GENOME-WIDE INTERGENIC REPRESSION BY POLYCOMB GROUP PROTEINS

By

HUNGOO LEE

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

and

The Graduate School of Biomedical Sciences

University of Medicine and Dentistry of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Biochemistry

Written under the direction of

Professor Vincenzo Pirrotta

And approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2014

ABSTRACT OF THE DISSERTATION

GENOME-WIDE INTERGENIC REPRESSION BY POLYCOMB GROUP PROTEINS

By HUNGOO LEE

Dissertation Director:

Vincenzo Pirrotta, Ph.D.

Polycomb group (PcG) proteins have been studied as epigenetic regulators of diverse developmental regulatory genes in various organisms. PcG proteins form two major complexes: the Polycomb repressive complex 2 (PRC2) which includes Enhancer of zeste (E(Z)) that methylates the lysine 27 residue of histone H3 (H3K27) and the Polycomb repressive complex 1 (PRC1) which preferentially binds to tri-methylated H3K27 (H3K27me3). Recent genome-wide analysis revealed that the PcG proteins stably bind to hundreds of genomic loci with strong H3K27me3 enrichment.

Unlike the focused enrichment of H3K27me3, di-methylation of H3K27 (H3 K27me2) is found on more than half of all nucleosomal H3, raising the interesting possibility that H3K27me2 might have a potential repressive role across the genome.

To address the possibility, we generated an *E(z)* temperature sensitive mutant cell line and examined the effects of E(Z) inactivation. Upon E(Z) inactivation, we observed a strong transcriptional increase from the H3K27me2-enriched intergenic regions along with significant increases of H3K27ac and H3K4me1. The intergenic de-repression was dependent on two dUTX-containing H3K27 de-methylase complexes that are associated with the increased active marks. Genome-wide analysis revealed that these effects are prevailing all over the genome, suggesting an antagonistic relationship between PRC2 and the dUTX complexes for regulating chromatin states genome-wide.

In addition, PRC1 involvement in the intergenic repression was indicated by increased intergenic transcription by knock-downs or a mutation of PRC1 components. Moreover, a significant level of H2Aub1, a repressive mark produced by PRC1, was found in silent intergenic regions.

These results extend the roles of PcG proteins by showing that PcG complexes not only function as master regulators of several hundred developmental regulators but also set a high inhibitory threshold to suppress the pervasive intergenic transcription all over the genome.

Acknowledgement

First of all, I would like to express my deepest gratitude to my advisor Dr. Vincenzo Pirrotta, for his support, guidance, and patience. I learned a lot from him not only the knowledge in diverse aspects of science but also the attitude of a sincere scientist. His endless enthusiasm for science has made him my role model and urged me to be a better scientist.

I am also very grateful to the committee members of my dissertation, Dr. Marc Gartenberg, Dr. Mike Hampsey, and Dr. Isaac Edery. During all committee meetings, they have encouraged me and provided helpful suggestions for my thesis projects.

I also would like to thank our lab members. Donna McCabe has been very supportive and willing to help me with my research and life. Former lab members Yuri Schwartz and Tanya Kahn taught me many experimental skills and allowed their time for helpful discussions. I also want to thank Sung-Yeon Park, Katsu Ohno, Huabing Li, Simon Guo, Corrado Caslini and Dalal Asker for their kind support and encouragement. Also, I thank Hangnoh Lee for many helpful discussions and friendship.

I deeply thank my family for their love and support. My parents were always supportive and encouraged me to pursue my dream, being a good scientist. Especially, I want to express my sincere love and appreciation to my wife, Hyosoo Lee, for her encouragement and support. She has been always there for me, even though she was having difficulty with raising our son and struggling with her own study. Finally, I would like to thank my son, Jay Bok Lee, who brings inexpressible joy to my life.

Table of Contents

List of figures

List of Tables

Chapter 1. Introduction

1.1 Polycomb group (PcG) and Trithorax group (TrxG) proteins

Genes encoding PcG proteins were first identified from genetic studies in *Drosophila melanogaster* as regulators that are required for the appropriate expression of homeotic genes. PcG proteins are widely conserved from flies to mammals, and are well-known transcriptional repressors for a wide variety of genes that are involved in embryonic development, X-inactivation, genetic imprinting and cell cycle progression (Oktaba *et al.*, 2008; Schuettengruber *et al.*, 2007; Schwartz and Pirrotta, 2007).

Trithorax group (TrxG) proteins are known to antagonize PcG proteins for regulating PcG targets (Klymenko *et al.*, 2006; Schwartz and Pirrotta, 2008)**,** and these two groups epigenetically regulate their targets by various histone modifications and maintain the transcriptional states of target genes throughout multiple cell divisions (Schwartz and Pirrotta, 2007).

1.1.1 PcG proteins

PcG proteins form two major complexes: the Polycomb repressive complex 2 (PRC2) which methylates the lysine 27 residue of histone H3 (H3K27) and the Polycomb repressive complex 1 (PRC1) which preferentially binds to tri-methylated H3K27 (H3K27me3) (Figure 1.1, Table 1.1).

The PRC2 complex includes Enhancer of Zeste (E(Z)), Extra sex combs (ESC), SU(Z)12 and p55. E(Z) is the only methyl-transferase for H3K27me2/3 in *Drosophila*. ESC (or its paralog ESCL) and SU(Z)12 contribute to the ability of the PRC2 complex to bind and

methylate H3K27 (Czermin *et al.*, 2002; Ketel *et al.*, 2005; Margueron *et al.*, 2009; Müller *et al.*, 2002). Besides, some other additional components of PRC2 can be found in isoforms of PRC2. RPD3, a histone de-acetylase, and Polycomb-like (PCL) were co-purified with isoforms of PRC2 complex (Tie *et al.*, 2003). PCL was reported to contribute to high H3K27me3 level specifically on PcG target regions (Nekrasov *et al.*, 2007). Another histone de-acetylase SIR2 was also found in larval PRC2. Since acetylated H3K27 (H3K27ac) is an epigenetic mark antagonistic to methylated H3K27, it is notable that these histone de-acetylases can interact with the PRC2 core complex.

The key components of *Drosophila* PRC1 complex are Polycomb (PC), Posterior sex combs (PSC) (or its paralog SU(Z)2), polyhomeotic (PH), and dRING (Saurin *et al.*, 2001; Shao *et al.*, 1999). PC is known to bind specifically to H3K27me3 (Fischle *et al.*, 2003) via its chromo-domain. dRING contains RING domain and functions as an E3 ubiquitin ligase that mono-ubiquitinates lysine 118 residue of histone H2A (H2Aub1) in *Drosophila* and in mammals (Cao *et al.*, 2005; de Napoles *et al.*, 2004; Wang *et al.*, 2004a). PSC also contains a RING domain and its mammalian homolog BMI1 is reported to enhance the H2A ubiquitination activity of a dRING homolog in mammal. (Buchwald *et al.*, 2006; Li *et al.*, 2006b)

In addition to PRC1, dRING-associated factor (dRAF) has been reported to have an important role in H2A ubiquitination in Drosophila (Lagarou *et al.*, 2008). dRAF complex contains dRING, PSC, and dKDM2, a H3K36 de-methylase. In mammals, many other complexes that contain dRING homologs, RING1A/B, have been identified. RYBP-PRC1, composed of RYBP, RING1B, and MEL-18/BMI-1, PSC homologs in

mammals, is also reported to be responsible for H2A mono-ubiquitination (Tavares *et al.*, 2012). BCOR and E2F6 complexes also contain RING1A/B and the PSC homologs NSPC1 and MBLR in mammals, respectively (Gearhart *et al.*, 2006; Ogawa *et al.*, 2002; Sánchez *et al.*, 2007; Trimarchi *et al.*, 2001). Since the BCOR complex includes KDM2B, one of mammalian homologs of dKDM2 (Gearhart *et al.*, 2006; Sánchez *et al.*, 2007), it is possible that dRAF complex in *Drosophila* might be related to the BCOR complex.

Besides PRC1 and PRC2, the PhoRC complex is reported to be involved in PcG repression. So far, only two components, PHO/PHOL and dSFMBT, are known to form PhoRC complex (Klymenko *et al.*, 2006). PHO and its paralog PHOL are known to be able to bind directly to DNA. dSFMBT contains multiple MBT-domains and binds specifically to mono- and di-methylated H3K9 and H4K20. It is also notable that physical interactions of PHO with both PRC1 and PRC2 have been reported (Mohd-Sarip *et al.*, 2002, 2006; Wang *et al.*, 2004b).

1.1.2 TrxG proteins

The TrxG protein Trithorax (TRX) is a histone methyltransferase (HMTase) that contains SET domain and is reported to trimethylate H3K4 (Petruk *et al.*, 2006; Smith *et al.*, 2004). TRX can be cleaved in two moieties enzymatically by Taspase1, and the cleaved N- and C-terminal moieties (TRX-N and TRX-C, respectively) can interact each other. TRX-N contains multiple PHD domains and TRX-C contains the catalytic SET domain (Capotosti *et al.*, 2007; Hsieh *et al.*, 2003; Takeda *et al.*, 2006).

Two other TrxG proteins, Trithorax-related (TRR) and ASH1, also have HMTase activity. TRR has been reported to be the strongest candidate for the writer of H3K4me1 (Ardehali *et al.*, 2011; Herz *et al.*, 2012). Methylation target of ASH1 is much less clear; it has been variously reported to methylate H3K4, H3K9, H4K20 or H3K36 *in vitro* (Beisel *et al.*, 2002; Tanaka *et al.*, 2007). A recent *in vivo* study suggests that ASH1 is an H3K36 di-methylase (Dorighi and Tamkun, 2013).

In addition to the SET-domain-containing HMTases, many other histone modification enzymes are involved in the regulation of PcG/TrxG target regions. CBP, a versatile acetyl-transferase, is responsible for the acetylation of H3K27 in *Drosophila* (Tie *et al.*, 2009), although the function of CBP is not restricted to PcG/TrxG target regions. One more important key player for the regulation of H3K27 methylation is dUTX, which is the sole known de-methylase for H3K27 in *Drosophila*. Interestingly, CBP and TRR can form two different complexes with dUTX (Mohan *et al.*, 2011; Tie *et al.*, 2012).

Besides, several TrxG proteins including BRM, MOR and OSA are related to SWI/SNF ATP-dependent chromatin-remodeling (Mohrmann and Verrijzer, 2005), indicating not only histone modifications but also nucleosome remodeling are involved in the PcG and TrxG regulation. Notably, BRM is a component of dUTX/CBP complex (Tie *et al.*, 2012), implying a cooperative process of histone modifications and chromatin remodeling.

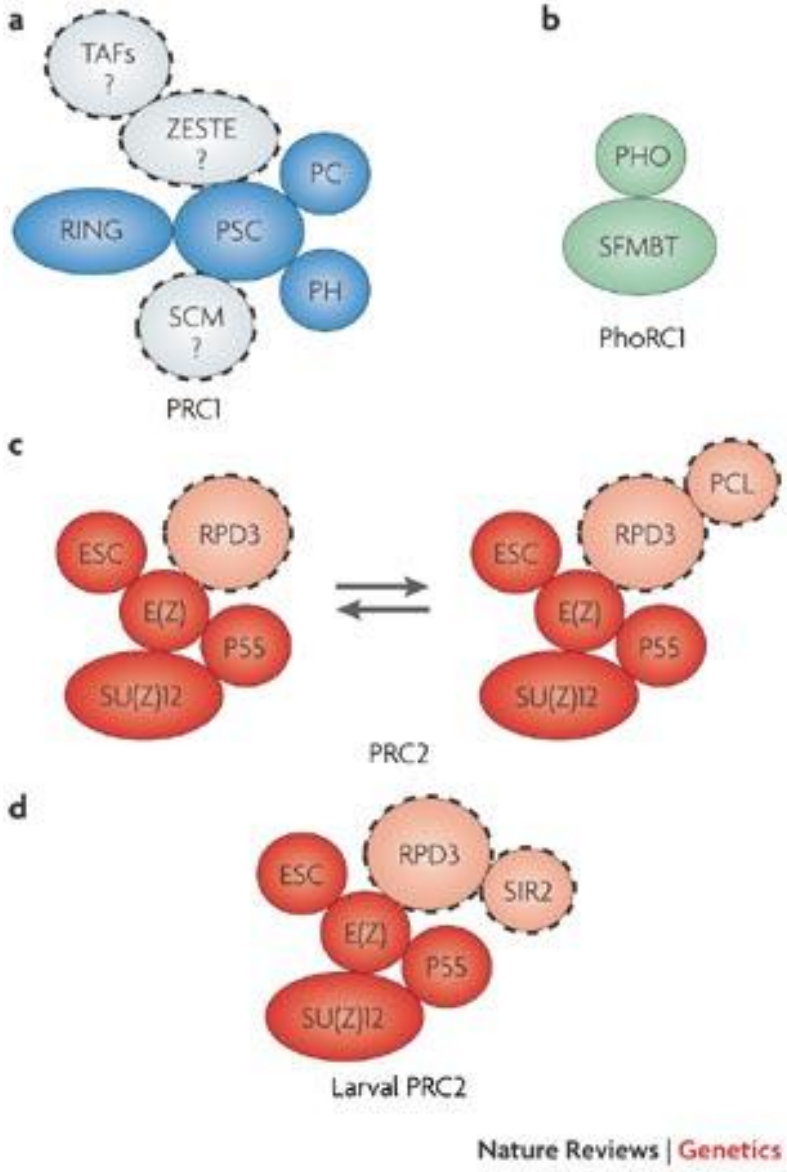


Figure 1.1 PcG complexes in *Drosophila*

Directly adapted from Schwartz and Pirrotta (2007). The solid spheres represent the core components of the PcG complexes, and the contacts between them reflect known interactions. Proteins that are known to associate with the complex but for which direct interaction partners are not known are shown as spheres with dashed borders. **(a)** The

PRC1 complex consists of a quaternary core that includes PC, PH, RING, and PSC. Besides,

Figure legend of Figure 1.1 continued.

a number of other proteins have been reported to be associated with the PRC1 complex. **(b)** For the PhoRC complex, only two directly interacting components, SFMBT (*Scm*-related gene containing four MBT domains) and PHO/PHOL (pleiohomeotic / pleiohomeotic-like), have been reported. **(c)** Several isoforms of PRC2 have been reported (Cao and Zhang, 2004), all of which have a core that consists of four proteins that are important for histone methyl-transferase activity (ESC, E(Z), P55 and SU(Z)12). In embryos, the 600 kDa PRC2 complex is the most abundant and includes the histone de-acetylase RPD3. Another 1 MDa embryonic complex contains both RPD3 and PCL (Polycomb-like). These two proteins seem to interact with each other directly. **(d)** In larvae, another PRC2 isoform was found to contain the histone de-acetylase SIR2.

Table 1.1 Components of the PcG / TrxG system.

Adapted from Schwartz and Pirrotta (2007).

Drosophila protein	Complex	Protein domains	Biochemical activity	Mammalian protein homologues
Polycomb group				
PC	PRC1	Chromodomain	Binding to trimethyl H3K27	NPCD, M33 (CBX2), CBX4, CBX6, CBX7, CBX8
PH	PRC1	SAM	?	PHC1, PHC2, PHC3
PSC	PRC1	RING	Cofactor for SCE	BMI1, MEL18
dRING(SCE)	PRC1	RING	E3 ubiquitin ligase specific to H2AK119	RING1A, RING1B
SCM	PRC1?	SAM, MBT, Zn-finger	?	SCMH1, SCML2
E(Z)	PRC2	SET	Methylation of H3K9, H3K27	EZH2, EZH1
ESC	PRC2	WD40	Cofactor for E(Z)	EED
ESCL	PRC2	WD40	Cofactor for E(Z)	EED
SU(Z)12	PRC2	Zn-finger	?	SUZ12
PCL	PRC2	PHD, Tudor	?	PHF19, MTF2 (M96)
PHO	PhoRC	Zn-finger	DNA binding	YY1, YY2
PHOL	?	Zn-finger	DNA binding	YY1, YY2
CG16975 (SFMBT)	PhoRC	MBT, SAM	Binding to mono- and dimethyl H3K9, H4K20	L3MBTL2, MBTD1
SU(Z)2	?	RING,	?	
SXC	?	?	?	?
ASX	?	PHD	?	ASXL1, ASXL2,
MXC	?	LA, RRM	?	Q9CUQ5
E(PC)	?	?	?	EPC1, EPC2
Trithorax group				
TRX	TRX-dCOMPASS-like, TAC1,	PHD, SET,	Methylation of H3K4	WBP7, MLL1
ASH1	?	SET, PHD, BAH	Methylation of H3K4, H3K9, H3K36, H4K20	ASH1L
ASH2	TRX/TRR/SET1-dCOMPASS-like	PHD, SPRY	?	ASH2L
TRR	TRR-dCOMPASS-like	SET,PHD,	Methylation of H3K4	MLL3/4
dUTX	TRR-dCOMPASS-like	JMJ,	De-methylation of H3K27	UTX, UTY
BRM	SWI/SNF	SNF2, HELICc, Bromo	ATP-dependent nucleosome sliding	SMARCA4
MOR	SWI/SNF	SWIRM, SAINT	Cofactor for BRM	SMARCC1, SMARCC2
OSA	SWI/SNF?	BRIGHT	?	ARID1B

Table legend of Table 1.1 continued.

ARID1B, AT-rich interactive domain 1B; ASH, absent, small, or homeotic discs; ASX, Additional sex combs; BRM, brahma; CBX, chromobox homologue; EED, embryonic ectoderm development; E(PC), Enhancer of Polycomb; ESC, extra sex combs; ESCL, extra sex combs like; E(Z), Enhancer of zeste; MLL, myeloid/lymphoid or mixed lineage leukaemia; MOR, moira; MTF, metal response element-binding transcription factor; MXC, multi sex combs; NPCD, neuronal pentraxin with chromodomain; PC, Polycomb; PCL, Polycomb-like; PH, polyhomeotic; PHC, polyhomeotic-like; PHF19, PHD-finger protein 19; PHO, pleiohomeotic; PHOL, pleiohomeotic-like; SCE, Sex combs extra; SCM, Sex comb on midleg; SFMBT, Scm-related gene containing four MBT domains; SU(Z), Supressor of zeste; SXC, super sex combs; TRR, Trithorax-related; TRX, Trithorax; WBP7, WW-domain binding protein 7; YY, Yin-Yang transcription factor.

1.2 PcG recruitment mechanisms

In *Drosophila,* PcG complexes bind to specific DNA elements called Polycomb response elements (PREs), which are known to be necessary and sufficient for the recruitment of PcG complexes (Muller and Kassis, 2006; Schwartz and Pirrotta, 2007) (Figure 1.2). Several PREs have been characterized by inserting transgenic constructs that contains reporter genes flanked by PREs, which get PcG-mediated repression (Brown *et al.*, 2005; Busturia and Bienz, 1993; Chiang *et al.*, 1995; Hodgson *et al.*, 2001; Horard *et al.*, 2000; Simon *et al.*, 1993). The core regions of PREs are depleted of histones (Kahn *et al.*, 2006; Mito *et al.*, 2007), and often contain binding sites for DNA-binding proteins including GAF, PHO/PHOL, PSQ, DSP1 and Zeste (Déjardin *et al.*, 2005; Faucheux *et al.*, 2003; Hodgson *et al.*, 2001; Horard *et al.*, 2000). Although the role of these factors is not so clear, it is possible that some of them might act as PcG complex recruiters. Especially PHO has been reported to interact with PRC1 and PRC2 (Mohd-Sarip *et al.*, 2002, 2006; Wang *et al.*, 2004b), and is a component of PhoRC complex, which largely overlaps with the binding sites of PRC1 and PRC2 (Oktaba *et al.*, 2008).

So far, it is not so clear how the PcG complexes can be recruited to PREs. One suggested model for PcG complexes recruitment is that PRC2 is initially recruited to PREs by DNA-binding proteins and tri-methylates H3K27, which in turn recruits PRC1 by its affinity specific to H3K27me3 (Cao and Zhang, 2004; Wang *et al.*, 2004b). In *Drosophila* polytene chromosomes, E(Z) inactivation can cause loss of PRC1 binding from many but not all of PREs, indicating that PRC1 binding to PREs is partially dependent on PRC2 and/or H3K27me3 (Czermin *et al.*, 2002; Ebert *et al.*, 2004). However, the core of many PREs, where the PRC1 binding is the strongest, shows

significantly lower H3K27me3 level than flanking regions due to the lack of histones (Kahn *et al.*, 2006; Mito *et al.*, 2007). It implies that initial PRC1 recruitment might not be mediated by H3K27me3 but by some DNA-binding proteins that are targeted to these histone-depleted core PREs.

Mammalian PcG recruitment seems to be much more complicated. Up to date, mammalian PREs are poorly defined. Instead, CpG islands have been suggested to have a role in PcG recruitment, as PRC2 binding is often overlapping with the CpG islands (Ku *et al.*, 2008). Recently, the BCOR complex, a non-canonical PRC1 complex that contains KDM2B, is reported to occupy the non-methylated CpG islands (Farcas *et al.*, 2012; He *et al.*, 2013; Wu *et al.*, 2013). In addition, many studies reported that non-coding RNAs are involved in PRC2 recruitment in mammalian systems (Beisel and Paro, 2011; Pandey *et al.*, 2008; Rinn *et al.*, 2007; Zhao *et al.*, 2008). Another study that used embryonic stem cells that lack EED (an ESC homolog in mammals) reported that the absence of H3K27me3 did not affect much the binding of PRC1 proteins (Vincenz and Kerppola, 2008), indicating that the PRC1 binding is independent of H3K27me3.

Since the observations for the PcG recruitment are varying, it is possible that different target regions might have different ways of recruitment including other mechanisms that are not fully characterized.

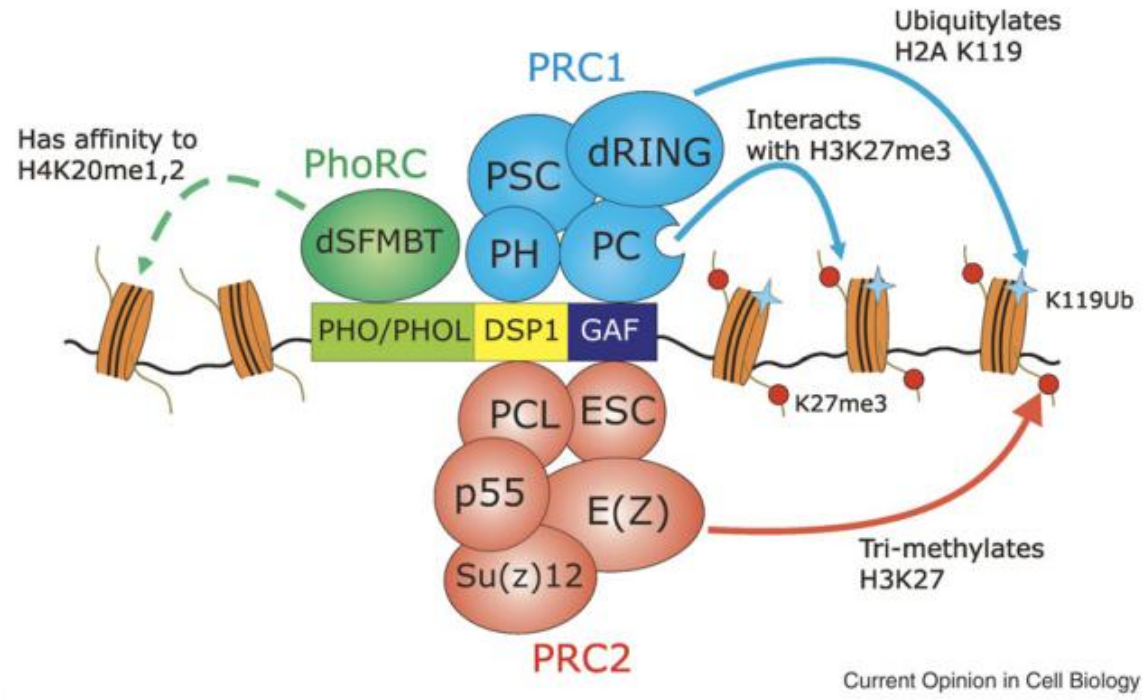


Figure 1.2 The PcG complexes binding on the PREs

Directly adapted from Schwartz & Pirrotta (2008). The three PcG complexes, PRC1, PRC2 and PhoRC are depicted in blue, red, and green spheres, respectively. In *Drosophila*, these complexes are thought to be recruited to the PRE by a combination of DNA binding proteins including PHO, DSP1, and GAF (rectangles). E(Z) methylates lysine 27 of histone H3, and the chromodomain of PC preferentially recognize the histone H3K27me3 mark. dRING mono-ubiquitylates the lysine 118 residue of histone H2A. dSFMBT is reported to have an affinity for mono- and di-methyl K20 of histone H4.

1.3 Multiple epigenetic states on PREs by PcG and TrxG proteins

Recent genome-wide location analysis revealed that PcG protein complexes can be found on hundreds of genomic loci including the Hox gene clusters (Boyer *et al.*, 2006; Bracken *et al.*, 2006; Lee *et al.*, 2006; Schwartz *et al.*, 2006). Furthermore, additional genome-wide analysis for TrxG proteins on several *Drosophila* cell lines identified multiple epigenetic chromatin states on PREs: repressed, active, balanced and void states (Schwartz *et al.*, 2010) (Figure 1.3).

In the repressed state, stable binding of PRC1 and PRC2 components can be observed at PREs with H3K27me3 widely distributed over a much larger surrounding region containing the gene body. Intriguingly, TRX binding is also observed at the core of the PREs in repressed state although the target regions are transcriptionally silenced.

In the active state, the genes are transcriptionally active and the binding of TrxG proteins including TRX-N, TRX-C and ASH1 can be found at the PREs. The most distinctive feature of the active chromatin state is that TRX-N and ASH1 are broadly distributed all over the gene body. It is notable that ASH1 can be found in active state but not in repressed state. H3K27ac, a histone marker antagonistic to H3K27me, also shows the same broad distribution with TRX-N and ASH1. H3K27ac can be found in many other actively transcribing non-PcG target gene regions, but the distribution in these non-PcG targets is focused on promoter regions.

The balanced state is an exceptional state that shows the mixed features of the active state and the repressed state; the binding of ASH1 and H3K27me3 enrichment can be observed at the same region, although ASH1 binding is confined to promoter regions, and the gene

is transcriptionally active. Recently, similar chromatin state, named “bivalent” state as H3K27me3 and H3K4me3 can be found at the same chromatin, was observed in mammalian embryonic stem cells (Bernstein *et al.*, 2006), although the transcription activity of the bivalent state genes is very weak. In the void state, PREs do not show the binding of either PcG proteins or TrxG proteins and the gene is transcriptionally inactive. These regions are defined as PREs because some of the cell lines showed the binding of PcG/TrxG proteins.

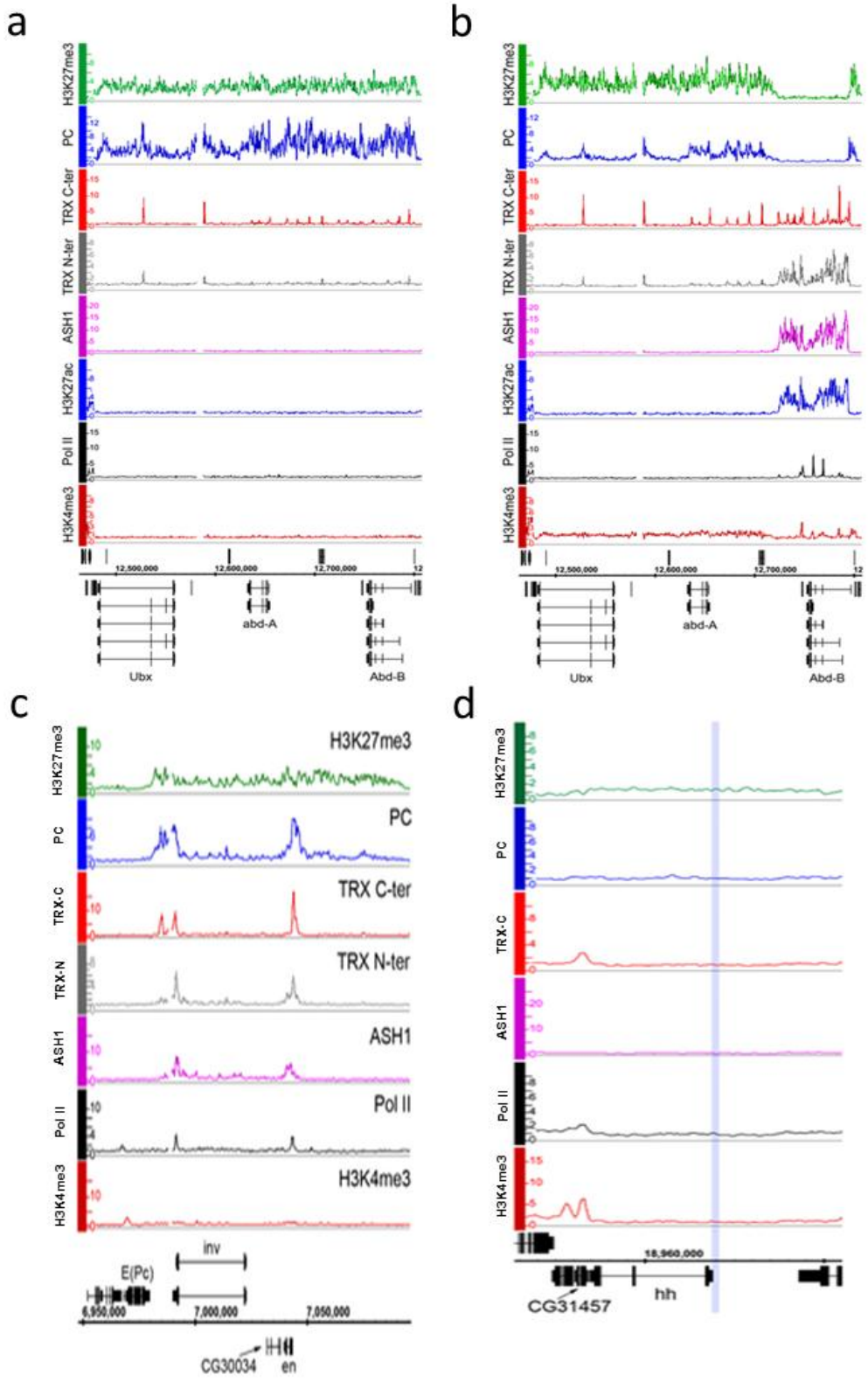


Figure 1.3 Four chromatin states regulated by PcG/TrxG proteins

Figure legend of Figure 1.3 continued.

