Not All H3K4 Methyllations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification

Highlights

- Mll2/COMPASS occupies and catalyzes H3K4me3 at non-TSS elements
- The CXXC domain of Mll2/COMPASS mediates its chromatin targeting
- Mll2 regulates transcription of some genes including PGC specification genes
- PGC specification requires Mll2/COMPASS’s methyltransferase activity

In Brief

Hu et al. analyzed the contribution of MLL2’s methyltransferase and CXXC domain in the trimethylation of H3K4 in mouse ES cells and find that while it trimethylates H3K4 at both bivalent gene promoters and non-TSS elements, it regulates transcription at a limited number of genes including those required for PGC specification.
Not All H3K4 Methyllations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification

Deqing Hu,1,2 Xin Gao,2 Kaixiang Cao,1 Marc A. Morgan,1 Gloria Mas,3,4 Edwin R. Smith,1 Andrew G. Volk,1,5 Elizabeth T. Bartom,1 John D. Crispino,1,5,7 Luciano Di Croce,3,4,6 and Ali Shilatifard1,2,7,8,*

1Department of Biochemistry and Molecular Genetics, Feinberg School of Medicine, Northwestern University, 320 East Superior Street, Chicago, IL 60611, USA
2Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, MO 64110, USA
3Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Dr. Aiguader 88, 08003 Barcelona, Spain
4Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain
5Division of Hematology and Oncology, Northwestern University Feinberg School of Medicine, 303 E. Superior St., Chicago, IL 60611
6Institució Catalana de Recerca i Estudis Avançats (ICREA), Pg. Lluis Companys 23, 08010 Barcelona, Spain
7Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, 320 E. Superior St., Chicago, IL 60611, USA
8Lead Contact
*Correspondence: ash@northwestern.edu
http://dx.doi.org/10.1016/j.molcel.2017.01.013

SUMMARY

The spatiotemporal regulation of gene expression is central for cell-lineage specification during embryonic development and is achieved through the combinatorial action of transcription factors/co-factors and epigenetic states at cis-regulatory elements. Here, we show that in addition to implementing H3K4me3 at promoters of bivalent genes, Mll2 (KMT2B)/COMPASS can also implement H3K4me3 at a subset of non-TSS regulatory elements, a subset of which shares epigenetic signatures of active enhancers. Our mechanistic studies reveal that association of Mll2’s CXXC domain with CpG-rich regions plays an instrumental role for chromatin targeting and subsequent implementation of H3K4me3. Although Mll2/COMPASS is required for H3K4me3 implementation on thousands of loci, generation of catalytically mutant MLL2/COMPASS demonstrated that H3K4me3 implemented by this enzyme was essential for expression of a subset of genes, including those functioning in the control of transcriptional programs during embryonic development. Our findings suggest that not all H3K4 trimethylations implemented by MLL2/COMPASS are functionally equivalent.

INTRODUCTION

Histone H3K4me3 is an evolutionarily conserved chromatin mark from yeast to mammals and is associated with diverse chromatin-based processes, such as chromatin remodeling, transcriptional initiation, histone acetylation, and DNA recombination (Li et al., 2006; Matthews et al., 2007; Vermeulen et al., 2010). In budding yeast, H3K4 methylation is deposited by Set1/COMPASS (complex of proteins associated with Set1) (Krogan et al., 2002; Miller et al., 2001; Roguev et al., 2001; Schneider et al., 2005; Shilatifard, 2012). Drosophila has three Set1-related H3K4 methyltransferases, dSet1, Trithorax (Trx), and Trithorax-related (Trr), each of which forms a COMPASS-like complex (Mohan et al., 2011; Shilatifard, 2012). Mammals have two close relatives for each of the three H3K4 methyltransferases identified in Drosophila, resulting in a total of six COMPASS-like complexes: Set1a and Set1b; Mll1 and Mll2; and Mll3 and Mll4 (Piunti and Shilatifard, 2016; Shilatifard, 2012). Functional diversity within the COMPASS family in metazoans includes deposition of distinct methylation states of H3K4 and targeting distinct genomic regions. The Set1 proteins within COMPASS mediate the bulk of promoter H3K4me3 through crosstalk with the H2B monoubiquitination machinery (Mohan et al., 2011; Wu et al., 2008). Mll1 is required for H3K4me3 at the promoters of a subset of genes in mouse embryonic fibroblasts (MEFs) (Wang et al., 2009), while Mll2 is responsible for implementation of H3K4me3 at bivalently marked gene promoters in mouse embryonic stem cells (mESC) (Hu et al., 2013b). Recently, the Drosophila ortholog of Mll1 and Mll2 was shown to implement H3K4me2 at Polycomb response elements (Rickels et al., 2016). In contrast, the enhancer mark H3K4me1 is largely implemented by Trr in Drosophila and Mll3 and Mll4 in mammals (Herz et al., 2012, 2014; Hu et al., 2013a; Morgan and Shilatifard, 2015).

From yeast to humans, a direct functional role for H3K4 methylation in transcription remains unclear. Set1/COMPASS is the only H3K4 methyltransferase in yeast and its deletion affects all three states of H3K4 methylation (Schneider et al., 2005; Shilatifard, 2012). Nevertheless, there is no widespread transcriptional alteration in the absence of Set1 in budding yeast (Miller et al., 2001). Likewise, in mammalian cells, the loss of H3K4me3 at promoters has minimal effects on steady-state and regulated transcriptional induction in mESC (Clouaire...
Figure 1. Mll2/COMPASS Catalyzes H3K4me3 at Non-TSS Mll2 Binding Sites
(A) Pie chart of genome-wide Mll2 distribution in mESC determined by ChIP-seq with ab CT2.
(B) Mll2 occupancy at TSS and non-TSS regions.

(legend continued on next page)
et al., 2012, 2014; Hu et al., 2013b). Therefore, the role of H3K4 methylation in regulating transcription and embryonic development remains elusive. In this study, we uncover an essential role for the catalytic activity of MII2/COMPASS in H3K4 methylation in the regulation of a limited number of genes, including at enhancers and promoters of genes encoding regulators of PGC specification.

RESULTS

MII2/COMPASS Occupies Both Promoters and Non-TSS Regulatory Elements in mESC

Histone H3K4me3 accumulates at promoter-proximal regions of active genes but can also be found with H3K27me3 at the lowly transcribed “bivalent” genes in ES cells (Azuara et al., 2006; Bernstein et al., 2006; Santos-Rosa et al., 2002). We previously demonstrated that the H3K4me3 at bivalent promoters in ES cells is implemented by the MII2 branch of the COMPASS family (Hu et al., 2013b). To gain a broader understanding of the role of MII2/COMPASS in transcriptional regulation during development, we generated antibodies recognizing two different epitopes in the C-terminal portion of MII2 (ab CT1 and more C-terminal ab CT2) (Figure S1A). We first confirmed the specificity of the two antibodies in the detection of endogenous MII2 protein by immunoblotting whole-cell extracts from mESC in which MII2 was depleted by RNAi (Figure S1B). We further validated the two antibodies with immunoprecipitation and found that components of MII2/COMPASS were co-immunoprecipitated with MII2 (Figures S1C and S1D).

We identified MII2 targets by ChIP-seq with each antibody (Figures 1A and S1E). A total of 19,822 binding regions (peaks) were identified with ab CT2 MII2, among which 70%, 14%, and 16% of peaks are localized to promoters, gene bodies, and intergenic regions, respectively (Figure 1A). The high percentage of MII2 occupancy at promoters was consistent with its activity at bivalent genes in mESC (Hu et al., 2013b). MII2 peaks localized within some gene bodies or at intergenic regions (non-TSS) demonstrated a lower occupancy than sites of MII2 occupancy overlapping transcription start sites (TSS) (Figure 1B). Similar results were observed when performing ChIP-seq with ab CT1 (Figures S1E and S1F). Inspection of non-TSS MII2 peaks near H3F3a and Foxj2 loci reveals that they are co-occupied with the active enhancer marks p300, H3K4me1, and H3K27ac (Figure 1C). Non-TSS MII2 peaks can be associated with p300, H3K4me1, H3K27ac, and H3K27me3 (Figures 1D and S1G). More of the non-TSS MII2-associated regions are enriched for the active enhancer marks of p300, H3K4me1, and H3K27ac, than for H3K27me3, a mark of poised enhancers (Rada-Iglesias et al., 2011) (Figures 1D, 1E, S1G, and S1H).

