimetry (ITC) or electron microscopy (EM) (data not shown). Our peptide pull-down results indicated that

SFMBT1 bound to various unmethylated or methylated histone N-terminal tails *in vitro*. How, then, is the SLC complex recruited in the physiological setting? We did find that among the various methyl marks, the ChIP-seq signal for SFMBT1, LSD1, and CoREST strongly correlated with that of the H3K4me2 mark, a substrate of LSD1 demethylase activity (Shi et al. 2004), both spatially across the genome (Fig. 2D,E) and temporally during the cell cycle (Fig. 5C,D). Based on the findings that SFMBT1 recruited LSD1 and CoREST to gene targets and that the MBT domains are important for this recruitment (Fig. 3), we propose that the MBT domains of SFMBT1 serve as a docking module, possibly through binding to H3K4me2, to tether the SLC complex to genomic targets (Fig. 7A). A similar situation was observed in the case of L3MBTL1, which binds to several mono- and dimethylated histone tails *in vitro* yet is specifically recruited by H4K20me1 *in vivo* (Kalakonda et al. 2008). This suggests that the true specificity of L3MBTL1 and SFMBT1 (and possibly other MBT proteins) can only be observed *in vivo*, possibly because it requires the contribution of other histone PTMs (Ruthenburg et al. 2007), higher-order chromatin structure (Li and Reinberg 2011), chromatin-associated proteins, or noncoding RNAs (Wang and Chang 2011). Moreover, SFMBT1 caused a decrease in H3K4me1 (but not H3K4me2) levels at sites that were also bound by LSD1 and CoREST (Fig. 2E). In light of the recent finding that LSD1 functions in decommissioning of enhancers (Whyte et al. 2012), it will be interesting to test whether SFMBT1 specifically enhances the demethylation activity of LSD1 toward H3K4me1, thus facilitating its enhancer-decommissioning function.

Function of SFMBT1 during mammalian spermatogenesis

The MBT domain was originally identified within a gene, *L(3)mbt*, whose loss of function causes neoplastic transformation in *Drosophila* brains (Bonasio et al. 2010) and derepression of germline-specific genes (Janic et al. 2010; Meier et al. 2012). Moreover, mice bearing a homozygous truncation of the SPM domain of *Scmh1* exhibited defects in spermatogenesis and failed to exclude *Polycomb* complexes from the XY body (Takada et al. 2007). These studies have suggested that MBT proteins play important roles during germ cell development through transcriptional repression and association with other *Polycomb* complexes.

Gene expression analyses in a panel of murine tissues revealed that both SFMBT1 and SFMBT2 are highly enriched in mouse testis. SFMBT2 displayed higher expression levels in several embryonic tissues, including placenta (Kuzmin et al. 2008; data not shown), but it was not further analyzed in this study. Importantly, LSD1 associated with SFMBT1 in two independent affinity purifications from mouse testis (Fig. 6C), and components of the SLC complex remain associated during sedimentation on a glycerol gradient (Fig. 6D), indicating the presence of a stable SLC complex in this primary tissue. It is worth noting that the homolog of LSD1 in *Caenorhabditis elegans* is

essential for transmission of the H3K4 modification in germ cells (Katz et al. 2009) and that protein levels of LSD1 were inversely correlated with those of H3K4me2 during mouse spermatogenesis (Fig. 6E; Godmann et al. 2007). These observations suggest that LSD1 demethylation activity might be important for mammalian spermatogenesis as well.

Consistent with the association of LSD1 and SFMBT1 with meiotic chromatin in a stage-specific manner, 75% of SFMBT1 targets were also occupied by LSD1 in pachytene stage spermatocytes (Supplemental Fig. S3C), suggesting that the SLC complex has important roles in the regulation of transcription in not only transformed human somatic cell lines, but also primary mouse tissues. Given that more LSD1 targets were identified (7292 ERs) compared with SFMBT1 (Supplemental Fig. S3C), LSD1 must also function in an SFMBT1-independent manner in germ cells, a conclusion further supported by differences in the distribution of SFMBT1 and LSD1 in prophase chromosomes, especially on the sex body (Fig. 6E).