Directly adapted from Schwartz & Pirrotta (2010). The distributions of PcG, TrxG, Pol II, and associated histone marks are shown. The Abd-B gene is in a repressed state in BG3 cells (a) and an active state in Sg4 cells (b). (c) the *inv-en* locus in a balanced state; H3K27me3 and PC are enriched in the *inv-en* locus, yet the binding of Pol II, TRX and ASH1 can also be observed. (d) The *hedgehog* (*hh*) gene is a well-known PcG target, but is in a void state in D23 cells, with no PC nor H3K27me3 but also lacking TRX, ASH1, Pol II, and H3K4me3.

1.4 PcG silencing mechanism : How do PcG complexes repress?

Multiple hypotheses have been proposed for the mechanisms of PcG-mediated repression. One general hypothesis for the PRC1-mediated repression mechanisms is that PRC1 reduces the accessibility of DNA in PcG repressed chromatin (Fitzgerald and Bender, 2001; McCall and Bender, 1996; Zink and Paro, 1995). Several studies suggested that PRC1 can inhibit the chromatin remodeling by TrxG-related SWI/SNF complexes (Francis *et al.*, 2001; Levine *et al.*, 2002; Shao *et al.*, 1999). In addition, PRC1-mediated transcription inhibition by chromatin compaction, which can impede access of polymerases or transcription factors, is suggested based on electron microscopic observations and *in vitro* experiments with reconstituted PRC1 complexes (Francis *et al.*, 2004; Grau *et al.*, 2011; King *et al.*, 2002). From their studies, PSC has been determined to be the main contributor for chromatin compaction as its positive charges can contribute to the compaction by strengthening the DNA-histone interaction.

On the other hand, some other studies suggested that PRC1 is interfering the transcription initiation steps (Dellino *et al.*, 2004; Lehmann *et al.*, 2012). They reported that RNA Polymerase II (Pol II), TATA binding protein, and transcription factor II D binding was still observed on the repressed state PREs, but further progression to elongation steps such as the Mediator recruitment was inhibited. It has been proposed that this inhibition is due to H2AK118 mono-ubiquitination (H2Aub1) by dRING (Lehmann *et al.*, 2012). Zhou et al. (2008) also suggested that H2Aub1 might inhibit Pol-II release at the early stage of elongation by preventing FACT recruitment at the promoter regions. Intriguingly, PR-DUB, a recently identified H2A-deubiquitinating complex (Scheuermann *et al.*, 2010), is reported to be required for PcG repression. Furthermore, synergistic effect of

dRING and PR-DUB mutants on loss of PcG target repression suggests that the writing and erasing of the H2Aub1 needs to be appropriately balanced for the PcG-mediated repression.

PRC2 and its methylation in H3K27 are also crucial in PcG silencing, as exemplified by a study that replacing wild-type histone H3 to the non-methylatable mutant H3K27R results in the failure to maintain PcG repression (Pengelly *et al.*, 2013). Still, its mechanistic contributions are not so clear. Although many lines of evidence support its function as a recruiting site for PRC1 as described before, it is still possible that methylation on H3K27 by itself might exert repressive function, as the methylation on H3K27 prevents the acetylation on it.

Importantly, some recent studies indicated that PRC2 complex alone can function as a repressor independent of PRC1 or vice versa. Iovino et al. (2013) reported that PRC2 is required for maintenance of oocyte fate by repressing expression of cell cycle genes, while PRC1 function is not required for oogenesis. Leeb et al., 2010 also showed that PRC1 and PRC2 can be recruited independently and repress their targets redundantly, at least in part, in mammalian embryonic stem cells. Their data show that a large set of genes is strongly de-repressed in embryonic stem cells with *Ring1B* and *Eed* double knock-out, but remains silenced in the absence of *Ring1B* or *Eed* alone. And the de-repression of subsets of genes showed different sensitivity to the absence of either *Ring1B* or *Eed*, suggesting that the repression in those genes is more dependent on one of the PcG complexes.

Although it is clear that there is a dependency between PRC1 and PRC2 for PcG repression in many PcG targets (Czermin *et al.*, 2002; Ebert *et al.*, 2004), these observations suggest their independent function for repression cannot be excluded.

These proposed diverse mechanisms for PcG-mediated repression are not necessarily mutually exclusive. Polycomb silencing might impact on multiple aspects of chromatin states or the transcription machinery so that the repression becomes more robust. Still, the details of the actual mechanisms for PcG-mediated repression remain to be further elucidated.

1.5 Transcription by RNA Pol II

The transcription cycle of RNA Pol II is largely divided into three steps: initiation, elongation, and termination. Transcription initiation is a complex process that involves multiple steps. The recruitment of Pol II and general transcription factors (GTFs) to a promoter results in the formation of the pre-initiation complex (PIC), DNA melting, and formation of the first few phosphodiester bonds of RNA (Nechaev and Adelman, 2011). The initiation step begins with the assembly of the PIC, which includes Pol II, the GTFs TFIIB, TFIID (which includes the TATA-binding protein), TFIIE, TFIIF and TFIIH. In addition to the GTFs, recruitment of Pol II to a promoter is greatly influenced by the Mediator complex, transcription factors, chromatin remodeling complexes and histone modifying enzymes (Li *et al.*, 2007; Roeder, 2005). After PIC assembly, the DNA at the transcription start site (TSS) is melted to form an open promoter complex, which is dependent on TFIIH (Kouzine *et al.*, 2013). Pol II subsequently synthesizes and releases short transcripts (abortive initiation). In the early elongation step when Pol II overcomes the abortive phase, it escapes from the promoter but might pause at a promoter-proximal region (promoter-proximal pausing). Release of paused Pol II finally leads to a productive elongation step.

During the transition from initiation to early elongation (promoter escape and pausing), the C-terminal domain (CTD) of the largest Pol II subunit is phosphorylated by the CDK7/Kin28 component of TFIIH specifically on Ser5 (Pol II Ser5p), which is thought to destabilize the interactions between Pol II and promoter-bound factors, facilitating promoter escape. The Ser5 phosophorylation also stimulates the co-transcriptional 5’ capping (Ho *et al.*, 1998). Then, the transition into productive elongation is triggered by

the recruitment of the P-TEFb, which phosphorylates Ser2 on the Pol II CTD. This relieves inhibition of early elongation caused by pausing factors like DRB sensitivity-inducing factor (DSIF) and negative elongation factor (NELF) (Cheng and Price, 2007; Marshall and Price, 1995; Renner *et al.*, 2001). The Pol II complex in the productive elongation step is remarkably stable, and can produce very long transcripts (Singh and Padgett, 2009).

Transcription termination can occur via distinct pathways. In most cases, termination is conducted by the canonical cleavage and poly-adenylation machinery. On the other hand, termination of several shorter transcripts, including snoRNA, utilize a poly(A) independent pathway by the Nrd1/Nab3/Sen1 complex (Steinmetz *et al.*, 2001, 2006).

Recently, transcription termination has been studied as an important step of transcriptional regulation that is coupled with the efficient recycling of Pol II. Several studies reported that some of the GTFs can remain associated with the promoter, forming a scaffold that enables re-initiation by Pol II during successive rounds of transcription (Hahn, 2004; Sarge and Park-Sarge, 2005). In yeast and mammals, the transcription initiation factor TFIIB is known to interact with the cleavage and poly-adenylation complex, which contains Ssu72, a Ser5 phosphatase. This interaction adjoins the promoter and terminator DNA to form a structure known as gene looping and allows Pol II recycling and rapid re-initiation (Ansari and Hampsey, 2005; Calvo and Manley, 2003; O'Sullivan *et al.*, 2004; Singh and Hampsey, 2007). Notably, Tan-Wong *et al.*, (2012) reported that the gene looping represses the antisense transcription from bidirectional promoters by Ssu72.

1.6 Potential importance of H3K27me2 in repressing pervasive transcription

Since PRC1 preferentially binds to H3K27me3, PcG functional studies have been focused mainly on H3K27me3 enriched regions. H3K27me1/2 are considered to be intermediate states and have had very few functional studies so far although they have a potential to function as repressive markers because they can prevent H3K27 being acetylated.

Especially, H3K27me2 might have an important biological role since about half of the total H3 is di-methylated and can be observed widespread throughout the genome. Ebert *et al.* (2004) performed polytene immunostaining and observed that H3K27me2 can be found broadly even at euchromatin and constitutes 40~60% of total H3 in *Drosophila*. Ebert et al. and other groups utilized comparative mass spectrometry techniques and observed high percentage of K27me2 (40~70%) in total histone H3 of fly and mammalian cells (Peters *et al.*, 2003; Robin *et al.*, 2007; Yuan *et al.*, 2011). Still, it is unclear where the H3K27me2 is enriched in the genome due to the lack of genome-wide localization studies.

Whether the widespread H3K27me2 has a repressive role is becoming a more interesting question since it might be relevant to the pervasive transcription across whole genome that has been reported recently in various species. Diverse genome-wide analyses of the eukaryotic transcriptome have revealed that the majority of the genome is transcribed, producing a large number of non-coding transcripts that arise from intergenic and intronic regions (Berretta and Morillon, 2009; Cheng *et al.*, 2005; Dinger *et al.*, 2009; Dutrow *et al.*, 2008; Li *et al.*, 2006a; Nagalakshmi *et al.*, 2008; Stolc *et al.*, 2004), suggesting the

possibility that most of the genome is capable of being transcribed. The human ENCODE project, for example, reported that 74% of regions analyzed by the project could be detected in primary transcripts identified by two independent technologies (Birney *et al.*, 2007). Manak *et al.* (2006) utilized tiling arrays to generate a series of transcription maps during *Drosophila* early embryonic development and their data indicates that most intergenic regions show transcription flags in various embryonic stages.

Pervasive transcription refers to the generation of a wide variety of RNAs that occurs pervasively across the genome, and the transcripts are distinct from those that have established functions like encoding proteins, tRNAs, rRNAs, snRNAs, and snoRNAs (Jensen *et al.*, 2013). Pervasive transcription includes several classes of transcripts from 5' and 3' nucleosome-depleted regions (NDRs) such as cryptic unstable transcripts (CUTs) and stable unannotated transcripts (SUTs) identified in yeast (Davis and Ares, 2006; Houalla *et al.*, 2006; Neil *et al.*, 2009; Wyers *et al.*, 2005; Xu *et al.*, 2009) and promoter upstream transcripts (PROMPTs), and promoter-associated small RNAs (PASRs) in mammals (Kapranov *et al.*, 2007; Preker *et al.*, 2011), many of which are the target of rapid degradation by exosomes. Recent studies also reported that a significant proportion of extragenic transcripts are originated from enhancers (Kim *et al.*, 2010; De Santa *et al.*, 2010). In addition, many diverse long intergenic non-coding RNAs are identified in mammals, most of which are not well-characterized.

Pervasive transcription is often associated with the regions that are highly accessible like promoter regions and enhancer regions, suggesting that most of it is likely due to leaky transcription initiation. Just as the NDR of a gene promoter allows the assembly of PIC, it is likely that the genomic regions that have higher DNA accessibility (e.g., active

enhancer regions) allow cryptic initiation. Chromatin not only condenses the genome, but also functions for the transcription regulation by regulating the DNA accessibility of transcription machinery and various transcription factors. Especially, promoter regions and enhancer regions are among the regions that are highly regulated by chromatin remodeling and histone modifications. If the abundant histone mark H3K27me2 has a repressive function, it might contribute to suppress the pervasive transcription at the chromatin level in more than half of the *Drosophila* genome. Some recent studies indicate that PRC2 might have a repressive role in extended regions of the *Drosophila* genome besides the H3K27me3 enriched regions. Recently, Chopra *et al.* (2011) reported that *esc* mutant embryo exhibited Pol II binding increase in unexpectedly large number of genes. Intriguingly, a significant proportion of those genes were considered to be non-PcG target genes as they are not enriched by H3K27me3. Although the study speculated that this phenomenon might be an indirect effect of PRC2 inactivation, it is possible that those genes are repressed by H3K27me2. In addition, Lee *et al.* (2010) reported the repressive role of H3K27me2 although this study focused only on a limited number of genes.

Chapter 2. Genome-wide intergenic repression by PcG proteins

2.1 Introduction

PcG proteins have been studied mainly as master epigenetic regulators for several hundreds of developmental regulatory genes that are associated with strong H3K27me3 enrichment. However, the biological roles of H3K27me2, which is also methylated by E(Z) and covers more than half of the *Drosophila* genome, are largely uncharacterized. Whether the H3K27me2 might have a repressive role would be a very important question since it might function as a suppressor of the pervasive transcription that can happen all over the genome.

To examine the repressive role of the methylation in H3K27 more expediently, we recently established *E(z)* temperature sensitive cell line (EZ2-2) which gives a unique opportunity to study the repressive function of PRC2 by transient inactivation which allows direct comparison before and after E(Z) inactivation. The EZ2-2 cell line carries homozygous $E(z)^{61}$ temperature sensitive mutation in which the H3K27 methyl transferase activity of E(Z) become significantly reduced at non-permissive temperature (31˚C) (see Materials and Methods for more details).

Loss of PRC1 or PRC2 function could also be obtained by RNAi treatments against PcG genes. However, it was very hard to observe strong de-repression effects on PcG target genes in some cell lines including S2 cells from modENCODE and Bg3 cells, which turned out to have high levels of *E(z)* and *Pc* expression, respectively. More importantly, even if we could get strong knock-down of *E(z)* enough to observe strong PcG target de-repression in another S2 cell line, it gives a paradoxical effect that the de-repression also

increases the *E(z)* mRNA level, as the loss of PRC2 function caused the de-repression of most genes as described later (see Results). In addition, very efficient knock-down of PRC1 and PRC2 components was difficult to achieve consistently perhaps because the effect of the knockdown negatively affects cell growth, which might have selective effects on the population if treated for a long time. So the E(Z) inactivation by the temperature sensitive mutant was a very effective and consistent way of studying the function of the H3K27 methylation, supported in part by RNAi knock-down studies.

Genome-wide mapping analysis of various histone modifications including H3K27me2/3 and total RNA-seq analysis revealed that the H3K27me2 is indeed present all over the genome, and E(Z) inactivation entailed strong decrease of H3K27me2/3 globally and caused global de-repression including most of intergenic regions and non-PcG target genes. The de-repression was accompanied by the gaining of active histone marks in most of the genome. In a nutshell, our results suggest that methylation on H3K27 by PRC2 increases the threshold of transcription all over the genome so that extensive pervasive transcription is suppressed.

2.2 Materials and Methods

E(z)^{61} mutant

The *E(z)*^{61} allele, also known as *E(z)*^{s2}, contains a single G to A transition that results in a Cys to Tyr substitution on the 603th amino acid (C603Y), and this point mutation lie within the Cys-rich region which is adjacent to the SET domain of the E(Z) protein (Carrington and Jones, 1996). In the past studies on the *E(z)*^{61} mutant embryos and larvae, 18-21℃ has been considered as a permissive temperature and 29℃ as a non-permissive temperature (Carrington and Jones, 1996; Jones and Gelbart, 1990; Ketel *et al.*, 2005). The flies bearing homozygous *E(z)*^{61} is viable and diplay wild-type morphology at 18℃ and lethal at 29℃ although approximately 1/5 to 1/3 survive to adulthood, indicating leaky activity of *E(z)*^{61} at 29℃ (Jones and Gelbart, 1990). *E(z)*^{61} embryos at 18℃ shows comparable levels of H3K27me2/3 with wild type embryos, while the levels are dramatically reduced in *E(z)*^{61} embryos at 29℃ (Ketel *et al.*, 2005). It is also reported that PRC2 core subunit assembly has not been affected by this mutation, but the mutant complex purified from the Sf9 cells grown at 29℃ showed at least 8~10 times reduced level of HMTase activity (Ketel *et al.*, 2005).

PcG mutant cell line establishment

PcG mutant cell lines are established by introducing a constitutively active form of Ras, Ras^{V12}, for promoting proliferation of the cells bearing PcG mutants (*Pc*^{3} and *E(z)*^{61}) as the way described in Simcox *et al.* (2008). Briefly, Ras^{V12} with an upstream UAS promoter (UAS-Ras^{V12}) was introduced to the PcG mutant fly lines by series of crosses

(Figure 2.1). In parallel, Act5C-GAL4 was also introduced to the PcG mutant lines independently. The resulting fly lines were crossed to generate embryos containing both UAS- Ras^{V12} and Act5C-GAL4 with homozygote PcG mutant background (Figure 2.2). Establishing stable cell lines from the embryos was conducted by the Amanda Simcox Lab as described in Simcox *et al.* (2008). I thank Dr. Amanda Simcox for helping us with generating stable PcG mutant cell lines. *Psc* and *Su(z)12* mutant cell lines were generated by Dr. Yuri Schwartz and Dr. Tanya Kahn. I thank them for allowing me to use those two cell lines before publication.

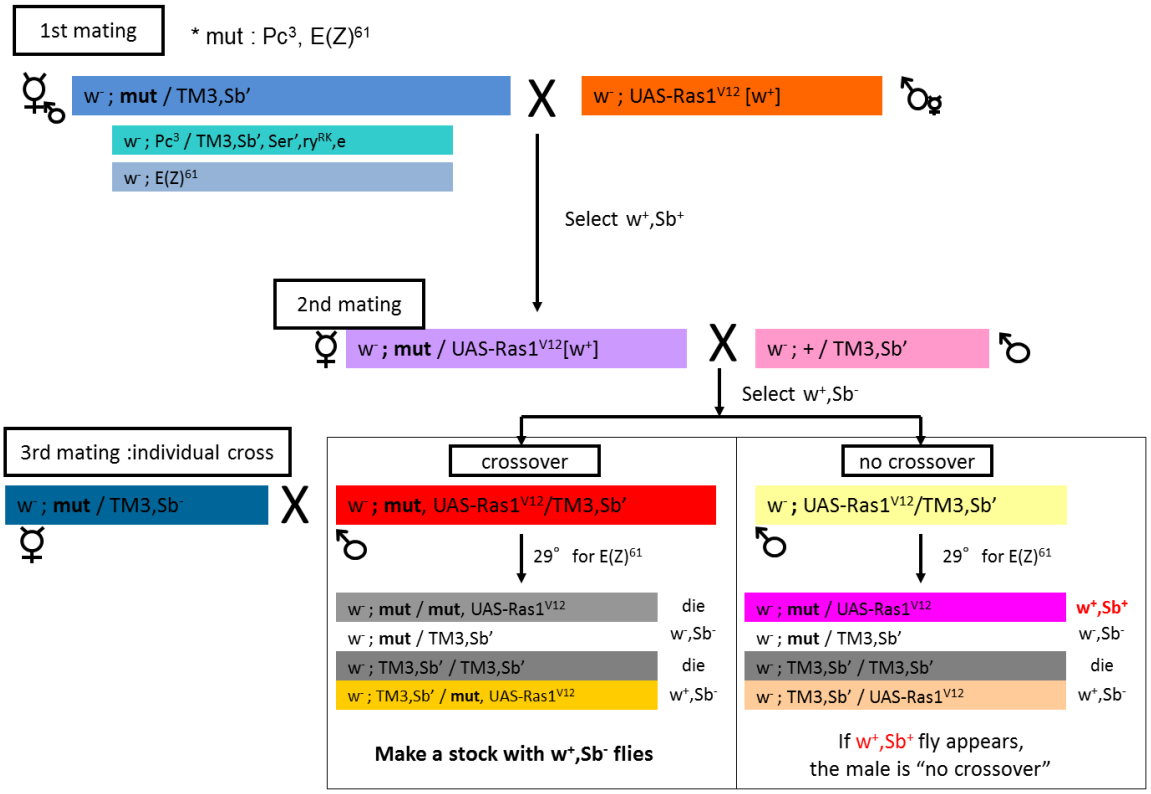


Figure 2.1 Generating fly lines for establishing PcG homozygous mutant cell lines

PcG mutant fly lines are crossed with the UAS-Ras1^{V12} line to generate the fly lines that contain both PcG mutants (*Pc*^{3} or *E(Z)*^{61}) and UAS-Ras1^{V12}. Same strategy has been used for generating the fly lines that contain both PcG mutants and Act5C-Gal4.

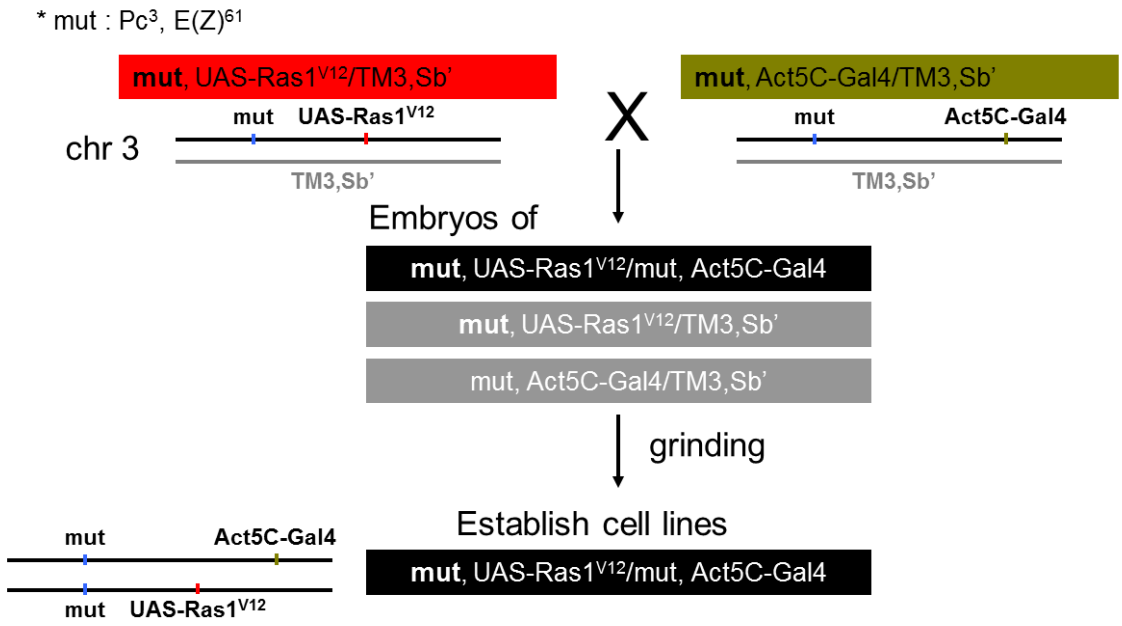


Figure 2.2 Establishing PcG homozygous mutant cell lines

The embryos that have the homozygous PcG mutations combined with UAS-Ras1^{V12} (red) and Act5C-Gal4 (green) were generated to establish the PcG homozygous mutant cell lines. The embryos that contain both UAS-Ras1^{V12} and Act5C-Gal4 are expected to have a full advantage of proliferation, so the established cell lines would be dominated by the cells from the homozygous mutant embryos (black).

E(z) temperature sensitive cell line (EZ2-2)

EZ2-2 cells, which are generated from homozygous $E(z)^{61}$ mutant embryos**,** grow well in firmly attached confluent monolayers at 25℃ (doubling in every 3-4 days) and grow very slowly at 18-21℃. In 31℃, the growth become slow after 4 days and cells start to form big clumps and eventually almost stop growing after 12 days but still can be maintained up to 24 days if passaged without dilution (monolayers can hardly be maintained after 12 days; clumps can be attached well in new media). EZ2-2 cells are extremely sensitive to the cell density so the well-established monolayers need to be maintained carefully. Even at 25℃, too low density of these cells can cause much slower cell division and formation of clumps, and too high density also causes large clumps which are hard to be dissociated once formed.

Ras3, a wild-type control cell line that are generated by the same way as the EZ2-2 cell line, grows well in monolayers at 25℃ and even at 31℃ if the cell density is appropriately maintained. Ras3 also can form clumps and be detached if overgrown.

Extensive tests for determining the non-permissive temperature condition have been conducted and 31-31.5℃ was the maximum temperature which the cells can be maintained reproducibly. Even occasional temperature raise over 31.5℃ due to unstable temperature controlling could cause the failure to maintain the cell culture.

Cell culture

The S2 cells were grown at 25^0C in 25 cm^2 T-flasks in Schneiders *Drosophila* medium (GIBCO) supplemented with 10% Fetal Bovine Serum (GIBCO), 100U/ml of Penicillin, 100ug/ml of Streptomycin sulfate and 292ug/ml of L-glutamine. The S2 cell line was obtained from Dr. Dessislava Dimova's Lab.

All of the PcG-mutant cell lines that have been established by the introduction of Ras^{V12} were grown in the same media that has been used for S2 cells. For passaging the adherent PcG-mutant cell lines and the wild-type Ras3 cell line, the media was completely removed and the cells were washed with Dulbecco's phosphate buffered saline (DPBS), then 1 ml of 0.05% trypsin was added per 25 cm^2 and incubated at R.T. for 3 to 4 minutes to make the cells detach from the bottom of the plate. After mild tapping, the 5ml media containing 10% FBS was added by gently blowing off with a pipette. Collected cells were centrifuged by 1500 rpm for 3 minutes and supernatant was removed. After gently tapping the tip of the tubes, cells were resuspended to appropriate amount of new media and transferred to new plates.

RNAi

The RNAi was performed as described by Clemens *et al.* (2000) with minor modifications. Cells at approximately 70% to 80% confluency were subjected to three consecutive treatments with corresponding dsRNA at a ratio of 50-100ug of dsRNA per T25 flask. Since it takes time for the knock-down effects on the PcG-encoding genes to be reflected to histone modifications, most of the RNAi experiments were performed at

least three consecutive treatments. The cells were grown in regular media for 4 days between each treatment. Four days after the third treatment the cells were harvested and used for preparation of total RNA and total nuclear proteins. For the knock-down experiments in non-permissive temperature with the EZ2-2 cell line, one additional round of the RNAi treatments was conducted 4 days before the temperature shifting to minimize the de-repression effect on the target genes by E(Z) inactivation.

Chromatin immunoprecipitation (ChIP) assays

For the crosslinked chromatin preparation, collected cells were crosslinked by adding of 36% fomaldehyde (Sigma) to a final concentration of 1.8%, and incubated for 10min at 25℃ while shaking gently. The reaction was stopped by adding the glycine to a final concentration of 0.125M on ice. The crosslinked cells were immediately pelleted at 4℃, and washed once each with ice cold 1× PBS, and ChIP wash buffer A (10mM HEPES pH7.6; 10mM EGTA pH8.0; 0.5mM EGTA pH8.0; 0.25% Triton X-100) for 10min, and then ChIP wash buffer B (10mM HEPES pH7.6; 200 mM NaCl; 1mM EDTA pH 8.0; 0.5mM EGTA pH 8.0; 0.01% Triton X-100) for 10min. The washed fixed cells were then pelleted, and frozen in liquid nitrogen after removing supernatant, and stored in -80℃. The fixed cells were washed twice with ice cold TE buffer (10mM Tris-HCl; 1mM EDTA pH8.0) and resuspended into TE buffer with 0.1% SDS to a concentration of 1×10^8 cells/mL, and then subjected to sonication with Biorupter (Diagenode). The power was set as “high”, and the sonication was divided by 5 sessions, each session with 0.5min ON and 0.5min OFF for 5min long, and changed the iced water in the chamber to prevent

overheating during each sessions. After sonication, the resulting lysate was further adjusted to RIPA buffer (140 mM NaCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 1% Triton X-100, 0.1% SDS, 0.1% sodium deoxycholate) and cleared by 5 min centrifugation at 14,000 rpm, divided in 400 µl aliquots and stored at -80℃ after frozen in liquid nitrogen. The samples were kept in ice during the whole procedure.

For immunoprecipitations, 400 µl of lysate were precleared by incubation with Sepharose beads conjugated to Protein A (Sigma). The cleared lysate was further incubated with 3-5 µg of the appropriate antibodies overnight at 4℃. The antibody complexes were precipitated by incubation with Protein A-Sepharose beads (Sigma or GE Healthcare) for 3h at 4℃. The beads were washed five times with 1 ml RIPA, once with 1 ml LiCl buffer (250 mM LiCl, 10 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.5% NP-40, 0.5% sodium deoxycholate), twice with 1 ml TE (10 mM Tris-HCl pH 8.0, 1 mM EDTA). After washing steps, RNase A was added to a final concentration of 50ug/mL and incubated at 37℃ for 30min. For reverse crosslinking, SDS (final 0.5%), NaCl (final 140mM) and Proteinase K (final 0.5mg/mL) were added and incubated overnight at 37℃, then transferred to 65℃ for 6hrs. The reverse-crosslinked DNA was then phenol-chloroform purified, and precipitated by 100% EtOH. After incubation overnight at -20℃, the immnunoprecipitated DNA pellet was washed by fresh 70% EtOH and dissolved either in 150ul distilled water for quantitative PCR (qPCR) analysis or in 12-50ul distilled water for ChIP-chip or ChIP-seq experiments.

