


SET/MLL family proteins in hematopoiesis and leukemia

Weiwei Yang^{1,2} · Patricia Ernst¹ 

Received: 29 September 2016 / Accepted: 20 October 2016 / Published online: 31 October 2016
© The Japanese Society of Hematology (outside the USA) 2016

Abstract Accumulating recent evidence supports the notion that many enzymes that modify histones are fundamental players in normal hematopoiesis as well as hematologic malignancies, and represent an important new class of drug targets. Histone H3 lysine 4 (H3K4) methylation plays several distinct roles in gene expression and is modulated by specific methyltransferases and demethylases. Recent progress has been made clarifying the unique biological roles of the enzymes that carry out H3K4 methylation, yet a detailed understanding of H3K4 methylation states in various genomic contexts and the diverse functions of the enzymes that perform these methylation events is incomplete, but developing rapidly. In this review, we summarize and discuss the general mechanisms of H3K4 methylation, and how the six main enzymes from the SET/MLL family (responsible for H3K4me1/me2/me3) function in hematopoiesis and in hematologic malignancies.

Keywords AML · Leukemia · Histone methyltransferase · MLL1 · MLL2 · Wbp7 · H3K4 methylation

Histone H3, lysine 4 methylation in animals

Despite the discovery of histone proteins in the 1800s, we have only recently begun to appreciate the immense

regulatory role that they play in DNA replication, repair, nuclear organization, and gene expression. Post-translational modifications of all four core histones, H2A, H2B, H3, and H4, are a major determinant of gene expression. Furthermore, post-translational modifications are dynamic and connected to signaling pathways that respond to intracellular and extracellular cues. In this review, we focus on histone H3 lysine 4 methylation (H3K4me) in hematopoietic lineage cells, in both normal physiologic settings, and in hematologic malignancies.

What is the function of this particular H3 modification? First, H3K4 can be mono-, di-, or tri-methylated, with all three modifications co-existing and distributed distinctly, not just representing transition states. H3K4me3/2/1 is associated with euchromatin rather than heterochromatin but plays several roles in active chromatin. Genome-wide chromatin immunoprecipitation-sequencing experiments (ChIP-seq) have illustrated that H3K4me3 is enriched at transcription start sites (TSS) of active genes (Fig. 1) in an asymmetric manner [1–3]. Recent studies examining the breadth of the H3K4me3 enrichment (rather than the absolute enrichment) across the TSS of genes have indicated that transcriptional fidelity and elongation may be enhanced at these genes [4, 5].

H3K4me2 also associates with the TSS, but is distributed distinctly in the gene body depending on the class of genes assessed. For example, a meta-analysis of ChIP-seq data identified three classes of genes based on H3K4me2 distribution, with a particular gene-body-biased distribution identifying highly regulated tissue-specific genes (Fig. 1; [6]). In contrast, H3K4me1 is enriched at enhancer element distant from the promoter, representing a defining feature of these regulatory elements [7].

In addition, recent studies have illuminated a new class of poised, tissue-specific gene-harboring H3K4me1 that

✉ Patricia Ernst
patricia.ernst@ucdenver.edu

¹ Department of Pediatrics and Pharmacology, University of Colorado, Denver/Anschutz Medical Campus, Aurora, CO 80045, USA

² Department of Genetics, Geisel School of Medicine at Dartmouth, Hanover, NH 03755, USA

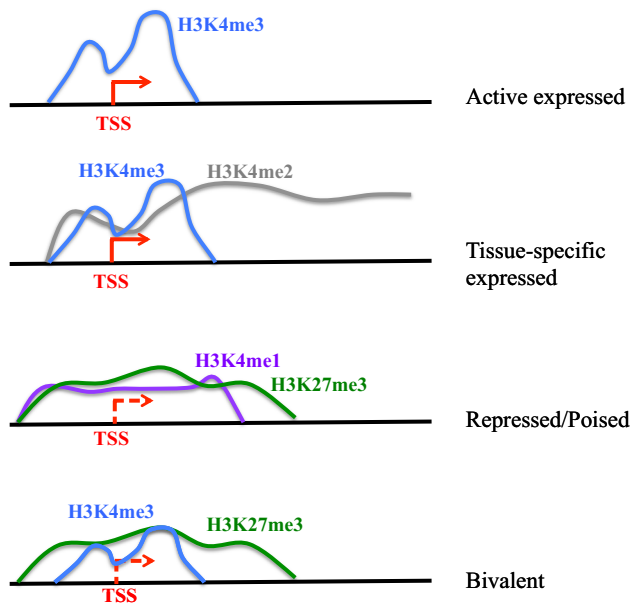


Fig. 1 H3K4 methylation patterns at four distinct categories of promoters. “Active expressed” gene promoters are enriched for H3K4me3 modification adjacent to the TSS, with the dip at the start site reflecting a nucleosome free region. “Tissue-specific expressed” genes are enriched for H3K4me2 from the promoter region into the gene body, and typically have an active expressed gene pattern of H3K4me3 as well. “Repressed/poised” gene promoters are associated with broad H3K4me1 across the entire promoter along with the H3K27me3 repressive mark; upon activation, H3K4me1 is converted to H3K4me3 and the repressive mark is replaced. “Bivalent” promoters are characterized by the coexistence of H3K27me3 and H3K4me3 and are primed for lineage-restricted expression or permanent repression. The *dashed arrow* at the TSS indicates no/low transcriptional output

extends broadly into the gene body (Fig. 1). These promoters are also enriched for H3K27me3 (a modification associated with silenced genes), but lack H3K4me3. These genes initially are not transcriptionally active, but undergo a conversion of H3K4 to the tri-methylated state upon differentiation and induction. This conversion then allows recruitment of H3K4me2/3-binding proteins that enhance gene expression, such as ING proteins [8].

Beyond the correlation with euchromatin, H3K4me3 is a defining feature of a regulatory state of a promoter termed “bivalent”. Originally defined in embryonic stem (ES) cells by the co-occurrence of H3K4me3 and H3K27me3 at histones surrounding the TSS, this combination of modifications reflects a state in which a gene is poised to become rapidly transcribed upon a lineage choice, but is currently not expressed or expressed at a low level [9]. The above two poised transcriptional states differ by the cell type in which they have been described as well as the enzymes and reader proteins involved; however, both may represent similar transitional states.

