



Published in final edited form as:

Mol Cell. 2016 July 21; 63(2): 318–328. doi:10.1016/j.molcel.2016.06.018.

An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements implemented by Trx/MLL/COMPASS

Ryan Rickels¹, Deqing Hu¹, Clayton K. Collings¹, Ashley R. Woodfin¹, Andrea Piunti¹, Man Mohan^{2,+}, Hans-Martin Herz^{2,#}, Evgeny Kvon³, and Ali Shilatifard^{1,4,*}

¹Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, 320 E. Superior Street, Chicago, IL 60611, USA

²Stowers Institute for Medical Research, Kansas City, MO 64110, USA

³Research Institute of Molecular Pathology (IMP), Vienna Biocenter VBC, Dr Bohr-Gasse 7, 1030 Vienna, Austria

⁴Robert H. Lurie NCI Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, 320 E. Superior St., Chicago, IL 60611, USA

Abstract

Polycomb Response Elements (PREs) are specific DNA sequences that stably maintain the developmental pattern of gene expression. *Drosophila* PREs are well characterized, whereas the existence of PREs in mammals remains debated. Accumulating evidence supports a model in which CGIs recruit Polycomb-Group complexes (PcG), however, which subset of CGIs are selected to serve as PREs is unclear. Trithorax (Trx) positively regulates gene expression in *Drosophila* and co-occupies PREs to antagonize Polycomb-dependent silencing. Here, we demonstrate that Trx-dependent H3K4 dimethylation (H3K4me₂) marks *Drosophila* PREs and maintains the developmental expression pattern of nearby genes. Similarly, the mammalian Trx homology, MLL1, deposits H3K4me₂ at CpG dense regions that could serve as PREs. In the absence of MLL1 and H3K4me₂, H3K27me₃ levels, a mark of Polycomb Repressive Complex 2 (PRC2), increase at these loci. By inhibiting PRC2-dependent H3K27me₃ in the absence of MLL1, we can rescue expression of these loci, demonstrating a functional balance between MLL1 and PRC2 activities at these sites. Thus, our study provides rules for identifying cell-type specific functional mammalian PREs within the human genome.

*Correspondence and proofs should be sent to the following address: Ali Shilatifard, Ph.D., Department of Biochemistry and Molecular Genetics, Northwestern University Feinberg School of Medicine, 320 E. Superior Street, Chicago, IL 60611, USA.

ASH@northwestern.edu.

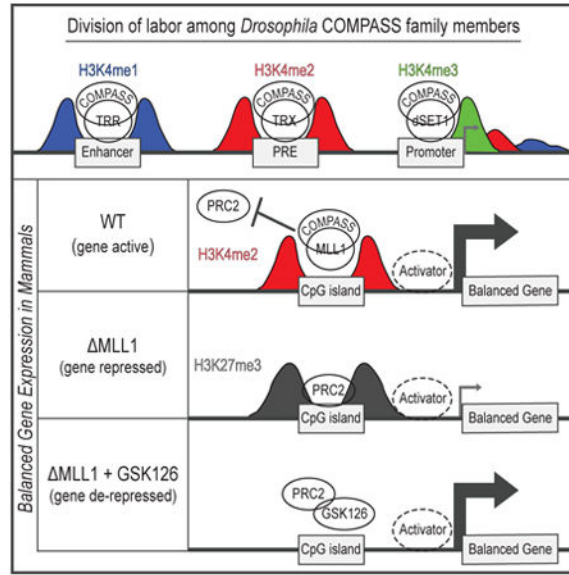
⁺Present address: Department of Biochemistry and Molecular Cell Biology, Shanghai Jiaotong University School of Medicine, Shanghai 200025, China

[#]Present address: Department of Cell and Molecular Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

Author Contributions: R.R. and A.S. conceived the project R.R. and D.H. designed and performed experiments with input from M.M., H.M.H., and A.P. R.R., A.P., and E.K. analyzed experimental data. C.K.C. and A.R.W. designed and performed genomics data analyses. R.R. and A.S. co-wrote the paper. All authors edited the manuscript.

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

Graphical abstract



Introduction

Histone H3K4-methylation by the COMPASS family (complex of proteins associated with Set1) is conserved from yeast to mammals and is closely associated with transcriptionally active chromatin (Shilatifard, 2012). Interest in yeast Set1 arose due to its ancestral homology with mammalian MLL1, a gene commonly translocated in childhood leukemia, and with *trithorax*, identified as a homeotic mutant and suppressor of *Polycomb* phenotypes in *Drosophila* (Kennison and Tamkun, 1988; Shilatifard, 2012). The COMPASS family members share several subunits in common, as well as unique factors believed to impart functional specificity (Herz et al., 2012; Mohan et al., 2011). Accumulating evidence supports a model in which proper transcriptional regulation requires a division of labor among COMPASS family members. For instance, Set1 is responsible for maintaining bulk levels of H3K4 di- and tri-methylation (H3K4me2/3) at transcriptionally active genes in both *Drosophila* and mammalian cells (Herz et al., 2012; Mohan et al., 2011; Wang et al., 2009; Wu et al., 2008). Meanwhile, MLL2 implements H3K4me3 at bivalently marked promoters in mouse embryonic stem (ES) cells, and Trx/MLL3/MLL4 were recently shown to catalyze H3K4-monomethylation at *Drosophila* and mammalian enhancers, respectively (Herz et al., 2012; Hu et al., 2013a; Hu et al., 2013b).

In *Drosophila*, Trx maintains the active transcription state of a target-gene in the absence of the initiating signal; thus, maintaining epigenetic memory of previous transcriptional states (Poux et al., 2002). This maintenance function is known to rely on Trx binding to a nearby Polycomb Response Element (PRE) (Klymenko and Muller, 2004; Schuettengruber et al., 2011). While *Drosophila* PREs are well-documented, the existence of comparable elements in mammals has been debated (Bauer et al., 2015; Levine et al., 2004; Mendenhall et al., 2010; Simon and Kingston, 2013). One theory gaining support is that CpG Islands (CGIs) can function as PREs via recruitment of PRC1 and PRC2 complexes (Klose et al., 2013;

Mendenhall et al., 2010; Tanay et al., 2007). In support of this idea, a recent report demonstrated that the vast majority of CGIs become bound by PcG complexes when transcription is globally inhibited, suggesting that any CGI can function as a PRE (Riising et al., 2014). MLL1 and MLL2 proteins contain CXXC domains that bind unmethylated CpG-rich sequences, but these factors do not require active transcription for their DNA-binding (discussed in this study) (Hu et al., 2013b).