Western blot analysis

Total nuclear protein was isolated by first lysing cells in hypotonic buffer containing 10% sucrose, 10mM Tris-Cl pH8.0, 10mM NaCl, 3mM $MgCl_2$, 2mM DTT , 0.2% Triton X100, 1mM PMSF and 1x Complete Protease Inhibitor (Roche) on ice, followed by centrifuge at 1500g for 3min at 4℃ and washing of the nuclear pellet with Nuclear Buffer (10mM Tris-Cl pH8.0, 1mM EDTA pH8.0, 140mM NaCl, 2mM DTT, 1mM PMSF and 1x Complete Protease Inhibitor). After resuspension into the Nuclear Buffer, the nuclei were further lysed with Sample Buffer (12mM Tris-HCl pH6.8, 5% glycerol, 0.4% SDS, 2.9mM 2-mercaptoethanol, 0.02% bromphenol blue) by strong pipetting and boiled for 10 min. Serial dilutions of protein samples were loaded on 14% SDS polyacrylamide gel, separated by electrophoresis, transferred to a PVDF membrane and detected by incubation with primary antibodies (1:2000 dilution) in Blocking Buffer (1% blocking reagent (Roche), 0.05% Tween-20 in 1x PBS) at 4℃ overnight with gentle shaking, and secondary antibodies conjugated with alkaline phosphatase. For the H2Aub1 antibody, the incubation buffer consisting of 5% BSA, 1x TBS, and 0.1% Tween-20 was used.

Antibodies

Antibodies used for ChIP and western blot analysis are as follows: anti-H3K27me2 (Diagenode pAB-046-050), anti-H3K27me3 (Abcam ab6002 for ChIP, Millipore 07-449 for western), anti-H3K27ac (Abcam ab4729), anti-H3K4me1 (Abcam ab8895), anti-H2AK118ub1 (Cell signaling D27C4), anti-PC (Pirrotta Lab, affinity purified), and anti-E(Z) (Pirrotta Lab, affinity purified).

RNA/DNA isolation and Reverse transcription (RT)

Total RNA was isolated using the TRIzol reagent according to the manufacturer's instructions (Invitrogen). Isolated total RNA was treated with RNase-free DNase I (Roche) for 20min at 37℃. DNase I was heat-inactivated by 10min incubation at 75℃, and further removed by phenol extraction. DNA isolation from the same TRIzol-treated samples was performed according to the manufacturer's instructions (Invitrogen) with minor modifications. To estimate the relative cell amount between multiple samples used for the RNA extraction, isolated DNA was quantified by qPCR with appropriate dilution to make the Ct value between 20 and 30.

Reverse transcription was performed with 1ug of total RNA using the first strand cDNA synthesis kit with random hexamers (0.2ug per reaction) according to the manufacturer's instructions (GE healthcare). Synthesized cDNA was purified by Qiaquick PCR purification kit (Qiagen).

Quantatitive PCR

Quantitative PCRs were done to quantify the copies of DNA in cDNA libraries from RT-PCR or immunoprecipitated material from ChIP assays. 2.5ul of the ChIP DNA (1/60 of total immunoprecipitated material) or purified RT-PCR product was used for each PCR reaction in a 10ul mix containing 10mM Tris-Cl pH8.3, 40mM KCl, 2.5mM MgCl2, 0.2mM dNTP, 100nM of target-specific primers, 1x EvaGreen dye (Biotium) and 1 unit of AmpliTaq Gold (Applied Biosystems). PCR was performed in 96-well plates with the Mx-3000P (Stratagene) or Realplex 2 (Eppendorf). 7-point standard curves were used for

each primer pairs by amplification of serial dilution of genomic DNA (four-fold dilutions for RT-qPCR) or the input DNA (two-fold dilutions for ChIP-qPCR) isolated from an aliquot of lysate that did not undergo immnunoprecipitation. Samples are appropriately diluted to be in the linear range of the standard curves. Primers were designed to amplify the regions between 100 and 200 base pairs of the target sequences. Primer sequences are available upon request.

Genome-wide distribution data overview

The genome-wide profiling of H3K27me3, H3K27ac, and H3K4me1 in EZ2-2 cells at permissive temperature (25℃) and non-permissive temperature (31℃) was performed by combining chromatin immunoprecipitation and DNA-microarray analysis (ChIP-chip) using Affymetrix GeneChip *Drosophila* Tiling 1.0R arrays. Chromatin immunoprecipitation followed by high-throughput DNA sequencing (ChIP-seq) was performed for generating the genome-wide distribution maps of H2Aub1 and PC in EZ2-2 at permissive and non-permissive temperature. ChIP-chip data from Sg4 cell line was generated by Dr. Yuri Schwartz, Dr. Tanya Kahn, and Dr. Vincenzo Pirrotta (PC, H3K27me3, Pol II ChIP-chip from Schwartz et al., 2010; H3K27me2 data was not published elsewhere). I thank them for allowing me to use their data. Genome-wide maps of H3K27me1, H3K27me2, H3K27me3, histone H3, PC, and Pol-II in Bg3 cells were obtained from the modENCODE project website (http://modencode.org/) (Celniker *et al.*, 2009).

For the genome-wide transcriptional profiling of EZ2-2 cells, high-throughput RNA sequencing (RNA-seq) with ribosomal RNA (rRNA)-depleted total RNA from EZ2-2 cells at permissive and non-permissive temperature was performed. RNA tiling array data for the genome-wide transcriptome profiling in Bg3 cells was obtained from modENCODE project (Celniker *et al.*, 2009).

For the genomic coordinates, *Drosophila melanogaster* Apr. 2006 (BDGP R5/UCSC dm3) assembly was used for the analysis of the genome-wide data.

Library construction for ChIP-chip

Preparation of labeled probes for ChIP-chip was performed as described in Schwartz *et al*. (2006), with the following modifications. Briefly, one-third of the immunoprecipitated material and, in parallel, 50ng of the ChIP input DNA were amplified using the WGA amplification kit (Sigma) by following the manufacturer's instructions with a few modifications. The fragmentation step was omitted since the amplified materials are fragmented in a separated step after amplification as described later. Eighteen cycles of amplification were performed, and the resulting PCR product was purified using QIAquick PCR purification kit (QIAGEN), and 2 µg of amplified ChIP material or Input DNA were fragmented to a size of 25-200 bp (peak at 75bp) by incubation with 0.5 U of DNase I (Roche) for 5-15 min at 25˚C. After heat inactivation of DNase I by 10min at 95˚C, the fragmented DNA was end-labeled with bio-11-ddATP (Perkin Elmer) using recombinant Terminal Transferase (Roche). The labeled probes were hybridized to GeneChip *Drosophila* Tiling 1.0R arrays (Affymetrix) in 0.1 M MES pH 6.6, 3 M Tetramethylammonium Chloride (TMA), 40 pM of control oligo B2 (Affymetrix), 100 µg/ml of sonicated salmon sperm DNA and 0.02% Triton X100.

ChIP-chip data processing

The TiMAT T2 1.2.0 open source tiling microarray analysis software package was used for primary ChIP-chip data processing. The microarray results are computed in terms of the ratio between the ChIP value and the input DNA value from a control microarray. The intensity ratios between ChIP and input DNA were calculated to generate the

genome-wide distribution maps. The raw intensity values were further smoothed by calculating mean ratio over a sliding window of 675 bp along the genomic DNA sequence. Integrated Genome Browser (Affymetrix) was used for visual inspection of the results.

Library construction for Next Generation Sequencing (NGS)

For the RNA-seq libraries, rRNA was depleted from the 5ug of DNase I-treated total RNA by RiboZero kit (Epicentre). After following purification steps by Agencourt RNAclean XP beads (Beckman Coulter), approximately 100-200ng of rRNA-depleted total RNA was recovered.

ChIP samples (5-10ng in 50ul distilled water) and rRNA depleted total RNA samples (100-200ng in 15ul RNase-free water) were submitted to Rutgers University Cell and DNA Repository (RUCDR) to construct the libraries for the sequencing. TruSeq ChIP Sample Preparation Kit (Illumina) was used for generating ChIP-seq libraries by following the manufacturer's instructions with a few modifications. Briefly, the size selection step was performed by selecting 250-500bp range after adaptor ligation, and the following amplification step was done with 15 cycles. For RNA-seq library construction, Ovation RNA‑Seq System V2 kit (NuGEN) was used. With the constructed libraries, 100bp paired-end sequencing was performed by Illumina HiSeq technology.

NGS data process and analysis

Sequence tags from ChIP-seq libraries were mapped uniquely in its best strata of alignment to *Drosophila* genome (FlyBase 5.22) using bowtie v0.12.7 (Langmead *et al.*, 2009). Sequence alignment information is further processed by SAMtools v0.1.18 (Li *et al.*, 2009) to generate a pileup of read bases which is used for obtaining the genome-wide profiling data format compatible with Integrated Genome Browser (Affymetrix). RNA sequence tags are mapped by the genesifter pipeline (www.genesifter.net), which utilizes BWA WTS PE (GATKv3) v2 (Li and Durbin, 2009) for the transcriptome alignment. Sequence tags that are multiply matching to the repetitive sequences were excluded in all ChIP-seq and RNA-seq results.

Normalization of the genome-wide data from sequencing (ChIP-seq and RNA-seq) between different conditions (25˚C and 31˚C) was done by calculating the correction factors based on ChIP-qPCR or RT-qPCR results. At least three regions (five regions for RNA-seq) were used for the normalization.

Defining PcG targets and non-PcG targets

Annotated FlyBase genes were classified as PcG targets and non-PcG targets based on the levels of PC and H3K27me3 for each cell line (Figure 2.3). All genes were sorted by maximum binding values of PC (PC_{max}) at gene bodies and 1kb flanking regions, and the genes with the PC_{max} and H3K27me3 values higher than the 2 standard deviations above the mean were defined as PcG target genes. Genes with the PC_{max} values lower than the average PC_{max} value of all genes were defined as non-PcG target genes. Non-PcG target

genes were further classified according to the expression level. Silent genes were defined as the genes that have no detectable RNA-seq tags. The same classification was applied to the intergenic regions.

EZ2-2 RNA-seq data from permissive temperature was used for the classification of coding/intergenic non-PcG target regions of EZ2-2 and Sg4 cell lines. For Bg3, published modENCODE whole-genome expression data from tiling arrays was used for obtaining the expression levels of coding regions.

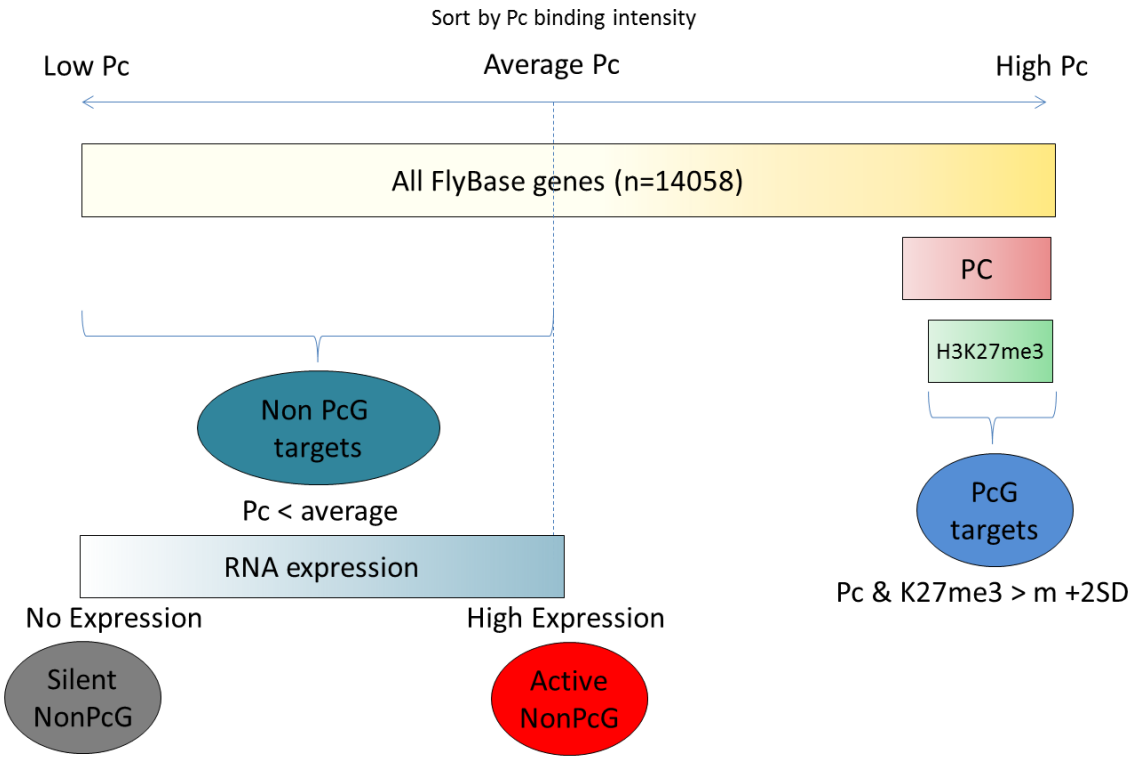


Figure 2.3 Classification of PcG targets and Non-PcG targets

For each cell line, annotated FlyBase genes are classified as PcG targets and non-PcG targets based on the levels of PC and H3K27me3 in genome-wide data, and non-PcG targets are further sub-grouped by their expression levels. The same classification strategy was applied to the intergenic regions.

Determining H3K27me2 enrichment

By analyzing the distribution of ChIP-chip signals, we could observe that two distinct bell-curved distributions are composing the whole signal distribution; one from H3K27me2-enriched regions and the other from low H3K27me2 regions (Figure 2.4). The highest 5% point value (1.06) from the distribution of low H3K27me2 regions was used as a threshold for determining enrichment of H3K27me2 in Bg3. For the Sg4, since the median value of the all signals was above the distribution of the low H3K27me2 regions, the median value (1.12) was used for determining enrichment.

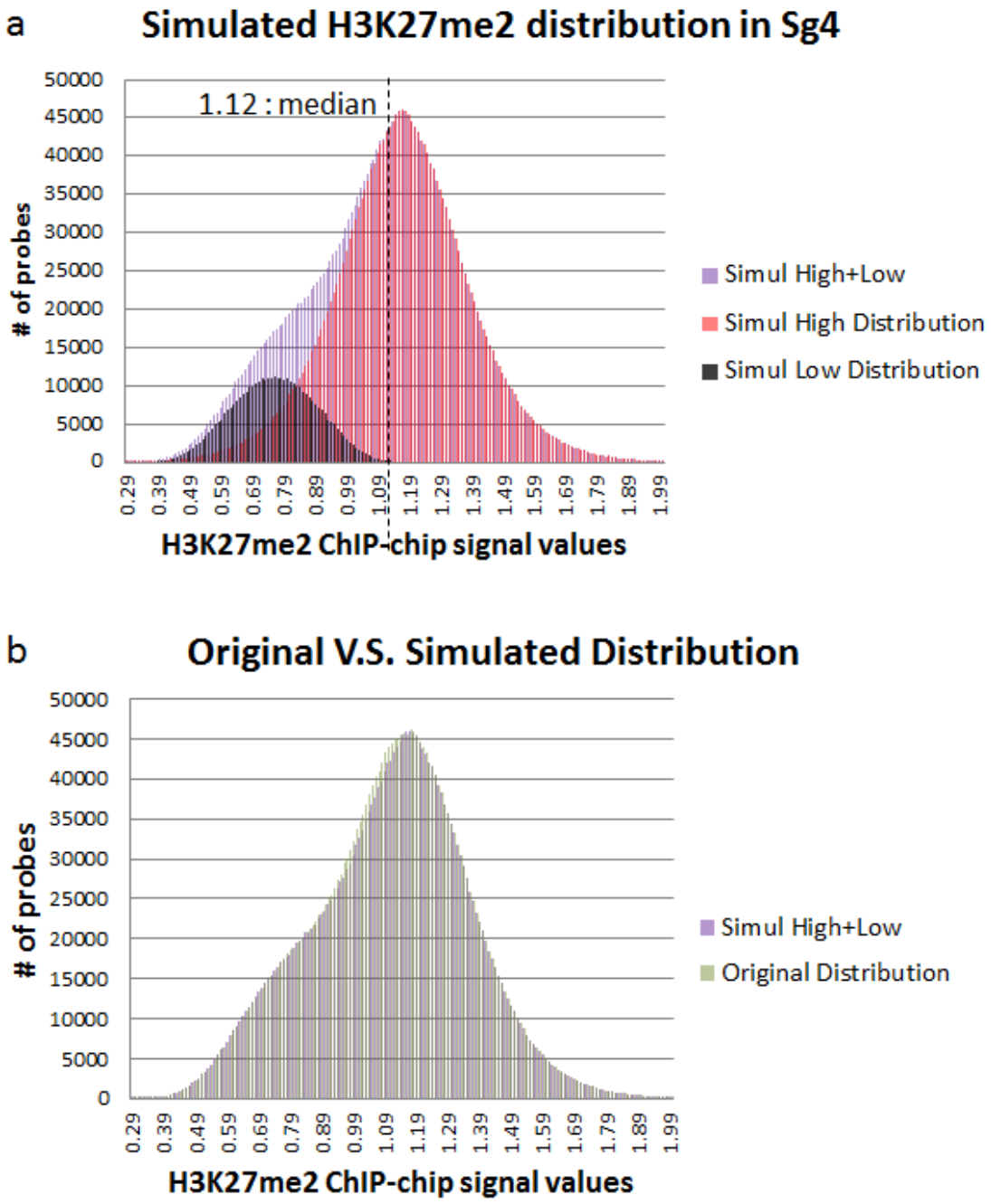


Figure 2.4 Determining the enrichment of H3K27me2

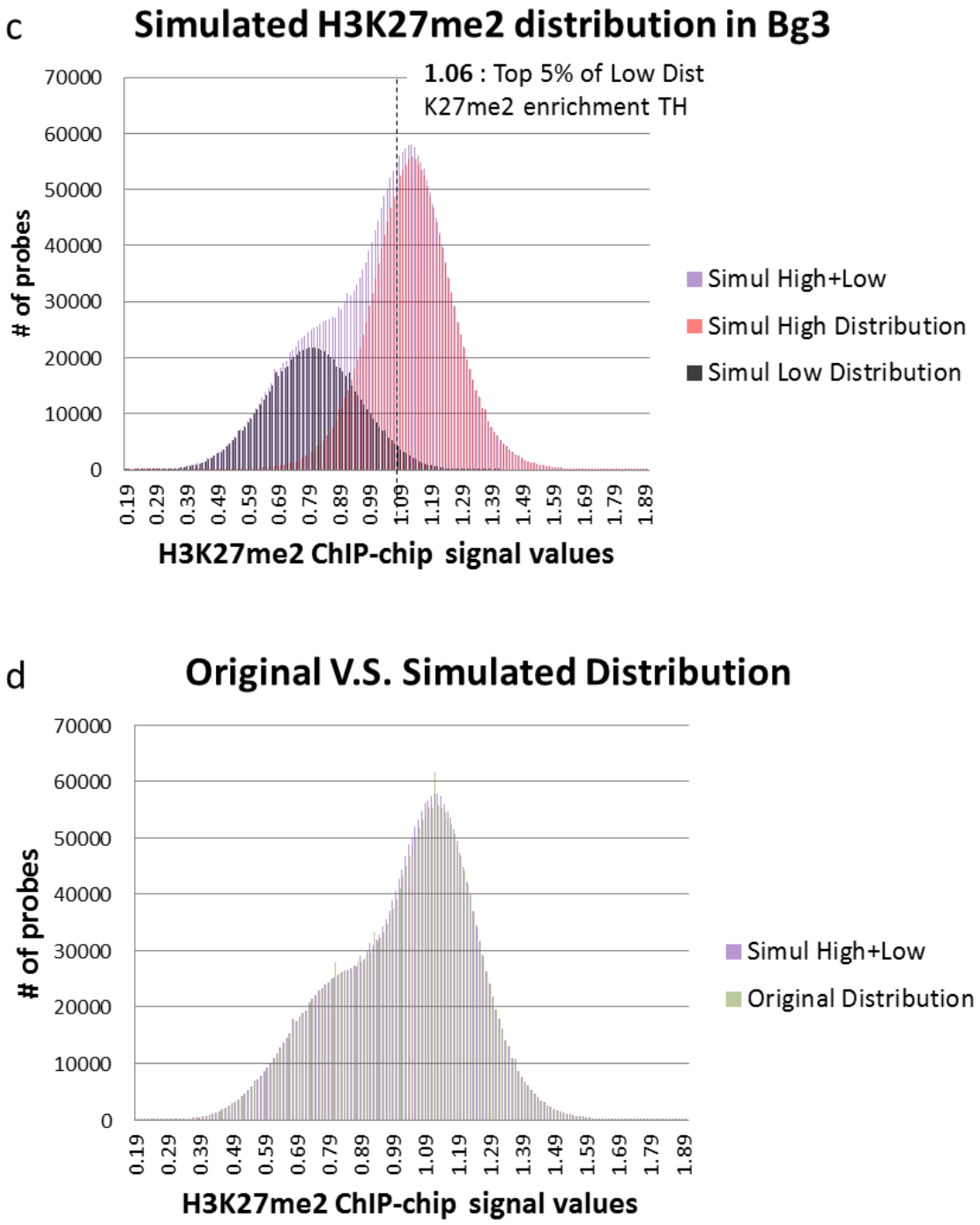


Figure 2.4 Determining the enrichment of H3K27me2 (Continued)

Figure legend of Figure 2.4 continued.

The overall distribution was obtained by collecting the IP/input signal values from all probes on the microarray. Two hypothetical symmetric bell-curved distributions were simulated based on the overall distribution for each cell line (a for Sg4 and c for Bg3). The enrichment levels of H3K27me2 were determined based on these distributions (dotted lines in a and c; see text for details). The combined distribution of the two simulated distributions is compared to the original distribution (b for Sg4 and d for Bg3).

2.3 Results

2.3.1 Widespread distribution of H3K27me2 and a ubiquitous low level of H3K27me3

In order to investigate the distribution of H3K27me2/3, we performed genome-wide mapping of H3K27me2 and H3K27me3 on Sg4 cell line (Y. Schwartz, T. Kahn, and V. Pirrotta) and the EZ2-2 cell line, which bears a homozygous temperature sensitive *E(z)* mutation, $E(z)^{61}$, using chromatin immunoprecipitation followed by hybridization to microarrays (ChIP-chip) or by sequencing (ChIP-seq) (see Materials and Methods for more details). In addition, we utilized the modENCODE ChIP-chip data (Celniker *et al.*, 2009). In accordance with the previous observation on polytene chromosomes (Ebert *et al.*, 2004), H3K27me2 was widespread across the genome; we could observe high levels of H3K27me2 at most sites except Pol II binding regions and regions strongly enriched in H3K27me3 (Figure 2.5). It covers most of long intergenic regions (77.7% in Sg4 and 71.3% in Bg3 of intergenic regions longer than 10kb; see Methods for determining the enrichment of H3K27me2) and many silent CG genes that have no obvious Pol II binding.

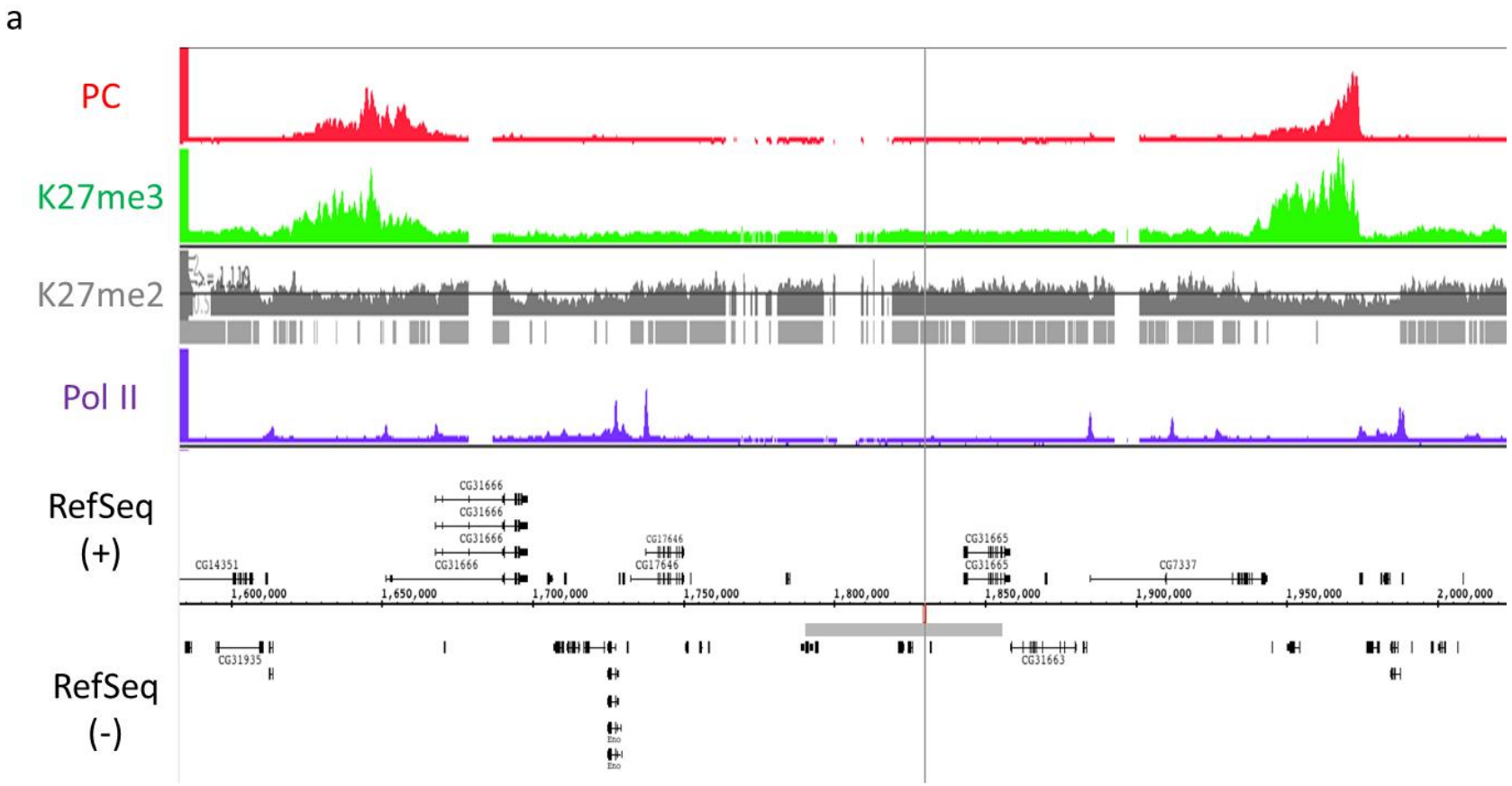


Figure 2.5 Widespread H3K27me2 distribution in Sg4 and Bg3 cell lines

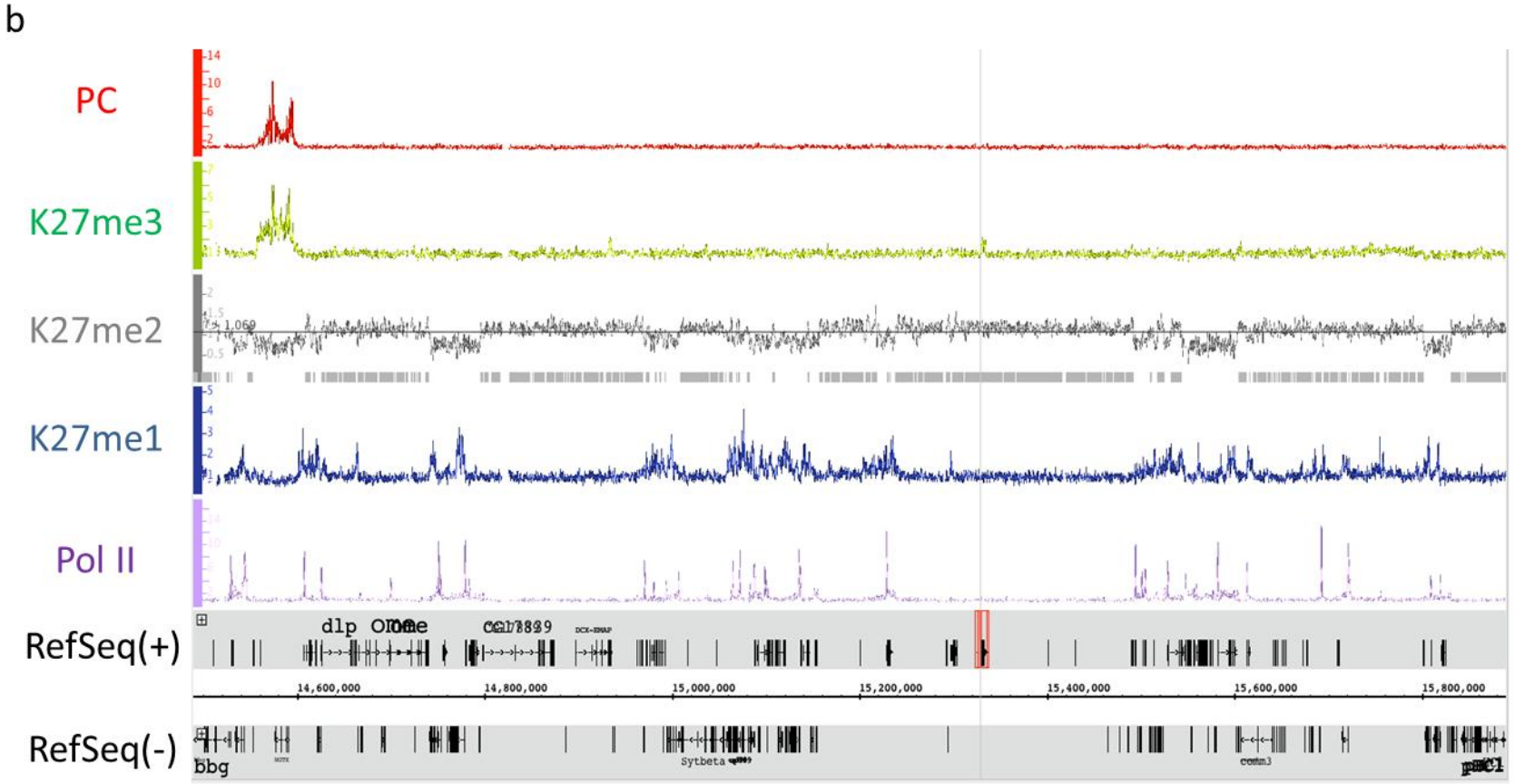


Figure 2.5 Widespread H3K27me2 distribution in Sg4 and Bg3 cell lines (Continued)

Figure legend of Figure 2.5 Continued.

