



ISSN: 1933-6934 (Print) 1933-6942 (Online) Journal homepage: https://www.tandfonline.com/loi/kfly20

Correspondence of Drosophila polycomb group proteins with broad H3K27me3 silent domains

Youngsook L. Jung, Hyuckjoon Kang, Peter J. Park & Mitzi I. Kuroda

To cite this article: Youngsook L. Jung, Hyuckjoon Kang, Peter J. Park & Mitzi I. Kuroda (2015) Correspondence of Drosophila polycomb group proteins with broad H3K27me3 silent domains, Fly, 9:4, 178-182, DOI: 10.1080/19336934.2016.1151988

To link to this article: https://doi.org/10.1080/19336934.2016.1151988

4	1	(1
Б			

Published online: 03 Mar 2016.



Submit your article to this journal 🕑





View related articles



View Crossmark data 🗹

ረት	Citing articles: 1	View citing articles	2
പ്പ	Citing articles: 1	View citing articles	

Correspondence of *Drosophila* polycomb group proteins with broad H3K27me3 silent domains

Youngsook L. Jung^{1,#}, Hyuckjoon Kang^{2,3,#}, Peter J. Park¹, and Mitzi I. Kuroda^{2,3,*}

¹Department of Biomedical Informatics; Harvard Medical School; Boston, MA 02115; ²Division of Genetics; Brigham and Women's Hospital; Boston, MA 02115;

³Department of Genetics; Harvard Medical School; Boston, MA 02115

[#]These authors contributed equally to this work.

The Polycomb group (PcG) proteins are key conserved regulators of development, initially discovered in Drosophila and now strongly implicated in human disease. Nevertheless, differing silencing properties between the Drosophila and mammalian PcG systems have been observed. While specific DNA targeting sites for PcG proteins called Polycomb response elements (PREs) have been identified only in Drosophila, involvement of non-coding RNAs for PcG targeting has been favored in mammals. Another difference lies in the distribution patterns of PcG proteins. In mouse and human cells, PcG proteins show broad distributions, significantly overlapping with H3K27me3 domains. In contrast, only sharp peaks on PRE regions are observed for most PcG proteins in Drosophila, raising the question of how large domains of H3K27me3, up to many tens of kilobases, are formed and maintained in Drosophila. In this Extra View, we provide evidence that PcG distributions on silent chromatin in Drosophila are considerably broader than previously detected. Using BioTAP-XL, a chromatin crosslinking and tandem affinity purification approach, we find a broad, rather than PRE-limited overlap of PcG proteins with H3K27me3, suggesting a conserved spreading mechanism for PcG in flies and mammals.

Since Polycomb group (PcG) proteins were first discovered in *Drosophila* as regulators that prevent inappropriate expression of Hox genes,¹ several PcG proteins have been characterized. Many of them can be classified into 2 principal complexes: PcG-repressive complex 1 (PRC1), involved in chromatin compaction, and

PRC2, which mediates histone H3K27 methylation. Drosophila PRC1 and PRC2 each have 4 core components. PRC1 consists of Polycomb (Pc), Polyhomeotic (Ph), dRING, and Posterior sex combs (Psc), while PRC2 is comprised of Enhancer of zeste (E(z)), Suppressor of zeste 12 (Su(z)12), Extra sex combs (Esc) and Nurf55.²⁻⁵ E(z), the catalytic subunit of PRC2, has a SET domain with lysine methyltransferase activity and is solely responsible for methylation of histone H3K27.4-6 Pc, a subunit of PRC1, contains a chromodomain that is able to specifically recognize the H3K27me3 histone mark added by PRC2.7 Therefore, H3K27me3 is strongly associated with PcG silencing.