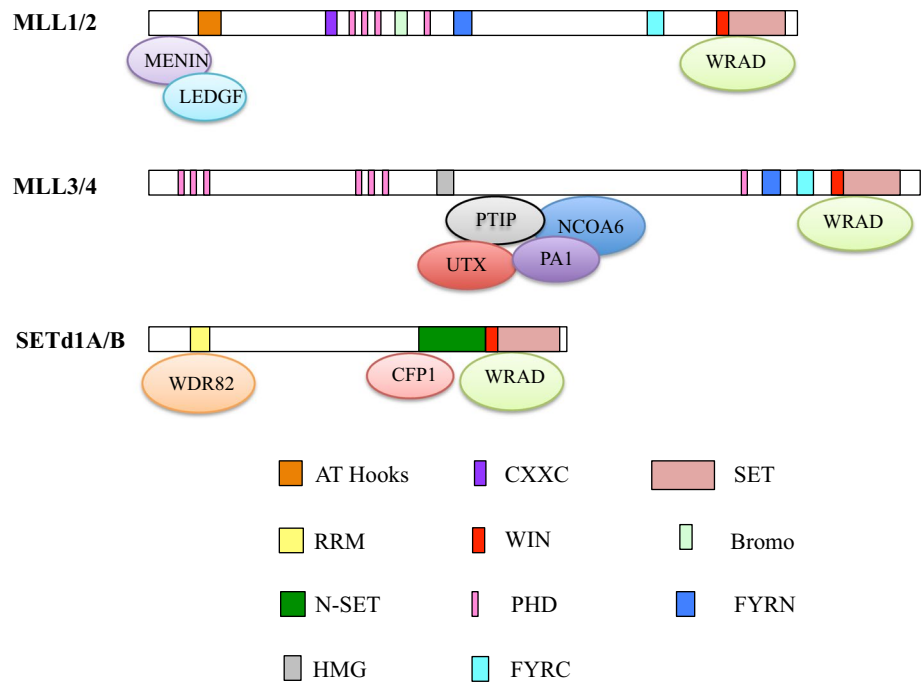
What do these “flavors” of chromatin do to gene regulation or other aspects of chromatin organization? Since H3K4me modification does not affect the charge or overall structure of the histone octamer, proteins that recognize the various methylation states in the histone tail have been sought (H3K4 “readers”). Many proteins that recognize each of the H3K4 methylation states have been identified, leading to a clearer picture of the variety of roles for H3K4 modifications [10]. For example, the plant homeodomain (PHD) domain of TAF3 (an integral subunit of the TFIID Pol II transcription initiation complex) binds H3K4me3 with high affinity, helping to recruit the pre-initiation complex to previously active TSSs [11]. In addition, connecting splicing to transcription has been attributed to the histone reader, CHD1, which binds H3K4me3 via its tandem chromo-domains. This interaction bridges key spliceosomal components to chromatin to regulate pre-mRNA-splicing processes [12]. In the tissue-specific repressed/poised promoters cited above [8], H3K4me1 prevents ING protein binding, but upon differentiation of myoblasts, conversion to H3K4me2/3 at these promoters allows for the binding of ING protein complexes that enable transcriptional activation. Finally, a non-transcriptional function of H3K4me3-enriched histone octamers is that B- and T-cell antigen-receptor recombination requires H3K4me3 recognition by the PHD domain of the RAG2 protein [13]. Binding to the antigen-receptor loci then enables gene rearrangement. The topic of chromatin readers interacting with H3K4 in its unmodified and methylated forms has been reviewed extensively [10, 14, 15], but these few examples illustrate that the distinct H3K4 modifications can recruit distinct activities that alter transcriptional status, chromatin access, transcriptional fidelity, or splicing outcome.

SET/MLL family histone methyltransferases share common enzymatic components but interact with distinct targeting components

The conserved *Su(var)3-9*, *Ezh2*, *Trithorax* (SET) domain in this family of proteins is similar to that of *S. cerevisiae* SET1 [16, 17]. SET1 is an H3K4 methyltransferase and is responsible for all three methylation states (mono-, di-, and tri-) of H3K4 [17–19]. There are three pairs of SET/MLL family members in mammals (Fig. 2): SETd1A/SETd1B, MLL1/MLL2, and MLL3/MLL4 [20]. Although the nomenclature has been confused between MLL2 and MLL4, we use here MLL2 for the protein encoded by the *Wbp7* or *Kmt2b* gene. MLL5, although harboring an SET domain, is enzymatically inactive [21] so is not included in this discussion.

Members of this SET/MLL family exhibit relatively moderate-to-weak enzymatic activity as recombinant domains *in vitro* but are markedly and variably enhanced by association with core subunits shared by the entire family. The

Fig. 2 SET/MLL family structure and components. Motifs and interacting partners unique to the paralog groups or those shared with all members are identified in the legend. The general structure of each paralog group is diagrammed with the members indicated at the left of each diagram



minimal subunit complex is comprised of WDR5, RbBP5, ASH2L, and DPY30 (WRAD, Fig. 2), which can enhance the histone methyltransferase (HMT) enzymatic activity of recombinant SET domains of all six family members [22–26]. Of note, most groups find a significant H3K4 mono- and di-methyl activity using recombinant proteins and histone peptide substrates, but some discrepancies exist when tri-methylation is assessed; these may be due to the use of bacterial versus eukaryote-produced complexes or due to the use of distinct substrates (peptides versus nucleosomes) in the assays. For example, the Cosgrove group tested each bacterial core SET/MLL complex (with WRAD) *in vitro* and found that MLL1/2 are predominantly mono- and di-methyltransferases and MLL3/4 act mostly as mono-methyltransferases, whereas SETd1A/B can catalyze all three methylation states similar to the yeast ortholog [25]. In addition to allosteric effects of WRAD on methyltransferase activity, the WDR5 component may help target the complex to chromatin. WDR5 interacts with histones [27] and lncRNAs [28], thus likely is involved in stabilizing chromatin interactions. In summary, interaction with these core complex components is clearly involved in determining the specificity of the individual enzymes, but how relative dependencies or modulations of the complex components *in vivo* translates to the distinct activities on target genes is a challenging question not easily answered with recombinant proteins.

