Published in final edited form as:

Semin Cell Dev Biol. 2010 April ; 21(2): 221–230. doi:10.1016/j.semcdb.2009.09.010.

MBT domain proteins in development and disease

Roberto Bonasio, Emilio Lecona, and Danny Reinberg

Howard Hughes Medical Institute and Department of Biochemistry, School of Medicine, New York University, New York, New York 10016, USA.

Abstract

The Malignant Brain Tumor (MBT) domain is a “chromatin reader”, a protein module that binds to post-translational modifications on histone tails that are thought to affect a variety of chromatin processes, including transcription. More specifically, MBT domains recognize mono- and di-methylated lysines at a number of different positions on histone H3 and H4 tails. Three *Drosophila* proteins, SCM, L(3)MBT and SFMBT contain multiple adjacent MBT repeats and have critical roles in development, maintenance of cell identity, and tumor suppression. Although they function in different pathways, these proteins all localize to chromatin *in vivo* and repress transcription by a currently unknown molecular mechanism that requires the MBT domains. The human genome contains several homologues of these MBT proteins, some of which have been linked to important gene regulatory pathways, such as E2F/Rb- and Polycomb-mediated repression, and to the insurgence of certain neurological tumors. Here, we review the genetics, biochemistry, and cell biology of MBT proteins and their role in development and disease.

1. Introduction

Eukaryotic DNA is packaged inside the nucleus in the form of chromatin, a macromolecular nucleoprotein complex involved in the functional regulation of the genome. Key players in the function of chromatin are histones, small basic proteins that form disc-like octamers around which the DNA is wrapped into nucleosomes and packaged inside the nucleus. The N-terminal (and to a smaller extent the C-terminal) tails of histones contain amino acid residues that are post-translationally modified with a number of chemical groups. These post-translational modifications, also termed “histone marks”, are believed to function as regulatory signals that influence the transcriptional status of the underlying genes, as well as other chromatin functions [1]. It is likely that a number of different histone marks act in concert to regulate transcription and other chromatin processes. For the purpose of this review, we will concentrate on one type of histone mark, the methylation of lysine residues. Lysine methylation exists in three different states: mono-, di- and tri-methylation. We will refer to these methylated states by the abbreviation me1, me2 and me3, preceded by the type of histone and the lysine residue number. For example, “H3K9me3” denotes tri-methylation of lysine 9 in histone H3.

Whether or not histone marks constitute an unambiguous code [2], it is beyond any doubt that they are recognized by specific protein modules or domains that reoccur in many

© 2009 Elsevier Ltd. All rights reserved.

Address for correspondence: Danny Reinberg, Ph.D., NYU School of Medicine, 522 First Avenue, Smilow #211, New York, New York 10016, Tel: 212-263-9036 • Fax: 212-263-9040, reinbd01@nyumc.org.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

nuclear proteins, formerly known as co-activators or co-repressors, and now understood to be part of a complex network of chromatin regulators. Analogous to the case of SH2 domains that recognize phosphorylation in signaling proteins, “chromatin reader” modules have evolved binding pockets or surfaces that recognize histone marks and neighboring residues [3], often with a remarkable degree of accuracy, to the point of distinguishing, for example, a di-methylated from a tri-methylated lysine. Among the motifs that bind histone marks, chromo-, bromo-, WD40 domains, as well as PHD fingers, have been intensively studied in the past few years and we refer the reader to other excellent reviews for details [3–5].

Here, we focus our attention on the Malignant Brain Tumor (MBT) domain, one that is less understood yet displays structural similarity to, and possibly an evolutionary relationship with, its better known brethren, chromo-, PWWP, and tudor domains [6]. The MBT domain binds prevalently mono- and di-methylated lysines and it is found in Polycomb group (PcG) proteins as well as in the *l(3)mbt* family of tumor suppressors, from which it derives its name. First, we will review the genetic screens that unveiled the founding members of these two gene families, and then we will take advantage of more recent functional and structural data to try to gain some insight into the potential mechanism and function of this fascinating group of proteins.

2. Genetic identification of MBT-containing proteins

When Wismar *et al.* cloned the cDNA encoding for the *lethal(3) malignant brain tumor* (*l(3)mbt*) gene in *Drosophila*, they noticed that it contained three tandem repeats of a previously unknown motif consisting of 99–103 amino acids. They christened this motif “*mbt-repeat*” and the name has remained to this day [7]. One year later, Bornemann *et al.* reported the molecular identification of *Sex comb on midleg* (*Scm*) [8], a known PcG gene [9, 10], and observed that the primary sequence of the predicted protein had two regions of homology with the *Drosophila l(3)mbt* protein (dL(3)MBT*): at the N-terminus it contained two copies of the *mbt-repeat*, and at the C-terminus it contained a domain previously found in dL(3)MBT and in the *Polyhomeotic* product dPH, another PcG family member. This second region of homology was called the “SPM domain” from the initials of the three proteins in which it had been observed (see below). Fast-forward fifteen years; we now know that several proteins encoded in the genomes of multi-cellular eukaryotes contain conserved MBT domains (Figure 1). With few exceptions, each of these proteins can be traced back to one of three *Drosophila* orthologues, the previously mentioned dL(3)MBT, dSCM, and a protein containing four MBT repeats known as dSFMBT [11] (Figure 1). All MBT proteins encoded in the fully sequenced *C. elegans*, *D. melanogaster*, and *H. sapiens* genomes are shown in Figure 1, and their domain organization is indicated. In the following section, we will review the work that led to the genetic and molecular identification of these genes, and discuss the implications for their function.

2.1 An MBT tumor suppressor in *Drosophila*

The temperature sensitive mutant *lethal(3) malignant brain tumor*, was isolated in a genetic screen for malignant transformations in the developing fly [12]. When shifted to the restrictive temperature during the first 6 hours of embryonic development, *l(3)mbt* mutants develop malignant overgrowth of the larval brain with 100% penetrance and die at the end of the third instar larva phase, without ever giving rise to an adult [12]. The brain is not the

*To avoid ambiguity, for the rest of the review we will add the “d” prefix when mentioning proteins from *Drosophila melanogaster*, and “h” for *Homo sapiens* unless the species to which we are referring is unambiguous in the context. For the gene names we will use standard nomenclature, and thus *Drosophila* genes will be indicated in small letter italics (except for capitalization of the first letter in the case of dominant mutations) and human genes will be in capitalized italics.

only tissue affected in this mutant; imaginal discs also are hyperplastic and their single-layer epithelium is abnormally folded. In contrast with the brain, the transformation of the imaginal discs is not malignant because the transformed cells do not give rise to secondary tumors upon transplantation [12]. Because most of the *Drosophila* larva is composed of post-mitotic or endo-mitotic cells, only a restricted number of cells and tissues that are actively cycling are susceptible to neoplastic transformation, namely the adult optic neuroblasts and ganglion mother cells in the larval brain, the imaginal discs, the gonads, and blood cells [13]. Therefore, the neoplastic transformation in *l(3)mbt* mutants is restricted to the brain not because the protein has a particular function in this organ, but because most other tissues are not prone to transformation in this model organism.

More *l(3)mbt* alleles were recovered in a screen for germ cell formation [14], and the defect was traced back to the inability of the mutants to carry out synchronous nuclear division that occurs in the first hour of embryonic development. Interestingly, all the *l(3)mbt* alleles recovered, including two nonsense mutations resulting in truncated proteins, are temperature sensitive suggesting that the underlying process which is regulated by *l(3)mbt* is temperature sensitive, not the gene product itself [14].

