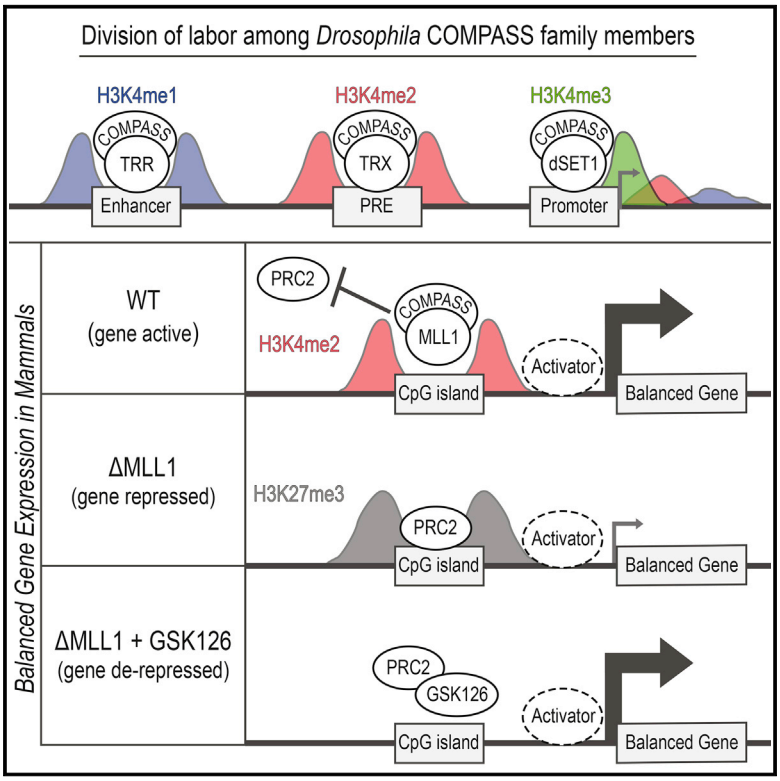
Short Article

Molecular Cell

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

Graphical Abstract



Highlights

- Trx-dependent H3K4me2 marks the majority of *Drosophila* Polycomb response elements
- This H3K4-dimethylase function is conserved in MLL1, the mammalian Trx homolog
- MLL1-dependent H3K4me2 occurs predominantly at CpG islands
- In HCT116 cells, ~300 active genes require MLL1 to block PRC2-dependent repression

Authors

Ryan Rickels, Deqing Hu, Clayton K. Collings, ..., Hans-Martin Herz, Evgeny Kvon, Ali Shilatifard

Correspondence

ash@northwestern.edu

In Brief

The extent to which mammalian CpG islands functionally resemble *Drosophila* Polycomb response elements (PREs) remains unclear. Rickels et al. describe a conserved role for Trx/MLL1 as an H3K4-dimethylase at PREs and CpG islands, respectively. HCT116 cells express ~300 genes that require MLL1, not for activation but to block PRC2-dependent repression.

Accession Numbers

GSE81795

Rickels et al., 2016, Molecular Cell *63*, 318–328
July 21, 2016
http://dx.doi.org/10.1016/j.molcel.2016.06.018

CellPress

Molecular Cell
Short Article

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

Ryan Rickels,[1] Deqing Hu,[1] Clayton K. Collings,[1] Ashley R. Woodfin,[1] Andrea Piunti,[1] Man Mohan,[2,5] Hans-Martin Herz,[2,6] Evgeny Kvon,[3] and Ali Shilatifard[1,4,*]