Beyond the shared components, each of the SET/MLL members participates also in unique protein interactions. For example, the N-terminus of both MLL1 and MLL2 bind to a Menin/LEDGF subcomplex (Fig. 2). This subcomplex plays a major role in chromatin targeting through

recognition of H3K36me2 by LEDGF (Fig. 2) [29–31]. MLL3 and MLL4, on the other hand, were originally purified in the Activating Signal Cointegrator Complex (ASCOM), and in addition to not binding Menin/LEDGF, they bind to PTIP, UTX, PA1, and NCOA6 [32–35]. SETd1A and SETd1B have the unique partners WDR82 and CFP1, which do not interact with other SET/MLL enzymes (Fig. 2). WDR82 bridges the SET complexes to Ser5-phosphorylated Polymerase II, thus actively transcribed genes, and also contributes to the stability of the protein complex [36]. Interestingly, MLL1 and MLL2 have a domain (CXXC, Fig. 2) that binds to non-methylated CpGs, whereas CFP1 accomplishes a similar targeting function *in trans* for the SET complexes. These examples of unique components for the paralog groups are important, as targeting these protein interactions has been pursued pharmacologically and through knockout approaches (discussed below). Many additional protein interaction partners have been described, which are unique to individual members or ortholog groups, but these are either tissue-specific or not integral components of the corresponding complexes.

Distinct functions of SET/MLL family methyltransferases in hematopoiesis and leukemia

MLL1 and MLL2 functions in hematopoiesis

Loss-of-function studies in the mouse were initiated before it was even clear how many family members existed or

what biochemical activities these genes encoded. For example, the first mouse knockouts disrupting *Mlll* were published from 1995 to 2001 [37–39], followed by the generation of conditional knockout alleles due to the early lethality of the germline knockouts [40–42].

Germline knockout studies established the following principles: first, *Mlll* was indeed functionally conserved with its *Drosophila* counterpart, Trithorax, in that it maintained proper expression domains of several clustered homeodomain (*Hox*) genes [37]. This functional similarity also placed MLL1 in a well-described genetic network in which Polycomb orthologs were predicted to oppose its function [20]. Homozygous knockout embryos from the different laboratories all exhibited different penetrance in that lethality occurred from pre-implantation to E12.5, but all groups noted homeotic transformation of the axial skeleton and mis-expression as well as reduced expression of *Hox* cluster genes in mutant embryos or mice [37–39]. Second, reduced production of blood lineages was noted in these germline knockouts, mainly manifesting as small fetal livers, reduced colony forming units (CFU) from fetal liver and yolk sac cells, and reduced engraftment activity from aorta-gonad-mesonephros (AGM) cells of knockout embryos, where the earliest definitive hematopoietic stem cells (HSCs) arise [38, 43, 44]. However, none of these studies addressed whether reduced hematopoiesis reflected a cell-intrinsic function of MLL1, since knockout HSCs developed in a mutant environment in germline knockout embryos. Years later, newly engineered mouse models addressed these questions and more precisely defined tissue-specific functions.

The use of conditional deletion for *Mlll* naturally first focused on hematopoietic-specific functions [40, 41, 45]. The number and variety of fetal hematopoietic populations in Vav-Cre-mediated (developmentally regulated, hematopoietic-specific) knockout fetal liver cells implied that defective endothelial cells or other support cells contributed to the severity of hematopoietic phenotypes observed in the germline knockouts [38, 41, 43, 45]. Specifically, HSPCs are numerically and phenotypically normal in Vav-Cre-mediated *Mlll*^{-/-} mid-gestation fetal livers, whereas the corresponding assays using germline knockout fetal liver cells showed a significant reduction in HSPC numbers and function [38]. Nonetheless, even these fetal liver *Mlll*^{-/-} HSPCs do not at all engraft adult recipient animals [41, 45]. Thus, these studies demonstrated that *Mlll* was required in a hematopoietic-intrinsic manner for the development of adult-transplantable HSCs [41, 45].

Further studies using inducible knockout strategies revealed that *Mlll*^{-/-} HSCs exhibited many specific defects, including failure to maintain G₀, lack of self-renewal without loss of differentiation, widespread transcriptional deregulation, and complete failure to engraft

recipients. Interestingly, distinct alleles produced milder results (no reduction in steady-state adult bone-marrow populations), suggesting a difference in severity of the alleles [40, 41].

The role of the HMT activity in these severe *Mlll*^{-/-} hematopoietic phenotypes was addressed by deleting the SET domain harboring the catalytic site. These knockout animals survive as homozygous mutants, and the initial report showed altered *Hox* gene expression in embryos and some gene-specific reductions in H3K4me1 [46]. Remarkably, adult homozygous mutant animals exhibited no overt hematopoietic or developmental phenotypes. In the hematopoietic system, no changes in gene expression, histone modifications, or function of HSCs could be demonstrated in animals lacking the MLL1 HMT activity [47]. These data revealed that although MLL1 plays an important and non-redundant role in maintaining hematopoietic stem and progenitor cells, it does so with activities other than its HMT activity.

Although *Mlll* does not appear to be involved in the homeostasis of mature T cells, several lines of evidence show that it plays a role in immunity and T-cell function. *Mlll* heterozygous animals on a Balb/C background exhibited initially normal effector T_{H2} responses and gene expression, but failed to maintain expression of T_{H2} cytokines and *Gata3* expression upon antigen recall [48]. Furthermore, animals in which *Mlll* is deleted only in CD4⁺ T cells fail to control helminth infections and fail to maintain IL-4 expression (Colby Zaph, Monash University, personal communication). The precise genes and mechanisms controlling these responses in T cells have not yet been elucidated. Given the wealth of recent information regarding epigenetic regulation of T-cell differentiation [49], this avenue of investigation promises to yield important mechanistic insights.

MLL2, such as MLL1, is expressed broadly. The two genes are paralogs likely originating from gene duplication, and both are considered Trithorax orthologs due to the overall similarity and domain structure [50]. In contrast to MLL1, MLL2 is never involved in chromosomal translocations, and in fact, the N-terminus of MLL2 cannot substitute for MLL1 in leukemia oncoproteins, likely due to differences in the CXXC domain [51–53]. Germline deletions in *Mll2* result in delayed development from at least E8 and lethality by E10.5 with neural tube defects and widespread apoptosis. MLL2 expression increases during oocyte maturation, and is in fact required for oocyte survival; remarkably, MLL2-deficient oocytes exhibit global loss of H3K4me3 and H3K4me2 and a slight increase in H3K4me1 [54]. This is the first tissue in which MLL2 was shown to be the dominant H3K4me2/3 methyltransferase, thus co-expression of other family members in oocytes apparently cannot compensate for loss of MLL2. Despite

this major role in oocytes, complete deletion of *Mll2* after E10.5 had no effect on the hematopoietic system or other organ homeostasis, as determined using a tamoxifen-inducible Rosa-ER^{T2}-Cre system [50].