Approximately 39% of non-TSS MII2-associated elements (2,465 out of 6,418) share marks with active enhancers (Figure 1F), half of which are further marked with H3K4me3 (Figure 1F). Similar percentages of co-occurrence of H3K4me3 with marks of active enhancers were observed with MII2 ab CT1 (Figure S1I) or with H3K4me3 ChIP-seq datasets generated in different mESC lines and by distinct laboratories (Clouaire et al., 2014; Denissov et al., 2014; Mikkelsen et al., 2007) (Figure S1J). Gene expression analysis revealed that genes closest to non-TSS MII2 sites had an overall higher expression level than all genes, consistent with them being enhancers (Figure 1G). Genes nearest non-TSS MII2 sites with H3K27ac had further increased expression, consistent with these being active enhancers (Figure 1G). However, the co-occurrence of H3K4me3 with these putative active enhancers was not significantly predictive of higher expression than H3K27ac alone (Figure 1G, group II versus group I), suggesting that for a majority of these putative enhancers, MII2-dependent H3K4me3 functions redundantly with H3K27ac and is not associated with driving increased expression in the ES cell state.

MII2 Catalyzes H3K4 Methylation at Non-TSS Sites

As MII2 functions as a methyltransferase for H3K4 and associates with non-TSS MII2-associated sites enriched for H3K4me3, we reasoned that MII2 may be the enzyme directly responsible for implementation of H3K4me3 at the non-TSS MII2-associated sites. To explore this hypothesis, we employed H3K4me3 ChIP-seq in mESC infected with lentiviral shRNA against GFP and MII2, respectively. Efficient knockdown of MII2 has previously been described (Hu et al., 2013b) and confirmed at the protein level (Figure S1B). MII2 depletion resulted in nearly complete loss of its occupancy around H3F3a and Foxj2, indicating the specificity of this antibody for ChIP-seq (Figure 1H). Histone H3K4me3 occupancy at enhancers of H3F3a and Foxj2 was similarly reduced after MII2 depletion.
Figure 2. Rescue Strategy for Dissecting Mll2 Activities in mESC
(A) Top: graphical representation of the Mll2 locus and two CRISPR/Cas9 target locations. Target sites are indicated in green and PAM sequences in red. Bottom: wild-type and mutant sequenced clones. Size of deletion indicated in base pairs (bp).

(legend continued on next page)
(Figure 1H). In contrast, H3K4me3 levels at the promoters of H3f3a and Foxj2 were unaffected (Figures 1H and 1I).

To further investigate the role of MII2 in H3K4me3 deposition at non-TSS MII2-associated sites, we performed global analysis of MII2 and H3K4me3 in shGFP and shMII2-treated cells. MII2 depletion resulted in broad reduction of H3K4me3 levels and MII2 occupancy at non-TSS MII2-associated sites targets (Figures 1J and 1K). To rule out the possibility that H3K4me3 at non-TSS sites was a crosslinking artifact due to enhancer-promoter interactions, ChIP-seq with native MNase-digested chromatin was performed (Figure S1K).

Cxxc1/Cpf1, a subunit found in Set1A and Set1B/COMPASS (Shilatifard 2012), is required for targeting Set1A and Set1B to promoters of active genes and establishing H3K4me3 in mESC (Clouaire et al., 2012). However, at non-TSS MII2-associated sites, H3K4me3 occupancy remained unchanged in the absence of Cxxc1 (Figure S1L), further indicating the specific requirement for MII2/COMPASS in H3K4me3 deposition at these regions. Direct regulation of H3K4me3 was further validated by determining MII2 and H3K4me3 occupancies at non-TSS MII2-bound regions identified by the CT1 MII2 antibody (Figure S1M). Together, our data established a requirement of MII2/COMPASS for H3K4me3 deposition at a subset of non-TSS sites in mESC.

Ectopically Expressed MII2 Depends on Its CXXC Domain for Recruitment to Chromatin

Menin associates with the MII1 and MII2 COMPASS-like complexes and is required for targeting MLL1 fusion proteins to genomic loci associated with leukemic transformation (Yokoyama et al., 2005). We performed ChIP-seq for Menin in mESC and observed that Menin co-occurs with MII2 at putative enhancers close to H3f3a and Foxj2 (Figure S2A). Despite co-occurring at many loci (Figure S2B), depletion of Menin by shRNA had no detectable effect on MII2 recruitment in mESC (Figures S2C and S2D), suggesting that Menin-dependent recruitment to chromatin could be cellular context dependent.

To identify the molecular mechanisms underlying MII2 targeting to chromatin, we used CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats) technology (Cong et al., 2013), to generate MII2 knockout (MII2KO) mESC, which could be further used in rescue experiments (Figure 2A). Western blot analysis of a representative mESC clone demonstrated a complete loss of MII2 protein in MII2KO mESC (Figure 2B), Consistent with the shRNA-mediated MII2 knockdown studies, MII2 knockout had no major effect on bulk levels of H3K4 methylation in mESC (Figure 2B). We first tested the requirement for the CXXC domain of MII2 for its targeting to chromatin, since the homologous domain in MII1 can bind unmethylated CpG-containing DNA and plays an essential role in the recruitment of MLL1 fusion proteins to HOXA9 in leukemia cells (Ayton et al., 2004; Cierpicki et al., 2010). Although most promoters in mammals are characterized by the presence of CpG islands (Guenther et al., 2007), about half of mammalian CpG islands are found in intragenic and intergenic regions (Illingworth et al., 2010).

We reintroduced wild-type, CXXC domain-deficient (ΔCXXC), or the catalytically mutant (Y2604A) versions of Halo-tagged human MLL2 (Figure 2C) into MII2KO mESC. Different rescue constructs were expressed to similar levels as endogenous MII2 (Figure 2D) and all of the proteins could assemble into COMPASS-like complexes (Figures 2E and S2E). While the wild-type MII2 construct was able to fully rescue H3K4me3 defect at non-TSS MII2-associated sites, the CXXC and Y2604A mutants could not (Figures 2F–2H and S2F). Although the Y2604A and wild-type Halo human MLL2 proteins were properly recruited to MII2-associated regions, the recruitment of the CXXC mutant was impaired (Figures 2I, 2J, and S2G). In line with a role for the CXXC domain in recruiting MII2 to chromatin, the non-TSS MII2 sites overlap CpG-rich regions (Figure 2G and 2I).

CRISPR-Mediated Genome Editing of Endogenous MII2’s CXXC and Catalytic Domains

We used CRISPR-mediated homologous recombination to mutate six of the eight conserved cysteine residues to alanine to abolish CXXC function of the endogenous MII2, and tyrosine 2602 (corresponding to tyrosine 2604 in human MLL2) to alanine to generate catalytically deficient MII2, respectively (Figures 3A, 3B, and S3A–S3F). Western blotting in the targeted mESC clones demonstrated that the amino acid substitutions have no observable impact on MII2 protein levels (Figure 3C). ChIP-seq for H3K4me3 in parental, MII2 knockout, and two independent clones each for the CXXC domain and catalytic mutants, demonstrated a requirement for MII2’s CXXC domain and catalytic activity for non-TSS H3K4me3 (Figures 3D, 3E, S3G, and S3H). ChIP-seq for MII2 demonstrated that the catalytic mutant, but not the CXXC mutants, was properly recruited to non-TSS...
Figure 3. CRISPR Editing Demonstrates Requirements for Mll2’s CXXC and Catalytic Domains in mESC
(A) Top: strategy for disrupting the catalytic activity of Mll2 in mESC. A homology repair template changes the catalytic tyrosine 2602 to alanine. The CRISPR-cas9 cleavage site is indicated with a red arrowhead. Primers (P1 to P6) used for genotyping are shown as bars with primer pairs color coded. Bottom: agarose gel electrophoresis of PCR genotyping products in clones before (N) and after (F) flippase (Flp)-mediated eviction of the Neo cassette. Note that PCR products are

(legend continued on next page)
Mll2-associated sites and bivalent promoters (Figures 3F, 3G, and S3I). Taken together, these findings establish an essential role for the CXXC domain and the methyltransferase activity of Mll2 in ES cells for chromatin targeting and H3K4me3 deposition at non-TSS Mll2-associated sites.