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

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

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

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

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

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

*Correspondence: ash@northwestern.edu

http://dx.doi.org/10.1016/j.molcel.2016.06.018

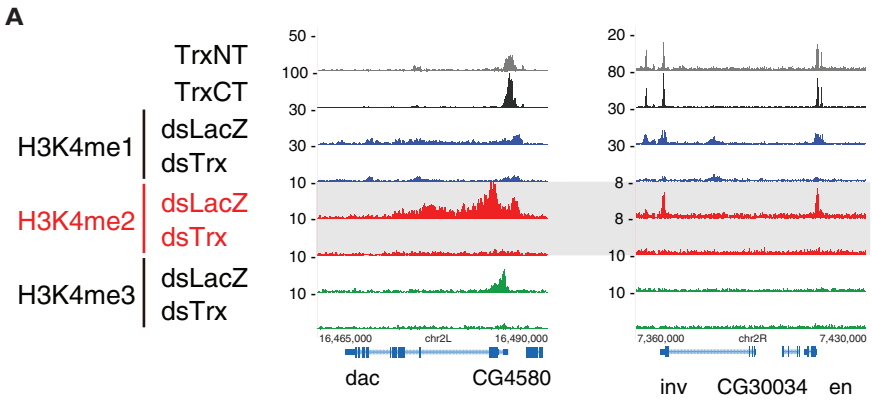
SUMMARY

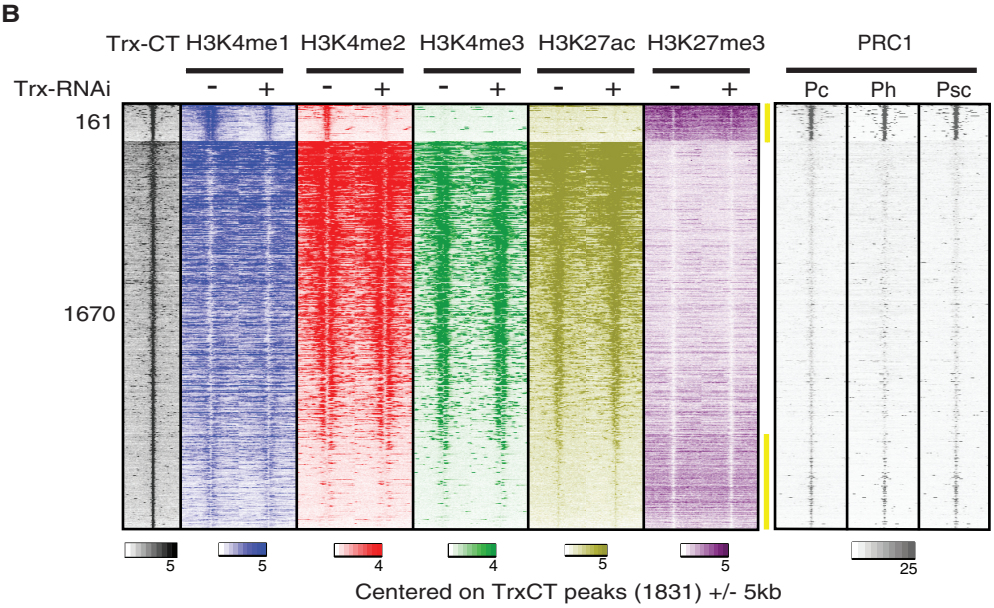
Polycomb response elements (PREs) are specific DNA sequences that stably maintain the developmental pattern of gene expression. *Drosophila* PREs are well characterized, whereas the existence of PREs in mammals remains debated. Accumulating evidence supports a model in which CpG islands recruit Polycomb group (PcG) complexes; however, which subset of CGIs is 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 homolog, 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.

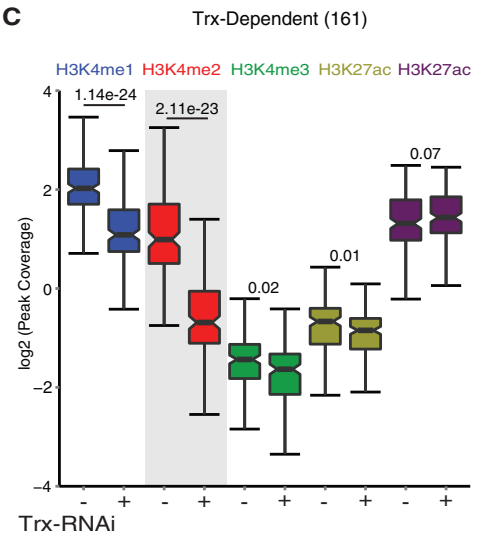
INTRODUCTION

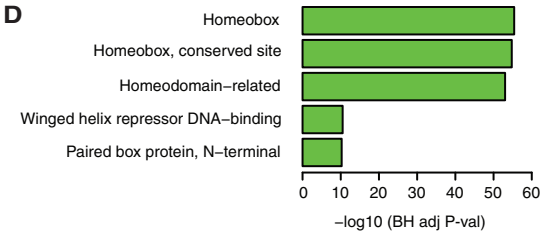
Histone H3K4 methylation by the COMPASS family (complex of proteins associated with Set1) is conserved from yeast to mammals and is closely associated with transcriptionally active chromatin (Shilatifard, 2012). Interest in yeast Set1 arose due to its ancestral homology with mammalian MLL1, a gene commonly 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 trimethylation (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 cells (mESCs), and Trr/MLL3/MLL4 recently were shown to catalyze H3K4 monomethylation at *Drosophila* and mammalian enhancers, respectively (Herz et al., 2012; Hu et al., 2013a, 2013b).

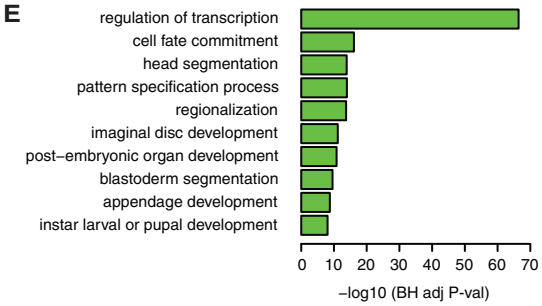
In *Drosophila*, Trx maintains the active transcription state of a target gene in the absence of the initiating signal, thus maintaining epigenetic memory of previous transcriptional states (Poux et al., 2002). This maintenance function is known to rely on Trx binding to a nearby Polycomb response element (PRE) (Klymenko and Müller, 2004; Schuettengruber et al., 2011). While *Drosophila* PREs are well documented, the existence of comparable elements in mammals has been debated (Bauer et al., 2016; 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 the 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 Polycomb group (PcG) complexes when transcription is globally inhibited, suggesting that any CGI can function as a PRE (Riising











(legend on next page)