While all CGIs may have the potential to function as PREs in the absence of transcription, it is unclear what mechanisms prevent PRC1/2 from accessing these sites. Mouse and human genomes contain roughly 20,000-30,000 CGIs, indicating these elements alone are not sufficient to impart specificity to MLL/PcG target binding (Illingworth et al., 2010). Here, we describe a previously unrecognized role for Trx as an H3K4-dimethylase at *Drosophila* PREs, and report the identification of over 2,800 cell-type specific human DNA elements that resemble *Drosophila* PREs with regard to their MLL1-dependent H3K4me2 and ability to counteract PRC2-silencing activity.

Results

Trx/COMPASS-dependent H3K4me2 chromatin signature marks PREs in *Drosophila*

The COMPASS family in *Drosophila* (dCOMPASS) has three Set1-related H3K4-methylases: dSet1, Trx, and Trr, each of which associates with several proteins to form COMPASS-like complexes (Herz et al., 2012; Mohan et al., 2012; Mohan et al., 2011). Prompted by the findings of Herz et al. that Trr catalyzes specific H3K4me1 at enhancers, we decided to interrogate the function of Trx in *Drosophila* S2 cells. Because Trx is enzymatically cleaved by Taspase into N-terminal (Trx-NT) and C-terminal (Trx-CT) fragments (Capotosti et al., 2007; Hsieh et al., 2003), we determined genome-wide occupancies of both termini with N- and C-terminus specific antibodies. To identify unique sites of Trx-dependent H3K4-methylation, we first performed genome-wide CHIP-seq for H3K4me1/2/3 after the reduction in Trx levels by RNAi (Figure S1A). At genes such as *dac*, *inv*, and *en*, we noticed a prominent loss of H3K4me2 upon Trx depletion (Figure 1). Genome-wide analysis identified 161 high-confidence Trx-CT (SET domain-containing) occupied sites at which H3K4me2 was significantly reduced following Trx-RNAi (Figure 1B and C; Figure S1C). Although Trr is shown to co-bind with Trx at these sites, H3K4me2 levels are not affected by Trr-depletion (Figure S1B). Surprisingly, these Trx-dependent sites also contain high levels of H3K27me3 in combination with H3K4me2 (Figure 1B). In *Drosophila*, H3K27me3 is nucleated at PREs where it is thought to aid in local transcriptional repression (Simon and Kingston, 2009). We found 88% of our Trx-dependent sites in these cells overlap with sites of Polycomb Repressive Complex 1 (PRC1) binding (Enderle et al., 2010) (Figure 1B). This is consistent with several reports that Trx and PRC1 co-occupy PREs regardless of their transcriptional state (Beisel et al., 2007; Chinwalla et al., 1995; Enderle et al., 2010; Schwartz et al., 2010; Simon and Kingston, 2013). Gene ontology analysis of the 127 nearest genes revealed an enrichment for known targets of PcG-dependent regulation, including homeodomain-encoding genes and other developmental transcription factors (Figures 1D and 1E). Thus, we have identified an alternative method, based on Trx-dependent H3K4-dimethylation, for identifying cell-type specific sequences

either defined operationally or by prediction to be Polycomb/Trithorax Response Elements (PREs) (Beisel et al., 2007; Enderle et al., 2010; Ringrose et al., 2003). We do not, however, exclude the possibility that some PREs will go undetected by our method; for instance, those reported to exist in a “void” state lacking any identifiable chromatin modification (Schwartz et al., 2010).

Trx-dependent H3K4-methylation at PREs and Enhancers

Our genome-wide analyses reveal that after Trx-depletion, PRE-associated genes with decreased H3K4me2 levels also exhibit reduced transcription (Figure 2A). This can be observed despite the fact that the majority of PRE-associated genes are already transcriptionally silent in S2 cells (Figure S2A). Gene ontology terms associated with the down-regulated genes are enriched for developmental processes (Figure 2B). Interestingly, Trx-CT not only remains bound, but also retains catalytic activity at repressed PREs. Track examples in Figure 1A provide a comparison of Trx-dependent H3K4-methylation at transcriptionally active (*dac*) or silenced PREs (*inv*, *en*). In the silent state, Trx-dependent H3K4 mono- and -dimethylation is confined to the PRE, while, in the transcriptionally active state H3K4me2 is spread across the gene body. The extent to which H3K4me2 spreads from the active PRE matches the transcriptional activity of the nearby Trx-regulated gene. For example, at the transcribed *apterous* (*ap*) locus, H3K4me2 spreads into the gene body and also across a ~20kb region upstream from the PRE (Figure 2C). This upstream intergenic region contains several cis-regulatory elements that most likely regulate *ap* expression patterns *in vivo* (Gohl et al., 2008; Kvon et al., 2014). Trx depletion coincides with a dramatic decrease in *ap* and *pnr* transcript levels, as well as a reduction in H3K4me2 and H3K27ac at the PRE and across the entire region encompassing these two genes and their putative regulatory elements (Figure 2C).

Within a single transcriptional domain, PREs may regulate the activity of several enhancers, each of which is necessary for robust expression of a given gene during development (Francis and Kingston, 2001; Simon and Kingston, 2013). To explore this further, we compared our 161 PREs with a published list of *Drosophila* enhancers identified by STARR-seq (Arnold et al., 2013). Although we do not find a significant overlap between PREs and STARR-seq enhancers, we do see examples of PREs in close or overlapping proximity to enhancers, such as at *Ptx1* and *Optix* (Figure 2C). This is somewhat unexpected, given the prior indications that PREs do not possess intrinsic enhancer activity (Poux et al., 2002; Simon et al., 1993). This finding, however, may help explain other prior results that some PREs can impart “positional information” to a reporter gene's late-embryonic expression pattern (Chan et al., 1994; Hagstrom et al., 1996).