One additional hematopoietic cell type provided a significant mechanistic insight into how *Mll2* functions in vivo and how it regulates target genes. Using the Rosa-ER^{T2}-Cre animals described above, *Mll2*^{-/-} bone marrow was used to differentiate macrophages in vitro, which were phenotypically indistinguishable from their wild-type counterparts. When challenged with lipopolysaccharide (LPS), which activates TLR4 signaling, these cells exhibited a specific defect in the induction of NFκB target genes. Surprisingly, this was not due to a direct effect on NFκB-dependent genes, but rather the specific underexpression of *Pigp*, a gene required for the complex that adds glycosylphosphatidylinositol (GPI) anchors to transmembrane proteins. Without this enzyme, CD14 was reduced at the cell surface and without its function as an LPS co-receptor, and TLR4 signaling was greatly reduced [55]. *Pigp* and several other MLL2-regulated genes were identified in this study, and all exhibited reduced H3K4me3 enrichment around their promoters, but a global reduction in H3K4me3 was not reported. Interestingly, most of these promoters also exhibited an increase in H3K27me3, a mark associated with repressed and bivalent genes, suggesting that one role of the H3K4me3 enrichment is to resist Polycomb complex-mediated repression. On the other hand, many other genes that were found to be hypo-H3K4 methylated were not changed in expression, suggesting that the role of MLL2 in maintaining its target genes is in some cases HMT-dependent and in other cases through other transcriptional effects of MLL2. Alternatively, the expression of some genes may simply not be as sensitive to lack of H3K4me3 depletion at the promoter. This latter concept is supported by the fact that 90 % of the genes exhibiting reduced H3K4me3 were not altered in transcript level [55].

MLL1 and MLL2 functions in leukemia

As discussed above, despite the similarity in gene structure and overlapping gene expression patterns, MLL1 and MLL2 have diversified to function differently (with respect to the requirement for HMT activity) and do regulate distinct sets of genes. These differences may underlie the fact that only MLL1 is involved in chromosomal translocations in leukemia. The mechanisms by which MLL-fusion oncoproteins transform cells have been reviewed extensively [56–59], but the contribution of the non-translocated, wild-type allele remains controversial. Several early studies suggested that the remaining wild-type MLL1 molecule was required for maintaining MLL-AF9 initiated leukemia or that targeting the HMT activity of MLL1 would be

therapeutically beneficial [60, 61]. However, studies using the HMT-inactive SET domain-deleted *Mll1* bone-marrow cells demonstrated that leukemia initiated by MLL-fusion oncoproteins was absolutely unimpeded in primary or secondary recipients of leukemia cells [47]. Furthermore, CRISPR-Cas9-mediated deletion of the SET domain in transformed cells did not impair their growth, suggesting that the HMT activity or possibly any activity of endogenous MLL1 is not needed for the fusion oncoproteins to transform cells [62]. The role of MLL2 in leukemia thus far is unknown, as is the issue of redundancy with MLL1 in MLL-AF9 leukemia.

MLL3 and MLL4 in hematopoiesis, leukemia, and lymphoma

MLL3/MLL4 in mammals was identified as HMTs in the ASCOM complex, which is a coactivator complex associated with nuclear receptors [32], but have since been demonstrated to interact with a wide variety of transcription factors in a similar complex (reviewed in [63]). Knockout of either gene is embryonic or perinatal lethal, but there is also evidence for redundancy between these genes from studies of adipogenesis and myogenesis, since co-deletion of both was required to observe a global decrease in H3K4me1. Thus, in preadipocytes, MLL3 and MLL4 were responsible for the majority of H3K4me1/2 but not me3 [34]. It is not yet clear if both enzymes are collectively responsible for all H3K4me1 at enhancers in all tissues. Finally, their enzymatic activity has also been associated with the repressive H3K4me1 at promoters discussed above [8].

MLL3 was also cloned as a region deleted in 7q- acute myelogenous leukemia (AML) [64], and interest in MLL3/MLL4 roles in hematopoiesis was revived when MLL4 was discovered to be mutated in 30–90 % of diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) [65, 66]. Most genomic alterations were predicted to be loss-of-function mutations, with about half mono-allelic and half bi-allelic. Therefore, the role of MLL4 was explicitly tested as a tumor suppressor in several murine leukemia models. First, knockdown of MLL4 in the context of a Vav-Bcl2 transgene or deletion of MLL4 in an activation-induced cytidine deaminase (AID) overexpression model accelerated lymphomagenesis, and even B-cell specific deletion of MLL4 alone resulted in lymphoma development [67, 68]. In normal B-cell differentiation, loss of MLL4 increased steady-state transitional B-cell numbers, enhanced germinal center formation, and enhanced proliferation in response to CD40 stimulation [66, 68]. Furthermore, in B-cells, the deletion of MLL4 is sufficient to observe a global reduction in H3K4me3/2/1 [68]. Therefore, MLL4 functions in a non-redundant manner to suppress mature B-cell proliferation act as a tumor suppressor in FL and DLBCL.

The role of the MLL3/MLL4 proteins in hematopoiesis appears generally similar as either knockdown of MLL3 or knockout of MLL4 in hematopoietic stem/multipotent progenitor cells (HSPCs) results in increased HSPC numbers [69, 70]. However, MLL4^{-/-} HSPCs exhibit reduced engraftment in secondary recipients, accompanied by high levels of reactive oxygen species [69]. These defects were attributed to the loss of MLL4-dependent genes that regulate and protect from oxidative stress. In the case of AML driven by an MLL-fusion oncoprotein, MLL4 loss limited leukemogenesis [69], in contrast to the role of MLL3 in the 7q- AML model of tumor repression [70] or role in FL and DLBCL discussed above. These distinct tumor enabling or tumor suppressive roles likely reflect the different cellular contexts and target genes in AML versus lymphoma.