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 paper) (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 we report the identification of over 1,400 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., 2011, 2012). 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 chromatin immunoprecipitation sequencing (ChIP-seq) for H3K4me1/2/3 after reducing 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 1A). Genome-wide analysis identified 161 high-confidence Trx-CT (SET domain-containing) occupied sites at which H3K4me2 was significantly reduced following Trx RNAi (Figures 1B and 1C; Figure S1C). Although Trr was shown to co-bind with Trx at these sites, H3K4me2 levels were not affected by Trr depletion (Figure S1B). Surprisingly, these Trx-dependent sites also contained 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., 2011) (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., 2011; Schwartz et al., 2010; Simon and Kingston, 2013). Gene ontology (GO) 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 PREs/Trithorax response elements (Beisel et al., 2007; Enderle et al., 2011; 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 revealed that, after Trx depletion, PRE-associated genes with decreased H3K4me2 levels also exhibited reduced transcription (Figure 2A). This could be observed despite the fact that the majority of PRE-associated genes were already transcriptionally silent in S2 cells (Figure S2A). GO terms associated with the downregulated genes are enriched for developmental processes (Figure 2B). Interestingly, Trx-CT not only remained bound but also retained 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 was confined to the PRE, while in the transcriptionally active state H3K4me2 was spread across the gene body. The extent to which H3K4me2 spread from the active PRE matched the transcriptional activity of the nearby Trx-regulated gene. For example, at the transcribed *apterous* (*ap*) locus, H3K4me2 spread into the gene body and also across a ~20-kb 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 coincided 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).

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* and *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 are shown. Profiles are centered on Trx-CT-occupied peaks (±5 kb) 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 overlaps with PRC1-defined PREs. PRC1 ChIP data were obtained from Enderle et al. (2011).

(C) The average occupancy levels, per base pair within each of the 161 Trx peaks, were determined for the five histone modifications and presented in the boxplot. The p values from two-sided Student's t tests comparing occupancy ± Trx RNAi are shown in the figure.

(D and E) For the nearest genes associated with the 161 sites (127 genes), the (D) top five Interpro protein domain enrichment results and (E) select top GO biological process results are displayed.

See also Figure S1.

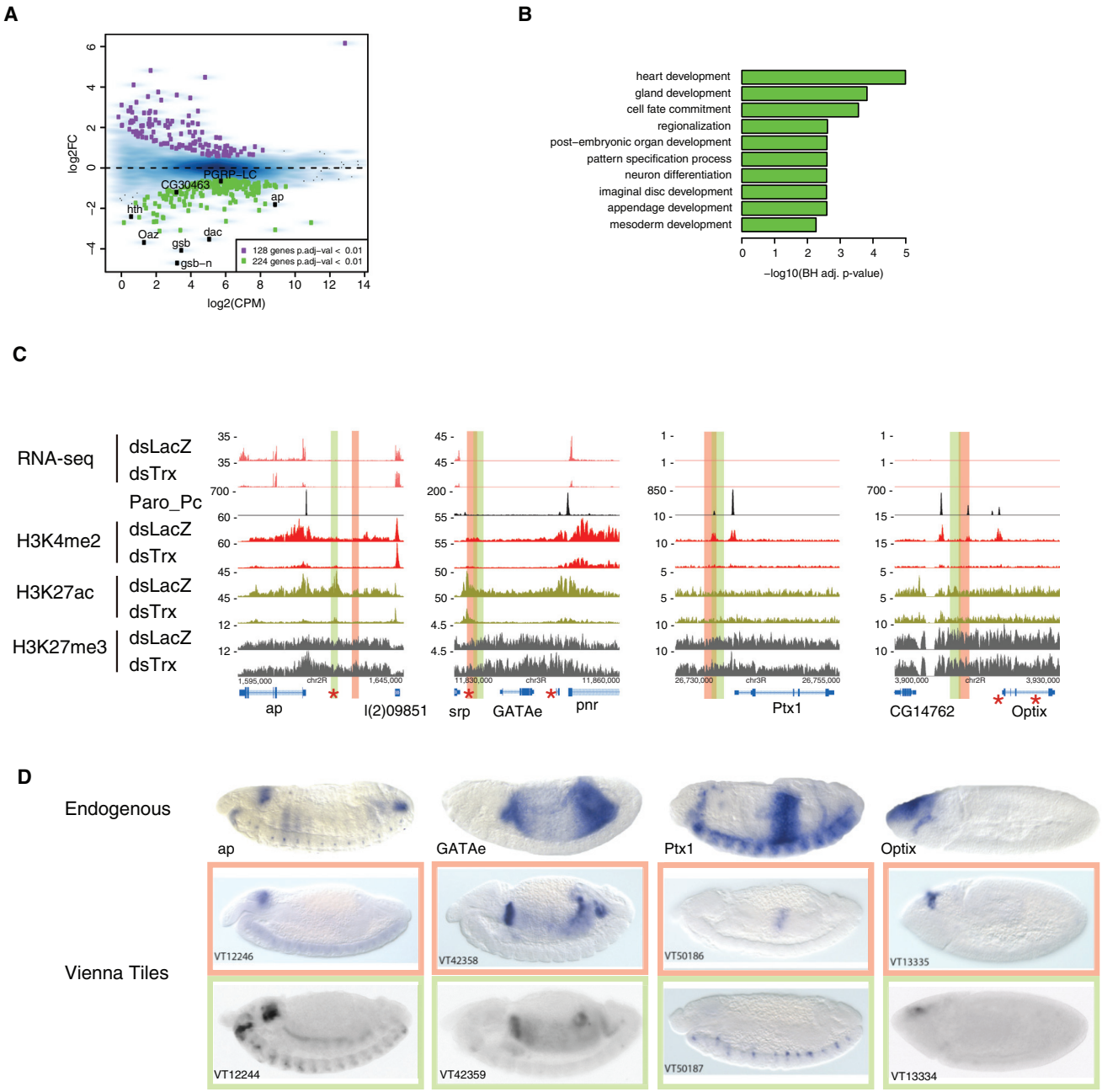


Figure 2. Trx Depletion Negatively Affects Transcription at PRE-Regulated Genes and Results in Decreased H3K4 Methylation and H3K27ac at Nearby Enhancers

(A) Analysis of RNA-seq after Trx RNAi. MA plot shows that, of the 224 downregulated genes (green boxes), eight 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 those of the endogenous gene. Colored boxes in (C) correspond to an ~2-kb 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* (stages 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).