We also observed overlap in the gene targeting and functions of SFMBT1 and LSD1 between somatic and germ cells. Although the measured occupancy of the SLC complex at replication-dependent histone loci was not as high as in HeLa S3 cells, we clearly detected both SFMBT1 and LSD1 at some of these histone genes in spermatocytes. Given that canonical histones must be replaced by testis-specific histones in pachytene spermatocyte, it is possible that the SLC complex participates in the repression of the former to facilitate deposition of the latter. Furthermore, SFMBT1 and LSD1 occupancy appeared to be anti-correlated with Pol II in different stages of meiotic germ cells, which was also observed with other SFMBT1–LSD1 gene targets in germ cells (Fig. 6F; data not shown).

In conclusion, our results show that the human *Polycomb* protein SFMBT1 functions in concert with LSD1 and CoREST to regulate transcriptional output, at least in part via chromatin compaction. This repressive system is deployed during the cell cycle to control in part the expression of replication-dependent histone genes and during mouse spermatogenesis. Thus, the SLC complex is a new member of the ever-growing family of chromatin-associated protein complexes that modulates chromatin function in somatic and germ cells (Fig. 7).

Materials and methods

Plasmids

SFMBT1, SFMBT2, LSD1, and CoREST were cloned from human cDNA, with truncations and mutant constructs generated by PCR. SFMBT1 and SFMBT2 were cloned in pcDNA4/TO-N-GAL4 (Vaquero et al. 2004), pINTO-C-HF (Gao et al. 2012), pAcHLT (BD Biosciences), pET30a, and pGEX6p1. LSD1 and CoREST were cloned in pFastBac (Invitrogen).

Cell culture, cell transfection, and infection

HeLa S3 (American Type Culture Collection) and 293T-REx (Invitrogen) cells were cultured in DMEM with 10% bovine

serum, 100 IU/mL penicillin, 100 µg/mL streptomycin, and 300 µg/mL L-glutamine. Empty vectors (pcDNA4-N-GAL4 and pINTO-C-HF) or plasmids for mammalian expression were transfected into 293T-REx cells using FuGENE 6 (Roche). Stable clones were selected with 500 µg/mL G418 (Sigma), 100 µg/mL Zeocin (Invitrogen), or 10 µg/mL blasticidin (InvivoGen) individually or in combinations, depending on the transfected plasmids. siRNAs were transfected into HeLa S3 cells using RNAiMAX (Invitrogen) according to the manufacturer's instructions. pGIPZ SFMBT1 lentiviral shRNA or control shRNA (Thermo Scientific) plasmids were cotransfected with packaging plasmids into 293 cells; lentiviruses were harvested, concentrated, and used for HeLa S3 cell infection; and cells were then cultured in medium with 5 µg/mL puromycin (Invitrogen) to select cells stably expressing shRNAs. For baculovirus infection, Sf9 cells were grown in SF900 III SFM (Gibco). Cells were harvested 72 h after infection for recombinant protein purification.

Cell synchronization and FACS analysis

HeLa S3 cells at 50% confluency were washed with PBS and incubated in medium containing 2 mM thymidine (Sigma) for 14 h. After the first thymidine block, cells were washed with PBS four times and cultured in fresh nonthymidine medium for 7 h, blocked again with 2 mM thymidine for 14 h, washed with PBS, and cultured in fresh medium for 1, 5, 9, or 13 h to be released in early S, late S, G2/M, or G1 phase. For FACS analysis, cells were washed with PBS, passed through needles to obtain single-cell suspensions, fixed with ice-cold 70% ethanol for 30 min, and stored at –20°C. Before FACS analysis, cells were washed with PBS and resuspended in PBS containing 1 µg/µL RNase A and 25 µg/mL propidium iodide (PI).

Affinity purification and gel filtration chromatography

For recombinant protein purification from baculovirus-infected Sf9 cells, cells were resuspended in BC buffer (20 mM Tris at pH 7.9, 4°C, 20% glycerol, 0.05 M KCl, 0.2 mM EDTA) containing 350 mM NaCl (BC350), 0.1% IGEAL CA-630, 0.2 mM PMSE, and 10 mM β-mercaptoethanol. After sonication, the cell lysate was centrifuged at 20,000g for 20 min, and the supernatant was used for affinity purification using the same buffer. In gel filtration chromatography experiments, recombinant SFMBT1, LSD1, and CoREST were mixed, loaded onto a Superose 6 PC 3.2/30 2.4-mL (GE Lifesciences) sizing column, and fractionated.