In the rapidly developing *Drosophila* embryo, Trx maintains active transcription of a target gene in a PRE-dependent manner (Cavalli and Paro, 1999; Poux et al., 2002). Although conclusive experimental evidence is still lacking, this function is hypothesized to depend on Trx lysine-methyltransferase activity (Schuettengruber et al., 2011). We wanted to determine whether or not PRE-proximal sites of Trx-dependent H3K4me2 are positioned near enhancer sequences that are active during embryogenesis. Indeed, the embryonic enhancer for *Optix* shown in Figure 2C and 2D was also identified as both a Trx-responsive PRE and

STARR-seq enhancer. In some cases, the pattern of an endogenous gene is composed of multiple patterns encoded by separate enhancers. For example, in late stage embryos, the full expression pattern of *Ptx1* is a composite of several inter- and intragenic enhancers (Figure 2D). The patterns driven by two enhancers flanking the *Ptx1* PRE contribute to its expression in the midgut and ventral nerve cord (peach and green box, respectively). Even in the repressed state, these *Ptx1* enhancers overlap with sharp peaks of Trx-dependent H3K4me1 and H3K4me2 (Figure 2D, third panel). At highly active Trx-regulated genes, such as *ap* and *pnr*, embryonic enhancers are enriched for H3K4me2 and H3K27ac, all of which are severely reduced after Trx-RNAi (Figure 2C, second panel). This result is consistent with the hypothesis that, while Trx-mediated H3K4me2 may not be necessary for initial gene activation, it can facilitate the ability of overlapping and nearby developmental enhancers to maintain transcription.

A conserved role for MLL1/COMPASS as H3K4-dimethylase at PRE-like regions in the Human genome

The human orthologs of Trx are MLL1 and MLL2 (Kmt2A and Kmt2B, respectively) (Mohan et al., 2012)(Shilatifard 2012). MLL1 is known to play a central role in the deposition of H3K4me3 at Hox gene promoters while MLL2 is required for H3K4me3 at bivalent gene clusters in embryonic stem cells (Hu et al., 2013b; Wang et al., 2009). To investigate whether the H3K4-dimethylase function of Trx is conserved in mammals, we depleted MLL1 in HCT116 cells and performed ChIP-seq for H3K4me1/2/3 (Figure 3A; Figure S3A). MLL2 is not expressed in HCT116 cells, obviating concerns regarding any compensatory function of MLL2 (Figure S3B). Although bulk H3K4-methylation is unaffected in the absence of MLL1 (Figure S3C), H3K4me2 levels are dramatically reduced at a total of 2,832 locations in the genome. After centering on total H3K4me2 peaks, we divided the data into TSS and non-TSS overlapping peaks. Intriguingly, only 51% (1,446) of MLL1-dependent H3K4me2 peaks overlap an annotated TSS, while the remaining 49% (1,386) do not overlap on annotated TSS. Our findings corroborate previous reports that MLL1 catalyzes H3K4me3 at target gene promoters (Wang et al., 2009). However, we reveal that H3K4me2 levels predominate over H3K4me3, and are more significantly affected at MLL1 target-TSS compared to MLL1-independent TSS (Figure 3C; Figure S3D). Upon MLL1-depletion, H3K27me3 levels are significantly increased at MLL1-dependent TSSs, consistent with a role for MLL1 as an antagonist of PRC2-mediated repression (Figures 3B and 3C; Figure S3F).

At the 1,386 MLL1-dependent non-TSS sites, we see identical changes with regard to H3K4-methylation; however, the effects on H3K27ac and H3K27me3 are less drastic compared to TSS sites (Figures 3D and 3E). To our surprise, the MLL1-dependent sites are significantly enriched for CGIs ($p = 2.99E-221$, Pearson's Chi-squared test) compared to MLL1-independent sites of intergenic H3K4me2 (Figure 3D). This suggests CGIs are part of a common recruitment mechanism for targeting MLL1 to both genic and intergenic loci. To further distinguish the 1,386 MLL1-dependent sites from the remaining 4,884 nonTSS sites containing H3K4me2, we performed motif enrichment analysis of the underlying DNA sequences. This analysis revealed that CTCF was the most significantly enriched motif at MLL1-dependent nonTSS sites (p -value = $1e-167$, reported by HOMER). This result was

confirmed by analyzing recently published ChIP-seq data for CTCF in HCT116 cells, which shows a significant enrichment (p-value < 2.2e-16, Pearson's Chi-squared test) for CTCF co-binding at MLL1-dependent nonTSS sites versus MLL1-independent sites (Figure 3D) (Maurano et al., 2015). This result suggests that CTCF might play a role in specifying the MLL1-dependent subset of CGIs.

Loss of MLL1-dependent H3K4me2 at non-TSS accompanies down-regulation of nearby target genes

MLL1 is known to positively regulate transcription of a diverse set of genes in different cell types and under distinct cellular environments (Caslini et al., 2009; Milne et al., 2005; Takeda, 2006). We performed RNA-seq to assess genome-wide transcriptional changes upon MLL1-depletion in HCT116 and identified 550 significantly down-regulated genes (Figure S4A). Next, we wanted to know whether the MLL1-dependent H3K4me2 sites tend to be located near down-regulated genes or distributed randomly. To evaluate this, we identified 2,434 genes nearest to MLL1-dependent sites (TSS and nonTSS) and calculated the overlap with the 550 down-regulated genes (Figure S4A). Of the 550 down-regulated genes, 8% (44) overlapped with nonTSS nearest genes, and 32% (179) overlapped with TSS nearest genes (Figure S4B). We believe this is a conservative estimate due to the fact that cis-regulatory elements do not necessarily regulate the nearest gene (Kvon et al., 2014). Nonetheless, GREAT analysis of the nearest genes to both MLL1-dependent TSS and nonTSS elements retrieved similar GO-terms, indicating these elements associate with the same set of genes (Figure S4C and S4D) (McLean et al., 2010). These results imply that MLL1/COMPASS-regulated genes utilize a combination of TSS and nonTSS-associated cis-regulatory elements to control transcription in mammalian cells.

Transcription of MLL1-dependent genes is balanced by PRC2 catalytic-activity: A predicted characteristic of mammalian PREs

Previous work in *Drosophila* cells demonstrates that a subset of Trx-regulated genes exist in a “balanced” state, in which their transcriptional output is simultaneously influenced by Trx and PcG complexes (Schwartz et al., 2010). In the presence of PcG proteins, these genes require Trx to maintain some level of transcriptional activity; however, this level of transcription is usually less than what would otherwise occur in the absence of PcG complexes. The subset of Trx/PcG-regulated genes that match this description is most likely cell type-specific (Schuettengruber et al., 2011; Schwartz et al., 2010).