The distributions of PC (red), H3K27me (me3 in green; me2 in gray; me1 in blue), and Pol II (purple) were mapped by ChIP-chip (a : Sg4 by Y. Schwartz, T. Kahn, and V. Pirrotta ; b : Bg3 by modENCODE project). The gaps in the binding profiles are due to the lack of unique oligonucleotides in the underlying DNA sequence. For the H3K27me2 tracks, enriched regions are indicated as gray bars below the profiles.

Recently, Filion *et al.* (2010) classified the whole genome into five chromatin types by integrating genome-wide maps of 53 chromatin components in *Drosophila* Kc cell line. Particularly, one type of chromatin (so called BLACK chromatin) covers 48% of the genome and is actively repressed, but it is distinct from other repressive chromatin types which are enriched in known repressive marks H3K9me2 and H3K27me3. The BLACK chromatin type contains gene-poor intergenic regions and >4000 silent or low level expressing genes. BLACK chromatin domains largely correspond with regions where we observe the H3K27me2 in genome-wide mapping data (Figure 2.6). Indeed, 87.2% of the BLACK chromatin regions was enriched with H3K27me2 in Sg4 cells and had highest level of H3K27me2 among the five chromatin types (Figure 2.7), suggesting that the active repression in BLACK chromatin is associated with the enriched H3K27me2 level.

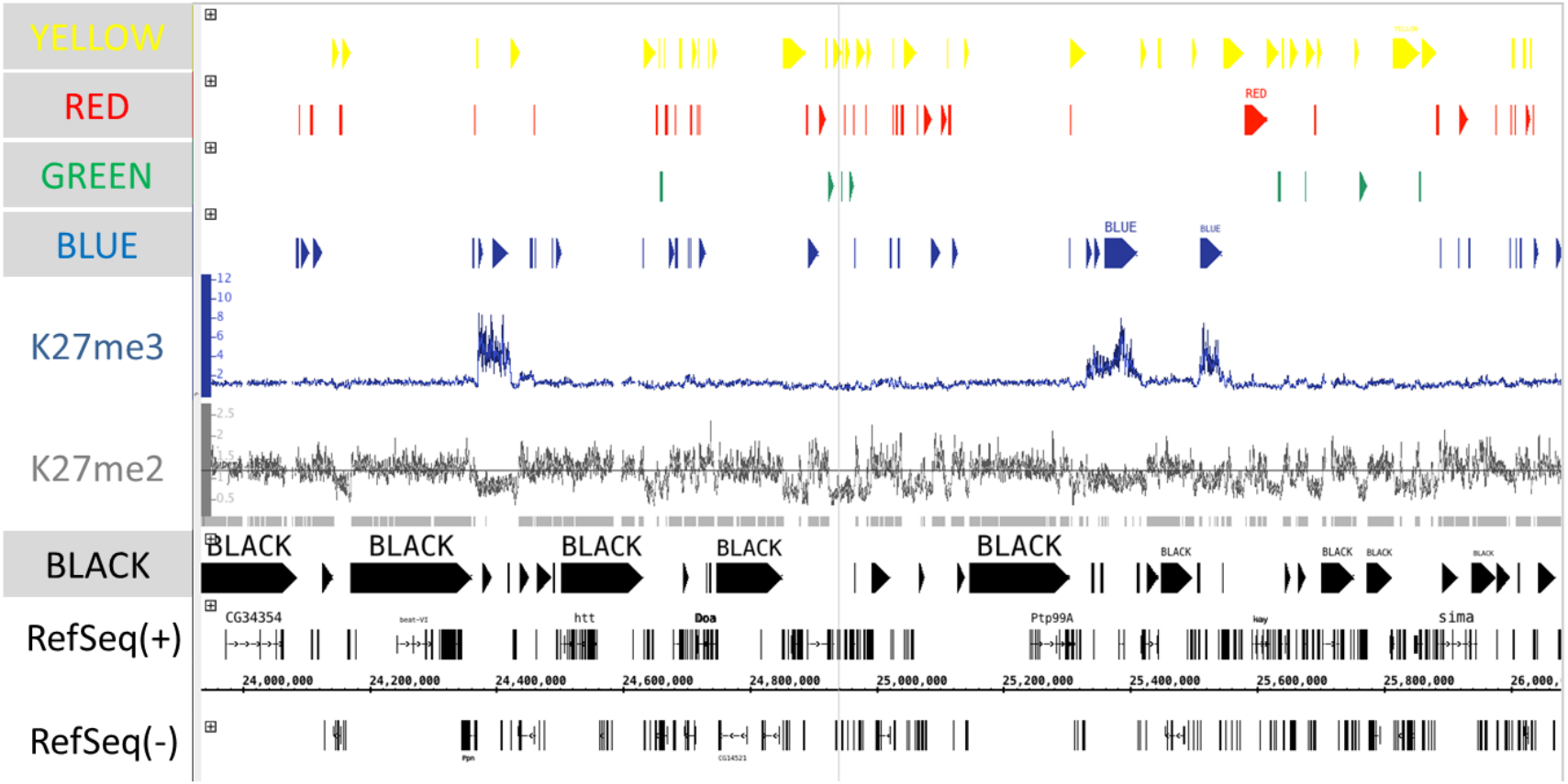


Figure 2.6 H3K27me2 distribution overlaps with BLACK chromatin

Figure legend of Figure 2.6 continued.

The profiles of H3K27me2/3 in Sg4 cell line are compared with the 5 types of chromatin in Kc cell line (colored arrows and bars; from Filion *et al*., 2010). YELLOW and RED chromatin states are transcriptionally active states. GREEN and BLUE chromatin states are transcriptionally repressed and enriched by H3K9me2 and H3K27me3, respectively. BLACK chromatin state is shown below the H3K27me2 track.

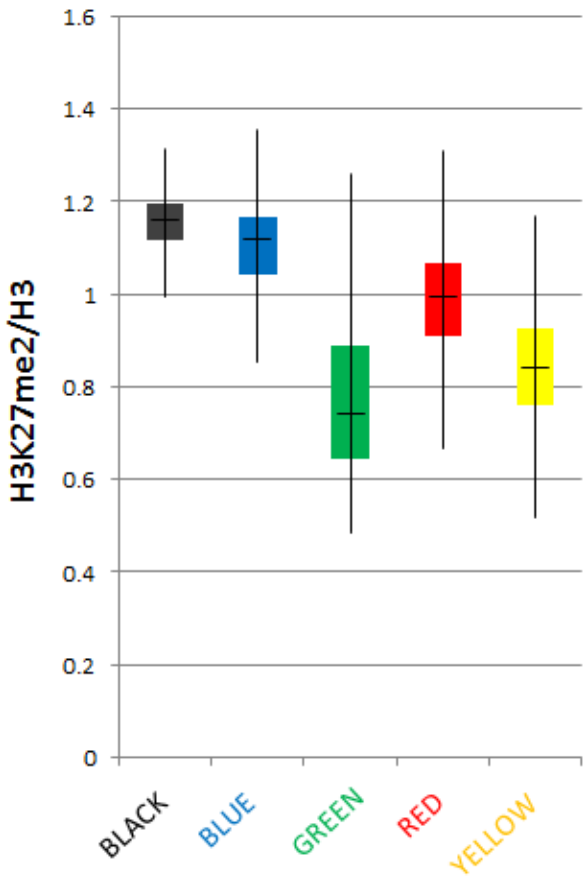


Figure 2.7 BLACK chromatin is enriched by H3K27me2

Schematic box plots of the average H3K27me2/H3 signal values (from Sg4) on the five types of chromatin regions (from Kc cells ; Filion *et al.*, 2010). The bottom and top of the boxes are the first and third quartiles, and the marks inside the boxes represent the median values. The ends of the whiskers represent the lowest value within 1.5 interquartile range (IQR) of the lower quartile, and the highest value within 1.5 IQR of the upper quartile.

In order to further confirm that the signals we are observing in H3K27me2/3 genome-wide data are genuine, we examined whether the level of H3K27me decreases when E(Z) is inactivated in the EZ2-2 cell line (See Materials and Methods for details). By western blot, we could observe strong decreases of global H3K27me2/3 levels in EZ2-2 at 31˚C compared with 25˚C (Figure 2.8).

ChIP-qPCR analysis on several intergenic regions and silent genes that are not enriched in H3K27me3 confirmed that H3K27me2 was enriched in those regions and showed a strong decrease upon E(Z) inactivation (Figure 2.9). Notably, H3K27me3 level also showed over ten-fold decrease in those regions indicating that a low but significant level of H3K27me3 coexists in H3K27me2 enriched regions. So, it seems that a significant proportion of the total H3K27me3 exists in these H3K27me2 enriched regions and outside of H3K27me3-rich PcG target genes. Since the decrease in H3K27me3 in these regions (10-fold) was far stronger than the H3K27me2 decrease (2- to 4-fold) upon E(Z) inactivation, the observed H3K27me3 signal is not likely to be due to the cross-reactivity of the antibody. So it is possible that the low level of H3K27me3 might contribute to the genome-wide repression.

At 31˚C, PcG target gene regions in the EZ2-2 cell line showed strong H3K27me3 decrease. Intriguingly, the level of H3K27me2 increased significantly at those regions (Figure 2.9). Since E(Z) binding at most PREs was diminished to background levels at 31˚C (Figure 2.10), this H3K27me2 increase is more likely to be the result of de-methylation of H3K27me3 than methylation from H3K27me1. This observation suggests that H3K27me3 is being constantly de-methylated by a de-methylase, most likely dUTX

since it is the sole known de-methylase for H3K27 in *Drosophila*, and stable E(Z) binding is required for maintaining high H3K27me3 level at PcG target genes.

Although the function of the PRC2 with mutated E(Z) is largely inactivated at 31˚C, it is likely that the methyl-transferase activity is not completely lost as more drastic loss of methylation occurs at 32.5˚C. To obtain the robust decrease of methylation in H3K27, EZ2-2 cells needed to be grown at least 8-12 days at 31˚C. During the growth at 31˚C, EZ2-2 cells gradually stopped dividing, and the estimated number of cell divisions in 12 days at 31˚C was less than two (see Materials and Methods). So the dilution of H3K27me2/3 by cell divisions might account for some proportion (2- to 4-fold) of the observed decrease in H3K27me2/3 levels. Still, the 10-fold decrease in H3K27me3 levels at intergenic regions indicates the active removal of H3K27me3 by de-methylation. For H3K27me2, it is not so evident since the decreased level is comparable to the expected range by dilution through cell divisions, but this relative weaker decrease of H3K27me2 is thought to be, at least in part, due to the K27me3 de-methylation, which causes increase in H3K27me2 as can be seen at PcG target regions (Figure 2.9). In addition, the H3K27me2/3 levels in PcG target and non-PcG target regions showed further decrease during 12-24 days after temperature shifting, although the EZ2-2 cells have almost stopped growing at that period, indicating that active de-methylation is involved in the decrease of H3K27me2/3.

The loss of H3K27me3 was varying among PcG target regions; the flanking regions of some strong PREs like bxd-PRE exhibited more residual H3K27me3 levels, especially near the core of the PREs, whereas some other PcG target regions lost the H3K27me3 almost completely.

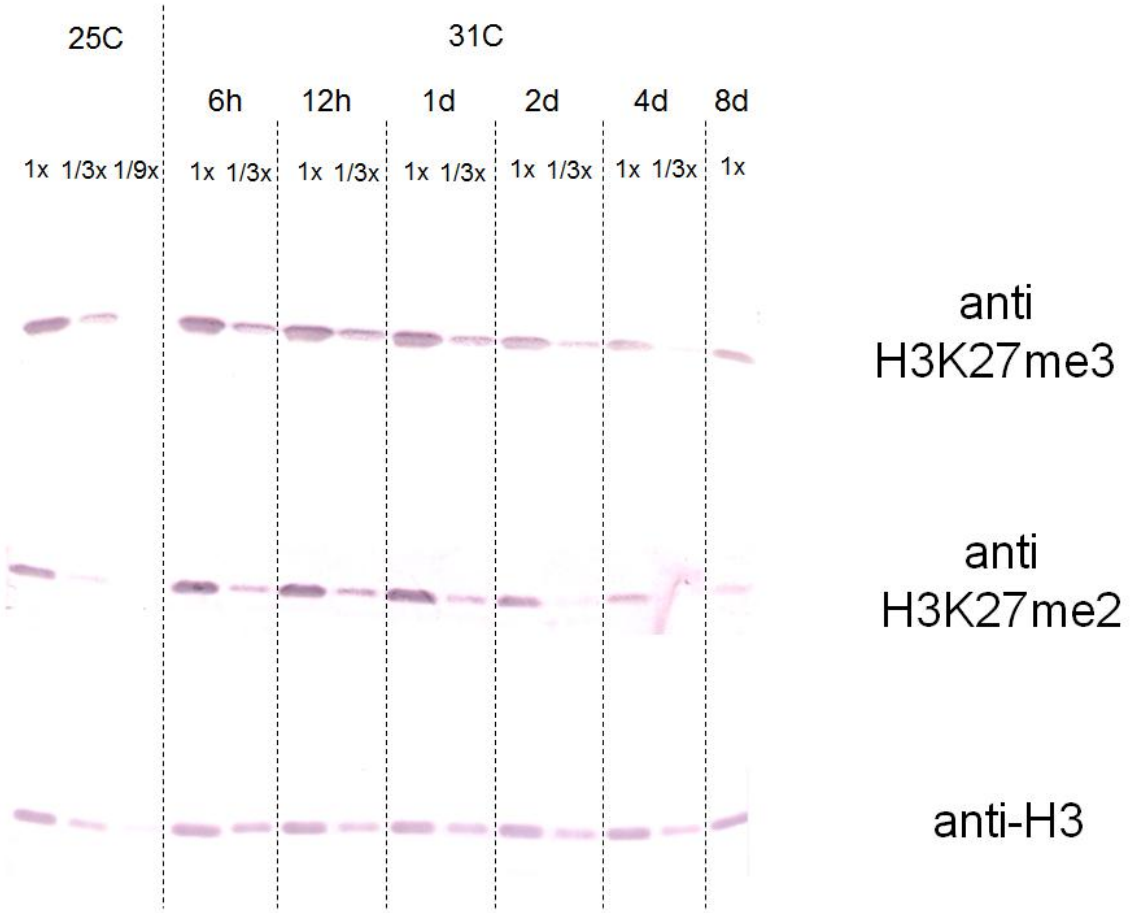


Figure 2.8 E(Z) inactivation of histone methyl-transferase activity by temperature shifting

EZ2-2 cells that have homozygous *E(z)*^{61} mutant are grown in 25℃ (permissive temperature) and 31℃ (non-permissive temperature), and the global H3K27me2/3 levels in nuclear extracts are determined by western blot analysis. Three-fold serial dilutions of lysate were loaded. Total H3 served as a loading control.

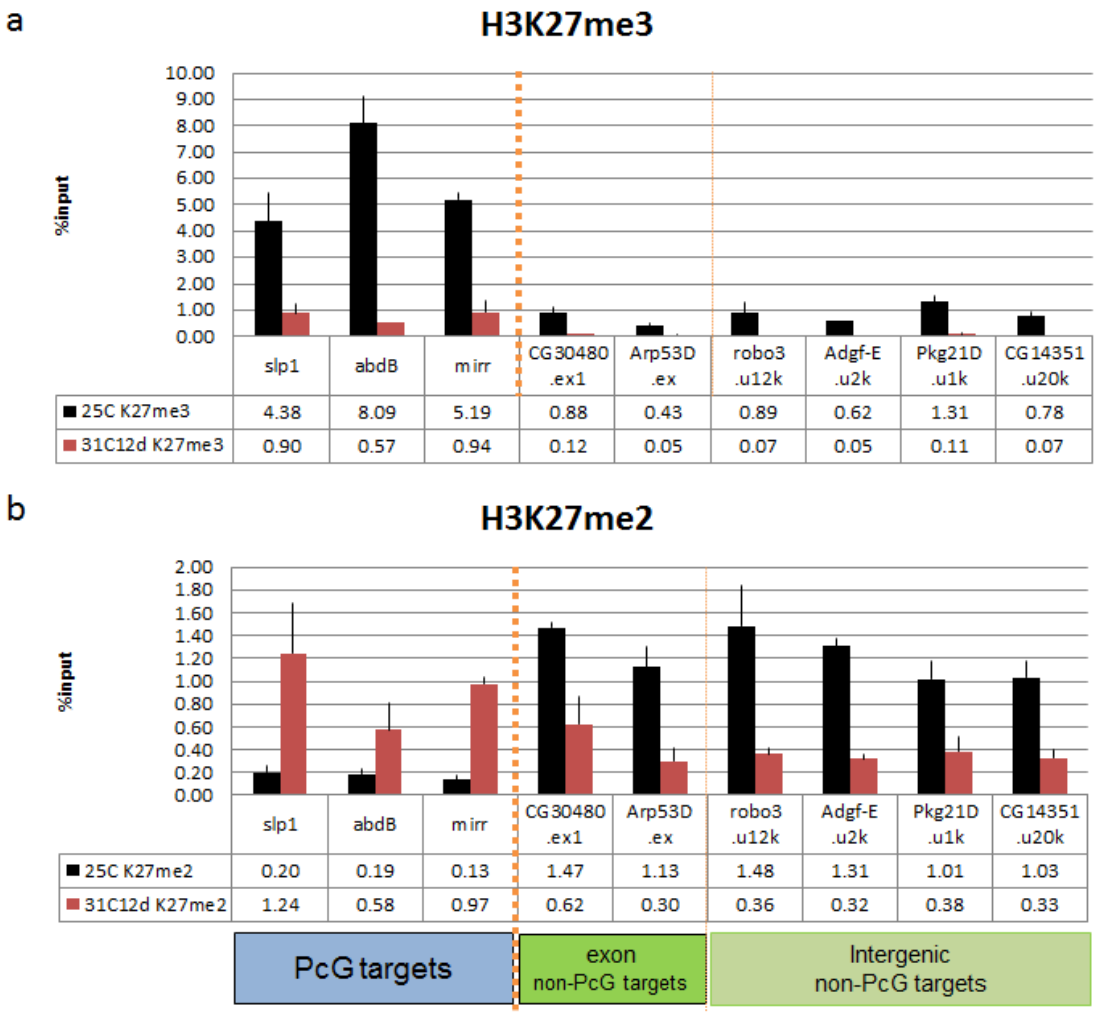


Figure 2.9 Changes in the levels of H3K27me2/3 after E(Z) inactivation at PcG targets and non-PcG targets

The levels of (a) H3K27me3 and (b) H3K27me2 at PcG targets and non-PcG targets (coding and non-coding regions) are measured by ChIP-qPCR from the EZ2-2 cells grown in 25℃ and 31℃. Results from *slp1*, *abdB*, and *mirr* regions represent the PcG target regions. Two exonic regions of non-PcG target genes, *CG30480* and *Arp53D*, and four intergenic non-PcG target regions (upstream regions of *robo3*, *Adgf-E*, *Pkg21D*, and *CG14351*; approximate distances to the transcription start sites are indicated) are

Figure legend of Figure 2.9 continued.

measured to check the effect of E(Z) inactivation in non-PcG target regions. Measured values normalized to input DNA (%input) are shown.

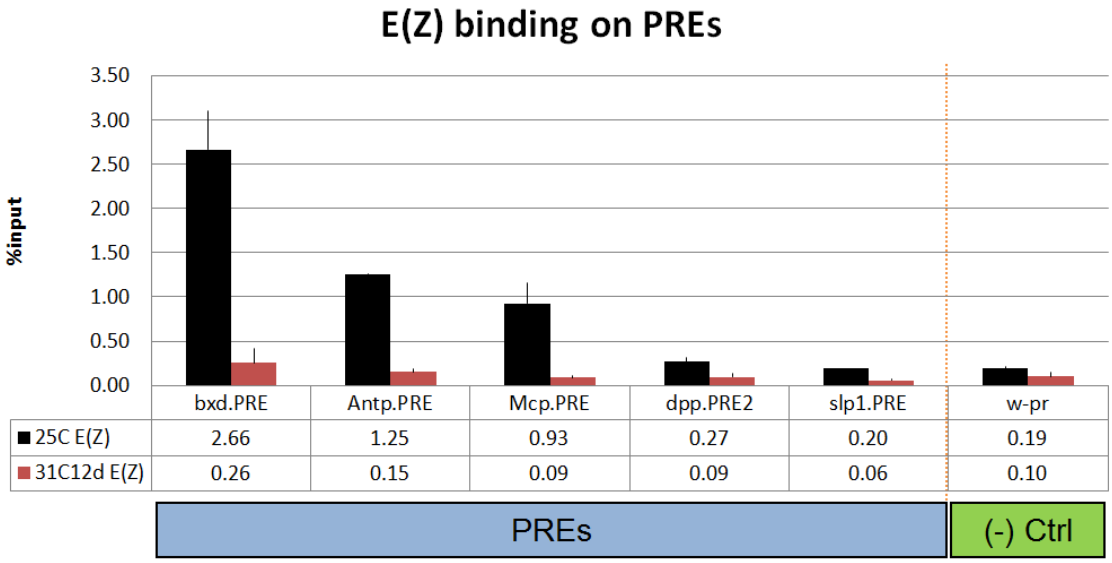


Figure 2.10 E(Z) binding decrease after E(Z) inactivation at PREs

The levels of E(Z) binding at PREs (and the white promoter as negative control) are measured by ChIP-qPCR from the EZ2-2 cells grown in permissive temperature (25℃) and non-permissive temperature (31-32.5℃).

To test whether the widespread H3K27me2 level has a transcriptional repressive role distinct from PcG repression which accompanies strong Pc binding and high H3K27me3 levels, we examined the correlation between the H3K27me2 level and transcription level of non-PcG targets. For coding regions, we classified the FlyBase annotated genes as PcG targets or non-PcG targets based on the levels of PC and H3K27me3 to distinguish from the repression by PRC1, and non-PcG targets were further classified by their expression levels based on RNA-seq data at the 25˚C (see Materials and Methods for details). Similar classification was applied to intergenic regions.

Supporting the idea of a repressive role for H3K27me2, non-PcG target genes with low level expression showed a higher level of H3K27me2 than actively transcribing genes in Sg4 (Figure 2.11) and modENCODE Bg3 cell line (Figure 2.12). It is also noteworthy that H3K27me1 level is relatively high on actively transcribing genes and low on weakly expressing genes. This inverse pattern of H3K27me1 with H3K27me2 indicates that the H3K27me2 on active genes might be actively removed by de-methylation, which suggests the possibility that dUTX is involved in the removal of H3K27me2 on active non-PcG target genes. However, it is also possible that higher nucleosome turn-over rates in active genes contributed to the high H3K27me1 levels by de novo methylation. Meta-analysis results after normalization relative to the H3 distribution still preserves the same pattern (Figure 2.12). Similar to coding regions, intergenic regions also showed higher H3K27me2 in silent or low level expression regions (Figure 2.13). PcG target regions were depleted for both H3K27me1 and H3K27me2 because of strong H3K27me3 enrichment (Figure 2.11-14).

In addition to high H3K27me2, H3K27me3 was also relatively higher at the silent coding and intergenic regions than active transcribing regions (Figure 2.14), which is in accordance with the observation of the low but significant level of H3K27me3 on H3K27me2-enriched intergenic regions in the ChIP-qPCR results (Figure 2.9).

In agreement with the notion that H3K27ac is associated with active transcription, H3K27ac level showed a good correlation with the expression levels and a negative correlation with H3K27me2/3 (Figure 2.15).

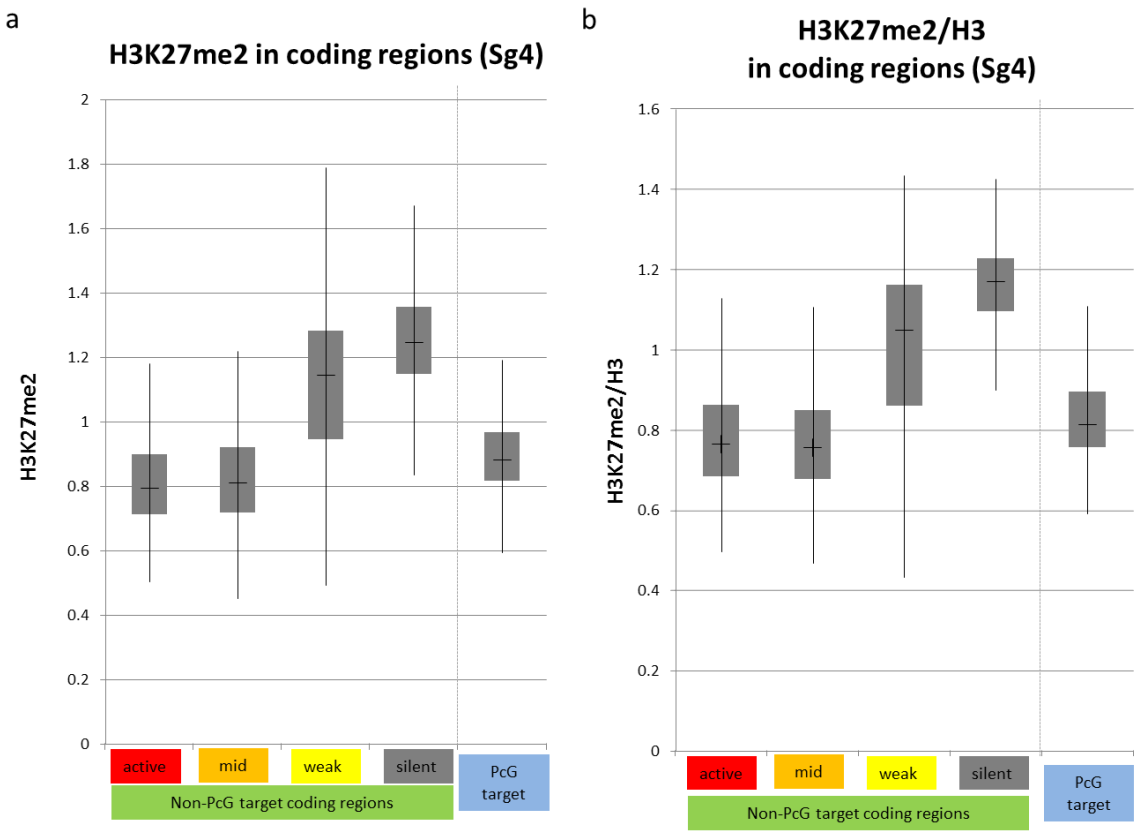


Figure 2.11 H3K27me2 levels in non-PcG target genes and PcG target genes in Sg4

(a) H3K27me2 levels in the PcG target genes and non-PcG target genes that are highly active (highest 10%), intermediately active (middle 10%), weakly expressing (bottom 10% of transcribing genes), and silent (no RNA-seq reads) were analyzed from Sg4 H3K27me2 ChIP-chip data (b) The same analysis was done with the normalized H3K27me2 values relative to H3.

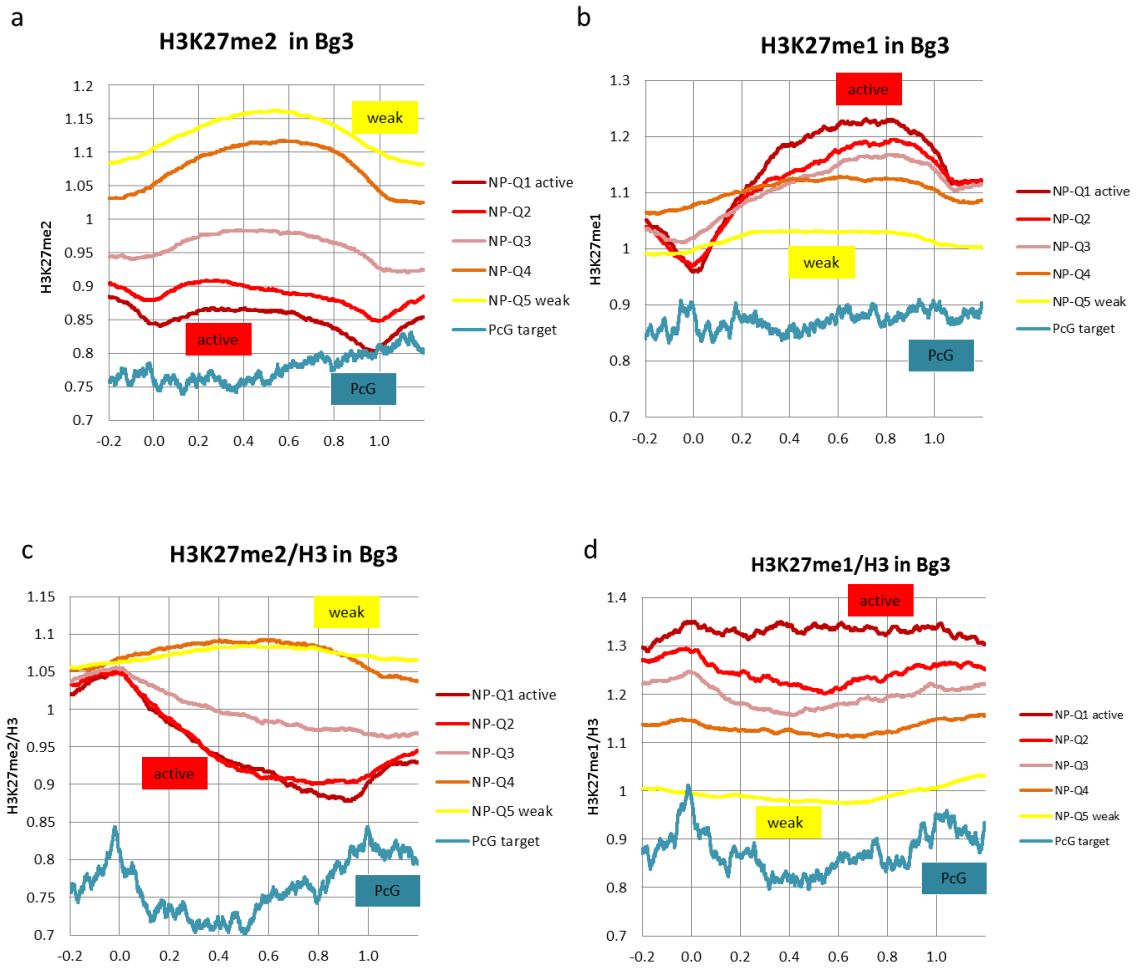


Figure 2.12 H3K27me1/2 levels in non-PcG target genes and PcG target genes in Bg3

Meta-gene profiles of (a) H3K27me2 and (b) H3K27me1 for the PcG target genes and five sub-groups of non-PcG target genes with different expression level (e.g., NP-Q1 for the top 20 percentile actively transcribing non-PcG target genes) in Bg3 cells are shown. The numbers on x-axis represent the positions relative to transcription start sites (0.0) and transcription end sites (1.0). The results after normalization relative to the H3 distribution are shown in (c) and (d).

