

NORTHWESTERN UNIVERSITY

**Catalytic and Non-Catalytic Functions of COMPASS Family
H3K4 Methyltransferases in Flies and Humans**

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Driskill Graduate Training Program in Life Sciences

By

RYAN ADAM RICKELS

EVANSTON, ILLINOIS

September 2019

Acknowledgments

I would first like to thank my mentor, Dr. Ali Shilatifard, for giving me every opportunity to succeed as a graduate student. I can truly say that I got everything I wanted out of graduate school, and I don't think that could have been possible if not for Ali. He taught me the necessity of hard work in science and to be a fearless experimenter, as so many of our endeavors do not go as we initially planned. I would also like to thank Dr. Scott Hawley for taking a chance on me as a young pupil, and imparting to me his many years of experience and wisdom in just one semester. I am grateful to my thesis committee members: Drs. Jason Brickner, Panagiotis Ntziachristos, Daniel Foltz, and Dale Dorsett for their guidance and support. I want to thank all Shilatifard lab members, especially Drs. Edwin Smith, Marc Morgan, and Andrea Piunti, who were always willing to talk science and entertain my many questions over the years. I would like to express gratitude to my two best childhood friends, Doug Griffey and Kirby Reitz, who have always cheered me on. I want to thank my two loving parents for buying me the chemistry set that sparked my interest in studying the natural world. They have always encouraged my interest in art and science, and gave me the freedom to explore and pursue my dreams. Lastly, I thank my beautiful wife, Amy, for her endless love and support from the very beginning. I could not have made it without you.

To My Daughter, Zelda.

TABLE OF CONTENTS

Acknowledgments	2
Dedication	3
Table of Contents	4
Introduction	6
I. References	16
 Chapter 1: An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS	
I. Abstract	21
II. Introduction	22
III. Results	23
IV. Discussion	43
V. Methods	45
VI. References	51
 Chapter 2: Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability	
I. Abstract	56
II. Introduction	57
III. Results	58

IV. Discussion	78
V. Methods	81
VI. References	86
Chapter 3: Identifying the Minimal Requirements for UTX Stability in Drosophila and Humans	90
I. Abstract	90
II. Introduction	91
III. Results	92
IV. Discussion	109
V. Methods	111
VI. References	113
Chapter 4: Concluding Remarks	115
I. References	121

Abstract

Methylation of histone 3 lysine 4 (H3K4) catalyzed by the COMPASS family of lysine methyltransferases is universally associated with eukaryotic transcription. However, despite thousands of published studies examining the deposition, dynamics, and genomic positions of this chromatin modification, there is no clear consensus as to the molecular function or biological significance of H3K4-methylation. Furthermore, deletion of any COMPASS family member causes severe developmental abnormalities in metazoans, whereas catalytic-inactivating mutations are typically non-lethal and produce remarkably milder phenotypes. While only a limited number of studies have attempted to deconvolute catalytic-dependent from catalytic-independent functions for COMPASS, the findings are consistent thus far with a growing body of work highlighting a functional disconnect between histone modifications and the enzymes that catalyze them. My thesis work focuses on using *Drosophila* as a tool to disentangle enzymatic from non-enzymatic functions for COMPASS members Trithorax (Trx) and Trithorax-related (Trr), and applying my findings in a mammalian system to test for conservation of function as well as medical relevance.

I performed the first genome-wide assessment of Trx-dependent H3K4-methylation and found H3K4 dimethylation (H3K4me₂) deposited by Trx is highly predictive of Polycomb Response Elements (PREs) in *Drosophila*. Although I was unsuccessful in determining whether or not Trx-dependent H3K4me₂ is necessary for PRE function, I demonstrate a conserved role for the mammalian homolog, MLL1, in catalyzing H3K4me₂ at CpG-rich sequences that functionally resemble PREs in the human genome. This study establishes that, in a given cell type, several hundred developmentally regulated genes require MLL1, not for transcriptional activation, but to maintain activation by blocking silencing by Polycomb Repressive Complex 2 (PRC2). These

results identify a subset of genes whose expression levels are balanced by MLL1 and PRC2, and challenges a passive model of PRC2 recruitment to transcriptionally silent promoters.

Lastly, my thesis work culminates with the discovery that enhancer-associated H3K4-monomethylation (H3K4me1) is not essential for *Drosophila* development, and is generally dispensable for enhancer function in mammalian cells as well. While *trr* deletion is recessive lethal, flies harboring a Trr catalytic-inactivating mutation develop to productive adulthood and only display mild wing-vein phenotypes when reared at higher temperatures, suggesting enhancer-associated H3K4me1 might play a role in buffering enhancer-promoter communication under environmental stress. Consistent with my findings in *Drosophila*, similar experiments in mouse embryonic stem cells demonstrate thousands of gene expression changes in the absence of Trr mammalian homologs, MLL3/MLL4; however, catalytic-inactivating mutations result in relatively few expression changes, suggesting a non-enzymatic function for Trr/MLL3/MLL4 is essential for their role in facilitating enhancer-mediated gene activation during development. Genetic complementation assays in *Drosophila* identified a Trr fragment of unknown function was able to rescue Trr-null lethality, and this fragment was shown to bind and stabilize Utx, an H3K27 demethylase known to promote enhancer activation. The Utx-binding domain is conserved in mammalian MLL3/4 and expression of an 80 residue 'minimal' peptide was demonstrated to bind UTX and block its degradation in the absence of full-length MLL3/4. *Utx* and *Kdm6a* are essential genes in flies and mice, respectively, suggesting that stabilizing this important chromatin modifier represents at least one important non-enzymatic function of Trr/MLL3/MLL4 that is critical for life. Taken together, this thesis on the catalytic and non-catalytic functions of COMPASS provides a more nuanced view of epigenetic regulation and, by testing assumptions, serves as a guide from which to contemplate the role of chromatin modifiers in the process of transcription.

Introduction

Examples of Epigenetics

Epigenetics is defined as “a stably heritable phenotype resulting from changes in a chromosome without alterations in the DNA sequence”¹. Heritable changes in chromosome structure can include DNA methylation, chemical modification of histones and other chromatin proteins, nucleosome positioning, non-coding RNA recruitment, sub-nuclear positioning, and three-dimensional changes in chromosome topology. One well-studied example is the process of centromere formation in higher eukaryotes, in which the centromere-specific CENPA histone variant is deposited independent of the underlying DNA sequence^{2,3}. This is demonstrated to be a self-propagating, epigenetic process as ectopic recruitment of the chromatin assembly factor, HJURP, to a non-centromeric loci is sufficient to deposit CENPA and establish a functional kinetochore *de novo*⁴. Additional experiments reveal how even transient CENPA recruitment can nucleate ectopic centromeres that are stably transmitted across several cell generations⁵.

Another famous example is the case of X chromosome inactivation, in which mammalian female cells achieve gene dosage compensation by randomly silencing one of two X chromosomes. While the mechanisms controlling this process are still under investigation, it is clear that a noncoding RNA, termed Xist, plays a critical role in both establishing and maintaining transcriptional repression of the inactive X chromosome (Xi)⁶⁻⁸. In the early blastocyst, both X chromosomes are initially active, and both transcribe low levels of Xist and Tsix, a noncoding RNA antagonist of Xist^{9,10}. Once one X chromosome is randomly selected for silencing, it will produce high levels of Xist, coating the Xi from which it was transcribed, while the non-silenced active X (Xa) continues to produce Tsix. Genetic and lineage tracing experiments using X-linked alleles to track both homologs demonstrate that all descendent cells will continue to silence the same X chromosome. Xist coating of the Xi leads to its compaction into a heterochromatic

structure, presumably through recruitment of various chromatin-modifying enzymes known to promote transcriptional silencing¹¹. One such enzyme, EZH2, associates with the Polycomb Repressive Complex 2 (PRC2) to catalyze trimethylation of Histone 3 Lysine 27 (H3K27me3) across the Xi¹². This post-translational modification is believed to function by recruiting the Polycomb Repressive Complex 1 (PRC1), which condenses local chromatin into a state that is refractory to transcription^{13,14}.

Histone Modification

Dozens of histone modifications have now been reported, as well as various enzymes that regulate their deposition and removal^{15,16}. Still, the molecular functions and/or biological significance of many of these modifications remain unclear¹⁷. Precise functional analysis is made even more challenging for some chromatin modifiers who function in more than one biological process, or possess catalytic-independent activities. Next generation sequencing technologies have allowed researchers to assemble genome-wide maps for various histone modifications in attempts to better understand their function by correlating particular genomic features (e.g. gene bodies, *cis*-regulatory elements, repetitive DNA sequences, etc.) with the accumulation of different histone modifications^{18,19}. For instance, H3K27me3 strongly correlates with transcriptionally repressed facultative chromatin²⁰, which is consistent with earlier work in *Drosophila* that showed mutations to *polycomb* can lead to de-repression of homeotic genes that cause segmental defects in the developing fly²¹.

COMPASS and H3K4 Methylation

In contrast, histone 3 lysine 4 trimethylation (H3K4me3) is an evolutionarily conserved feature of transcriptionally active chromatin, from single-cell eukaryotes to metazoans²². This modification was initially discovered in *Tetrahymena thermophila* and predicted to play a role in gene

activation by virtue of its abundance in the transcriptionally active macronuclei²³.

Furthermore, depletion of H3K4me3 in fission yeast was shown to correlate with reduced levels of H3 acetylation, a modification also believed to promote transcription²⁴. Set1 was soon identified as the sole H3K4-methylase in *Saccharomyces cerevisiae*, which functions within a large multimeric protein complex named COMPASS (Complex of Proteins Associated with Set1)^{25,26}. *set1*-deleted yeast (*set1*Δ) are viable, suggesting H3K4-methylation is not strictly required for gene transcription. *set1*Δ yeast do, however, display reduced growth rates compared with WT strains²⁷, as well as small defects in cell-cycle checkpoint inhibition²⁸. While the promoters of nearly all RNA Polymerase II (Pol II) -transcribed genes are normally enriched for H3K4me3 across species²⁹⁻³², only a small number of genes are quantitatively reduced in their expression when *set1* is deleted in yeast, suggesting H3K4me3 might only be important for gene activation in certain contexts^{25,33}. This is also consistent with recent reports that Set1-mediated H3K4-methylation in yeast is essential for propagating an epigenetic transcriptional “memory” of previous exposures to inositol³⁴. In this experiment, prior exposure to inositol was shown to potentiate the *INO1* gene for future re-activation, concomitant with increased H3K4me2 in the gene; however, *set1*Δ cells lacking H3K4-methylation do not show this heightened response to repeated inositol treatment. It will be interesting to see which, if any, other inducible genes display this kind of H3K4-dependent “memory”, and to understand how this mechanism evolved to help cells adapt to changes in their environment.

While many questions remain regarding the true biological function of yeast Set1, the regulatory mechanisms governing its catalytic activity are less enigmatic. Set1 is not enzymatically active on its own, but requires four core subunits: Wdr5/Cps30, Rbbp5/Cps50, Ash2L/Cps60, Dpy30/Cps25 (referred to as WRAD) to catalyze H3K4me3 *in vitro*^{26,35}, and at least two additional proteins: CXXC1/Cps40 and Wdr82/Cps35 to deposit correct patterns of H3K4 mono-

, di-, and trimethylation *in vivo*³⁶⁻³⁸. Among these, Wdr5 and Rbbp5 are both required for assembly and stability of the COMPASS complex. With the exception of Wdr82/Cps35, all COMPASS components are encoded by non-essential genes in *S. cerevisiae*²². While the cell growth-essentiality of Wdr82/Cps35(Swd2) is due to its additional role in 3' RNA processing, this COMPASS subunit is also critical for stimulating correct H3K4me2/3 levels in yeast³⁹.

The genomic recruitment and catalytic activity of Set1/COMPASS is also dependent on additional protein complexes, and even other histone modifications. A combination of genetic and biochemical screening identified the Paf1 (Polymerase Associated Factor) complex to be responsible for bringing Set1 to chromatin-bound RNA polymerase II⁴⁰, along with Rad6, an E2 ubiquitin-conjugating enzyme, and Bre1, an E3 ubiquitin ligase⁴¹. Together, Rad6/Bre1 catalyze monoubiquitination of H2B lysine 123, and this modification is required for deposition of H3K4-methylation by Set1⁴², although the molecular details underpinning this form of “histone crosstalk” are still murky⁴³. Interestingly, H2B-monoubiquitination (H2Bub) is also essential for catalysis of H3K79me2/3 by Dot1, an additional chromatin modifier recruited by Paf1 to transcriptionally active chromatin^{44,45}. Both H3K79-methylation and H2B-monoubiquitination are enriched within gene bodies and are strongly associated with the elongating form of RNA Polymerase II⁴⁶. Set1 may travel with elongating Pol II, as evidenced by strong H3K4me3 levels at promoters followed by a trail of H3K4me2 and H3K4me1 further into the gene body⁴⁷. Thus, Set1 may also function co-transcriptionally to regulate the process of elongation; however, considering that Set1 is non-essential in yeast, this function may be redundant for the majority of active genes under laboratory conditions. The perplexing relationship between yeast Set1 and transcription elongation has clinical implications for a subset of aggressive childhood leukemia. Through spontaneous chromosomal translocation, a human homolog of Set1, called MLL (Mixed Lineage Leukemia), was found to form chimeric fusion proteins in certain blood

cancers. Work from our lab identified multiple MLL-fusion partners as being components of a separate transcription elongation complex, termed the Super Elongation Complex, which promotes the release of paused Pol II into production elongation⁴⁸⁻⁵⁰. These studies strongly suggest that polymerase pause-release represents a crucial point of genetic regulation in development and disease pathogenesis.

The large number of identified histone modifications, as well as their correlations with genomic features and transcriptional states, lead to popularization of the “histone code” hypothesis, which postulates that sequential histone modifications constitute a combinatorial code to bring about downstream gene expression changes¹⁶. While addition or subtraction of a few histone modifications are demonstrated to be causal in their effects on transcription⁵¹⁻⁵³, these examples seem to be the exception rather than the rule¹⁷. Instead, a growing body of research suggests the role of several histone modifications in regulating transcription may only be secondary to less-understood catalytic-independent functions of the enzymes themselves⁵⁴⁻⁶⁰. In some cases, the catalytic and non-catalytic functions are purely context-dependent, and I will discuss these examples later. In the case of yeast Set1, the enzymatic versus non-enzymatic functions are more difficult to separate. One interesting report demonstrates Set1’s ability to “sense” the presence of H3K4-methylation through its SET domain, which creates a feedback to stabilize the enzyme⁶¹. In this study, an epitope-tagged Set1 protein containing a catalytic-inactivating point mutation was found to be unstable when expressed in the Δ Set1 background, but stable in WT cells. Expressing a catalytic-active Set1 in Δ Set1 cells restored H3K4me3 levels, but also stabilized the catalytic-inactive Set1, as shown by western blot. Furthermore, deletion of Paf1, mutation of H3K4, or chemical inhibition of transcription all elicited the same destabilizing effect on Set1. These results indicate that Set1 stability is highly regulated by its catalytic product; therefore, any potential non-enzymatic function of Set1 will depend indirectly on its enzymatic

activity. It is unclear how these feedback mechanisms might operate in higher eukaryotes, where additional COMPASS family enzymes implement H3K4-methylation at distinct genomic loci.

Expansion of COMPASS in Multicellular Animals

Through a series of genome duplications, the COMPASS family has increased in number and diversified in function during metazoan evolution. Whereas Set1 represents the sole H3K4-methylase in unicellular organisms, three homologous enzymes exist in *Drosophila*: Set1, Trithorax (Trx), and Trithorax-related (Trr). Each of the three Set1 homologous assemble into distinct COMPASS-like complexes, each with complex-specific subunits that regulate the complex's function and stability in unique ways, in addition to the 4 core WRAD subunits required for methyltransferase activity *in vitro* (Figure 1A). In contrast to COMPASS in *S. Cerevisiae*, deletion of any COMPASS family member or accessory subunit is recessive lethal in *Drosophila* (phenotypes described at Flybase.org). Many of these mutants die during embryogenesis, making it difficult to precisely determine which biological processes are disrupted to produce that phenotype. Each of the three COMPASS-like complexes in *Drosophila* are represented by two paralogs in mammals (Figure 1A) and genetic knock-out of any one of these six COMPASS members is lethal to mice [reviewed by Piunti and Shilatifard, 2016].

While our molecular understanding is incomplete as to how each COMPASS member functions during development, our studies in *Drosophila* have helped to provide a clearer idea as to how deposition of H3K4-methylation is regulated by different COMPASS enzymes, genome-wide (Figure 1B). For instance, we know that Set1 is responsible for catalyzing bulk H3K4me3 levels, and this modification co-localizes almost exclusively with transcriptionally active promoters in *Drosophila*^{19,62,63}. Although Set1-deleted flies are homozygous-lethal, Set1 catalytic-activity has

never been formally demonstrated to be required for viability. Another 'branch' of the COMPASS family in *Drosophila* is represented by Trx, the homolog of mammalian MLL1 and MLL2. Mutations of the *trx* gene were first described as early as 1930, following isolation of a four-winged fly resulting from a transformation of the metathoracic segment into an additional mesothoracic segment⁶⁴. CHIP-sequencing experiments identified a role for Trx in catalyzing H3K4me2 at Polycomb response elements (PREs) in a process that regulates expression of segmental-identity genes during development⁶⁵. The background and details of this study will be further described in Chapter 1 of this dissertation, while Chapters 2 and 3 will primarily focus on the catalytic-dependent and –independent functions of Trr, the third 'branch' of COMPASS. We now know bulk H3K4me1 levels are catalyzed by Trr in flies and MLL3/4 in mammals, and this modification is predominantly deposited at enhancer elements^{66,67}. Many questions persist regarding the biological role of H3K4me1 in regulating enhancer function, and these quandaries will be addressed in Chapter 2.

Figure 1.

(A) Conservation of the COMPASS family from yeast to flies to humans. Methyltransferase enzymes shown in red. WRAD components colored green. Complex-specific subunits shown in turquoise. (B) Diagram representing the division of labor among COMPASS family members in catalyzing various degrees of H3K4-methylation at different kinds of *cis*-regulatory elements.

References

- 1 Berger, S. L., Kouzarides, T., Shiekhatar, R. & Shilatifard, A. An operational definition of epigenetics. *Genes Dev* **23**, 781-783, doi:10.1101/gad.1787609 (2009).
- 2 Allshire, R. C. & Karpen, G. H. Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nat Rev Genet* **9**, 923-937, doi:10.1038/nrg2466 (2008).
- 3 Cleveland, D. W., Mao, Y. & Sullivan, K. F. Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* **112**, 407-421 (2003).
- 4 Barnhart, M. C. *et al.* HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J Cell Biol* **194**, 229-243, doi:10.1083/jcb.201012017 (2011).
- 5 Mendiburo, M. J., Padeken, J., Fulop, S., Schepers, A. & Heun, P. *Drosophila* CENH3 is sufficient for centromere formation. *Science* **334**, 686-690, doi:10.1126/science.1206880 (2011).
- 6 Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X chromosome. *Nature* **349**, 38-44, doi:10.1038/349038a0 (1991).
- 7 Brown, C. J. *et al.* Localization of the X inactivation centre on the human X chromosome in Xq13. *Nature* **349**, 82-84, doi:10.1038/349082a0 (1991).
- 8 Brown, S. D. XIST and the mapping of the X chromosome inactivation centre. *Bioessays* **13**, 607-612, doi:10.1002/bies.950131112 (1991).
- 9 Lee, J. T., Davidow, L. S. & Warshawsky, D. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* **21**, 400-404, doi:10.1038/7734 (1999).
- 10 Lee, J. T. & Lu, N. Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* **99**, 47-57 (1999).
- 11 Augui, S., Nora, E. P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet* **12**, 429-442, doi:10.1038/nrg2987 (2011).
- 12 Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750-756, doi:10.1126/science.1163045 (2008).
- 13 Gao, Z. *et al.* PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol Cell* **45**, 344-356, doi:10.1016/j.molcel.2012.01.002 (2012).
- 14 Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**, aad9780, doi:10.1126/science.aad9780 (2016).

- 15 Black, J. C., Van Rechem, C. & Whetstine, J. R. Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol Cell* **48**, 491-507, doi:10.1016/j.molcel.2012.11.006 (2012).
- 16 Strahl, B. D. & Allis, C. D. The language of covalent histone modifications. *Nature* **403**, 41-45, doi:10.1038/47412 (2000).
- 17 Henikoff, S. & Shilatifard, A. Histone modification: cause or cog? *Trends Genet* **27**, 389-396, doi:10.1016/j.tig.2011.06.006 (2011).
- 18 Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nat Biotechnol* **28**, 817-825, doi:10.1038/nbt.1662 (2010).
- 19 Kharchenko, P. V. *et al.* Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**, 480-485, doi:10.1038/nature09725 (2011).
- 20 Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708, doi:10.1038/nrm2763 (2009).
- 21 Kennison, J. A. & Tamkun, J. W. Dosage-dependent modifiers of polycomb and antennapedia mutations in *Drosophila*. *Proceedings of the National Academy of Sciences* **85**, 8136-8140, doi:10.1073/pnas.85.21.8136 (1988).
- 22 Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81**, 65-95, doi:10.1146/annurev-biochem-051710-134100 (2012).
- 23 Strahl, B. D., Ohba, R., Cook, R. G. & Allis, C. D. Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc Natl Acad Sci U S A* **96**, 14967-14972, doi:10.1073/pnas.96.26.14967 (1999).
- 24 Noma, K. & Grewal, S. I. Histone H3 lysine 4 methylation is mediated by Set1 and promotes maintenance of active chromatin states in fission yeast. *Proc Natl Acad Sci U S A* **99 Suppl 4**, 16438-16445, doi:10.1073/pnas.182436399 (2002).
- 25 Miller, T. *et al.* COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A* **98**, 12902-12907, doi:10.1073/pnas.231473398 (2001).
- 26 Roguev, A. *et al.* The *Saccharomyces cerevisiae* Set1 complex includes an Ash2 homologue and methylates histone 3 lysine 4. *EMBO J* **20**, 7137-7148, doi:10.1093/emboj/20.24.7137 (2001).
- 27 Briggs, S. D. *et al.* Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in *Saccharomyces cerevisiae*. *Genes Dev* **15**, 3286-3295, doi:10.1101/gad.940201 (2001).
- 28 Corda, Y. *et al.* Interaction between Set1p and checkpoint protein Mec3p in DNA repair and telomere functions. *Nat Genet* **21**, 204-208, doi:10.1038/5991 (1999).
- 29 Bernstein, B. E. *et al.* Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc Natl Acad Sci U S A* **99**, 8695-8700, doi:10.1073/pnas.082249499 (2002).
- 30 Eisenberg, J. C. & Shilatifard, A. Histone H3 lysine 4 (H3K4) methylation in development and differentiation. *Dev Biol* **339**, 240-249, doi:10.1016/j.ydbio.2009.08.017 (2010).

- 31 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics* **39**, 311-318, doi:10.1038/ng1966 (2007).
- 32 Santos-Rosa, H. *et al.* Methylation of histone H3 K4 mediates association of the Isw1p ATPase with chromatin. *Mol Cell* **12**, 1325-1332 (2003).
- 33 Margaritis, T. *et al.* Two distinct repressive mechanisms for histone 3 lysine 4 methylation through promoting 3'-end antisense transcription. *PLoS Genet* **8**, e1002952, doi:10.1371/journal.pgen.1002952 (2012).
- 34 D'Urso, A. *et al.* Set1/COMPASS and Mediator are repurposed to promote epigenetic transcriptional memory. *Elife* **5**, doi:10.7554/eLife.16691 (2016).
- 35 Krogan, N. J. *et al.* COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric silencing of gene expression. *J Biol Chem* **277**, 10753-10755, doi:10.1074/jbc.C200023200 (2002).
- 36 Dehe, P. M. *et al.* Protein interactions within the Set1 complex and their roles in the regulation of histone 3 lysine 4 methylation. *J Biol Chem* **281**, 35404-35412, doi:10.1074/jbc.M603099200 (2006).
- 37 Schlichter, A. & Cairns, B. R. Histone trimethylation by Set1 is coordinated by the RRM, autoinhibitory, and catalytic domains. *EMBO J* **24**, 1222-1231, doi:10.1038/sj.emboj.7600607 (2005).
- 38 Schneider, J. *et al.* Molecular regulation of histone H3 trimethylation by COMPASS and the regulation of gene expression. *Mol Cell* **19**, 849-856, doi:10.1016/j.molcel.2005.07.024 (2005).
- 39 Cheng, H., He, X. & Moore, C. The essential WD repeat protein Swd2 has dual functions in RNA polymerase II transcription termination and lysine 4 methylation of histone H3. *Mol Cell Biol* **24**, 2932-2943, doi:10.1128/mcb.24.7.2932-2943.2004 (2004).
- 40 Krogan, N. J. *et al.* The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol Cell* **11**, 721-729 (2003).
- 41 Schneider, J., Dover, J., Johnston, M. & Shilatifard, A. Global proteomic analysis of *S. cerevisiae* (GPS) to identify proteins required for histone modifications. *Methods Enzymol* **377**, 227-234, doi:10.1016/S0076-6879(03)77013-X (2004).
- 42 Dover, J. *et al.* Methylation of histone H3 by COMPASS requires ubiquitination of histone H2B by Rad6. *J Biol Chem* **277**, 28368-28371, doi:10.1074/jbc.C200348200 (2002).
- 43 Lee, J. S., Smith, E. & Shilatifard, A. The language of histone crosstalk. *Cell* **142**, 682-685, doi:10.1016/j.cell.2010.08.011 (2010).
- 44 Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-793, doi:10.1038/ng.646 (2010).
- 45 Wood, A. *et al.* Bre1, an E3 ubiquitin ligase required for recruitment and substrate selection of Rad6 at a promoter. *Mol Cell* **11**, 267-274 (2003).
- 46 Schulze, J. M. *et al.* Linking cell cycle to histone modifications: SBF and H2B monoubiquitination machinery and cell-cycle regulation of H3K79 dimethylation. *Mol Cell* **35**, 626-641, doi:10.1016/j.molcel.2009.07.017 (2009).

- 47 Wood, A. *et al.* Ctk complex-mediated regulation of histone methylation by COMPASS. *Mol Cell Biol* **27**, 709-720, doi:10.1128/MCB.01627-06 (2007).
- 48 Lin, C. *et al.* AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of MLL chimeras, can link transcription elongation to leukemia. *Mol Cell* **37**, 429-437, doi:10.1016/j.molcel.2010.01.026 (2010).
- 49 Mohan, M., Lin, C., Guest, E. & Shilatifard, A. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. *Nat Rev Cancer* **10**, 721-728, doi:10.1038/nrc2915 (2010).
- 50 Smith, E., Lin, C. & Shilatifard, A. The super elongation complex (SEC) and MLL in development and disease. *Genes Dev* **25**, 661-672, doi:10.1101/gad.2015411 (2011).
- 51 Hilton, I. B. *et al.* Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* **33**, 510-517, doi:10.1038/nbt.3199 (2015).
- 52 Klann, T. S. *et al.* CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol* **35**, 561-568, doi:10.1038/nbt.3853 (2017).
- 53 Shogren-Knaak, M. *et al.* Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-847, doi:10.1126/science.1124000 (2006).
- 54 Copur, O. & Muller, J. Histone Demethylase Activity of Utx Is Essential for Viability and Regulation of HOX Gene Expression in Drosophila. *Genetics* **208**, 633-637, doi:10.1534/genetics.117.300421 (2018).
- 55 Dorighi K M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular Cell* (2017).
- 56 Hu, D. *et al.* Not All H3K4 Methylations Are Created Equal: Mll2/COMPASS Dependency in Primordial Germ Cell Specification. *Mol Cell* **65**, 460-475 e466, doi:10.1016/j.molcel.2017.01.013 (2017).
- 57 Illingworth, R. S. *et al.* The E3 ubiquitin ligase activity of RING1B is not essential for early mouse development. *Genes Dev* **29**, 1897-1902, doi:10.1101/gad.268151.115 (2015).
- 58 Rickels, R. *et al.* Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat Genet* **49**, 1647-1653, doi:10.1038/ng.3965 (2017).
- 59 Terranova, R., Agherbi, H., Boned, A., Meresse, S. & Djabali, M. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A* **103**, 6629-6634, doi:10.1073/pnas.0507425103 (2006).
- 60 Wang, C. *et al.* UTX regulates mesoderm differentiation of embryonic stem cells independent of H3K27 demethylase activity. *Proc Natl Acad Sci U S A* **109**, 15324-15329, doi:10.1073/pnas.1204166109 (2012).
- 61 Soares, L. M., Radman-Livaja, M., Lin, S. G., Rando, O. J. & Buratowski, S. Feedback control of Set1 protein levels is important for proper H3K4 methylation patterns. *Cell Rep* **6**, 961-972, doi:10.1016/j.celrep.2014.02.017 (2014).