Genome-wide analyses using chromatin immunoprecipitation (ChIP) have shown that PcG proteins are localized at hundreds of specific sites (known and presumptive PREs) and H3K27me3 silent domains are formed at many of these PcG target regions.^{8,9} Nevertheless, the binding profiles have also revealed that the H3K27me3 domains are distributed broadly around PRE sites, while PcG proteins, even E(z) responsible for H3K27me3, show sharp peaks on chromatin, suggesting PcG protein binding is confined to narrow PRE sites.⁸ Drosophila PREs often contain diverse combinations of consensus sequences for DNA binding proteins such as Pho, SP1/KLF, GAF, Psq, Dsp1, Grh and Zeste.¹⁰ In mammals, however, no defined PREs containing specific sequence motifs have been found so far. Consistently, ChIP binding profiles of mammalian PcG proteins lack the sharp peaks on specific chromatin sites seen in

Keywords: BioTAP-XL, *Drosophila*, H3K27me3, polycomb Group (PcG), PREs

*Correspondence to: Mitzi I. Kuroda Email: mkuroda@genetics.med.harvard.edu

Submitted: 12/14/2015

Revised: 02/01/2016

Accepted: 02/03/2016

http://dx.doi.org/10.1080/19336934.2016.1151988

Extra View to: Kang HJ, McElroy KA, Jung YL, Alekseyenko AA, Zee BM, Park PJ, Kuroda MI (2015). Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in Drosophila. Genes Dev. 29: 1136-1150.

Drosophila; instead they show a broad distribution significantly overlapping with H3K27me3 domains,^{11,12} accounting for the formation of broad H3K27me3 domains in mammals. This difference has led to the following question: How are the broad H3K27me3 domains formed in spite of the confined PcG binding on narrow PREs in *Drosophila*? It has been suggested that some components of PcG complexes on PREs may interact with flanking nucleosomes, triggering the formation of a loop domain, thereby bringing the neighboring nucleosomes into the vicinity of PRC2 on PREs.^{13,14}

To further understand the functional mechanisms of PRC1 and PRC2 in silencing, we recently utilized a tandem affinity purification approach, BioTAP-XL, with Pc and E(z) as bait proteins for PRC1 and PRC2 complexes, respectively.¹⁵ The crosslinking-coupled tandem affinity purificanot only improves tion strategy preservation of bait protein-protein interactions, but also allows the analysis of target DNA in parallel.¹⁶⁻¹⁸ We found that Sex comb on midleg (Scm), previously known as a substoichiometric component of PRC1,^{3,19} is the only protein strongly enriched with both complexes.¹⁵ Surprisingly, we also found through BioTAP-XL DNA sequencing that Pc, E(z) and Scm show a broad distribution on silent chromatin and overlap significantly with the broad domains of H3K27me3, though they also still show sharp peaks on PREs. Herein, we further investigate the binding patterns of these PcG proteins on silent chromatin.

First, we examined the profiles of Polycomb proteins and H3K27me3 in S2 cells from the modENCODE (model organism encylopedia of DNA elements) consortium (Fig. 1). The broad regions significantly enriched by H3K27me3 often lacked binding of Polycomb proteins such as E(z), Pc or Pcl which displayed much narrower peaks compared to H3K27me3. An example is shown on chr3R, proximal to the Hox gene cluster (Fig. 1A). The domain of 12.9-13.5 Mbp is mostly transcriptionally silent and is significantly enriched with H3K27me3 except for the topological domain boundaries where active marks such as H3K36me3 are enriched. Interestingly, Polycomb proteins profiled from the modENCODE project do not co-localize with H3K27me3 for most of the regions. However, Polycomb protein profiles mapped through BioTAP-XL largely overlap with H3K27me3 and exhibit broader enrichment patterns compared to those from the modENCODE consortium (Fig. 1A). It is also noteworthy that both H3K27me3 and PcG proteins mapped by BioTAP-XL fill the entire regions inside of topological domains, which was not previously observed as further discussed below. In general, profiles of Polycomb proteins analyzed using BioTAP-XL are mutually exclusive with transcriptionally active regions enriched for H3K36me3 (Fig. 1B). This is expected for proteins that play mainly a silencing role; potential roles in activation are not examined here.^{15,20} To further test for the specificity of our PcG-BioTAP profiles, we compared them to previous data for MSL3, a member of the MSL (Male-Specific Lethal) complex which is required for X chromosome dosage compensation in Drosophila. MSL3-BioTAP specifically shows high enrichment over the transcriptionally active regions on chrX (Fig. 1C). This demonstrates that BioTAP-XL profiles reflect the distribution of each specific bait protein, rather than any shared affinity of the epitope tag.¹⁷