For TAP of SFMBT1- and SFMBT2-associated polypeptides, 293T-REx cells stably transfected with empty vector and SFMBT expression constructs were treated with 0.1 ng/mL dox for 24 h. Cells were then resuspended in buffer A (10 mM Tris at pH 7.9, 1.5 mM MgCl₂, 10 mM KCl, 0.5 mM DTT, 0.2 mM PMSE, 1 µg/mL aprotinin, leupeptin, and pepstatin A), lysed on ice for 30 min, and homogenized with dounce homogenizer A. Nuclei were isolated by centrifugation at 10,000g for 10 min and resuspended in buffer C (20 mM Tris at pH 7.9 at 4°C, 25% glycerol, 420 mM NaCl, 1.5 mM MgCl₂, 0.2 mM EDTA, 0.5 mM DTT, 0.2 mM PMSE, 1 µg/mL aprotinin, leupeptin, and pepstatin A). After 30 min of incubation at 4°C, the nuclear extract was obtained by centrifugation at 40,000g for 30 min and taking the supernatant. Nuclear extract was then incubated with Flag M2 beads (Sigma) and rotated overnight at 4°C. The beads were then washed with 5× BC300 containing 0.05% IGEAL CA-630 and eluted with 250 µg/mL Flag peptides (Sigma). The Flag eluate was then incubated with HA beads (Roche) for 4 h at 4°C. The HA beads were washed as described in the Flag purification and eluted with 0.1 M glycine (pH 2.5) at room temperature for 10 min or 500 µg/mL HA peptides overnight at 4°C. To identify

polypeptides, the final eluates were digested with proteases and subjected to tandem mass spectrometry analysis for peptide identification. The “number of peptides identified” in Figure 1A represents the number of peptides that were identified for each indicated protein.

Antibodies

Rabbit antibodies against human or mouse SFMBT proteins were generated using synthesized peptides or His-tagged recombinant proteins and affinity-purified with GST-tagged recombinant protein. Rabbit antibody against human LSD1 was generated using a His-tagged full-length LSD1 protein, which is used in ChIP-seq in HeLa S3 cells. The following antibodies were used: CASZ1 (Rockland, 600-401-B62), EZH2 (Kuzmichev et al. 2004), CoREST (Bethyl Laboratories, A300-130A), LSD1 (Bethyl Laboratories, A300-216A), L3MBTL3 (Bethyl Laboratories, A311-052A), GAL4 (Millipore, 06-262), HA (Covance, MMS-101P), Histone H3 (Abcam, ab1791), H3K4me1 (Abcam, ab8895), H3K4me2 (Millipore, 07-030), H3K9me3 (Abcam, ab8898), H3K27me2/3 (Kuzmichev et al. 2004), Pol II (Santa Cruz Biotechnology, sc-899 N-20), RING1B (Abcam, ab3832), SCMHI (Abcam, ab56908), and SYCP3 (Holloway et al. 2008).

Histone peptide pull-down

Twenty-five micromoles of biotinylated histone peptides (Anaspec) was incubated with 2.5 μ g of Flag-tagged SFMBT proteins (purified from Sf9 cells) (see above) in 500 μ L of BC350 containing 0.1% IGEPAL CA-630 and 0.2 mM PMSF for 2 h on a rotating wheel at 4°C. Ten microliters of streptavidin agarose beads (Millipore) was then added, incubated for 1 h, and washed with 500 μ L of peptide pull-down buffer five times. The pulled-down proteins were then loaded on SDS-PAGE and analyzed by Western blot.

Luciferase activity assay

293T-REx-luciferase (Vaquero et al. 2004) cells stably transfected with pcDNA4/TO-GAL4 or pcDNA4/TO-GAL4-SFMBT protein constructs were seeded at 10^5 cells per well in 24-well plates. Cells were treated with 1–10 ng/mL dox to obtain comparable protein expression (Fig. 3D, bottom). For luciferase assay, cells were lysed by adding 250 μ L of ice-cold lysis buffer (0.2% Triton X-100, 100 mM potassium phosphate at pH 7.8, 1 mM DTT) and shaking for 15 min at 4°C. The cell lysate was centrifuged at 20,000g for 10 min, and the supernatant was assayed for luciferase activity using luciferase assay substrate (Promega).