- 62 Ardehali, M. B. *et al.* Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *EMBO J* **30**, 2817-2828, doi:10.1038/emboj.2011.194 (2011).
- 63 Hallson, G. *et al.* dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. *Genetics* **190**, 91-100, doi:10.1534/genetics.111.135863 (2012).
- 64 Ingham, P. W. trithorax and the regulation of homeotic gene expression in Drosophila: a historical perspective. *Int J Dev Biol* **42**, 423-429 (1998).
- 65 Rickels, R. *et al.* An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS. *Mol Cell* **63**, 318-328, doi:10.1016/j.molcel.2016.06.018 (2016).
- 66 Herz, H. M. *et al.* Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & Development* **26**, 2604-2620, doi:10.1101/gad.201327.112 (2012).
- 67 Hu, D. *et al.* The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and Cellular Biology* **33**, 4745-4754, doi:10.1128/mcb.01181-13 (2013).

CHAPTER 1:**An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements implemented by Trx/MLL/COMPASS¹****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.

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 chromatin². 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*^{2,3}. The COMPASS family members share several subunits in common, as well as unique factors believed to impart functional specificity^{4,5}. 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 cells⁴⁻⁷. Meanwhile, MLL2 implements H3K4me3 at bivalently marked promoters in mouse embryonic stem (ES) cells, and Trr/MLL3/MLL4 were recently shown to catalyze H3K4-monomethylation at *Drosophila* and mammalian enhancers, respectively^{4,8,9}.

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 states¹⁰. This maintenance function is known to rely on Trx binding to a nearby Polycomb Response Element (PRE)^{11,12}. While *Drosophila* PREs are well-documented, the existence of comparable elements in mammals has been debated¹³⁻¹⁶. One theory gaining support is that CpG Islands (CGIs) can function as PREs via recruitment of PRC1 and PRC2 complexes¹⁶⁻¹⁸. 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 PRE¹⁹. 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)⁹.

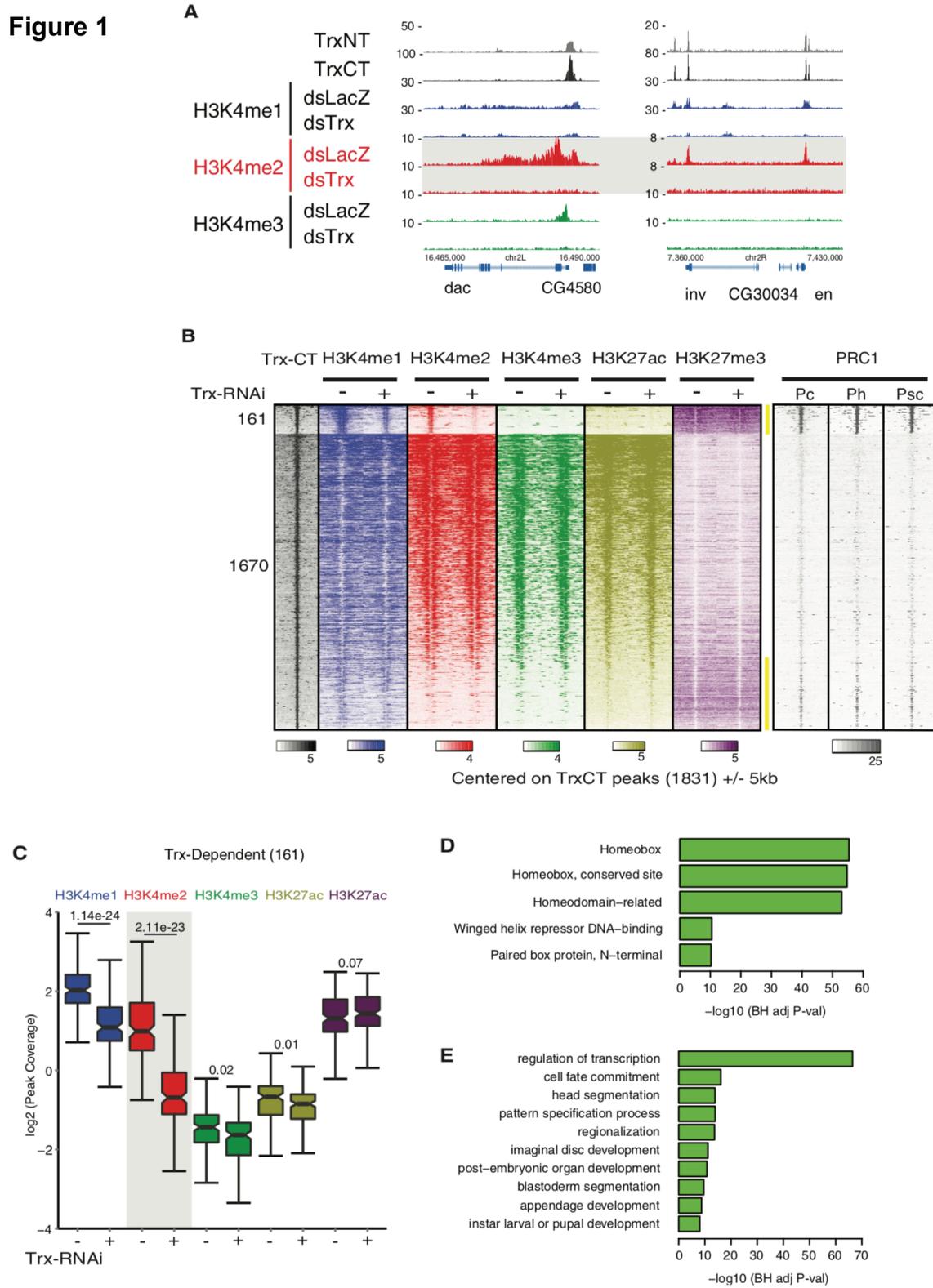
While all CGIs may have the potential to function as PREs in the absence of transcription, it is unclear what mechanisms prevent PRC1/2 from accessing these sites. Mouse and human genomes contain roughly 20,000-30,000 CGIs, indicating these elements alone are not sufficient to impart specificity to MLL/PcG target binding²⁰. 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

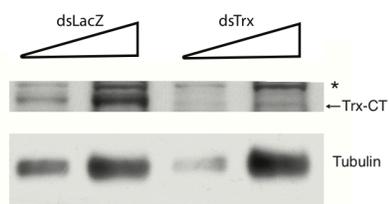
Trx/COMPASS-dependent H3K4me2 chromatin signature marks PREs in *Drosophila*-The COMPASS family in *Drosophila* (dCOMPASS) has three Set1-related H3K4-methylases: dSet1, Trx, and Trr, each of which associates with several proteins to form COMPASS-like complexes^{4,5,21}. Prompted by the findings of Herz 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^{22,23}, 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 1F). 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,C,H). Although Trr is shown to co-bind with Trx at these sites, H3K4me2 levels are not affected by Trr-depletion (Figure 1G). 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 repression²⁴. We found

88% of our Trx-dependent sites in these cells overlap with sites of Polycomb Repressive Complex 1 (PRC1) binding²⁵ (Figure 1B). This is consistent with several reports that Trx and PRC1 co-occupy PREs regardless of their transcriptional state^{14,25-28}. 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)^{25,26,29}. 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 modification²⁸.

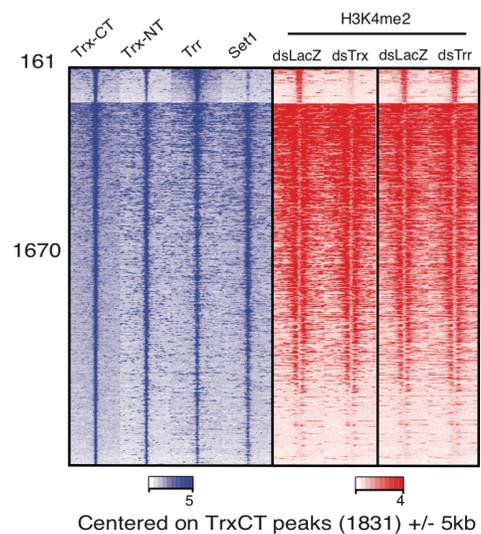
Figure 1



F



G



H

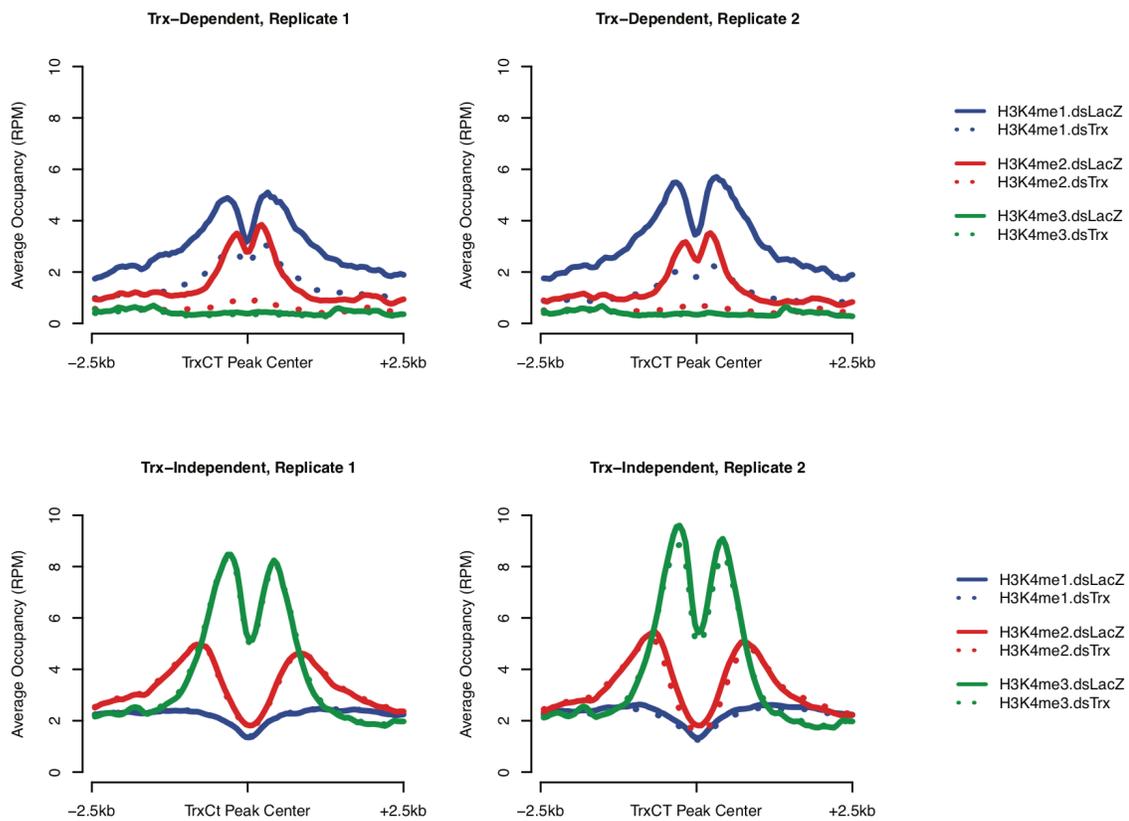


Figure 1.**A divalent, Trx-dependent H3K4me2 chromatin signature marks PREs in *Drosophila*.**

(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 ²⁵. (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. (F) Western blot showing reduction of Trx protein levels after 6-day RNAi. Asterisk indicates a non-specific band. (G) Occupancy profiles of Trx-CT, Trx-NT, Trr, dSet1. Occupancy levels of H3K4me2 +/- Trx-RNAi or Trr-RNAi show changes in H3K4me2 are Trr-independent. Profiles are centered on Trx-CT occupied peaks (+/- 5kb) and sorted in descending order of H3K4me2 occupancy in wild-type cells. (H) For two biological replicates, the average H3K4me1/2/3 occupancy levels +/- dsTrx are plotted with respect to H3K4me2 peak centers for the 161 Trx-dependent sites and for the 1670 Trx-independent sites where no significant change in H3K4me2 is observed.

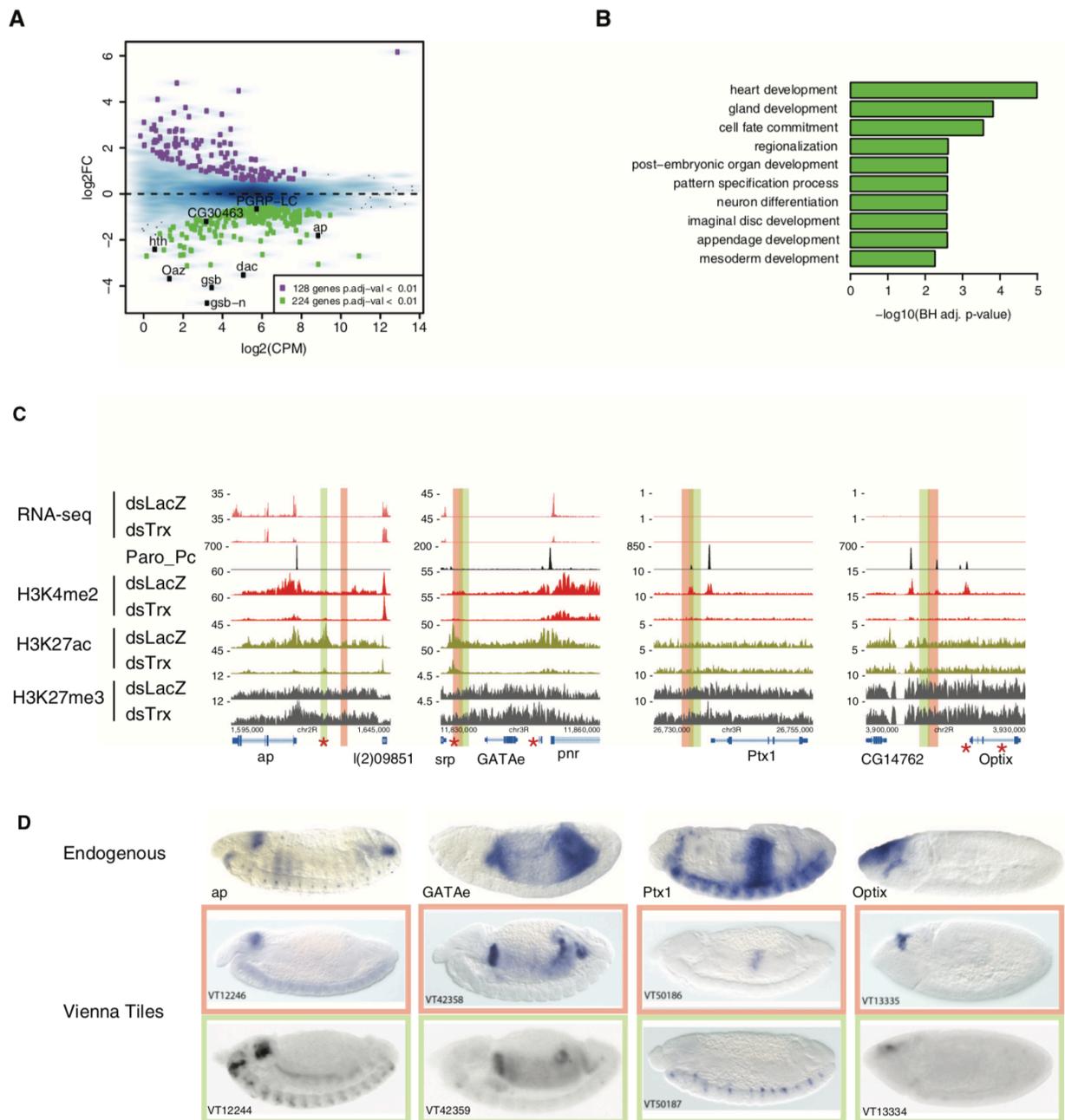
Trx-dependent H3K4-methylation at PREs and Enhancers-Our genome-wide analyses reveal that after Trx-depletion, PRE-associated genes with decreased H3K4me2 levels also exhibit reduced transcription (Figure 2A). This can be observed despite the fact that the majority of PRE-associated genes are already transcriptionally silent in S2 cells (Figure 2E). 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*^{30,31}. 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^{14,32}. To explore this further, we compared our 161 PREs with a published list of *Drosophila* enhancers identified by STARR-seq³³. 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^{10,34}. 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^{35,36}.

In the rapidly developing *Drosophila* embryo, Trx maintains active transcription of a target gene in a PRE-dependent manner^{10,37}. Although conclusive experimental evidence is still lacking, this function is hypothesized to depend on Trx lysine-methyltransferase activity¹². 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.

Figure 2



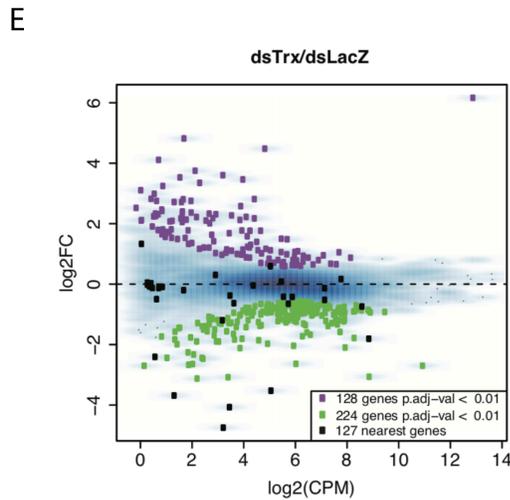


Figure 2.

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

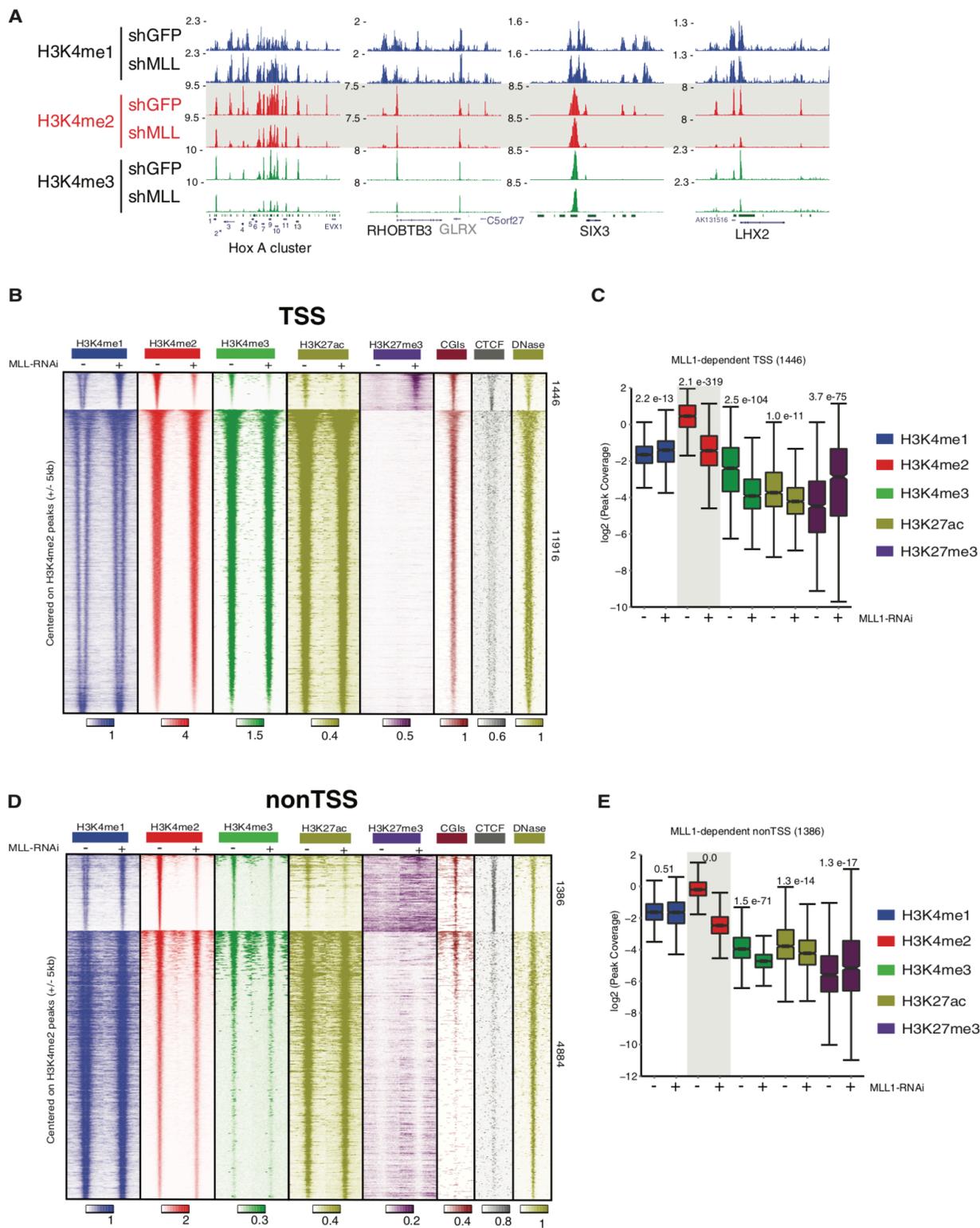
stages 13-14, except for *optix* (stage 8-9). Vienna Tile embryo images are from ³¹. Red asterisks indicate STARR-seq enhancers in S2 cells ³³. (E) MA plot showing differential gene expression after Trx-RNAi. Black boxes indicate 127 nearest genes to the 161 PREs (the majority of which are not expressed in S2 cells).

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) ²¹(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 ^{6,9}. 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 and 3F). MLL2 is not expressed in HCT116 cells, obviating concerns regarding any compensatory function of MLL2 (Figure 3G). Although bulk H3K4-methylation is unaffected in the absence of MLL1 (Figure 3H), 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 ⁶. 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 and 3I). 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).

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)³⁸. This result suggests that CTCF might play a role in specifying the MLL1-dependent subset of CGIs.

Figure 3



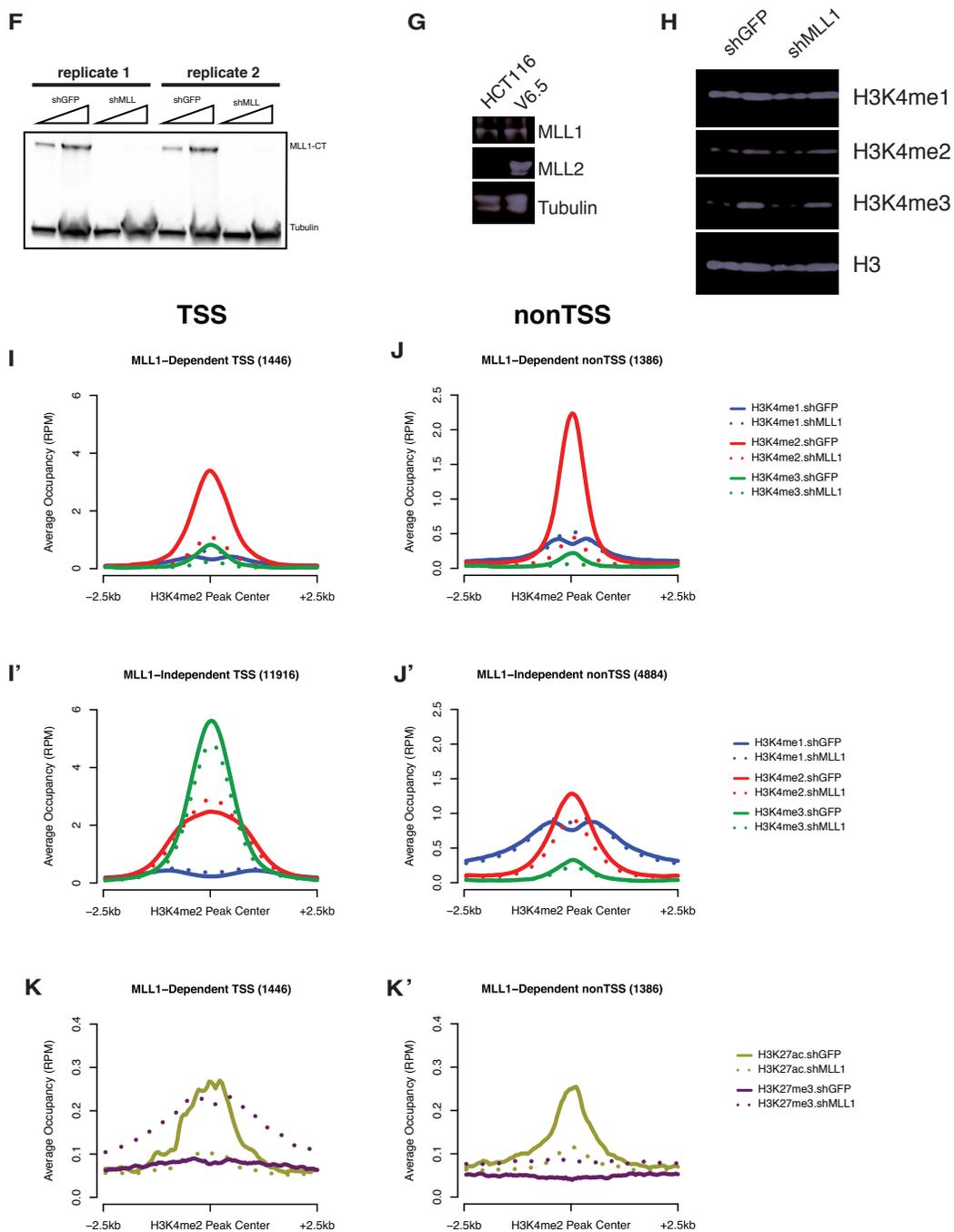


Figure 3.**MLL1-dependent H3K4me2 marks PRE-like sequences in the human genome.**

(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 +/- shMLL 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. (F) Western blot showing 2 replicates of shMLL1 efficacy, (G) lack of MLL2 protein expression in HCT116 cells, and (H) lack of global-decreases in H3K4me-methylation following MLL1-depletion. (I) Average H3K4me1/2/3 occupancy levels +/- shMLL are plotted with respect to H3K4me2 peak centers for 1446 MLL-dependent sites and (I') 11916 MLL-independent sites. Note the predominance of H3K4me2 at MLL-dependent TSS compared with MLL-independent TSS. (J) Average H3K4me1/2/3 occupancy levels +/- shMLL at nonTSS. Note, increased level of H3K4me2 compared to (J'). (K) Average H3K27ac/me3 levels +/- shMLL at 1446 MLL-dependent TSS and (K') MLL-dependent nonTSS sites.