The *Drosophila l(3)mbt* gene encodes for a proline-rich protein of 1477 amino acids that contains a C₂/C₂ zinc finger domain (aa 730 – 752), three repeats of the 100 amino acid long MBT domain (aa 819 – 1133) [7], and a region of homology with dSCM and dPH called the SPM domain (SCM, PH and MBT homology domain) [8]) (Figure 1), which might form a surface for protein-protein interactions such as hetero- and homo-dimerization [15, 16]. The SPM domain belongs to the extended family of sterile-alpha-motif (SAM) domains, also known as helix-loop-helix (bHLH) domains [17, 18], found in more than 200 human proteins including almost half of the transcription factors belonging to the ETS family [19]. All known mutations that cause the malignant phenotype in *Drosophila* map to or near the three MBT repeats, suggesting that they are indeed important for L(3)MBT function [14].

2.2 MBT domains in the Polycomb group

Shortly after the molecular identification of *l(3)mbt*, Bornemann *et al.* reported that another *Drosophila* gene, *Sex comb on midleg (Scm)*, encoded a protein containing two tandem MBT repeats at the N-terminus, three putative C₂/C₂ zinc fingers, and a C-terminal SPM domain [8]. *Scm* was originally identified as a gene required for the proper maintenance of expression patterns of homeotic genes during development [10], and was therefore classified as a *Polycomb* group (PcG) repressor. *Scm* mutant embryos have a classic PcG developmental phenotype, in which anterior segments of the body plan assume the identity of more posterior ones (e.g. thorax into abdomen) [10], due to the loss of repression of genes in the *bithorax* locus [20]. The phenotype is even more pronounced when the maternal contribution of wild-type *Scm* protein to the zygote is abolished. In these *Scm*^{-/-} embryos developing from germline-specific *Scm* null mothers, all body segments are completely transformed to A8, the most posterior abdominal segment [9], implicating dSCM as a critical component of the repressive machinery required for transcriptional memory and maintenance of cell identity. Importantly, as is the case for *l(3)mbt*, most of the recovered mutations in *Scm* cluster at or near the MBT repeats, again underscoring their importance for the cellular function of MBT proteins [21]. Mutants in the first MBT repeat of *Scm*, such as the *Su(z)302*, *ET50*, and *R5-13* alleles (caused by missense mutations D215N, G275E and a four amino acid deletion, respectively), though behaving as hypomorphic alleles in isolation, show stronger interactions with the *Polycomb* mutant *Pc*³ than the null alleles, suggesting that a nonfunctional first MBT repeat may create a dominant negative protein that exacerbates mild defects in PcG complexes [21].

2.3 A slew of orthologues and paralogues

The discovery of *l(3)mbt* and *Scm* and their apparent similarity in domain organization was followed by a quest for homologues in the fly and other organisms, including humans. The *Drosophila* genome encodes for one more MBT-containing protein, dSFMBT that interacts with the PcG protein dPHO to form the PHO repressive complex (PhoRC, see section 4.1) [11]. *Drosophila* SFMBT contains four MBT repeats and an SPM domain and, like dSCM, is a PcG member. In *C. elegans* three genes encode MBT repeats: *lin-61*, *mbtr-1* and *sor-3* (Figure 1). LIN-61 and MBTR-1 both contain four MBT repeats, but no other discernible protein motif. *lin-61* is a class B synthetic multivulva (synMuv) gene and therefore is in the same genetic pathway as the Rb/E2F family of transcriptional repressors, critical for vulva development [22]. MBTR-1 resembles LIN-61 in domain organization, but it does not participate in the synMuv signaling pathway and nothing else is known about its function [22]. On the other hand, *sor-3* encodes for a protein containing only one MBT domain that has no sequence orthologues in any other organism, but from a functional perspective resembles PcG proteins given that in its absence the pattern of *Hox* gene expression is not correctly maintained, even after being correctly established [23].

In the case of mammals, the analysis of MBT-containing genes and their function is complicated by the existence of several closely related paralogues in each family, likely results of ancestral gene duplication, as well as multiple promoters and splice variants. To date, four human genes have been identified for their homology to *l(3)mbt*: *L3MBTL1* (also known as *L3MBTL1*) [24], *L3MBTL2* (previously known as *H-l(3)mbt-l*) [25], *L3MBTL3* (previously known as *MBT1*) [26], and *L3MBTL4*. It is important to note that, unlike the other three *l(3)mbt*-like human genes, *L3MBTL2* does not possess an SPM domain, has four rather than three MBT domains, and by sequence homology it appears to be more closely related to human *MBTD1* (another 4-MBT repeat gene) and *Drosophila Sfmbl* than to *l(3)mbt* (Figure 1). Two human genes bear close resemblance to *Scm*, *SCMH1* on chromosome 1p34 [27] and *SCML2* on chromosome Xp22 [28]. The previously identified *SCML1*, also on Xp22, encodes for a small protein that is homologous to dSCM only in its SPM domain but contains no MBT domains [29]; however, its cellular function might still be related to the MBT family of proteins, as it could interact via its SPM domain with hSCMH1 or hSCML2 and possibly serve the purpose of a dominant negative isoform. Finally, two human genes, *SFMBT1* and *SFMBT2*, encoding proteins with four MBT domains and a C-terminal SPM domain are commonly considered *Sfmbl* homologues, although the CLUSTALW algorithm places them in a common branch with *l(3)mbt*, despite the presence of an extra MBT domain (Figure 1) [30–32]. Two additional human proteins, PHF20 and its close homologue PHF20L1 (C20orf104), were reported to contain a single N-terminal MBT domain paired with a Tudor domain [33], but unbiased protein domain predictions reveal either an MBT or a Tudor domain in this position (preceding a second, unambiguous Tudor domain), depending on which algorithm is used [34]. Therefore, PHF20 and PHF20L1 might contain a transitional or hybrid domain, displaying features of both MBT and Tudor. Until they are characterized further, we will refrain from considering PHF20 and PHF20L1 *bona fide* MBT proteins.

3. Structure and function of the MBT domain

As soon as the MBT domain was linked to development via PcG protein, and tumor suppression via dL(3)MBT, the quest for its structure and biochemical function began. MBT domains with or without their putative physiological ligand(s) have been crystallized and their structure solved at atomic resolution. At the time of writing, a search for “MBT” in the Protein Data Bank retrieves 30 atomic structures containing the MBT domain. From the analysis of these structures, as well as from sequence homology studies, it has emerged that the MBT domain is evolutionary related to the Royal family of chromatin binding domains,

which also includes the Tudor, plant Agenet, Chromo, and PWWP domains [6]. These domains can recognize and bind histone marks and are in general found in proteins that bind to chromatin. In fact, MBT domains from all the proteins analyzed to date can bind to methyl-lysine residues, in particular to methylated lysines in the context of histone tails. The spatial arrangement of multiple MBT domains also presents interesting features and may form an interacting surface in addition to the methyl-lysine binding pocket. In the next section we will review the structure and biochemical function(s) of the MBT domains.

3.1 Binding to methyl-lysine

Individual MBT domains contain two distinct structural motifs: a 30–50 amino acid long N-terminal “arm” formed by at least one alpha and one 3_{10} helix, and a longer (60 to 80 amino acids) C-terminal beta subunit “core” formed by five beta strands arranged in a barrel topology, a 3_{10} helical turn connecting $\beta 3$ and $\beta 4$, an alpha helix, and C-terminal sixth beta strand [35]. Importantly, the arm of each repeat does not pack against its own beta subunit core, but it forms extended interactions with the preceding MBT domain (see section 3.3), so that typically all MBT repeats in a given protein need to be expressed as a unit in order to obtain a stable structure [36]. The basic architecture of the MBT domain was first described for human L3MBTL [35, 37], but it is conserved in all other crystal structures solved, including MBT repeats from hSCML2 [36], dSCM [38], hL3MBTL2 [39], and dSFMBT [40]. Differences between the structures of these proteins are to be found in the spatial organization of the multiple MBT repeats, rather than in the structure of the repeats themselves. This will be discussed in the next section.