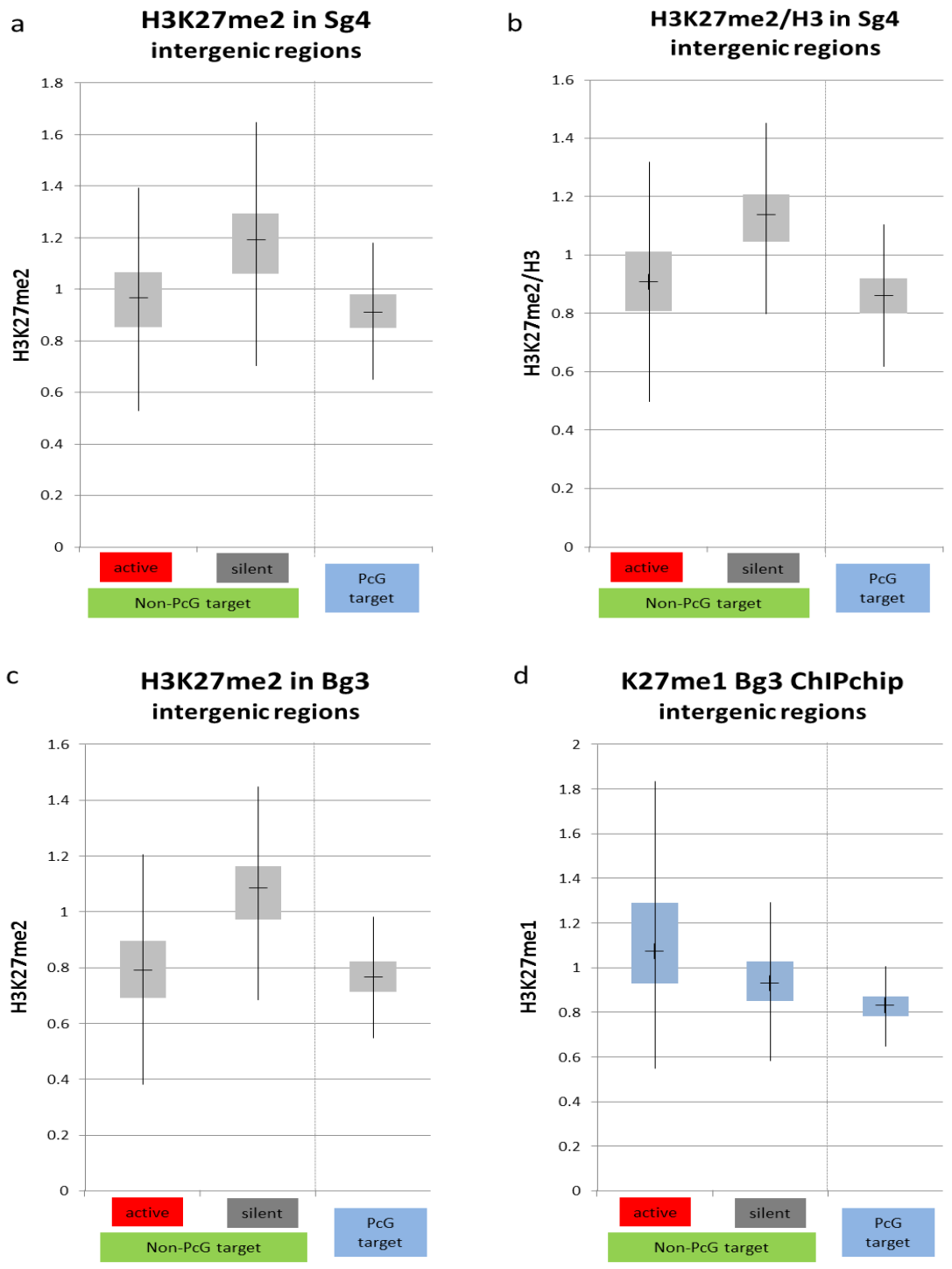


Figure 2.13 H3K27me2 levels in intergenic non-PcG target and PcG target regions in Sg4 and Bg3

(a) H3K27me2 levels in the intergenic PcG target regions and non-PcG target regions that are highly active (highest 10%) and silent (no RNA-seq reads) in Sg4 cells were

Figure legend of Figure 2.13 continued.

analyzed. (b) The H3K27me2 levels in Sg4 cells after normalization relative to the H3 distribution are shown. For Bg3, the levels of (c) H3K27me2 and (d) H3K27me1 in the same sub-groups of intergenic regions are shown. The RNA-seq result from EZ2-2 in 25˚C is used for determining the sub-groups of intergenic non-PcG target regions.

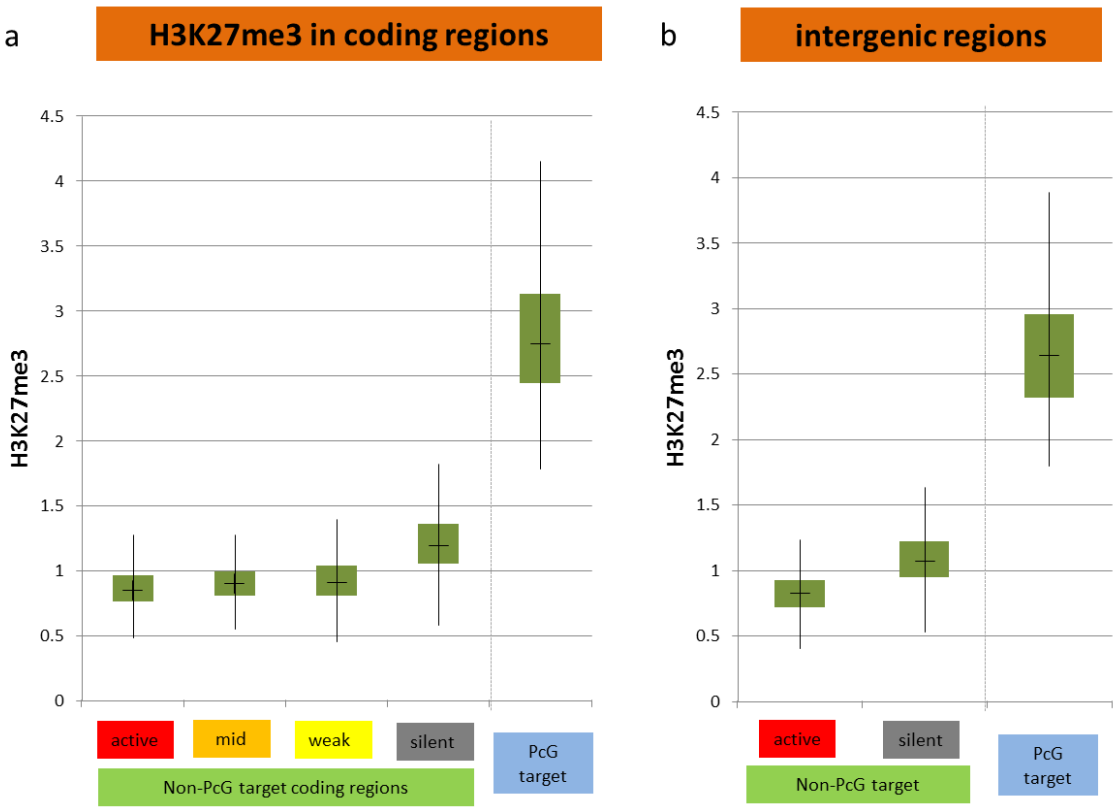


Figure 2.14 H3K27me3 levels of non-PcG target and PcG target regions in EZ2-2 cells

H3K27me3 levels of the non-PcG target and PcG target regions in (a) coding regions and (b) intergenic regions were analyzed using the ChIP-chip data from EZ2-2 cells at 25˚C.

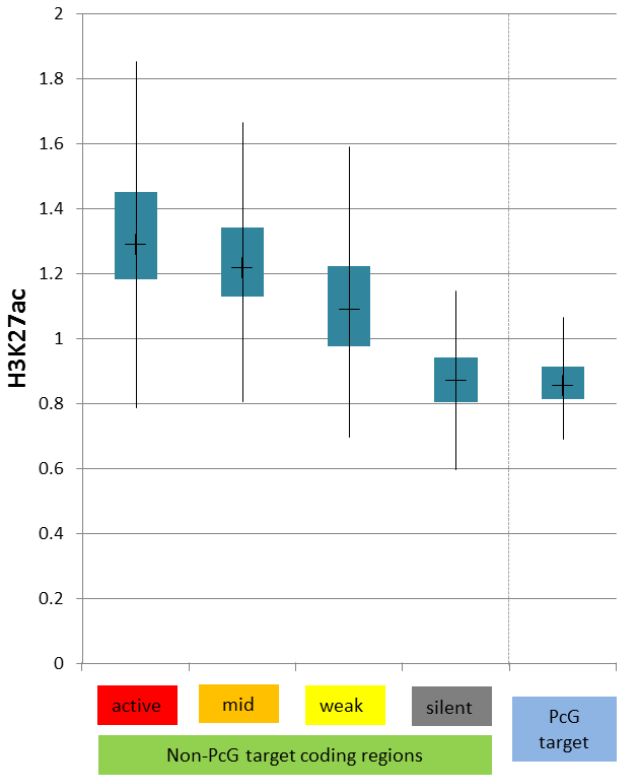


Figure 2.15 H3K27ac levels in non-PcG target and PcG target regions in EZ2-2 cells

H3K27ac levels in the non-PcG target genes and PcG target genes were analyzed using the ChIP-chip data from EZ2-2 cells at 25˚C.

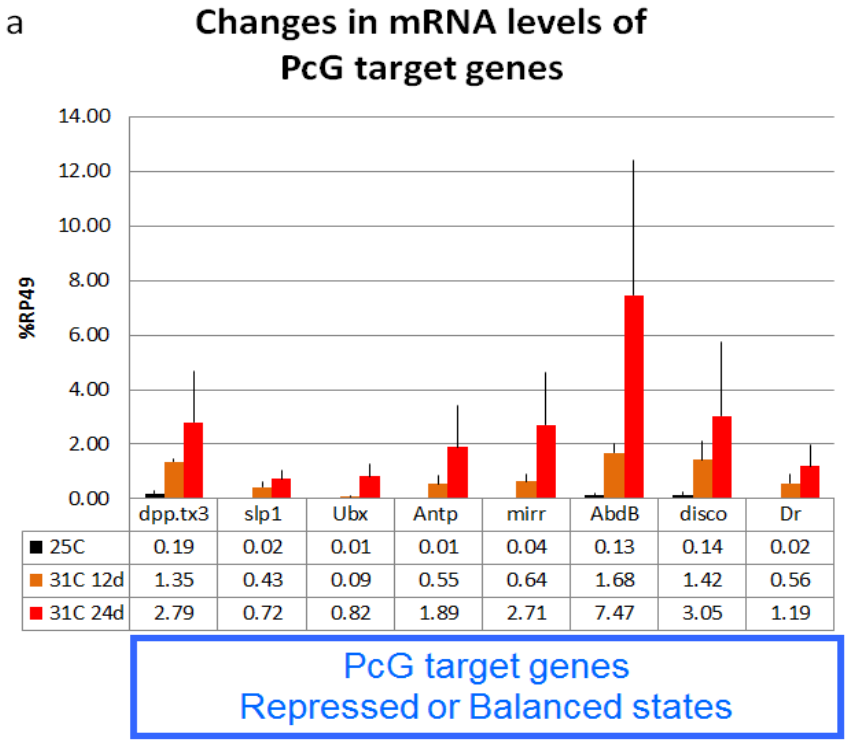
2.3.2 E(Z) inactivation causes global de-repression

At 31˚C, we observed very strong de-repression of all tested H3K27me3-enriched PcG target genes (Figure 2.16a) along with significant decrease of H3K27me3 level. More importantly, we could observe significant transcriptional level increase on all tested non-PcG intergenic regions (Figure 2.17) and silent genes (Figure 2.16b) when H3K27me levels decreased by E(Z) inactivation, which suggests that E(Z) represses the intergenic regions and silent non-PcG target genes through genome-wide H3K27me2.

To further investigate the extensive repressive role of E(Z) in genome-wide, we compared the paired-end RNA-seq reads from the total RNA of EZ2-2 cells at 25˚C and 31˚C. RNA-seq results confirmed that this intergenic de-repression is happening all over the genome (Figure 2.18-20, Table 2.1). At 25˚C, 12.3% of total intergenic regions showed detectable transcription and the coverage increased to 37.7% when E(Z) function is impaired by temperature shifting to 31˚C (Figure 2.18, Table 2.1). Considering the fact that many of the intergenic regions whose transcription could be detected by qPCR had no reads in our RNA-seq, this result is likely to be under-estimated due to the insufficient sequencing depth. Coding regions also showed increased coverage from 49.5% to 71.8%, indicating that a large proportion of silent genes are also de-repressed. In fact, 67% (1538 genes) of the 2294 silenced non-PcG genes that had no reads at 25˚C exhibited RNA-seq reads at 31˚C. Furthermore, most of the transcribing non-PcG target genes also showed increased transcription upon E(Z) inactivation (Figure 2.20), indicating that, even when transcriptionally active, non-PcG target genes are not completely independent of H3K27 methylation. More than 90% of the non-PcG target genes with detectable expression levels at 31˚C (4856 genes from total 5314 non-PcG target genes that had >0.5 FPKM

and Pc binding below average) exhibited more than a 2-fold increase in their expression. In general, among non-PcG target genes, lower level transcribing genes showed stronger expression increase upon E(Z) inactivation but PcG target genes showed the strongest de-repression. These results suggest that repression by H3K27me happens in a gradual manner all over the genome. Since H3K27me2/3 are more enriched in weakly expressing genes, so the impact of H3K27me decrease is more dramatic on low level transcribing genes. These observations also suggest that H3K27 methylation is the cause, not the consequence of transcriptional inactivity, arguing against the possibility that the methylation on H3K27 occurs opportunistically when a region has low transcriptional activity.

To eliminate the possibility that this global de-repression is due to the high temperature (e.g. heat shock effect or higher accessibility at higher temperature), we tested the intergenic expression levels of a control wild type cell line (Ras3) at 31˚C and found no significant increase (Figure 2.17). In addition, we could observe the intergenic de-repression in another independent *Drosophila* cell line (S2 cells) when E(Z) is depleted by RNAi (Figure 2.21).



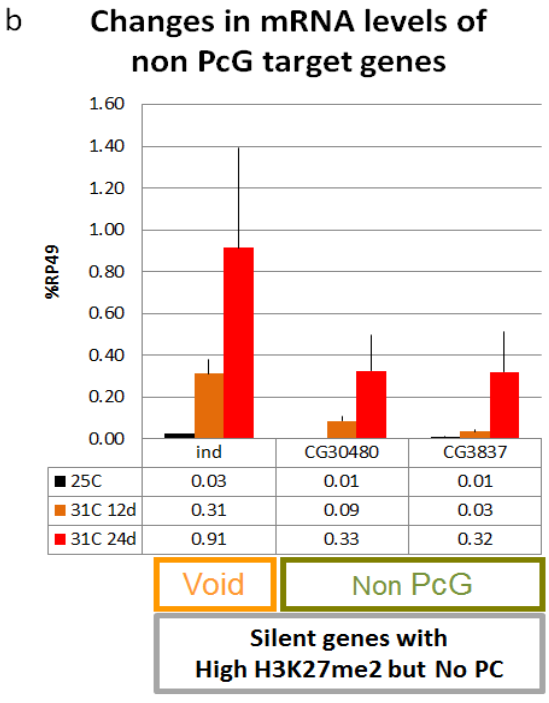


Figure 2.16 Changes in the expression level of PcG target genes and non-PcG target genes

Figure legend of Figure 2.16 continued.

Expression levels of (a) H3K27me3-enriched PcG target genes and (b) H3K27me2-enriched non-PcG target genes in EZ2-2 cells at 25℃ (black) and 31℃ (after 12 days as orange and 24 days as red) are shown as the values relative to those of RP49 (also known as RpL32). The data from at least two independent experiments are averaged, and error bars indicate standard deviations.

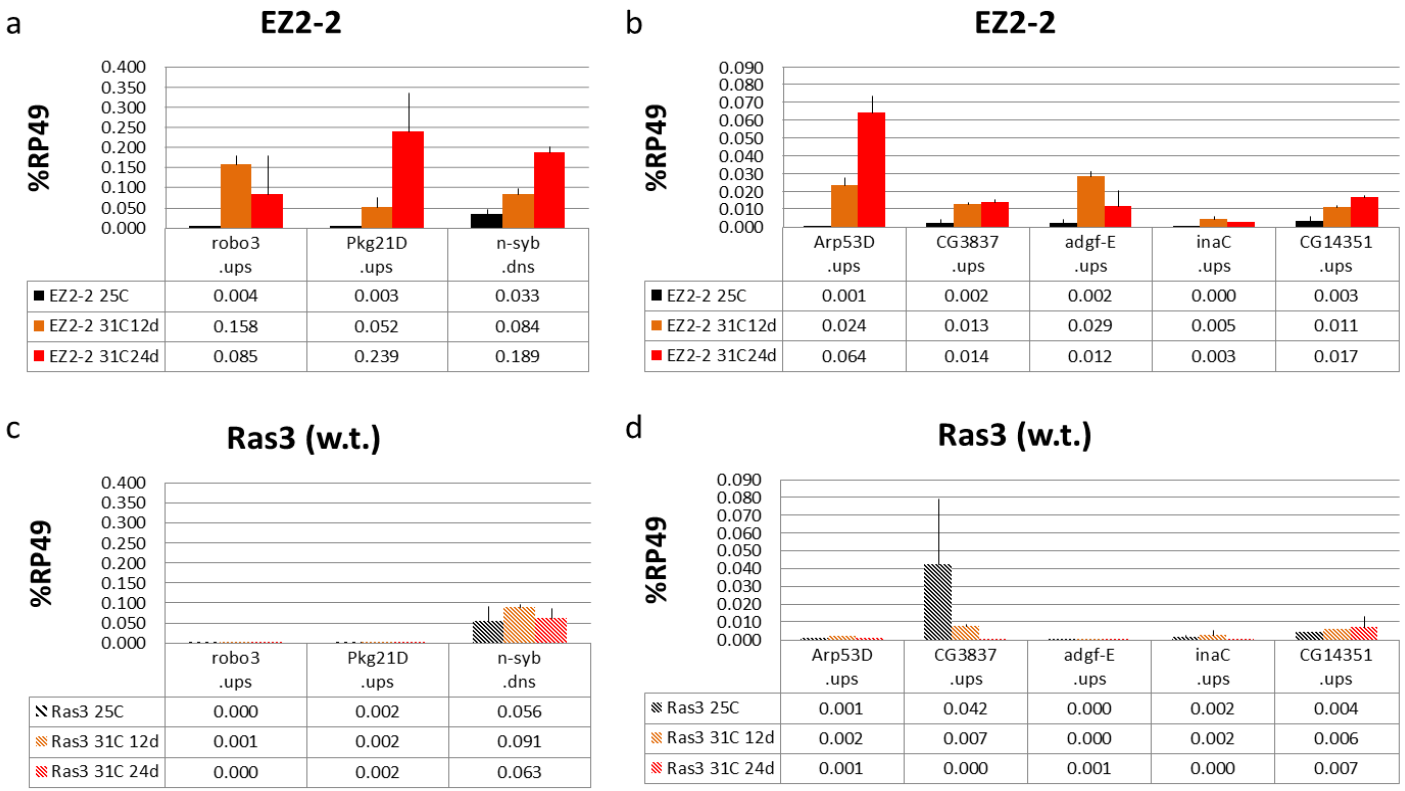


Figure 2.17 Changes in the expression level of intergenic non-PcG target regions

(a and b) Expression levels of H3K27me2-enriched intergenic non-PcG target regions in EZ2-2 cells at 25℃ (black) and 31℃ (after 12 days as orange and 24 days as red) are shown as the values relative to those of RP49. (c and d) Expression levels in a wild type cell line (Ras3; diagonally striped bars) with the same conditions are shown as controls.

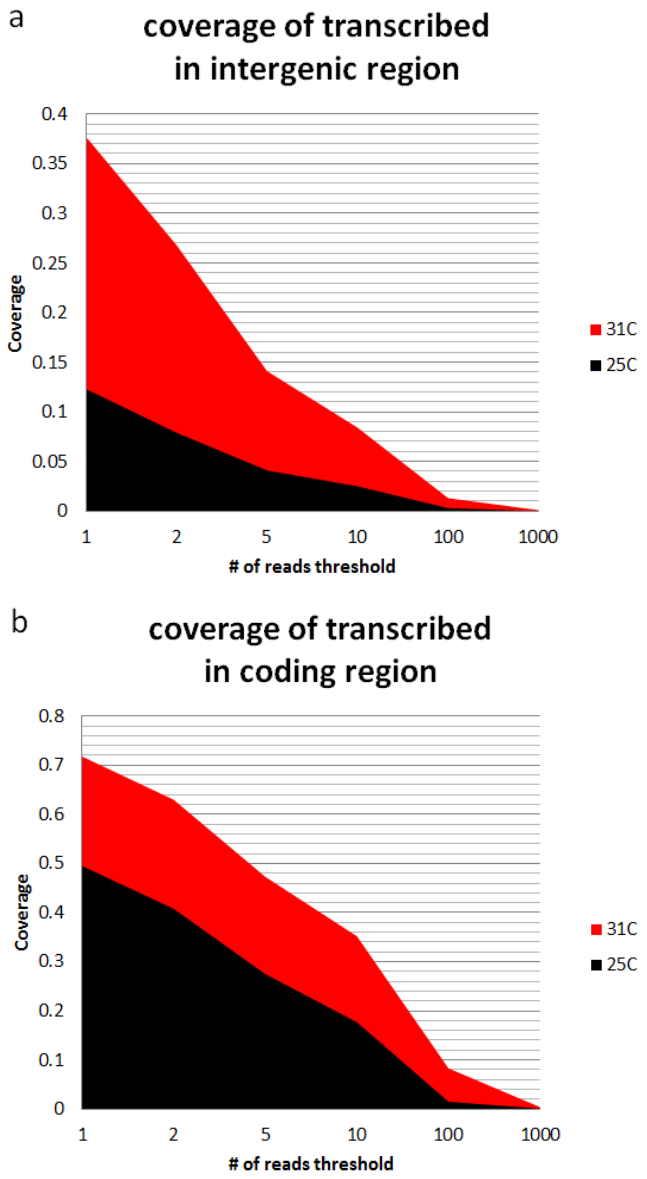


Figure 2.18 Genome-wide coverage of expression tags in EZ2-2 cells at 25˚C and 31˚C

The proportion of transcribing regions in *Drosophila* genome (a for intergenic regions; b for coding regions) was estimated by calculating the percentage of bins (25bp size) that have equal or more than the threshold number of reads.

	25℃	31℃
Intergenic region	12.3%	37.7%
Coding region	49.5%	71.8%
whole genome	33.4%	57.3%

Table 2.1 Proportion of transcribing regions in EZ2-2 cells at 25℃ and 31℃

The percentage of bins that have equal or more than one read from intergenic and coding regions in EZ2-2 cells at **25℃ and 31℃** are shown for estimating the proportion of transcribing regions in *Drosophila* genome. Since the RNA-seq reads were obtained by 100bp paired-end sequencing and the alignments at the repetitive sequences were excluded, the uniquely mapped RNA-seq reads are considered to be very reliable.

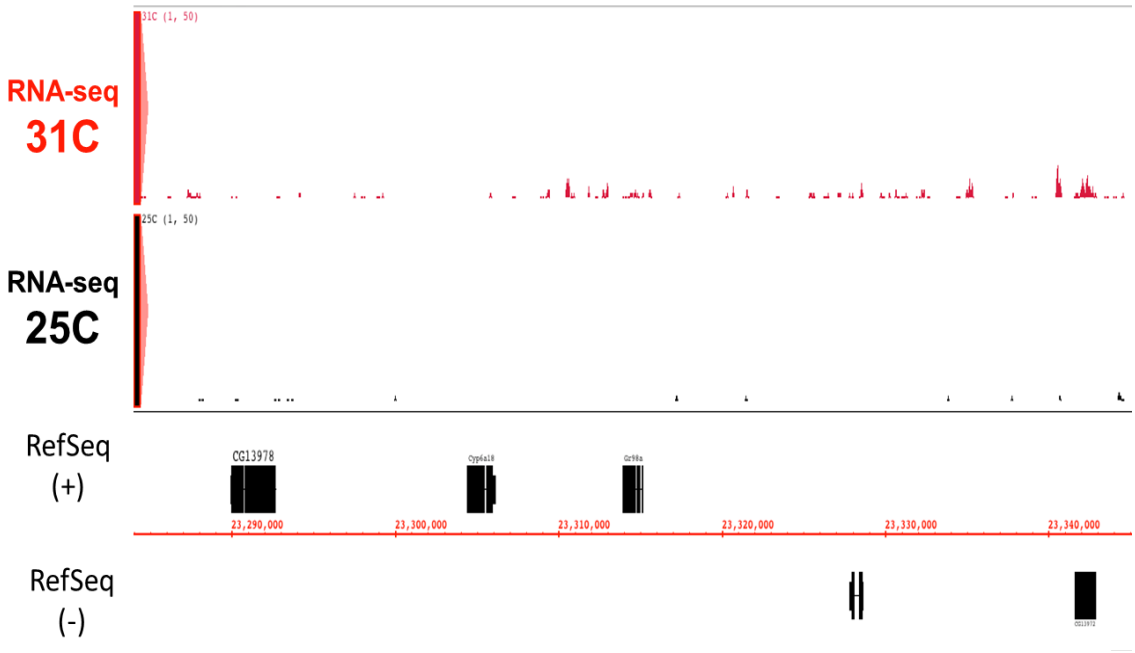


Figure 2.19 De-repression in intergenic regions after E(Z) inactivation

Mapped RNA-seq results from EZ2-2 cells at 25℃ (black; lower track) and 31℃ (red; upper track) temperature in a genomic locus that contains long intergenic regions with some silent genes are shown.

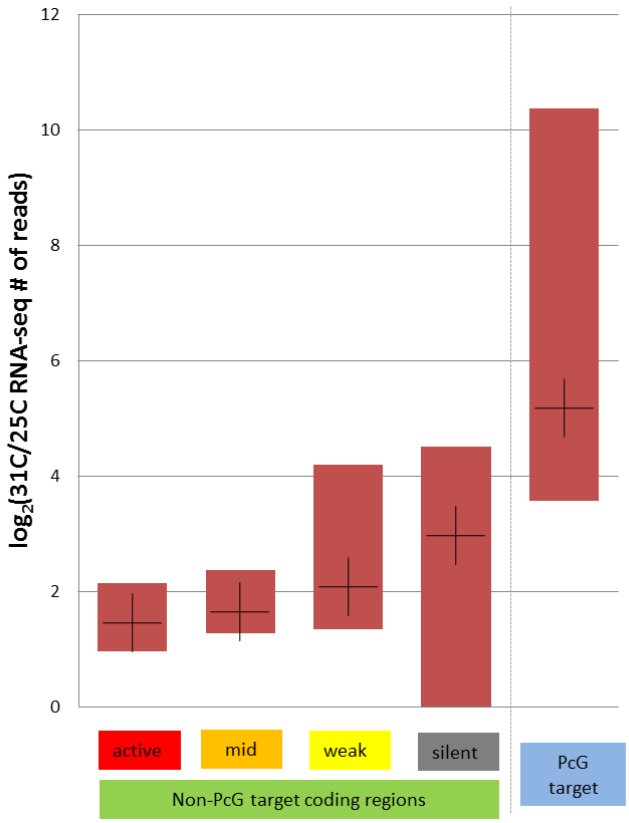


Figure 2.20 De-repression levels in coding regions after E(Z) inactivation

Logged ratio between the expression levels at 31℃ and 25℃ is calculated for each gene in the non-PcG targets with different expression levels and PcG targets. Since the silent genes have no RNA-seq reads at 25℃, one expression tag read was added to the expression values (number of RNA-seq tags) in order to estimate the approximate fold differences between 31℃ and 25℃. The bottom and top of each box represent the first and third quartiles, and the mark inside the box represents the median value.