See also Figure S2.

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 self-transcribing active regulatory region-sequencing (STARR-seq) (Arnold et al., 2013). Although we did not find a significant overlap between PREs and STARR-seq enhancers, we did see examples of PREs in close or overlapping proximity to enhancers, such as at *Ptx1* and *Optix* (Figure 2C). This was 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 Figures 2C and 2D also was 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 contributed to its expression in the midgut and ventral nerve cord (peach and green boxes, respectively). Even in the repressed state, these *Ptx1* enhancers overlapped 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 were enriched for H3K4me2 and H3K27ac, all of which were 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.

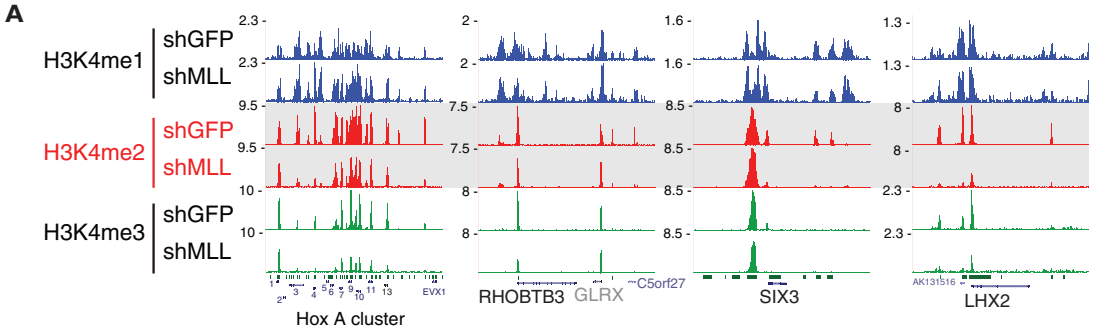
A Conserved Role for MLL1/COMPASS as H3K4 Dimethylase at PRE-like Regions in the Human Genome

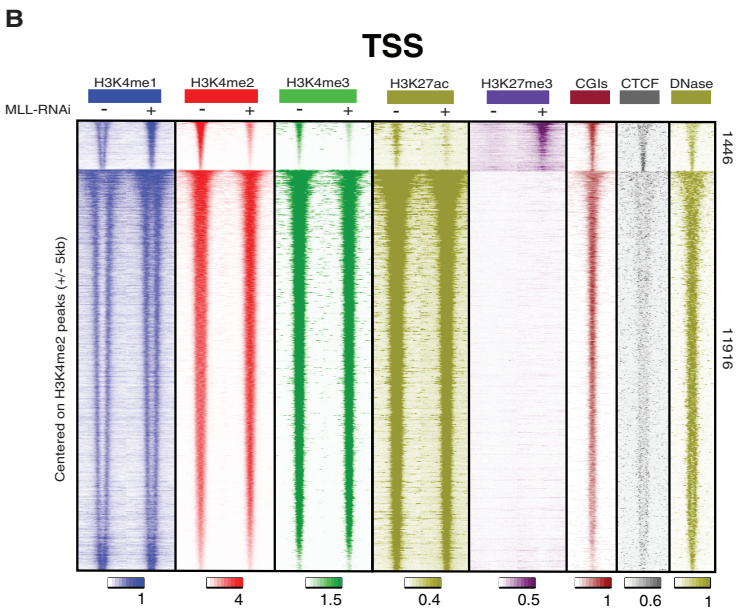
The human orthologs of Trx are MLL1 and MLL2 (Kmt2A and Kmt2B, respectively) (Mohan et al., 2012; Shilatifard, 2012). MLL1 is known to play a central role in the deposition of H3K4me3 at Hox gene promoters, while MLL2 is required for H3K4me3 at bivalent gene clusters in ESCs (Hu et al., 2013b; Wang et 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 was not expressed in HCT116 cells, obviating concerns regarding any compensatory function of MLL2 (Figure S3B). Although bulk H3K4 methylation was unaffected in the absence of MLL1 (Figure S3C), H3K4me2 levels were dramatically reduced at a total of 2,832 locations in the genome. After centering on total H3K4me2 peaks, we divided the data into transcription start site (TSS)- and non-TSS-overlapping peaks. Intriguingly, only 51% (1,446) of MLL1-dependent H3K4me2 peaks overlapped an annotated TSS, while the remaining 49% (1,386) did not overlap an 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 predominated over H3K4me3 and were more significantly affected at MLL1 target TSSs compared to MLL1-independent TSSs (Figure 3C; Figure S3D). Upon MLL1 depletion, H3K27me3 levels were 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).