The beta subunit core of each MBT repeat forms a binding pocket suggestive of potential interactions with multiple ligands. Indeed, in the first solved structure of hL3MBTL, each of the three pockets bound to morpholino rings from the buffering agent MES [35]. Considering the homology of the MBT repeats with other domains involved in methyl-lysine recognition, and the involvement of MBT proteins in development and tumor suppression, histone tails bearing methylated lysines became instant candidates for recognition by MBT proteins. In fact, isolated MBT domains bind *in vitro* to a number of histone tail peptides bearing mono- or di-methylated lysines, discriminating against unmethylated or tri-methylated lysines [33, 38, 41] (Table 1). Interestingly, in co-crystal structures of MBT repeats and peptide ligands, it is evident that only one of the MBT domain accommodates a methyl-lysine in its pocket, namely the 2nd MBT repeat in hL3MBTL [37, 42], hSCML2 [43] and dSCM [38], and the 4th repeat in hL3MBTL2 [39] and dSFMBT [40]. The methyl-lysine binding pocket is lined by a conserved triad of aromatic amino acids that form an “aromatic cage”, and the interaction with the ligand is buttressed by a highly conserved aspartate residue that forms a hydrogen bond with a methyl-ammonium proton [42].

The preference for Kme1 and Kme2 versus unmethylated lysine (Kme0) and Kme3 is explained by the features of the ligand binding pocket. First, a narrow entry channel to the pocket sterically excludes bulky Kme3 residues; second, the hydrogen bond and ion pair interaction between the conserved aspartate residue in the pocket and the lysine ligand is only possible when a free methyl-ammonium or dimethyl-ammonium proton is available, increasing the affinity for Kme1 and Kme2 versus Kme3; third, the surface of the pocket is lined by hydrophobic amino acids, which favors interactions with the more hydrophobic Kme1 and Kme2, versus Kme0 [42].

The structural differences between the ligand-binding pockets of chromo- and MBT domains are highlighted in Figure 2a and tentatively explain the selective recognition of Kme3 (chromo) versus Kme1/2. In this regard, the MBT domain is more similar to the Tudor domain, which also displays a ligand binding pocket rather than a surface groove. A

surface binding groove, as found in chromodomains, allows for a better fit with the residues adjacent to the methyl-lysine, achieving a high degree of context-dependent specificity in peptide recognition and allowing for the discrimination, for example, between H3K9me3 and H3K27me3 [44] (Figure 2a, right). In contrast, the histone tail residues adjacent to the methyl-lysine do not form extensive contacts with the surface of the MBT domain, explaining the lack of specificity of these binding interactions *in vitro* [39] (Figure 2a, left). This is not to say that the MBT domain lacks specificity *in vivo*. For example, H4K20me1, but not H3K9me1/2, recruits hL3MBTL to chromatin, despite the fact that peptides containing these histone marks bind to the MBT repeats with similar affinities *in vitro* [41]. Clearly, the context in which the methylated lysine is present *in vivo* contributes to binding specificity in ways that cannot be recapitulated *in vitro*. In addition, we note that a surface groove similar to the one observed in the chromodomain is also present in the vicinity of the ligand binding pocket of the MBT domain (Figure 2a, left). It is possible that *in vivo* this groove accommodates yet to be discovered peptide ligands.

Finally, it is worth noting that a potential regulatory mechanism for ligand binding in an MBT pocket was initially observed in the crystal structure of hL3MBTL. When the crystals were grown with hL3MBTL protein pre-incubated with GTP, the pocket in the third repeat could not bind a MES molecule because its entrance became obstructed by arginine 461 [35]. It is hard to assess the importance of this observation given that the only MBT repeat known to bind a physiological ligand in hL3MBTL is the second repeat, not the third one; nonetheless, a similar gating mechanism was observed in crystals of hSCML2, in which phenylalanine 213 swings open and allows access of the second repeat binding pocket to methyl-lysine [43]. Although only based on crystal structures, these observations suggest that, in some cases, binding to methyl-lysine may be regulated allosterically or may induce conformational changes in the MBT domain that could potentially be transmitted to the other domains.

3.2 Functional significance of multiple MBT repeats

The observation that only one out of two, three or four repeats of the same structural domain appears to bind to a ligand begs the question of what could be the function of the other repeats. Although it is conceivable that they only provide structural support for the one repeat that interacts with histone marks, some observations strongly argue against this simple interpretation. First and foremost, the critical aspartate residue that forms a hydrogen bond with methyl-lysine is conserved across the board in most MBT repeats, from *Drosophila* to man, even when the aromatic triad that forms the pocket in methyl-lysine binding repeats is missing. Moreover, the *Su(z)302* mutation in *Drosophila Scm*, causing posterior transformation in the homozygous state and synthetic lethality with *Polycomb*, arises from the substitution of this conserved aspartate in the first MBT repeat with an asparagine [21], but the first repeat is not the one that binds methyl-lysine in the crystal structure [38]. Also to be considered are the observations that the first pocket of hL3MBTL can bind to a Pro-Ser-Thr sequence [42] and that conserved residues in the third domain contribute to binding affinity for H1bK26me1 [45]. Based on this evidence, we suggest the possibility that the MBT repeats that do not bind histone peptides *in vitro*, might recognize yet to be discovered ligands, either in a constitutive or regulated manner.

3.3 Multi-domain organization

Although each MBT domain encodes an N-terminal “arm” and a C-terminal “core”, each arm packs against the core of a different sister repeat. This gives rise to the geometries outlined in Figure 2b. In SCM family proteins, the two MBT repeats form a saddle-like interlocked structure [36]; in L(3)MBT family proteins, the three MBT repeats form a clover-like, three-leaved propeller structure with a central channel measuring $15 \text{ \AA} \times 10 \text{ \AA}$

[35]; and in hL3MBTL2 and dSFMBT, the four MBTs form an irregular rhombus composed of a three-leaved propeller similar to the one observed in hL3MBTL, with the extra MBT repeat grafted asymmetrically onto one side [39, 40]. The interdigitation of the N-terminal arms, which are very flexible [35], with the more stable cores from different MBT repeats allows, at least in theory, for communication between repeats so that the conformation in one could be affected by a ligand binding event in another. This kind of conformational change might help explain the paradoxical observations of functional mutants in pockets that do not bind ligands *in vitro* (see section 3.1). Along the same lines, it is possible that the contribution of the third MBT repeat of hL3MBTL to H1bK26me binding [45] might be due to this type of communication between domains. It is important to note, though, that NMR spectroscopic studies of hSCML2 did not uncover conformational change in the first MBT repeat pocket upon binding of the ligand to the second MBT repeat [43].

In other propeller structures the central cavity typically forms protein interaction, ligand binding or catalytic sites [46], but such functions have yet to be attributed to the cavity observed in 3- and 4-MBT proteins. Nonetheless, Wang *et al.* observed that a sulfate ion was bound to an electropositive patch in the channel formed by the three MBT repeats of hL3MBTL, and noted that the size of the channel would be sufficient to accommodate a nucleotide co-factor or, more intriguingly, a single stranded nucleic acid, with the backbone phosphate replacing the sulfate ion observed in the structure [35].