Chromatin accessibility assay

Chromatin accessibility was assayed using the EpiQ chromatin analysis kit (Bio-Rad, 172-5400) according to the manufacturer's instructions. In brief, cells were harvested and then lysed, and their genomic DNA was extracted using the provided reagents with or without nuclease treatment. The degree of chromatin accessibility in our experiments (exp) was measured in comparison with a reference (ref) compact locus *RHO* and an open control (con) locus *GAPDH*, according to the following formula:

$$1 - \frac{2^{\Delta Ct(ref)}}{2^{\Delta Ct(exp)}} \cdot \frac{1 - \frac{2^{\Delta Ct(ref)}}{2^{\Delta Ct(con)}}}{1 - \frac{2^{\Delta Ct(ref)}}{2^{\Delta Ct(con)}}}$$

$\Delta Ct = Ct(digested) - Ct(undigested)$. $\Delta Ct(digested)$ is the Ct value obtained from nuclease-digested genomic DNA, and

$\Delta Ct(undigested)$ is the Ct value obtained from samples without nuclease digestion.

ChIP assay, ChIP-seq, and RNA sequencing (RNA-seq) library construction

ChIP assays were performed as described previously (Takahashi et al. 2000; Gao et al. 2012). Chromatin-immunoprecipitated DNA was subjected to qPCR analysis and/or library construction for sequencing. ChIP-seq libraries were prepared according to the manufacturer's instructions (Illumina) and as described (Asp et al. 2011; Gao et al. 2012) with a few modifications. Briefly, immunoprecipitated DNA was end-repaired, A-tailed, and ligated to custom adapters with T4 ligase (Enzymatics, 603-HC-L). After quantification, libraries were sequenced at a concentration of 7 pM on Illumina Genome Analyzer IIx or of 10 pM on Illumina HiSeq. When HiSeq was used for sequencing, a custom barcoding system was used to sequence more than one sample per lane. For ChIP-qPCR, primers were designed surrounding ERs identified from ChIP-seq and at control regions. qPCRs were performed on the Stratagene Mx3005p platform using SYBR Green Master PCR mix (Invitrogen) according to the manufacturer's instructions. The percentage of input was calculated as 2 to the power of the difference in Ct value between the ChIP samples and the input chromatin.

RNA was isolated using TRIzol (Invitrogen). PolyA⁺ RNA was further isolated using Oligo(dT)₂₅ Dynabeads (Invitrogen). Ribosomal RNA was removed from the polyA⁺ flowthrough using the RiboMinus kit (Invitrogen). The polyA⁺ and Ribo⁻ fractions were converted to cDNA while preserving strand information using the dUTP method (Parkhomchuk et al. 2009). Libraries were constructed from cDNA with the same protocol as described above for ChIP-seq libraries.

High-throughput sequencing analysis

ChIP-seq analysis was performed as described previously with some modifications (Gao et al. 2012). Sequenced reads were mapped with Bowtie using parameter -v2 -m4 -b (Langmead et al. 2009). Significantly ($P < 0.01$) enriched regions (ERs) for each ChIP-seq data set were identified with QSESEQ (Asp et al. 2011; Micsinai et al. 2012) and filtered using stringent parameters (peak width ≥ 80 , total read counts ≥ 10 , and enrichment ≥ 2.5). GO and functional term association analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (Huang et al. 2009a,b) and the Genomic Regions Enrichment of Annotation Tool (GREAT) (McLean et al. 2010).

Sequenced reads from RNA-seq experiments were mapped to the reference genome with Bowtie using parameters -v2 -m40. Reads were assigned to genes using the DEGseq package (<http://www.bioconductor.org>), and an M-A plot using normalized reads per kilobase per million (RPKM) was generated in R.

The ChIP-seq data for H3K4me2 (GSM733734), H3K4me3 (GSM733682), H3K9ac (GSM733756), H3K27ac (GSM733684), H3K27me3 (GSM733696), H3K36me3 (GSM733711), H3K79me2 (GSM733669), H4K20me1 (GSM733689), and Pol II (GSM733759) were downloaded from the ENCODE Project. All other sequencing data have been deposited to the Gene Expression Omnibus as SuperSeries GSE45489.