To probe the colocalization pattern of H3K27me3 and Polycomb proteins with respect to topologically associating domains (TADs) genome-wide, we compared the enrichment of H3K27me3 and Polycomb proteins around the topological boundaries from Hi-C data in embryos (Fig. 2A).^{21,22} H3K27me3 is spread within many TADs and is confined within the domains, as previously observed,²¹ while Polycomb group proteins profiled by modENCODE generally do not fill TADs. Surprisingly, PcG protein profiles in our work using BioTAP-XL display similarity to H3K27me3-they are broadly distributed within the topological domains and are strongly overlapping with H3K27me3 genome-wide. Next, we compared the enrichments of previously profiled PcG proteins and BioTAP-XL profiles with the significantly enriched regions of H3K27me3 (ChIP over input enrichment z score > 3) (Fig. 2B). The comparison showed that the PcG profiles

from modENCODE are not enriched for most regions within H3K27me3 enrichment, while the BioTAP-XL profiles are comparable to that of H3K27me3. In addition, the genome-wide correlation of these factors reveals that H3K27me3 is clustered most closely with the BioTAP-XL rather than modENCODE PcG profiles (Fig. 2C). To further compare the broadness of PcG enrichment from our BioTAP-XL analyses with those from typical ChIP-seq, we examined embryonic pulldowns for which sequencing data are available. When the significant peaks were determined as the regions where the Poisson rate of IP reads is over 3-fold higher than that of input (Fig. 2D), the peaks of PcG proteins enriched by BioTAP-XL are significantly broader than those from embryonic ChIP-seq data profiled by modENCODE, including for Pc.

Given the similar broad binding patterns of Pc, E(z) and Scm PcG proteins in S2 cells and/or embryos using BioTAP-XL, one might question whether technical artifacts related to the BioTAP tag are an explanation. However, these broad distributions are not observed in a non-specific manner but are significantly and precisely overlapped with H3K27me3, showing the enrichment of the whole region inside topological domains and depletion from the domain boundaries. In addition, Bio-TAP-XL DNA sequencing data from MSL3, a factor uninvolved with PcG silencing, shows that binding signals of MSL3-BioTAP do not overlap with tagged PcG proteins, indicating that there are no serious technical issues with use of the BioTAP epitope which might result in significant systemic artifacts. Rather, use of the identical BioTAP tags (Protein A and Biotinylation sequences) on the 3 different PcG proteins, and the contrasting results with MSL3 addresses any potential problem resulting from differential sensitivity and specificity between protein-specific antibodies.

If not due to epitope tagging, how might we reconcile our contrasting results? Many technical factors including antibody quality, fixation conditions and sonication parameters have historically been issues to be considered for the generation of high-quality ChIP-seq data. Briefly, a typical ChIP experiment utilizes 1% formaldehyde



Figure 1. Enrichment profiles of Pc, E(z), Pcl, H3K27me3, and H3K36me3 (mapped by modENCODE), and Scm, Pc and MSL3 (mapped using BioTAP-XL) in S2 cells. A. Log2 fold enrichment profiles (IP over input) in the regions of 12.9-13.5 Mb on chr3R. The top triangle displays the topological domains (TADs) from Hi-C data in embryos²¹ showing higher contact regions with a red color. Red dashed horizontal lines indicate where fold enrichment is 0.5. Dashed vertical lines represent topological domain boundaries, which are typically conserved between different cell types.²³ B. Enrichment profiles in the domain of 11.0-11.2Mb on chr3R where most regions are transcriptionally active. C. Enrichment profiles in the domain of 11.14-11.48 Mb on chrX around the *roX2* gene where MSL complex proteins bind.

fixation, subsequent sonication to produce soluble crosslinked chromatin, and one step affinity purification using antibodies directed against a protein of interest. Alternatively, BioTAP-XL starts with cell lysis, exposing nuclei to 3% formaldehyde, followed by sonication to produce soluble crosslinked chromatin. Subsequently, the 2-step affinity purification sequentially exploits 2 very strong interactions, first protein A-IgG, and then biotin-streptavidin. The second step allows very stringent washing (0.2% SDS + 6M urea), retaining the tightly bound specific factors, while removing any remaining non-specific interactions. Thus, the BioTAP-XL procedure has been optimized for both high yield and increased signal to noise enrichment. For these reasons, we favor an explanation in which our results reflect an increased ability of Bio-TAP-XL to capture a wide range of interactions on chromatin. In the case of PcG

proteins, this would include the well-documented stable binding found at PREs, as well as potentially more transient interactions within broad H3K27me3 domains.⁸