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 environments³⁹⁻⁴¹. We performed RNA-seq to assess genome-wide transcriptional changes upon MLL1-depletion in HCT116 and identified 550 significantly down-regulated genes (Figure 4E). 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 4E). Of the 550 down-regulated genes, 8% (44) overlapped with nonTSS nearest genes, and 32% (179) overlapped with TSS nearest genes (Figure 4F). We believe this is a conservative estimate due to the fact that cis-regulatory elements do not necessarily regulate the nearest gene³¹. 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 4G and 4H)⁴². 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

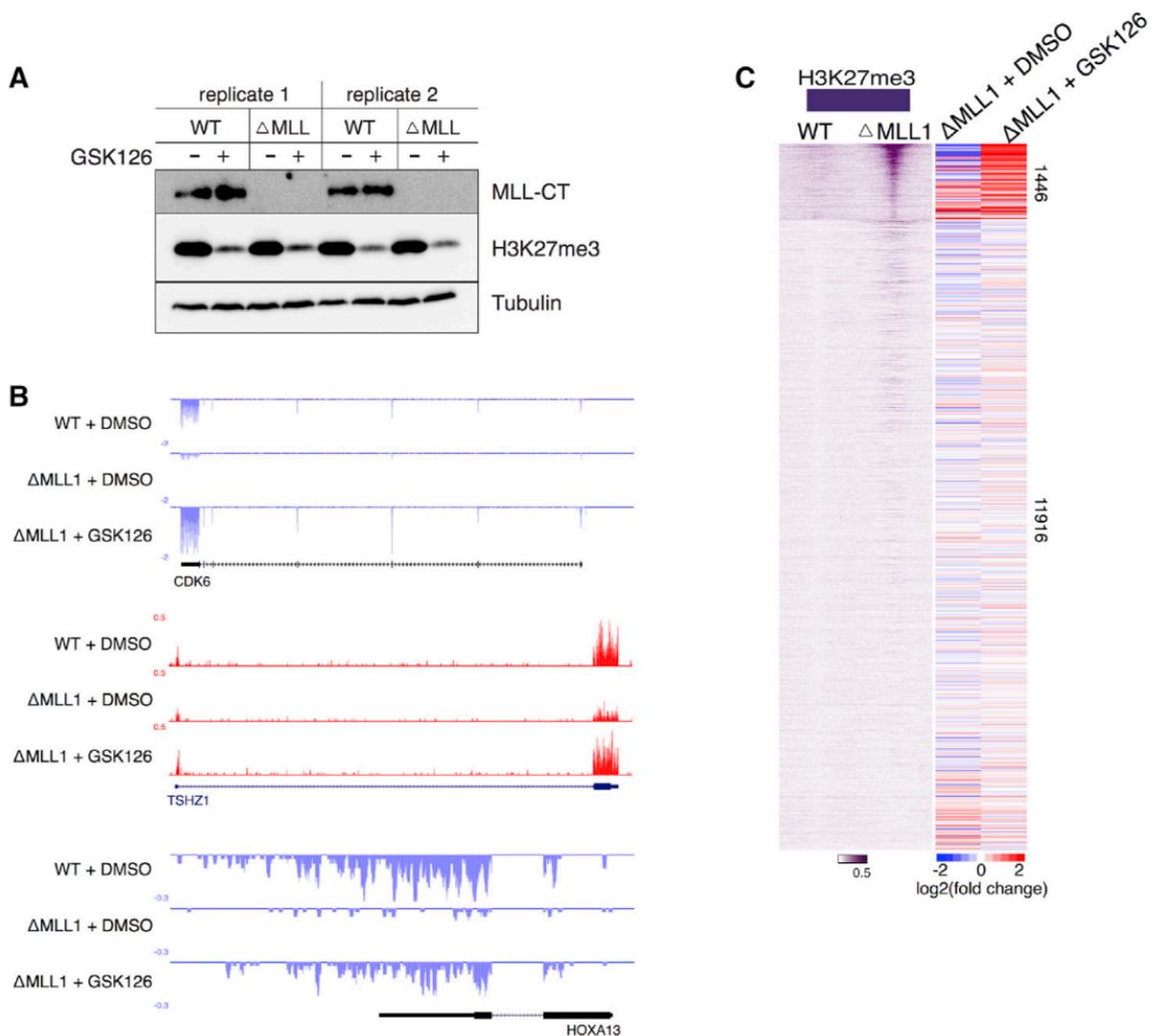
Previous work in *Drosophila* cells demonstrates that a subset of Trx-regulated genes exist in a “balanced” state, in which their transcriptional output is simultaneously influenced by Trx and PcG complexes²⁸. In the presence of PcG proteins, these genes require Trx to maintain some level of transcriptional 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^{12,28}.

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).

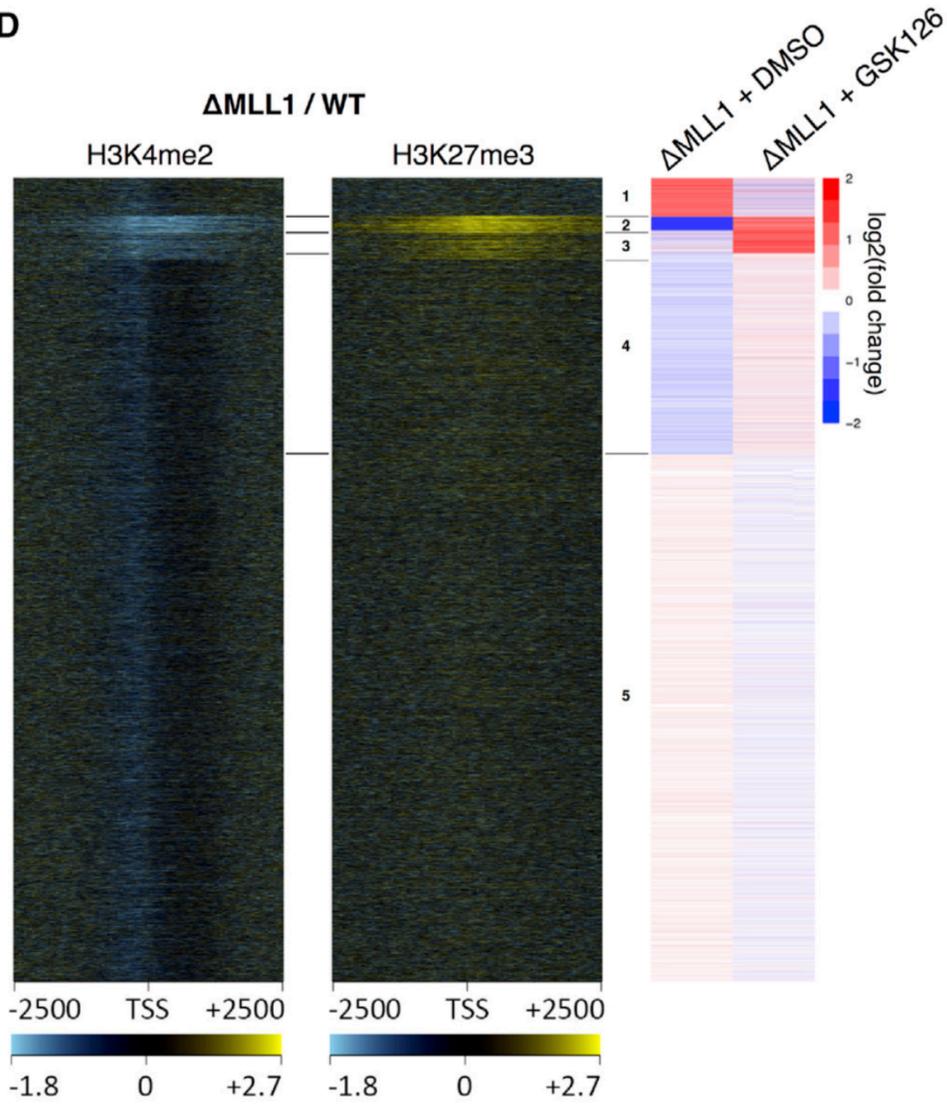
To validate the specificity of our H3K4me2-based method of PRE-like element identification, we extended our analysis genome-wide to consider differential expression of all protein-coding genes. Using K-means clustering to partition genes based on differential expression in the MLL1^{NULL} cells +/- GSK126, 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^{6,43}. 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).

Figure 4



D



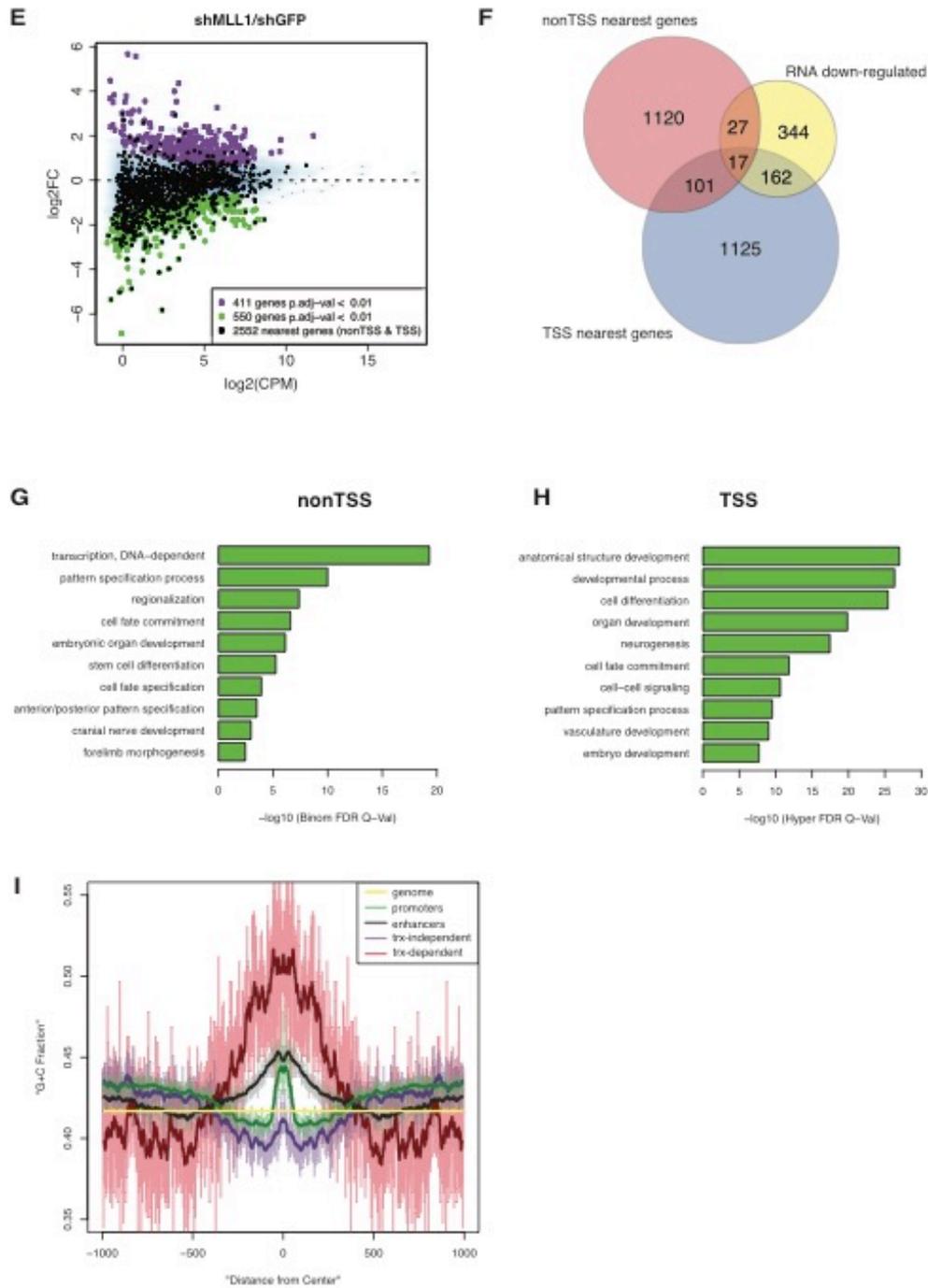


Figure 4.**MLL1 target-genes are transcriptionally balanced by PRC2.**

(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). (E) MA-plot showing differentially expressed genes after MLL-depletion. Black boxes represent 2552 nearest genes to PRE-like sites. Note the majority of black boxes overlap genes negatively affected by shMLL versus up-regulated genes. (F) Venn diagram showing overlap between the 550 down-regulated genes, the 1405 genes with reduced H3K4me2 at their TSS, and the 1265 nearest genes to nonTSS sites of MLL1-dependent H3K4me2. (G) GREAT GO-term enrichment for genes associated with MLL-dependent TSSs are enriched for developmental genes. (H) GREAT analysis of nearby genes associated with MLL-dependent nonTSS sites are enriched for similar GO-terms, including transcription factor genes, indicating the TSS and nonTSS sites co-regulate the same set of genes. (I) Frequency profiles of G+C content surrounding various *Drosophila* regulatory elements. Raw G+C fractions of occurrence were generated from forward and reverse complement sequences aligned to the

transcription start sites of all dm3 Ensembl 70 transcripts (green) and to the centers of STARR-seq enhancers (gray) (Arnold et al., 2013), trx- independent sites (purple), and trx-dependent sites (red). Smoothed profiles are represented by darker colors and were generated using 21 bp moving averages. The average G+C content in the genome is also plotted (yellow).

Discussion

From flies to humans, the COMPASS family of H3K4-methylases have increased in number and diversified in function². 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 humans⁴³, the identification of vertebrate PREs that function similarly to those in *Drosophila* is still under investigation^{16,44}. 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 discrete 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^{19,45}. 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 4I).

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 mESCs⁹. 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^{46,47}.

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 cells⁴⁸. 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 genes¹¹. One important implication of this finding is that, in the absence of MLL1, on-going 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 CGIs¹⁹. 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 in⁴.

Cell culture and antibodies.

HCT116 cells were grown in DMEM medium with 10% FBS. GSK126 was administered for 96 hours at [5uM]⁴⁹. 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 µg/ml Polybrene (Sigma) for 24 hr followed by 2 µg/ml of puromycin for 72 h before harvest.

HCT116 ChIP

ChIP samples were prepared as previously described⁵⁰, 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.

Drosophila S2 ChIP-seq

Drosophila S2 cells (two T75 flasks per antibody) were cross-linked in 1% paraformaldehyde for 10 min at room temperature. Samples were quenched in 225 mM Glycine. Following centrifugation at 2000 g for 5 min at 4°C, the supernatant was aspirated. The cell pellet was resuspended in 5 mL of Orlando/Paro buffer (10 mM Tris HCl at pH 7.5, 10 mM EDTA, 0.5 mM EGTA, 0.25% Triton X-100, 0.5 mM DTT, protease inhibitors [complete, EDTA-free; Roche]) and centrifuged at 2000g for 5 min at 4°C. The supernatant was aspirated, and the wash step with Orlando/Paro buffer was repeated another two times. The cell pellet (~200 µL for two T75 flask) was resuspended in 6 mL of RIPA buffer (25 mM Tris at pH 7.5, 140 mM NaCl, 1% Triton X-100, 1 mM EDTA, 0.1% SDS, 0.1% Na-deoxycholate, 0.5% N-lauroylsarcosine, 0.5 mM DTT, protease inhibitors [complete, EDTA-free; Roche]) and was sonicated at 60% output for 10 min (30 sec on / 90 sec off for 20 cycles) (Misonix 3000) in 15-mL hard plastic tubes. Sonicated chromatin was centrifuged at 20,817 g for 20 min at 4°C, and the supernatant was kept. Thirty microliters of the sonicated chromatin was kept for gel analysis (200 - 600 bp), and 60 µL was used as an input control. Sizing samples were reverse-cross-linked overnight at 65°C by adding 70 µL of RIPA and 3 µL of proteinase K (30 mg/mL), and input samples were reverse-cross-linked by adding 40 µL of RIPA, incubated at 65°C overnight, and processed/purified in the

same way as the ChIP samples (see below). The remaining chromatin was incubated overnight at 4°C with the respective antibody on a nutator. Sixty microliters of protein A agarose (Invitrogen) was washed in 5 mL of RIPA buffer and centrifuged at 200 g for 2 min at 4°C. The supernatant was aspirated, and the chromatin sample was added and incubated for 2 h at 4°C on a nutator. After centrifugation at 500 g for 2 min at 4°C, the protein A agarose was transferred to a 1.5-mL tube, washed with 1 mL of RIPA buffer, incubated for 5 min at room temperature on a nutator, and centrifuged at 500 g for 2 min at 4°C. The supernatant was aspirated, and the previous washing steps were repeated another seven times. Elution was performed on a nutator for 20 min at room temperature with 300 mL of elution buffer (0.1 M NaHCO₃, 1% SDS) and the sample was centrifuged at 500 g for 2 min at room temperature. The supernatant was kept, and the elution step was repeated. Elution fractions were pooled and reverse cross-linked overnight at 65°C. Samples were then diluted two-fold in TE, treated with 1 µL RNase A (Sigma) for 1 h at 37°C, then treated with 5 µL proteinase K at 50°C for 2 h. DNA was isolated with the Qiagen PCR purification kit and eluted in 50 µL Buffer EB. Up to 10ug eluted DNA was used to prepare ChIP-seq libraries using TruSeq universal adapters (Illumina) and KAPA Library Preparation Kit (KAPA Biosystems). SPRIselect Reagent was used to select samples between 200-400 bp for next- generation sequencing (Illumina HiSeq).

RNAi in S2

S2 cells were maintained at 2-10e6 cells/mL in SFX medium (containing 1% penicillin/streptomycin) prior to RNAi treatment. Cells were plated at 5e5 cells/mL in 20mL SFX per T75 flask and treated with 100 µg dsRNA for 5.5 days. dsRNA was generated using T7 RiboMAXTM Large Scale RNA Production Kit (Promega). Primers used to prepare dsRNA were:

LacZ forward, TAATACGACTCACTATAGGGAGGAATGCTTAATCAGTGA GGCACC; LacZ reverse, TAATACGACTCACTATAGGGAGGAA AGCCATACCAAACGACGAGC. Trr_1 forward,

TAATACGACTCACTAT AGGGCGGAGACTCGCCTGGCAGCTTCTGC; Trr_1 reverse,
 TAATACGACTCACTATAGGGCCTGGTTGGTGACAAGC GCTACACG; Trr_2 forward,
 TTAATACGACTCACTATAGG GAGAAAGACGGAGCTGCTTCTCGGA; Trr_2 reverse, TTAA
 TACGACTCACTATAGGGAGACATCAGCTGGGTTTTTCATC TTGG; Trx_1 forward,
 TAATACGACTCACTATAGGGGCCAG TGTGTCCAAGTGCTATGCCC; Trx_1 reverse,
 TAATACGACTC ACTATAGGGGCGCTGGCATCCACTTCCATCGTCG; Trx_2 forward,
 TAATACGACTCACTATAGGGGCAATGCAGCAG ATCAAAAA; and Trx_2 reverse,
 TAATACGACTCACTATAG GGTTCGATTCATCACCAACAGGA.

Cell culture and antibodies.

HCT116 cells were grown in DMEM medium with 10% FBS. GSK126 was administered for 96 hours at [5uM] (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).

Lentivirus-mediated RNA interference (RNAi).

Parental HCT116 (MLL3null) cells were infected with lentivirus either for green fluorescent protein (GFP) control short hairpin RNA (shRNA) or MLL1 shRNA in the presence of 8 µg/ml Polybrene (Sigma) for 24 h (target sequence for MLL1: GCCAAGCACTGTCGAAATTAC) followed by 2 µg/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.

Data Acquisition and Processing

Drosophila RNA-seq and ChIP-seq samples (except replicate 2 for H3K27ac and H3K27me3) were sequenced with the Illumina HiSeq technology. All other samples were sequenced with the

Illumina NextSeq technology. Output data from HiSeq and NextSeq sequencing systems were processed with the Casava v1.8 and bcl2fastq software tools, respectively. Sequence quality was assessed using FastQC v 0.11.2 (Andrew 2010) for all samples and quality trimming was done using the FASTX toolkit. RNA-seq and ChIP-seq reads were aligned to the hg19 and dm3 genomes using TopHat v2.0.9. (Kim et al., 2013) and Bowtie v0.12.9 (Langmead et al., 2009), and only uniquely mapped reads with a two-mismatch threshold were considered for downstream analysis. Gene annotations from Ensembl 72 were used for HCT116 cells, and gene annotations from Ensembl 70 were used for S2 cells. To calculate coverage for ChIP-seq data, reads were extended to 150 bases in the 5' to 3' direction and then normalized to the total number of reads per million (RPM). ChIP-seq peaks were called with MACS v1.4.2 (Zhang et al., 2008) using default parameters with an FDR cutoff of 0.05%. External sequence data were acquired from GEO, and raw reads and aligned in the same way as internally sequenced samples. Data generated for this study are available under GEO accession number GSE81795. ChIP-seq data for PRC1 components in S2 cells come from GEO accession number GSE24521 (Enderle et al., 2010). ChIP-seq data for CTCF and DNase Hypersensitivity in HCT116 cells come from GEO accession number GSE50610 (Maurano et al., 2015).

Differential Gene Expression Analysis

Gene count tables were constructed using Ensembl gene annotations and used as input for edgeR 3.0.8 (Robinson et al., 2010). Genes with Benjamini-Hochburg adjusted p-values less than 0.01 were considered differentially expressed.

Differential H3K4me2 Occupancy Analysis

To identify sites where knockdown of Trx significantly decreased H3K4me2 occupancy in S2 cells, read counts of H3K4me2 were quantified under TrxCt peaks from two biological replicates for each condition, and edgeR 3.0.8 was used to perform differential occupancy analysis. A pvalue cutoff of $1e-3$ and a log₂ fold change cutoff of -1 were the criteria used to define a site

with significant H3K4me2 loss. Similarly, to identify sites where knockdown of MLL1 significantly decreased H3K4me2 occupancy in HCT116 cells, read counts of H3K4me2 were quantified under H3K4me2 peaks from two biological replicates for each condition, and differential occupancy was evaluated by edgeR in the same manner. HCT116 H3K4me2 peaks were separated into TSS and non-TSS groups based on whether or not they overlapped regions within 500 bp of a TSS. Associated heatmap data were centered at TrxCt or H3K4me2 peak centers, binned into 25 bp windows, and sorted by H3K4me2 occupancy based on control samples.

Gene Enrichment Analysis

Interpro and GO Biological Process results for Figures 1D, 1E, and 2B were obtained using DAVID (Huang et al., 2008) while GO Biological Process results for Figures 4G and 4H were obtained using GREAT (McLean et al., 2010). For the 1446 TSS MLL1-dependent sites, the 'single nearest gene' association rule setting was used, and the hypergeometric test results were presented. For the 1386 nonTSS MLL1-dependent sites, default parameters were used for identifying cis-regulatory function, and the binomial test results were presented.

CTCF and CpG Island Enrichment Analysis

Enrichment of CTCF and CpG islands at nonTSS MLL1-dependent sites were evaluated using Pearson Chi-squared tests with the `chisq.test` R package (<http://www.r-project.org/>). Specifically, count data were arranged in 2X4 contingency tables such that rows represented sites of H3K4me2 loss or no change while columns represented the presence or absence of a feature at TSS or nonTSS H3K4me2 peaks. The presence of CTCF or a CpG island was defined if any overlap of a CTCF peak or CpG island (defined by the UCSC genome browser coordinates) occurred with an H3K4me2 peak. Observed versus expected count tables indicated that both CTCF and CpG islands at nonTSS MLL1-dependent sites were observed ~3 times more than expected (data not shown). Additionally, motif analysis by the HOMER software (Heinz et al.,

2010) was used to show significant enrichment of the CTCF motif at nonTSS MLL1-dependent sites using CpG islands as background.

Accession Numbers

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

References

- 1 Rickels, R. *et al.* An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS. *Mol Cell* **63**, 318-328, doi:10.1016/j.molcel.2016.06.018 (2016).
- 2 Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81**, 65-95, doi:10.1146/annurev-biochem-051710-134100 (2012).
- 3 Kennison, J. A. & Tamkun, J. W. Dosage-dependent modifiers of polycomb and antennapedia mutations in *Drosophila*. *Proceedings of the National Academy of Sciences* **85**, 8136-8140, doi:10.1073/pnas.85.21.8136 (1988).
- 4 Herz, H. M. *et al.* Enhancer-associated H3K4 monomethylation by Trithorax-related, the *Drosophila* homolog of mammalian Mll3/Mll4. *Genes & Development* **26**, 2604-2620, doi:10.1101/gad.201327.112 (2012).
- 5 Mohan, M. *et al.* The COMPASS Family of H3K4 Methylases in *Drosophila*. *Molecular and Cellular Biology* **31**, 4310-4318, doi:10.1128/mcb.06092-11 (2011).
- 6 Wang, P. *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* **29**, 6074-6085, doi:10.1128/mcb.00924-09 (2009).
- 7 Wu, M. *et al.* Molecular Regulation of H3K4 Trimethylation by Wdr82, a Component of Human Set1/COMPASS. *Molecular and Cellular Biology* **28**, 7337-7344, doi:10.1128/mcb.00976-08 (2008).
- 8 Hu, D. *et al.* The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and Cellular Biology* **33**, 4745-4754, doi:10.1128/mcb.01181-13 (2013).
- 9 Hu, D. *et al.* The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. *Nat Struct Mol Biol* **20**, 1093-1097, doi:10.1038/nsmb.2653 (2013).
- 10 Poux, S., Horard, B., Sigrist, C. J. & Pirrotta, V. The *Drosophila* trithorax protein is a coactivator required to prevent re-establishment of polycomb silencing. *Development* **129**, 2483-2493 (2002).
- 11 Klymenko, T. & Muller, J. The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. *EMBO Rep* **5**, 373-377, doi:10.1038/sj.embor.7400111 (2004).

- 12 Schuettengruber, B., Martinez, A. M., Iovino, N. & Cavalli, G. Trithorax group proteins: switching genes on and keeping them active. *Nat Rev Mol Cell Biol* **12**, 799-814, doi:10.1038/nrm3230 (2011).
- 13 Bauer, M., Trupke, J. & Ringrose, L. The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma*, doi:10.1007/s00412-015-0539-4 (2015).
- 14 Simon, J. A. & Kingston, R. E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* **49**, 808-824, doi:10.1016/j.molcel.2013.02.013 (2013).
- 15 Levine, S. S., King, I. F. & Kingston, R. E. Division of labor in polycomb group repression. *Trends Biochem Sci* **29**, 478-485, doi:10.1016/j.tibs.2004.07.007 (2004).
- 16 Mendenhall, E. M. *et al.* GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet* **6**, e1001244, doi:10.1371/journal.pgen.1001244 (2010).
- 17 Klose, R. J., Cooper, S., Farcas, A. M., Blackledge, N. P. & Brockdorff, N. Chromatin sampling--an emerging perspective on targeting polycomb repressor proteins. *PLoS Genet* **9**, e1003717, doi:10.1371/journal.pgen.1003717 (2013).
- 18 Tanay, A., O'Donnell, A. H., Damelin, M. & Bestor, T. H. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc Natl Acad Sci U S A* **104**, 5521-5526, doi:10.1073/pnas.0609746104 (2007).
- 19 Riising, E. M. *et al.* Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol Cell* **55**, 347-360, doi:10.1016/j.molcel.2014.06.005 (2014).
- 20 Illingworth, R. S. *et al.* Orphan CpG islands identify numerous conserved promoters in the mammalian genome. *PLoS Genet* **6**, e1001134, doi:10.1371/journal.pgen.1001134 (2010).
- 21 Mohan, M., Herz, H.-M. & Shilatifard, A. SnapShot: Histone Lysine Methylase Complexes. *Cell* **149**, 498-498.e491, doi:10.1016/j.cell.2012.03.025 (2012).
- 22 Capotosti, F., Hsieh, J. J. D. & Herr, W. Species Selectivity of Mixed-Lineage Leukemia/Trithorax and HCF Proteolytic Maturation Pathways. *Molecular and Cellular Biology* **27**, 7063-7072, doi:10.1128/mcb.00769-07 (2007).
- 23 Hsieh, J. J. D., Ernst, P., Erdjument-Bromage, H., Tempst, P. & Korsmeyer, S. J. Proteolytic Cleavage of MLL Generates a Complex of N- and C-Terminal Fragments That Confers Protein Stability and Subnuclear Localization. *Molecular and Cellular Biology* **23**, 186-194, doi:10.1128/mcb.23.1.186-194.2003 (2003).
- 24 Simon, J. A. & Kingston, R. E. Mechanisms of polycomb gene silencing: knowns and unknowns. *Nat Rev Mol Cell Biol* **10**, 697-708, doi:10.1038/nrm2763 (2009).
- 25 Enderle, D. *et al.* Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. *Genome Research* **21**, 216-226, doi:10.1101/gr.114348.110 (2010).
- 26 Beisel, C. *et al.* Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. *Proceedings of the National Academy of Sciences* **104**, 16615-16620, doi:10.1073/pnas.0701538104 (2007).
- 27 Chinwalla, V., Jane, E. P. & Harte, P. J. The Drosophila trithorax protein binds to specific chromosomal sites and is co-localized with Polycomb at many sites. *EMBO J* **14**, 2056-2065 (1995).