4. Cellular function of MBT proteins

Genetic and structural studies link the role of MBT proteins in transcriptional regulation to the recognition of lower methylation states of lysine residues on histone tails. How can we begin to explain the role of these proteins at the cellular level? We propose here that the experimental evidence supports a classification of MBT proteins into two families: 1) proteins related to *Drosophila* SCM that have two MBT repeats and an SPM domain plus one or more zinc fingers, and are genetically and biochemically implicated in the repression of developmental genes and maintenance of cell identity; and 2) proteins related to *Drosophila* L(3)MBT that have three MBT repeats, are involved in chromatin processes affecting the cell cycle and associate with E2F/Rb complexes. In this scheme, dSFMBT-related proteins would be placed somewhere in the middle of these two categories because the *Drosophila* *Sfmbt* null mutant displays a classic *Polycomb* phenotype [11], yet dSFMBT regulates the function of E2F/Rb complexes [47]. Similarly, the human orthologue L3MBTL2 forms a complex reminiscent of PRC1 (P. Trojer and D. Reinberg, unpublished results), but it also associates with E2F6 [48]. Interestingly, a functional connection between the Polycomb and Rb/E2F repression systems has been proposed before [49], and it may very well be that MBT proteins contribute to the link between these two pathways. In the next section we will review the biochemistry and cell biology experiments that support this functional classification.

4.1 Protein complexes containing MBT proteins

It was observed early on that dSCM and its homologues contain a C-terminal SPM domain related to the SAM domain found in ETS transcription factors, and that this domain mediates *in vitro* interactions with the similar SPM domain found in PH, the product of the PcG gene *Polyhomeotic* [15, 50]. Kim *et al.* reported that the SPM domain forms helical polymers *in vitro* [16], suggesting a model for the assembly of large stretches of PcG proteins on chromatin. Although this model awaits confirmation *in vivo*, the link between dSCM and Polycomb Repressive Complex 1 (PRC1) is well established given that dSCM and hSCMH1 are associated with PRC1 in nuclear extracts from *Drosophila* [51] and HeLa cells [52]. Less is known about the other human dSCM homologue SCML2, though this protein too appears to form a complex in the nucleus with so far unidentified partners (E.L.

and D.R., unpublished results). In contrast with other PRC1 core components, dSCM is only present in sub-stoichiometric quantities in the purified complex [53] and the bulk of dSCM separates from dPH upon gel filtration chromatography [54], suggesting that its incorporation into PRC1 may be regulated (e.g. only occurring on chromatin at target genes) or that the interaction is inherently weak.

Genetic and biochemical evidence connect the L(3)MBT family to the E2F/Rb pathway of transcriptional repression. *Drosophila* L(3)MBT was recovered in affinity purifications of the Myb-MuvB super-complex [55] (also known as the dREAM complex [56]), which contains proteins of the Rb and E2F families and represses transcription of developmentally regulated genes. Although Myb is commonly associated with transcriptional activation, dL(3)MBT was only found in a repressive Myb supercomplex also containing Rb/E2F [55]. dL(3)MBT was also identified as a regulator of the E2F/Rb pathway in a genome-wide RNAi screen [47]. This connection holds true in organisms other than *Drosophila*. The *C. elegans* orthologue of *Sfinbt* and *L3MBTL2*, *lin-61* (Figure 1) belongs to the synMuvB class of genes, which puts it in the same pathway as the worm orthologues of Rb and E2F [22]. In humans, L3MBTL co-purifies with Rb, as well as HP1 and is found at E2F target genes, like *MYC* and *CYCLIN E1* [45], while L3MBTL2 is found in a super-complex containing E2F6 and HP1, among other proteins [48]. These data strongly suggest that L(3)MBT-like proteins function in transcriptional repression of at least a subset of the large group of genes controlled by the E2F/Rb pathway, which may explain their function as tumor suppressors.

Lastly, we consider the case of SFMBT. The *Drosophila* protein was identified as a factor stably associated with the PcG members PHO and PHOL, the only PcG proteins with a well-characterized DNA binding activity [11]. The *Sfinbt* null homozygotes die as larvae and homozygous imaginal disc cells show derepression of HOX genes [11]. In addition, dSFMBT interacts with dSCM in heterologous insect cell extracts, they co-occupy some PcG responsive promoters, and mutations in *Scm* and *Sfinbt* interact synergistically in deregulating expression of certain HOX genes in imaginal discs [40]. Taken together, these observations place SFMBT in the PcG pathway of gene repression, but the same RNAi screen that identified dL(3)MBT as a regulator of E2F/Rb responsive genes also recovered *Sfinbt* [47], suggesting that its product may function in more than one repressive pathway. In contrast to the case in *Drosophila*, the mammalian SFMBTs do not form stable complexes with YY1, the PHO orthologue [57], although YY1 and hSFMBT2 interact upon their overexpression in 293 cells [31]. On the other hand, L3MBTL2, which is slightly closer to dSFMBT than to dL(3)MBT by sequence (Figure 1a), associates with E2F6. It is important to note, though, that hL3MBTL2 lacks the C-terminal SPM domain found in dSFMBT, which is instead present in hSFMBT1 and hSFMBT2. It is possible that sequence analysis alone may be misleading in this case, and that hL3MBTL2 might be closer in function to dL(3)MBT than to dSFMBT, whose true functional homologues could be hSFMBT1 and hSFMBT2 (or hMBTD1). Only a thorough biochemical characterization of hSFMBT-containing complexes in mammalian cells is likely to shed light on this problem.

4.2 MBT proteins and transcriptional repression

As we await further insights into the biochemical composition and cellular roles of complexes containing MBT proteins, we can at least make a final statement regarding their ultimate function inside the nucleus. All MBT proteins tested repress transcription of endogenous and reporter genes, after being recruited to their promoters *in vivo*. This has been shown in intact flies for dSCM [58] and dSFMBT [11], and in cultured cells for hL3MBTL [59] and hSFMBT1 [60]. In the two cases in which it was tested (hL3MBTL and hSFMBT1), the MBT domains were required for repression, even when the endogenous recruitment system, whichever it may be, was bypassed by artificially tethering hL3MBTL and hSFMBT1 to a promoter using the GAL4/UAS system. This suggests that the

recognition of the histone mark(s) does not serve only as a recruitment signal, if it serves as such at all (see below), but is directly involved in the repression mechanism, consistent with the observation that hL3MBTL compacts chromatin *in vitro*, via recognition of H4K20me1 or H1bK26me1/2 on different nucleosomes [45]. A histone mark-dependent compaction mechanism, ultimately leading to decreased accessibility of the transcription machinery to chromatin and resulting in gene repression, is a fascinating model for the function of the MBT domains and merits being tested *in vivo* and with other MBT-containing proteins. In this regard, MBT protein complexes may function similarly to PRC1, which was shown to compact chromatin *in vitro*, though in that case histone modifications were not a pre-requisite for compaction [61].

MBT proteins might also repress target genes indirectly by recruiting other repressive complexes. For example, dSCM might recruit PRC1 by interacting with dPH via the respective SPM domains, and hL3MBTL and hL3MBTL2 both co-purify with HP1 a well known constituent of repressive chromatin structures [62]. These possibilities are schematized in Figure 3.

4.3 Recruitment of MBT proteins to target genes

How do MBT proteins reach the target genes that they are set to repress? This is a recurrent question in the chromatin field, as many protein complexes with established functions in chromatin processes lack subunits containing sequence-specific DNA binding activity, and yet are found at precise positions on chromosomes *in vivo*. The mechanism for recruitment of MBT proteins is no less mysterious. Here, we mention a few possibilities, outlined in Figure 3.