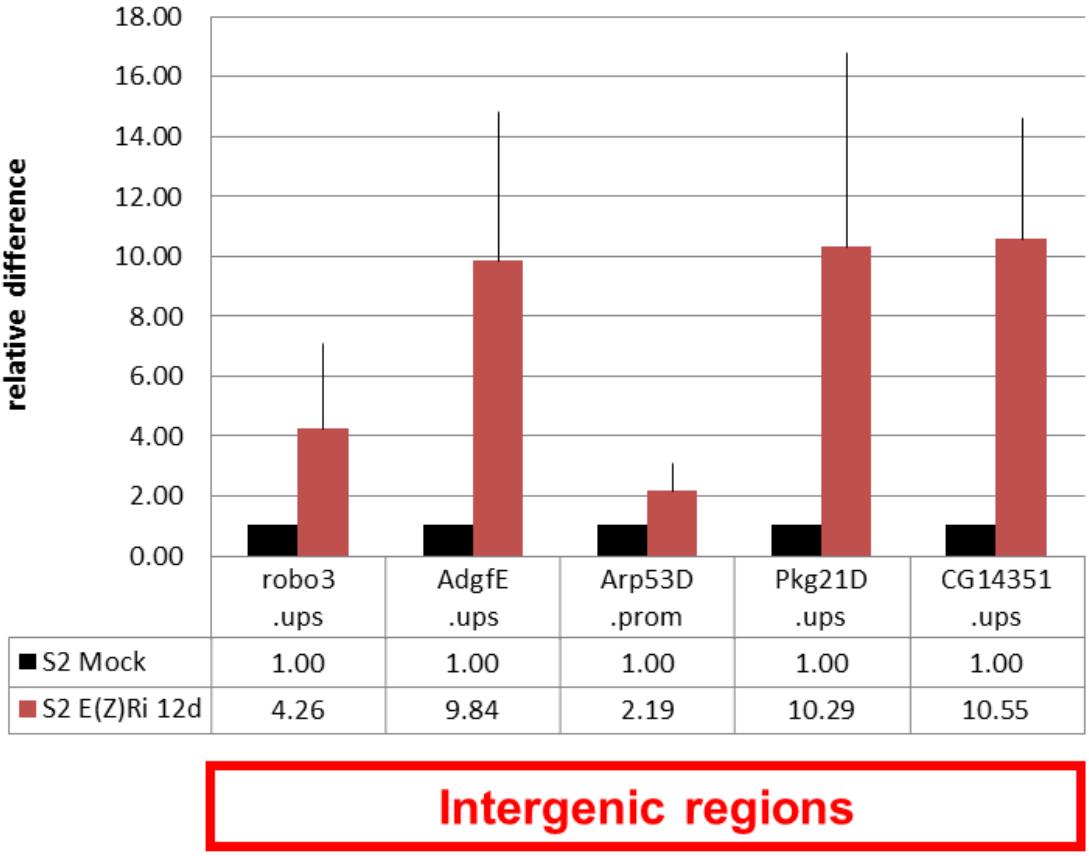


Figure 2.21 De-repression in intergenic regions after E(Z) depletion in S2 cells

Expression levels in intergenic regions were measured by RT-qPCR analysis with total RNA from mock-treated (control; black) or E(Z)-depleted (red) S2 cells. Changes in the expression levels are shown as relative differences.

2.3.3 De-repression in intergenic regions is dependent on dUTX-containing complexes

To further examine whether the decrease of H3K27me is the actual cause of the de-repression in intergenic regions, we knocked down dUTX, the sole known K27 de-methylase in *Drosophila*, and observed significant attenuation of the de-repression. All tested intergenic regions showed more than a 2-fold decrease of de-repression (Figure 2.22). This result suggests that the de-repression upon E(Z) inactivation is dependent on dUTX function that removes pre-existing H3K27 methylation. As observed from RNA-seq results, E(Z) inactivation caused the transcriptional increase of most genes, including the UTX gene. So it is likely that the increased level of UTX also contributes to the genome-wide de-repression. Considering the fact that the mRNA level of UTX in knock-down experiments at 31˚C is comparable to the level at 25˚C, the difference of the de-repression levels between mock-treated and UTX knock-down cells at 31˚C are thought to be the extent that caused by the increased UTX level. It is true that *E(z)* mRNA level also increases at 31˚C but the product is largely inactive due to the temperature-sensitive mutation.

Along with the decrease of H3K27me, we also found that H3K27ac and H3K4me1 levels were significantly increased in intergenic non-PRE regions at 31˚C (Figure 2.23).

CBP, a versatile acetyl-transferase, is responsible for the acetylation of H3K27 in *Drosophila* (Tie *et al.*, 2009). For H3K4me1, recent studies show that TRR is the strongest candidate by comparing the effect of knock-down for several SET domain-containing proteins including dSET1, TRX, and TRR (Ardehali *et al.*, 2011; Herz *et al.*,

2012). Interestingly, CBP and TRR can form two different complexes with dUTX (Mohan *et al.*, 2011; Tie *et al.*, 2012). So, it is very likely that these complexes are causing de-repression by de-methylating H3K27me and writing H3K27ac and H3K4me1 when E(Z) activity is compromised.

To confirm the involvement of CBP and TRR in intergenic de-repression, we knocked down TRR or CBP at 31˚C and found that the de-repression level became reduced (Figure 2.22). Interestingly, the intergenic regions tested behaved in different ways for each knock-down; e.g., some intergenic regions are strongly affected by RNAi against CBP but weakly by TRR RNAi or vice versa. These results could be further confirmed by qPCR analysis of several adjacent intergenic regions (Figure 2.24-25) and indicate that de-repression of some regions is more dependent on CBP function and other regions are more dependent on TRR function.

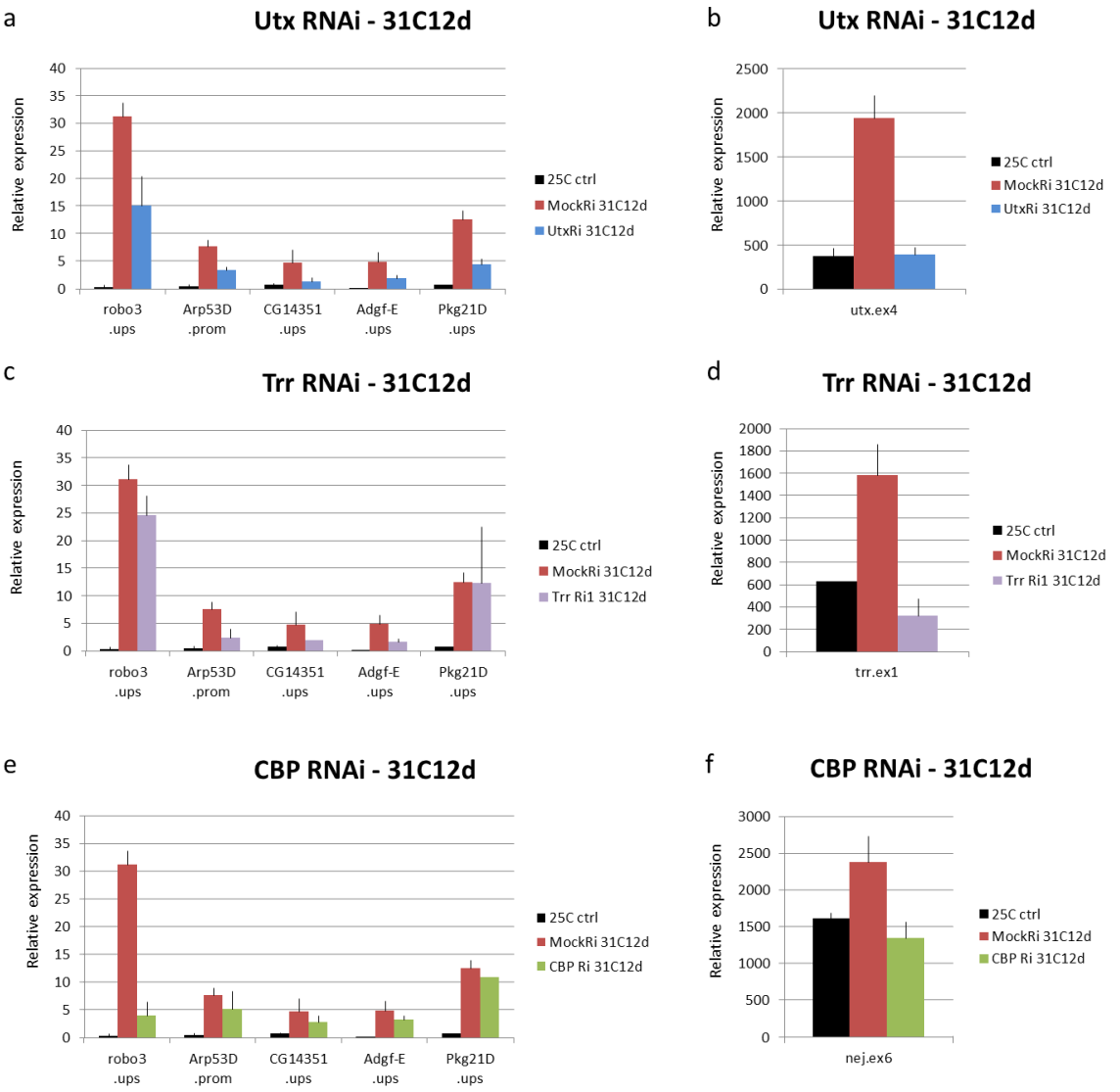


Figure 2.22 Attenuation of the de-repression in intergenic regions by the knock-down of UTX, TRR, and CBP

Expression levels in intergenic regions were measured by RT-qPCR analysis with total RNA from EZ2-2 cells at 25℃ and at 31℃ incubated with double-stranded RNA targeting LacZ (control), (a) UTX, (c) TRR or (e) CBP. The measured expression levels are normalized by the number of cells that are determined by the DNA amount.

Figure legend of Figure 2.22 continued.

The efficiency of the knock-down experiments was monitored by the mRNA levels of the targets (b, d, and f).

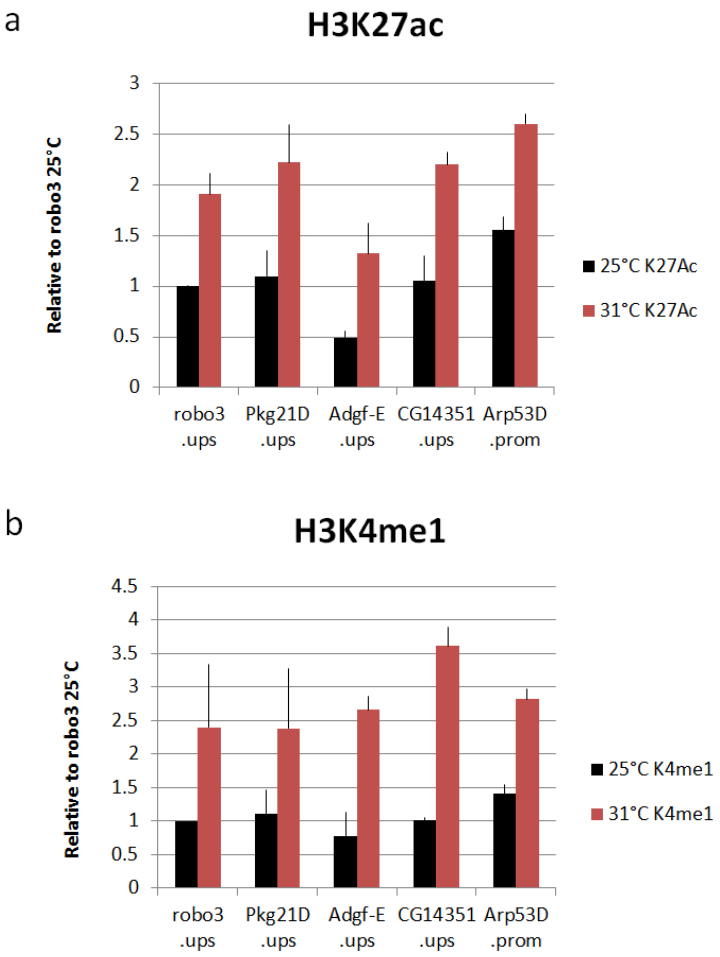


Figure 2.23 Changes in the levels of active histone marks at intergenic regions after E(Z) inactivation

The levels of (a) H3K27ac and (b) H3K4me1 in intergenic regions are determined by ChIP-qPCR from the EZ2-2 cells grown in 25℃ (black) and 31℃ (red).

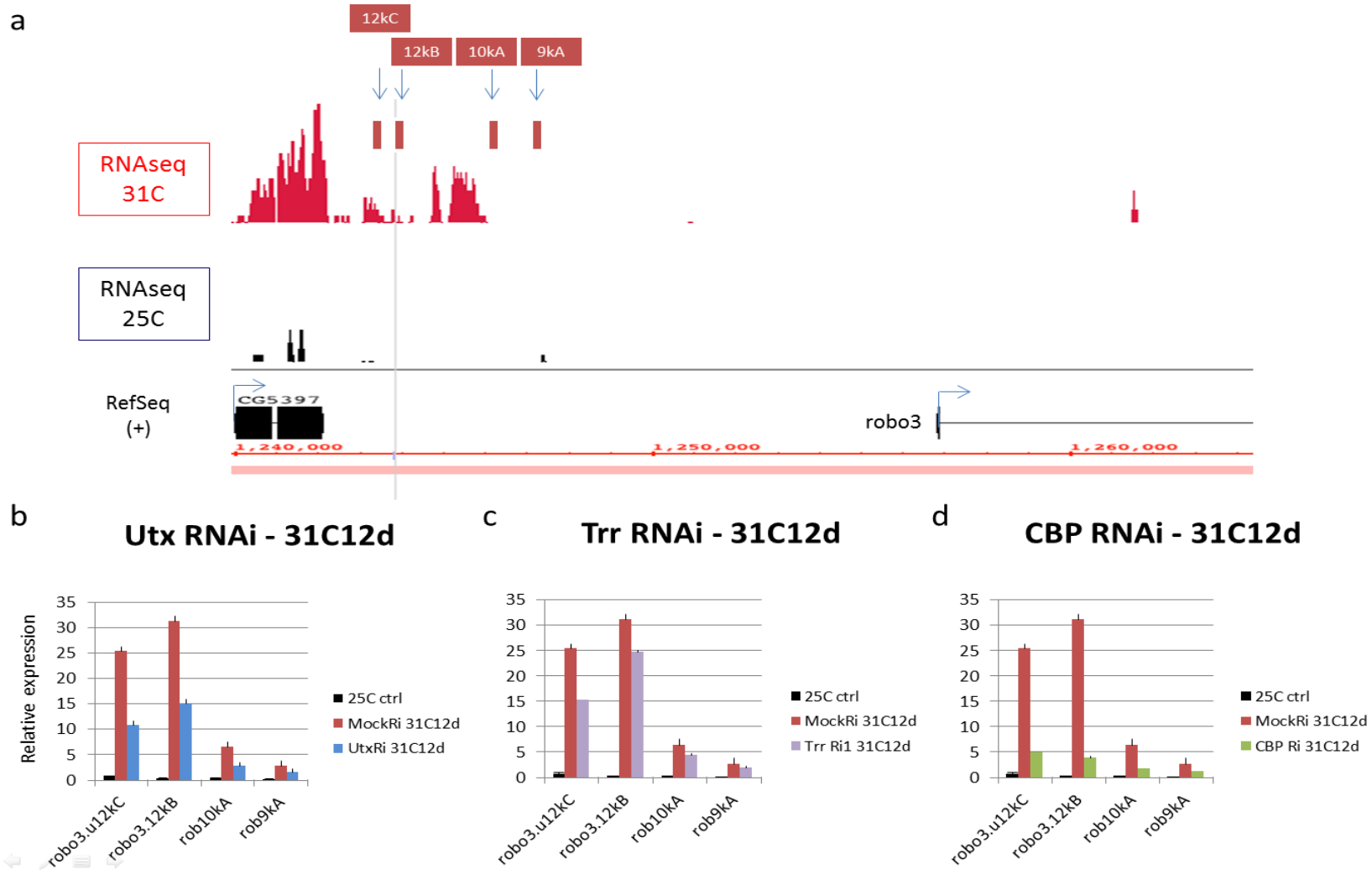


Figure 2.24 Knock-down effects of UTX, TRR, and CBP at the robo3 upstream regions

Figure legend of Figure 2.24 continued

RNA-seq tags on the *robo3* upstream region at 25℃ (black) and 31℃ (red) were shown in (a). Amplicons used for qPCR analysis are indicated as red bars. The transcription direction of the annotated genes is indicated by the arrows at the transcription start sites. Expression levels in *robo3* upstream regions were measured by RT-qPCR analysis with total RNA from EZ2-2 cells at 25℃ and at 31℃ incubated with double-stranded RNA targeting LacZ (control), (b) UTX, (c) TRR or (d) CBP.

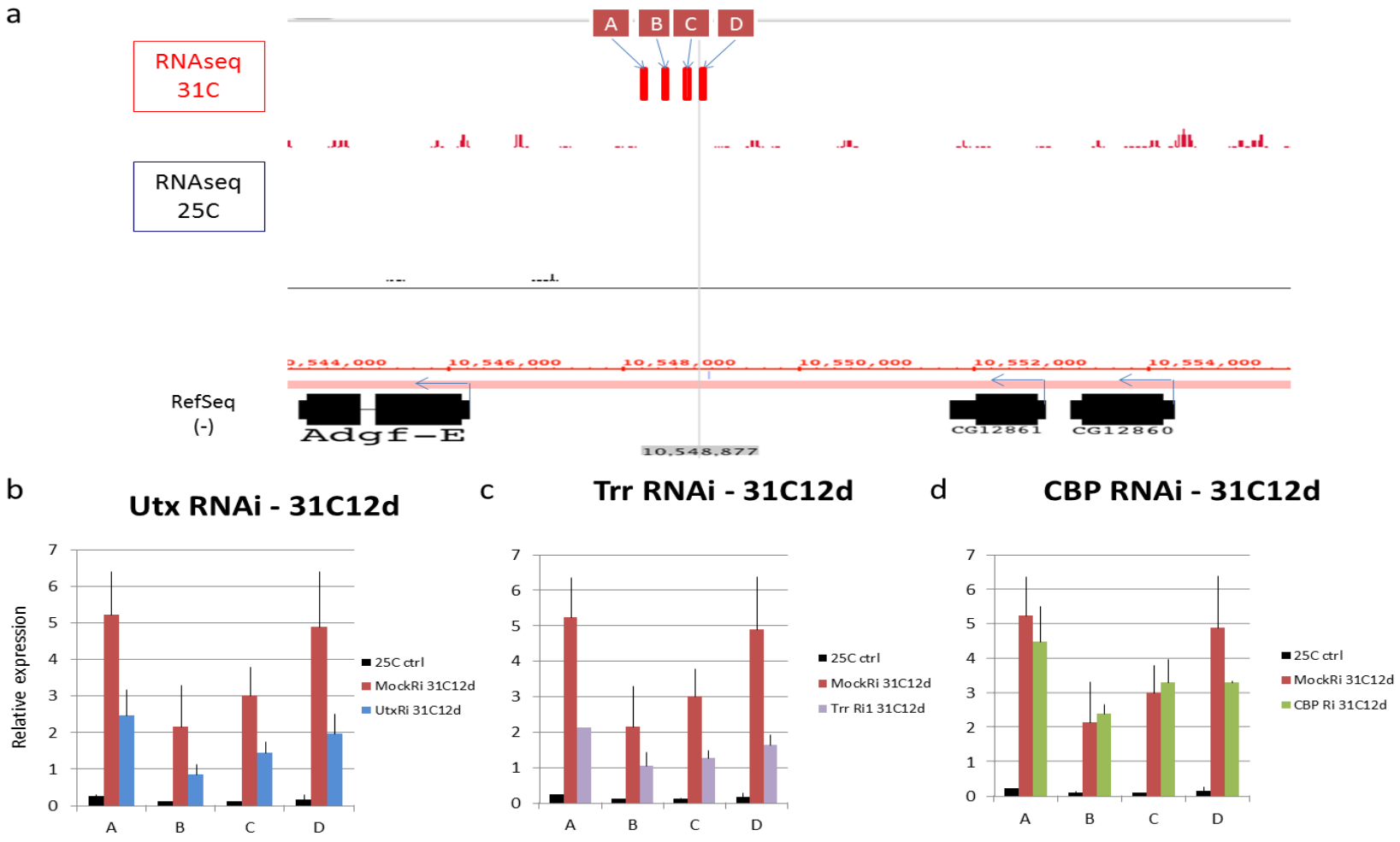


Figure 2.25 Knock-down effects of UTX, TRR, and CBP at the Adgf-E upstream regions

Figure legend of Figure 2.25 continued.

RNA-seq tags on the *Adgf-E* upstream region at 25℃ (black) and 31℃ (red) were shown in (a). Amplicons used for qPCR analysis are indicated as red bars. The transcription direction of the annotated genes is indicated by the arrows at the transcription start sites. Expression levels in *Adgf-E* upstream regions were measured by RT-qPCR analysis with total RNA from EZ2-2 cells at 25℃ and at 31℃ incubated with double-stranded RNA targeting LacZ (control), (b) UTX, (c) TRR or (d) CBP.

2.3.4 Global increase of H3K27ac and H3K4me1

As the increased level of H3K27ac and H3K4me1 was observed in all tested intergenic regions, we further examined whether the increase of these marks is also happening in a genome-wide manner as global H3K27me decreases by E(Z) inactivation. We performed ChIP-chip experiments on those two marks and observed a genome-wide increase in both cases (Figure 2.26). Interestingly, the ratio between the values at 31˚C and at 25˚C showed strong negative correlation with the distribution at 25˚C (-0.65 for H3K27ac, -0.61 for H3K4me1: Pearson correlation value, Figure 2.27), indicating that a proportionally stronger increase preferentially occurred at the regions that had low levels at 25˚C. And this increase pattern is in accordance with that of the RNA expression level; the lower the expression level at 25˚C, the stronger the increase at 31˚C, indicating that the change in the active marks correlates well with change in transcriptional level.

As a result of the ubiquitous increase of H3K27ac and H3K4me1, regions enriched in the two active marks at 25˚C become less distinguishable from depleted regions. Since these two active marks are known to be associated with active enhancers (Creyghton *et al.*, 2010; Heintzman *et al.*, 2009), this situation can be considered as showing that most regions in the genome are becoming enhancer-like regions upon E(Z) inactivation. As active enhancers not only stimulate the transcription of nearby genes but also make transcripts of their own (Core *et al.*, 2012; Kim *et al.*, 2010), this might result in the transcriptional increase all over the genome regardless of coding regions and non-coding regions that we are observing here upon E(Z) inactivation. So, this global increase of H3K27ac and H3K4me1 after E(Z) inactivation suggests the interesting possibility that E(Z) antagonizes the writing of those two marks, by complexes that also contain dUTX,

in most intergenic regions so that only particular regions can be selected and utilized as enhancers.

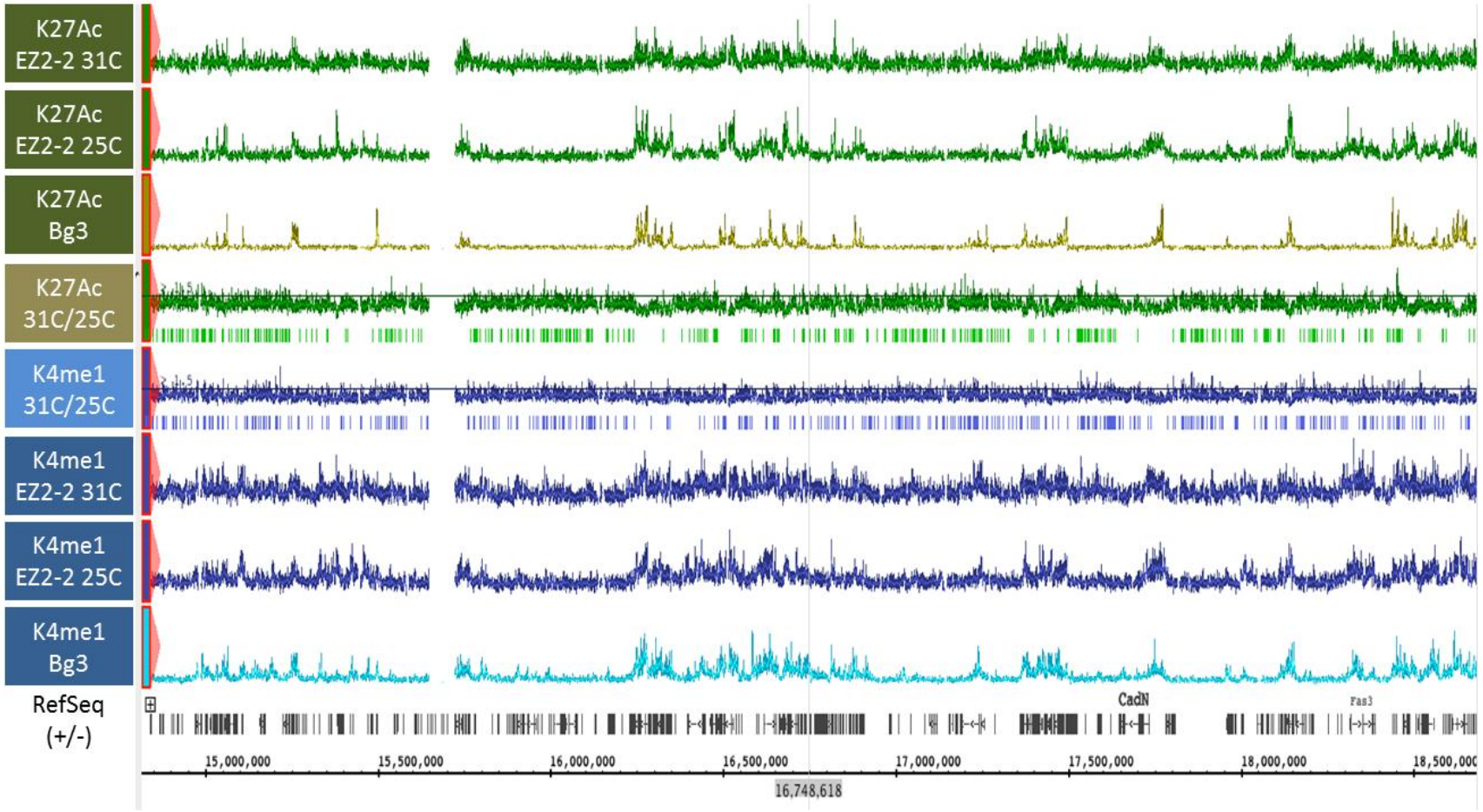


Figure 2.26 H3K27ac and H3K4me1 distribution in EZ2-2 and Bg3 cell lines

Figure legend of Figure 2.26 continued.

The distributions of H3K27ac (first three tracks; EZ2-2 (31℃ and 25℃) as dark green and Bg3 as light green) and H3K4me1 (last three tracks; EZ2-2 (31℃ and 25℃) as dark blue and Bg3 as light blue) were mapped by ChIP-chip (ChIP-chips for Bg3 cell lines by Y. Schwartz, T. Kahn, and V. Pirrotta). The ratio between the signals from 31℃ and those from 25℃ of H3K27ac and H3K4me1 are shown in the two tracks between H3K27Ac tracks and H3K4me1 tracks. The gaps in the binding profiles are due to the lack of unique oligonucleotides in the underlying DNA sequence.

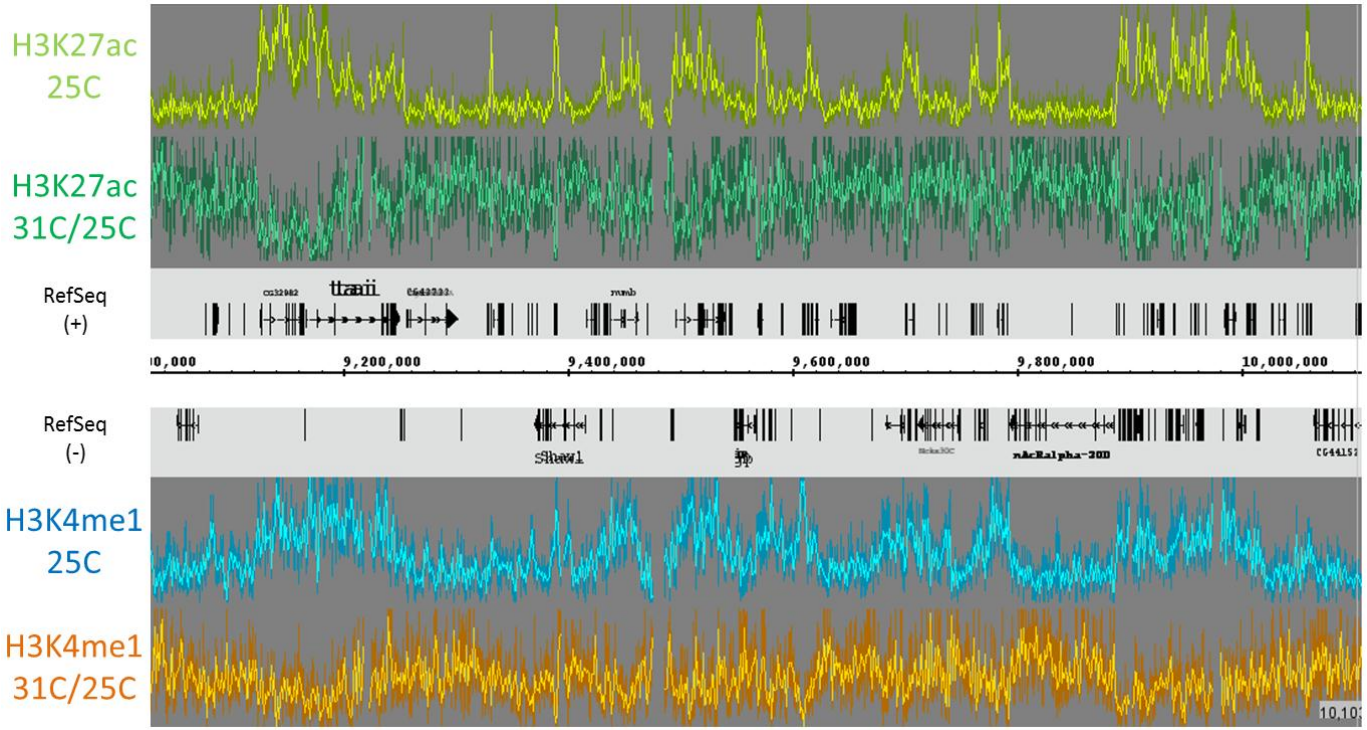


Figure 2.27 Increase of the active histone marks H3K27ac and H3K4me1 negatively correlates with those levels at 25˚C

Distribution of H3K27ac at 25˚C is shown in light green. Ratio between the H3K27ac levels at 31˚C and 25˚C is shown in dark green. Same analysis for H3K4me1 is shown at the two bottom tracks. H3K4me1 distribution at 25˚C is shown as light blue, and the ratio between the H3K4me1 levels at 31˚C and 25˚C is shown as orange.