- 28 Schwartz, Y. B. *et al.* Alternative Epigenetic Chromatin States of Polycomb Target Genes. *PLoS Genetics* **6**, e1000805, doi:10.1371/journal.pgen.1000805 (2010).
- 29 Ringrose, L., Rehmsmeier, M., Dura, J.-M. & Paro, R. Genome-Wide Prediction of Polycomb/Trithorax Response Elements in *Drosophila melanogaster*. *Developmental Cell* **5**, 759-771, doi:10.1016/s1534-5807(03)00337-x (2003).
- 30 Gohl, D., Muller, M., Pirrotta, V., Affolter, M. & Schedl, P. Enhancer Blocking and Transvection at the *Drosophila* apterous Locus. *Genetics* **178**, 127-143, doi:10.1534/genetics.107.077768 (2008).
- 31 Kvon, E. Z. *et al.* Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature*, doi:10.1038/nature13395 (2014).
- 32 Francis, N. J. & Kingston, R. E. Mechanisms of transcriptional memory. *Nature Reviews Molecular Cell Biology* **2**, 409-421, doi:10.1038/35073039 (2001).
- 33 Arnold, C. D. *et al.* Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* **339**, 1074-1077, doi:10.1126/science.1232542 (2013).
- 34 Simon, J., Chiang, A., Bender, W., Shimell, M. J. & O'Connor, M. Elements of the *Drosophila* Bithorax Complex That Mediate Repression by Polycomb Group Products. *Developmental Biology* **158**, 131-144, doi:10.1006/dbio.1993.1174 (1993).
- 35 Chan, C. S., Rastelli, L. & Pirrotta, V. A Polycomb response element in the *Ubx* gene that determines an epigenetically inherited state of repression. *EMBO J* **13**, 2553-2564 (1994).
- 36 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* **10**, 3202-3215, doi:10.1101/gad.10.24.3202 (1996).
- 37 Cavalli, G. & Paro, R. Epigenetic inheritance of active chromatin after removal of the main transactivator. *Science* **286**, 955-958 (1999).
- 38 Maurano, M. T. *et al.* Role of DNA Methylation in Modulating Transcription Factor Occupancy. *Cell Rep* **12**, 1184-1195, doi:10.1016/j.celrep.2015.07.024 (2015).
- 39 Caslini, C., Connelly, J. A., Serna, A., Broccoli, D. & Hess, J. L. MLL Associates with Telomeres and Regulates Telomeric Repeat-Containing RNA Transcription. *Molecular and Cellular Biology* **29**, 4519-4526, doi:10.1128/mcb.00195-09 (2009).
- 40 Milne, T. A. *et al.* From The Cover: MLL associates specifically with a subset of transcriptionally active target genes. *Proceedings of the National Academy of Sciences* **102**, 14765-14770, doi:10.1073/pnas.0503630102 (2005).
- 41 Takeda, S. Proteolysis of MLL family proteins is essential for Taspase1-orchestrated cell cycle progression. *Genes & Development* **20**, 2397-2409, doi:10.1101/gad.1449406 (2006).
- 42 McLean, C. Y. *et al.* GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* **28**, 495-501, doi:10.1038/nbt.1630 (2010).
- 43 Bracken, A. P. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes & Development* **20**, 1123-1136, doi:10.1101/gad.381706 (2006).
- 44 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343-349, doi:10.1038/nature09784 (2011).

- 45 Cooper, S. *et al.* Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep* **7**, 1456-1470, doi:10.1016/j.celrep.2014.04.012 (2014).
- 46 Li, H. B. *et al.* Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in *Drosophila melanogaster*. *Mol Cell Biol* **31**, 616-625, doi:10.1128/MCB.00849-10 (2011).
- 47 Li, H. B., Ohno, K., Gui, H. & Pirrotta, V. Insulators target active genes to transcription factories and polycomb-repressed genes to polycomb bodies. *PLoS Genet* **9**, e1003436, doi:10.1371/journal.pgen.1003436 (2013).
- 48 Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**, aad9780, doi:10.1126/science.aad9780 (2016).
- 49 McCabe, M. T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108-112, doi:10.1038/nature11606 (2012).
- 50 Lee, T. I., Johnstone, S. E. & Young, R. A. Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat Protoc* **1**, 729-748, doi:10.1038/nprot.2006.98 (2006).

Methods References

Enderle, D., Beisel, C., Stadler, M.B., Gerstung, M., Athri, P., and Paro, R. (2010). Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. *Genome Research* **21**, 216-226.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589.

Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* **14**, R36.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25.

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.

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.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139-140.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137.

Chapter 2

Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability¹

Abstract

Histone H3 lysine 4 monomethylation (H3K4me1) is an evolutionarily conserved feature of enhancer chromatin catalyzed by the Trr/MLL3/4-COMPASS family²⁻⁴. Here we demonstrate that *Drosophila* embryos expressing catalytically deficient Trr-COMPASS eclose and develop to a productive adulthood. Parallel experiments with a Trr allele that augments enzyme product specificity reveal that conversion of H3K4me1 at enhancers to H3K4me2 and H3K4me3 is also compatible with life and results in minimal changes in gene expression. Similarly, loss of mammalian MLL3 and MLL4 catalytic SET domain in embryonic stem cells does not disrupt self-renewal capability of the ES cells. Trr catalytic mutant alleles manifest subtle developmental phenotypes when subjected to temperature stress or altered cohesin levels. Collectively, our findings suggest that metazoan development can occur in the context of Trr/COMPASS with H3K4me1 enzymatic deficiency, and points to a possible role for H3K4me1 on cis-regulatory elements in specific settings to fine-tune transcriptional regulation in response to environmental stress.

Introduction

Transcriptional enhancers are cis-regulatory elements that potentiate transcriptional output even when separated from their cognate promoter by megabases of intervening DNA⁵⁻⁸. Enhancer-promoter communication is thought to occur through long-range looping mechanisms facilitated by cohesin complexes and other nuclear factors⁹. Enrichment of specific histone modifications, in particular H3K27-acetylation (H3K27ac) and H3K4me1, is an evolutionarily conserved feature of enhancer chromatin^{10,11}. Previous work from our laboratory established the initial link between Trr/COMPASS in *Drosophila* and its mammalian homologues MLL3/4-COMPASS as the enhancer H3K4 monomethylases required for enhancer promoter communication during development¹⁻². Subsequent studies confirmed our original findings that MLL3/4-COMPASS are H3K4 monomethylases functioning on enhancers, and that they are essential for embryonic development in mammals^{12,13}. Importantly, recent studies implicate defective MLL3/4 function in the pathogenesis of various forms of cancer¹⁴. However, the specific requirements for H3K4me1 catalytic activity during organismal development are largely undefined.

Results

To investigate the role of Trr catalyzed H3K4me1 at cis-regulatory elements in *Drosophila*, we complemented the embryonic lethal *trr*¹ null allele¹⁵ with various *trr* rescue transgenes. Using site-specific integration, we rescued *trr*¹ lethality using a 12kb transgenic genomic construct encompassing the *trr* locus and its associated regulatory elements. We also introduced specific amino acid substitutions into the Trr SET domain at highly conserved residues that disrupt Set1 methyltransferase activity in yeast (Figure 1B). A Trr cysteine-to-alanine (C2398A) mutation is catalytically deficient when reconstituted *in vitro* (Figure 1C), and cell lysates from C2398A (referred to as Trr-C/A) larval tissues demonstrate a large reduction in bulk H3K4me1 levels without affecting Trr, Trx, or dSet1 protein stability (Figure 1D). This closely resembles the effect of Trr-RNAi on H3K4me1 levels, and is consistent with our original findings that Trr predominantly catalyzes H3K4me1 in *Drosophila*².

We next tested the importance of Trr product-specificity by mutating the “F/Y switch” position, which converts Set1/COMPASS to a more efficient H3K4-di- and trimethylase, *in vitro* and *in vivo*^{16,17}. Indeed, introducing this mutation at the corresponding residue (Y2383F, here on referred to as Trr-Y/F) in Trr’s SET domain shifts its enzymatic activity towards that of an H3K4-di and –trimethylase without affecting protein stability (Figures 1B-D).

Intriguingly, both the catalytic-deficient (Trr-C/A) and –hyperactive (Trr-Y/F) mutations rescue the recessive lethal *trr*¹ allele, producing fertile adults with normal life-span and no gross abnormalities (Figure 1E). Because loss of Trr results in embryonic lethality, we were intrigued that disrupting its catalytic activity produces no apparent phenotype, given the strong evolutionary conservation of the COMPASS SET domain (Figure 1B). Sanger-sequencing

confirmed the intended mutations in both genomic DNA and mRNA (Figures 1G and 1H). We removed the *trr-rescue* transgene through genetic crosses and confirmed that *trr*¹ recessive lethality persists, thus ruling out the possibility that recombination occurred on the chromosome X to produce a functional Trr enzyme (Figure 1F). Two additional lethal alleles, *trr*^{C2375X} and *trr*^{K662X}¹⁸, were also rescued in the same manner to confirm complementation is not unique to *trr*¹ (data not shown). These observations suggest an essential methylase-independent role for Trr in regulating enhancer-mediated processes critical for *Drosophila* development.

Figure 1

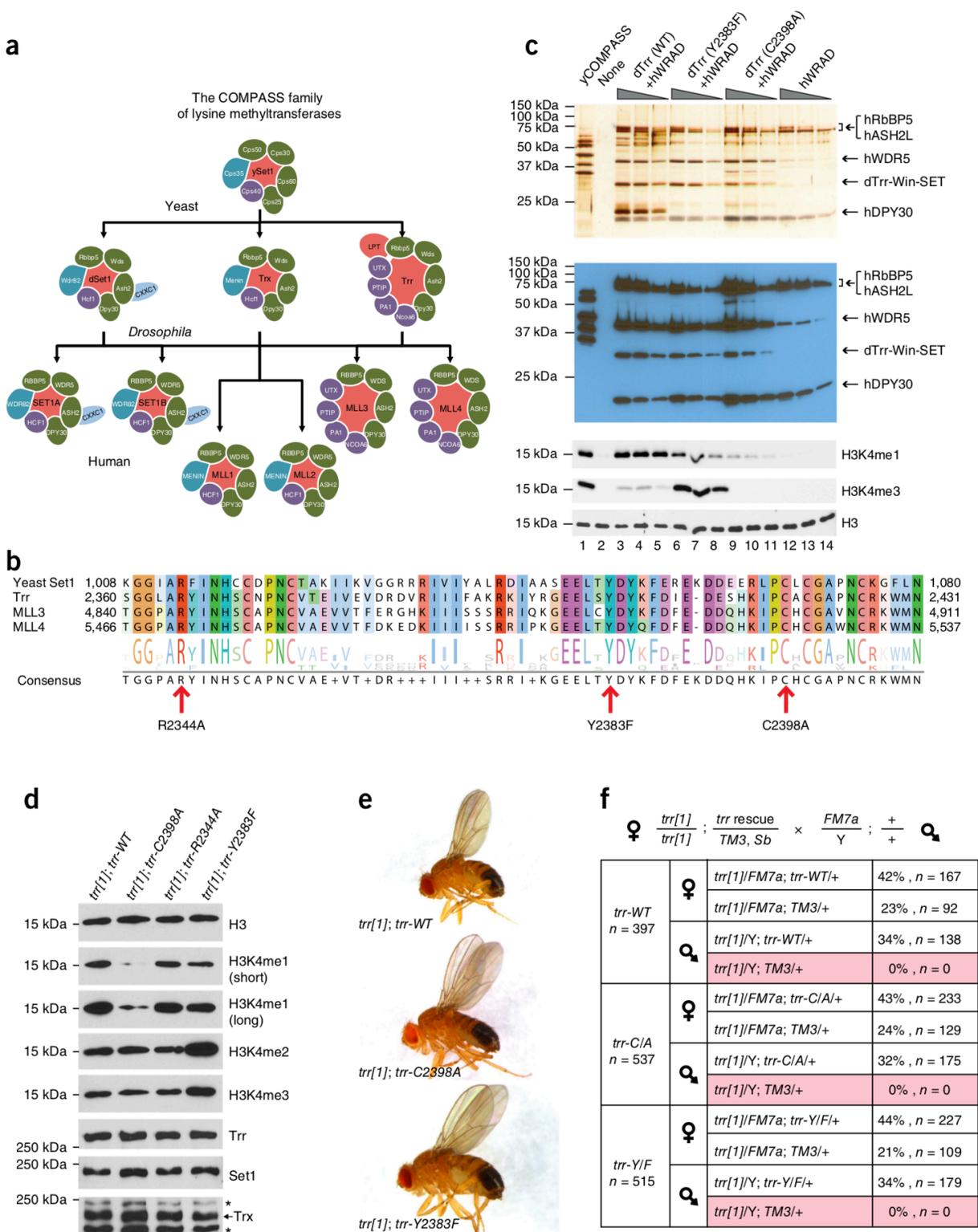


Figure 1.**The catalytic activity of Trr is dispensable for viability.**

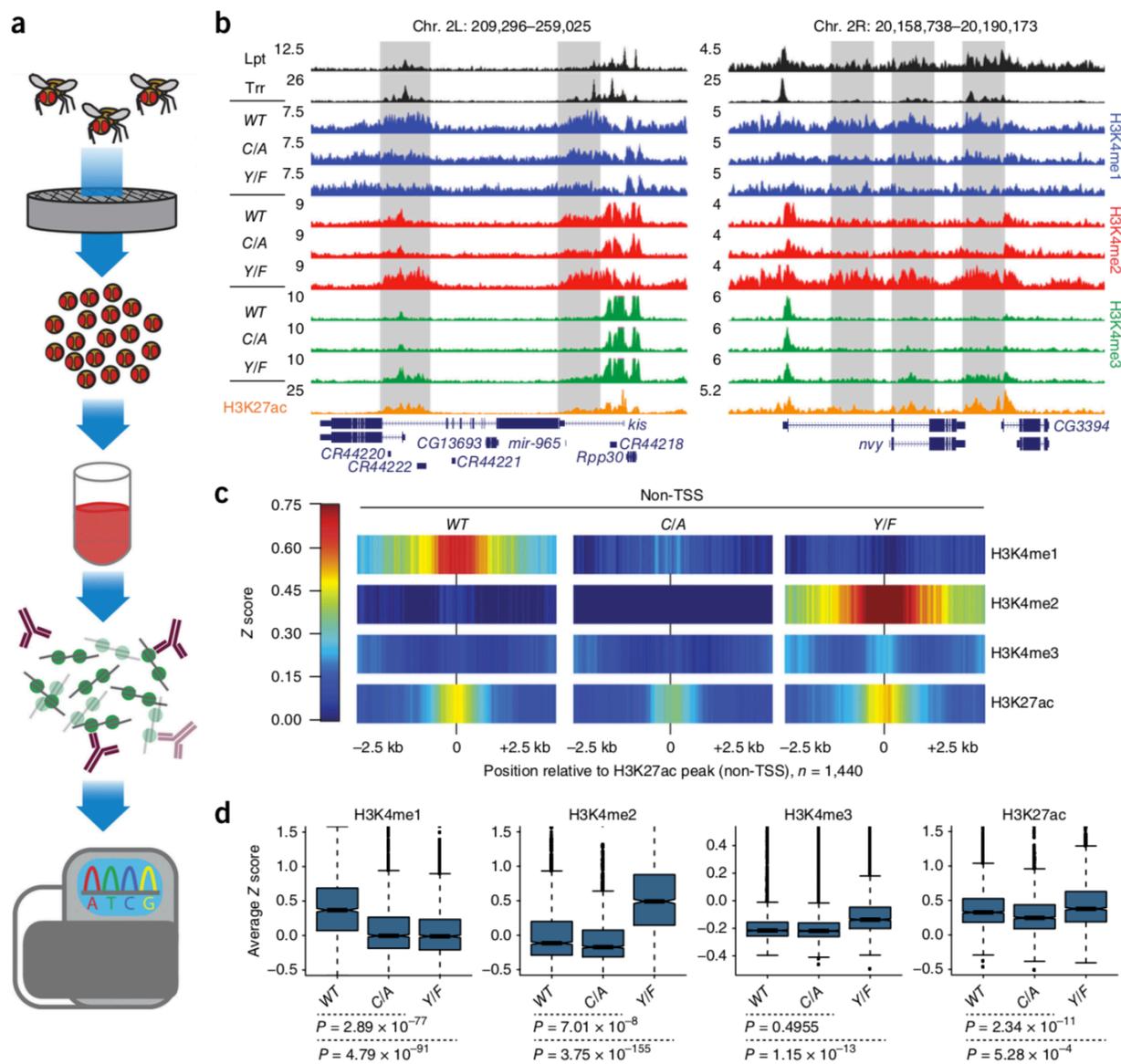
(A) The COMPASS family of lysine-methyltransferases is conserved from yeast to humans. Common subunits are shown in green, shared subunits in purple, and complex-specific subunits in blue. (B) Alignment of the SET domains from Trr, MLL3, MLL4 and yeast Set1. Single amino acid substitutions were introduced into the SET domain of Trr at the positions highlighted by red arrows. Alignment plots were created in Jalview. (C) *Drosophila* Trr SET domain was reconstituted using Baculovirus in SF9 cells along with flag-tagged human COMPASS core subunits (WRAD): WDR5, RbBP5, ASH2L, and DPY30. Silver-staining and anti-flag western blots show the purified components, and western blots for H3K4me1 and H3K4me3 demonstrate diminished catalytic activity of the C2398A mutation and hyperactivating effect of Y2383F on Trr activity *in vitro*. (D) Western blot of lysates from wing imaginal discs. The catalytically deficient *trr-C2398A* construct is not able to restore H3K4me1 (lane 2, panels 2-3, short and long exposures) while the hyperactive *trr-Y2383F* construct exhibits increased levels in H3K4me2 (lane 4, panel 4) and H3K4me3 (lane 4, panel 5). Lane 1: *trr[1] ; ; trr-WT*. Lane 2: *trr[1] ; ; trr-C2398A*. Lane 3: *trr[1] ; ; trr-R2344A*. Lane 4: *trr[1] ; ; trr-Y2383F*. Panel 1: α -H3. Panel 2: α -H3K4me1, short exposure. Panel 3: α -H3K4me1, long exposure. Panel 4: α -H3K4me2. Panel 5: α -H3K4me3. Panel 6: α -Trr. Panel 7: α -Set1. Panel 8: α -Trx. The arrow marks the Trx band and the asterisks indicate unspecific bands. The R2344A mutation showed no effect on bulk H3K4me1 levels and was not included in subsequent experiments. (E) Lethality of the *trr* null allele *trr[1]* can be rescued by various genomic *trr* constructs containing the *trr* gene and putative regulatory regions. Only male rescued flies are shown. From top to bottom: *trr-WT*, *trr-C2398A* (inactive), and *trr-Y2383F* (hyperactive). (F) Genetic analyses confirm the *trr[1]* allele remains hemizygous lethal in the absence of either of the three *trr-rescue* transgenes. Lack of *trr[1]* males are highlighted in pink. (G) Chromatograms confirming the correct mutations

incorporated into trr[1] genomic DNA and expressed as mRNA; relevant codons are highlighted in green. The catalytically inactive (C2398A) mutant is a TGC-to-GCC conversion, and the catalytically hyperactive (Y2383F) mutant is a TAC-to-TTC conversion. Note the overlapping peaks due to wild-type sequence at the endogenous trr locus (which is transcribed but truncates after 88 codons; only 3.6% of the full Trr protein). (H) The null trr[1] allele has a point mutation (CAG>TAG) that generates a premature stop codon after the first 88 residues (Q89X). Chromatograms are shown confirming the nonsense mutation (red peak = T) present in mRNA from all three trr[1] rescue lines (marked by a red arrowhead). Note the overlapping peaks due to lack of Q89X mutation in the three trr rescue constructs integrated into chromosome 3L.

To test whether our Trr catalytic mutations specifically affect H3K4-methylation at *cis*-regulatory elements, we mapped the genomic distribution of these modifications using ChIP-seq in adult fly brain tissues (Figure 2A and 2E). Track examples of two representative genes, *kis* and *nvy*, show reduction of H3K4me1 at intronic and intergenic regions in both *trr-C/A* and *trr-Y/F* mutant brain tissues (Figure 2B). These changes overlap with sites of H3K27ac, a known mark of transcriptionally active chromatin, as well as binding of Trr and Lpt (a component of Trr/COMPASS¹). Remarkably, these same sites exhibit conversion of H3K4me1 to H3K4me2/3 in *trr-Y/F* catalytic-hyperactive flies (Figure 2B and 2C).

To confirm that these effects occur genome-wide, we identified H3K27ac peaks and divided them into 5367 transcription start sites (TSS) and 1440 nonTSS sites. ChIP-seq reads were converted to Z-scores and displayed as density bar-plots centered on nonTSS H3K27ac peaks. At these 1440 putative enhancers, H3K4me1 is either diminished in the *trr-C/A* mutation, or converted to H3K4me2/3 in the *trr-Y/F* allele (Figures 2C-D and 2F). Interestingly, enhancer H3K27ac levels are modestly but reproducibly correlated with H3K4-methylation levels, decreasing in the *trr-C/A* and increasing slightly in the *trr-Y/F*. The effects caused by these *trr* catalytic mutations are specific to enhancers, and are not observed at TSS regions. By contrast, H3K4me1 in the *trr-C/A* is slightly increased at TSSs, consistent with Trr-RNAi experiments² (Figure 2G). We performed ChIP-seq for both Trr and Lpt components of COMPASS in our transgenic flies and found that their genome-wide binding is not substantially affected (Figure 2H). To confirm our findings in a different tissue-type, we repeated these experiments for H3K4me1/3 and H3K27ac using larval wing imaginal discs, and obtained similar results to brain tissues (Figures 2I-K), thus corroborating our previous findings that Trr functions predominantly at enhancers¹.

Figure 2



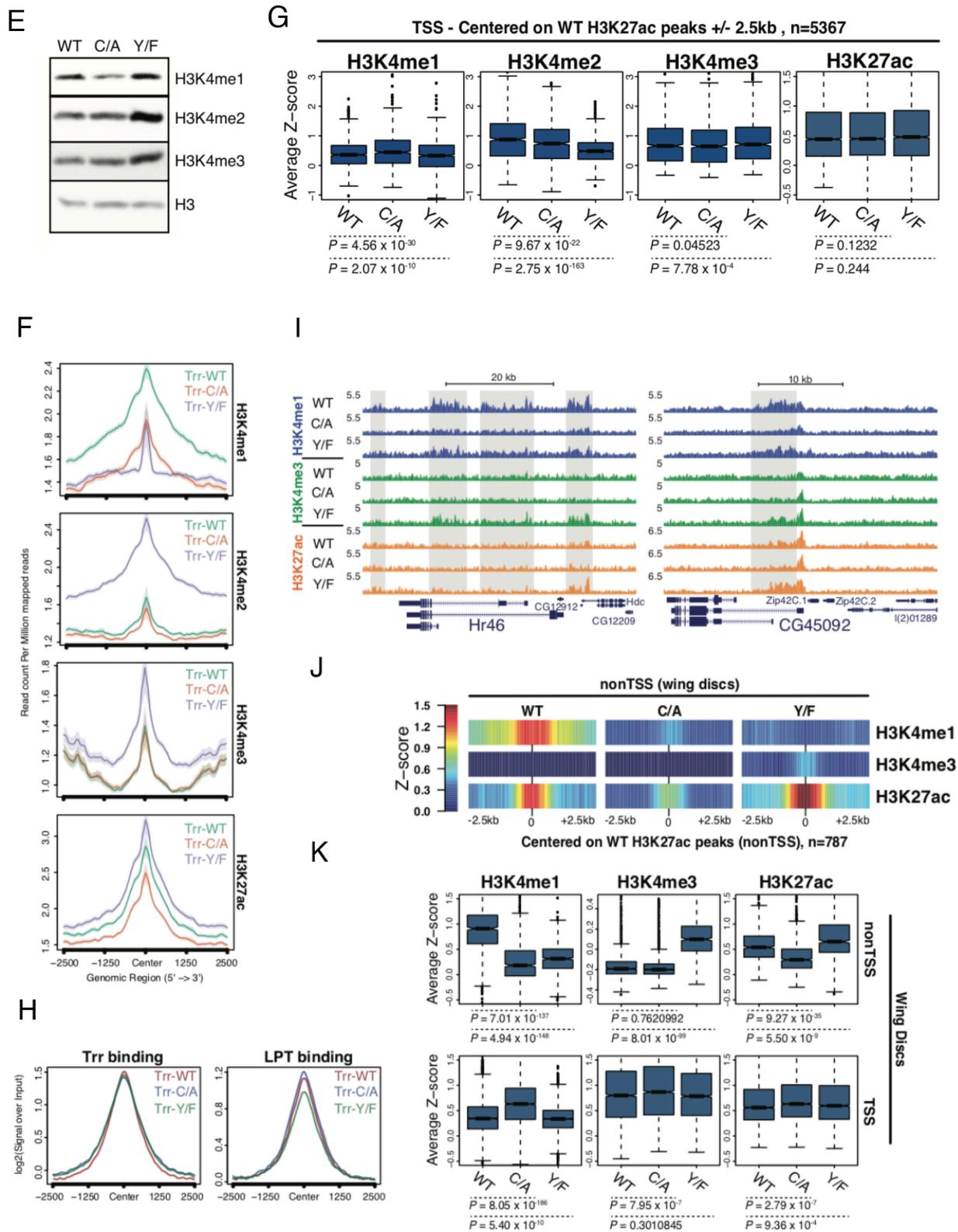


Figure 2.**Enhancer chromatin modification is specifically affected by Trr catalytic mutations.**

(A) Graphical depiction of ChIP-seq experimental workflow using adult fly brain tissues, as described in the methods section. (B) Representative track examples highlighting changes (gray boxes) in enhancer H3K4-methylation. Note the reductions in H3K4me1 in the *trr-C/A* (inactive), and the conversion of H3K4me1 to H3K4me2/3 in the *trr-Y/F* (hyperactive) specifically at H3K27ac-marked nonTSS sites. Image representative of two independent experiments. (C) Density bar-plots comparing H3K4-methylation Z-scores at 1440 active enhancers. Plots are centered on H3K27ac (nonTSS) peaks, +/- 2.5kb. Note the reductions in H3K4me1 in the *trr-C/A* and conversion of H3K4me1 to H3K4me2/3 in the hyperactive *trr-Y/F*. The Z-score scale for H3K4me3 is -0.25 to 0.5, and for H3K27ac is 0 to 1.5. (D) Box plots quantifying the H3K4-methylation changes shown in (C). Upper/lower boundaries represent the interquartile range (IQR). Upper whiskers signify the third quartile (Q3) + 1.5*IQR. Lower whiskers signify Q1 - 1.5*IQR. Dots represent outliers. P-values were calculated by conducting t-tests. (E) Western blots from lysates of adult fly brains confirm H3K4me1 bulk reduction in *trr-C/A*, and H3K4me2/3 bulk increase in *trr-Y/F*, similar to that shown in wing discs. (F) ChIP-seq data from adult fly brains plotted as average profiles, centered on non-TSS H3K27ac peaks; n = 1,440. (G) Box plots of Z-score averages comparing H3K4 methylation and H3K27ac at 5,367 active promoters in adult brain tissue. Plots are centered on H3K27ac peaks overlapping annotated TSSs, ± 2.5 kb. P values are shown below. In comparison to the effects at enhancers, H3K4 methylation levels often show the opposite effect at TSSs. This is most likely an artifact due to redistribution of sequencing reads in the *trr* mutants. (H) Average meta-peak profiles for both Trr and Lpt at 6,807 genomic loci demonstrate that neither of the catalytic mutations substantially affects chromatin binding. (I) ChIP-seq track examples from wing imaginal discs demonstrate reduced H3K4me1 and H3K27ac in *trr-C/A*, as well as increased H3K4me3 and H3K27ac in *trr-*

Y/F at putative enhancers, similar to adult brains. (J) Density bar plots comparing H3K4 methylation and H3K27ac Z scores at 787 active enhancers in wing imaginal discs. Plots are centered on wild-type H3K27ac (non-TSS) peaks, ± 2.5 kb. Note the reductions in H3K4me1 in *trr-C/A* and conversion of H3K4me1 to H3K4me3 in hyperactive *trr-Y/F*. The Z-score scale for H3K4me1 and H3K27ac is 0 to 1.5 and for H3K4me3 is 0 to 1. Note the changes in H3K27ac that track with increased H3K4 methylation. (K) Box plots of ChIP Z-score averages, centered on wild-type H3K27ac peaks, were calculated the same as in G. Quantitative differences in H3K4me1/3 and H3K27ac are shown in wing imaginal disc enhancers ($n = 787$), similar to adult brain tissues. Changes in histone modification at promoters (TSSs) are negligible ($N = 4077$).