By immunohistochemistry, *Drosophila* SCM co-localizes with the PRC1 component dPH on polytene chromosomes, including at the finely regulated *bithorax* locus [15]. Because the two proteins interact via their SPM domains [15] and *Scm* mutants lacking the SPM domain are functional null alleles [54, 63], it is possible, yet so far unproven, that PRC1 recruits dSCM to its own target genes. To this end, comparison of genome-wide distributions of dSCM and PRC1 in the fly and in mammals would be very informative. To date the only genome-wide Chromatin IP (ChIP) data available is for *Drosophila* SFMBT [64]. Unlike dSCM, dSFMBT does have a stably associated partner with DNA binding activity, the PcG protein dPHO (see section 4.1), but the distribution of dSFMBT and dPHO only overlap by 50%, and the sites bound only by dSFMBT do not contain the DNA motif recognized by dPHO, suggesting alternative mechanisms for dSFMBT recruitment [64]. Intuitively, it may seem that particular histone marks or combinations thereof recognized by the MBT domains should provide positional information contributing to recruitment, but the low affinities of these interactions suggest that they are more likely to function in synergy with other recruitment mechanisms. As dSFMBT binds to several histone marks *in vitro* [11], these modifications are good candidates to function as part of a recruitment signal, but their genome-wide profile has not been compared with that of dSFMBT. In addition, dSFMBT and its close human homologue L3MBTL2 bind to SUMO (a ubiquitin-like post-translational tag implicated in many nuclear processes) and regulate the repressive activity of SUMOylated transcription factor Sp3 [65, 66], offering a potential dPHO-independent recruitment mechanism.

The only other MBT protein for which recruitment mechanisms have been proposed is hL3MBTL. hL3MBTL can be brought to promoters *in vivo* by the transcriptional repressor TEL, via a SAM-SPM domain interaction [59]. It was also shown using an artificial tethering system that the enzymatic activity of PR-SET7, which catalyzes H4K20me1 [67], is sufficient to recruit hL3MBTL to a reporter gene [41]. This result is surprising, given that the hL3MBTL-H4K20me1 interaction alone has very low affinity [37, 42], but it must be

considered that in the tethering system used for these experiments, hundred of molecules of PR-SET7 are recruited to the target locus, catalyzing the deposition of the H4K20me1 on a long stretch of DNA. This might indeed be sufficient to recruit H4K20me1-binding proteins such as hL3MBTL, but it remains to be shown that the mark alone suffices as a recruitment signal on endogenous loci. Interestingly, the repressive effect of hL3MBTL requires the MBT domains even when the recruitment step is artificially bypassed by fusion to the GAL4 DNA binding domain [59], suggesting that the function of the MBT domains in hL3MBTL1 cannot be limited to recruitment, and that they might need to bind to the histone mark(s) in order to compact chromatin and cause repression [45]. Similar to dSFMBT and hL3MBTL2, hL3MBTL is also recruited to a promoter bound by SUMOylated Sp3 [66], but in this case a direct binding to SUMO was not shown and the recruitment of hL3MBTL might have been secondary to the formation of compacted chromatin at this particular locus. Nonetheless, the link between SUMOylation and MBT proteins is intriguing, as it appears to be conserved through evolution. The *C. elegans* MBT protein SOR-3 (see section 2.3) interacts genetically with SOP-2 [23], a PcG protein that depends on SUMOylation to repress HOX genes [68]. The significance of this connection remains unclear, and the importance of the SUMO pathway in recruitment of MBT proteins in various organisms is a promising avenue for future investigation.

4.4 Genetic mouse models

To date, only two targeted mutations of MBT proteins have been reported in the mouse. Mice bearing a homozygous truncation of *Scmh1* lacking the SPM domain display a weak *Polycomb* phenotype, and a more pronounced defect in spermatogenesis, but no sign of malignancies [63]. It is worth noting that the homologous gene *Scml2* is on the $\times$ chromosome, which is silenced during meiosis. If *Scmh1* and *Scml2* can compensate for each other, the cellular phenotype of an *Scmh1* null mutation should be more pronounced during male meiosis, which is exactly what is observed. Alternatively, mouse *Scmh1* might have evolved a male meiosis-specific function, not apparent or not existent in *Drosophila*. *L3mbtl3* knock-out mice die at E18 due to anemia, secondary to the accumulation of undifferentiated myeloid progenitors [26]. Though such a phenotype is reminiscent of the accumulation of undifferentiated cells in myeloid malignancies, the embryonic lethality thwarted tumorigenicity studies, which will require the generation of conditional knock-out animals. Conditional knock-out animals might also provide further insight into the function of the other MBT proteins in mammals and we believe will be more useful than germline knock-outs, as many MBT proteins are likely to carry out essential cellular functions predictive of early lethality in the case of homozygous mutant embryos.

5. The L(3)MBT family as tumor suppressor genes in humans

Arguably, interest in the MBT family of proteins was also fueled by the strong tumorigenic phenotype observed in *l(3)mbt* mutant flies. It was originally observed that the human *L3MBTL* gene mapped to 20q12, a region amplified in some breast cancer [69] and deleted in hematopoietic malignancies such as myeloproliferative disorders, myelodysplastic syndromes, myeloid leukemias, and polycythemia vera [70, 71]. These correlations and the observation that over-expression of *L3MBTL* caused severe mitotic defects leading to multinucleated cells [24], suggested that alterations in *L3MBTL* may be responsible for tumor development or progression. However, a detailed analysis of DNA sequences and mRNA abundance in a panel of leukemic cell lines heterozygous for 20q deletions did not reveal any significant mutations or alterations in expression levels of the retained *L3MBTL* allele [71, 72], nor any preferential loss of the imprinted and silenced allele [72, 73]. These last experiments are obviously limited by the choice of disease and tissue analyzed, and they cannot exclude a role for *L3MBTL* or other MBT proteins in the etiology of certain types of cancer. For example, *L3MBTL3* also maps to a chromosomal region, 6q23, frequently

altered in acute leukemia cells [26, 74], and a recent associative study reported an infrequent but significant rate of focal hemizygous and homozygous deletions of *L3MBTL2*, *L3MBTL3*, and *SCML2* in medulloblastoma [75], renewing the interest in MBT proteins and their link with cancer.

6. Closing remarks and future directions

MBT proteins have risen to the stage of chromatin regulation as transcriptional repressors capable of recognizing mono- and di-methylated lysines on histone N-terminal tails. The experimental evidence gathered from a combination of genetic, biochemical, and cell biology approaches point to critical functions of MBT proteins in biological processes as fundamental as the maintenance of cell identity and body pattern during development, as well as the regulation of mitosis and tumor suppression.

As we learn more about these proteins, the complexes that contain them, their recruitment to chromatin and the ensuing downstream effects, several open questions remain that we believe are critical to understanding their function. How are MBT proteins specifically recruited to their target genes? Does recruitment only occur via interactions with sequence-specific transcription factors, or does recognition of histone marks via the MBT domain also play a role? If the MBT domain is not required for recruitment, what is its role in repression? Does the chromatin compaction model shown for L3MBTL apply to other MBT proteins? What is the molecular nature of the link between the PcG and E2F/Rb repression pathways [49]?