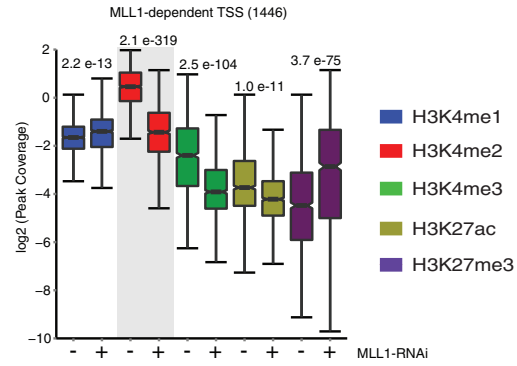
At the 1,386 MLL1-dependent non-TSSs, we saw identical changes with regard to H3K4 methylation; however, the effects on H3K27ac and H3K27me3 were less drastic compared to TSSs (Figures 3D and 3E). To our surprise, the MLL1-dependent sites were significantly enriched for CGIs ($p = 2.99e{-}221$, Pearson's chi-square 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 non-TSSs 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 non-TSSs (p value = 1e−167, reported by Hypergeometric Optimization of Motif Enrichment [HOMER]). This result was confirmed by analyzing recently published ChIP-seq data for CTCF in HCT116 cells, which showed a significant enrichment (p value < 2.2e−16, Pearson's chi-square test) for CTCF co-binding at MLL1-dependent non-TSSs 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.

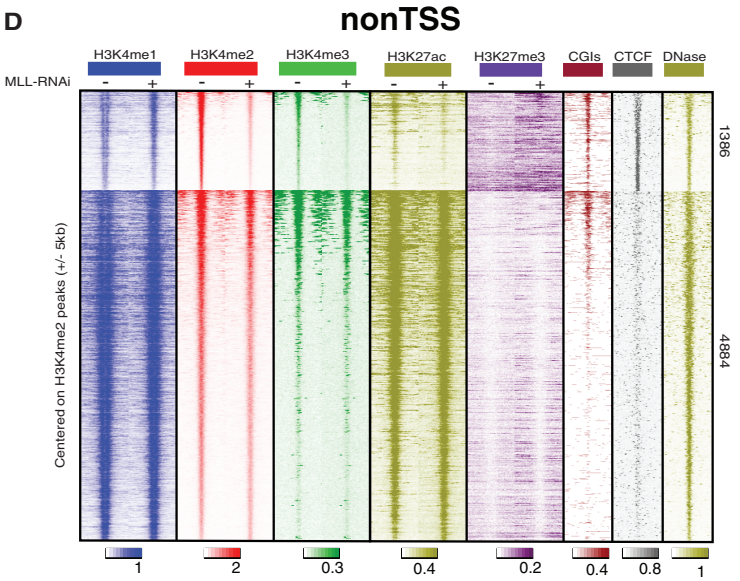
Loss of MLL1-Dependent H3K4me2 at Non-TSSs Accompanies Downregulation of Nearby Target Genes

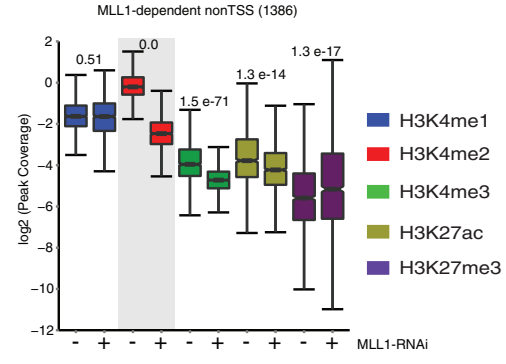
MLL1 is known to positively regulate transcription of a diverse set of genes in different cell types and under distinct cellular environments (Caslini et al., 2009; Milne et al., 2005; Takeda et al., 2006). We performed RNA sequencing (RNA-seq) to assess genome-wide transcriptional changes upon MLL1 depletion in HCT116, and we identified 550 significantly downregulated genes (Figure S4A). Next we wanted to know whether the MLL1-dependent H3K4me2 sites tend to be located near downregulated genes or distributed randomly. To evaluate this, we identified 2,434 genes nearest to the MLL1-dependent sites (TSS and non-TSS), and we calculated the overlap with the 550 downregulated genes (Figure S4A). Of the 550 downregulated genes, 8% (44) overlapped with non-TSSs nearest genes and 32% (179) overlapped with TSSs 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, Genomic Regions Enrichment of Annotations Tool (GREAT) analysis of the nearest genes to











(legend on next page)

both MLL1-dependent TSS and non-TSS elements retrieved similar GO terms, indicating these elements associate with the same set of genes (Figures S4C and S4D) (McLean et al., 2010). These results imply that MLL1/COMPASS-regulated genes utilize a combination of TSS- and non-TSS-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 matches 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 exists 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 decreased in the *MLL1*-null but returned to near wild-type (WT) levels upon treatment with EZH2 inhibitor GSK126 (Figures 4B and 4C). As expected, the MLL1-dependent TSSs that gained the most H3K27me3 associated with the largest decreases in gene expression upon deletion of MLL1 (Figure 4C). Interestingly, the majority of genes with MLL-dependent H3K4me2 at their TSSs were de-repressed following GSK126 treatment, indicating PRC2 and H3K27me3 played 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 revealed 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 recruited PRC2 to their TSSs after MLL1 deletion, as evidenced by increased H3K27me3 at these specific locations (Figure 4D). Importantly, cluster 2 was significantly enriched ($p = 3.88e{-}28$, hypergeometric test) for MLL1-dependent H3K4me2. Cluster 3 closely resembled cluster 2 with regard to MLL1-dependent H3K4me2, increased H3K27me3, and transcriptional de-repression following GSK126 treatment. However, the 488 genes in this group were repressed by PRC2, despite the presence of MLL1, and were thus more lowly expressed in WT cells compared to cluster 2 (Figure 4D). Cluster 1 showed increased gene expression after removing MLL1; however, there were no changes in H3K4me2 or H3K27me3, indicating the transcriptional changes seen in this group were 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 et al., 2006; Wang et al., 2009). It is interesting how even modest gene expression changes in the MLL1-independent clusters 1, 4, and 5 were 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 has increased in number and diversified in function (Shilatifard, 2012). However, we continue to uncover remarkable similarities in the 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 et al., 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 we 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

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.