To explore the transcriptional consequences of altering H3K4 methylation at enhancers, we performed RNA-seq on both adult brains and larval wing imaginal discs from our transgenic flies (Figure 3A). Despite the alterations of enhancer histone methylation, gene expression in adult brain tissue is largely unaffected in the *trr-C/A* and *trr-Y/F* flies (Figures 3E and 3F). This is predicted, as neither transgenic line demonstrates apparent abnormalities in their behavior or head morphology. In wing imaginal discs, gene-expression profiles for the two catalytic mutants are highly similar to *trr-WT* (>99%, Pearson correlation) (Figure 3B). By contrast, reducing total Trr levels by RNAi, using either *engrailed* or *T80* Gal4 drivers, elicits major changes in gene expression (Figure 3C) suggesting that there are H3K4 methylation-independent functions of Trr in developmental gene regulation. We performed unsupervised hierarchical clustering of the 3168 differentially expressed genes from our six datasets (adjusted $p < 0.01$) and plotted Z-score transformed read counts for comparison. Control samples (LacZ-RNAi) and the three different *trr-rescue* alleles exhibit highly similar patterns of gene expression, whereas Trr-RNAi depleted

samples form a separate cluster with dramatically altered RNA expression profiles (Figure 3C). These results support our hypothesis that the major functions of Trr in transcription regulation and organismal development are not strictly dependent on H3K4me1 methyltransferase activity.

By integrating CHIP-seq and wing disc RNA-seq data, we observed a link between subtle differences in gene expression (Figure 3C) and altered H3K4-methylation levels at nearby enhancers. Centering on total H3K27ac peaks (n=5908), we performed k-means clustering (k=3) (Figure 3D). Cluster #3 (bottom cluster) exhibits the strongest alteration in nonTSS enhancer H3K4-methylation (Figure 3D). Diminished enhancer H3K4me1 in *trr-C/A* flies is associated with decreased expression of the nearest gene ($p = 1.22 \times 10^{-42}$, hypergeometric test), conversely *trr-Y/F* conversion of H3K4me1 to H3K4me2/3 is associated with increased expression of the nearest gene ($p = 6.25 \times 10^{-22}$, hypergeometric test) (Figure 3D). Despite these observed changes in gene expression, all three *trr¹*-rescued fly lines produce adults with morphologically normal wings (Figure 1E), whereas both *trr-RNAi* lines die during pupation, prohibiting analysis of wing development.

Figure 3

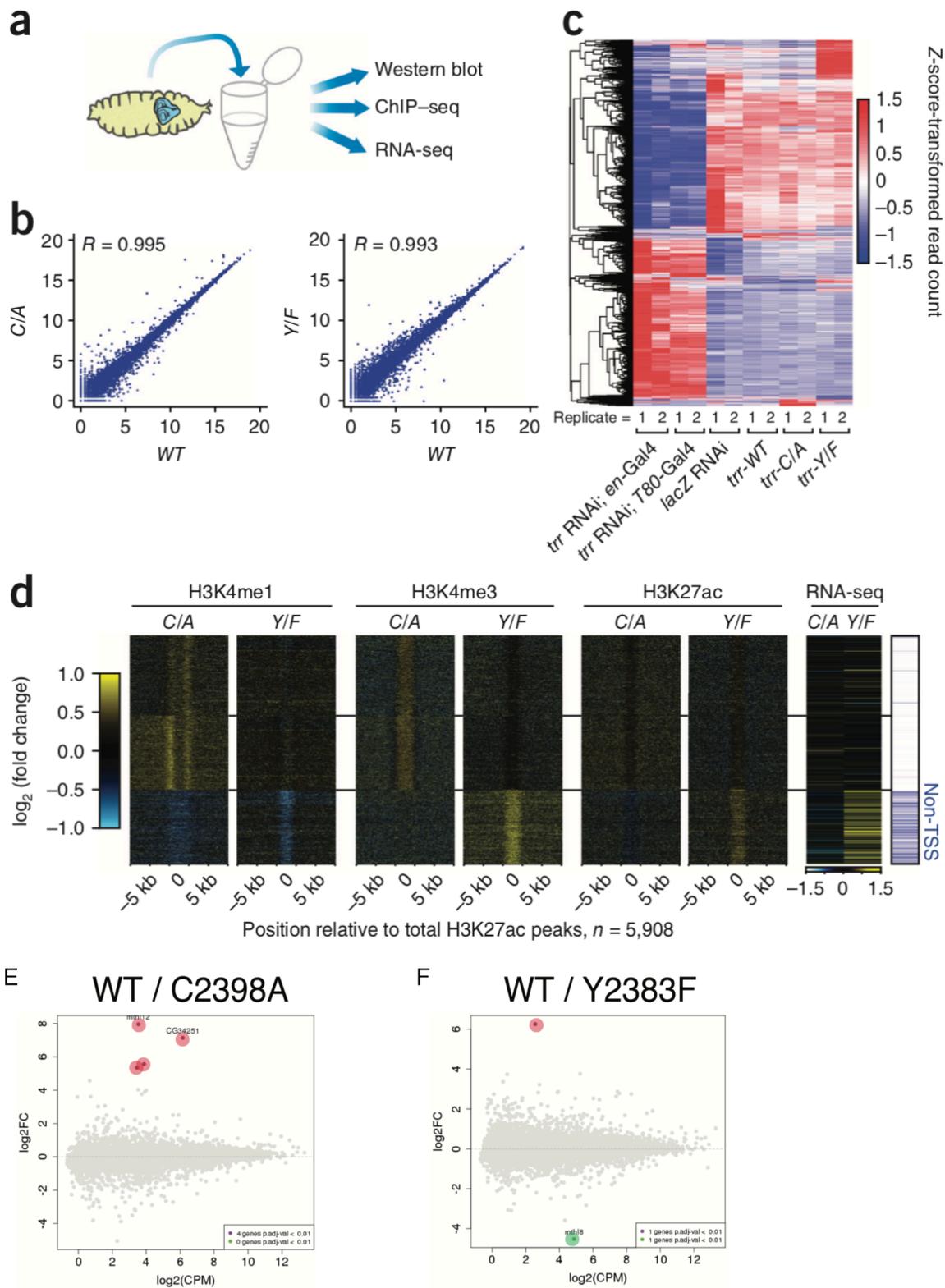


Figure 3.**Gene expression is only modestly affected by Trr-dependent enhancer methylation.**

(A) Graphical depiction of 3rd instar larval wing imaginal discs and experiments performed with these tissues. (B) Gene expression analysis shows >99% correlation (Pearson) between *trr*¹ flies rescued with *trr-WT* and flies rescued with either *trr-C/A* (inactive) or *trr-Y/F* (hyperactive). (C) *trr*-RNAi is expressed in wing discs using separate Gal4 drivers, *en* and *T80*, and gene expression profiles are compared with three *trr*¹-rescue lines. Hierarchical clustering of differentially expressed genes segregates *trr*-RNAi samples from *trr* catalytic mutants. Two independent biological replicates from each sample are shown in the heatmap displaying the total number of differentially expressed genes among these 6 samples. Note the similarities among *trr*¹-rescue lines versus *trr*-RNAi. (D) K-means clustering (k=3) of H3K4me1/3 and H3K27ac ChIP-seq data displayed as log₂ fold change heat maps centered on total H3K27ac peaks (+/- 5kb). Cluster 3 (bottom) is over-enriched for nonTSS sites (p=3.61 x 10⁻⁴⁰⁴, hypergeometric test) most affected by Trr catalytic mutations. RNA-seq log₂ fold changes show the effects of enhancer chromatin modifications on nearby gene expression. NonTSS sites are depicted as blue in the far-right panel, whereas TSS are white. Wing disc ChIP-seq experiments were performed once. (E) RNA-seq analysis from adult brains displayed as MA plots comparing *trr*-WT with *trr*-C/A. Note that only four genes are differentially expressed in *trr*-C/A. (F) RNA-seq analysis from adult brains displayed as MA plots comparing *trr*-WT with *trr*-Y/F. Note that only two genes are differentially expressed in *trr*-Y/F.

After detailed analysis we were able to uncover subtle phenotypes in the three *trr-rescue* lines. When maintained at the elevated temperature of 29°C, *trr-C/A* flies display an additional L3/L4 cross-vein, and this phenotype disappears when *trr-C/A* is expressed in a *trr*⁺ background (Figure 4A). Also, *trr-Y/F* females have much stronger pigmentation of their 7th abdominal segment, whereas the majority of *trr-C/A* females lack this pigmentation (Figure 4B). We also observe a mild bristle phenotype in the *trr-Y/F* line when maintained at 29°C. The occurrence of supernumerary thoracic macrochaetae (bristles) was significantly higher ($p = 1.19 \times 10^{-17}$, t-test with unequal variance) in the *trr-Y/F* than in *trr-C/A* flies ($p = 0.36$) (Figure 4D). Remarkably, this phenotype also manifests in mutants of the H3K4me3-demethylase, *lid*^{19,20}. Bulk levels of H3K4me3 are increased in *lid* mutants, similar to what we observe in *trr-Y/F* flies, suggesting this bristle phenotype results from increased H3K4me3, presumably through the Notch pathway. Together, these observations demonstrate a wide tolerance for differential H3K4-monomethylation at developmental enhancers in *Drosophila*, and suggest this modification is important for fine-tuning enhancer activity, especially under temperature stress.

In eukaryotes, the cohesin complex is essential for sister chromatid cohesion during mitosis and also plays a pivotal role in facilitating enhancer-promoter communication by bringing these elements into close physical proximity²¹⁻²⁵. Nipped-B is responsible for loading cohesin onto chromatin, and over-expression of this gene in *Drosophila* causes abdominal segmentation defects in adult flies²⁶. Interestingly, we detect abdominal segmentation defects in the *trr-Y/F* flies at low penetrance. If this is due to enhancer over-activation through hypermethylation, then Nipped-B overexpression in the *trr-Y/F* background should exacerbate this phenotype. Indeed, crossing *da-GAL4:UAS-Nipped-B* with *trr-Y/F* increases segmentation defect penetrance in a *trr*⁺ background by ~12-fold relative to UAS-Nipped-B alone (Figure 4C). By contrast, *trr-C/A* does not dramatically increase the segmentation defect penetrance of UAS-Nipped-B (Figure

Figure 4. Subtle phenotypes of *trr* catalytic mutants.

(A) Histograms quantitate increased frequency of additional L3/L4 cross-vein phenotype in *trr*¹;; *trr-C/A* versus +;; *trr-C/A* when grown at 29°C. The x-axis shows whether one or both wings display the additional cross-vein. 50 adult flies were scored for each genotype. Scale bar = 0.5mm. (B) Female 7th abdominal segment pigmentation is modulated in a Trr-methylation-dependent manner. Note, increased frequency of strong pigmentation in *trr-Y/F*, and decreased pigmentation in *trr-C/A* compared with *trr-W/T*. Red arrowhead designates 7th abdominal segment. Number of flies scored is listed for each genotype. Scale bar = 0.25mm. (C) *trr-Y/F* flies display subtle abdominal segmentation defects, which are exacerbated by Nipped-B over-expression in the *trr*⁺ background. Note the roughly 12-fold increase in penetrance when Nipped-B levels are increased in the *trr*⁺;;*trr-Y/F* background. Number of flies counted per genotype is: 267, 177, 223, 192, 113, 103, and 89, with respect to their order shown in the graph. Scale bar = 0.25mm. (D) *Trr-Y/F* adults have extranumerary thoracic macrochaetae when grown at higher temperatures. When grown at 29 °C, *trr-Y/F* flies exhibit extranumerary thoracic macrochaetae, as compared with either *trr-WT* or *trr-C/A*. Note that this phenotype is less penetrant in the wild-type *trr* background (right). Sixty flies were scored per genotype.

To test the conservation of enhancer H3K4me1 function between *Drosophila* Trr and mammalian MLL3 and MLL4 (encoded by *KMT2C* and *KMT2D*, respectively^{2,3,27}), we used CRISPR/Cas9 to generate several mESC clones containing deletions of both the MLL3 and MLL4 SET domains (referred to as MLL3/4-ΔSET) (Figures 5A). These mutant cells exhibit bulk reductions in H3K4me1 (Figures 5B). In contrast to a recent report claiming MLL4 protein stability is dependent on its methyltransferase activity²⁸, we observe no protein stability defects in our MLL3/4-ΔSET clones (Figure 5C). In addition, while this manuscript was under revision, another study in mESCs used single amino acid catalytic-inactivating MLL3/4 point mutations, and also observed no effect on MLL3/4 protein stability²⁹. We attribute these differences to the use of a potentially destabilizing triple-amino acid mutant versus the single point mutations or the deletion of the entire SET domain. Collectively, these results are in agreement that MLL3/4 regulates enhancer function largely through a methylation-independent mechanism²⁹.

By integrating our datasets with those of Wysocka and colleagues²⁹, we observe significant reductions in levels of H3K4me1/2 and H3K27ac at MLL3/4-bound enhancers, consistent with multiple ChIP-seq studies in MLL3/4 double-KO mESCs^{29,30} (Figures 5D and 5E). K-means clustering (k=2) of the ChIP-seq data reveals that sites of MLL3/4-dependent H3K4-methylation are weakly associated with reductions in nearby gene expression, similar to results shown in Figure 3D (Figure 5F). Interestingly, the effects on nearby transcription are more evident in MLL3/4-ΔSET versus MLL3/4-point mutants, but less severe than the double-KO cells, indicating the MLL3/4 point-mutants may retain some residual catalytic-activity (Figure 5F). Additionally, MLL3/4-ΔSET mESCs stain positive for alkaline phosphatase activity, whereas MLL3/4 double-KO cells do not³⁰, thus the consequences of disrupting MLL3/4 methyltransferase activity are likely less severe than removing the entire gene product (Figure

5G). These results agree with our findings in *Drosophila* and point to a vital methylase-independent function for the Trr/MLL3/MLL4 COMPASS family *in vivo*.

Figure 5

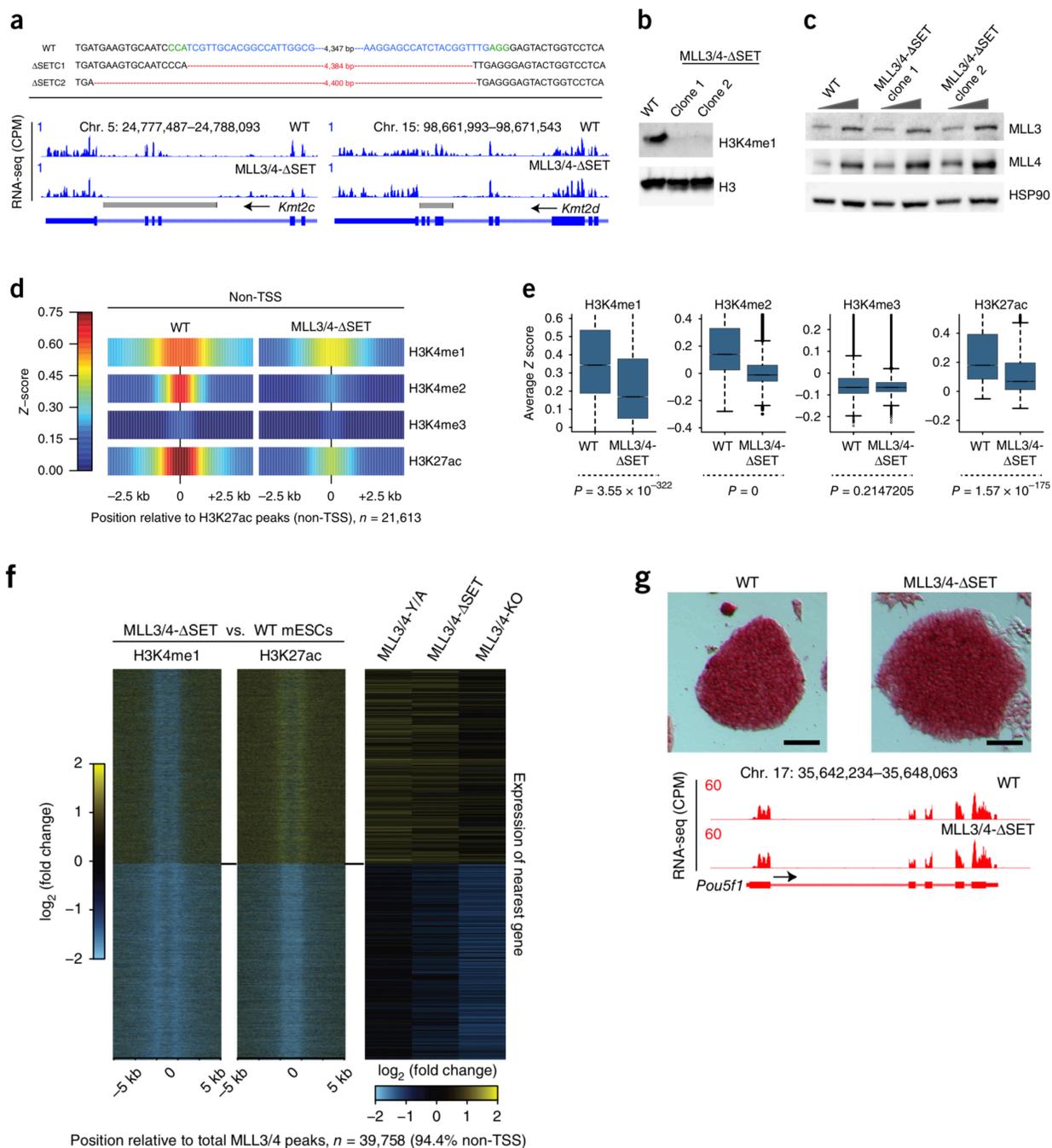


Figure 5.**MLL3/4-catalytic activity is not required for enhancer function in mouse ES cells.**

(A) DNA sequence and track examples from RNA-sequencing showing the CRISPR deleted regions to generate double MLL3/4- Δ SET homozygous mESC lines. Deletion of these three 3' exons of *KMT4C* and *KMT4D* create truncations shortly after the WDR5-interacting motif.

(B) Western blots showing decreased H3K4me1 in both MLL3/4- Δ SET clones. (C) Western blots showing our SET domain deletions do not reduce MLL3/4 protein stability. Image representative of at least two independent experiments. (D) Density bar-plots of Z-score transformed ChIP-seq reads are centered on nonTSS H3K27ac peaks (n=21613) and show decreased enhancer H3K4me1/2 and H3K27ac in MLL3/4- Δ SET cells. (E) Box plots quantifying changes in histone modifications displayed in (D). Upper/lower boundaries represent the interquartile range (IQR). Upper whiskers signify the third quartile (Q3) + 1.5*IQR. Lower whiskers signify Q1 - 1.5*IQR. Dots represent outliers. P-values were calculated by conducting t-tests. Results are representative of two independent experiments.

(F) K-means clustered heatmaps (k=2) centered on total MLL3/4 peaks from Wysocka and colleagues²⁹ (n=39758, 94.4% enriched for nonTSS) display log₂ fold-changes of H3K4me1 and H3K27ac in MLL3/4- Δ SET versus WT mESCs. The accompanying heatmap shows log₂ fold changes in nearby gene expression, comparing the published MLL3/4 point-mutation (MLL3/4-Y/A) and MLL3/4-KO²⁹ with our MLL3/4- Δ SET cells provides evidence that deletion of the SET domain is less consequential than a double-KO. Cluster #2 shows 831 significantly (adjusted p-value < 0.01) down-regulated genes in MLL3/4- Δ SET cells versus 3419 down-regulated in MLL3/4-KO cells. Average log₂ fold gene-expression changes between 2 independent replicates are shown. (G) MLL3/4- Δ SET mESCs retain their self-renewal capacity, express pluripotency factors such as Pou5f1, and stain positive for alkaline phosphatase activity, contrary to what is reported for MLL3/4-double-KO cells³⁰. Scale-bar = 100 microns.

Discussion

The concept that methylation of histone proteins can influence transcriptional regulation was proposed over 50 years ago; however, the mechanistic details of how several histone modifications affect transcription outcomes are still lacking. This is certainly the case in yeast, where deletion of Set1 abolishes all H3K4-methylation without significantly affecting gene expression or cell viability^{31,32}. While the role of Set1 in transcription remains disputed, a recent study shows Set1-dependent H3K4me2 near the *INO1* promoter is necessary for rapid induction of this gene following inositol withdraw³³. This represents a bona fide example of epigenetic memory as the H3K4me2 remains in the absence of *INO1* activation. However, it remains unclear as to why Set1 deposits H3K4-methylation at essentially all RNA Polymerase II transcribed genes, if only to provide this memory function at a few inducible genes. Similarly, mouse development and fertility were shown not to depend on MLL1 methyltransferase activity³⁴. In contrast to MLL1^{-/-} mice that die during embryogenesis, a homozygous MLL1^{ΔSET} mutation produced only minor homeotic transformations in otherwise healthy adult mice³⁴. These studies hint at a subtler role for H3K4-methylation in fine-tuning gene expression which appear secondary to the methylase-independent functions of the COMPASS family enzymes.

From yeast to humans, the COMPASS family has multiplied in number and diversified in function to create a division of labor among its different family members^{2,35,36}. *Drosophila* has proven to be an especially useful model for studying the COMPASS family of H3K4-methyltransferases as the three sets of mammalian paralogs (Set1a/b, MLL1/2, MLL3/4) are each represented by just one enzyme in fruit flies (dSet1, Trx, Trr), allowing us to deduce their basic biological functions without potential compensatory redundancies²⁷. In this study, we have investigated the functional significance of enhancer-marked H3K4me1, implemented by Trr, and were surprised to discover this specific chromatin modification is apparently

dispensable for fly development, homeostasis, and fertility. This was demonstrated by rescuing *Trr*-null lethality with a catalytic-dead allele (*trr-C/A*) in which enhancer H3K4-monomethylation is significantly reduced (Figures 2B-D). Also, by altering *Trr*'s product specificity with a catalytic-hyperactive mutation (*trr-Y/F*), we showed that enhancer-marked H3K4me1 is efficiently converted to H3K4me2/3, without affecting fly viability. However, once *trr-Y/F* flies were cultured at higher temperatures did we begin to observe minor defects, such as bristle duplications, which resemble phenotypes obtained when H3K4me3 levels are increased by *lid* loss-of-function mutations^{19,20}. These observations support a model in which *Trr*-dependent H3K4-methylation only functions in fine-tuning the activity of some enhancers, particularly under stress conditions, while *Trr* binding remains absolutely critical for enhancer function. It is intriguing how an increase of H3K4me2/3 levels at enhancers in *trr-Y/F* flies can phenocopy abdominal segmentation defects observed when *Nipped-B* is overexpressed. This genetic interaction suggests that enhancer activity can be “over-activated” by higher degrees of H3K4-methylation, similar to excessive cohesin deposition.

Although the morphological defects we observe are subtle, we cannot exclude the possibility that either of the *trr* catalytic mutants are defective for some undetermined biological process or sensitive to untested environmental factors. However, considering the fact that recessive *trr* NULL mutations result in embryonic lethality, it is clear that the functions for *Trr* critical to organismal development do not depend on its methyltransferase activity. Further experimentation will require the mapping of critical domains to understand how these portions of *Trr* work in regulating transcription. Certainly, one of these domains will interact with LPT (Lost PHDs of Trr), which corresponds to the N-terminal half of MLL3/MLL4 and is also shown to stably interact with *Trr*²⁷. LPT was recently demonstrated to be one of the strongest transcriptional activators, out of 812 factors tested, when targeted to a reporter gene in

combination with several other transcription factors³⁷. Importantly, the stability of both LPT and UTX is dependent on Trr, making it likely that one essential function of Trr is to recruit these additional activating/co-activating factors to enhancers².

In the last few years, increased attention has been paid to the mammalian *trr* homologs, MLL3/MLL4 (KMT2C and KMT2D, respectively) as both of these enzymes were recently found to be mutated in a diverse array of cancer types, and are among the most frequently mutated in human tumors^{14,38}. While disease-associated mutations of MLL3/MLL4 are scattered throughout the length of these genes, there is a slight enrichment for mutations to occur within sequences encoding the N-terminal PHD fingers of MLL3 and the C-terminal SET domain of MLL4³⁹. Even in the case of SET domain mutations, the role of MLL3/MLL4 methyltransferase activity in the pathogenesis of these diseases remains unclear. The SET domain contains not only the catalytic pocket, but also several binding surfaces necessary for interaction between COMPASS enzymes and their regulatory subunits: Wdr5, Rbbp5, Dpy30, and Ash2L⁴⁰⁻⁴². Thus, mutation of the SET domain itself has the potential to indirectly disrupt histone-methylation by destabilizing the entire protein complex. Indeed, one of the major challenges moving forward will be to deconvolute the structural versus catalytic functions of the SET domain among COMPASS family members. Intriguingly, the H3K27me2/3-demethylase, UTX (KDM6A), is also among the most highly mutated genes in human cancers, many of which overlap those associated with MLL3/MLL4 mutation^{14,38,43,44}. The fact that UTX also exists in a complex with MLL3 and MLL4 underscores the clinical importance of better understanding the integral functions served by these two COMPASS-like complexes⁴⁵.

Methods

Fly Stocks

Genomic *trr* rescue flies were generated using *pattB* plasmid for site-specific integration on 3R (89E11) and injections performed by BestGene (strain 9744). Transgenic flies were crossed to *trr*¹ and then made homozygous for the *trr*¹ allele and *trr-rescue* construct. UAS-Nipped-B was overexpressed using a *da*-Gal4 driver line. *T80*-Gal4 (1878), *en*-Gal4 (33557), UAS-*trr*-RNAi (29563) were purchased from Bloomington Drosophila Stock Center.

HMTase analysis of the reconstituted Trr/hWRAD complex *in vitro*

Trr complex was reconstituted by co-transfecting Sf9 cells with the cocktail of five baculoviruses expressing Trr-WinSET, hRBBP5, hASH2L, hWDR5, and hDPY30, respectively, and purified with M2 agarose resin (sigma-aldrich). All five components are N-terminally FLAG-tagged. HMTase assay was performed in 20µL of 50mM Tris-HCl pH8.8, 20mM KCl, 5mM MgCl₂, 0.5mM DTT at 37°C for 1hr with 1µg of recombinant histone H3 (NEB), 200µM of S-adenosyl-methionine, and near equal amounts of Trr + hWRAD.