Mouse tissue analysis and PND experiment

For antibody-mediated affinity purification, testicles were dissected from adult C57BL/6 mice, lysed in tissue lysis buffer (50 mM Tris at pH 7.4, 400 mM NaCl, 0.01% IGEPAL, 5 mM EDTA,

5 mM NaF, 1 mM DTT, 2 mM PMSF, 1 μ g/mL aprotinin, leupeptin and pepstatin A), and homogenized using Tissue Tearor. Samples were incubated for 15 min on ice and centrifuged at 20,000g for 10 min. Tissue lysate was obtained by taking the supernatant and carefully avoiding the fat layer on top.

In gene expression analysis experiments, mouse tissues were dissected from adult C57BL/6 mice and quickly ground in a mortar containing liquid nitrogen. RNA was then extracted from tissue powder. For the PND experiment, testes were removed from newborn C57BL/6 mice every other day from postnatal day 10 (P10) to P26 and from 6-wk-old adult mice. Testis tissue lysate was obtained as described above, or testis was homogenized in TRIzol (Invitrogen) to obtain RNA for RT-qPCR analysis.

Chromosome spreading and immunofluorescent staining

Testicles were removed and decapsulated from adult C57BL/6 mice and quickly incubated in hypotonic sucrose extraction buffer (HEB, containing 50 mM sucrose) for 30–60 min on ice. A small section of tubules was removed and placed in 22 μ L of sucrose bubble on a depression slide. Tubules were then minced and added to slides coated with 1% paraformaldehyde. The slides were dried slowly in a humidified chamber for 3 h, washed in PBS containing Photo-flo (Kodak, EMS), and processed for immunofluorescence staining. Briefly, slides were blocked with PBS containing either Photo-flo, 0.1% Triton X, or ADB, each for 10 min. They were stained with primary antibodies in moist hybridization chambers overnight at room temperature. The slides were blocked as described above, incubated with secondary antibodies for 1 h at 37°C, and washed with PBS containing Photo-flo. The sample slides were then air-dried and mounted with Prolong Gold Antifade (Invitrogen).

Mouse germ cell fractionation

Testes were removed from adult C57BL/6 mice, decapsulated, and incubated in Krebs buffer containing 2 mg/mL collagenase for 15 min in a 34°C shaking water bath. Cells were allowed to settle by standing for 5 min at room temperature and washed with Krebs buffer. Cells were resuspended in Krebs buffer containing 2.5 mg/mL trypsin and 2 μ g/mL DNase I, added to 0.5% BSA, and centrifuged at 500g for 2 min at 18°C. The cell pellet was washed in Krebs, resuspended in 0.5% BSA, and filtered through a Falcon 35-2350 cap filter. The isolated cells from mouse testis were then loaded on a sedimentation chamber and fractionated using a STA-PUT sucrose gradient isolation system (0.5%–4% sucrose solution). An aliquot of each fraction was observed under the microscope to verify germ cell morphology. The fractions that contained the same germ cell population were pooled and resuspended in PBS. For ChIP experiments, 1% formaldehyde was added to the germ cell fractions, which were then incubated on rotating wheels for 10 min at room temperature. Cross-linking was stopped by adding 0.125 M glycine and washing with cold PBS.

Acknowledgments

We are grateful for the generous help from Dr. Dinshaw Patel and Dr. Zhongjun Chen on structural analysis and quantitative binding, Dr. Steve Gambelin and Dr. Valeria De Marco on quantitative binding experiments, and Dr. Or Gozani and Dr. Dan Levy on peptide array experiments. We thank Dr. Brian Dynlacht and Dr. Patrik Asp for sharing their ChIP-seq protocols and helpful discussion on ChIP-seq experiments, and the mass spectrometry facility at Rutgers University and University of Medicine and Dentistry of New Jersey for polypeptide identification.

Dr. Yang Shi and Dr. Nima Mosammaparast kindly provided LSD1 antibodies. Dr. Prabhakara Reddi provided fractionated spermatocytes and spermatids for our initial analysis. CBX8 antibody was a kind gift from Dr. Kristian Helin. We appreciate Dr. Lynne Vales and Dr. Patrick Trojer for critical reading of our manuscript and helpful comments. This work was supported by grants from the National Institute of Health (GM-64844 and R37-37120) and the Howard Hughes Medical Institute to D.R. R.B. was supported by a Helen Hay Whitney Foundation post-doctoral fellowship and by the Helen L. and Martin S. Kimmel Center for Stem Cell Biology Post-Doctoral Fellow Award. F.S. was supported by the American-Italian Cancer Foundation Post-Doctoral Research Fellowship.