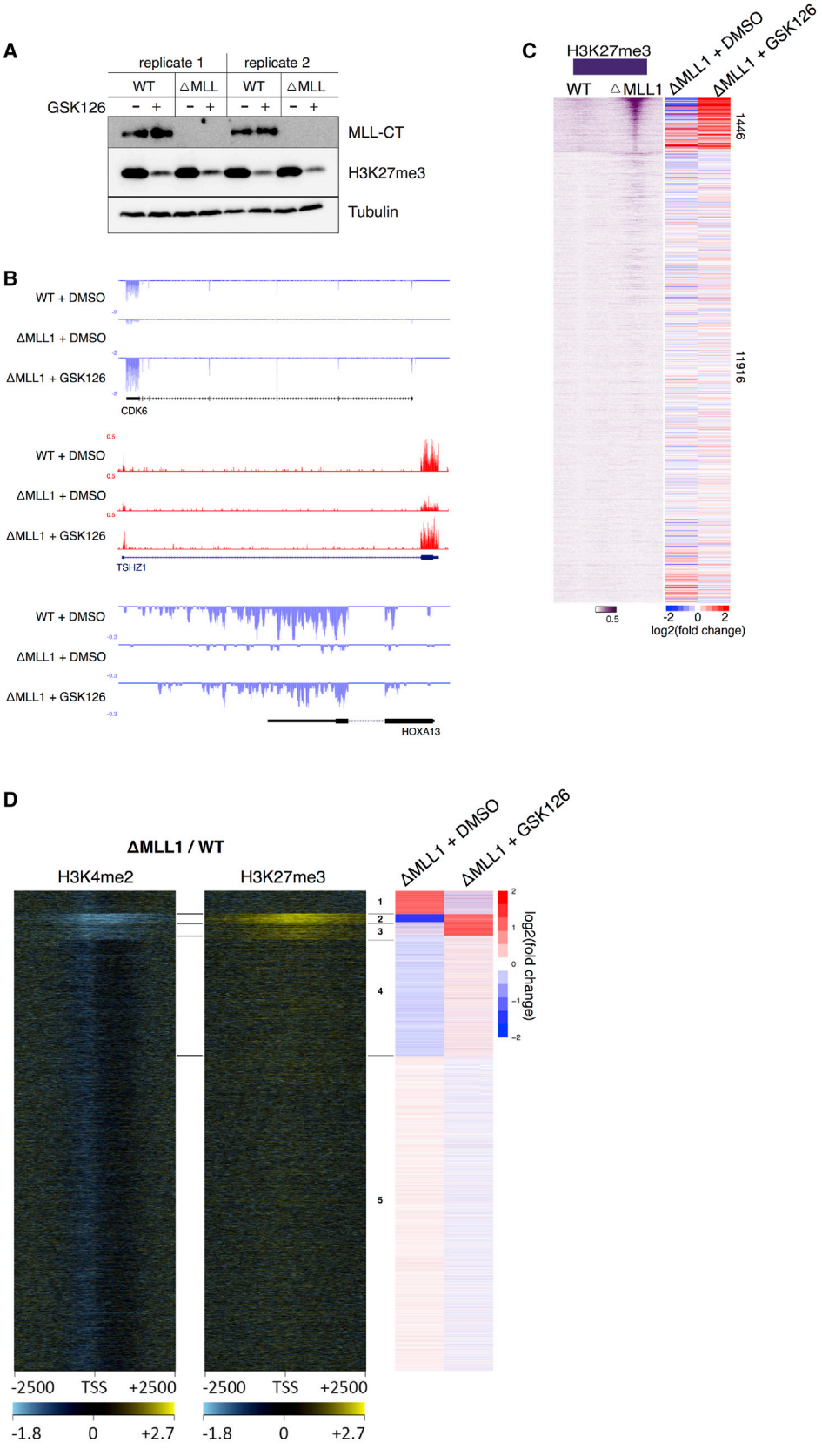
(B) Coverage profiles for H3K4me1/2/3, H3K27ac, and H3K27me3 ± shMLL1. Heatmaps are centered on H3K4me2 peaks and ranked by decreasing H3K4me2 signal at TSS-overlapping sites. 1,446 genes show significantly decreased H3K4me2 and increased H3K27me3 levels following MLL1 knockdown. Note the focused CTCF pattern at MLL1-dependent loci compared to 11,916 other TSSs.

(C) Boxplots quantifying histone modification levels shown in (B). The p values from two-sided Student's t tests comparing occupancy ± MLL1 RNAi are shown above plots.

(D) Coverage profiles of histone modifications ± shMLL at intergenic sites not overlapping with an annotated TSS. Note the association of H3K4me2 loss with regions of high CTCF occupancy.

(E) Boxplots quantifying data shown in (D). H3K4me2 is most significantly affected among the modifications considered. The p values as calculated in (C) are displayed above plots.

See also Figure S3.



(legend on next page)

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 CGI-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, 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 Müller, 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 also will 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.

EXPERIMENTAL PROCEDURES

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 with 10% fetal bovine serum (FBS). GSK126 was administered for 96 hr at 5 μM (McCabe et al., 2012). H3K4me1, H3K4me2, H3K4me3, and H3K27me3 polyclonal antibodies were generated in the A.S. lab. H3K27ac monoclonal antibodies were purchased from Cell Signaling Technologies (8173). *MLL1*-null cells were created using CRISPR-Cas9 with homologous recombination to create a gene trap.