Mll2 Is Required for Expression of Regulators of Primordial Germ Cell Specification during Differentiation

Mll2 is essential for embryonic development but is not required to maintain the pluripotent state (Glaser et al., 2006; Lubitz et al., 2007). To explore Mll2’s role during early development, we conducted RNA-seq analysis in parental and Mll2KO mESC and during their differentiation (Figure S4A). In agreement with a dispensable role for Mll2 in self-renewal, RNA-seq analysis in Mll2KO mESC identified 286 genes that were downregulated and 327 genes that were upregulated (Figure 4A). Functional annotation of downregulated genes revealed enrichment for different classes of genes including those involved in germ cell development and gamete generation (Castrillon et al., 2000; Fujisawa et al., 1994; Noce et al., 2001; Ohinata et al., 2005; West et al., 2009; Yamaji et al., 2008) (Figure 4B; Table S1). In contrast, genes upregulated in Mll2KO cells were enriched in categories such as brain and pancreas development, and metabolic processes.

We conducted an additional RNA-seq analysis in parental, Mll2KO mESC, and Mll2KO mESC with the Crispr-Cas9 system (Figure S4A). Together, these analyses suggest that Mll2 may have a role in maintaining the pluripotent state (Table S2). We note that the dependency on Mll2 is most obvious during the differentiation studies as the dependency on Mll2 in the ES cell state can vary with growth conditions.

Although loss of Mll2 does not affect self-renewal of mESC grown under serum + LIF condition, Mll2KO mES cells exhibit faster mRNA depletion for Nanog and Klf4 during differentiation (Figure S4E). Supporting the accelerated differentiation of the Mll2KO mESC, we observed that many markers of endoderm, mesoderm, and epidermis have significantly higher expression at different days of Mll2KO EBs than in wild-type EBs (Figures S4F–S4H; Table S2). Taken together, these studies suggest a role for Mll2 in PGC induction and the proper orchestration of differentiation for PGC and other lineages.

Both the CXXC Domain and the H3K4 Methyltransferase Activity of Mll2 Are Required for PGC Induction from mESC

To better understand the requirement of Mll2 for expression of genes involved in PGC specification during differentiation, we compared the gene expression profile in parental, Mll2 knockout, CXXC mutant, and catalytically deficient Mll2 mESC (Figure S5A). Hierarchical clustering of differentially expressed genes across these cell lines reveal that expression changes in CXXC and catalytically deficient mutant MiLL2 cells are similar to Mll2KO cells (Figure S5A). The expression of PGC determinants such as Prdm14 and Prdm1 were decreased in mESC lacking the DNA binding or the methyltransferase activity of endogenous Mll2 (Figures S5B and S5B).

To further determine the role of Mll2 in PGC specification during mESC differentiation, we differentiated parental, Mll2 knockout, CXXC, and catalytically deficient mutant Mll2 mESC to EBs (Figure S5C). Similar to Mll2KO EBs, EBs derived from CXXC and catalytically deficient mutant Mll2 mESC were heterogeneous in morphology and displayed dark regions in the center of their EBs, although the extent was less severe as compared to Mll2KO EBs (Figure S5C). We analyzed RNA-seq data from EBs differentiated from parental, Mll2KO, CXXC, and catalytically deficient MiLL2 mutant mESC and observed reduced expression of Prdm1 and Prdm14 in day 6 EBs derived from Mll2_mCXXC, Mll2_Y2602A, and Mll2 KO EBs, indicating an essential role for the CXXC domain and the full enzymatic activity for Mll2-dependent PGC formation (Figures S5D, S5C, and S5D).

As EBs represent spontaneous differentiation of ES cells and the number of PGC in EBs was potentially low, we decided to test the function and respective mechanisms (CXXC domain

unsuccessfully amplified with primers P1 and P2 in two targeted ES cell clones due to the large Neo cassette but can be amplified after FLP-mediated excision of Neo, with the PCR product migrating slower than the wild-type product due to a remaining FRT site.

(B) Top: strategy for mutating the CXXC domain of Mll2 in mESC. CRISPR-cas9 cleavage site primer pairs indicated as in (A). Bottom: agarose gel analysis of PCR genotyping in (A).

(C) Western analysis of Mll2 protein levels in targeted ES cell clones before Neo excision (N1–N2) but mutant Mll2 is present at wild-type levels in four clones (F1–F4) after FLP-mediated excision of Neo.

(D) UCSC genome browser track example of H3K4me3 occupancy near the H3F3A locus in parental, Mll2 knockout, CXXC, and catalytic mutant MiLL2 mESC.

(E) Heatmap representation of H3K4me3 occupancy at Mll2 non-TSS sites in parental, and the indicated mutant Mll2 mESC.

(F) UCSC genome browser track example of Mll2 occupancy near the H3F3A locus in parental and the indicated mutant Mll2 mESC.

(G) Heatmap analysis of Mll2 occupancy at Mll2 non-TSS sites in parental, and the indicated mutant Mll2 mESC. ChIP-seq signal represented as in Figure 1J. See also Figure S3.
Figure 4. MII2 is Required for Full Expression of PGC Genes during Differentiation

(A) MA plot of gene expression changes between parental and MII2KO mESC as determined by RNA-seq. The x axis shows the log2-normalized counts per million (CPM) of averaged expression in the two conditions (A) and the y axis shows the log2-fold change in the two conditions (M). Significantly upregulated and downregulated genes are highlighted in red and green, respectively (FDR < 0.001).

(legend continued on next page)
and full methyltransferase activity) of Mll2 in PGC induction in a more controlled model of PGC differentiation system. mESC of parental, Mll2KO, Mll2_mCXXC, and Mll2_Y2602A grown under N2B27+2i+LIF conditions were first induced to differentiate into Epiblast-like cells (EpiLC) followed by addition of multiple cytokines to dictate PGCLC specification (Figure 5E, top). As seen in EBs, expression of PGC determinants Prdm1, Prdm14, and Dppa3 was considerably decreased while markers for endoderm and ectoderm were upregulated in PGCCLC aggregates derived from Mll2KO, Mll2_mCXXC, and Mll2_Y2602A mESC, further supporting the essential role for Mll2 and its CXXC domain and catalytic activity in PGC induction (Figure 5E, bottom). As expected, the number of PGCLCs derived from Mll2KO, Mll2_mCXXC, and Mll2_Y2602A cells was decreased 4-fold as demonstrated by flow cytometric analysis of PGC markers (Hayashi and Saitou, 2013) (Figure 5F). Collectively, our findings support a role for Mll2, its ability to bind DNA, and its methyltransferase activity in PGC specification during ES cell differentiation.

**Regulation of Enhancer Activity by Mll2/COMPASS during PGC Specification**

Among the 6,418 non-TSS Mll2-associated sites, 58% of them were enriched for enhancer marks (Figure 1F). We observed that genes closest to non-TSS Mll2-associated sites had higher expression than all genes; however, loss of Mll2 had no general effect on the expression of the nearest genes in ES cells (Figure S6A). Of the 4,684 genes nearest non-TSS Mll2-associated sites, only 43 genes were significantly downregulated by Mll2 loss (Figure S6B), including genes encoding PGC regulators. The low number of downregulated genes nearest non-TSS sites is consistent with the observation from Figure 1G that the presence of H3K4me3 at putative active enhancers conferred no general advantage on gene expression when compared to other active enhancers. Together, these findings indicate that loss of H3K4me3 at non-TSS sites does not have a general instructive role in transcriptional activation, but instead depends on other, yet unidentified chromatin contexts. However, we cannot rule out that the minimal effect of loss of Mll2 and H3K4me3 on nearby gene expression was the result of a low level of H3K4me3 implemented by other enzymes being sufficient for gene activation, as we rarely observed a complete lack of H3K4me3 at non-TSS Mll2-associated sites upon Mll2 deletion.