SETd1A/SETd1B in hematopoiesis and leukemia

Despite the fact that SETd1A and SETd1B are likely to be the most broadly acting, transcription-coupled H3K4 methyltransferases, very little is known about the *in vivo* functions of these enzymes compared to their paralog MLL1 until recently. Both SETd1A and SETd1B knockout embryos develop to blastocysts and implant, but SETd1A blastocysts exhibit a defective inner cell mass and fail to produce ES cell clones. On the other hand, SETd1B null embryos exhibited growth defects from E7.5 and die before E11.5. In ES cells engineered with inducible (Rosa-ERT²) Cre, SETd1A also exhibited a more severe phenotype, with growth arrest and cell death occurring as soon as protein reduction was observed. Commensurate with reduced SETd1A protein reduction, a global reduction in H3K4me_{3/2/1} was observed in ES cells, but the mark was not completely depleted, suggesting that the remaining H3K4 methylation in ES cells is due to SETd1B or other methyltransferases. However, parallel deletion of SETd1B in the same study did not result in perceivable global H3K4 methylation changes [71]. These data illustrated that the SET proteins certainly play distinct roles in gene expression, but may also exhibit some redundancy in maintaining global H3K4me₃ levels.

The Huang group has focused on the role of SETd1A in hematopoietic cells using an independently produced conditional knockout allele. Using the Mx1-Cre strain to inducibly delete SETd1A in bone-marrow cells, they noted reduced B-cell numbers, particularly inducing a partial block in differentiation from the pro-B to pre-B-cell transitions. It is not clear whether H3K4 methylation was globally reduced from these studies, but reduction of H3K4me₃ at important B-cell regulators and the immunoglobulin heavy chain locus was suggested to underlie the block in differentiation [72]. Similarly, erythroid-specific SETd1A knockout partially blocked erythropoiesis, resulting in

a mild anemia. Reduced expression of lineage regulators, such as *Gata1* and *Tall1*, was noted, as was reduced H3K4me₃ at the promoters of these genes. Although several sequence-specific transcriptional regulators have been implicated in the recruitment of SETd1A, including USF1 in this study, the rules for targeting this complex to specific lineage target genes are not yet clear [73].

Knockouts of several unique and shared SET/MLL complex components have been insightful, particularly in comparison with knockouts of the core HMTs. For example, by knocking out the CFP1 (also called CXXC protein, binds non-methylated CpGs) in murine ES cells, H3K4me₃ peak reduction was observed at about 50 % of TSSs. Furthermore, H3K4me₃ is acquired in nuclear territories and regions of the genome in which it does not appear in wild-type cells, resulting in ectopic transcription [74, 75]. Therefore, CFP1 is critical for directing H3K4 methylation activity to proper genomic locations. In addition, a global reduction in DNA methylation was also observed, most likely due to reduced expression of DNMT1 [76]. Differentiation cannot occur in these CFP1^{-/-} ES cells, but rescue of the knockout ES cells with different CFP1 mutants indicates that either the PHD domain binding to H3K4me_{1/2/3} or the CpG binding activity are required to restore H3K4me distribution and ES differentiation [75]. One of the remarkable observations from CFP1^{-/-} ES cells is that although the most highly expressed genes are those that suffer the greatest H3K4me₃ loss, yet significant loss of transcription at those loci did not occur [74]. Thus, a proper targeting of SETd1A/B complexes is required for differentiation and embryo development, but the relationship between transcription and H3K4me₃ is not a simple one.

Specifically, within the hematopoietic system, knockout of CFP1 using the Mx1-Cre inducible transgene resulted in bone-marrow failure and animal lethality within 2 weeks, loss of lineage-restricted progenitors, and mature cells. Hematopoietic stem cells and primitive progenitors, however, expand in the absence of CFP1 [77]. Interestingly, deletion of SETd1A in HSPCs revealed a similar phenotype. Upon SETd1A deletion, an accumulation of short-term HSCs and multipotent progenitors (MPPs) was observed, but these cells were not able to engraft secondary recipients. Cell-cycle analysis of SETd1A^{-/-} HSCs suggested an accumulation in the G₁ phase of the cell cycle, consistent with the overall reduction in proliferation observed by Chun et al. ([72] and Claudia Waskow, Technical University of Dresden, personal communication).

Conditional knockout of the WRAD component, DPY30, resulted in both shared and distinct phenotypes comparing to the SETd1A and CFP1 models above, likely due to the fact that DPY30 interacts with all six MLL/SET family members. In contrast to CFP1 loss, the overall levels of H3K4me_{3/2/1} were significantly reduced in

bone-marrow HSPCs. However, Mx1-Cre-mediated deletion resulted in bone-marrow failure and animal death in the same time frame as CFP1 loss and a significant accumulation of LT-HSCs and ST-HSCs in *Dpy30*^{-/-} bone marrow or cells in primary recipient bone marrow [78]. In this conditional knockout, however, neither abnormal cell proliferation nor decreased HSPC apoptosis were observed. Rather, the authors observe a block in HSC differentiation based on the inability of *Dpy30*^{-/-} bone marrow to yield differentiated colonies in vitro or to differentiate post-transplant in vivo. Furthermore, despite the accumulation in HSCs in conditionally *Dpy30*^{-/-} bone marrow, in a chimeric setting, these cells fail to contribute long term to even the HSPC pool [78]. Gene expression studies reveal downregulation of HSC genes by gene-set enrichment analysis, but the identity of those genes is very distinct from those deregulated in the *Mll1* knockout, which is the most comparably analyzed SET/MLL family member. The fact that either CFP1, SETd1A, or DPY30 loss each produces an increase in HSPCs distinguishes all three from the *Mll1* or *Mll2* hematopoietic knockouts, as does the limited gene expression data. Despite the overlapping enzymatic activity, it is very clear that the regulation and gene-specific action of these two paralog groups are very distinct from each other.

Discussion and future perspectives

Despite similar enzymatic activity and C-terminal cofactors, the overall domain structure, and widespread expression, it is now clear that each of the SET/MLL family members performs unique biological functions in vivo. While this may not be surprising, it does present challenges if pharmacologic inhibition of a unique family member is desired. Although subtle features of the HMT activity requirements may be exploited for relatively selective targeting, it is still not clear in many cases whether the HMT activity of each of the family members is the key activity that should be targeted. Nonetheless, the powerful combination of biochemical/structural analyses and in vivo manipulation will likely clarify which activities of these protein complexes are critical in the particular biological setting. However, given the large size and wide variety of protein interactions of all of these complexes nucleated by the SET/MLL family member, this process will likely require significant investment.