Lentivirus-Mediated RNAi

Parental HCT116 (*MLL3*-null) cells were infected with lentivirus either for GFP control short hairpin RNA (shRNA) or MLL1 shRNA in the presence of 8 μg/ml Polybrene (Sigma) for 24 hr, followed by 2 μg/ml puromycin for 72 hr 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. Single-guide RNAs (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 hr.

ACCESSION NUMBERS

The accession number for the sequencing data reported in this paper is GEO: GSE81795.

Figure 4. MLL1 Target Genes Are Transcriptionally Balanced by PRC2

(A) Western blot shows lack of MLL1 protein in *MLL1*-null HCT116 cells and diminished H3K27me3 levels following 4-day GSK126 treatment.

(B) Representative RNA-seq track examples show MLL1 target genes whose expression is rescued in *MLL1*-null cells following GSK126 treatment.

(C) Left panel: heatmaps, ordered identically to that in Figure 2C, showing increased TSS-associated H3K27me3 in *MLL1*-null cells. Right panels: heatmaps show 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 the 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) (Shen et al., 2014). See also Figure S4.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2016.06.018.

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 genomic data analyses. R.R. and A.S. co-wrote the paper. All authors edited the manuscript.

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 A.S. laboratory were supported in part by a grant from the NIH (R35CA197569). A.P. is an EMBO postdoctoral fellow (ALTF 372-2015) and his work in the A.S. lab is supported by AIRC and Marie Curie Actions – People – COFUND.

Received: January 24, 2016
Revised: April 7, 2016
Accepted: June 10, 2016
Published: July 21, 2016

REFERENCES

Arnold, C.D., Gerlach, D., Stelzer, C., Boryń, L.M., Rath, M., and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science *339*, 1074–1077.

Bauer, M., Trupke, J., and Ringrose, L. (2016). The quest for mammalian Polycomb response elements: are we there yet? Chromosoma *125*, 471–496.

Beisel, C., Buness, A., Roustan-Espinosa, I.M., Koch, B., Schmitt, S., Haas, S.A., Hild, M., Katsuyama, T., and Paro, R. (2007). Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. Proc. Natl. Acad. Sci. USA *104*, 16615–16620.

Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H., and Helin, K. (2006). Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. *20*, 1123–1136.

Capotosti, F., Hsieh, J.J.D., and Herr, W. (2007). Species selectivity of mixed-lineage leukemia/trithorax and HCF proteolytic maturation pathways. Mol. Cell. Biol. *27*, 7063–7072.

Caslini, C., Connelly, J.A., Serna, A., Broccoli, D., and Hess, J.L. (2009). MLL associates with telomeres and regulates telomeric repeat-containing RNA transcription. Mol. Cell. Biol. *29*, 4519–4526.

Cavalli, G., and Paro, R. (1999). Epigenetic inheritance of active chromatin after removal of the main transactivator. Science *286*, 955–958.

Chan, C.S., Rastelli, L., and Pirrotta, V. (1994). A Polycomb response element in the Ubx gene that determines an epigenetically inherited state of repression. EMBO J. *13*, 2553–2564.

Chinwalla, V., Jane, E.P., and Harte, P.J. (1995). The Drosophila trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. EMBO J. *14*, 2056–2065.

Cooper, S., Dienstbier, M., Hassan, R., Schermelleh, L., Sharif, J., Blackledge, N.P., De Marco, V., Elderkin, S., Koseki, H., Klose, R., et al. (2014). Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Rep. *7*, 1456–1470.

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., and Paro, R. (2011). Polycomb preferentially targets stalled promoters of coding and non-coding transcripts. Genome Res. *21*, 216–226.

Francis, N.J., and Kingston, R.E. (2001). Mechanisms of transcriptional memory. Nat. Rev. Mol. Cell Biol. *2*, 409–421.

Gohl, D., Müller, M., Pirrotta, V., Affolter, M., and Schedl, P. (2008). Enhancer blocking and transvection at the Drosophila apterous locus. Genetics *178*, 127–143.

Hagstrom, K., Muller, M., and Schedl, P. (1996). Fab-7 functions as a chromatin domain boundary to ensure proper segment specification by the Drosophila bithorax complex. Genes Dev. *10*, 3202–3215.

Herz, H.M., Mohan, M., Garruss, A.S., Liang, K., Takahashi, Y.H., Mickey, K., Voets, O., Verrijzer, C.P., and Shilatifard, A. (2012). Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev. *26*, 2604–2620.

Hsieh, J.J.D., Ernst, P., Erdjument-Bromage, H., Tempst, P., and Korsmeyer, S.J. (2003). Proteolytic cleavage of MLL generates a complex of N- and C-terminal fragments that confers protein stability and subnuclear localization. Mol. Cell. Biol. *23*, 186–194.

Hu, D., Gao, X., Morgan, M.A., Herz, H.M., Smith, E.R., and Shilatifard, A. (2013a). The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol. Cell. Biol. *33*, 4745–4754.

Hu, D., Garruss, A.S., Gao, X., Morgan, M.A., Cook, M., Smith, E.R., and Shilatifard, A. (2013b). The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat. Struct. Mol. Biol. *20*, 1093–1097.