Conventional biochemistry, i.e. the purification of complexes containing MBT proteins, should be integrated with genome-wide chromatin IP profiling to gain a better picture of the molecular associations of MBT proteins inside the cells. At the same time, genetic approaches in model organisms, extremely valuable so far, should be extended now that we have molecular data at our disposal. Can targeted mutations that abolish binding of the MBT domains to their physiological ligand rescue null and hypomorph alleles of *Scm* and *l(3)mbt* in *Drosophila*? Is the PHO-independent recruitment of dSFMBT affected by targeted mutation in its MBT domains? What is the effect of deleting *Scm*, *l(3)mbt*, and *Sfmbt* homologues in mice? The answers to these and other questions will shed much needed light on the cellular role of MBT proteins, the mechanism by which they exert transcriptional repression, and their role in cancer formation in model organisms and man.

References

1. Kouzarides T. Chromatin Modifications and Their Function. *Cell*. 2007; 128:693–705. [PubMed: 17320507]
2. Jenuwein T, Allis CD. Translating the histone code. *Science*. 2001; 293:1074–1080. [PubMed: 11498575]
3. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ. How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat Struct Mol Biol*. 2007; 14:1025–1040. [PubMed: 17984965]
4. Ruthenburg AJ, Allis CD, Wysocka J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol Cell*. 2007; 25:15–30. [PubMed: 17218268]
5. Sims RJ, Reinberg D. Histone H3 Lys 4 methylation: caught in a bind? *Genes Dev*. 2006; 20:2779–2786. [PubMed: 17043307]
6. Maurer-Stroh S, Dickens NJ, Hughes-Davies L, Kouzarides T, Eisenhaber F, Ponting CP. The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci*. 2003; 28:69–74. [PubMed: 12575993]

7. Wismar J, Löffler T, Habtemichael N, Vef O, Geissen M, Zirwes R, et al. The *Drosophila melanogaster* tumor suppressor gene *lethal(3)malignant brain tumor* encodes a proline-rich protein with a novel zinc finger. *Mech Dev.* 1995; 53:141–154. [PubMed: 8555106]
8. Bornemann D, Miller E, Simon J. The *Drosophila* Polycomb group gene *Sex comb on midleg (Scm)* encodes a zinc finger protein with similarity to polyhomeotic protein. *Development.* 1996; 122:1621–1630. [PubMed: 8625848]
9. Breen TR, Duncan IM. Maternal expression of genes that regulate the bithorax complex of *Drosophila melanogaster*. *Dev Biol.* 1986; 118:442–456. [PubMed: 3098596]
10. Jürgens G. A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature.* 1985; 316:153–155.
11. Klymenko T, Papp B, Fischle W, Köcher T, Schelder M, Fritsch C, et al. A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.* 2006; 20:1110–1122. [PubMed: 16618800]
12. Gateff E, Löffler T, Wismar J. A temperature-sensitive brain tumor suppressor mutation of *Drosophila melanogaster*: developmental studies and molecular localization of the gene. *Mech Dev.* 1993; 41:15–31. [PubMed: 8507589]
13. Gateff E. Malignant neoplasms of genetic origin in *Drosophila melanogaster*. *Science.* 1978; 200:1448–1459. [PubMed: 96525]
14. Yohn CB, Pusateri L, Barbosa V, Lehmann R. *l(3)malignant brain tumor* and three novel genes are required for *Drosophila* germ-cell formation. *Genetics.* 2003; 165:1889–1900. [PubMed: 14704174]
15. Peterson AJ, Kyba M, Bornemann D, Morgan K, Brock HW, Simon J. A domain shared by the Polycomb group proteins *Scm* and *ph* mediates heterotypic and homotypic interactions. *Mol Cell Biol.* 1997; 17:6683–6692. [PubMed: 9343432]
16. Kim CA, Sawaya MR, Cascio D, Kim W, Bowie JU. Structural organization of a *Sex-comb-on-midleg/polyhomeotic* copolymer. *J Biol Chem.* 2005; 280:27769–27775. [PubMed: 15905166]
17. Ponting CP. SAM: a novel motif in yeast sterile and *Drosophila* polyhomeotic proteins. *Protein Sci.* 1995; 4:1928–1930. [PubMed: 8528090]
18. Qiao F, Bowie JU. The many faces of SAM. *Sci STKE.* 2005; 2005:re7. [PubMed: 15928333]
19. Sharrocks AD. The ETS-domain transcription factor family. *Nat Rev Mol Cell Biol.* 2001; 2:827–837. [PubMed: 11715049]
20. Simon J, Chiang A, Bender W. Ten different Polycomb group genes are required for spatial control of the *abdA* and *AbdB* homeotic products. *Development.* 1992; 114:493–505. [PubMed: 1350533]
21. Bornemann D, Miller E, Simon J. Expression and properties of wild-type and mutant forms of the *Drosophila* *sex comb on midleg (SCM)* repressor protein. *Genetics.* 1998; 150:675–686. [PubMed: 9755199]
22. Harrison MM, Lu X, Horvitz HR. LIN-61, one of two *Caenorhabditis elegans* malignant-brain-tumor-repeat-containing proteins, acts with the DRM and NuRD-like protein complexes in vulval development but not in certain other biological processes. *Genetics.* 2007; 176:255–271. [PubMed: 17409073]
23. Yang Y, Sun Y, Luo X, Zhang Y, Chen Y, Tian E, et al. Polycomb-like genes are necessary for specification of dopaminergic and serotonergic neurons in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA.* 2007; 104:852–857. [PubMed: 17215367]
24. Koga H, Matsui S, Hirota T, Takebayashi S, Okumura K, Saya H. A human homolog of *Drosophila* *lethal(3)malignant brain tumor (l(3)mbt)* protein associates with condensed mitotic chromosomes. *Oncogene.* 1999; 18:3799–3809. [PubMed: 10445843]
25. Wismar J. Molecular characterization of *h-l(3)mbt-like*: a new member of the human *mbt* family. *FEBS Lett.* 2001; 507:119–121. [PubMed: 11682070]
26. Arai S, Miyazaki T. Impaired maturation of myeloid progenitors in mice lacking novel Polycomb group protein *MBT-1*. *EMBO J.* 2005; 24:1863–1873. [PubMed: 15889154]
27. Berger J, Kurahashi H, Takihara Y, Shimada K, Brock HW, Randazzo F. The human homolog of *Sex comb on midleg (SCMH1)* maps to chromosome 1p34. *Gene.* 1999; 237:185–191. [PubMed: 10524249]