References

- Asp P, Blum R, Vethantham V, Parisi F, Micsinai M, Cheng J, Bowman C, Kluger Y, Dynlacht BD. 2011. Genome-wide remodeling of the epigenetic landscape during myogenic differentiation. *Proc Natl Acad Sci* **108**: E149–E158.
- Ballas N, Battaglioli E, Atouf F, Andres ME, Chenoweth J, Anderson ME, Burger C, Moniwa M, Davie JR, Bowers WJ, et al. 2001. Regulation of neuronal traits by a novel transcriptional complex. *Neuron* **31**: 353–365.
- Banaszynski LA, Allis CD, Lewis PW. 2010. Histone variants in metazoan development. *Dev Cell* **19**: 662–674.
- Beisel C, Paro R. 2011. Silencing chromatin: Comparing modes and mechanisms. *Nat Rev Genet* **12**: 123–135.
- Berger SL. 2007. The complex language of chromatin regulation during transcription. *Nature* **447**: 407–412.
- Bonasio R, Lecona E, Reinberg D. 2010. MBT domain proteins in development and disease. *Semin Cell Dev Biol* **21**: 221–230.
- Cai Y, Jin J, Yao T, Gottschalk AJ, Swanson SK, Wu S, Shi Y, Washburn MP, Florens L, Conaway RC, et al. 2007. YY1 functions with INO80 to activate transcription. *Nat Struct Mol Biol* **14**: 872–874.
- Campos EI, Reinberg D. 2009. Histones: Annotating chromatin. *Annu Rev Genet* **43**: 559–599.
- Christensen KL, Patrick AN, McCoy EL, Ford HL. 2008. The six family of homeobox genes in development and cancer. *Adv Cancer Res* **101**: 93–126.
- Cohen PE, Pollard JW. 2001. Regulation of meiotic recombination and prophase I progression in mammals. *Bioessays* **23**: 996–1009.
- Cohen PE, Pollack SE, Pollard JW. 2006. Genetic analysis of chromosome pairing, recombination, and cell cycle control during first meiotic prophase in mammals. *Endocr Rev* **27**: 398–426.
- Gao Z, Zhang J, Bonasio R, Strino F, Sawai A, Parisi F, Kluger Y, Reinberg D. 2012. PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* **45**: 344–356.
- Garcia-Bassets I, Kwon YS, Telese F, Prefontaine GG, Hutt KR, Cheng CS, Ju BG, Ohgi KA, Wang J, Escoubet-Lozach L, et al. 2007. Histone methylation-dependent mechanisms impose ligand dependency for gene activation by nuclear receptors. *Cell* **128**: 505–518.
- Godmann M, Auger V, Ferraroni-Aguiar V, Di Sauro A, Sette C, Behr R, Kimmins S. 2007. Dynamic regulation of histone H3 methylation at lysine 4 in mammalian spermatogenesis. *Biol Reprod* **77**: 754–764.
- Hakimi MA, Bochar DA, Chenoweth J, Lane WS, Mandel G, Shiekhhattar R. 2002. A core-BRAF35 complex containing histone deacetylase mediates repression of neuronal-specific genes. *Proc Natl Acad Sci* **99**: 7420–7425.
- Harrison MM, Lu X, Horvitz HR. 2007. LIN-61, one of two *Caenorhabditis elegans* malignant-brain-tumor-repeat-con-