Chromatin Immunoprecipitation (ChIP)

ChIP-seq experiments in adult brains were performed as follows. Approximately 10,000 adult flies (4-6 days after eclosion) were flash-frozen in a 50mL tube, vortexed to decapitate heads, and separated by passing through 710 µm and 425 µm sieves. Heads were cross-linked for 10 minutes while homogenizing with a 50mL dounce (loose pestle) in Buffer A1 (15mM HEPES pH 7.5, 15mM NaCl, 60mM KCl, 4mM MgCl₂, 0.5% Triton X-100, 0.5mM DTT) plus 2%

paraformaldehyde and protease inhibitor (Sigma P8340) added fresh. Glycine was added to final concentration of 225mM to quench fixative and the mixture was then passed through a 70 μ m cell-strainer to remove debris. Brain cells were pelleted for 5 minutes at 2000 x g, 4°C, and washed once with Orlando/Paro buffer (10mM Tris pH7.5, 10mM EDTA, 0.5mM EGTA, 0.25% Triton X-100, 0.5mM DTT, protease inhibitors) and pelleted again at 2000 x g. Final cell pellet was resuspended in sonication buffer (10mM Tris pH 8.0, 1mM EDTA, 0.1% SDS, protease inhibitor) and chromatin was fragmented to between 200-600 bp using a Covaris E220 bath sonicator. Chromatin was incubated with antibodies overnight at 4°C, immunoprecipitated the following day by incubating with Protein A/G agarose beads (SantaCruz), washed six times in RIPA buffer (25mM Tris pH7.5, 140mM NaCl, 1% Triton X-100, 1mM EDTA, 0.1% SDS, 0.1% Na-deoxycholate, 0.5mM DTT), and eluted (0.1M NaHCO₃, 1% SDS). After cross-link reversal and Proteinase K digestion overnight at 65°C, DNA was purified using Qiagen PCR purification spin-columns.

Larval wing imaginal disc ChIP-seq (~100 discs/ChIP) performed similar to previous reports⁴⁶, except chromatin was sheared using a Covaris E220 bath sonicator.

Qiagen RNeasy kits were used for all RNA-purification.

Antibodies

Antibodies recognizing H3K4me1, H3K4me2, H3K4me3, Trr, Lpt, Trx, dSet1, MLL3, and MLL4 were generated in the Shilatifard lab^{2,36} and anti-H3K27ac was purchased from CST (D5E4, Rb mAb #8173).

Plasmid Cloning

CRISPR sgRNA constructs were cloned into pX330. pX330-U6-Chimeric_BB-CBh-hSpCas9 was a gift from Feng Zhang (Addgene plasmid # 42230)⁴⁷. The following oligos were heat-denatured, annealed by gradual cooling and cloned into the BbsI site of pX330:

Generation of CRISPR mutations

Mouse v6.5 ES cells were cultured in 2i media plus LIF and electroporated with 25ug of each sgRNA construct and 10ug of CAG-EGFP-IRES-Puro (a kind gift from the Hitoshi Niwa lab). One day post electroporation, cells were selected for 24 hours with 1ug/ml Puromycin. Cells were allowed to recover for 4 days and then plated at low density (1,000 cells per 10cm dish) and allowed to form colonies. Single colonies were picked into 96-well plates and allowed to grow to confluence at which point plates were split in half for freezing and lysis. Cells were lysed in DNA extraction buffer (10mM Tris, pH 8.5, 50mM KCl, 1.5mM MgCl₂, 0.45% NP-40, 0.45% Tween-20, 0.5 mg/ml proteinase K) and incubated overnight at 55°C. Proteinase K was inactivated by heating to 95°C for 12 minutes and DNA was analyzed by PCR.

NGS Data Processing

RNA-seq and ChIP-seq samples were sequenced with the Illumina NextSeq technology, and output data were processed with the bcl2fastq software tool. Sequence quality was assessed using FastQC v 0.11.2 (Andrew 2010), and quality trimming was done using the FASTX toolkit. RNA-seq and ChIP-seq reads were aligned to the mm9 and dm3 genomes using TopHat v2.0.9 and Bowtie v0.12.9, respectively, and only uniquely mapped reads with a two-mismatch threshold were considered for downstream analysis. Gene annotations from Ensembl 67 were used for mouse cells, and gene annotations from Ensembl 70 were used for Drosophila cells.

Output bam files were converted into bigwig track files to display coverage throughout the genome (in RPM) using the GenomicRanges package⁴⁸.

RNA-seq Analysis

Gene count tables were used as input for edgeR 3.0.8⁴⁹. Genes with Benjamini-Hochburg adjusted p-values less than 0.01 were considered to be differentially expressed. Heatmaps displaying gene expression levels transformed into Z-scores were generated using the pheatmap R package, and the rows (genes) and/or columns (samples) in these heatmaps were subjected to unsupervised hierarchical clustering.

ChIP-seq Analysis

Peaks were called with the MACS v1.4.2 software⁵⁰ using default parameters. H3K27ac peaks were separated into TSS and nonTSS groups based on whether or not they overlapped regions within 500 bp of a TSS. Density bar plots were generated using unpublished perl and R scripts written by Yaping Lui, which incorporated some UCSC genome browser tools. For the density bar plots, mean RPM values for each sample were computed along the genome in 10-bp bins, input was subtracted, and ChIP-seq RPM values were transformed into Z-scores by subtracting the mean RPM value across the genome and dividing by the standard deviation of the genome-wide mean RPM value. Subsequently, these Z-scores were aligned to H3K27ac nonTSS and TSS peaks. Metaplots and heatmaps and were generated using ngsplot⁵¹. Metaplots show log fold changes relative to input, and the heatmaps show log fold changes relative to wild type. K-means clustering was also performed using ngsplot, and nearest-gene log fold changes in gene expression (from the RNA-seq edgeR output) corresponding to the clustered peaks in the heatmaps were determined using in house scripts and visualized with Java TreeView⁵².

Statistical Analysis

For statistical analyses, R and Microsoft Excel were used. Appropriate statistical tests were used for all data where a statistical analysis was reported. An F test was performed to determine whether different groups had the same variance or not. For ChIP-seq Z-score analysis, t tests were calculated between the area-under-the-curve values. P values <0.05 were considered to be statistically significant. All experiments were conducted in unblinded conditions.

Data Availability Statement

RNA-seq and ChIP-seq raw data are available in the Gene Expression Omnibus (GEO) database: GEO-GSE95781

References

- 1 Rickels, R. *et al.* Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat Genet* **49**, 1647-1653, doi:10.1038/ng.3965 (2017).
- 2 Herz, H. M. *et al.* Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & Development* **26**, 2604-2620, doi:10.1101/gad.201327.112 (2012).
- 3 Hu, D. *et al.* The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and Cellular Biology* **33**, 4745-4754, doi:10.1128/mcb.01181-13 (2013).
- 4 Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81**, 65-95, doi:10.1146/annurev-biochem-051710-134100 (2012).
- 5 Banerji, J., Rusconi, S. & Schaffner, W. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299-308 (1981).
- 6 Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166**, 358-368, doi:10.1016/j.cell.2016.05.025 (2016).
- 7 Levine, M., Cattoglio, C. & Tjian, R. Looping back to leap forward: transcription enters a new era. *Cell* **157**, 13-25, doi:10.1016/j.cell.2014.02.009 (2014).
- 8 Smith, E. & Shilatifard, A. Enhancer biology and enhanceropathies. *Nat Struct Mol Biol* **21**, 210-219, doi:10.1038/nsmb.2784 (2014).
- 9 Dorsett, D. Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr Opin Genet Dev* **9**, 505-514 (1999).
- 10 Creyghton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A* **107**, 21931-21936, doi:10.1073/pnas.1016071107 (2010).
- 11 Heintzman, N. D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics* **39**, 311-318, doi:10.1038/ng1966 (2007).
- 12 Lee, J. E. *et al.* H3K4 mono- and di-methyltransferase MLL4 is required for enhancer activation during cell differentiation. *Elife* **2**, e01503, doi:10.7554/eLife.01503 (2013).
- 13 Piunti, A. & Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**, aad9780, doi:10.1126/science.aad9780 (2016).
- 14 Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501, doi:10.1038/nature12912 (2014).
- 15 Sedkov, Y. *et al.* Molecular genetic analysis of the Drosophila trithorax-related gene which encodes a novel SET domain protein. *Mech Dev* **82**, 171-179 (1999).
- 16 Takahashi, Y. H. *et al.* Regulation of H3K4 trimethylation via Cps40 (Spp1) of COMPASS is monoubiquitination independent: implication for a Phe/Tyr switch by the catalytic domain of Set1. *Mol Cell Biol* **29**, 3478-3486, doi:10.1128/MCB.00013-09 (2009).

- 17 Collins, R. E. *et al.* In vitro and in vivo analyses of a Phe/Tyr switch controlling product specificity of histone lysine methyltransferases. *J Biol Chem* **280**, 5563-5570, doi:10.1074/jbc.M410483200 (2005).
- 18 Kanda, H., Nguyen, A., Chen, L., Okano, H. & Hariharan, I. K. The Drosophila ortholog of MLL3 and MLL4, trithorax related, functions as a negative regulator of tissue growth. *Mol Cell Biol* **33**, 1702-1710, doi:10.1128/MCB.01585-12 (2013).
- 19 Eissenberg, J. C. *et al.* The trithorax-group gene in Drosophila little imaginal discs encodes a trimethylated histone H3 Lys4 demethylase. *Nat Struct Mol Biol* **14**, 344-346, doi:10.1038/nsmb1217 (2007).
- 20 Gildea, J. J., Lopez, R. & Shearn, A. A screen for new trithorax group genes identified little imaginal discs, the Drosophila melanogaster homologue of human retinoblastoma binding protein 2. *Genetics* **156**, 645-663 (2000).
- 21 Kagey, M. H. *et al.* Mediator and cohesin connect gene expression and chromatin architecture. *Nature* **467**, 430-435, doi:10.1038/nature09380 (2010).
- 22 Michaelis, C., Ciosk, R. & Nasmyth, K. Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell* **91**, 35-45 (1997).
- 23 Phillips-Cremins, J. E. *et al.* Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281-1295, doi:10.1016/j.cell.2013.04.053 (2013).
- 24 Rollins, R. A., Korom, M., Aulner, N., Martens, A. & Dorsett, D. Drosophila nipped-B protein supports sister chromatid cohesion and opposes the stromalin/Scc3 cohesion factor to facilitate long-range activation of the cut gene. *Mol Cell Biol* **24**, 3100-3111 (2004).
- 25 Rollins, R. A., Morcillo, P. & Dorsett, D. Nipped-B, a Drosophila homologue of chromosomal adherins, participates in activation by remote enhancers in the cut and Ultrabithorax genes. *Genetics* **152**, 577-593 (1999).
- 26 Gause, M., Misulovin, Z., Bilyeu, A. & Dorsett, D. Dosage-sensitive regulation of cohesin chromosome binding and dynamics by Nipped-B, Pds5, and Wapl. *Mol Cell Biol* **30**, 4940-4951, doi:10.1128/MCB.00642-10 (2010).
- 27 Mohan, M. *et al.* The COMPASS Family of H3K4 Methylases in Drosophila. *Molecular and Cellular Biology* **31**, 4310-4318, doi:10.1128/mcb.06092-11 (2011).
- 28 Jang, Y., Wang, C., Zhuang, L., Liu, C. & Ge, K. H3K4 Methyltransferase Activity Is Required for MLL4 Protein Stability. *J Mol Biol*, doi:10.1016/j.jmb.2016.12.016 (2016).
- 29 Dorigi K M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular Cell* (2017).
- 30 Wang, C. *et al.* Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc Natl Acad Sci U S A* **113**, 11871-11876, doi:10.1073/pnas.1606857113 (2016).
- 31 Miller, T. *et al.* COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. *Proc Natl Acad Sci U S A* **98**, 12902-12907, doi:10.1073/pnas.231473398 (2001).

- 32 Nislow, C., Ray, E. & Pillus, L. SET1, a yeast member of the trithorax family, functions in transcriptional silencing and diverse cellular processes. *Mol Biol Cell* **8**, 2421-2436 (1997).
- 33 D'Urso, A. *et al.* Set1/COMPASS and Mediator are repurposed to promote epigenetic transcriptional memory. *Elife* **5**, doi:10.7554/eLife.16691 (2016).
- 34 Terranova, R., Agherbi, H., Boned, A., Meresse, S. & Djabali, M. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A* **103**, 6629-6634, doi:10.1073/pnas.0507425103 (2006).
- 35 Hallson, G. *et al.* dSet1 is the main H3K4 di- and tri-methyltransferase throughout Drosophila development. *Genetics* **190**, 91-100, doi:10.1534/genetics.111.135863 (2012).
- 36 Rickels, R. *et al.* An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS. *Mol Cell* **63**, 318-328, doi:10.1016/j.molcel.2016.06.018 (2016).
- 37 Stampfel, G. *et al.* Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**, 147-151, doi:10.1038/nature15545 (2015).
- 38 Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-339, doi:10.1038/nature12634 (2013).
- 39 COSMIC.
- 40 Li, Y. *et al.* Structural basis for activity regulation of MLL family methyltransferases. *Nature* **530**, 447-452, doi:10.1038/nature16952 (2016).
- 41 Takahashi, Y. H. *et al.* Structural analysis of the core COMPASS family of histone H3K4 methylases from yeast to human. *Proc Natl Acad Sci U S A* **108**, 20526-20531, doi:10.1073/pnas.1109360108 (2011).
- 42 Herz, H. M., Garruss, A. & Shilatifard, A. SET for life: biochemical activities and biological functions of SET domain-containing proteins. *Trends Biochem Sci* **38**, 621-639, doi:10.1016/j.tibs.2013.09.004 (2013).
- 43 Morgan, M. A. & Shilatifard, A. Chromatin signatures of cancer. *Genes Dev* **29**, 238-249, doi:10.1101/gad.255182.114 (2015).
- 44 Ntziachristos, P. *et al.* Contrasting roles of histone 3 lysine 27 demethylases in acute lymphoblastic leukaemia. *Nature* **514**, 513-517, doi:10.1038/nature13605 (2014).
- 45 Issaeva, I. *et al.* Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol* **27**, 1889-1903, doi:10.1128/MCB.01506-06 (2007).
- 46 Papp, B. & Muller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev* **20**, 2041-2054, doi:10.1101/gad.388706 (2006).
- 47 Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823, doi:10.1126/science.1231143 (2013).
- 48 Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118, doi:10.1371/journal.pcbi.1003118 (2013).

- 49 Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140, doi:10.1093/bioinformatics/btp616 (2009).
- 50 Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
- 51 Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284, doi:10.1186/1471-2164-15-284 (2014).
- 52 Saldanha, A. J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-3248, doi:10.1093/bioinformatics/bth349 (2004).

Methods-only References

30. Papp, B. & Muller, J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev* **20**, 2041-54 (2006).
31. Rickels, R. *et al.* An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS. *Mol Cell* **63**, 318-28 (2016).
32. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-23 (2013).
33. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput Biol* **9**, e1003118 (2013).
34. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-140 (2009).
35. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).
36. Shen, L., Shao, N., Liu, X. & Nestler, E. ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. *BMC Genomics* **15**, 284 (2014).
37. Saldanha, A.J. Java Treeview--extensible visualization of microarray data. *Bioinformatics* **20**, 3246-8 (2004).

Chapter 3:

Identifying the Minimal Requirements for UTX Stability in *Drosophila* and Humans

Abstract

We recently reported catalytic-inactivating mutations of Trr, the enhancer-associated COMPASS H3K4-monomethylase, is compatible with proper development and homeostasis in *Drosophila*. Similar experiments in mouse embryonic stem cells also demonstrate non-enzymatic roles for MLL3/4, as deletion of the catalytic SET domain causes only minor changes in gene expression compared with MLL3/4 whole-gene deletions. Here, we describe a minimal portion of Trr capable of rescuing Trr-null lethality and we show this region is able to bind and stabilize Utx *in vivo*. We mapped the corresponding homologous sequence in both human MLL3 and MLL4 to ~80 residue Utx-Stabilization-Domain (USD) that binds and promotes UTX stability in the absence of MLL3/4. Interestingly, nuclear UTX stability is greatly enhanced in the presence of USD fused with the N-terminal MLL4 HMG-box, which is the domain of Lpt previously shown to co-purify with Trr in *Drosophila*. Tumor sequencing studies highlight MLL3 (*KMT2C*), MLL4 (*KMT2D*), and UTX (*KDM6A*) as genes highly mutated across a wide variety of malignancies. Our results point to COMPASS-dependent stabilization of UTX as an essential non-catalytic function Trr/MLL3/MLL4, and suggests stabilizing UTX could be a therapeutic strategy in treating cancers with MLL3/4 loss-of-function mutations.

Introduction

Trithorax-related (*trr*), MLL3 (*KMT2C*) and MLL4 (*KMT2D*) represent the 'branch' of COMPASS family lysine methyltransferases responsible for catalyzing histone 3 lysine 4 monomethylation (H3K4me1) at enhancer chromatin in *Drosophila* and mammals, respectively^{1,2}. These enzymes assemble into complexes with components common to all COMPASS members, as well as complex-specific subunits that enable unique functions, and this subunit composition is conserved from *Drosophila* to humans³. In particular, the H3K27-demethylase, UTX (*KDM6A*), is specific to Trr/MLL3/MLL4 complexes and is believed to function in the process of enhancer activation by facilitating acetylation of H3K27 by CBP/p300^{1,4-7}. Indeed, depletion of either Utx or Trr in *Drosophila* results in decreased levels of H3K27ac, as well as H3K4me1, underscoring the functional interdependence between Trr and Utx in modifying chromatin structure *in vivo*¹. Trr is critical for maintaining stability of Utx, as well as Lpt, a protein homologous to the N-terminal half of MLL3/4, which is expressed from a separate gene but complexes with Trr/COMPASS in the nucleus^{1,7}. Thus, the function and stability of these two enzymes are intimately entwined, and this relationship is conserved from flies to mammals. Genetic links between MLL4 and UTX were initially identified in a rare pediatric congenital disorder called Kabuki Syndrome, where the majority of patients show mutations in either *KMT2D* or *KDM6A*⁸⁻¹¹. More recently, *KMT2C*, *KMT2D*, and *KDM6A* are all found to be frequently mutated across a broad swath of human malignancies, highlighting a connection between regulators of enhancer histone modification and cancer pathogenesis¹²⁻¹⁸.

Results

We recently demonstrated that loss of H3K4me1, or conversion to H3K4me2/3, at enhancers is compatible with proper development and adult homeostasis in *Drosophila melanogaster*¹⁹.

Using complementation assays in a lethal *Trr*-null background, we rescued viability by expressing *Trr* with either catalytic-defective (C2398A) or catalytic-hyperactive (Y2383F) point mutations, each producing unique yet subtle phenotypes only observable under temperature-stress. Consistent with data showing H3K4me1 is largely dispensable for enhancer function *in vivo*, our group and others also reported similar catalytic-inactivating mutations to MLL3/4 only minimally effects gene expression compared to MLL3/4-null deletions, which cause down-regulation of thousands of distal target genes in mouse embryonic stem cells (mESCs)^{19,20}.

Measurement of chromatin accessibility in mESCs by ATAC-seq showed reduced enhancer accessibility in the absence of MLL3/4, with no change in the catalytic-dead mutants.

Furthermore, enhancer-associated RNA (eRNA) synthesis at MLL3/4-bound enhancers was diminished only in the double-knockout cell line, concomitant with increased RNA polymerase II pausing at associated gene promoters²⁰. This is in agreement with a growing body of research suggesting enhancer-mediated gene activation functions by stimulating transcription elongation^{21,22}. Thus, localization of MLL3/4, but not its catalytic-activity, is required to maintain open chromatin at enhancers, which indicates a catalytic-independent function of *Trr*/MLL3/MLL4 is critical to its role in development.

Identification of an essential *Trr* domain sufficient to rescue *Trr*-Null associated lethality

In search of this catalytic-independent function, we returned to *Drosophila* complementation assays to identify which domain of *Trr* is required to rescue viability in the *trr*¹ background.

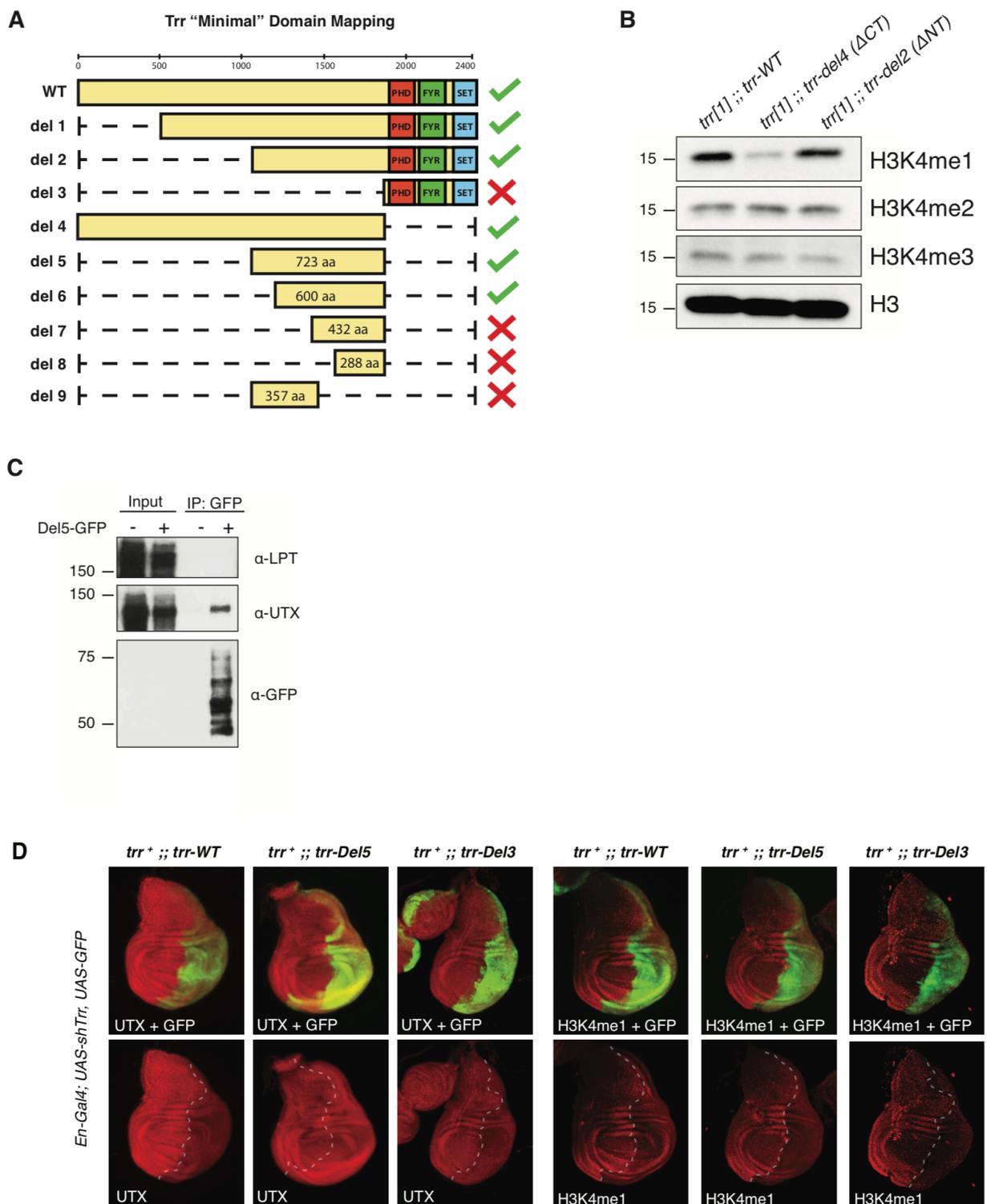
We created a series of *Trr* domain-deletions constructs, each integrated into the same genomic site, to generate a suite of transgenic fly lines to assay for complementation. Males from each

line were crossed with *trr*¹ females and the progeny were assessed to determine whether or not individual domain-deletions could rescue viability, as previously described¹⁹. While our previous studies demonstrate the catalytic activity of Trr is not required for viability, we were surprised to find that deletion of the entire C-terminus (*del4*), which includes the SET domain, was able to rescue lethality (Figure 1A). Consistent with our previous reports, bulk reductions in H3K4me1, as measured by western blot, have no obvious effect on *Drosophila* development¹⁹. Interestingly, when reared at higher temperatures (29°C), Trr-del4 flies phenocopy the catalytic-dead Trr-C2398A mutation by producing ectopic L3/L4 wing cross-veins, while N-terminal deletions do not (Figure 1F and data not shown). This supports our theory that H3K4me1 is important in fine-tuning enhancer-mediated gene expression, especially under temperature stress. While N-terminal deletions, such as Trr-del2, readily rescue viability, we found this mutation dramatically reduces female fertility in the *trr*¹ background (data not shown), which will be further described in a forth-coming publication. Ultimately, our genetic complementation strategy successfully identifies a 600 amino acid region of unknown function, located roughly in the middle of Trr, that rescues Trr-null lethality (Figure 1A).

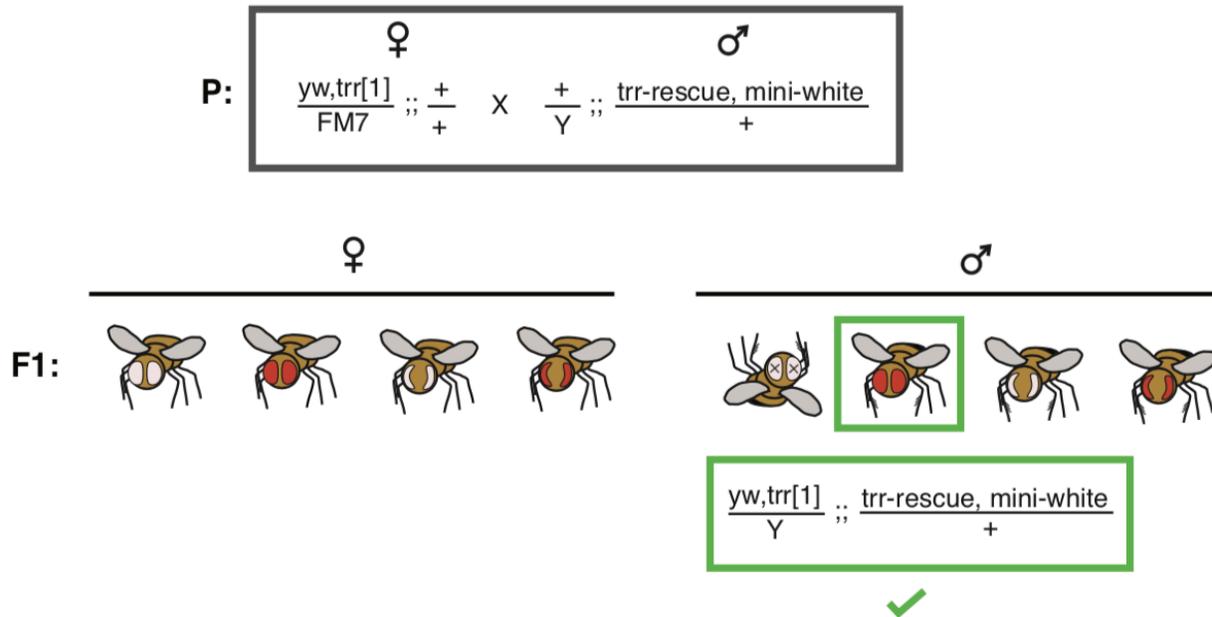
To study the biochemical function of this region, we expressed a GFP-tagged peptide corresponding to the del5 rescue fragment in *Drosophila* S2 cells, and performed protein purification to identify interacting partners. The protein products of two essential genes, *Lpt* and *Utx*, are known to form a complex with Trr, however, only Utx was confirmed by western blot to interact with the Trr-del5 fragment (Figure 1C). This finding is consistent with a previous report that showed, in human cells, UTX interacts with a 489aa portion near the C-terminus of MLL4¹⁴. Our lab and others have previously reported that Trr and MLL4 are responsible for stabilizing Utx protein in flies and mammals, respectively^{1,20}. To test if this region of Trr is sufficient to stabilize Utx *in vivo*, we expressed either full-length Trr-WT, Trr-del3, or Trr-del5 in a *trr*^{WT}; *en-*

Gal4>Trr-RNAi background, and then performed immunofluorescence microscopy to assess the effect on Utx stability in wing imaginal discs. The En-Gal4 system drives Trr-RNAi expression in the posterior half of imaginal discs, providing an internal control for comparison across all three genotypes. Importantly, both the endogenous *ttr* and Trr-WT are sensitive to Trr-RNAi, while Trr-del3 and -del5 are insensitive. In all three genotypes, knockdown of endogenous *ttr* is evident by reduced H3K4me1, as previously published¹. Likewise, Utx levels are diminished in the Trr-WT and Trr-del3, but unaffected in Trr-del5 (Figure 1D). Taken together, our findings suggest that stabilization of Utx is an essential catalytic-independent function of Trr.