28. Montini E, Buchner G, Spalluto C, Andolfi G, Caruso A, den Dunnen JT, et al. Identification of SCML2, a second human gene homologous to the *Drosophila* sex comb on midleg (*Scm*): A new gene cluster on Xp22. *Genomics*. 1999; 58:65–72. [PubMed: 10331946]
29. van de Vosse E, Walpole SM, Nicolaou A, van der Bent P, Cahn A, Vaudin M, et al. Characterization of SCML1, a new gene in Xp22, with homology to developmental polycomb genes. *Genomics*. 1998; 49:96–102. [PubMed: 9570953]
30. Usui H, Ichikawa T, Kobayashi K, Kumanishi T. Cloning of a novel murine gene *Sfmbt*, *Scm*-related gene containing four mbt domains, structurally belonging to the Polycomb group of genes. *Gene*. 2000; 248:127–135. [PubMed: 10806358]
31. Kuzmin A, Han Z, Golding MC, Mann MR, Latham KE, Varmuza S. The PcG gene *Sfmbt2* is paternally expressed in extraembryonic tissues. *Gene Expr Patterns*. 2008; 8:107–116. [PubMed: 18024232]
32. Frankenberg S, Smith L, Greenfield A, Zernicka-Goetz M. Novel gene expression patterns along the proximo-distal axis of the mouse embryo before gastrulation. *BMC Dev Biol*. 2007; 7:8. [PubMed: 17302988]
33. Kim J, Daniel J, Espejo A, Lake A, Krishna M, Xia L, et al. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep*. 2006; 7:397–403. [PubMed: 16415788]
34. Trojer P, Reinberg D. Beyond histone methyl-lysine binding: how malignant brain tumor (MBT) protein L3MBTL1 impacts chromatin structure. *Cell Cycle*. 2008; 7:578–585. [PubMed: 18256536]
35. Wang WK, Tereshko V, Boccuni P, MacGrogan D, Nimer SD, Patel DJ. Malignant brain tumor repeats: a three-leaved propeller architecture with ligand/peptide binding pockets. *Structure*. 2003; 11:775–789. [PubMed: 12842041]
36. Sathyamurthy A, Allen MD, Murzin AG, Bycroft M. Crystal structure of the malignant brain tumor (MBT) repeats in Sex Comb on Midleg-like 2 (SCML2). *J Biol Chem*. 2003; 278:46968–46973. [PubMed: 12952983]
37. Min J, Allali-Hassani A, Nady N, Qi C, Ouyang H, Liu Y, et al. L3MBTL1 recognition of mono- and dimethylated histones. *Nat Struct Mol Biol*. 2007; 14:1229–1230. [PubMed: 18026117]
38. Grimm C, de Ayala Alonso AG, Rybin V, Steuerwald U, Ly-Hartig N, Fischle W, et al. Structural and functional analyses of methyl-lysine binding by the malignant brain tumour repeat protein Sex comb on midleg. *EMBO Rep*. 2007; 8:1031–1037. [PubMed: 17932512]
39. Guo Y, Nady N, Qi C, Allali-Hassani A, Zhu H, Pan P, et al. Methylation-state-specific recognition of histones by the MBT repeat protein L3MBTL2. *Nucleic Acids Res*. 2009
40. Grimm C, Matos R, Ly-Hartig N, Steuerwald U, Lindner D, Rybin V, et al. Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J*. 2009
41. Kalakonda N, Fischle W, Boccuni P, Gurvich N, Hoya-Arias R, Zhao X, et al. Histone H4 lysine 20 monomethylation promotes transcriptional repression by L3MBTL1. *Oncogene*. 2008; 27:4293–4304. [PubMed: 18408754]
42. Li H, Fischle W, Wang W, Duncan EM, Liang L, Murakami-Ishibe S, et al. Structural Basis for Lower Lysine Methylation State-Specific Readout by MBT Repeats of L3MBTL1 and an Engineered PHD Finger. *Mol Cell*. 2007; 28:677–691. [PubMed: 18042461]
43. Santiveri CM, Lechtenberg BC, Allen MD, Sathyamurthy A, Jaulent AM, Freund SM, et al. The malignant brain tumor repeats of human SCML2 bind to peptides containing monomethylated lysine. *J Mol Biol*. 2008; 382:1107–1112. [PubMed: 18706910]
44. Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev*. 2003; 17:1870–1881. [PubMed: 12897054]
45. Trojer P, Li G, Sims RJ, Vaquero A, Kalakonda N, Boccuni P, et al. L3MBTL1, a Histone-Methylation-Dependent Chromatin Lock. *Cell*. 2007; 129:915–928. [PubMed: 17540172]
46. Fulop V, Jones DT. Beta propellers: structural rigidity and functional diversity. *Curr Opin Struct Biol*. 1999; 9:715–721. [PubMed: 10607670]
47. Lu J, Ruhf ML, Perrimon N, Leder P. A genome-wide RNA interference screen identifies putative chromatin regulators essential for E2F repression. *Proc Natl Acad Sci USA*. 2007; 104:9381–9386. [PubMed: 17517653]

48. Ogawa H, Ishiguro K, Gaubatz S, Livingston DM, Nakatani Y. A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science*. 2002; 296:1132–1136. [PubMed: 12004135]
49. Dahiya A, Wong S, Gonzalo S, Gavin M, Dean DC. Linking the Rb and polycomb pathways. *Mol Cell*. 2001; 8:557–569. [PubMed: 11583618]
50. Tomotsune D, Takihara Y, Berger J, Duhl D, Joo S, Kyba M, et al. A novel member of murine Polycomb-group proteins, Sex comb on midleg homolog protein, is highly conserved, and interacts with RAE28/mph1 in vitro. *Differentiation*. 1999; 65:229–239. [PubMed: 10653359]
51. Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, et al. Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell*. 1999; 98:37–46. [PubMed: 10412979]
52. Levine SS, Weiss A, Erdjument-Bromage H, Shao Z, Tempst P, Kingston RE. The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. *Mol Cell Biol*. 2002; 22:6070–6078. [PubMed: 12167701]
53. Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE. A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. *Nature*. 2001; 412:655–660. [PubMed: 11493925]
54. Peterson AJ, Mallin DR, Francis NJ, Ketel CS, Stamm J, Voeller RK, et al. Requirement for sex comb on midleg protein interactions in Drosophila polycomb group repression. *Genetics*. 2004; 167:1225–1239. [PubMed: 15280237]
55. Lewis PW, Beall EL, Fleischer TC, Georlette D, Link AJ, Botchan MR. Identification of a Drosophila Myb-E2F2/RBF transcriptional repressor complex. *Genes Dev*. 2004; 18:2929–2940. [PubMed: 15545624]
56. Korenjak M, Taylor-Harding B, Binné UK, Satterlee JS, Stevaux O, Aasland R, et al. Native E2F/RBF complexes contain Myb-interacting proteins and repress transcription of developmentally controlled E2F target genes. *Cell*. 2004; 119:181–193. [PubMed: 15479636]
57. Wu S, Shi Y, Mulligan P, Gay F, Landry J, Liu H, et al. A YY1-INO80 complex regulates genomic stability through homologous recombination-based repair. *Nat Struct Mol Biol*. 2007; 14:1165–1172. [PubMed: 18026119]
58. Roseman RR, Morgan K, Mallin DR, Roberson R, Parnell TJ, Bornemann DJ, et al. Long-range repression by multiple polycomb group (PcG) proteins targeted by fusion to a defined DNA-binding domain in Drosophila. *Genetics*. 2001; 158:291–307. [PubMed: 11333237]
59. Boccuni P, MacGrogan D, Scandura JM, Nimer SD. The human L(3)MBT polycomb group protein is a transcriptional repressor and interacts physically and functionally with TEL (ETV6). *J Biol Chem*. 2003; 278:15412–15420. [PubMed: 12588862]
60. Wu S, Trievel RC, Rice JC. Human SFMBT is a transcriptional repressor protein that selectively binds the N-terminal tail of histone H3. *FEBS Lett*. 2007; 581:3289–3296. [PubMed: 17599839]
61. Francis NJ, Kingston RE, Woodcock CL. Chromatin compaction by a polycomb group protein complex. *Science*. 2004; 306:1574–1577. [PubMed: 15567868]
62. Maison C, Almouzni G. HP1 and the dynamics of heterochromatin maintenance. *Nat Rev Mol Cell Biol*. 2004; 5:296–304. [PubMed: 15071554]
63. Takada Y, Isono K, Shinga J, Turner JM, Kitamura H, Ohara O, et al. Mammalian Polycomb Scmh1 mediates exclusion of Polycomb complexes from the XY body in the pachytene spermatocytes. *Development*. 2007; 134:579–590. [PubMed: 17215307]
64. Oktaba K, Gutiérrez L, Gagneur J, Girardot C, Sengupta AK, Furlong EE, et al. Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. *Dev Cell*. 2008; 15:877–889. [PubMed: 18993116]
65. Stielow B, Sapetschnig A, Krüger I, Kunert N, Brehm A, Boutros M, et al. Identification of SUMO-dependent chromatin-associated transcriptional repression components by a genome-wide RNAi screen. *Mol Cell*. 2008; 29:742–754. [PubMed: 18374648]
66. Stielow B, Sapetschnig A, Wink C, Krüger I, Suske G. SUMO-modified Sp3 represses transcription by provoking local heterochromatic gene silencing. *EMBO Rep*. 2008; 9:899–906. [PubMed: 18617891]
67. Nishioka K, Rice JC, Sarma K, Erdjument-Bromage H, Werner J, Wang Y, et al. PR-Set7 is a nucleosome-specific methyltransferase that modifies lysine 20 of histone H4 and is associated with silent chromatin. *Mol Cell*. 2002; 9:1201–1213. [PubMed: 12086618]