Consistent with this, and the conserved functions between Trx and MLL1, we have identified a subset of MLL1-regulated genes that exist in a similar state of “balanced” expression. Using CRISPR-CAS9, we generated an HCT116 MLL1^{NULL} cell line that closely resembles the effects of MLL1-RNAi on H3K4 and H3K27-methylation (Figures 4A and 4C). Within the MLL1-dependent TSS-overlapping sites, a fraction of genes decrease in the MLL1^{NULL}, but return to near WT levels upon treatment with EZH2-inhibitor (GSK126) (Figures 4B and 4C). As expected, the MLL1-dependent TSSs that gain the most H3K27me3 associate with the largest decreases in gene expression upon deletion of MLL1 (Figure 4C). Interestingly, the majority of genes with MLL-dependent H3K4me2 at their

TSS are de-repressed following GSK126 treatment, indicating PRC2 and H3K27me3 play a dominant role at most of these 1,446 sites (Figure 4C).

To validate the specificity of our H3K4me2-based method of PRE-like element identification, we extended our analysis genome-wide to consider differential expression of all protein-coding genes. Using K-means clustering to partition genes based on differential expression in the MLL1^{NULL} cells +/- GSK126, our analysis reveals cluster 2 as a distinct set of genes that visibly fit the “balanced” model of MLL1/PcG genetic regulation (Figure 4D). In agreement with Figure 4C, these 295 genes specifically recruit PRC2 to their TSS after MLL1-deletion, as evidenced by increased H3K27me3 at these specific locations (Figure 4D). Importantly, cluster 2 is significantly enriched ($p=3.88e-28$, hypergeometric test) for MLL1-dependent H3K4me2. Cluster 3 closely resembles cluster 2 with regards to MLL1-dependent H3K4me2, increased H3K27me3, and transcriptional de-repression following GSK126 treatment. However, the 488 genes in this group are repressed by PRC2, despite the presence of MLL1, and are thus more lowly expressed in WT cells compared to cluster 2 (Figure 4D). Cluster 1 shows increased gene expression after removing MLL1; however, there are no changes in H3K4me2 or H3K27me3, indicating the transcriptional changes seen in this group are an indirect effect of MLL1 deletion (Figure 4D). These secondary effects can be explained by many previous reports that MLL1 regulates a cohort of master regulatory transcription factors (Bracken, 2006; Wang et al., 2009). It is interesting how even modest gene expression changes in the MLL1-independent clusters 1,4, and 5 are reversed upon GSK126 treatment. This demonstrates that, while only ~300 genes in HCT116 cells directly depend on MLL1 for maintained activation, fluctuating expression of “balanced” genes can perpetuate transcriptional changes throughout the genome (Figure 4D).

Discussion

From flies to humans, the COMPASS family of H3K4-methylases have increased in number and diversified in function (Shilatifard, 2012). However, we continue to uncover remarkable similarities in rules of transcriptional regulation between these species. While the identities of target genes regulated by Trx and MLL1 are remarkably conserved between flies and humans (Bracken, 2006), the identification of vertebrate PREs that function similarly to those in *Drosophila* is still under investigation (Margueron and Reinberg, 2011; Mendenhall et al., 2010). In this work, we describe an alternate method for identifying *Drosophila* PREs based on Trx-dependent H3K4me2, and apply this same approach to identifying mammalian DNA elements that resemble PREs in many respects. The most striking characteristic in common between the two systems is the high level of Trx/MLL1-dependent H3K4me2 at discreet sites surrounding known Trx/MLL1 target-genes, regardless of their transcriptional state. We find mammalian PRE-like sequences are enriched for CGIs, DNA elements already shown to recruit PcG complexes following gene silencing (Cooper et al., 2014; Riising et al., 2014). Although CGIs, by the strict definition, do not exist in *Drosophila*, the conservation of CXXC-domains in Trx and MLL1/2 suggests that high G/C-content could be part of a conserved recruitment mechanism. By analyzing the underlying DNA sequence content, we demonstrate that *Drosophila* PREs are indeed G/C-rich elements, with an average G/C-content higher than either promoters or enhancers (Figure S4E).

In *Drosophila* cells, we noticed a puzzling “divalent” chromatin signature where silent PREs are marked by both H3K4me2 and H3K27me3 (Figure 1B). This particular combination is similar to “bivalent” mammalian promoters, whose H3K4me3 is deposited by MLL2 in mESCs (Hu et al., 2013b). Although we have not yet grasped the significance of bivalent chromatin during development, our comparisons between Trx and MLL1/2 may provide some clues as to how this branch of the COMPASS family has diversified to fulfill similar tasks. For instance, while MLL2-dependent H3K4-trimethylation is truly specific to CpG island-containing bivalent promoters in mESCs, our study did not uncover any evidence for MLL1-dependent bivalency in HCT116 cells. Perhaps Trx's function in implementing H3K4-methylation at “divalent” PREs is now split between MLL2-dependent H3K4me3 at bivalent promoters during early embryogenesis and MLL1-dependent H3K4me2 at PRE-like sequences during later stages of development.

While our identification of over 2,800 PRE-like sequences in human HCT116 cells will contribute to the ongoing work of understanding mammalian epigenetic regulation, several questions still remain. For instance, what is the functional significance of having H3K4me2 at PREs and PRE-like sequences? Can a high ratio of TSS-associated H3K4me2-to-H3K4me3 predict “balanced” genes in other cell types? How can MLL1-binding discriminate among the thousands of unmethylated CGIs throughout the genome? Does CTCF impart specificity to MLL1-binding at the intergenic CGIs? Several reports in *Drosophila* have implicated CTCF in mediating the repressive properties of PREs by enabling long-range PRE-PRE interactions (Li et al., 2011; Li et al., 2013).

Our experiments inhibiting the catalytic activity of EZH2 in MLL1^{NULL} cells have allowed us to test the “balanced” model of MLL1/PcG regulation in human cells (Piunti and Shilatifard, 2016). In this study, we identified 295 active genes (cluster 2) that do not require MLL1 for activation, but only to counteract PRC2-dependent silencing (Figures 4B and 4D), consistent with studies of Trx function at Hox genes (Klymenko and Muller, 2004). One important implication of this finding is that, in the absence of MLL1, ongoing transcription at those genes is not sufficient to block PRC2 recruitment. It will be important for future studies to investigate the sequential order in which these events occur. For instance, when MLL1 is removed from these sites, is PRC2 recruitment a passive or active process? It will also be necessary to understand how the mechanism of PRC2 recruitment to transcriptionally active CpG sequences differs from that of silenced CGIs (Riising et al., 2014). Another interesting aspect of our data concerns cluster 3, and MLL1's inability to block PRC2-repression at those sites. A lot of work remains to be done; however, our study has begun to shed light on the outstanding question of how MLL1/COMPASS and PcG complexes participate to control gene expression in human health and disease.