We identified putative enhancer regions in mESC that required Mll2 for their H3K4me3 near PGC determinants such as Prdm14 and Prdm1 (Figure 6A). To determine whether they function as true enhancers for regulation of Prdm1 and Prdm14 transcription, we deleted these genomic elements by CRISPR/Cas9 (Figures 6B, 6C, and S6C). As expected, deletion of these elements led to decreased expression of Prdm1 and Prdm14, indicating that they are bona fide enhancers for Prdm1 and Prdm14 (Figures 6D and S6D). In addition to H3K4me3 loss in Mll2KO cells, enhancers of Prdm1 and Prdm14 exhibited reduced levels of p300 and H3K27ac, two marks associated with active enhancers, while the general enhancer mark, H3K4me1, was largely unaffected (Figure S6E). However, global analysis of H3K4me1 at 6,418 non-TSS Mll2-associated sites demonstrated a very weak, but significant increase after Mll2 loss (Figures S6F and S6G). This could reflect some degree of competition between Mll2/COMPASS and other COMPASS family members such as the monomethylases MLL3 and MLL4 (Hu et al., 2013a).

One mechanism underlying enhancer action is the physical association with targeted gene promoters (Dekker et al., 2002; Ong and Corces, 2011). To explore whether there are physical contacts between enhancers and promoters of PGC determinants, we performed circularized chromosome capture (4C) coupled with high-throughput sequencing in mESC using genomic regions near the promoters of Prdm1 and Prdm14 as viewpoints. Significant 4C-seq signal from both viewpoints was observed at enhancer regions with Mll2-dependent H3K4me3, suggesting that Mll2-regulated enhancers indeed loop into close proximity with promoters of PGC determinant genes (Figure 6E).

To determine whether Mll2, its CXXC domain or its catalytic activity play a direct role in the regulation of cis-regulatory elements of Prdm1 and Prdm14, their enhancers and promoters were individually cloned into luciferase reporter constructs and co-transfected with Halo-tagged human MLL2 or its mutant derivatives in HEK293 cells. We observed that the activity of enhancers and promoters of Prdm1 and Prdm14 can be stimulated by expression of wild-type MLL2. This MLL2 activity was abolished by either deletion of its CXXC domain or mutation of the SET domain, indicating that the stimulation of enhancer and promoter activity by MLL2 is dependent on the DNA binding activity of its CXXC domain and the catalytic activity of its SET domain (Figure 6F).

The requirement of Mll2’s methyltransferase activity and DNA binding activity for Prdm1 and Prdm14 enhancer function prompted us to test whether establishment of enhancer-promoter interaction requires Mll2 or its methyltransferase activity. We performed 4C with the Prdm14 promoter as viewpoint in wild-type, KO, and CXXC and catalytic Mll2 mutant mESC. We found that Mll2 deletion, inhibition of its chromatin recruitment, or mutation of its catalytic domain has no observable impact on the interaction between the enhancers and the promoter of Prdm14 (Figure S6H). However, further studies are needed to understand the role of Mll2 in PGC specification.
Figure 5. Mll2’s CXXC Domain and Catalytic Activity Function in PGC Specification

(A) Heatmap of gene expression in mESC after Mll2 knockout, mutation of the catalytic SET domain, or functional disruption of Mll2’s CXXC domain. Differentially expressed genes are separated into two groups based on their up- or down-regulation in Mll2 knockout cells.

(B) UCSC genome browser tracks of total RNA-seq in wild-type and indicated Mll2 mutant conditions. Top four tracks are from Figure 4C.

(C) Representative phase contrast images of 6 day EBs derived from parental, Mll2KO, CXXC, and catalytically deficient mESC.

(legend continued on next page)
better understand the role of H3K4 methylation in the regulation of enhancer and promoter communication at other loci. To investigate a role of Mll2 and Mll2-dependent enhancers in PGC specification, we differentiated parental, Mll2KO, Prdm1, and Prdm14 enhancer-deleted mESC into PGCLC in vitro followed by flow cytometric analysis of PGCLC markers. As seen with Mll2KO mESC, Prdm14 enhancer deletion also leads to inefficient specification of PGCLC (Figure 6G). However, we did not observe a defect in PGCLC differentiation with the deletion of the distal Prdm1 enhancer (En1, data not shown), suggesting that the second enhancer (En2) residing within the gene body of Prdm1 may compensate for the loss of function of En1 during differentiation. In summary, our results demonstrate a molecular mechanism in which the recruitment of Mll2/COMPASS onto CpG island embedded enhancers establishes H3K4me3 and activates factors essential for PGC specification (Figure 7).

**DISCUSSION**

Here, we have demonstrated that Mll2/COMPASS occupies some promoters and non-TSS regulatory elements and implements H3K4me3 at these loci. We found that Mll2 is required for the expression of several genes including the master regulators required for primordial germ cell specification during ES cell differentiation. The DNA binding activity of the CXXC domain and the catalytic activity are indispensable for Mll2’s function in the regulation of PGC specification during differentiation, suggesting that the recruitment of the catalytic activity of Mll2/COMPASS to these loci is central for its specific transcriptional regulatory properties. Given that Mll2/COMPASS can methylate thousands of promoters and non-TSS regulatory elements, but only a subset of genes require Mll2/COMPASS and its catalytic activity in H3K4 trimethylation for the regulation of gene expression, our findings suggest that not all H3K4-methylated sites are functionally equal.

In line with an active regulatory role for H3K4me3 in PGC gene expression, deletion of Kdm5b, a H3K4me3 demethylase, leads to a global increase of H3K4me3 and a failure to silence the expression of PGC genes (Schmitz et al., 2011). Together, these findings point to an essential role of H3K4me3 in PGC establishment during embryonic differentiation. Since we observe that Mll2 is required for H3K4me3 at both promoters and enhancers of PGC genes, determining the degree to which H3K4me3 changes at promoters or enhancers of these genes is an important area of future investigation.

Nanog is a core pluripotency factor required for both the pluripotent mESC and the unipotent PGCs. Ectopic expression of Nanog in EpiLCs is sufficient to induce PGCLCs formation without addition of exogenous cytokines (Murakami et al., 2016). Mechanistic studies revealed that Nanog associates with the enhancers of Prdm1 and Prdm14 and is required for their activation during PGC induction (Murakami et al., 2016). As we noticed an accelerated reduction in Nanog expression in Mll2KO mESC during differentiation (Figure S4E), lowered expression of Nanog may partially account for the failure to fully activate Prdm1 and Prdm14 in the absence of Mll2.

CpG islands are found at promoters of bivalent genes targeted by Mll2 in mESC (Bernstein et al., 2006; Hu et al., 2013b) and artificial integration of synthetic CpG islands into the mESC genome can create novel bivalent domains that are independent of Cxxc1-containing Set1A/Set1B/COMPASS-like complexes (Wachter et al., 2014). We observed that non-TSS regulatory elements targeted by Mll2 in mESC also show enrichment for CpG dinucleotides (Figures 2G and 2I). Mll2 harbors a CXXC domain, which can bind specifically to nonmethylated CpG (Voo et al., 2000) and this domain is essential for targeting Mll2 to these regions.

Since both Mll1 and the Cxxc1 subunit of Set1A- and Set1B/COMPASS have CXXC domains, it is unclear how Mll2 is preferentially recruited for H3K4me3 at non-TSS Mll2-associated sites in ES cells. More than half of non-TSS Mll2-associated sites are enriched for the enhancer marker of H3K4me1. We previously demonstrated that Mll3/Mll4/COMPASS-like complexes are recruited to enhancers for H3K4 monomethylation (Herz et al., 2012; Hu et al., 2013a). Determining how enhancers specifically choose Mll2 for their H3K4 trimethylation and Mll3/Mll4 for their H3K4 monomethylation will be important for understanding functional differences in different cis-regulatory elements and how they function to regulate pluripotency and development.