Gene-targeting studies of some of the associated WRAD and N-terminal-interacting proteins have provided evidence that all in vivo functions do not necessarily depend on these interacting partners. For example, a comparison of the CFP1, DPY30, and SETd1A loss-of-function models discussed above suggests that the loss of any of these proteins

results in overlapping phenotypes in HSPCs. One might have predicted that *Dpy30*^{-/-} bone marrow would exhibit a phenotype representing the loss of all 6 SET/MLL family members, but instead this knockout resembles most the SETd1A knockout ([40, 72, 78], Claudia Waskow, personal communication). These observations are consistent with the fact that the HMT activity of MLL1 is not required for its role in HSPCs, thus may not be affected by DPY30 loss [47]. Similarly, hematopoietic-specific *Menin* knockout is not as severe as the corresponding *Mll1* knockout [79], suggesting that this physical interaction is not required for all in vivo functions.

With respect to gene regulation, some mechanistic generalities can be made. Most evidences in multiple tissues, including the hematopoietic system, implicate the MLL3/MLL4 proteins in performing H3K4me1 globally, which are predominantly associated with enhancer function, as well as the recently described developmentally poised gene class (Fig. 1). In contrast, the MLL1/MLL2 and SETd1A/SETd1B families have been associated with H3K4me2/3 around TSSs, with the individual protein playing the dominant role highly dependent on the tissue surveyed. However, global and gene-specific histone modification changes in the various knockout settings should even be interpreted with caution. For example, the presence or absence of H3K4 demethylases in the individual tissues surveyed may have an enormous impact on the final H3K4me status. In addition, if H3K4mono methylases are deleted, the observation of H3K4me2/me3 reduction may simply reflect the lack of an appropriate substrate for the di- and tri-methylases to act upon, thus an indirect effect may be misinterpreted as a direct effect on H3K4me2/3 activity.

One important question to answer in the future is how each of these complexes is targeted to act at specific loci in the genome, and whether there are scenarios in which generic maintenance of the histone mark is important for maintaining gene expression can be accomplished by multiple enzymes. Although a few cases of tissue-specific transcriptional regulators recruiting individual complexes to particular promoter/enhancer elements have been reported, this aspect of SET/MLL protein function remains an area of active investigation. Given the number of chromatin-binding motifs in all of the family members (Fig. 2), the targeting of SET/MLL complexes even within a particular cell type is likely combinatorial between recruitment by tissue-specific transcriptional regulators and localized chromatin modifications, as exemplified by the non-methyl CpG binding activities.

As far as the individual roles of these family members in hematologic malignancies, MLL1 and MLL3/MLL4 are the members for which there is a significant body of the literature showing their involvement. MLL1 is disrupted by chromosomal translocations, MLL3 deleted in 7q- AML,

and MLL4 acting as a tumor suppressor in lymphoma. Nonetheless, it is likely that all family members participate as contributors in other hematologic malignancies, but the specific roles are just beginning to be worked out using the genetic models discussed here.

Acknowledgements We would like to thank the members of our lab, Claudia Waskow, Hao Jiang, Etienne Danis, and Simone Riedel for critical comments. This work was partially supported by Grants from the NIH (OD019716 and HL090036) and a grant from the Alex's Lemonade Stand Foundation.