Methods and Materials

S2 cells

RNAi, CHIP-seq, and RNA-seq were performed with low passage S2 cells, as described in (Herz et al., 2012).

Cell culture and antibodies

HCT116 cells were grown in DMEM medium with 10% FBS. GSK126 was administered for 96 hours at [5 μ M] (McCabe et al., 2012). H3K4me1, H3K4me2, H3K4me3, and H3K27me3 polyclonal antibodies were generated in the Shilatifard lab. H3K27ac monoclonal antibodies were purchased from Cell Signaling Technologies (cat# 8173). MLL1-null cells were created using CRISPR-Cas9 with homologous recombination to create a gene trap.

Lentivirus-mediated RNA interference (RNAi)—Parental HCT116 (MLL3^{null}) cells were infected with lentivirus either for green fluorescent protein (GFP) control short hairpin RNA (shRNA) or MLL1 shRNA in the presence of 8 μ g/ml Polybrene (Sigma) for 24 hr followed by 2 μ g/ml of puromycin for 72 h before harvest.

HCT116 ChIP

ChIP samples were prepared as previously described (Lee et al., 2006), with the exception that fixed chromatin was fragmented to 200-600 bp in length with a Covaris E220 bath sonicator.

Genome Editing

MLL1^{NULL} HCT116 cells were generated using CRISPR/Cas9 to create a gene-trap. sgRNAs were cloned into pX459 (Addgene: 62988) and transfected (with donor DNA) into HCT116 using Lipofectamine 2000 and selecting with Puromycin (2 mg/ml) for 24 hours.

Accession Numbers

Sequencing data have been deposited at the GEO under the accession number GSE81795.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Drs. Edwin Smith and Marc Morgan for conversations throughout the study and for critical reading of the manuscript and Laura Shilatifard for editorial assistance. Studies on *Drosophila* Trx and mammalian MLL1 in the Shilatifard laboratory were supported in part by a grant from the National Institutes of Health R35CA197569.

References

- Arnold CD, Gerlach D, Stelzer C, Boryn LM, Rath M, Stark A. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science*. 2013; 339:1074–1077. [PubMed: 23328393]
- Bauer M, Trupke J, Ringrose L. The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma*. 2015
- Beisel C, Buness A, Roustan-Espinosa IM, Koch B, Schmitt S, Haas SA, Hild M, Katsuyama T, Paro R. Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. *Proceedings of the National Academy of Sciences*. 2007; 104:16615–16620.
- Bracken AP. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & Development*. 2006; 20:1123–1136. [PubMed: 16618801]

- Capotosti F, Hsieh JJD, Herr W. Species Selectivity of Mixed-Lineage Leukemia/Trithorax and HCF Proteolytic Maturation Pathways. *Molecular and Cellular Biology*. 2007; 27:7063–7072. [PubMed: 17698583]
- Caslini C, Connelly JA, Serna A, Broccoli D, Hess JL. MLL Associates with Telomeres and Regulates Telomeric Repeat-Containing RNA Transcription. *Molecular and Cellular Biology*. 2009; 29:4519–4526. [PubMed: 19528237]
- Cavalli G, Paro R. Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science*. 1999; 286:955–958. [PubMed: 10542150]
- Chan CS, Rastelli L, Pirrotta V. A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J*. 1994; 13:2553–2564. [PubMed: 7912192]
- Chinwalla V, Jane EP, Harte PJ. The *Drosophila* trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. *EMBO J*. 1995; 14:2056–2065. [PubMed: 7744011]
- Cooper S, Dienstbier M, Hassan R, Schermelleh L, Sharif J, Blackledge NP, De Marco V, Elderkin S, Koseki H, Klose R, et al. Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep*. 2014; 7:1456–1470. [PubMed: 24857660]
- Enderle D, Beisel C, Stadler MB, Gerstung M, Athri P, Paro R. Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. *Genome Research*. 2010; 21:216–226. [PubMed: 21177970]
- Francis NJ, Kingston RE. Mechanisms of transcriptional memory. *Nature Reviews Molecular Cell Biology*. 2001; 2:409–421. [PubMed: 11389465]
- Gohl D, Muller M, Pirrotta V, Affolter M, Schedl P. Enhancer Blocking and Transvection at the *Drosophila* *apterous* Locus. *Genetics*. 2008; 178:127–143. [PubMed: 18202363]
- Hagstrom K, Muller M, Schedl P. Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the *Drosophila* bithorax complex. *Genes & Development*. 1996; 10:3202–3215. [PubMed: 8985188]
- Herz HM, Mohan M, Garruss AS, Liang K, Takahashi Yh, Mickey K, Voets O, Verrijzer CP, Shilatifard A. Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. *Genes & Development*. 2012; 26:2604–2620. [PubMed: 23166019]
- Hsieh JJD, Ernst P, Erdjument-Bromage H, Tempst P, Korsmeyer SJ. Proteolytic Cleavage of MLL Generates a Complex of N- and C-Terminal Fragments That Confers Protein Stability and Subnuclear Localization. *Molecular and Cellular Biology*. 2003; 23:186–194. [PubMed: 12482972]
- Hu D, Gao X, Morgan MA, Herz HM, Smith ER, Shilatifard A. The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and Cellular Biology*. 2013a; 33:4745–4754. [PubMed: 24081332]
- Hu D, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, Shilatifard A. The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol*. 2013b; 20:1093–1097. [PubMed: 23934151]
- Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP. Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet*. 2010; 6:e1001134. [PubMed: 20885785]
- Kennison JA, Tamkun JW. Dosage-dependent modifiers of polycomb and antennapedia mutations in *Drosophila*. *Proceedings of the National Academy of Sciences*. 1988; 85:8136–8140.
- Klose RJ, Cooper S, Farcas AM, Blackledge NP, Brockdorff N. Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS Genet*. 2013; 9:e1003717. [PubMed: 23990804]
- Klymenko T, Muller J. The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep*. 2004; 5:373–377. [PubMed: 15031712]
- Kvon EZ, Kazmar T, Stampfel G, Yáñez-Cuna JO, Pagani M, Schernhuber K, Dickson BJ, Stark A. Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature*. 2014