2.3.5 PRC1 involvement in intergenic repression

The chromodomain of the Polycomb (PC) protein contained in PRC1 binds preferentially to H3K27me3 but still has a significant affinity for H3K27me2. We have found low but detectable levels of H3K27me3 throughout the genome as well as abundant levels of H3K27me2, both of which might attract the presence of PRC1. To test whether PRC1 is also involved in the genome-wide intergenic repression, we checked the expression level of intergenic regions in a *Pc* mutant cell line and found significantly higher levels compared to a wild type cell line (Ras3) (Figure 2.28). Knock-down experiments of PRC1 components on S2 cells also confirmed that PRC1 is also responsible for intergenic repression (Figure 2.29).

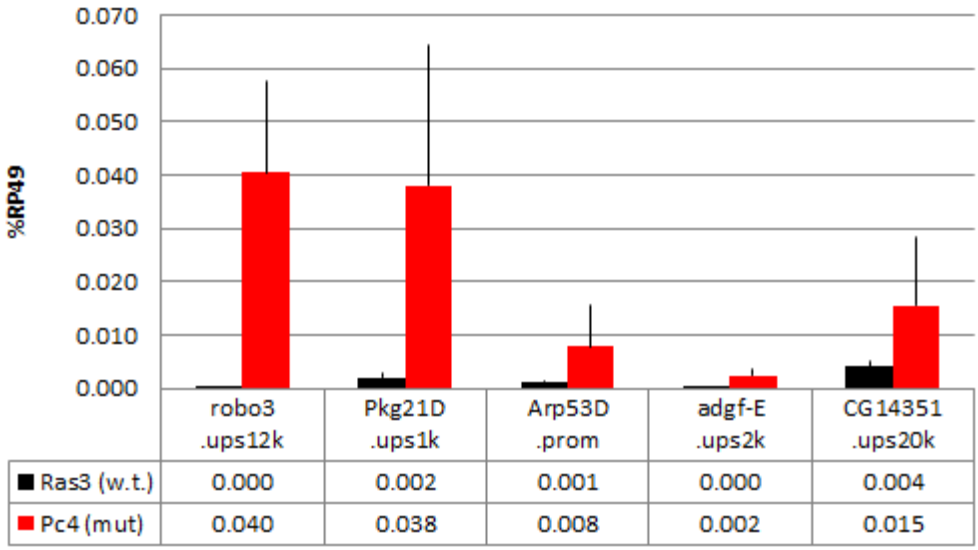


Figure 2.28 Intergenic expressions in *Pc* mutant cells

Expression levels in intergenic regions were measured by RT-qPCR analysis with total RNA from wild type (Ras3; black) and *Pc* mutant (Pc4; red) cells. The expression levels are shown as relative values to those of RP49.

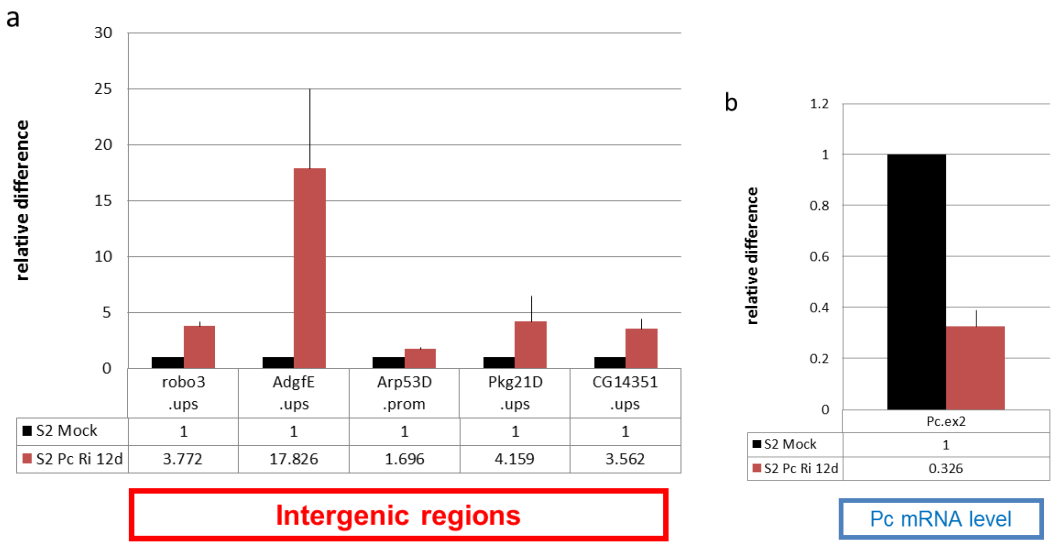


Figure 2.29 De-repression at intergenic regions after PC depletion in S2 cells

(a) Expression levels in intergenic regions were measured by RT-qPCR analysis with total RNA from mock-treated (control; black) and PC-depleted (red) S2 cells. Changes in the expression levels are shown as relative differences. (b) The efficiency of the knock-down experiments was monitored by the mRNA level of *Pc*.

The simplest way to explain these observations is that PRC1 represses the intergenic regions by binding to H3K27me2/3 at least transiently. Even though the level of H3K27me3 is low on intergenic regions, it is thought to be significant amount since the level can decrease up to 10-fold on E(Z) inactivation by temperature shifting (Figure 2.9).

The PRC1 complex includes dRING, an E3 Ubiquitin ligase that is responsible for H2A K118 mono-ubiquitination (H2Aub1). If PRC1 binds to H3K27me in intergenic regions and stays long enough to write the H2Aub1, H2Aub1 enrichment would be found in intergenic regions. To check this possibility, we validated one commercially available H2Aub1 antibody by western blotting and ChIP experiments upon dRing depletion (Figure 2.30-31). ChIP-qPCR results confirmed that H2Aub1 level is significantly high in intergenic regions and are strongly depleted when dRing is knocked down (Figure 2.31) on S2 cells. Further genome-wide analysis by ChIP-seq for H2Aub1 confirmed that many of silent intergenic regions are occupied by significant levels of H2Aub1 in EZ2-2 cell line (Figure 2.32-34).

There is still a possibility that H2Aub1 on intergenic regions is written by dRing-containing complex other than PRC1, e.g., dRAF complex (Lagarou *et al.*, 2008). Or PRC1 could access the whole genome independently of H3K27me3 in the same manner as E(Z): by random hit-and-run. In either case, H3K27me decrease would not affect the H2Aub1 level on the intergenic regions. In order to test these possibilities, we compared the intergenic H2Aub1 levels on EZ2-2 cell line at 25˚C and 31˚C. Upon E(Z) inactivation by temperature shifting, a global H2Aub1 level decrease of about 30% was observed by western blot analysis, and it is unlikely that the decrease is due to the temperature since wild-type cells do not show any decrease in global H2Aub1 level at

31˚C (Figure 2.35). Further ChIP-qPCR and ChIP-seq results showed that many PREs had strong H2Aub1 level decrease and most of the silent intergenic regions exhibited mild decrease of H2Aub1 level on EZ2-2 at 31˚C (Figure 2.33-34), suggesting that the H2Aub1 on intergenic regions is at least partially dependent on H3K27me. In addition, the H2Aub1 level in non-PcG intergenic regions showed a correlation with H3K27me3 level (Figure 2.36). We cannot exclude the possibility that this correlation is due to the depletion of histones or high turn-over rate on transcriptionally active intergenic regions, but it is not likely since most intergenic regions even at 31˚C have levels of transcription too low to cause significant histone depletion. These observations support the idea that PRC1 recruitment might be contributed in part by a ubiquitous low level H3K27me3.

Altogether, these results suggest that E(Z) represses transcription genome-wide in multiple ways via H3K27me2/3; 1) by antagonizing active histone marks (H3K27ac, H3K4me1) 2) by recruiting PRC1 all over the genome.

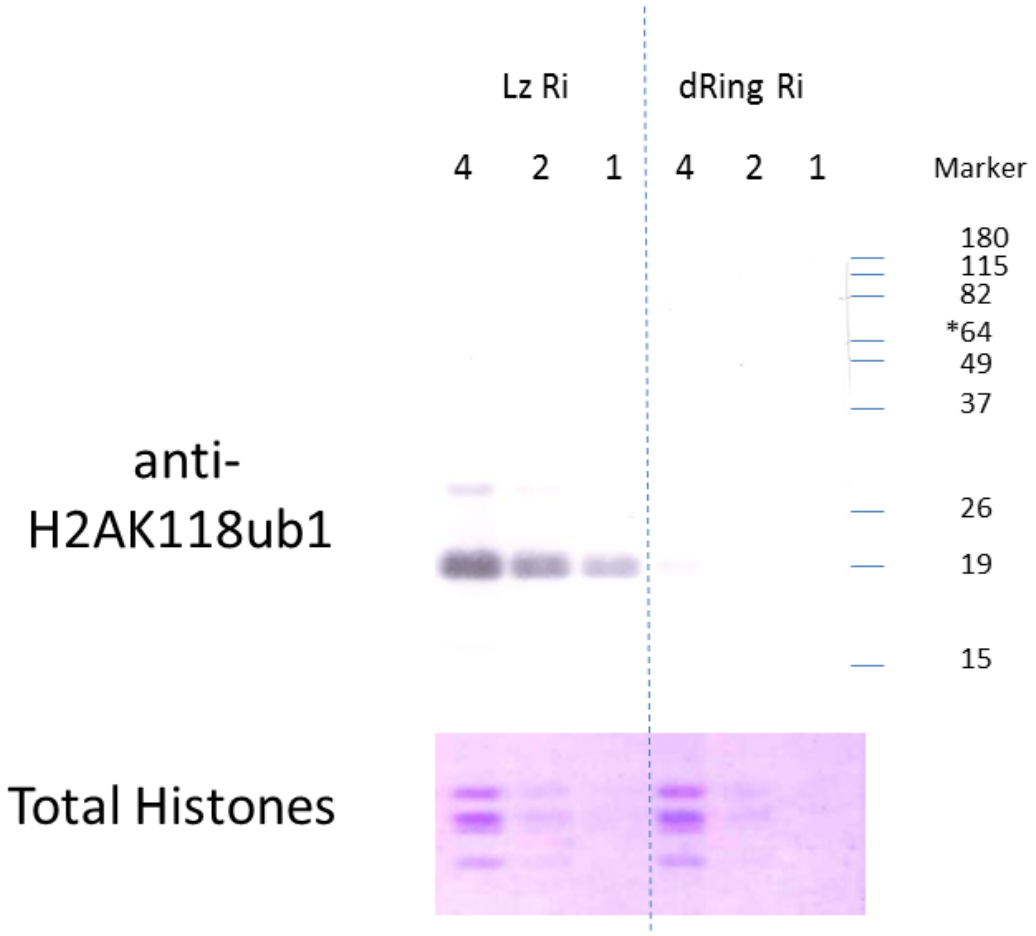
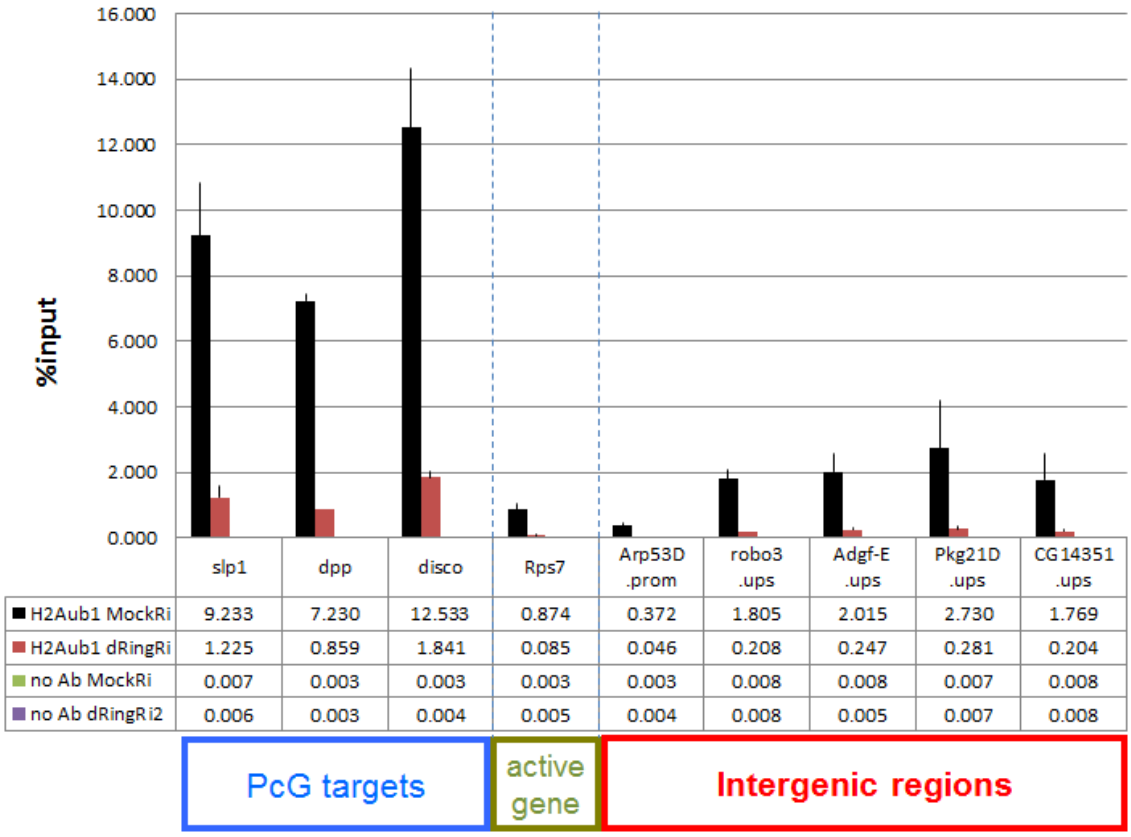


Figure 2.30 Decrease of global H2Aub1 level after dRing depletion

The global H2AK118ub1 levels in nuclear extracts of the S2 cells that are treated with dsRNA targeting LacZ (control) or dRing are determined by western blot analysis. Two-fold serial dilutions of lysate were loaded. Total histones detected by coomassie blue staining served as loading controls.



	slp1	dpp	disco	Rps7	Arp53D .prom	robo3 .ups	Adgf-E .ups	Pkg21D .ups	CG14351 .ups
H2Aub1 MockRi	9.233	7.230	12.533	0.874	0.372	1.805	2.015	2.730	1.769
H2Aub1 dRingRi	1.225	0.859	1.841	0.085	0.046	0.208	0.247	0.281	0.204
no Ab MockRi	0.007	0.003	0.003	0.003	0.003	0.008	0.008	0.007	0.008
no Ab dRingRi2	0.006	0.003	0.004	0.005	0.004	0.008	0.005	0.007	0.008

Figure 2.31 H2Aub1 enrichment in PcG and intergenic non-PcG target regions

ChIP assays are performed to measure the H2AK118ub1 levels in S2 cells that are treated with dsRNA targeting LacZ (control) or dRing. H2Aub1 levels in six PcG target regions (*slp1*, *dpp*, and *disco*), one active gene region (Rps7) and five intergenic non-PcG target regions (upstream of *Arp53D*, *robo3*, *Adgf-E*, *Pkg21D*, and *CG14351*) were measured by qPCR.

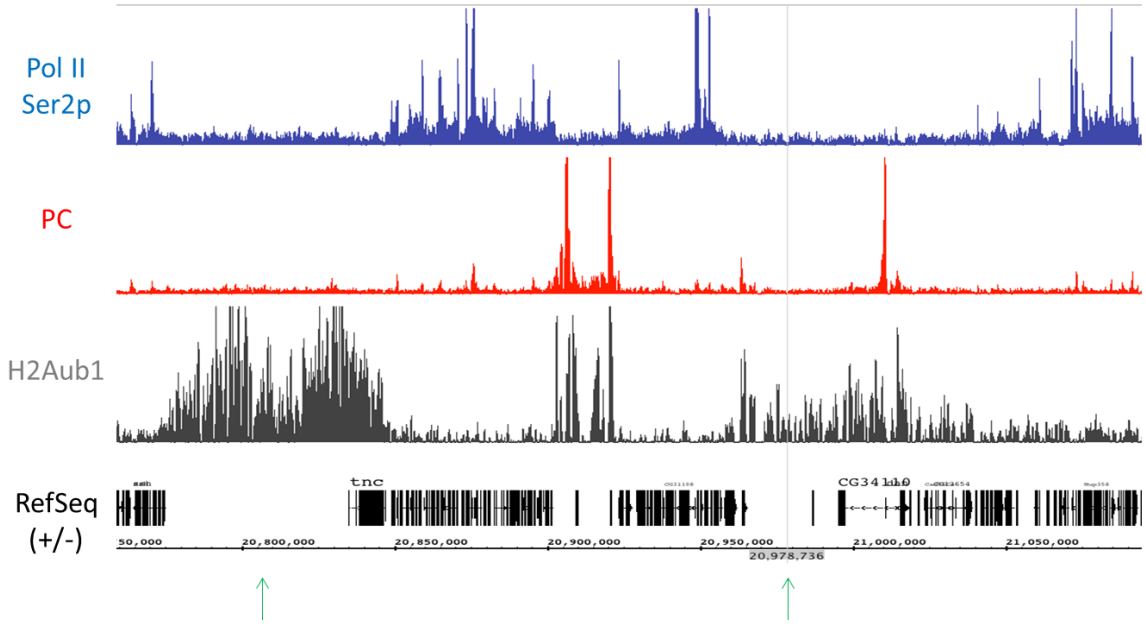


Figure 2.32 H2Aub1 distribution in PcG and intergenic non-PcG target regions

The distributions of Ser-2 phosphorylated Pol II (blue), PC (red), and H2Aub1 (gray) in EZ2-2 cells at 25˚C were mapped by ChIP-seq. Merged positions of annotated transcripts from both strands are indicated in the RefSeq track (black). Selected intergenic regions that shows strong H2Aub1 enrichment without stable PC binding were indicated by arrows

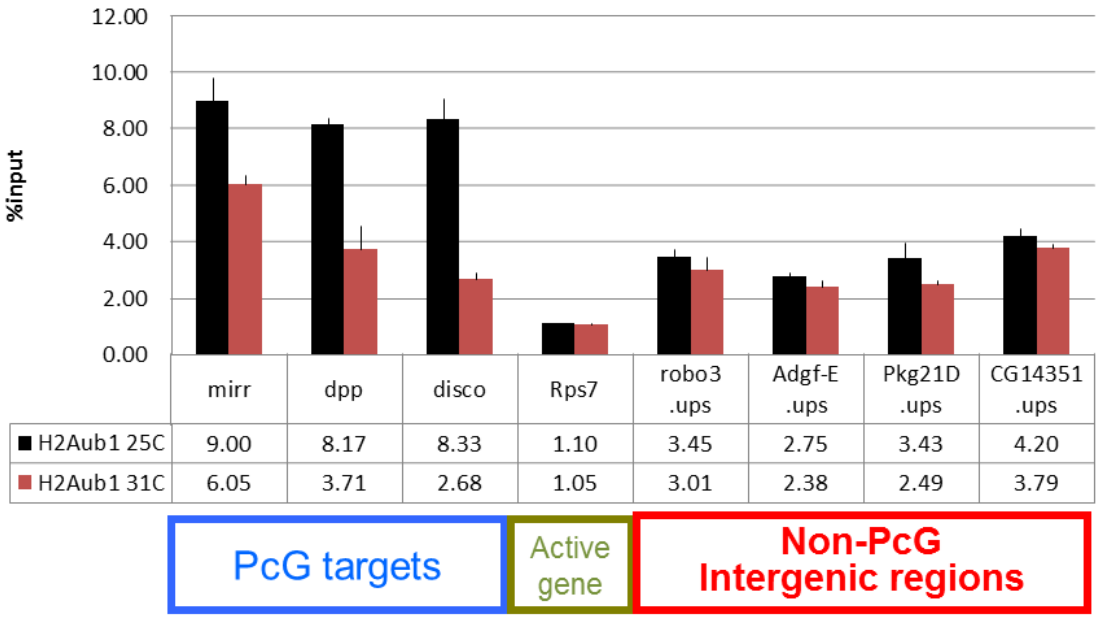


Figure 2.33 Changes in H2Aub1 levels at PcG and intergenic non-PcG target regions after E(Z) inactivation

The H2Aub1 levels of PcG and non-PcG target regions in EZ2-2 cells at 25˚C (black) and 31˚C (red) were measured by ChIP-qPCR.

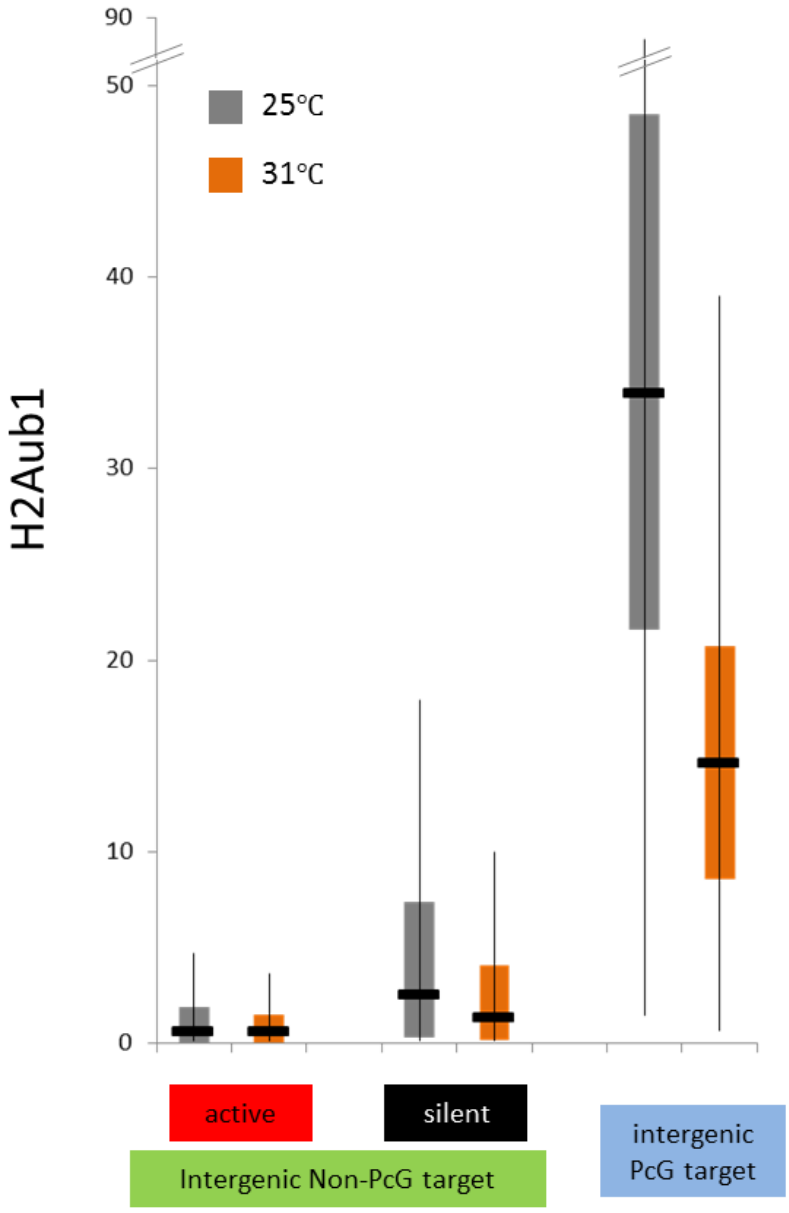


Figure 2.34 Changes in H2Aub1 levels at intergenic regions after E(Z) inactivation

The H2Aub1 levels of intergenic non-PcG (actively transcribing regions and silent regions) and PcG target regions in EZ2-2 cells at 25˚C (gray) and 31˚C (orange) were analyzed from ChIP-seq data.

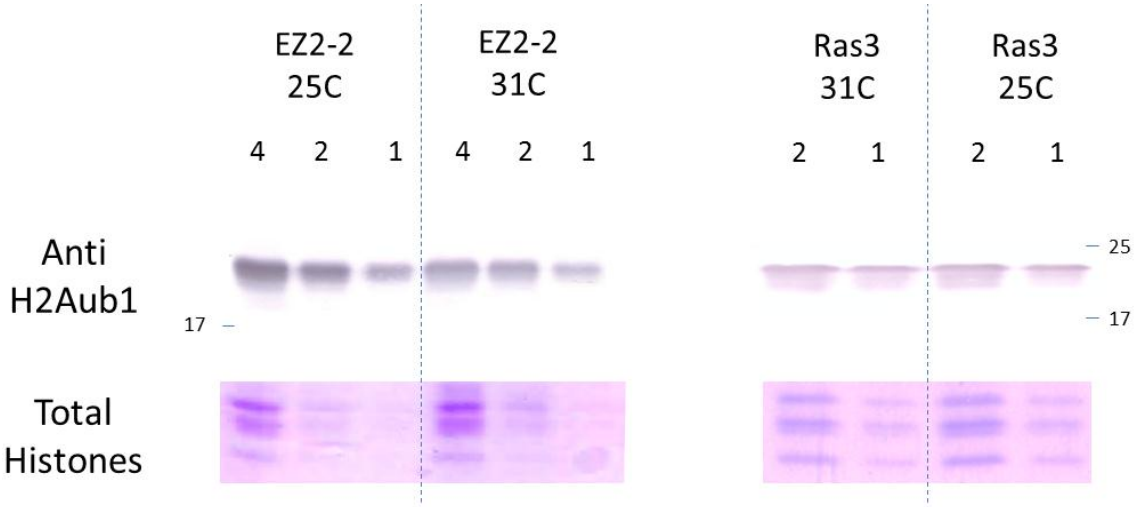


Figure 2.35 Changes in global H2Aub1 levels after E(Z) inactivation

The global H2AK118ub1 levels in nuclear extracts of the EZ2-2 cells and Ras3 (w.t. as control) cells from 25˚C and 31˚C are determined by western blot analysis. Two-fold serial dilutions of lysate were loaded. Total histones detected by coomassie blue staining served as loading controls.

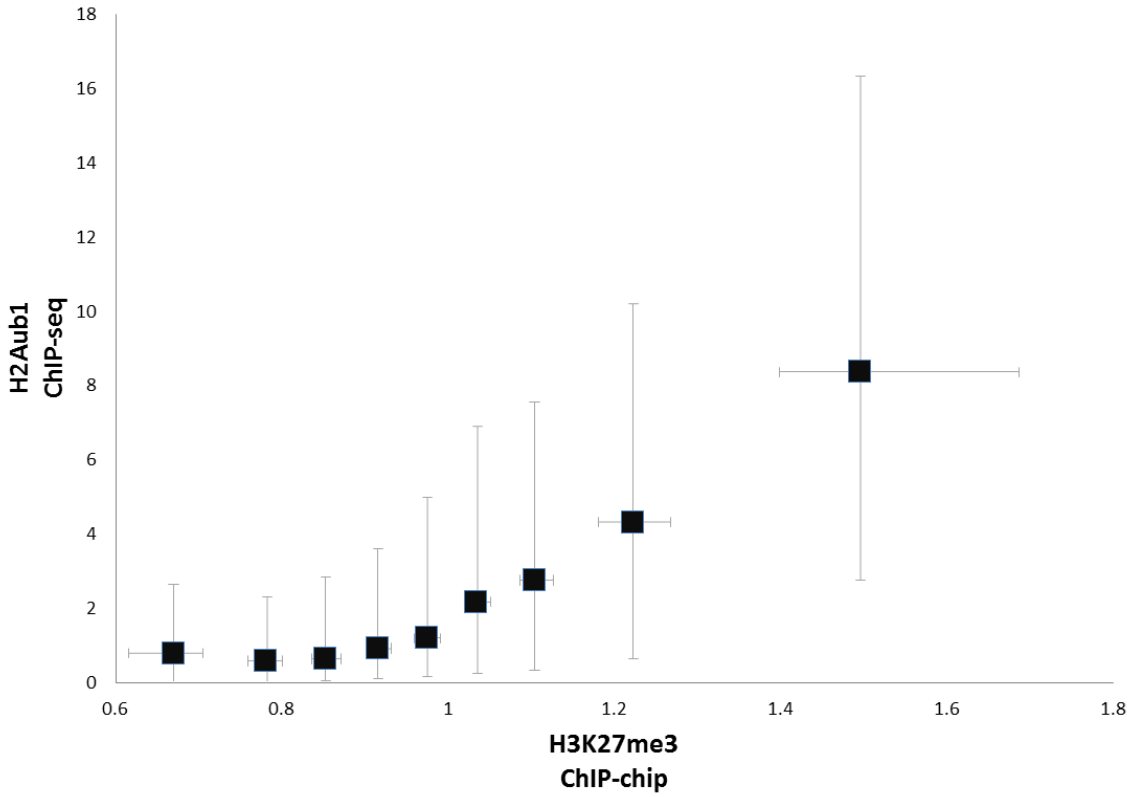


Figure 2.36 Correlation between H2Aub1 levels and H3K27me3 levels in intergenic regions

The H3K27me3 levels and H2Aub1 levels of all intergenic non-PcG target regions in EZ2-2 cells at 25˚C were analyzed from genome-wide distribution data. All non-PcG intergenic regions are sorted by H3K27me3 level and divided by nine groups each of which contains one thousand intergenic regions. The quartile values of the collected H2Aub1 levels for those groups are shown. Black rectangle marks represent the median values, and the negative and positive error bars represent the first and third quartiles.

2.4 Discussion

Genome-wide analysis of H3K27me2 distribution indicates that most of the genome except transcriptionally active or H3K27me3-enriched PcG target regions is covered by H3K27me2. Further ChIP-qPCR analysis on many H3K27me2-enriched intergenic regions and silent coding regions before and after the E(Z) inactivation confirmed that the H3K27me2 signal is genuine and E(Z) is responsible for it. Expression analysis by RT-qPCR and RNA-seq also shows that the H3K27me2-enriched regions start to transcribe when E(Z) is inactivated.

Although it is obvious that E(Z) writes the H3K27me2, we were not able to detect E(Z) binding on these H3K27me2 enriched regions whereas E(Z) binding on PcG target regions can be observed on the PREs. It is also notable that E(Z) binding can be found in sharp peaks at the PRE sites while H3K27me3 distribution forms a broad domain over at least several kilobases. One way to interpret the discrepancy would be that the transient binding might not be able to be detected by ChIP experiments. Since we can observe the H3K27me2 in more than half of the genome, PRC2 is thought to be able to access all over the genome transiently. One possible way to do so would be a random hit and run. There is another possibility that E(Z) interacts with PCNA and methylates histones during DNA replication (Hansen *et al.*, 2008).