Figure 1



E



F

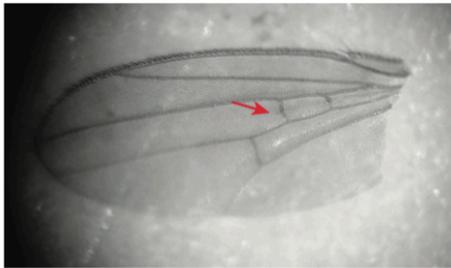


Figure 1.

A UTX-interacting region of Trr is sufficient to rescue Trr-null viability.

(A) The diagram depicts full-length Trr coding sequence and the various deletion constructs assayed for complementation with the *trr*¹ lethal allele. Green check marks designate which constructs rescue *trr*¹ viability according to the complementation assay illustrated in Figure 1E.

(B) Western blots for H3K4-methylation confirm bulk H3K4me1 decrease in the *del4* rescue line, which lack the catalytic SET domain. (C) Immunoprecipitation of the Del5-GFP fragment in *Drosophila* S2 cells co-purifies UTX, but not LPT. (D) Immunofluorescence microscopy in larval

wing imaginal discs confirms the “minimal” Del5 fragment is sufficient to stabilize UTX in the absence of full-length TRR. Endogenous TRR is depleted by RNAi in the posterior compartment (GFP labeled), leading to reductions in H3K4me1. However, UTX levels are not affected upon expression of Del5, which is insensitive to Trr-RNAi. (E) Cartoon depiction of the *trr*¹ complementation assay. Virgin females carrying *trr*¹/FM7 are crossed to males carrying the *trr*-rescue, *mini-white* transgene on the third chromosome. Presence of red-colored, non-bar eyes in the male F1 progeny indicates the transgene was able to rescue *trr*¹-associated lethality. (F) *trr*¹;*trr-Del4* flies reared at 29°C display ectopic L3/L4 cross-vein phenotypes identical to catalytic-dead *trr-C/A* flies, as previously reported.

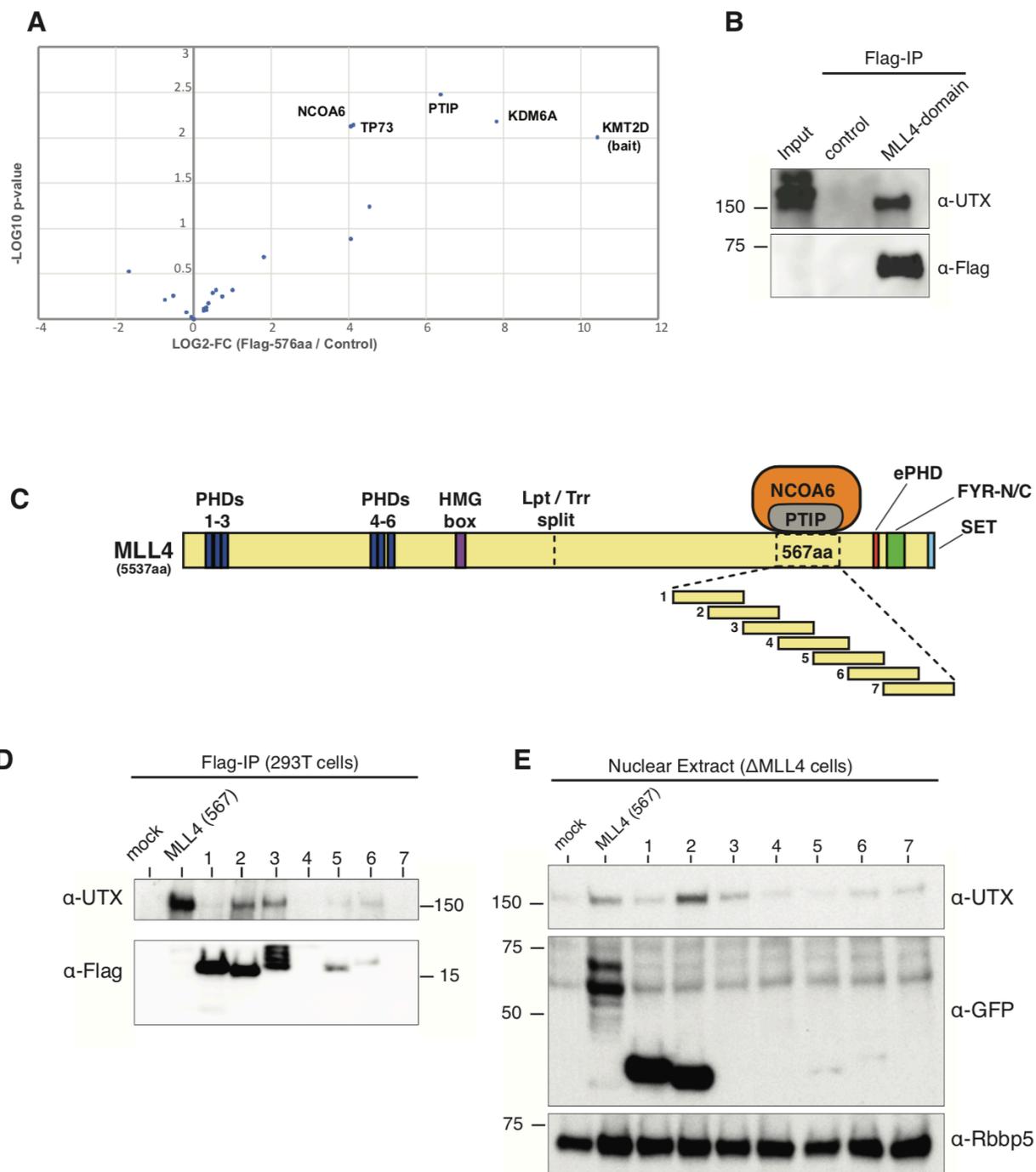
Identification of a conserved MLL4-UTX binding domain

To apply our findings from *Drosophila* in a mammalian system, we began by expressing a Flag-tagged homologous 567aa region of MLL4 in 293Trex cells, followed by protein purification to identify interacting factors. Mass-spectrometry from three independent experiments identified UTX (*KDM6A*) as the most significantly enriched protein interaction, compared with an IgG control, followed by PTIP and NCOA6, which are also known components of MLL3/4 COMPASS-like complexes (Figure 2A). TP73 was also enriched similar to NCOA6, but, to our knowledge, has yet to be reported as a biochemical interactor with either UTX or MLL4. We confirmed the MLL4 567aa region co-IPs with UTX by western blot (Figure 2B). It is important to note that, just as we did not detect Lpt in the Trr-del5 pull-down, we also do not detect endogenous MLL4 peptides in our mass-spectrometry. We performed size-exclusion chromatography with nuclear extracts from doxycycline-inducible 293Trex cells expressing a GFP-tagged MLL4 567aa fragment and found the vast majority of UTX co-elutes with MLL4 in uninduced cells. However, in the presence of the MLL4 domain, UTX co-elutes in later fractions

where GFP-tag is detected (Figure 2F). These results indicate the MLL4-UTX binding domain is not incorporated into the endogenous MLL4 complex, but instead competes for UTX binding.

While an approximate location of the MLL4-UTX binding domain was previously reported¹⁴, we began narrowing down the interaction surface to better understand how MLL4 stabilizes UTX *in vivo*. This was accomplished by creating a series of seven constructs that tile across the 567aa MLL4 domain, such that each fragment of ~150aa overlaps its neighbors by ~50% (Figure 2C). The Flag-tagged series was expressed in 293T cells, and co-IP/western blot experiments show fragments #2 and #3 both pull-down UTX (Figure 2D). Using CRISPR/Cas9 to delete the MLL4 promoter in HCT116 cells, we generated a Δ MLL4 cell line that display significantly lower levels of UTX compared to parental HCT116 (Figure 2G). After expressing the GFP-tagged series in this cell line, we observe fragment #2 provides some stabilizing effect on endogenous UTX levels (Figure 2E).

Figure 2



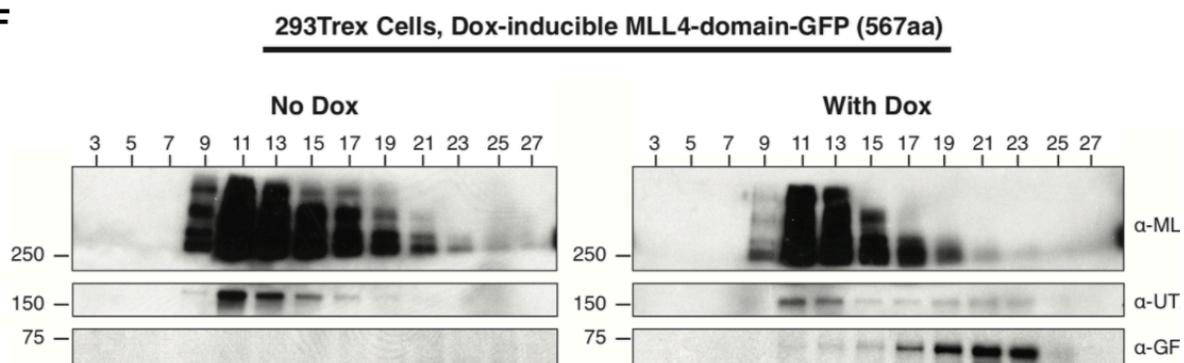
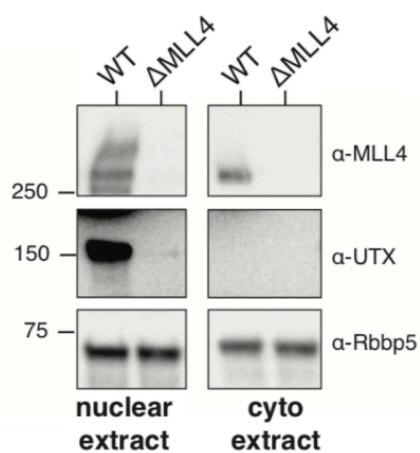
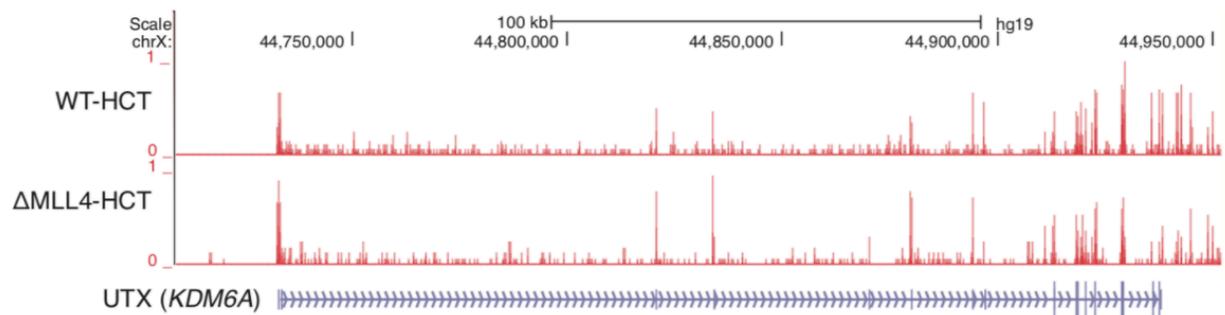
F**G****H**

Figure 2.**MLL4/UTX-binding interactions are important for UTX protein stability.**

(A) Volcano plot of mass-spectrometry analysis reveals UTX is the most significantly enriched factor in the MLL4-flag (567aa) purification, among other known MLL4/COMPASS subunits. Common contaminants [CRAPome] were filtered out and remaining hits are plotted as $-\log_{10}$ p-value (T-test) versus \log_2 fold-change. (B) Western blots confirm UTX interaction with MLL4-flag (567aa). (C) Diagram of the full-length MLL4 protein structure including annotated domains and approximate location of Lpt/Trr gene split. To fine-map the interacting region, a series of constructs express overlapping fragments tiled across the UTX-interaction domain. (D) Flag-tagged fragments #1-7 were transfected in 293T cells and purified by immunoprecipitation. UTX interaction is most evident with fragments #2 and #3. (E) Fragments #1-7 transfected in MLL4-KO HCT116 cells in which UTX levels are diminished due to loss of stabilizing interaction with MLL4. Note, the full-length 567aa domain, as well as fragment #2, provide some stabilizing effect on UTX. (F) Doxycycline-inducible 293Trex cells expressing the GFP-MLL4-UTX-interacting domain (567aa) were fractionated by size-exclusion chromatography. UTX co-elutes with endogenous MLL4 complex in cells without dox; however, induced expression of the MLL4-GFP fragment competes for UTX binding, as evidenced by co-elution in later fractions. (G) Western blots comparing MLL4 and UTX protein levels in WT versus MLL4-KO HCT116 cells, in which UTX levels are diminished in the absence of MLL3 or MLL4. (H) RNA-sequencing track example shows UTX (*KDM6A*) mRNA levels are unchanged in the absence of MLL4.

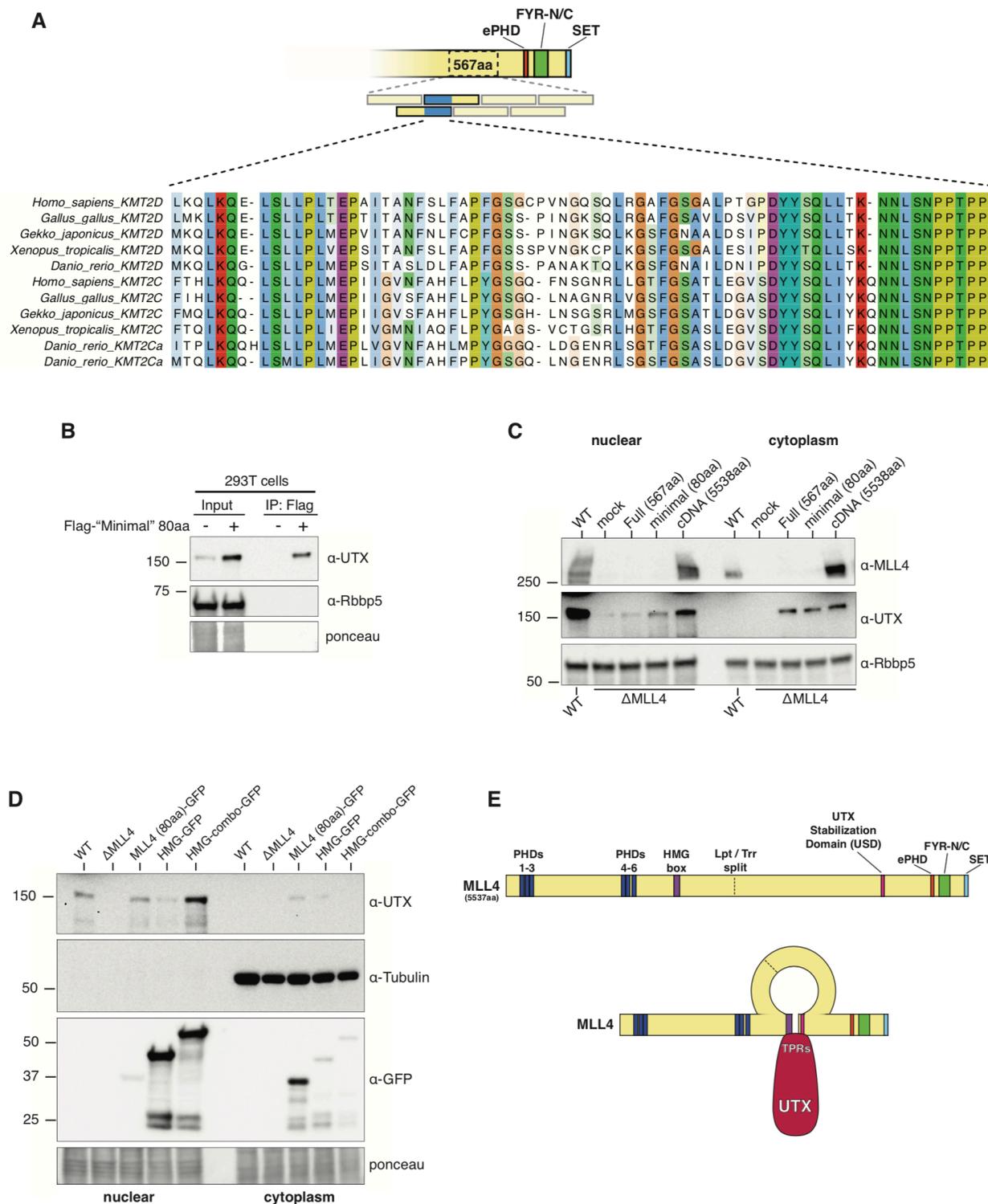
The MLL4 HMG-box domain is important for UTX stability

Based on those results, we focused our attention on a portion that contains the overlapping sequence between fragments #2 and #3. This “minimal” 80aa region is conserved in vertebrates and sufficient to bind, as well as stabilize, UTX in 293T cells (Figures 3A and 3B). However, when transiently transfected into Δ MLL4 cells, neither the original 567aa or “minimal” 80aa fragment were able to fully stabilize UTX to the extent observed with a full-length MLL4 cDNA, which achieves UTX levels similar to WT-HCT116 (Figure 3C). This suggests an additional MLL4 domain is also required to provide full UTX stability in the nucleus and, based on combined genetic and biochemical data, we hypothesized that the MLL4 HMG-box could be important for this function. First, our genetic complementation assays in *Drosophila* were always conducted in an *lpt*^{WT} background, indicating the N-terminal half of MLL4 could play a role in maintaining UTX stability. Second, a previous study in *Drosophila* showed that deletion of the Lpt HMG-box disrupts Lpt’s ability to pull-down Trr, although direct binding was never demonstrated²³. Third, we have used CRISPR in HCT116 to delete the N-terminus from MLL4, which contains the PHD domains, while leaving the HMG-box intact. Bulk UTX levels are not disrupted in these cells, whereas deleting the entire C-terminus, including the UTX-binding domain, results in strong depletion of UTX from the nucleus (Figure 3F). To test this hypothesis in Δ MLL4 cells, we transiently expressed a GFP-tagged MLL4 HMG-box domain alone, or fused with the 80aa UTX-binding domain, and observed a dramatic increase in nuclear UTX stability with the ‘HMG-combo’, comparable to WT levels (Figure 3D). Surprisingly, the HMG-box alone had a small effect on promoting UTX stability, suggesting this domain might also contact UTX directly (Figure 3D). A similar MLL3 HMG-combo fusion with the MLL3 HMG-box + 63aa UTX-binding domain also stabilized UTX similar to WT levels (Figure 3G). Thus far, all of our constructs used for protein expression in cell culture have necessarily included an SV40 nuclear localization signal (NLS); however, we found inclusion of the HMG-box in the fusion protein

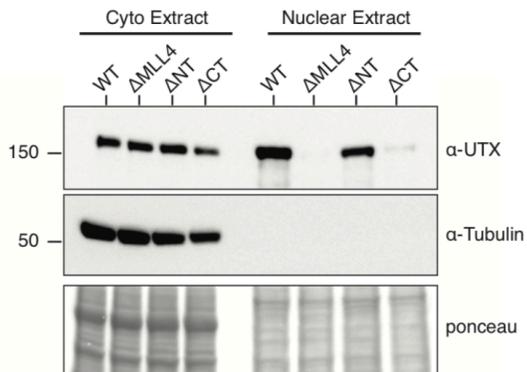
obviates the need for an NLS, indicating this domain might be important for MLL4 translocation into the nucleus (Figure 3G).

To better understand the biochemical interactions between MLL3/4 and UTX, we determined the approximate location of UTX's interaction with MLL4. Flag-tagged UTX was divided into three overlapping segments: NT, mid, CT, and IP/western analysis revealed the NT portion pulled down endogenous MLL4 (Figure 3H). This region of UTX contains at least six TPR domains, a structural motif known to mediate protein-protein interactions. Likewise, Flag-IP after co-transfection with Flag-UTX-NT and MLL4-HMG-combo-GFP demonstrates these regions also interact (Figure 3H). Taken together, we propose a model in which the MLL4 HMG-box and UTX-binding domain come together to stabilize UTX (Figure 3E), which is consistent with published genetic and biochemical data from *Drosophila*^{1,23}.

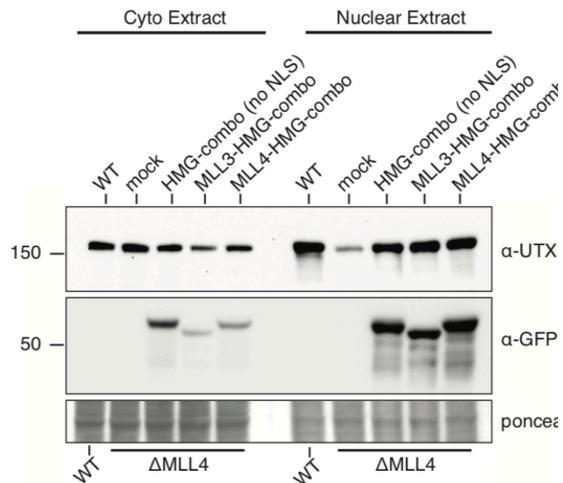
Figure 3



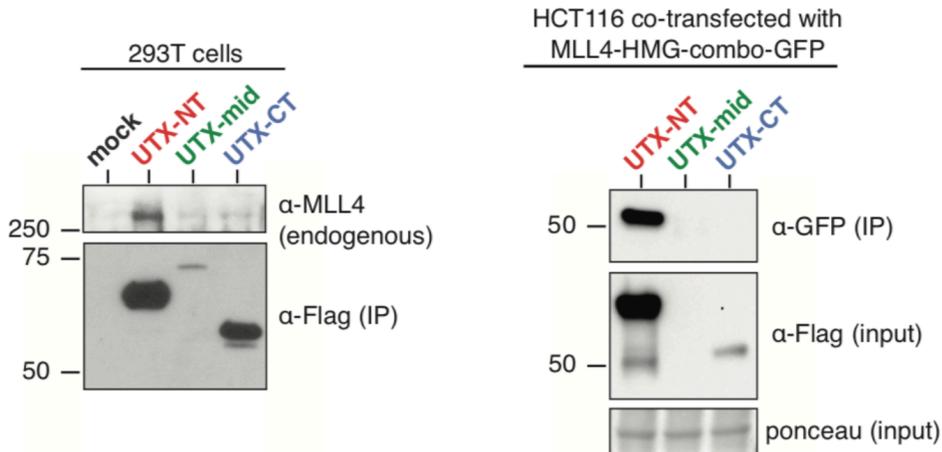
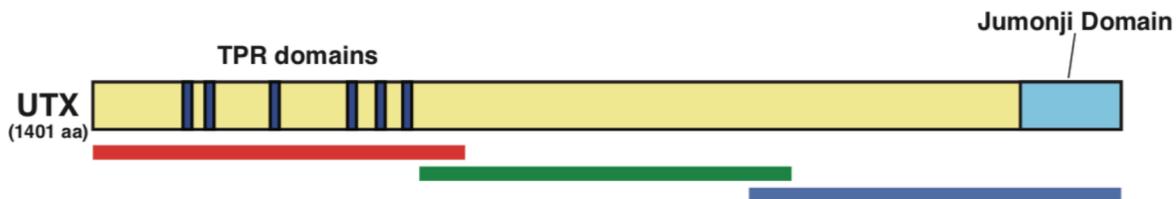
F



G



H



“minimal” 80aa domain fused with the MLL4 HMG-box domain (HMG-combo) fully restores nuclear UTX levels. Note the HMG-box alone has a minor stabilizing effect on UTX. (E) A diagram of the full-length MLL4 protein including annotated domains with approximate location of the Lpt/Trr gene split. Genetic and biochemical evidence suggests UTX is stabilized through intramolecular interactions between the MLL4 HMG-box and UTX-interacting domain. (F) Western blots comparing nuclear UTX levels in WT, Δ MLL4, MLL4- Δ NT, and MLL4- Δ CT HCT116 cells. (G) Western blots show the MLL4-HMG-combo expression construct does not require an exogenous NLS to fully restore nuclear UTX levels in Δ MLL4 cells. Also, an MLL3-HMG-combo restores UTX levels similar to the MLL4-HMG-combo. (H) Immunoprecipitation experiments reveal UTX most likely interacts with MLL3/4 through its N-terminal TPR domains. Flag-tagged constructs were created to express overlapping NT, mid, and CT fragments of UTX for transfection and immunoprecipitation. The NT-fragment pulls down endogenous MLL4 in 293T cells, or MLL4-HMG-combo when co-transfected in HCT116 cells. (I) Alignments of the same peptide sequence shown in Figure 3A also including five *Drosophila* species and multiple invertebrate species for improved alignment. Note the interrupting sequences present in *Drosophila*, which are predicted to form extended coils by HHPred secondary sequence prediction software (<https://toolkit.tuebingen.mpg.de/#/tools/hhpred>).

Numerous studies now demonstrate tumor suppressor functions for both MLL3/4 and UTX in various tissue contexts^{15,17,18,24-26}; however, questions remain as to how these enzymes function in that capacity, and how particular mutations disrupt those functions. It is also unclear to what extent MLL3/4's tumor suppressive function depends on UTX. To test whether or not stabilization of UTX is sufficient to slow cell proliferation in the absence of MLL3/4, we modified our Δ MLL4 HCT116 cell line (already mutated for MLL3) to include a doxycycline-inducible GFP-MLL4-HMG-combo transgene, and demonstrate robust UTX nuclear stabilization

approximately 3 days following doxycycline treatment (Figure 4A). After 7 days of doxycycline treatment, we observe a ~25% reduction in cell growth compared with no doxycycline controls. We also compared the parental Δ MLL4 cells +/- doxycycline, and do not observe a significant difference in their proliferation (Figure 4B).