- Lee TI, Johnstone SE, Young RA. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc.* 2006; 1:729–748. [PubMed: 17406303]
- Levine SS, King IF, Kingston RE. Division of labor in polycomb group repression. *Trends Biochem Sci.* 2004; 29:478–485. [PubMed: 15337121]
- Li HB, Muller M, Bahechar IA, Kyrchanova O, Ohno K, Georgiev P, Pirrotta V. Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in *Drosophila melanogaster*. *Mol Cell Biol.* 2011; 31:616–625. [PubMed: 21135119]
- Li HB, Ohno K, Gui H, Pirrotta V. Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. *PLoS Genet.* 2013; 9:e1003436. [PubMed: 23637616]
- Margueron R, Reinberg D. The Polycomb complex PRC2 and its mark in life. *Nature.* 2011; 469:343–349. [PubMed: 21248841]
- Maurano MT, Wang H, John S, Shafer A, Canfield T, Lee K, Stamatoyannopoulos JA. Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Rep.* 2015; 12:1184–1195. [PubMed: 26257180]
- McCabe MT, Ott HM, Ganji G, Korenchuk S, Thompson C, Van Aller GS, Liu Y, Graves AP, Della Pietra A 3rd, Diaz E, et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature.* 2012; 492:108–112. [PubMed: 23051747]
- McLean CY, Bristol D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.* 2010; 28:495–501. [PubMed: 20436461]
- Mendenhall EM, Koche RP, Truong T, Zhou VW, Issac B, Chi AS, Ku M, Bernstein BE. GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* 2010; 6:e1001244. [PubMed: 21170310]
- Milne TA, Dou Y, Martin ME, Brock HW, Roeder RG, Hess JL. From The Cover: MLL associates specifically with a subset of transcriptionally active target genes. *Proceedings of the National Academy of Sciences.* 2005; 102:14765–14770.
- Mohan M, Herz HM, Shilatifard A. SnapShot: Histone Lysine Methylase Complexes. *Cell.* 2012; 149:498–498.e491. [PubMed: 22500810]
- Mohan M, Herz HM, Smith ER, Zhang Y, Jackson J, Washburn MP, Florens L, Eissenberg JC, Shilatifard A. The COMPASS Family of H3K4 Methylases in *Drosophila*. *Molecular and Cellular Biology.* 2011; 31:4310–4318. [PubMed: 21875999]
- Piunti A, Shilatifard A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science.* 2016; 352:aad9780. [PubMed: 27257261]
- Poux S, Horard B, Sigrist CJ, Pirrotta V. The *Drosophila* trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. *Development.* 2002; 129:2483–2493. [PubMed: 11973279]
- Riising EM, Comet I, Leblanc B, Wu X, Johansen JV, Helin K. Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell.* 2014; 55:347–360. [PubMed: 24999238]
- Ringrose L, Rehmsmeier M, Dura JM, Paro R. Genome-Wide Prediction of Polycomb/Trithorax Response Elements in *Drosophila melanogaster*. *Developmental Cell.* 2003; 5:759–771. [PubMed: 14602076]
- Schuettengruber B, Martinez AM, Iovino N, Cavalli G. Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol.* 2011; 12:799–814. [PubMed: 22108599]
- Schwartz YB, Kahn TG, Stenberg P, Ohno K, Bourgon R, Pirrotta V. Alternative Epigenetic Chromatin States of Polycomb Target Genes. *PLoS Genetics.* 2010; 6:e1000805. [PubMed: 20062800]
- Shilatifard A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu Rev Biochem.* 2012; 81:65–95. [PubMed: 22663077]
- Simon J, Chiang A, Bender W, Shimell MJ, O'Connor M. Elements of the *Drosophila* Bithorax Complex That Mediate Repression by Polycomb Group Products. *Developmental Biology.* 1993; 158:131–144. [PubMed: 8101171]

- Simon JA, Kingston RE. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol.* 2009; 10:697–708. [PubMed: 19738629]
- Simon JA, Kingston RE. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell.* 2013; 49:808–824. [PubMed: 23473600]
- Takeda S. Proteolysis of MLL family proteins is essential for Taspase1-orchestrated cell cycle progression. *Genes & Development.* 2006; 20:2397–2409. [PubMed: 16951254]
- Tanay A, O'Donnell AH, Damelin M, Bestor TH. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc Natl Acad Sci U S A.* 2007; 104:5521–5526. [PubMed: 17376869]
- Wang P, Lin C, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM, et al. Global Analysis of H3K4 Methylation Defines MLL Family Member Targets and Points to a Role for MLL1-Mediated H3K4 Methylation in the Regulation of Transcriptional Initiation by RNA Polymerase II. *Molecular and Cellular Biology.* 2009; 29:6074–6085. [PubMed: 19703992]
- Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, Shilatifard A. Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS. *Molecular and Cellular Biology.* 2008; 28:7337–7344. [PubMed: 18838538]