So far, we have performed DNA sequencing of only 3 Drosophila PcG proteins (Pc, E(z) and Scm) using BioTAP-XL, and thus we do not know whether other Drosophila PcG proteins similarly distribute broadly on chromatin. We also cannot exclude the possibility that the discrepancy of chromatin binding broadness between E(z) in S2 cells from the modEN-CODE consortium and BioTAP-E(z) in embryos may be due to the difference of cell types. However, Pc was analyzed in S2 cells and in embryos by both methods: standard antibody pulldown by modEN-CODE and BioTAP-XL tandem affinity by our group. Thus, the contrast between the specific results obtained studying Pc is strong evidence that additional PcG

proteins may not be limited to narrow peaks at presumptive PRE regions. That at least some *Drosophila* PcG proteins have broad binding patterns, significantly overlapping with H3K27me3, suggests that mechanisms underlying the spreading of the H3K27me3 mark over broad PcGsilenced regions will be conserved between *Drosophila* and mammals.

Data accessibility

Pc, Scm and E(z) profiles mapped through BioTAP-XL were obtained from GSE66183. MSL3 BioTAP-XL data were downloaded from GSE56101. The accession numbers for modENCODE data are as follows: GSE20804 (Pc), GSE20765 (Pcl), GSE20769 (E(z)), GSE20781 (H3K27me3), GSE20785 (H3K36me3) in S2 cells, and GSE47232 (Pc),



Figure 2. Genome-wide comparison of enrichment patterns of Polycomb group proteins between modENCODE profiles and BioTAP-XL profiles in S2 cells and embryos. A. Enrichment of H3K27me3, Pc, Pcl, E(z) (from modENCODE) compared to Scm and Pc (using BioTAP-XL) in S2 cells, over scaled topological domains with a margin of 10 kb. Each row represents one topological domain (N = 1088) and rows are sorted by intensities of H3K27me3. Dashed vertical lines indicate the boundaries of the domains. Red: enriched. Blue: depleted. The topological boundary information was obtained from Ho *et al.*²² B. Distribution of the average fold enrichment values in H3K27me3 peaks for Pc, PcI, E(z) (modENCODE), and Scm and Pc (BioTAP-XL) in S2 cells. C. Genome-wide Pearson correlation coefficients for H3K27me3, H3K36me3, Pc, PcI, E(z) (modENCODE), and Scm and Pc (BioTAP-XL) from S2 cells using 500 bp bins. D. Comparison of peak broadness for Psc, Pc (modENCODE ChIP-seq), and Scm, Pc and E(z) (BioTAP-XL) in embryos. ****: *P*-value < 10^{-15} . Even for the same factor, Pc, the profile mapped by BioTAP-XL shows significantly broader peaks than that from ChIP-seq.

GSE47235 (Psc), GSE47230 (H3K27me3) and GSE47256 (H3K36me3) in embryos.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We are grateful to A. Alekseyenko for insightful discussions, and K. McElroy and H. Wallace for critical reading of the manuscript.

Funding

This work was supported by the National Institutes of Health (GM101958 to MIK).

References

- Lewis EB. A gene complex controlling segmentation in Drosophila. Nature 1978; 276:565-70; PMID:103000; http://dx.doi.org/10.1038/276565a0
- Shao Z, Raible F, Mollaaghababa R, Guyon JR, Wu CT, Bender W, Kingston RE. Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell 1999; 98:37-46; PMID:10412979; http://dx.doi.org/ 10.1016/S0092-8674(00)80604-2
- 3. Saurin AJ, Shao Z, Erdjument-Bromage H, Tempst P, Kingston RE. A Drosophila Polycomb group complex

includes Zeste and dTAFII proteins. Nature 2001; 412:655-60; PMID:11493925; http://dx.doi.org/ 10.1038/35088096

- Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, Pirrotta V. Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 2002; 111:185-96; PMID:12408863; http://dx.doi.org/ 10.1016/S0092-8674(02)00975-3
- Muller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, Simon JA. Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 2002; 111:197-208; PMID:12408864; http://dx.doi.org/10.1016/S0092-8674(02)00976-5
- Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, Zhang Y. Role of histone H3 lysine 27 methylation in Polycombgroup silencing. Science 2002; 298:1039-43;