The molecular basis for H3K4me3-dependent transcriptional activation of genes during differentiation is currently unknown. Histone H3K4me3 has been previously reported to participate in many biological processes including nucleosome remodeling, transcription initiation, pre-mRNA splicing, and DNA damage response through binding to distinct effector proteins (Li et al., 2006; Matthews et al., 2007; Shi et al., 2006; Sims et al., 2005, 2007; Vermeulen et al., 2007, 2010; Wysocka et al., 2006). Various repressor proteins have been found to bind unmodified H3K4, including BHC80 (component of CoREST), DNMT3L (in complex with DNA de novo methyltransferase DNMT3A and DNMT3B), and UHRF1 (in complex with the DNA maintenance methyltransferase DNMT1) (Lan et al., 2007; Nady et al., 2011; Ooi et al., 2007). H3K4me3 interferes with the binding of these repressors and thus may be involved in inhibition of DNA methylation. In addition, H3K4me3 was recently shown to keep DNMT3A in an autoinhibitory state and removal of histone methylation leads to enzymatic activation of DNMT3A (Guo et al., 2015). Therefore, H3K4me3 may recruit or repel a specific set of downstream effectors from enhancers near PGC determinant genes for transcriptional activation during differentiation.

Our results demonstrate that Mll2/COMPASS is required for the steady-state expression of several different genes during
Figure 6. Mll2 and Putative Enhancers of PGC Determinants

(A) ChIP-seq genome browser tracks showing Mll2-dependent H3K4me3 at two putative enhancers (En1 and En2) each near Prdm1 and Prdm14 in mESC. A non-bound element (NBE) at each locus is also indicated.

(legend continued on next page)
differentiation including genes involved in PGC specification during development. Removal of H3K4me3 from bivalent chromatin by Mll2 depletion in mESC skews cells toward differentiation and leads to prominent defects during development (Glaser et al., 2006; Lubitz et al., 2007). As seen in the pluripotent ES cells, PGCs also maintain a bivalent domain at promoters of key regulators of all three germ layers (Lesch and Page, 2014). We propose that Mll2 is not only required for the establishment of PGCs during differentiation through H3K4 methylation of key PGC genes but it may also be essential for maintaining the bivalent chromatin state for normal PGC function.

**STAR METHODS**

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
  - Cell Lines
  - Expression Plasmids and short hairpin RNAs
- **METHOD DETAILS**
  - Antibody Generation
  - Lentivirus-based Knockdowns, Rescue Experiments and Generation of Embryoid Bodies
  - CRISPR-mediated Knockouts and Knockins
  - PGCLC Induction
  - FACS Analysis
  - Immunoprecipitation
  - Dual Luciferase Assay
  - 4C Sequencing
  - Size Exclusion Chromatography
  - Quantitative RT- and ChIP- PCR
  - ChIP-seq
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
  - ChIP-Seq Analysis
  - Peak Calling
  - Replicate Information
  - ChIP-seq Heatmaps and Enrichment Profiles
  - RNA-Seq Analysis
  - 4C-Seq Analysis
  - Data Sources
- **DATA AND SOFTWARE AVAILABILITY**

(B) Schematic of CRISPR/Cas9 design for candidate enhancer deletion. Black arrows indicate the location of genomic regions targeted by CRISPR guide RNAs. P1, P2, P3, and P4 represent primers used for genotyping. For each construct, the protospacer sequence and the Cas9-specific proximal-adjacent motif (PAM) are underlined in green and red, respectively.

(C) Genotyping of mESC transfected with candidate enhancer-flanking CRISPR constructs identified five clones with deletion of Prdm14 putative enhancers and three clones with deletion of the distal putative Prdm1 enhancer.

(D) RNA-seq genome browser tracks of Prdm1 and Prdm14 in parental and three representative clones of the Prdm1 or Prdm14 enhancer-deleted mESC cells.

(E) Physical interactions of Mll2-regulated putative enhancers with promoters of Prdm1 and Prdm14 as revealed by circularized chromosome conformation capture with high-throughput sequencing (4C) assay. The median and 20th and 80th percentiles of a sliding 10 kb window determine the main trend line. Color scale represents enrichment relative to the maximum median value at a resolution of 12 kb.

(F) Luciferase reporter assay of the role of putative enhancers of Prdm14 (left) and Prdm1 (right) in transcriptional activation in the presence of wild-type, ΔCXXC, or catalytically deficient mutant human MLL2. Enhancers (En), non-binding element (NBE), or promoter (Pr) regions indicated in (A) were cloned into a luciferase reporter vector and co-transfected into HEK293 cells with the indicated MLL2 construct. After 48 hr, cells were harvested, and luciferase activity was determined. Relative luciferase induction is plotted as fold change compared to cells transfected with empty control vectors. Error bars indicate the standard deviation of two biological experiments.

(G) Parental, Mll2KO or Prdm14 enhancer-deleted mESC were differentiated into PGCLCs in vitro. After 6 days, PGCLCs were quantified by FACS. Percentage of PGCLCs induced from parental, Mll2KO, and Prdm14 enhancer-deleted mESC in a representative experiment and the average results from two independent quantifications are presented. Error bars indicate the standard deviation of two biological experiments. p values are calculated using the t test. See also Figure S6.

---

**Figure 7. A Working Model for Mll2/COMPASS during Development**

First panel: Mll2/COMPASS can be recruited to both enhancers and promoters through CXXC-dependent binding of CpG-rich regions and implements H3K4me3 for the expression of master regulators of PGC specification. Second panel: during differentiation, Mll2−/− cells are not capable of transcriptional activation of PGC factors, resulting in defective PGC specification. Third panel: Mll2ΔCXXCΔCXXC is incompetent to occupy enhancers and promoters of PGC regulators, leading to compromised PGC specification. Fourth panel: catalytically deficient Mll2Y2602A/Y2602A is recruited to enhancers and promoters of PGC regulators but fails to implement H3K4 methylation and leads to compromised PGC specification.

---

472 Molecular Cell 65, 460–475, February 2, 2017
SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures and three tables and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2017.01.013.

AUTHOR CONTRIBUTIONS


ACKNOWLEDGMENTS

We thank Stacy Marshall and Emily Rendleman for help with Illumina sequencing. We are grateful to the Stowers Molecular Biology core and Tissue Culture core for assistance with the generation of some of the NGS data and maintenance of cell lines. We thank Andrea Punti for critical reading of the manuscript and Laura Shilatifard for editorial assistance. D.H. was supported by the Robert H. Lurie Comprehensive Cancer Center – Translational Bridge Program Fellowship in Lymphoma Research. G.M. received the support of la Convocatoria de Ayudas Fundación BBVA a Investigadores, Innovadores y Creadores Culturales. A.V. was supported by NIH training grant T32CA080621. These studies were further supported by grants from the Spanish “Ministerio de Educacio´ n y Ciencia” (SAF2013-48926-P) and AGAUR to L.D.C., and NIH grants R01CA101774 to J.D.C., R50CA211428 to E.R.S., and R35CA197569 to A.S.

Received: March 10, 2016
Revised: July 16, 2016
Accepted: January 4, 2017
Published: February 2, 2017

SUPPORTING CITATIONS

The following references appear in the Supplemental Information: Xu et al. (2011); Zhang et al. (2015).