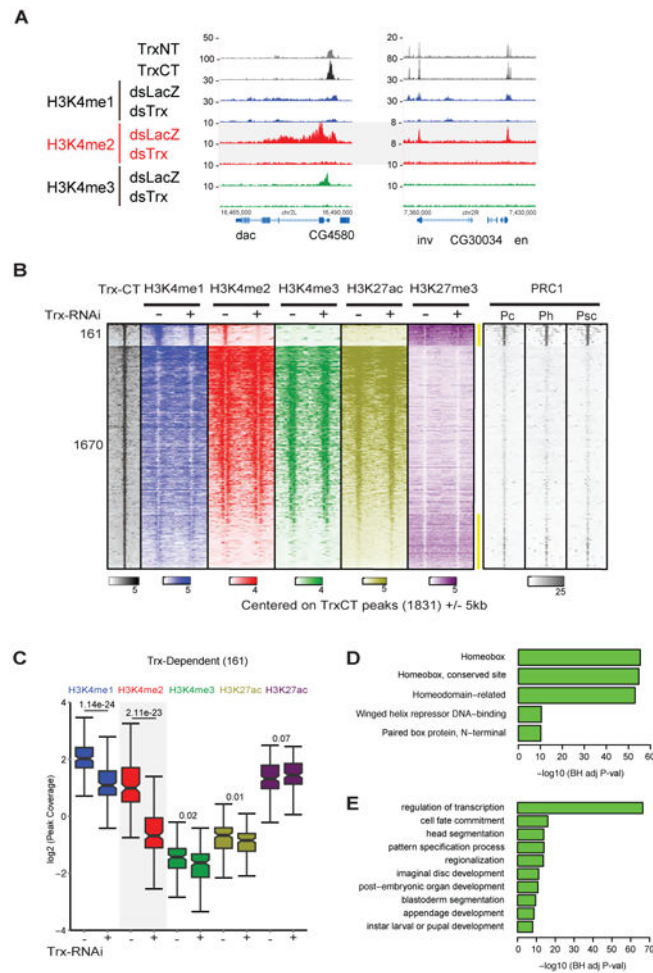


Figure 1. A divalent, Trx-dependent H3K4me2 chromatin signature marks PREs in *Drosophila* (A) Track examples of PREs at *dachshund* (*dac*), *invected* (*inv*), and *engrailed* (*en*). Both Trx and PRC1 remain bound at the PRE regardless of the gene's transcriptional activity. In the active state, Trx-dependent H3K4-methylation is spread across the gene body (*dac*), while in the repressed state, H3K4me1/2 is confined to the PRE (*inv*, *en*). (B) Analysis of ChIP-seq data after Trx-RNAi. Occupancy levels of Trx-CT and for H3K4me1/2/3, H3K27ac, and H3K27me3 +/- Trx-RNAi. Profiles are centered on Trx-CT occupied peaks (+/- 5kb) and sorted in descending order of H3K4me2 occupancy in WT cells. Group 1 (161 peaks) is distinguished by significantly decreased H3K4me2 after Trx-RNAi and overlap with PRC1-defined PREs. PRC1 ChIP data was obtained from (Enderle et al., 2010). (C) The average occupancy levels per bp within each of the 161 Trx peaks were determined for the five histone modifications and presented in the boxplot. P-values from two-sided student T-tests comparing occupancy +/- Trx-RNAi are shown in the figure. (D,E) For the nearest genes associated with the 161 sites (127 genes), the top 5 Interpro protein domain enrichment results and select top GO biological process results are displayed.

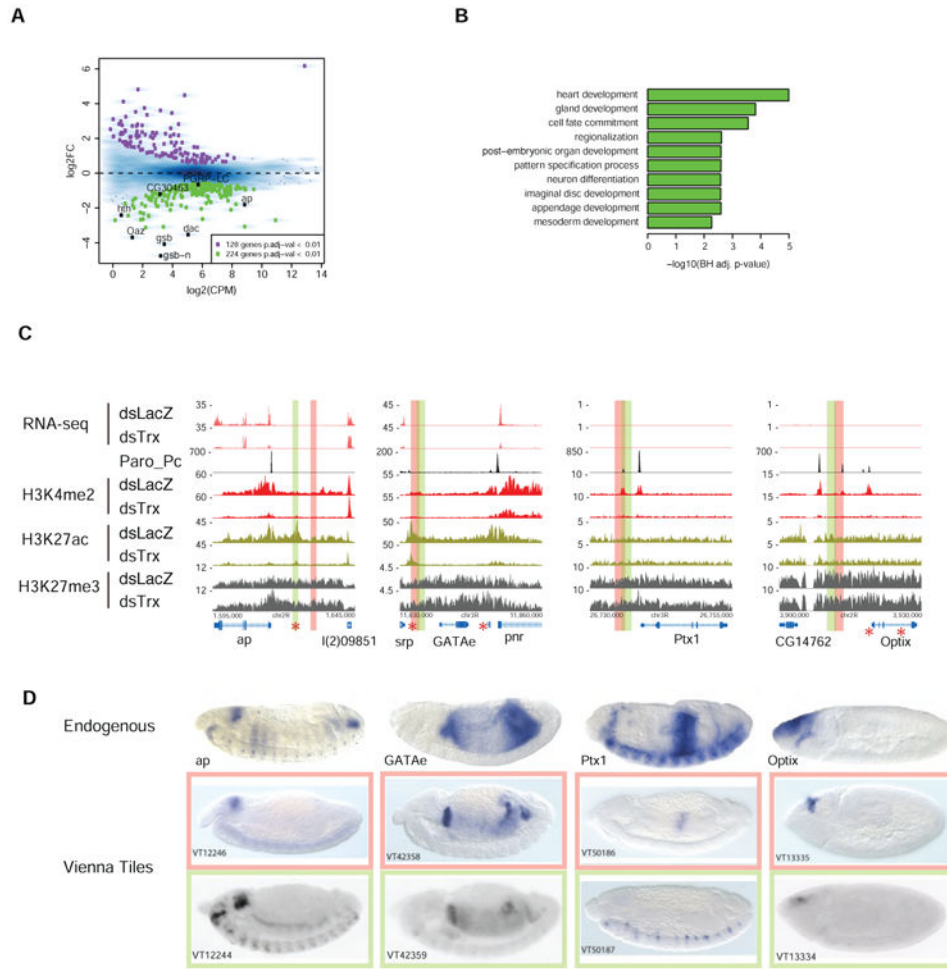


Figure 2. Trx-depletion negatively affects transcription at PRE-regulated genes, and results in decreased H3K4-methylation and H3K27ac at nearby enhancers

(A) Analysis of RNA-seq after Trx-RNAi. MA-plot shows of the 224 down-regulated genes (green boxes), 8 overlap with 161 PRE-nearest genes (black boxes). (B) For the nearest genes associated with the 161 sites (127 genes), select top GO biological process results are displayed. (C) Track examples of decreased transcription at *pnr* and *ap* following Trx-RNAi, and accompanying changes in chromatin modification at these genes, as well as at *Ptx1* and *Optix*. PRC1 remains bound at the PRE whether the gene is active or repressed. STARR-seq enhancers identified in S2 cells (red star) and embryonic enhancers (red and green columns) show the close, sometimes overlapping, proximity to PREs. Note the effect of Trx-RNAi on H3K4me2 and H3K27ac at *ap* enhancers. (D) PRE-associated enhancers give embryonic expression patterns that resemble that of the endogenous gene. Colored boxes in C correspond to a ~2kb fragment from the Vienna Tiles library that drives the corresponding LacZ expression patterns shown in D. Note, the *Ptx1* endogenous expression pattern is a composite of the two Vienna Tiles (peach and green boxes) that partially overlap at the *Ptx1* PRE. Embryos are stages 13-14, except for *optix* (stage 8-9). Vienna Tile embryo images are from (Kvon et al., 2014). Red asterisks indicate STARR-seq enhancers in S2 cells (Arnold et al., 2013).