Figure 4

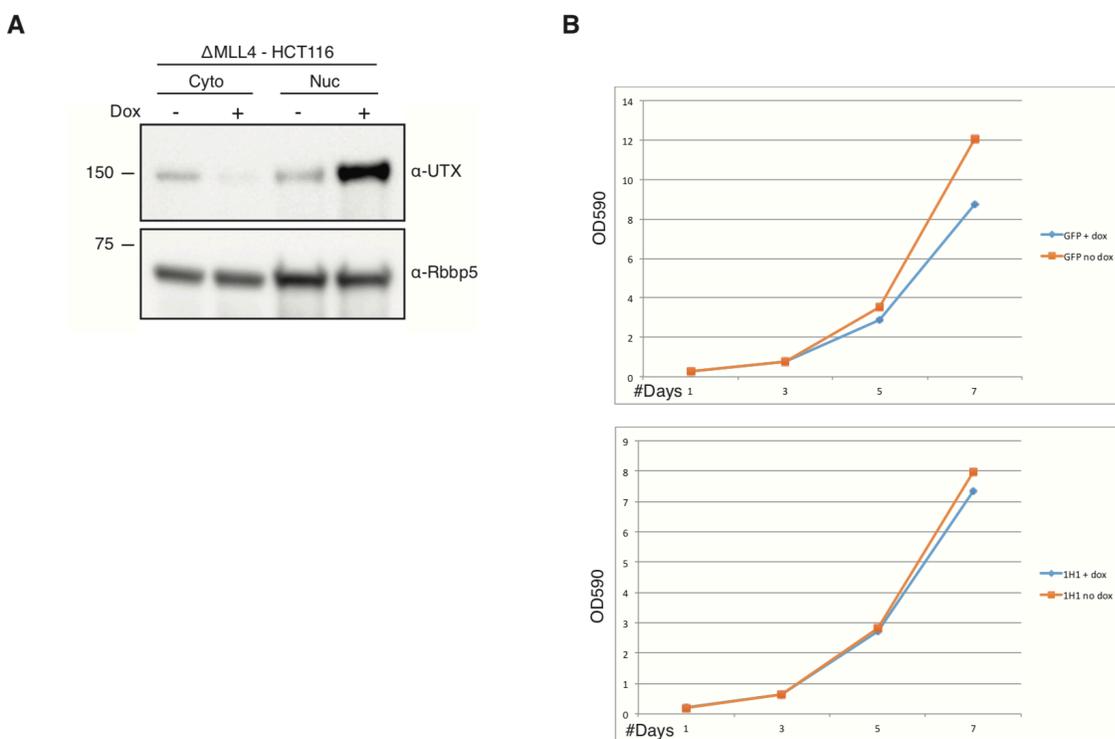


Figure 4.

Stabilization of UTX can slow proliferation in a Δ MLL3/4 cell culture model.

(A) Doxycycline-induced expression of MLL4-HMG-combo-GFP transgene in HCT116- Δ MLL3/4 cells is able to restore UTX stability within 72 hours of induction. (B) Cell growth assays +/- doxycycline demonstrate HCT116- Δ MLL3/4 cells proliferate slower when UTX protein levels are stabilized, compared with Δ MLL3/4 cells lacking the transgene.

Discussion

In the last few years, loss-of-function mutations of *KMT2C*, *KMT2D*, and *KDM6A* are increasingly recognized as driving events in a variety of cancers and developmental disorders^{17,27}. Here, we demonstrate that stabilization of UTX in the absence of MLL3/4 may restore some tumor suppressive functions normally fulfilled by UTX as part of COMPASS. Still, several questions remain with regard to the mechanism of UTX-dependent anti-proliferative effects. Similar to Trr/MLL3/MLL4, catalytic-dependent and –independent biological functions have been demonstrated for UTX in controlling various biological processes, both in *Drosophila* and mammals. Our experiments were performed in a catalytic-active UTX^{WT} background; therefore, we cannot conclude at this time which of these functions is responsible for slowing effects on cell growth following induced stabilization of UTX. Furthermore, it is unclear how UTX is recruited to enhancers, as this protein does not have a recognizable DNA-binding domain. One recent study demonstrates MLL3/UTX recruitment occurs through interactions between MLL3 and Bap1, and disrupting this interaction abrogates UTX enhancer binding²⁵. Bap1-dependent recruitment appears to be specific to MLL3, as MLL4 interactions were not detected. ChIP-sequencing experiments will determine whether or not stabilizing UTX in the absence of MLL3/4 is sufficient to correctly re-localize UTX to chromatin, and if so, what proportion of those binding sites are restored. Our study raises the exciting prospect that a pharmacologic strategy to stabilize UTX could be therapeutically beneficial for pathologies, such as bladder cancer, in which MLL3/4 and UTX mutations tend to be mutually exclusive^{12,28}. Developing a compound with those properties will also teach us about the contribution of the MLL4 HMG-box in simply promoting UTX stability versus recruitment to chromatin. Utx is demonstrated to have catalytic-dependent and –independent functions in *Drosophila*²⁹; however, three homologs UTX, UTY, and JMJD3 exist in mammals, and it remains to be determined which of these functions are redundant among these factors. Finally, our study demonstrates that a key function of Trr is to

stabilize Utx in *Drosophila*, and restoring UTX tumor suppressive functions might be a novel therapeutic strategy for treating cancers in which UTX is destabilized due to deleterious mutations in MLL3/4.

Methods

Fly (*Drosophila melanogaster*) stocks.

Genomic *trr* rescue flies were generated using *pattB* plasmid for site-specific integration on 3R (89E11) and injections performed by BestGene (strain 9744). Transgenic flies were crossed to *trr*⁻¹ flies and then made homozygous for the *trr*⁻¹ allele and *trr* rescue construct.

Cell Fractionation and Immunoprecipitation

Cytosolic extracts were made using hypotonic buffer (10mM HEPES pH7.9, 1.5mM MgCl₂, 10mM KCl) with 0.2% Triton X-100. Nuclear extracts were made with high-salt buffer (20mM HEPES pH7.9, 25% glycerol, 1.5mM MgCl₂, 0.2mM EDTA, 350mM NaCl). Anti-Flag M2 agarose (Sigma) was used for all Flag-immunoprecipitation, and non-specific proteins were removed with wash buffer (10mM HEPES pH7.9, 1.5mM MgCl₂, 300mM NaCl, 10mM KCl, 0.2% Triton X-100). SDS-PAGE was performed with 4-20% mini-PROTEAN TGX precast gels (Biorad) and transferred to nitrocellulose membrane for western blotting.

Protein Sequence Alignment

Full length sequences were aligned with MUSCLE v3.8.31, trimmed and manually adjusted in AliView, and displayed with Jalview with Clustal colors with a coloring by conservation score of 40 for vertebrates and 20 for the combined vertebrate and invertebrate alignment.

Mass-Spectrometry

Sample preparation and mass-spectrometry of immunoprecipitated protein complexes were performed as described (Hickox et al. 2017).

CRISPR–Cas9 gene editing

Plasmids expressing Cas9 and guide RNAs (gRNAs) were constructed in pX330 (Cong et al. 2013). Human MLL4-CT deletion gRNAs were GTGGTGTTCGGCGGGTTACTC and GCATCCATTTCCGACAATTCC. MLL4-NT deletion gRNAs were GGAGCAGCTTTTGTACGAGC and GGGACATCTCCATCGTGATA. Guide-RNAs for MLL4 promoter deletions to generate Δ MLL4 HCT116 cells were GAGGGGACTGATATGCACCGG and GTGCATGGTCGGCAGGCGTAT. HCT116 cells were electroporated and single clones selected as described (Morgan 2017). Cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum, penicillin and streptomycin.

Size Exclusion Chromatography

Size exclusion separation of 293T cell extracts were performed with a Superose 6 column (GE Healthcare) on a SMART HPLC system (Amersham).

References

- 1 Herz, H. M. *et al.* Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. *Genes & Development* **26**, 2604-2620, doi:10.1101/gad.201327.112 (2012).
- 2 Hu, D. *et al.* The MLL3/MLL4 Branches of the COMPASS Family Function as Major Histone H3K4 Monomethylases at Enhancers. *Molecular and Cellular Biology* **33**, 4745-4754, doi:10.1128/mcb.01181-13 (2013).
- 3 Shilatifard, A. The COMPASS Family of Histone H3K4 Methylases: Mechanisms of Regulation in Development and Disease Pathogenesis. *Annu. Rev. Biochem.* **81**, 65-95, doi:10.1146/annurev-biochem-051710-134100 (2012).
- 4 Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731-734, doi:10.1038/nature06145 (2007).
- 5 Cho, Y. W. *et al.* PTIP associates with MLL3- and MLL4-containing histone H3 lysine 4 methyltransferase complex. *J Biol Chem* **282**, 20395-20406, doi:10.1074/jbc.M701574200 (2007).
- 6 Issaeva, I. *et al.* Knockdown of ALR (MLL2) reveals ALR target genes and leads to alterations in cell adhesion and growth. *Mol Cell Biol* **27**, 1889-1903, doi:10.1128/MCB.01506-06 (2007).
- 7 Mohan, M. *et al.* The COMPASS Family of H3K4 Methylases in Drosophila. *Molecular and Cellular Biology* **31**, 4310-4318, doi:10.1128/mcb.06092-11 (2011).
- 8 Lederer, D. *et al.* Deletion of KDM6A, a histone demethylase interacting with MLL2, in three patients with Kabuki syndrome. *Am J Hum Genet* **90**, 119-124, doi:10.1016/j.ajhg.2011.11.021 (2012).
- 9 Miyake, N. *et al.* MLL2 and KDM6A mutations in patients with Kabuki syndrome. *Am J Med Genet A* **161A**, 2234-2243, doi:10.1002/ajmg.a.36072 (2013).
- 10 Miyake, N. *et al.* KDM6A point mutations cause Kabuki syndrome. *Hum Mutat* **34**, 108-110, doi:10.1002/humu.22229 (2013).
- 11 Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-793, doi:10.1038/ng.646 (2010).
- 12 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315-322, doi:10.1038/nature12965 (2014).
- 13 Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333-339, doi:10.1038/nature12634 (2013).
- 14 Kim, J. H. *et al.* UTX and MLL4 coordinately regulate transcriptional programs for cell proliferation and invasiveness in breast cancer cells. *Cancer Res* **74**, 1705-1717, doi:10.1158/0008-5472.CAN-13-1896 (2014).
- 15 Ntziachristos, P. *et al.* Genetic inactivation of the polycomb repressive complex 2 in T cell acute lymphoblastic leukemia. *Nat Med* **18**, 298-301, doi:10.1038/nm.2651 (2012).
- 16 van Haaften, G. *et al.* Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet* **41**, 521-523, doi:10.1038/ng.349 (2009).
- 17 Wang, L. & Shilatifard, A. UTX Mutations in Human Cancer. *Cancer Cell* **35**, 168-176, doi:10.1016/j.ccell.2019.01.001 (2019).
- 18 Zhang, J. *et al.* Disruption of KMT2D perturbs germinal center B cell development and promotes lymphomagenesis. *Nat Med* **21**, 1190-1198, doi:10.1038/nm.3940 (2015).
- 19 Rickels, R. *et al.* Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat Genet* **49**, 1647-1653, doi:10.1038/ng.3965 (2017).

- 20 Dorigi K M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular Cell* (2017).
- 21 Chen, F. X., Smith, E. R. & Shilatifard, A. Born to run: control of transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol* **19**, 464-478, doi:10.1038/s41580-018-0010-5 (2018).
- 22 Rickels, R. & Shilatifard, A. Enhancer Logic and Mechanics in Development and Disease. *Trends Cell Biol* **28**, 608-630, doi:10.1016/j.tcb.2018.04.003 (2018).
- 23 Chauhan, C., Zraly, C. B., Parilla, M., Diaz, M. O. & Dingwall, A. K. Histone recognition and nuclear receptor co-activator functions of *Drosophila* cara mitad, a homolog of the N-terminal portion of mammalian MLL2 and MLL3. *Development* **139**, 1997-2008, doi:10.1242/dev.076687 (2012).
- 24 Ortega-Molina, A. *et al.* The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med* **21**, 1199-1208, doi:10.1038/nm.3943 (2015).
- 25 Wang, L. *et al.* Resetting the epigenetic balance of Polycomb and COMPASS function at enhancers for cancer therapy. *Nat Med* **24**, 758-769, doi:10.1038/s41591-018-0034-6 (2018).
- 26 Wu, Q. *et al.* In vivo CRISPR screening unveils histone demethylase UTX as an important epigenetic regulator in lung tumorigenesis. *Proc Natl Acad Sci U S A* **115**, E3978-E3986, doi:10.1073/pnas.1716589115 (2018).
- 27 Morgan, M. A. & Shilatifard, A. Chromatin signatures of cancer. *Genes Dev* **29**, 238-249, doi:10.1101/gad.255182.114 (2015).
- 28 Fantini, D. *et al.* A Carcinogen-induced mouse model recapitulates the molecular alterations of human muscle invasive bladder cancer. *Oncogene* **37**, 1911-1925, doi:10.1038/s41388-017-0099-6 (2018).
- 29 Copur, O. & Muller, J. Histone Demethylase Activity of Utx Is Essential for Viability and Regulation of HOX Gene Expression in *Drosophila*. *Genetics* **208**, 633-637, doi:10.1534/genetics.117.300421 (2018).

Chapter 4

Concluding Remarks

In summary, we have demonstrated that H3K4me1, an evolutionary conserved histone modification associated with transcriptional enhancer chromatin, is apparently dispensable for normal enhancer function and organismal development under laboratory conditions¹. These results strongly suggest H3K4me1 is generally not instructive for enhancer function and simply reflects Trr occupancy at those genomic regions. However, minor wing-vein phenotypes observed in either *trr-C2398A* or *trr-del4* flies reared under temperature stress indicate this histone modification might function in certain contexts to support enhancer-promoter communication. Future studies might employ *Drosophila* as a tool to further dissect this phenomenon. At 29°C, the H3K4me1-dependent L3/L4 cross-vein phenotype is highly penetrant in both the *trr-C2398A* and *trr-del4* flies. By crossing *trr-del4* to a genetic deletion collection and screening for enhancers or suppressors of the wing-vein phenotype at 29°C, one could potentially identify factors that mediate enhancer function in an H3K4me1-dependent manner.

Two recent studies in mESCs identify the Cohesin complex, as well as the BAF chromatin remodeling complex as H3K4me1-associated proteins at enhancers^{2,3} and provide evidence that MLL3/4-dependent H3K4me1 modulates long-range enhancer chromatin interactions. Enhancer looping is certainly demonstrated to occur in *Drosophila*⁴, and although we did not examine enhancer looping or chromatin topology in our study, it is difficult to believe the *trr-C2398A* flies could develop properly if those processes were significantly disrupted. In an attempt to reconcile these findings, I believe H3K4me1 might *influence* enhancer looping; however, the functional significance of diminishing that influence may only manifest as a phenotype in certain contexts, such as environmental stress. It would be interesting to determine whether or not loss of enhancer-associated H3K4me1 impacts organismal fitness in a

natural setting with changing environmental conditions, increased pathogen exposure, and food scarcity. As a starting point, one could introduce equal numbers of *trr-WT* and *trr-C2398A* flies, which are genetically identical except for two bases, into a vivarium with changing temperature, humidity, etc., and then use DNA-sequencing to quantify the ratio of *trr-WT* to *trr-C2398A* alleles after an extended period of time.

It remains to be determined whether or not MLL3/4 catalytic-inactivating mutations are also compatible with proper development in mammals. Our group and the Wysocka lab both conclude that MLL3/4, but not MLL3/4-dependent H3K4me1, are required for enhancer function in mESCs⁵; however, these cells were never used to generate a mouse. One group previously generated an MLL3- Δ SET mouse model and showed the perinatal lethal phenotype matched that of a homozygous MLL3-KO⁶; however these results are somewhat inconclusive, as the lethal phenotype appears to vanish in other genetic backgrounds⁷. Heterozygous mutations in the MLL4 SET domain are strongly associated with Kabuki syndrome in humans⁸, which suggests that even if MLL3/4 catalytic-inactive mice appear to develop normally, it will still be important to test for cognitive deficiencies in these animals. Whatever the outcome, it is not unreasonable to expect the phenotype associated with either MLL3 or MLL4 catalytic-inactivating mutations will be less severe than a complete gene knockout. I make this prediction based on similar studies of MLL1 in which mice expressing MLL1- Δ SET were shown to be completely viable, albeit with some interesting skeletal defects, while *Mll1*-null embryos die due to a complete failure of hematopoiesis⁹.

These studies certainly fit into a growing body of work illustrating differences between null and catalytic mutations of chromatin regulators¹⁰. One example is the PRC1 component, Sce (RING1B in mammals), an E3 ubiquitin ligase required for ubiquitination of H2A (H2Aub) and

PRC1-dependent gene silencing in *Drosophila*. Sce is required for Polycomb-dependent silencing during development; however, transcriptional repression was not affected either by Sce catalytic-inactivating mutations or by creating H2A lysine-to-arginine mutations at residues ubiquitinated by Sce. Flies carrying these genetic alterations did not display phenotypes characteristic of *sce* mutants, indicating the critical function of Sce in maintaining PRC1-dependent repression is separable from its enzymatic capabilities¹¹. In the case of Utx, numerous studies in flies and mice suggest the H3K27-demethylase activity of Utx is required in some contexts while dispensable in others^{12,13}. For instance, the same group who published the *sce* story showed in *Drosophila* Utx catalytic-mutants lose Hox gene expression and die during embryogenesis, which phenocopies Utx-null mutations¹⁴. Utx is required throughout *Drosophila* development; however, flies with a maternally deposited catalytic-active Utx progress through embryogenesis and develop into morphologically normal adults, indicating the demethylase activity of Utx is required during very early development and the non-enzymatic function prevails thereafter. Intriguingly, the life-span of those adult flies is dramatically shorter, raising the possibility that loss of Utx demethylase activity causes unseen developmental phenotypes that do not manifest until adulthood, or perhaps Utx enzymatic activity is important for adult homeostasis. *Kdm6a* is X-linked in mammals and, while several groups have described catalytic-independent roles for Utx in various tissue contexts, perhaps the most striking is a phenotypic comparison between *Kdm6a* deletions in male versus female mice. Homozygous *Kdm6a*-null mutations are embryonic lethal in females while hemizygous *Kdm6a*-null males can survive to adulthood due to a male-specific Utx homolog, Uty, which is highly similar to UTX but lacks detectable H3K27-demethylase activity¹⁵.

These examples underscore the necessity of using multiple approaches in our efforts to uncouple the biological roles of histone modifiers from histone modifications and, in some

cases, an unexpected result might teach of us something fundamental about the enzymes in question. I already described the remarkable findings from Buratowski and colleagues demonstrating that yeast SET1 protein stability is dependent on the presence of H3K4-methylation¹⁶. The SET domain is required for this feedback process, which suggests the SET domain is not only important for catalyzing, but also sensing its catalytic product. My own experience attempting to generate COMPASS family catalytic alleles indicates the SET domain structure is closely linked with enzyme stability. The recessive lethal *trx*^{Z11} allele contains a G3601S mutation in the SET domain, and this was described in the literature as evidence that loss Trx-dependent histone methylation recapitulates loss of Trx, although the *trx*^{Z11} protein product was never shown to be stable¹⁷. I used CRISPR/Cas9 to create this single residue substitution in *Drosophila* S2 cells and found the G3601 mutation destabilizes Trx (data not shown). Similarly, attempts by two independent labs to generate CRISPR cell lines with MLL3/4 SET domain deletions both resulted in MLL3/4 protein instability^{5,18}; however, we were fortunately able to create these mutations without affecting protein levels. Comparing the specific deletions revealed each lab generated unique deletions due to using different guide-RNA sequences, and in some cases, removing just a few extra residues corresponds with drastic reductions in MLL3 or MLL4 protein levels. These combined observations hint at an additional role for the SET domain in regulating enzyme stability, probably through its interactions with chromatin.

Just as enhancer-associated H3K4me1 is not generally required for enhancer function, our *Drosophila* studies demonstrate that conversion of H3K4me1 to H3K4me2/3 is also tolerated. Intriguingly though, quantitative increases in gene expression are detected at enhancer-proximal genes in the *trr*-Y2383F compared with *trr*-WT, which suggests that, while H3K4me3 is not instructive to initiate transcription, perhaps it can provide a 'boost' when deposited at sites

that have already initiated transcription. This is consistent with work from Adrian Bird's lab showing that deletion of *Cfp1* in mESCs causes ectopic deposition of H3K4me3 by Set1a/b at thousands of new sites throughout the genome, and in cases where ectopic H3K4me3 peaks overlap with active enhancers, as determined by GRO-seq signal, neighboring gene expression was also increased¹⁹. This thought-provoking study also demonstrates that loss of promoter-associated H3K4me3 has minimal consequences for transcription, and that H3K4me3 is not sufficient to initiate transcription at ectopic sites, which raises the question of how H3K4me3 is able to further stimulate enhancer activity. It would be interesting to repeat this experiment in a Set1a/b double-catalytic mutant to test whether these enhancer effects are really due to H3K4me3 or as a consequence of ectopic Set1a/b recruitment.

This is a very exciting time for the chromatin and transcription fields because the more simplistic, albeit conceptually attractive, models that placed undue importance on the modifications versus the modifiers are slowly being replaced by more nuanced perspectives that view histone modifications as a reflection of the dynamic processes that regulate transcription. This perspective raises many new questions. For instance, the methyltransferase activity of Trr, and potentially other COMPASS enzymes, is not required for viability, then why is the SET domain the most highly conserved portion of these proteins? If H3K4-methylation plays a role in fine-tuning enhancer function under temperature stress, as we proposed, then what is the molecular mechanism that allows it to function in that capacity?

The phase separation model of enhancer function postulates a multi-molecular assembly of chromatin proteins, including nucleosomes, engage in multivalent, low-affinity interactions which increase local concentrations of transcriptional regulators to stimulate gene expression²⁰. If this model is correct, then enhancer-associated H3K4me1, and any factors that bind H3K4me1,

represent just one of many cross-links contributing to formation of the multi-molecular assembly. Perhaps the effect of losing enhancer H3K4-methylation is negligible for nuclear condensate formation, while under temperature stress the condensate is less stable, leading to enhancer dysfunction in certain tissues. If this is correct, then what role does MLL3/4 play in the process? Clearly, MLL3/4 function as an important scaffold to stabilize UTX, but are other chromatin proteins also stabilized in a similar manner by MLL3/4? What other nuclear functions does UTX support outside of its H3K27-demethylase activity?

Many functions have been ascribed to MLL3/4 including transcriptional activation, transcriptional repression, and regulation of genome stability; however, these conclusions are drawn either from genetic mutations or RNAi knock-down experiments, and lack the temporal resolution necessary to determine the acute effects of disrupting MLL3/4 or MLL3/4 catalytic-activity versus secondary or tertiary effects. Future experiments could make use of auxin-inducible degron systems, or similar systems that induce rapid protein degradation, combined with techniques that measure nascent transcription (GRO-seq, PRO-seq, etc.) to examine acute effects that occur within one or two hours. Alternatively, one could use SILAC or isobaric labeling techniques to look at proteomic changes over time to identify additional factors that rely on MLL3/4 for stability. It would be interesting to degrade MLL3 or MLL4, or both, in the presence of an un-degradable catalytic-dead MLL3/4 to assess transcriptional differences that result over the course of one or two cell cycles. Similar approaches could be applied to UTX, whose functions outside of regulating H3K27-methylation are largely unknown. The mechanisms surrounding MLL3/4-dependent stabilization of UTX should continue to be investigated, as restoring UTX tumor suppressive function might present a new therapeutic approach for treating MLL3/4-dependent tumors. Targeted CRISPR screens to carry out saturation mutation of *KDM6A* fused to a fluorescent reporter might allow high-throughput

identification of UTX-destabilizing mutations. This novel approach has the potential to identify a degron within UTX that is normally blocked by MLL3/4 to prevent UTX degradation, and might also provide functional links between UTX stability and specific *KDM6A* mutations found in tumor samples. These are just a few ideas of how new technologies might be applied to address some outstanding questions raised by my thesis work. There is no shortage of perplexity in the chromatin field^{21,22}; however, continued development of new techniques will allow us to ask more precise questions in our experiments and disentangle the complex interplay between histone modifiers and their catalytic products in regulating gene expression.

References

- 1 Rickels, R. *et al.* Histone H3K4 monomethylation catalyzed by Trr and mammalian COMPASS-like proteins at enhancers is dispensable for development and viability. *Nat Genet* **49**, 1647-1653, doi:10.1038/ng.3965 (2017).
- 2 Local, A. *et al.* Identification of H3K4me1-associated proteins at mammalian enhancers. *Nat Genet* **50**, 73-82, doi:10.1038/s41588-017-0015-6 (2018).
- 3 Yan, J. *et al.* Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res* **28**, 387, doi:10.1038/cr.2018.18 (2018).
- 4 Ghavi-Helm, Y. *et al.* Enhancer loops appear stable during development and are associated with paused polymerase. *Nature* **512**, 96-100, doi:10.1038/nature13417 (2014).
- 5 Dorighi K M. *et al.* Mll3 and Mll4 Facilitate Enhancer RNA Synthesis and Transcription from Promoters Independently of H3K4 Monomethylation. *Molecular Cell* (2017).
- 6 Lee, J. *et al.* Targeted inactivation of MLL3 histone H3-Lys-4 methyltransferase activity in the mouse reveals vital roles for MLL3 in adipogenesis. *Proc Natl Acad Sci U S A* **105**, 19229-19234, doi:10.1073/pnas.0810100105 (2008).
- 7 Arcipowski, K. M., Bulic, M., Gurbuxani, S. & Licht, J. D. Loss of Mll3 Catalytic Function Promotes Aberrant Myelopoiesis. *PLoS One* **11**, e0162515, doi:10.1371/journal.pone.0162515 (2016).
- 8 Ng, S. B. *et al.* Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* **42**, 790-793, doi:10.1038/ng.646 (2010).
- 9 Terranova, R., Agherbi, H., Boned, A., Meresse, S. & Djabali, M. Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. *Proc Natl Acad Sci U S A* **103**, 6629-6634, doi:10.1073/pnas.0507425103 (2006).

- 10 Aubert, Y., Egolf, S. & Capell, B. C. The Unexpected Noncatalytic Roles of Histone Modifiers in Development and Disease. *Trends Genet*, doi:10.1016/j.tig.2019.06.004 (2019).
- 11 Pengelly, A. R., Kalb, R., Finkl, K. & Muller, J. Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev* **29**, 1487-1492, doi:10.1101/gad.265439.115 (2015).
- 12 Wang, C. *et al.* UTX regulates mesoderm differentiation of embryonic stem cells independent of H3K27 demethylase activity. *Proc Natl Acad Sci U S A* **109**, 15324-15329, doi:10.1073/pnas.1204166109 (2012).
- 13 Wu, Q. *et al.* In vivo CRISPR screening unveils histone demethylase UTX as an important epigenetic regulator in lung tumorigenesis. *Proc Natl Acad Sci U S A* **115**, E3978-E3986, doi:10.1073/pnas.1716589115 (2018).
- 14 Copur, O. & Muller, J. Histone Demethylase Activity of Utx Is Essential for Viability and Regulation of HOX Gene Expression in Drosophila. *Genetics* **208**, 633-637, doi:10.1534/genetics.117.300421 (2018).
- 15 Wang, L. & Shilatifard, A. UTX Mutations in Human Cancer. *Cancer Cell* **35**, 168-176, doi:10.1016/j.ccell.2019.01.001 (2019).
- 16 Soares, L. M., Radman-Livaja, M., Lin, S. G., Rando, O. J. & Buratowski, S. Feedback control of Set1 protein levels is important for proper H3K4 methylation patterns. *Cell Rep* **6**, 961-972, doi:10.1016/j.celrep.2014.02.017 (2014).
- 17 Smith, S. T. *et al.* Modulation of heat shock gene expression by the TAC1 chromatin-modifying complex. *Nat Cell Biol* **6**, 162-167, doi:10.1038/ncb1088 (2004).
- 18 Jang, Y. *et al.* H3.3K4M destabilizes enhancer H3K4 methyltransferases MLL3/MLL4 and impairs adipose tissue development. *Nucleic Acids Res* **47**, 607-620, doi:10.1093/nar/gky982 (2019).
- 19 Clouaire, T. *et al.* Cfp1 integrates both CpG content and gene activity for accurate H3K4me3 deposition in embryonic stem cells. *Genes Dev* **26**, 1714-1728, doi:10.1101/gad.194209.112 (2012).
- 20 Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23, doi:10.1016/j.cell.2017.02.007 (2017).
- 21 Howe, F. S., Fischl, H., Murray, S. C. & Mellor, J. Is H3K4me3 instructive for transcription activation? *Bioessays* **39**, 1-12, doi:10.1002/bies.201600095 (2017).
- 22 Rada-Iglesias, A. Is H3K4me1 at enhancers correlative or causative? *Nat Genet* **50**, 4-5, doi:10.1038/s41588-017-0018-3 (2018).