References

- Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, et al. Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell*. 2005;120(2):169–81.
- Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007;39(3):311–8.
- Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, et al. High-resolution profiling of histone methylations in the human genome. *Cell*. 2007;129(4):823–37.
- Chen K, Chen Z, Wu D, Zhang L, Lin X, Su J, et al. Broad H3K4me3 is associated with increased transcription elongation and enhancer activity at tumor-suppressor genes. *Nat Genet*. 2015;47(10):1149–57.
- Benayoun BA, Pollina EA, Ucar D, Mahmoudi S, Karra K, Wong ED, et al. H3K4me3 breadth is linked to cell identity and transcriptional consistency. *Cell*. 2014;158(3):673–88.
- Pekowska A, Benoukraf T, Ferrier P, Spicuglia S. A unique H3K4me2 profile marks tissue-specific gene regulation. *Genome Res*. 2010;20(11):1493–502.
- Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, Wysocka J. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature*. 2011;470(7333):279–83.
- Cheng J, Blum R, Bowman C, Hu D, Shilatifard A, Shen S, et al. A role for H3K4 monomethylation in gene repression and partitioning of chromatin readers. *Mol Cell*. 2014;53(6):979–92.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell*. 2006;125(2):315–26.
- Yun M, Wu J, Workman JL, Li B. Readers of histone modifications. *Cell Res*. 2011;21(4):564–78.
- Vermeulen M, Mulder KW, Denissov S, Pijnappel WW, van Schaik FM, Varier RA, et al. Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell*. 2007;131(1):58–69.
- Sims RJ 3rd, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, et al. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. *Mol Cell*. 2007;28(4):665–76.
- Ji Y, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell*. 2010;141(3):419–31.
- Gatchalian J, Kutateladze TG. PHD fingers as histone readers. In: Zhou M-M, editor. *Histone recognition*. Berlin: Springer International Publishing; 2015.
- Andrews FH, Strahl BD, Kutateladze TG. Insights into newly discovered marks and readers of epigenetic information. *Nat Chem Biol*. 2016;12(9):662–8.
- Nislow C, Ray E, Pillus L. SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Mol Biol Cell*. 1997;8(12):2421–36.
- Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, Dent SY, et al. Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev*. 2001;15(24):3286–95.
- Krogan NJ, Dover J, Khorrami S, Greenblatt JF, Schneider J, Johnston M, et al. COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem*. 2002;277(13):10753–5.
- Nagy PL, Griesenbeck J, Kornberg RD, Cleary ML. A trithorax-group complex purified from *Saccharomyces cerevisiae* is required for methylation of histone H3. *Proc Natl Acad Sci USA*. 2002;99(1):90–4.
- Schuettengruber B, Martinez AM, Iovino N, Cavalli G. Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol*. 2011;12(12):799–814.
- Sebastian S, Sreenivas P, Sambasivan R, Cheedipudi S, Kandalla P, Pavlath GK, et al. MLL5, a trithorax homolog, indirectly regulates H3K4 methylation, represses cyclin A2 expression, and promotes myogenic differentiation. *Proc Natl Acad Sci USA*. 2009;106(12):4719–24.
- Dou Y, Milne TA, Ruthenburg AJ, Lee S, Lee JW, Verdine GL, et al. Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nat Struct Mol Biol*. 2006;13(8):713–9.
- Patel A, Dharmarajan V, Vought VE, Cosgrove MS. On the mechanism of multiple lysine methylation by the human mixed lineage leukemia protein-1 (MLL1) core complex. *J Biol Chem*. 2009;284(36):24242–56.
- Jiang H, Shukla A, Wang X, Chen WY, Bernstein BE, Roeder RG. Role for Dpy-30 in ES cell-fate specification by regulation of H3K4 methylation within bivalent domains. *Cell*. 2011;144(4):513–25.
- Shinsky SA, Monteith KE, Viggiano S, Cosgrove MS. Biochemical reconstitution and phylogenetic comparison of human SET1 family core complexes involved in histone methylation. *J Biol Chem*. 2015;290(10):6361–75.
- Zhang Y, Mittal A, Reid J, Reich S, Gamblin SJ, Wilson JR. Evolving catalytic properties of the MLL family SET domain. *Structure*. 2015;23(10):1921–33.
- Wysocka J, Swigut T, Milne TA, Dou Y, Zhang X, Burlingame AL, et al. WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development. *Cell*. 2005;121(6):859–72.
- Wang KC, Yang YW, Liu B, Sanyal A, Corces-Zimmerman R, Chen Y, et al. A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature*. 2011;472(7341):120–4.
- Hughes CM, Rozenblatt-Rosen O, Milne TA, Copeland TD, Levine SS, Lee JC, et al. Menin associates with a trithorax family histone methyltransferase complex and with the hox8 locus. *Mol Cell*. 2004;13(4):587–97.
- Yokoyama A, Cleary ML. Menin critically links MLL proteins with LEDGF on cancer-associated target genes. *Cancer Cell*. 2008;14(1):36–46.
- Zhu L, Li Q, Wong SH, Huang M, Klein BJ, Shen J, et al. ASH1L links histone H3 lysine 36 dimethylation to MLL leukemia. *Cancer Discov*. 2016;6(7):770–83.
- Goo YH, Sohn YC, Kim DH, Kim SW, Kang MJ, Jung DJ, et al. Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. *Mol Cell Biol*. 2003;23(1):140–9.
- Cho YW, Hong T, Hong S, Guo H, Yu H, Kim D, et al. PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem*. 2007;282(28):20395–406.