REFERENCES


## STAR★METHODS

### KEY RESOURCES TABLE

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Antibodies</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rabbit polyclonal anti-Menin</td>
<td>Bethyl Laboratories</td>
<td>Cat# A300-105A; RRID: AB_2143306</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-RbBP5</td>
<td>Bethyl Laboratories</td>
<td>Cat# A300-109A; RRID: AB_210551</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-ASH2L</td>
<td>Bethyl Laboratories</td>
<td>Cat# A300-489A; RRID: AB_451024</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-WDR5</td>
<td>Bethyl Laboratories</td>
<td>Cat# A302-429A; RRID: AB_1944302</td>
</tr>
<tr>
<td>Rabbit monoclonal anti-WDR82</td>
<td>Cell Signaling Technology</td>
<td>Cat# 9971S; RRID: AB_1944302</td>
</tr>
<tr>
<td>Mouse monoclonal anti-beta Tubulin</td>
<td>DSHB</td>
<td>Cat# E7; RRID: AB_2315513</td>
</tr>
<tr>
<td>Mouse monoclonal anti-HaloTag</td>
<td>Promega</td>
<td>Cat# G9211</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-p300</td>
<td>Santa Cruz</td>
<td>Cat# sc-585</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-H3K27ac</td>
<td>Abcam</td>
<td>Cat# ab4729</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-MLL2_ab CT1</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-MLL2_ab CT2</td>
<td>This study</td>
<td>N/A</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-H3K4me1</td>
<td>Hu et al., 2013a</td>
<td>N/A</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-H3K4me2</td>
<td>Hu et al., 2013a</td>
<td>N/A</td>
</tr>
<tr>
<td>Rabbit polyclonal anti-H3K4me3</td>
<td>Hu et al., 2013a</td>
<td>N/A</td>
</tr>
<tr>
<td>PE-conjugated anti-mouse/rat CD61 Antibody</td>
<td>BioLegend</td>
<td>Cat# 104308</td>
</tr>
<tr>
<td><strong>Antihuman/Mouse SSEA-1 eFluor 660 Antibody</strong></td>
<td>Affymetrix eBioscience</td>
<td>Cat# 50-8813-42</td>
</tr>
<tr>
<td><strong>Chemicals, Peptides, and Recombinant Proteins</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N-2 Supplement (100X)</td>
<td>ThermoFisher</td>
<td>Cat# 17502-048</td>
</tr>
<tr>
<td>B-27 Supplement (50X), minus vitamin A</td>
<td>ThermoFisher</td>
<td>Cat# 12587010</td>
</tr>
<tr>
<td>ESGRO Leukemia Inhibitory Factor (LIF)</td>
<td>Millipore</td>
<td>Cat# ESG1107</td>
</tr>
<tr>
<td>MEK inhibitor, PD0325901</td>
<td>Stemgent</td>
<td>Cat# 04-0006</td>
</tr>
<tr>
<td>GSK-3β inhibitor, CHIR99021</td>
<td>Stemgent</td>
<td>Cat# 04-0004</td>
</tr>
<tr>
<td>Human Plasma Fibronectin Purified Protein</td>
<td>Millipore</td>
<td>Cat# FC010</td>
</tr>
<tr>
<td>KnockOut Serum Replacement</td>
<td>ThermoFisher</td>
<td>Cat# 10828028</td>
</tr>
<tr>
<td>bFGF</td>
<td>ThermoFisher</td>
<td>Cat# 13256-029</td>
</tr>
<tr>
<td>Activin A</td>
<td>Peprotech</td>
<td>Cat# 120-14</td>
</tr>
<tr>
<td>Nunc dishes 96-well low cell-binding plate</td>
<td>Sigma</td>
<td>Cat# Z721093</td>
</tr>
<tr>
<td>Glasgow’s MEM (GMEM)</td>
<td>ThermoFisher</td>
<td>Cat# 11710-035</td>
</tr>
<tr>
<td>MEM Non-Essential Amino Acids Solution (NEAA)</td>
<td>ThermoFisher</td>
<td>Cat#11140-050</td>
</tr>
<tr>
<td>L-Glutamine (200 mM)</td>
<td>ThermoFisher</td>
<td>Cat# 25030081</td>
</tr>
<tr>
<td>Sodium Pyruvate (100 mM)</td>
<td>ThermoFisher</td>
<td>Cat# 1360070</td>
</tr>
<tr>
<td>Penicillin-Streptomycin (10,000 U/mL)</td>
<td>ThermoFisher</td>
<td>Cat# 15140122</td>
</tr>
<tr>
<td>Gibco 2-Mercaptoethanol</td>
<td>ThermoFisher</td>
<td>Cat# 21985023</td>
</tr>
<tr>
<td>BMP-4</td>
<td>R&amp;D Systems</td>
<td>Cat# 314-BP-010</td>
</tr>
<tr>
<td>BMP8a</td>
<td>R&amp;D Systems</td>
<td>Cat# 1073-BP-010</td>
</tr>
<tr>
<td>SCF</td>
<td>R&amp;D Systems</td>
<td>Cat# 455-MC-010</td>
</tr>
<tr>
<td>EGF</td>
<td>R&amp;D Systems</td>
<td>Cat# 2028-EG-200</td>
</tr>
<tr>
<td>Protein A/G PLUS-Agarose</td>
<td>Santa Cruz</td>
<td>Cat# sc-2003</td>
</tr>
<tr>
<td>DNase I (RNase-free)</td>
<td>New England BioLabs</td>
<td>Cat# M0303S</td>
</tr>
<tr>
<td>Micrococcal nuclease (MNase)</td>
<td>Worthington</td>
<td>Cat# 9013-53-0</td>
</tr>
</tbody>
</table>

(Continued on next page)
### Critical Commercial Assays

<table>
<thead>
<tr>
<th>REAGENT or RESOURCE</th>
<th>SOURCE</th>
<th>IDENTIFIER</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dual-Luciferase Reporter Assay System</td>
<td>Promega</td>
<td>Cat#E1910</td>
</tr>
<tr>
<td>PerfeCTa SYBR Green SuperMix Reaction Mixes, Quanta Biosciences</td>
<td>VVR</td>
<td>Cat# 01414-144</td>
</tr>
<tr>
<td>TruSeq Stranded Total RNA LT - (with Ribo-ZeroTM Human/Mouse/Rat) - Set A</td>
<td>Illumina</td>
<td>RS-123-2201</td>
</tr>
<tr>
<td>TruSeq Stranded Total RNA LT - (with Ribo-ZeroTM Human/ Mouse/Rat) - Set B</td>
<td>Illumina</td>
<td>RS-123-2202</td>
</tr>
<tr>
<td>High-Throughput Library Preparation Kit Standard PCR Amp Module - 96 rxn</td>
<td>KAPA Biosystems</td>
<td>KK8234</td>
</tr>
</tbody>
</table>

### Deposited Data

| Raw and analyzed data | This study | GEO: GSE78708 |

### Experimental Models: Cell Lines

| V6.5 mouse embryonic stem cells | Novus Biologicals | Cat#NBP1-41162 |
| HEK293 | ATCC | Cat#CRL-1573 |
| Puromycin-Resistant Mouse Embryonic Fibroblasts (MEF) | Stem Cell Technologies | Cat# #00325 |

### Recombinant DNA

| pFENHK | Promega | N/A |
| pFENHK_Halo_MLL2 | Promega | N/A |
| pFENHK_Halo_MLL2(ΔCXCC) | This study | N/A |
| pFENHK_Halo_MLL2(Y2604A) | This study | N/A |
| pLKO.1 | Addgene | Cat# 10878 |
| PX459 | Addgene | Cat# 48139 |

### Sequence-Based Reagents

| shMenin: ACATATTGCTGCCGGAATTG3 | This study | N/A |
| shMll2: GGAGAACTCTGATTGAGAAAGHu et al., 2013b | N/A |
| 4C sequencing Prdm14 locus Forward: AATGATACGGCGACCACCGAGATCTACAATTTCTTCCCTACAGAGGTCTTCCGATCTNNNNNNTGACTAAGACCAAAGTTCCAAGACTGCA | This study | N/A |
| 4C sequencing Prdm14 locus Reverse: CAAGCAGAAGACGGCATACGAGATCCGGAGTTCAAAATAGCA | This study | N/A |
| 4C sequencing Prdm1 locus Forward: AATGATACGGCGACCACCGAGATCTACAATTTCTTCCCTACAGAGGTCTTCCGATCTNNNNNNTGACTAAGACCAAAGTTCCAAGACTGCA | This study | N/A |
| 4C sequencing Prdm1 locus Reverse: CAAGCAGAAGACGGCATACGAGATCCGGAGTTCAAAATAGCA | This study | N/A |
| Mll2KO-crgRNA-Left: 5'GGAGGGGCGCCCGCGC3' | This study | N/A |
| Mll2KO-crgRNA-Right: 5'CTTGCGGCCCCGCGC3' | This study | N/A |
| Mll2-mCXXC-crgRNA: 5'TCGGCGTCTTCTGTGTCG3' | This study | N/A |
| Mll2-Y2602A-crgRNA: 5'TGTAACGCGAACATCGATGC3' | This study | N/A |

(Continued on next page)
CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests may be directed to, and will be fulfilled by the lead contact corresponding author Ali Shilatifard (ash@northwestern.edu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell Lines

HEK293 cells and mouse embryonic fibroblasts (MEF) were maintained in DMEM plus 10% FBS. Mouse V6.5 embryonic stem cells were cultured in 0.1% gelatin-coated tissue culture flasks with irradiated MEF feeder cells. Cells were grown in DMEM supplemented with 15% FBS (HyClone), 2 mM L-glutamine, 0.1 mM nonessential amino acids (StemCell Technologies) and 1,000 U/ml recombinant LIF (Millipore, ESG1107).