68. Zhang H, Smolen GA, Palmer R, Christoforou A, van den Heuvel S, Haber DA. SUMO modification is required for in vivo Hox gene regulation by the *Caenorhabditis elegans* Polycomb group protein SOP-2. *Nat Genet.* 2004; 36:507–511. [PubMed: 15107848]
69. Guan XY, Xu J, Anzick SL, Zhang H, Trent JM, Meltzer PS. Hybrid selection of transcribed sequences from microdissected DNA: isolation of genes within amplified region at 20q11-q13.2 in breast cancer. *Cancer Res.* 1996; 56:3446–3450. [PubMed: 8758910]
70. Bench AJ, Nacheva EP, Hood TL, Holden JL, French L, Swanton S, et al. Chromosome 20 deletions in myeloid malignancies: reduction of the common deleted region, generation of a PAC/BAC contig identification of candidate genes UK Cancer Cytogenetics Group (UKCCG). *Oncogene.* 2000; 19:3902–3913. [PubMed: 10952764]
71. MacGrogan D, Kalakonda N, Alvarez S, Scandura JM, Boccuni P, Johansson B, et al. Structural integrity and expression of the L3MBTL gene in normal and malignant hematopoietic cells. *Genes Chromosomes Cancer.* 2004; 41:203–213. [PubMed: 15334543]
72. Bench AJ, Li J, Huntly BJ, Delabesse E, Fourouclas N, Hunt AR, et al. Characterization of the imprinted polycomb gene L3MBTL, a candidate 20q tumour suppressor gene, in patients with myeloid malignancies. *Br J Haematol.* 2004; 127:509–518. [PubMed: 15566354]
73. Li J, Bench AJ, Vassiliou GS, Fourouclas N, Ferguson-Smith AC, Green AR. Imprinting of the human L3MBTL gene, a polycomb family member located in a region of chromosome 20 deleted in human myeloid malignancies. *Proc Natl Acad Sci USA.* 2004; 101:7341–7346. [PubMed: 15123827]
74. Sundareshan TS, Prabhash K, Bapsy PP. Deletion of 6q23 as sole abnormality in acute myelocytic leukemia. *Cancer Genet Cytogenet.* 2003; 143:87–88. [PubMed: 12742161]
75. Northcott PA, Nakahara Y, Wu X, Feuk L, Ellison DW, Croul S, et al. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nat Genet.* 2009

Figure 1. Evolutionary relationship and domain organization of MBT proteins

On the left, the putative evolutionary relationship between MBT containing proteins is shown in a cladogram, based on primary sequence similarity. All known MBT proteins in the *C. elegans*, *D. melanogaster*, and *H. sapiens* genomes are indicated. On the right a schematic representation of domain organization in the same MBT proteins is shown. The different conserved domains are color-coded according to the legend. When known, the MBT repeat that binds to histone tails is indicated with a yellow asterisk.

Figure 2. Structural features of the MBT repeats

(A) The crystal structure of the 2nd MBT repeat of hL3MBTL is shown in complex with the ligand H4K20me2 (left) and compared with a similar structure obtained for the chromodomain of dHP1 in complex with H3K9me3 (right). The two different modes of binding, pocket insertion versus surface groove are illustrated. Figures were generated with UCSF chimera and POV raytracing software using atomic coordinates from PDB entries 2PQW and 1Q3L. Surface colors indicate hydrophobicity according to the Kyte-Doolittle scale, ranging from blue for the most hydrophilic to red for the most hydrophobic. (B) Schematic depiction of higher order organization of 2, 3, and 4 tandem MBT repeats, as found in SCM, L(3)MBT and SFMBT family proteins. Structurally homologous domains across families are in the same color. Known binding sites for histone peptides are marked with an arrow. The atomic coordinates used were from PDB entries 2R5A, 1OZ2 and 3F70.

Figure 3. Recruitment and function of MBT proteins

Putative mechanisms for the recruitment of MBT proteins to chromatin (left) and their repressive effect on transcription (right) are schematized. See text for details.

Table 1

Binding specificities of MBT proteins

Protein	Peptides	Method	Reference
dSCM-2MBT (aa 174–435, R277C)	H3K4me1 (aa 2–15)	ITC ^a	Grimm et al., 2007
	H3K9me1 (aa 2–15)		
	H3K27me1 (aa 18–35)		
	H3K36me1 (aa 28–44)		
	H4K20me1 (aa 12–27)		
hSCML2-2MBT^b	H1bme ^b	NMR ^c	Santiveri et al., 2008
	H3K4me ^b		
	H3K9me ^b		
	H3K27me ^b		
	H3K36me ^b		
	H4K20me ^b		
hL3MBTL1-3MBT	H1bK26me1/2 (aa 20–37)	<i>In vitro</i> pull down	Trojer et al., 2007
	H4K20me1/2 (aa 16–25)		
hL3MBTL1-3MBT (aa 200–522)	H4K20me1/2 (aa 17–25)	ITC	Min et al., 2007
	H1bK26me1/2 (aa 21–32)		
	H3K4me1/2		
	H3K9me1/2		
	H3K27me1/2		
hL3MBTL1-3MBT (aa 206–519)	H1.4K26me1/2 (aa 18–32)	Fluorescence polarization	Li et al., 2007
	H3K4me1/2 (aa 1–15)		
	H3K9me1/2 (aa 1–15)		
	H3Kme1 scrambled		
	H3K27me1/2 (aa 19–35)		
	H3K36me1/2 (aa 28–43)		
	H4K20me1/2 (aa 12–27)		
hL3MBTL2-4MBT (aa 170–625)	H4Kme1 scrambled	ITC	Guo et al., 2009
	H3K4me1		
	H3K9me1/2		
	H3K27me1/2		
dSFMBT-4MBT (aa 531–980)	H4K20me1/2 (aa 15–25)	Fluorescence polarization	Klymenko et al., 2006
	H3K9me1/2 (aa 1–15)		
	3K9me3 (aa 1–15)		
dSFMBT-4MBT (aa 535–977)	H4K20me1/2 (aa 12–27)	ITC Fluorescence polarization	Grimm et al., 2009
	H3K4me1/2 (aa 1–15)		
	H3K9me1/2 (aa 1–15)		

Protein	Peptides	Method	Reference
	H3K27me1/2 (aa 19–35)		
	H3K36me1/2 (aa 28–43)		
	H4K20me1/2 (aa 12–27, 17–23)		
	H4Kme1 scrambled		

^aITC, Isothermal Titration Calorimetry

^bThe authors do not indicate the degree of methylation of the peptides or the fragment of hSCML2 used

^cNMR, Nuclear Magnetic Resonance