34. Lee JE, Wang C, Xu S, Cho YW, Wang L, Feng X, et al. H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife*. 2013;2:e01503.
35. Patel SR, Kim D, Levitan I, Dressler GR. The BRCT-domain containing protein PTIP links PAX2 to a histone H3, lysine 4 methyltransferase complex. *Dev Cell*. 2007;13(4):580–92.
36. Lee JH, Skalnik DG. Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. *Mol Cell Biol*. 2008;28(2):609–18.
37. Yu BD, Hess JL, Horning SE, Brown GA, Korsmeyer SJ. Altered Hox expression and segmental identity in Mll-mutant mice. *Nature*. 1995;378(6556):505–8.
38. Yagi H, Deguchi K, Aono A, Tani Y, Kishimoto T, Komori T. Growth disturbance in fetal liver hematopoiesis of Mll-mutant mice. *Blood*. 1998;92(1):108–17.
39. Ayton P, Sneddon SF, Palmer DB, Rosewell IR, Owen MJ, Young B, et al. Truncation of the Mll gene in exon 5 by gene targeting leads to early preimplantation lethality of homozygous embryos. *Genesis*. 2001;30(4):201–12.
40. Jude CD, Climer L, Xu D, Artinger E, Fisher JK, Ernst P. Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors. *Cell Stem Cell*. 2007;1(3):324–37.
41. McMahon KA, Hiew SY, Hadjur S, Veiga-Fernandes H, Menzel U, Price AJ, et al. Mll has a critical role in fetal and adult hematopoietic stem cell self-renewal. *Cell Stem Cell*. 2007;1(3):338–45.
42. Denissov S, Hofmeister H, Marks H, Kranz A, Ciotta G, Singh S, et al. Mll2 is required for H3K4 trimethylation on bivalent promoters in embryonic stem cells, whereas Mll1 is redundant. *Development*. 2014;141(3):526–37.
43. Hess JL, Yu BD, Li B, Hanson R, Korsmeyer SJ. Defects in yolk sac hematopoiesis in Mll-null embryos. *Blood*. 1997;90(5):1799–806.
44. Ernst P, Fisher JK, Avery W, Wade S, Foy D, Korsmeyer SJ. Definitive hematopoiesis requires the mixed-lineage leukemia gene. *Dev Cell*. 2004;6(3):437–43.
45. Gan T, Jude CD, Zaffuto K, Ernst P. Developmentally induced Mll1 loss reveals defects in postnatal haematopoiesis. *Leukemia*. 2010;24(10):1732–41.
46. Terranova R, Agherbi H, Boned A, Meresse S, Djabali M. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci USA*. 2006;103(17):6629–34.
47. Mishra BP, Zaffuto KM, Artinger EL, Org T, Mikkola HK, Cheng C, et al. The histone methyltransferase activity of MLL1 is dispensable for hematopoiesis and leukemogenesis. *Cell Rep*. 2014;7(4):1239–47.
48. Yamashita M, Hirahara K, Shinnakasu R, Hosokawa H, Norikane S, Kimura MY, et al. Crucial role of MLL for the maintenance of memory T helper type 2 cell responses. *Immunity*. 2006;24(5):611–22.
49. Onodera A, Nakayama T. Epigenetics of T cells regulated by Polycomb/Trithorax molecules. *Trends Mol Med*. 2015;21(5):330–40.
50. Glaser S, Schaft J, Lubitz S, Vintersten K, van der Hoeven F, Tufteland KR, et al. Multiple epigenetic maintenance factors implicated by the loss of Mll2 in mouse development. *Development*. 2006;133(8):1423–32.
51. Bach C, Mueller D, Buhl S, Garcia-Cuellar MP, Slany RK. Alterations of the CxxC domain preclude oncogenic activation of mixed-lineage leukemia 2. *Oncogene*. 2009;28(6):815–23.
52. Risner LE, Kuntimaddi A, Lokken AA, Achille NJ, Birch NW, Schoenfelt K, et al. Functional specificity of CpG DNA-binding CXXC domains in mixed lineage leukemia. *J Biol Chem*. 2013;288(41):29901–10.
53. Muntean AG, Tan J, Sitwala K, Huang Y, Bronstein J, Connelly JA, et al. The PAF complex synergizes with MLL fusion proteins at HOX loci to promote leukemogenesis. *Cancer Cell*. 2010;17(6):609–21.
54. Andreu-Vieyra CV, Chen R, Agno JE, Glaser S, Anastassiadis K, Stewart AF, et al. MLL2 is required in oocytes for bulk histone 3 lysine 4 trimethylation and transcriptional silencing. *PLoS Biol*. 2010;8(8):e1000453.
55. Austenaa L, Barozzi I, Chronowska A, Termanini A, Ostuni R, Prosperini E, et al. The histone methyltransferase Wbp7 controls macrophage function through GPI glycolipid anchor synthesis. *Immunity*. 2012;36(4):572–85.
56. Bach C, Slany RK. Molecular pathology of mixed-lineage leukemia. *Future Oncol*. 2009;5(8):1271–81.
57. Li BE, Ernst P. Two decades of leukemia oncogene epistasis: the MLL1 paradigm for epigenetic deregulation in leukemia. *Exp Hematol*. 2014;42(12):995–1012.
58. Chen CW, Armstrong SA. Targeting DOT1L and HOX gene expression in MLL-rearranged leukemia and beyond. *Exp Hematol*. 2015;43(8):673–84.
59. Marschalek R. MLL leukemia and future treatment strategies. *Arch Pharm (Weinheim)*. 2015;348(4):221–8.
60. Thiel AT, Blessington P, Zou T, Feather D, Wu X, Yan J, et al. MLL-AF9-induced leukemogenesis requires coexpression of the wild-type Mll allele. *Cancer Cell*. 2010;17(2):148–59.
61. Cao F, Townsend EC, Karatas H, Xu J, Li L, Lee S, et al. Targeting MLL1 H3K4 methyltransferase activity in mixed-lineage leukemia. *Mol Cell*. 2014;53(2):247–61.
62. Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol*. 2015;33(6):661–7.
63. Lee S, Roeder RG, Lee JW. Roles of histone H3-lysine 4 methyltransferase complexes in NR-mediated gene transcription. *Prog Mol Biol Transl Sci*. 2009;87:343–82.
64. Ruault M, Brun ME, Ventura M, Roizes G, De Sario A. MLL3, a new human member of the TRX/MLL gene family, maps to 7q36, a chromosome region frequently deleted in myeloid leukemia. *Gene*. 2002;284(1–2):73–81.
65. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature*. 2011;476(7360):298–303.
66. Pasqualucci L, Trifonov V, Fabbri G, Ma J, Rossi D, Chiarenza A, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. *Nat Genet*. 2011;43(9):830–7.
67. Ortega-Molina A, Boss IW, Canela A, Pan H, Jiang Y, Zhao C, et al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med*. 2015;21(10):1199–208.
68. Zhang J, Dominguez-Sola D, Hussein S, Lee JE, Holmes AB, Bansal M, et al. Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med*. 2015;21(10):1190–8.
69. Santos MA, Faryabi RB, Ergen AV, Day AM, Malhowski A, Canela A, et al. DNA-damage-induced differentiation of leukaemic cells as an anti-cancer barrier. *Nature*. 2014;514(7520):107–11.
70. Chen C, Liu Y, Rappaport AR, Kitzing T, Schultz N, Zhao Z, et al. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell*. 2014;25(5):652–65.
71. Bledau AS, Schmidt K, Neumann K, Hill U, Ciotta G, Gupta A, et al. The H3K4 methyltransferase Setd1a is first required at the epiblast stage, whereas Setd1b becomes essential after gastrulation. *Development*. 2014;141(5):1022–35.
72. Tusi BK, Deng C, Salz T, Zeumer L, Li Y, So CW, et al. Setd1a regulates progenitor B-cell-to-precursor B-cell development through histone H3 lysine 4 trimethylation and Ig heavy-chain rearrangement. *FASEB J*. 2015;29(4):1505–15.

73. Li Y, Schulz VP, Deng C, Li G, Shen Y, Tusi BK, et al. Setd1a and NURF mediate chromatin dynamics and gene regulation during erythroid lineage commitment and differentiation. *Nucleic Acids Res.* 2016;44(15):7173–88.
74. Clouaire T, Webb S, Skene P, Illingworth R, Kerr A, Andrews R, et al. Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev.* 2012;26(15):1714–28.
75. Mahadevan J, Skalnik DG. Efficient differentiation of murine embryonic stem cells requires the binding of CXXC finger protein 1 to DNA or methylated histone H3-Lys4. *Gene.* 2016;. doi:[10.1016/j.gene.2016.08.048](https://doi.org/10.1016/j.gene.2016.08.048) [**Epub ahead of print**].
76. Carlone DL, Lee JH, Young SR, Dobrota E, Butler JS, Ruiz J, et al. Reduced genomic cytosine methylation and defective cellular differentiation in embryonic stem cells lacking CpG binding protein. *Mol Cell Biol.* 2005;25(12):4881–91.
77. Chun KT, Li B, Dobrota E, Tate C, Lee JH, Khan S, et al. The epigenetic regulator CXXC finger protein 1 is essential for murine hematopoiesis. *PLoS One.* 2014;9(12):e113745.
78. Yang Z, Shah K, Khodadadi-Jamayran A, Jiang H. Dpy30 is critical for maintaining the identity and function of adult hematopoietic stem cells. *J Exp Med.* 2016 [**Epub ahead of print**].
79. Li BE, Gan T, Meyerson M, Rabbitts TH, Ernst P. Distinct pathways regulated by menin and by MLL1 in hematopoietic stem cells and developing B cells. *Blood.* 2013;122(12):2039–46.