Expression Plasmids and short hairpin RNAs

Human full-length MLL2 cDNA was subcloned into pFN205K EF1a plasmid with an amino-terminal Halo tag between Sgf I and Pme I sites. Halo-tagged MLL2 (ΔCXXC) and MLL2 (Y2604A) plasmids were generated by deleting amino acids from 958 to 1000 of the CXXC domain or replacing tyrosine 2604 within the SET domain, respectively (Dillon et al., 2005). Plasmids and shRNA sequences are provided in the Key Resources Table.
shRNAs (listed in Key Resources Table) were cloned into the pLKO.1 vector (Addgene). Mouse V6.5 ES cells were infected with lentiviruses expressing shRNAs for 24 hr before selection with 2 μg/ml puromycin for 2 d. For ChIP-seq and RNA-seq experiments, cells were grown for one passage without feeder cells for 30 min before harvesting.

**METHOD DETAILS**

**Antibody Generation**

Histone H3K4me1, H3K4me2, H3K4me3 were generated in our lab and have been used and described in our previous studies (Hu et al., 2013a; Hu et al., 2013b). Carboxyl-terminal antibodies to human MLL2 were made as follows: CT1 refers to rabbits #9172 and #9173 and were generated with the MLL2 peptide spanning amino acids 2071-2220; CT2 refers to rabbit #9175 and was generated with the MLL2 peptide spanning amino acids 2264-2413. The peptides were expressed as His-tag fusion proteins in pET-16b vector and purified on NTA-agarose according to QIAGEN’s protocol.

**Lentivirus-based Knockdowns, Rescue Experiments and Generation of Embryoid Bodies**

Lentivirus-mediated knockdown has been described previously (Hu et al., 2013b). The hanging drop method was used to form embryoid bodies (EBs) by culturing 1000 mouse ES cells in 25 μl differentiation medium (without LIF) on the lid of 150 mm culture dish for the indicated number of days. For generation of human MLL2-rescue mouse ES cells, constructs were linearized with Mlu I restriction enzyme and electroporated into wild-type or Mll2 KO mouse ES cells with the Gene Pulser Xcell system (Bio-Rad) with a setting of 250V, 500 μF capacitance. 24 hr post electroporation, cells were selected with 400 μg/ml G418 for 3 days followed by growing for one passage for 30 min to remove feeder cells before harvesting.

**CRISPR-mediated Knockouts and Knockins**

sgRNAs were designed using the CRISPRtool (http://crispr.mit.edu) and cloned into PX459 vector (Addgene, 48139). Targeting vector and linearized donor vector were co-electroporated into V6.5 mouse embryonic stem cells and 24 hr later followed by 2 days of puromycin selection. For generation of CXXC and catalytically deficient mutants of endogenous Mll2, cells were continued to be selected in the presence of neomycin drug for a week. Targeted single-ES cell colonies were screened by PCR with primers (see Figure 4A and 4B) and analyzed by agarose-gel electrophoresis. PCR products were sequenced to further validate the genomic editing.

**PGCLC Induction**

PGCLCs induction was conducted following a previously described protocol (Hayashi et al., 2011; Hayashi and Saitou, 2013). Briefly, mouse ES cells cultured in N2B27 medium (N-2: ThermoFisher Scientific, 17502-048; B27: ThermoFisher, 12587-010) plus LIF and 2i (1mM MEK inhibitor, Stemgent, PD0325901; 3mM GSK-3β inhibitor, Stemgent, CHIR99021) were trypsinized and 1 × 10⁵ cells were plated onto one well of 12-well plate pre-coated with 5ug/ml plasma fibronectin (Millipore, FC010). After two days of differentiation in N2B27 medium supplemented with 1% knockout serum replacement (KSR, ThermoFisher, 10828-028), 20 ng/ml activin A (Peprotech, cat.no 120-14) and 12 ng/ml bFGF (ThermoFisher, 13256-029), the resulting EpiLCs were trypsinized and 2 × 10⁴ cells were plated into one well of 96-well low cell-binding plate (Sigma, Z721093) for 4 days of differentiation into PGCLCs in serum-free GMEM (ThermoFisher, 11710-035) medium supplemented with 15% KSR, 1 x NEAA (ThermoFisher, 11140-050), 1 x L-glutamine (ThermoFisher, 25030081), 1 x sodium pyruvate (ThermoFisher, 1360070), 1 x penicillin-streptomycin (ThermoFisher, 15140122), 55 nM 2-Mercaptoethanol (ThermoFisher, 21985023), 1,000 U of LIF, 500 ng/ml BMP4 (R&D, 314-BP-010), 500 ng/ml BMP8a (R&D, 1073-BP-010), 100 ng/ml SCF (R&D, 455-MC-010) and 50 ng/ml EGF (R&D, 2028-EG-200). At the end of day 4 of differentiation, PGCLCs aggregates were either harvested for RNA extraction or dissociated for flow cytometric analysis.

**FACS Analysis**

For quantification of PGCLCs, differentiated aggregates were trypsinized at 37°C for 8 min and quenched with MEF medium. The dissociated cells were washed two times in FACS buffer (PBS with 0.1% BSA) and stained with PE-conjugated anti-mouse/rat CD61 (BioLegend, 104308, 1:200 dilutions) and anti-human/mouse SSEA1 eFluor®660 (eBioscience, 50-8813-42, 1:200 dilutions) antibodies for 1 hr on ice followed by three washes with FACS buffer. Stained cells were fixed in 4% paraformaldehyde for 30 min on ice and washed with FACS buffer for three times before analysis was performed with a FACSCantoll cytometer (BD Bioscience).

**Immunoprecipitation**

Immunoprecipitations were performed as previously described (Hu et al., 2013c). Briefly, V6.5 mouse ES cells were lysed in RIPA (20 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1% NP40, 1% sodium deoxycholate, 0.1% SDS, 1 mM dithiothreitol [DTT]) and protease inhibitors (Sigma). Cleared lysates were incubated with the indicated antibodies and protein A/G PLUS agarose (Santa Cruz, sc-2003). After washing beads, immunoprecipitates were eluted with loading buffer for SDS-PAGE.
**Dual Luciferase Assay**

For dual luciferase assays, promoter and putative enhancers of Prdm1 and Prdm14 were cloned into pGL4.12[luc2CP] and the pGL4.25[luc2CP/minP] reporter vectors, respectively. The reporter plasmids were transfected with Lipofectamine 2000 into HEK293 cells grown in 24-well plates, together with pRL-SV40 internal control plasmid, and either empty, wild-type, CXXC or catalytically deficient versions of human Halo-tagged MLL2 for 48 hr before harvesting for passive lysis (Promega, E1910), and quantitation of luciferase activity on a GloMax-multi fluorometer (Promega). The total amount of transfected plasmids per well of transfection was adjusted to be equal to 1ug/well for all samples.

**4C Sequencing**

4C sequencing was performed as previously described (van de Werken et al., 2012a). Briefly, \(10^7\) cells were fixed with 2% formaldehyde at room temperature and digested with HindIII (NEB) overnight at 37°C. The digested chromatin was diluted in 1× T4 DNA ligase buffer (Roche) with addition of T4 DNA ligase (Roche) and incubated overnight at 16°C. Ligated chromatin was reverse-cross-linked to generate the 3C template, which was further digested with NlaIII (NEB). Samples were then diluted and ligated with T4 DNA ligase (Roche) to generate the 4C template. 4C-PCR was performed with the primers in the Key Resources Table (NNNNNN indicates location of index sequence) and next generation sequencing was performed on Nextseq 500 (Illumina).