Illingworth, R.S., Gruenewald-Schneider, U., Webb, S., Kerr, A.R., James, K.D., Turner, D.J., Smith, C., Harrison, D.J., Andrews, R., and Bird, A.P. (2010). Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet. *6*, e1001134.

Kennison, J.A., and Tamkun, J.W. (1988). Dosage-dependent modifiers of polycomb and antennapedia mutations in Drosophila. Proc. Natl. Acad. Sci. USA *85*, 8136–8140.

Klose, R.J., Cooper, S., Farcas, A.M., Blackledge, N.P., and Brockdorff, N. (2013). Chromatin sampling–an emerging perspective on targeting polycomb repressor proteins. PLoS Genet. *9*, e1003717.

Klymenko, T., and Müller, J. (2004). The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep. *5*, 373–377.

Kvon, E.Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J.O., Pagani, M., Schernhuber, K., Dickson, B.J., and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. Nature *512*, 91–95.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. Nat. Protoc. *1*, 729–748.

Levine, S.S., King, I.F., and Kingston, R.E. (2004). Division of labor in polycomb group repression. Trends Biochem. Sci. *29*, 478–485.

Li, H.B., Müller, M., Bahechar, I.A., Kyrchanova, O., Ohno, K., Georgiev, P., and Pirrotta, V. (2011). Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in Drosophila melanogaster. Mol. Cell. Biol. *31*, 616–625.

Li, H.B., Ohno, K., Gui, H., and Pirrotta, V. (2013). Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. PLoS Genet. *9*, e1003436.

Margueron, R., and Reinberg, D. (2011). The Polycomb complex PRC2 and its mark in life. Nature *469*, 343–349.

Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K., and Stamatoyannopoulos, J.A. (2015). Role of DNA Methylation in Modulating Transcription Factor Occupancy. Cell Rep. *12*, 1184–1195.

McCabe, M.T., Ott, H.M., Ganji, G., Korenchuk, S., Thompson, C., Van Aller, G.S., Liu, Y., Graves, A.P., Della Pietra, A., 3rd, Diaz, E., et al. (2012). EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature *492*, 108–112.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Mendenhall, E.M., Koche, R.P., Truong, T., Zhou, V.W., Issac, B., Chi, A.S., Ku, M., and Bernstein, B.E. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet. *6*, e1001244.

Milne, T.A., Dou, Y., Martin, M.E., Brock, H.W., Roeder, R.G., and Hess, J.L. (2005). MLL associates specifically with a subset of transcriptionally active target genes. Proc. Natl. Acad. Sci. USA *102*, 14765–14770.

Mohan, M., Herz, H.M., Smith, E.R., Zhang, Y., Jackson, J., Washburn, M.P., Florens, L., Eissenberg, J.C., and Shilatifard, A. (2011). The COMPASS family of H3K4 methylases in Drosophila. Mol. Cell. Biol. *31*, 4310–4318.

Mohan, M., Herz, H.-M., and Shilatifard, A. (2012). SnapShot: Histone lysine methylase complexes. Cell *149*, 498–498.e1.

Piunti, A., and Shilatifard, A. (2016). Epigenetic balance of gene expression by Polycomb and COMPASS families. Science *352*, aad9780.

Poux, S., Horard, B., Sigrist, C.J., and Pirrotta, V. (2002). The Drosophila trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. Development *129*, 2483–2493.

Riising, E.M., Comet, I., Leblanc, B., Wu, X., Johansen, J.V., and Helin, K. (2014). Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol. Cell *55*, 347–360.

Ringrose, L., Rehmsmeier, M., Dura, J.-M., and Paro, R. (2003). Genome-wide prediction of Polycomb/Trithorax response elements in Drosophila melanogaster. Dev. Cell *5*, 759–771.

Schuettengruber, B., Martinez, A.M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. Nat. Rev. Mol. Cell Biol. *12*, 799–814.

Schwartz, Y.B., Kahn, T.G., Stenberg, P., Ohno, K., Bourgon, R., and Pirrotta, V. (2010). Alternative epigenetic chromatin states of polycomb target genes. PLoS Genet. *6*, e1000805.

Shen, L., Shao, N., Liu, X., and Nestler, E. (2014). ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics *15*, 284.

Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu. Rev. Biochem. *81*, 65–95.

Simon, J.A., and Kingston, R.E. (2009). Mechanisms of polycomb gene silencing: knowns and unknowns. Nat. Rev. Mol. Cell Biol. *10*, 697–708.

Simon, J.A., and Kingston, R.E. (2013). Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Mol. Cell *49*, 808–824.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila bithorax complex that mediate repression by Polycomb group products. Dev. Biol. *158*, 131–144.

Takeda, S., Chen, D.Y., Westergard, T.D., Fisher, J.K., Rubens, J.A., Sasagawa, S., Kan, J.T., Korsmeyer, S.J., Cheng, E.H., and Hsieh, J.J. (2006). Proteolysis of MLL family proteins is essential for taspase1-orchestrated cell cycle progression. Genes Dev. *20*, 2397–2409.

Tanay, A., O'Donnell, A.H., Damelin, M., and Bestor, T.H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. Proc. Natl. Acad. Sci. USA *104*, 5521–5526.

Wang, P., Lin, C., Smith, E.R., Guo, H., Sanderson, B.W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L.M., et al. (2009). 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. Mol. Cell. Biol. *29*, 6074–6085.

Wu, M., Wang, P.F., Lee, J.S., Martin-Brown, S., Florens, L., Washburn, M., and Shilatifard, A. (2008). Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol. Cell. Biol. *28*, 7337–7344.