PMID:12351676; http://dx.doi.org/10.1126/science. 1076997

- Fischle W, Wang Y, Jacobs SA, Kim Y, Allis CD, Khorasanizadeh S. Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev 2003; 17:1870-81; PMID:12897054; http://dx.doi.org/ 10.1101/gad.1110503
- Schwartz YB, Kahn TG, Nix DA, Li XY, Bourgon R, Biggin M, Pirrotta V. Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat Genet 2006; 38:700-5; PMID:16732288; http://dx.doi.org/ 10.1038/ng1817
- Tolhuis B, de Wit E, Muijrers I, Teunissen H, Talhout W, van Steensel B, van Lohuizen M. Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in Drosophila melanogaster. Nat Genet 2006; 38:694-9; PMID:16628213; http://dx.doi.org/ 10.1038/ng1792
- McElroy KA, Kang H, Kuroda MI. Are we there yet? Initial targeting of the Male-Specific Lethal and Polycomb group chromatin complexes in Drosophila. Open Biol 2014; 4:140006; PMID:24671948; http://dx.doi. org/10.1098/rsob.140006
- Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev 2006; 20:1123-36; PMID:16618801; http://dx.doi. org/10.1101/gad.381706
- Schwartz YB, Pirrotta V. Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet 2007; 8:9-22; PMID:17173055; http:// dx.doi.org/10.1038/nrg1981

- Kahn TG, Schwartz YB, Dellino GI, Pirrotta V. Polycomb complexes and the propagation of the methylation mark at the Drosophila ubx gene. J Biol Chem 2006; 281:29064-75; PMID:16887811; http://dx.doi. org/10.1074/jbc.M605430200
- Papp B, Muller J. Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev 2006; 20:2041-54; PMID:16882982; http://dx.doi.org/10.1101/gad.38-8706
- Kang H, McElroy KA, Jung YL, Alekseyenko AA, Zee BM, Park PJ, Kuroda MI. Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in Drosophila. Genes Dev 2015; 29:1136-50; PMID:26063573; http://dx.doi.org/ 10.1101/gad.260562.115
- Alekseyenko AA, Gorchakov AA, Kharchenko PV, Kuroda MI. Reciprocal interactions of human C10orf12 and C17orf96 with PRC2 revealed by Bio-TAP-XL cross-linking and affinity purification. Proc Natl Acad Sci U S A 2014; 111:2488-93; PMID:24550272; http://dx.doi.org/10.1073/pnas.14-00648111
- Alekseyenko AA, Gorchakov AA, Zee BM, Fuchs SM, Kharchenko PV, Kuroda MI. Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs. Genes Dev 2014; 28:1445-60; PMID:24990964; http://dx.doi.org/ 10.1101/gad.241950.114
- Alekseyenko AA, McElroy KA, Kang H, Zee BM, Kharchenko PV, Kuroda MI. BioTAP-XL: Cross-linking/Tandem Affinity Purification to Study DNA Targets, RNA, and Protein Components of Chromatin-

Associated Complexes. Curr Protoc Mol Biol 2015; 109:21 30 1-21 30 2; PMID:25559106

- Peterson AJ, Mallin DR, Francis NJ, Ketel CS, Stamm J, Voeller RK, Kingston RE, Simon JA. Requirement for sex comb on midleg protein interactions in Drosophila polycomb group repression. Genetics 2004; 167:1225-39; PMID:15280237; http://dx.doi.org/ 10.1534/genetics.104.027474
- Schaaf CA, Misulovin Z, Gause M, Koenig A, Gohara DW, Watson A, Dorsett D. Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. PLoS Genet 2013; 9: e1003560; PMID:23818863; http://dx.doi.org/ 10.1371/journal.pgen.1003560
- Sexton T, Yaffe E, Kenigsberg E, Bantignies F, Leblanc B, Hoichman M, Parrinello H, Tanay A, Cavalli G. Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 2012; 148:458-72; PMID:22265598; http://dx.doi.org/ 10.1016/j.cell.2012.01.010
- Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A, et al. Comparative analysis of metazoan chromatin organization. Nature 2014; 512:449-52; PMID:25164756; http://dx.doi.org/10.1038/nature13415
- Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012; 485:376-80; PMID:22495300; http://dx.doi.org/10.1038/nature11082