**Size Exclusion Chromatography**

HEK293 cells were transfected with the indicated plasmids by Lipofectamine 2000 (ThermoFisher, 11668027). Three days after transfection, nuclear extracts were prepared and concentrated with Millipore Ultrafree centrifugal filters (10 kDa nominal molecular mass limit) before applying onto a Superose 6 2.0/300 column (GE Healthcare) that had been equilibrated with buffer (25 mM Tris, pH 7.4, 150 mM NaCl, 10% glycerol, and 1mM EDTA). Fractions were separately collected and analyzed by western blotting with respective antibodies.

**Quantitative RT- and ChIP-PCR**

Total RNA was isolated from cells or EBs using Trizol, treated with DNAase I (NEB, M0303S), and reversed transcribed using SuperScript III and random primers (Invitrogen). Resulting cDNA was analyzed by quantitative PCR using Green SuperMix for iQ (VVR, 01414-144) on a MyIQ Systems (Bio-Rad). For other quantitative ChIP-PCR, isolated DNA was analyzed using SYBR green on a CFX connect Real-Time PCR detection system (Bio-Rad).

**ChIP-seq**

ChIP assays were performed as previously described (Lee et al., 2006). Briefly, 2-5 \(10^7\) cells were cross-linked with 1% paraformaldehyde at room temperature for 10 min and quenched by glycin. Fixed chromatin was sheared to generate chromatin fragments of 200-600 bp with a Misonix 3000 sonicator and immunoprecipitated with the indicated antibodies. For native ChIP-seq, nuclei were isolated from 1× \(10^7\) cells and digested with 10U MNase (Worthington, 9013-53-0) for 15min followed by full-speed centrifugation at 4°C for 30 min. Supernatants containing mononucleosomes were used for ChIP with antibodies against H3K4me3. All ChIP-seq libraries were prepared with the KAPA HTP library preparation kit (KAPA Biosystems, KK8234) and sequenced on either the Illumina HiSeq 2500 or NextSeq 500.

**QUANTIFICATION AND STATISTICAL ANALYSIS**

Methods and software for calculating \(p\) values are described in the relevant sections below and in the figure legends. Briefly, error bars for qPCR represent two technical replicates in a representative experiment of at least three biological replicates. FDR values for enrichment in ChIP-seq experiments were determined with MACS (Zhang et al., 2008). \(P\) values for estimating the significance of peak overlap in ChIP-seq data were determined with GAT (Heger et al., 2013). \(P\) values associated with boxplots were calculated with the Wilcoxon rank-sum test. \(P\) values for Gene Ontology were determined with the hypergeometric test using the software Metascape (Tripathi et al., 2015). \(P\) values for Venn Diagrams were calculated using the hypergeometric test in R. Adjusted \(p\) values for differential expression were determined with the Benjamini-Hochberg method in EdgeR 3.0.8 (Robinson et al., 2010).

**ChIP-Seq Analysis**

Sequencing reads were aligned to the mouse genome (UCSC mm9) using the Bowtie aligner v0.12.9 allowing uniquely mapping reads only and allowing up to two mismatches (see Table S3 for alignment statistics) (Langmead et al., 2009). Reads were extended to 150 bases toward the interior of the sequenced fragment and normalized to total reads aligned (reads per million; RPM). External sequencing data was acquired from GEO and ArrayExpress as raw reads and aligned in the same way as internally sequenced samples.

**Peak Calling**

Peak detection was performed with MACS v1.4.2 (Zhang et al., 2008). Associated control samples were used to determine statistical enrichment at a \(p < 1e-8\) and FDR < 0.05. The enrichment of external data (not including H3K4me3) was determined at \(p < 1e-5\) and
FDR < 0.05. Enrichment for H3K27me3 was determined with the broad domain peak detector SICER v1.1 (Zang et al., 2009) at the FDR < 1e-8, window size of 200, and gap size of 600.

Non-promoter peaks are the peaks that reside at least 1kb away from a transcription start site (TSS), with the remaining peaks being the TSS peaks. Gene annotations and TSS information were from Ensembl 67. Peaks were assigned to the features of its nearest gene. Co-bound peaks were called if the peak regions of two samples overlap. The significance of peak overlap was tested by GAT (Heger et al., 2013).

**Replicate Information**

Mll2 ChIP-seq with two independent antibodies were analyzed separately to gain comprehensive information on its genome-wide distribution and to increase our confidence in the specificity of two antibodies. The overlapping high-confidence peak regions of two biological replicates of H3K4me3 were used for further analysis. The overlapping peaks were merged into one region as the union of all. One replicate is generated from this study, the other is from GEO: GSE48172.

**ChIP-seq Heatmaps and Enrichment Profiles**

For ChIP-seq heatmaps, regions of interest were depicted at the center of each peak or nearest TSS and sorted by occupancy. Regions spanning 5 kb on either side of the indicated feature were binned into 25 bp windows. Coverage is normalized into reads per million (RPM). Occupancies were calculated as the mean coverage under each peak region. P values for boxplots were determined with the Wilcoxon rank sum test. For ChIP-Seq enrichment profiles, regions of interest were shown for each factor as a binary value of enriched/not enriched. The regions were centered at the center of each peak and sorted by the shortest distance of peak center to an annotated TSS. Regions shown are oriented 5’ to 3’ corresponding to the orientation of the nearest gene. Regions spanning 50 kb on either side of feature indicated were binned into 200 bp windows. Each line represents a peak and its surrounding region.

**RNA-Seq Analysis**

Sequencing reads were aligned to the mouse genome UCSC mm9 (see Table S1 for alignment statistics) and to gene annotations from Ensembl 67 using TopHat v2.0.13 (Trapnell et al., 2009), using option –g 1 –x 1. R package edgeR 3.0.8 was used to perform differential expression analysis at FDR < 1e-3 (Robinson et al., 2010). To estimate sample dispersion better, all the RNaseq samples are combined to fit one generalized linear model. Genes that expressed (> = 1 RPM) in less than 2 samples are filtered out. External sequencing data was downloaded from GEO48172 as raw reads and aligned in the same way as internally sequenced samples. Gene Ontology (GO) analysis of differentially expressed genes in Mll2 KO mouse ES cells was performed by Metascape (Tripathi et al., 2015). MA plots are generated by edgeR: A is logCPM, M is logFC. x axis indicates averaged expression level (log2 normalized counts). y axis represents log2 fold changes of normalized counts of Mll2 knockout (KO) cells versus the parental cells.

**4C-Seq Analysis**

Reads were demultiplexed by allowing no mismatches in the library index and 1 mismatch in viewpoint sequences. The library indexes were trimmed before mapping. 4cseqpipe was used for alignment, normalization and generating near-cis domainograms (van de Werken et al., 2012b).

**Data Sources**

Data generated for this study includes ChIP-seq data for H3K4me3, MLL2, shMll2_MLL2, H3K4me3 in Mll2WT_Vec, Mll2KO_Vec, Mll2KO_MLL2WT, Mll2KO_MLL2(CXXC) and Mll2_MLL2(Y2604A) ES cells, H3K9ac, H3K27ac and Pol II in ES cells, and H3K4me3 in MEF cells. Other datasets come from previously published studies. Replicates for H3K4me3, shMll2_H3K4me3 and RNA-seq in ES cells are from GEO: GSE48172. p300, H3K4me1 and H3K27ac ChIP-seq data are from GEO: GSE24164 (Creyghton et al., 2010). H3K27me3 ChIP-seq data are from GEO: GSE12241 (Mikkelsen et al., 2007). H3K4me3 ChIP-seq in Cxxc1(Cfp1) WT and KO ES cells are from ArrayExpress repository under accession E-ERAD-79 (Reynolds et al., 2012). CpG islands were downloaded from UCSC browser mm9.

**DATA AND SOFTWARE AVAILABILITY**

ChIP-seq and RNA-seq datasets have been deposited at GEO: GSE78708.