- taining proteins, acts with the DRM and NuRD-like protein complexes in vulval development but not in certain other biological processes. *Genetics* **176**: 255–271.
- Holloway JK, Booth J, Edelman W, McGowan CH, Cohen PE. 2008. MUS81 generates a subset of MLH1–MLH3-independent crossovers in mammalian meiosis. *PLoS Genet* **4**: e1000186.
- Huang DW, Sherman BT, Lempicki RA. 2009a. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**: 1–13.
- Huang DW, Sherman BT, Lempicki RA. 2009b. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**: 44–57.
- Humphrey GW, Wang Y, Russanova VR, Hirai T, Qin J, Nakatani Y, Howard BH. 2001. Stable histone deacetylase complexes distinguished by the presence of SANT domain proteins CoREST/kiaa0071 and Mta-L1. *J Biol Chem* **276**: 6817–6824.
- Janic A, Mendizabal L, Llamazares S, Rossell D, Gonzalez C. 2010. Ectopic expression of germline genes drives malignant brain tumor growth in *Drosophila*. *Science* **330**: 1824–1827.
- Kalakonda N, Fischle W, Boccuni P, Gurvich N, Hoya-Arias R, Zhao X, Miyata Y, Macgrogan D, Zhang J, Sims JK, et al. 2008. Histone H4 lysine 20 monomethylation promotes transcriptional repression by L3MBTL1. *Oncogene* **27**: 4293–4304.
- Katz DJ, Edwards TM, Reinke V, Kelly WG. 2009. A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**: 308–320.
- Kim CA, Sawaya MR, Cascio D, Kim W, Bowie JU. 2005. Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. *J Biol Chem* **280**: 27769–27775.
- Kimmins S, Sassone-Corsi P. 2005. Chromatin remodelling and epigenetic features of germ cells. *Nature* **434**: 583–589.
- Klymenko T, Papp B, Fischle W, Kocher T, Schelder M, Fritsch C, Wild B, Wilm M, Muller J. 2006. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* **20**: 1110–1122.
- Kouzarides T. 2007. Chromatin modifications and their function. *Cell* **128**: 693–705.
- Kuzmichev A, Jenuwein T, Tempst P, Reinberg D. 2004. Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol Cell* **14**: 183–193.
- Kuzmin A, Han Z, Golding MC, Mann MR, Latham KE, Varmuza S. 2008. The PcG gene *Sfmbt2* is paternally expressed in extraembryonic tissues. *Gene Expr Patterns* **8**: 107–116.
- Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**: R25.
- Li G, Reinberg D. 2011. Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* **21**: 175–186.
- Li B, Carey M, Workman JL. 2007. The role of chromatin during transcription. *Cell* **128**: 707–719.
- Li H, Fischle W, Wang W, Duncan EM, Liang L, Murakami-Ishibe S, Allis CD, Patel DJ. 2007. Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol Cell* **28**: 677–691.
- Lu J, Ruhf ML, Perrimon N, Leder P. 2007. A genome-wide RNA interference screen identifies putative chromatin regulators essential for E2F repression. *Proc Natl Acad Sci* **104**: 9381–9386.
- Marzluff WF, Gongidi P, Woods KR, Jin J, Maltais LJ. 2002. The human and mouse replication-dependent histone genes. *Genomics* **80**: 487–498.
- Marzluff WF, Wagner EJ, Duronio RJ. 2008. Metabolism and regulation of canonical histone mRNAs: Life without a poly(A) tail. *Nat Rev Genet* **9**: 843–854.
- Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. 2003. The Tudor domain ‘Royal Family’: Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* **28**: 69–74.
- McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**: 495–501.
- Meier K, Mathieu EL, Finkernagel F, Reuter LM, Scharfe M, Doehlemann G, Jarek M, Brehm A. 2012. LINT, a novel dL(3)mbt-containing complex, represses malignant brain tumour signature genes. *PLoS Genet* **8**: e1002676.
- Micsinai M, Parisi F, Strino F, Asp P, Dynlacht BD, Kluger Y. 2012. Picking ChIP-seq peak detectors for analyzing chromatin modification experiments. *Nucleic Acids Res* **40**: e70.
- Min J, Allali-Hassani A, Nady N, Qi C, Ouyang H, Liu Y, MacKenzie F, Vedadi M, Arrowsmith CH. 2007. L3MBTL1 recognition of mono- and dimethylated histones. *Nat Struct Mol Biol* **14**: 1229–1230.
- Muller J, Kassis JA. 2006. Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr Opin Genet Dev* **16**: 476–484.
- Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. 2002. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* **296**: 1132–1136.
- Osley MA. 1991. The regulation of histone synthesis in the cell cycle. *Annu Rev Biochem* **60**: 827–861.
- Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, Lehrach H, Soldatov A. 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Res* **37**: e123.
- Peterson AJ, Kyba M, Bornemann D, Morgan K, Brock HW, Simon J. 1997. A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. *Mol Cell Biol* **17**: 6683–6692.
- Ruthenburg AJ, Li H, Patel DJ, Allis CD. 2007. Multivalent engagement of chromatin modifications by linked binding modules. *Nat Rev Mol Cell Biol* **8**: 983–994.
- Sarma K, Reinberg D. 2005. Histone variants meet their match. *Nat Rev Mol Cell Biol* **6**: 139–149.
- Sathyamurthy A, Allen MD, Murzin AG, Bycroft M. 2003. Crystal structure of the malignant brain tumor (MBT) repeats in Sex Comb on Midleg-like 2 (SCML2). *J Biol Chem* **278**: 46968–46973.
- Schwartz YB, Pirrotta V. 2008. Polycomb complexes and epigenetic states. *Curr Opin Cell Biol* **20**: 266–273.
- Shi Y, Lan F, Matson C, Mulligan P, Whetstone JR, Cole PA, Casero RA. 2004. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**: 941–953.
- Simon JA, Kingston RE. 2009. Mechanisms of polycomb gene silencing: Knowns and unknowns. *Nat Rev Mol Cell Biol* **10**: 697–708.
- Spitz F, Furlong EE. 2012. Transcription factors: From enhancer binding to developmental control. *Nat Rev Genet* **13**: 613–626.
- Stein GS, van Wijnen AJ, Stein JL, Lian JB, Montecino M, Zaidi SK, Braastad C. 2006. An architectural perspective of cell-cycle control at the G1/S phase cell-cycle transition. *J Cell Physiol* **209**: 706–710.
- Takada Y, Isono K, Shinga J, Turner JM, Kitamura H, Ohara O, Watanabe G, Singh PB, Kamijo T, Jenuwein T, et al. 2007. Mammalian Polycomb Scmh1 mediates exclusion of Polycomb complexes from the XY body in the pachytene spermatocytes. *Development* **134**: 579–590.
- Takahashi Y, Rayman JB, Dynlacht BD. 2000. Analysis of promoter binding by the E2F and pRB families in vivo:

- Distinct E2F proteins mediate activation and repression. *Genes Dev* **14**: 804–816.
- Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. 2007. How chromatin-binding modules interpret histone modifications: Lessons from professional pocket pickers. *Nat Struct Mol Biol* **14**: 1025–1040.
- Trojer P, Li G, Sims RJ III, Vaquero A, Kalakonda N, Boccuni P, Lee D, Erdjument-Bromage H, Tempst P, Nimer SD, et al. 2007. L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell* **129**: 915–928.
- Trojer P, Cao AR, Gao Z, Li Y, Zhang J, Xu X, Li G, Losson R, Erdjument-Bromage H, Tempst P, et al. 2011. L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. *Mol Cell* **42**: 438–450.
- Tsai MC, Manor O, Wan Y, Mosammamaparast N, Wang JK, Lan F, Shi Y, Segal E, Chang HY. 2010. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**: 689–693.
- Vaquero A, Scher M, Lee D, Erdjument-Bromage H, Tempst P, Reinberg D. 2004. Human SirT1 interacts with histone H1 and promotes formation of facultative heterochromatin. *Mol Cell* **16**: 93–105.
- Wang KC, Chang HY. 2011. Molecular mechanisms of long noncoding RNAs. *Mol Cell* **43**: 904–914.
- Wang WK, Tereshko V, Boccuni P, MacGrogan D, Nimer SD, Patel DJ. 2003. Malignant brain tumor repeats: A three-leaved propeller architecture with ligand/peptide binding pockets. *Structure* **11**: 775–789.
- Whyte WA, Bilodeau S, Orlando DA, Hoke HA, Frampton GM, Foster CT, Cowley SM, Young RA. 2012. Enhancer decommitment by LSD1 during embryonic stem cell differentiation. *Nature* **482**: 221–225.
- Wu S, Shi Y, Mulligan P, Gay F, Landry J, Liu H, Lu J, Qi HH, Wang W, Nickoloff JA, et al. 2007a. A YY1–INO80 complex regulates genomic stability through homologous recombination-based repair. *Nat Struct Mol Biol* **14**: 1165–1172.
- Wu S, Trievel RC, Rice JC. 2007b. Human SFMBT1 is a transcriptional repressor protein that selectively binds the N-terminal tail of histone H3. *FEBS Lett* **581**: 3289–3296.
- You A, Tong JK, Grozinger CM, Schreiber SL. 2001. CoREST is an integral component of the CoREST–human histone deacetylase complex. *Proc Natl Acad Sci* **98**: 1454–1458.