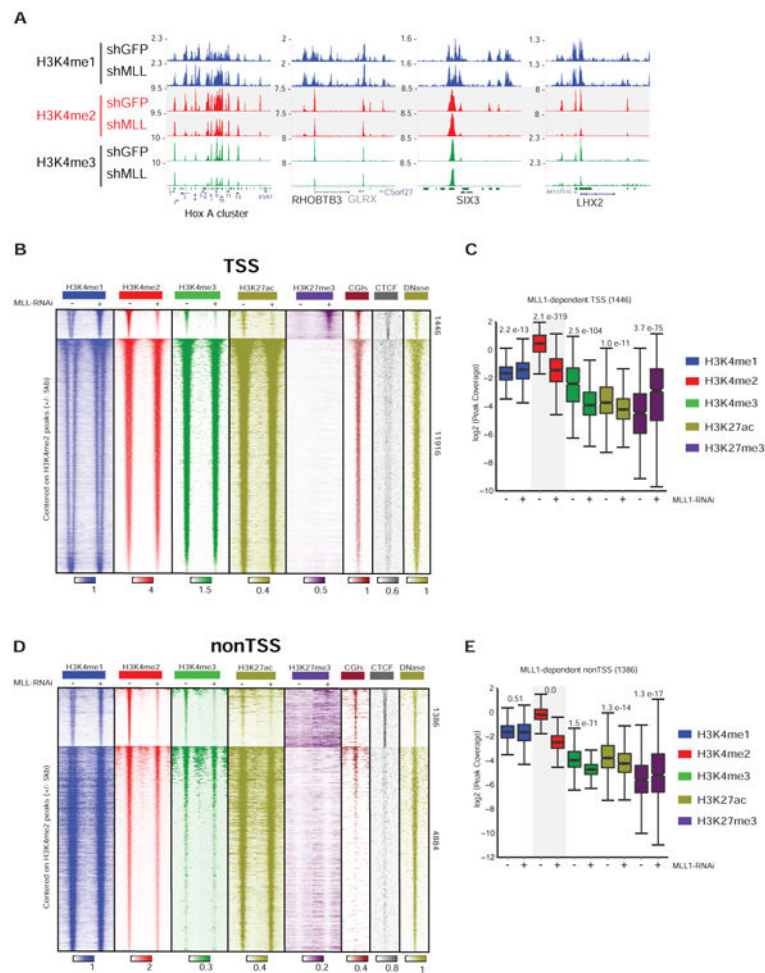


Figure 3. MLL1-dependent H3K4me2 marks PRE-like sequences in the human genome
 (A) Genome-browser track examples of H3K4me1/2/3 ChIP-seq +/- shMLL1 highlighting the loss of H3K4me2 at numerous CGIs near several homeodomain-containing genes. Green boxes represent CGIs. (B) Coverage profiles for H3K4me1/2/3, H3K27ac, and H3K27me3 +/- shMLL1. Heat maps are centered on H3K4me2 peaks and ranked by decreasing H3K4me2 signal at TSS-overlapping sites. 1446 genes show significantly decreased H3K4me2 and increased H3K27me3 levels following MLL1 knockdown. Note the focused CTCF pattern at MLL1-dependent loci compared to 11916 other TSS. (C) Box plots quantifying histone modification levels shown in (B). P-values from two-sided student T-tests comparing occupancy +/- MLL1-RNAi are shown above plots. (D) Coverage profiles of histone modifications +/- shMLL1 at intergenic sites not overlapping with an annotated TSS. Note the association of H3K4me2 loss with regions of high CTCF occupancy. (E) Box plots quantifying data shown in (D). H3K4me2 is most significantly affected among the modifications considered. P-values, as calculated in (C) are displayed above plots.

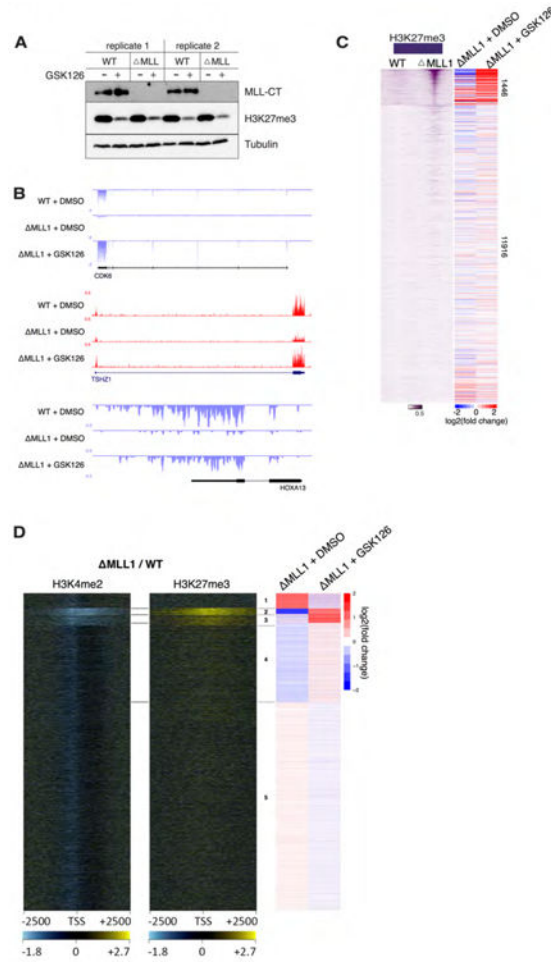


Figure 4. MLL1 target-genes are transcriptionally balanced by PRC2

(A) Western blot showing lack of MLL1 protein in MLL1^{NULL} HCT116 cells and diminished H3K27me3 levels following 4 day GSK126 treatment. (B) Representative RNA-seq track examples of MLL1 target-genes whose expression is rescued in MLL1^{NULL} cells following GSK126 treatment. (C) Left panel, heatmaps ordered identically to that in Fig. 2C, showing increased TSS-associated H3K27me3 in MLL1^{NULL} cells. Right panels, heatmaps showing the corresponding log₂ fold-changes in gene expression in MLL1^{NULL} and MLL1^{NULL} + GSK126, compared to WT cells. Note the substantial changes in gene expression coinciding with MLL1-dependent group. (D) K-means clustering was used to partition all protein-coding genes by their log₂ fold-changes in expression following MLL1^{NULL} treatment with GSK126. To the left are heatmaps showing decreased H3K4me2 and increased H3K27me3 in MLL1^{NULL} cells. Notice that changes in these modifications correspond exactly with the MLL1-dependent “balanced” genes (cluster 2) and other remaining MLL1 target-genes that are de-repressed following GSK126 treatment (cluster 3).