The widespread increase of H3K27ac and H3K4me1 after E(Z) inactivation led us to the idea that dUTX complexes (CBP/dUTX and TRR/dUTX) might have access throughout the genome, probably as the same manner with PRC2, and compete with PRC2 all over the genome. Supporting this idea, the genome-wide pattern of increased levels of these

two active marks corresponded well to the levels of transcription de-repression (Figure 2.15, Figure 2.20, Figure 2.27), which are also dependent on the dUTX-containing complexes (Figure 2.22). The active removal of H3K27me2 by de-methylation in actively transcribing non-PcG targets is also supported by the fact that H3K27me2 is lower and H3K27me1 is high in the gene body of active genes (Figure 2.12). One study of the H3K27 de-methylation process by mammalian UTX during muscle differentiation suggests that UTX spreads to de-methylate nucleosomes along the gene body when the target genes become de-repressed (Seenundun *et al.*, 2010), which further supports the idea that the removal of H3K27me marks by UTX occurs through accessing across the gene body. Still, we cannot exclude the possibility that the high H3K27me1 is due to de novo methylation on new nucleosomes that are deposited by nucleosome turnover.

Based on these observations, we hypothesize that E(Z) and dUTX can roam all over the genome, and the competition between E(Z) and dUTX in methylating and de-methylating H3K27 results in the genome-wide H3K27me2. Since tri-methylation of H3K27 by PRC2 is slower than di-methylation, writing H3K27me3 is thought to require much longer residence of PRC2, accounting for the widespread H3K27me2 in most of the genome and some remaining of H3K27me3 in H3K27me2 enriched regions. And the H3K27me2-enriched chromatin state serves as a default repressive state that reduces transcriptional noise from intergenic regions and down-regulates gene activity all over the genome.

Although our results demonstrate that the H3K27me2/3 represses most of the genome, there is still an open question that remains to be answered: how does the methylation of H3K27 contribute to the transcription repression?

One interesting possibility is that decreased H3K27me2/3 levels result in the increased DNA accessibility. One important role of H3K27me that is observed in our results is antagonizing active histone marks including H3K27ac and H3K4me1. The active enhancer regions, where H3K27ac and H3K4me1 are enriched, are known to be highly accessible, as measured by nuclease hypersensitivity mapping (Gross and Garrard, 1988), and exhibit high nucleosome turn-over rate. These active marks might contribute to the increased DNA accessibility, which in turn results in more Pol II and/or other transcriptional activators access to the cryptic promoters in intergenic regions.

In accordance with this idea, Ostuni *et al.* (2013) observed concomitant increase of H3K27ac and H3K4me1 with increased DNA accessibility during the activation of latent enhancers. Notably, CBP/dUTX complex contains BRM, a SWI/SNF ATP-dependent chromatin remodeler, which can open up the chromatin so the DNA become more accessible (Peterson, 2002; Smith *et al.*, 2003; Vignali *et al.*, 2000). So, it is very possible that the increased DNA accessibility after E(Z) inactivation contributes to the increased transcription in the silent intergenic and coding regions.

In addition, there can be possible indirect effects of E(Z) inactivation that can contribute to the de-repression of intergenic regions. Since many of the PcG target genes are transcriptional regulators, it is possible that the de-repressed transcription activators might contribute to the increase in some intergenic transcription. However, it is not likely that these transcription activators bind all over the genome and increase the transcription from all over the intergenic regions as we observed in RNA-seq and RT-qPCR results. Increase in the expression of Pol II subunits might also affect the global de-repression although the expected levels of de-repression from it (2- to 5-fold which we can expect

from the increased expression levels of Pol II subunits) would be far less than that we are observing on silent intergenic or coding regions (5- to 50-fold).

Still, many questions need to be answered to understand the actual mechanisms of intergenic de-repression upon E(Z) inactivation. Does the novel transcription arise from defined coding regions? How long are the novel transcripts? Where does the transcription start? To understand the observed intergenic transcription, it is important to characterize the intergenic RNA species.

RNA-seq results show that many expression tags appear at most intergenic regions after E(Z) inactivation. Notably, most of them do not form distinct peaks but are often scattered at very low level (Figure 2.19). Since the de-repression in many intergenic regions could be detected robustly from multiple independent experiments by RT-qPCR, which uses 100-200bp size amplicons for the detection, these intergenic transcripts seem to be relatively stable and a significant proportion of them must be longer than 200bp. The scattered weak signals seen in RNA-seq could be due to insufficient sequencing depth as we often failed to see the expression tags where the intergenic de-repression is obvious in RT-qPCR results (e.g. *Adgf-E* upstream region; Figure 2.25). Or they could be due to extensive degradation by exosome or related surveillance mechanisms.

Since the intergenic regions showed an increase of H3K27ac and H3K4me1, which can be found in active enhancers, it is possible that the transcripts from the intergenic de-repression might be similar to those from active enhancers. Kim *et al.* (2010) observed that un-phosphorylated and Ser5 phosphorylated Pol II can be found at extragenic enhancers and transcribes enhancer RNAs (eRNAs) bi-directionally within the 2kb

flanking regions (cumulative expression tags peak at around 500bp) of the enhancers where H3K4me1 is enriched. Although it is difficult to estimate the length of the eRNAs since the expression tags are also scattered around the enhancer regions, these observations suggest that the majority of eRNAs are approximately 500bp long and the maximal length can be up to 2kb. Notably, these eRNAs mostly lack poly(A) tails. These observations indicate that the eRNAs might be the products from the Pol II transcription cycle somewhere between the early elongation step and the productive elongation step.

Some results from our preliminary study for the Pol II binding at the intergenic regions with several different antibodies against various modified CTD versions Pol II further indicated that the intergenic transcripts might be the products of the Pol II machinery in the early-elongation step. ChIP-qPCR results of several intergenic regions showed slightly increased total Pol II binding (using an antibody against Rpb3) although the detected signals are very weak compared to those from gene promoters. Interestingly, only Ser-5 phosphorylated Pol-II showed marginal increase while Ser-2 phosphorylated Pol-II and un-phosphorylated Pol II (8WG16) did not show increase at 31˚C, indicating that most of the gained Pol II in intergenic regions is Ser5 phosphorylated Pol II. Since the phosphorylation at Ser5 occurs during the transition between initiation and early elongation step of transcription cycle, Pol II at the intergenic regions at 31˚C is thought to be the state after the initiation step. And the fact that Ser-2 phosphorylated Pol-II did not show increase implies that the Pol II has not reached to the productive elongation step.

Further analysis for characterizing the RNAs from intergenic regions would be necessary to understand the repressive mechanism by methylated H3K27 in intergenic regions. Identifying the location of transcription start sites would let us know whether these

transcripts are originated from distinct cryptic promoters or from random genomic sites. Especially the information for the length of the transcripts would give clues to understand the state of Pol II machinery at intergenic regions.

Our results suggest that PRC1 is also involved in intergenic repression; PC knock-down increased the intergenic transcription and a *Pc* mutant cell line exhibited significant level of intergenic transcription. PRC1 is thought to have a repressive function by itself through H2A mono-ubiquitination and/or chromatin compaction. Since the PRC1 recruitment is at least partly dependent on H3K27me3 at many PREs, the possibility that the low but significant level of H3K27me3 in intergenic regions contribute to the PRC1 recruitment needed to be examined. For this, determining H2Aub1 level on the intergenic regions was crucial since PRC1 can generate the H2Aub1 mark by its component dRING.

So far, the genome-wide distribution of H2Aub1 in *Drosophila* could not be properly analyzed because the specific antibody for the mono-ubiquitination of H2AK118 in *Drosophila* was not available. Lagarou *et al.* (2008) used a commercially available monoclonal H2Aub1 antibody E6C5 which has been used for mammalian cells, but Scheuermann *et al.* (2010) claimed that they were not able to specifically monitor H2Aub1 levels in *Drosophila* with this antibody. Similar failure to detect H2Aub1 in *Drosophila* with the E6C5 antibody was experienced in our laboratory and was reported by other researchers (unpublished). In this study, we were able to map the genome-wide distribution of H2Aub1 by using a newly verified antibody that specifically recognizes H2Aub1 in *Drosophila*.

Interestingly, H2Aub1 enrichment was not limited to PRE regions and can be found on many non-PcG target regions including silent intergenic regions. dRING is so far the only known mono-ubiquitinase for H2AK118 and knock-down of dRING virtually abolishes detectable H2Aub1. Since the distribution of dRING is mostly overlapping with PREs, these H2Aub1 on non-PcG targets are likely to be generated by transient binding of dRING-containing complexes.

The distribution of H2Aub1 was correlated with H3K27me3 levels in non-PcG intergenic regions (Figure 2.36), and the H2Aub1 levels in the silent intergenic regions decreased after E(Z) inactivation suggesting that PRC1 may be transiently recruited to the intergenic regions by H3K27me3. It is also possible that highly enriched H3K27me2 might also contribute to the PRC1 recruitment since PC still has a significant affinity to H3K27me2 (Fischle *et al.*, 2003). Still, the observed strong de-repression in the intergenic regions after E(Z) inactivation by temperature shifting is likely to be mostly due to the decrease of H3K27me2/3 itself since the decrease in the level of H2Aub1 is proportionally lower and arguably insufficient to account for it (Figure 2.33-35).

Several additional observations on the global H2Aub1 levels in various PcG mutant cell lines were conducted to clarify the sources of the H2Aub1. Consistent with the result from the EZ2-2 cell line at 31˚C, the global H2Aub1 level in a *Su(z)12* mutant cell line was approximately 30% less than that of wild type cells (Figure 2.37), indicating the proportion that is dependent on PRC2 function . In a *Pc* mutant cell line, the global H2Aub1 level was about a half of the level in the wild-type cell line (Figure 2.37). Since dRing is responsible for most (or possibly all) of the total H2Aub1 levels as can be seen from dRing knock-down experiments (Figure 2.30-31), the other half of the H2Aub1 is

thought to be written by the other dRing-containing complex(es) such as dRAF complex. Intriguingly, a *Psc/Su(z)2* mutant cell line, which has a deletion removing the *Psc-Su(z)2* locus, also showed a reduction of H2Aub1 level similar to that of the *Pc* mutant cell line (Figure 2.37). As both PRC1 and dRAF complexes include PSC as a subunit, it was an unexpected result. These results suggest that there might be some other dRING-containing complex that does not rely on PSC. In fact, a gene named *lethal(3)73Ah* encodes a protein that is paralogous to PSC/SU(Z)2 and mammalian MEL-18/BMI-1 (Irminger-Finger and Nöthiger, 1995) (Figure 2.38). So it is a very interesting possibility that the L(3)73AH protein might form a complex with dRING and ubiquitinate H2AK118 in *Drosophila*.

Further analysis would be necessary to characterize the source of the H2Aub1 in intergenic regions. Since *Pc* mutant cells showed a significant level of transcription in the intergenic regions and a mild decrease of H2Aub1 level was observed after E(Z) inactivation, at least some proportion of H2Aub1 in intergenic regions is likely to be written from PRC1. Further experiments including knocking down of dKDM2, a core component of dRAF complex, or L(3)73AH and monitoring H2Aub1 levels and transcriptional levels on intergenic regions would be necessary to understand the H2Aub1-mediated intergenic repression.

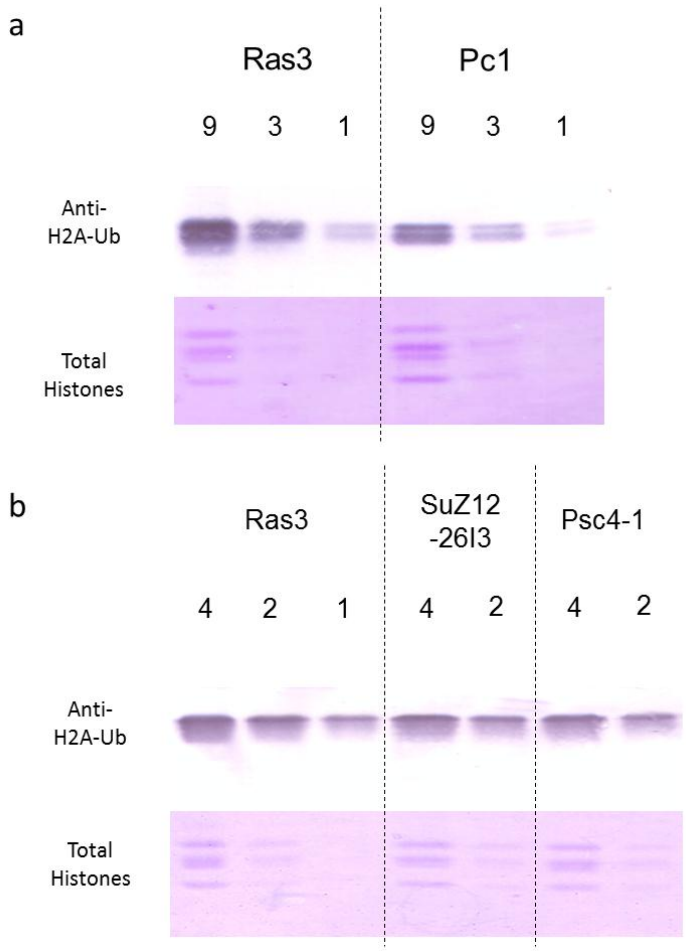


Figure 2.37 Global H2Aub1 levels in *Pc*, *Psc*, and *Su(z)12* mutant cells

The global H2AK118ub1 level of the Ras3 (w.t. as control) cells are compared with (a) *Pc*, (b) *Su(z)12* and *Psc* mutant cells. Three-fold serial dilutions of lysate from *Pc* mutant cells and two-fold serial dilutions of lysates from *Su(z)12* and *Psc* mutant cells were loaded. Total histones detected by coomassie blue staining served as loading controls.

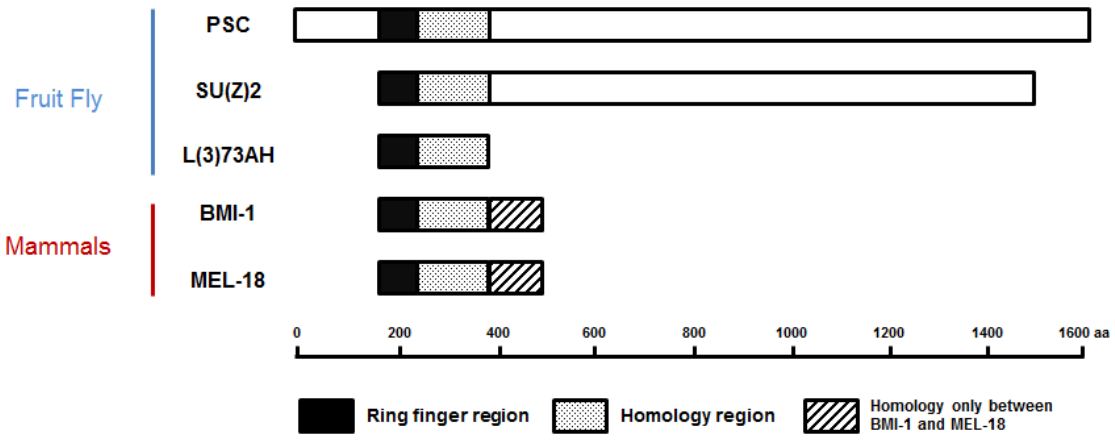


Figure 2.38 Schematic presentation of the ring finger proteins

Directly adapted from Irminger-Finger and Nöthiger (1995). Five ring finger proteins, PSC, SU(Z)2, L(3)73AH, MEL-18, and BMI-1, and relative locations of homologous regions are shown. Ring finger region and homology region are common to all five proteins, the C-terminal proline-serine rich region is present only in MEL-18 and BMI-1.

From the evolutionary point of view, it is possible that PRC2 has exerted a repressive role for the suppression of pervasive transcription from ancient organisms and PRC1 was developed later for more robust and stronger repression on selected, potentially more important regions. This scenario fits well with the fact that the stable binding of PcG proteins can be observed only at the PREs where H3K27me3 is maintained by the stably bound PRC2, while genome-wide H3K27me2 distribution, which can be considered as a default repressive chromatin state, is produced by PRC2 random access in *Drosophila*. Moreover, PRC2 is much more conserved than PRC1 in various organisms. In *C. elegans* there are no well-defined homologs of PRC1 components although there are some candidates for RING-containing proteins (Karakuzu *et al.*, 2009; Sanchez-Pulido *et al.*, 2008) whereas PRC2 components are well-conserved. Another interesting possibility is that the H2Aub1 ubiquitinase complexes containing dRING and PSC might have been functioning independently of H3K27me3 in ancient organisms and the chromodomain-containing PC was invented later to exert more robust repression by both H3K27me3 and H2Aub1 for the regions that needs to be regulated extensively which are now recognized as PREs.

Chapter 3. Summary and Conclusion

Unlike usual histone modification marks, our genome-wide data shows that H3K27me2 is enriched in more than half of the genome in various *Drosophila* cell lines. The H3K27me2-enriched regions encompass most of silent intergenic and coding regions, suggesting that this mark might have a transcriptional repressive role. Indeed, inactivation of E(Z), the sole H3K27me2/3 methyl-transferase in *Drosophila*, results in genome-wide de-repression; a significant proportion of silent intergenic and coding regions exhibit transcription, indicating that PRC2 function is repressing transcription from most of the genome.

Since the H3K27me2-enriched regions are covering more than half of the genome where the transcription is generally absent, this chromatin state might be considered as a default chromatin state when there is no active/repressive signal for transcription. This repressive default state becomes largely perturbed by E(Z) inactivation; as H3K27me decreases, most of the H3K27me2-enriched regions showed de-repression. The global de-repression was dependent on the function of the two dUTX-containing H3K27 de-methylase complexes, which are also the writers of active histone marks, H3K27ac and H3K4me1. Further genome-wide analysis on these active marks revealed that these marks increase globally when E(Z) function is impaired, suggesting that the PRC2 and dUTX-containing complexes are antagonistic to each other and compete over the whole genome.

So, it is tempting to hypothesize that E(Z) constantly methylates H3K27 all over the genome and the dUTX-containing complexes de-methylate it, and the competition between the E(Z) and dUTX results in the widespread H3K27me2 as a dynamic

equilibrium in most of the genome. In this manner, methylation of H3K27 increases the threshold for transcription activation and maintains genome-wide repression on the intergenic regions by antagonizing dUTX and other histone modifiers for active histone marks (H3K27ac and H3K4me1) (Figure 3.1).

In addition to PRC2, several lines of evidence indicated that PRC1 is also involved in the intergenic repression. Knock-down or mutation in PRC1 genes caused higher transcription levels in intergenic regions. Moreover, significant level of H2Aub1, a repressive mark that can be written by PRC1, in most intergenic regions was observed.

Although it is not so clear whether this PRC1-mediated repression in intergenic regions is dependent to PRC2 function, our observations suggest partial dependency between the PRC1 and PRC2 for the intergenic repression; the H2Aub1 level on the intergenic regions mildly decreased when E(Z) is inactivated; the level of H3K27me3 in these regions showed correlation with the observed H2Aub1 level, suggesting that PRC1 might interact with most of the intergenic regions through the presence of H3K27me3 and contribute to the repression.

Based on these observations, we conclude that PcG proteins raise the threshold for intergenic transcription by H3K27me2/3 and H2Aub1 and contribute to reduce the transcriptional noise and to create a specific requirement for activators to promote gene activity (Figure 3.2).

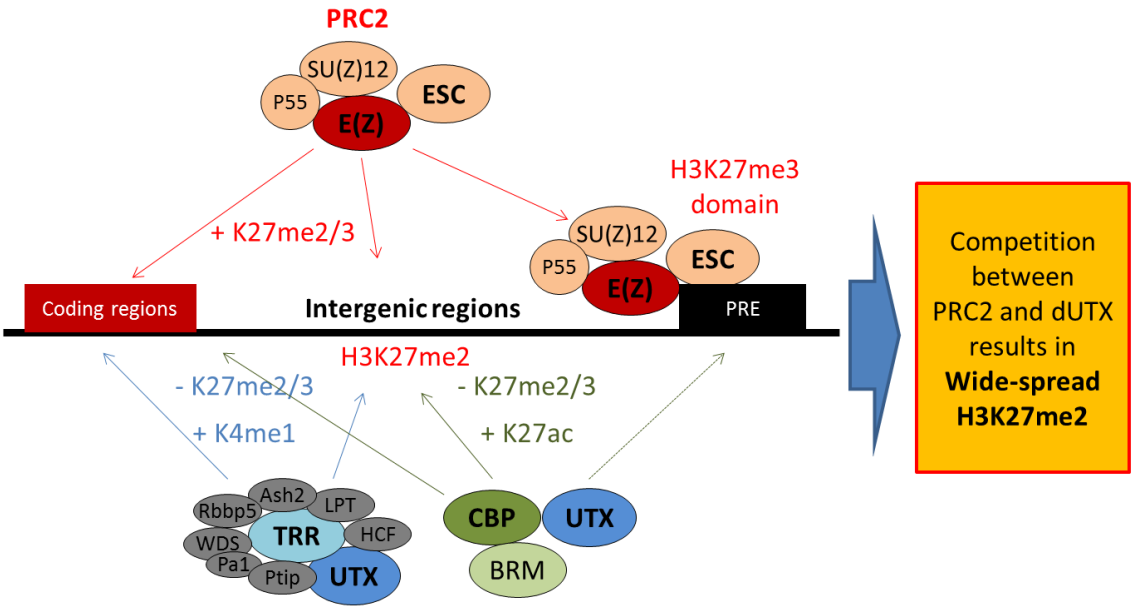


Figure 3.1 Model for genome-wide repression by PRC2

The PRC2 not only represses PcG target regions by H3K27me3 but also suppresses the transcription in most intergenic regions and coding regions by H3K27me2/3. Increased levels of H3K27ac and H3K4me1 after E(Z) inactivation in most of the genome suggest that the PRC2 and dUTX-containing complexes are antagonistic to each other and compete over the whole genome.

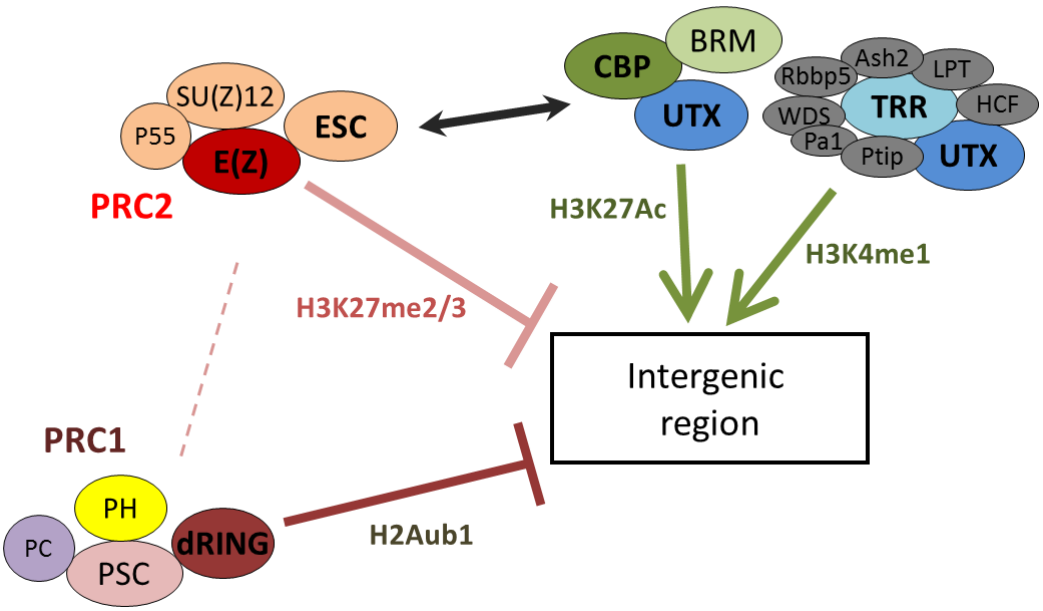


Figure 3.2 Intergenic repression by PRC1 and PRC2

Both PRC2 and PRC1 raise the threshold for intergenic transcription by H3K27me2/3 and H2Aub1, respectively.

References

Ansari, A., and Hampsey, M. (2005). A role for the CPF 3'-end processing machinery in RNAP II-dependent gene looping. Genes Dev. *19*, 2969–2978.

Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J. *30*, 2817–2828.

Beisel, C., and Paro, R. (2011). Silencing chromatin: comparing modes and mechanisms. Nat. Rev. Genet. *12*, 123–135.

Beisel, C., Imhof, A., Greene, J., Kremmer, E., and Sauer, F. (2002). Histone methylation by the Drosophila epigenetic transcriptional regulator Ash1. Nature *419*, 857–862.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., *et al.* (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Berretta, J., and Morillon, A. (2009). Pervasive transcription constitutes a new level of eukaryotic genome regulation. EMBO Rep. *10*, 973–982.

Birney, E., Stamatoyannopoulos, J.A., Dutta, A., Guigó, R., Gingeras, T.R., Margulies, E.H., Weng, Z., Snyder, M., Dermitzakis, E.T., Thurman, R.E., *et al.* (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Boyer, L.A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L.A., Lee, T.I., Levine, S.S., Wernig, M., Tajonar, A., Ray, M.K., *et al.* (2006). Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature *441*, 349–353.

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

Brown, J.L., Grau, D.J., DeVido, S.K., and Kassis, J.A. (2005). An Sp1/KLF binding site is important for the activity of a Polycomb group response element from the Drosophila engrailed gene. Nucleic Acids Res. *33*, 5181–5189.

Buchwald, G., van der Stoop, P., Weichenrieder, O., Perrakis, A., van Lohuizen, M., and Sixma, T.K. (2006). Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. EMBO J. *25*, 2465–2474.

Busturia, A., and Bienz, M. (1993). Silencers in abdominal-B, a homeotic Drosophila gene. EMBO J. *12*, 1415–1425.

Calvo, O., and Manley, J.L. (2003). Strange bedfellows: polyadenylation factors at the promoter. Genes Dev. *17*, 1321–1327.

Cao, R., and Zhang, Y. (2004). The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. Curr. Opin. Genet. Dev. *14*, 155–164.

Cao, R., Tsukada, Y., and Zhang, Y. (2005). Role of BMi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol. Cell *20*, 845–854.

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

Carrington, E., and Jones, R. (1996). The Drosophila Enhancer of zeste gene encodes a chromosomal protein: examination of wild-type and mutant protein distribution. Development *122*, 4073–4083.

Celniker, S.E., Dillon, L.A.L., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., Kellis, M., Lai, E.C., Lieb, J.D., MacAlpine, D.M., *et al.* (2009). Unlocking the secrets of the genome. Nature *459*, 927–930.

Cheng, B., and Price, D.H. (2007). Properties of RNA polymerase II elongation complexes before and after the P-TEFb-mediated transition into productive elongation. J. Biol. Chem. *282*, 21901–21912.

Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., *et al.* (2005). Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. Science *308*, 1149–1154.

Chiang, A., O'Connor, M., Paro, R., Simon, J., and Bender, W. (1995). Discrete Polycomb-binding sites in each parasegmental domain of the bithorax complex. Development *121*, 1681–1689.

Chopra, V.S., Hendrix, D.A., Core, L.J., Tsui, C., Lis, J.T., and Levine, M. (2011). The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. Mol. Cell *42*, 837–844.

Clemens, J.C., Worby, C.A., Simonson-Leff, N., Muda, M., Maehama, T., Hemmings, B.A., and Dixon, J.E. (2000). Use of double-stranded RNA interference in Drosophila cell lines to dissect signal transduction pathways. Proc. Natl. Acad. Sci. U. S. A. *97*, 6499–6503.

Core, L.J., Waterfall, J.J., Gilchrist, D.A., Fargo, D.C., Kwak, H., Adelman, K., and Lis, J.T. (2012). Defining the status of RNA polymerase at promoters. Cell Rep. *2*, 1025–1035.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., *et al.* (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. U. S. A. *107*, 21931–21936.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002). Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. Cell *111*, 185–196.

Davis, C.A., and Ares, M. (2006). Accumulation of unstable promoter-associated transcripts upon loss of the nuclear exosome subunit Rrp6p in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. U. S. A. *103*, 3262–3267.

Déjardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., and Cavalli, G. (2005). Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature *434*, 533–538.

Dellino, G.I., Schwartz, Y.B., Farkas, G., McCabe, D., Elgin, S.C.R., and Pirrotta, V. (2004). Polycomb silencing blocks transcription initiation. Mol. Cell *13*, 887–893.

Dinger, M.E., Amaral, P.P., Mercer, T.R., and Mattick, J.S. (2009). Pervasive transcription of the eukaryotic genome: functional indices and conceptual implications. Brief. Funct. Genomic. Proteomic. *8*, 407–423.

Dorighi, K.M., and Tamkun, J.W. (2013). The trithorax group proteins Kismet and ASH1 promote H3K36 dimethylation to counteract Polycomb group repression in Drosophila. Development.

Dutrow, N., Nix, D.A., Holt, D., Milash, B., Dalley, B., Westbroek, E., Parnell, T.J., and Cairns, B.R. (2008). Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. Nat. Genet. *40*, 977–986.

Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. Genes Dev. *18*, 2973–2983.

Farcas, A.M., Blackledge, N.P., Sudbery, I., Long, H.K., McGouran, J.F., Rose, N.R., Lee, S., Sims, D., Cerase, A., Sheahan, T.W., *et al.* (2012). KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife *1*, e00205.

Faucheux, M., Roignant, J.-Y., Netter, S., Charollais, J., Antoniewski, C., and Théodore, L. (2003). batman Interacts with polycomb and trithorax group genes and encodes a BTB/POZ protein that is included in a complex containing GAGA factor. Mol. Cell. Biol. *23*, 1181–1195.

Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., *et al.* (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212–224.

Fischle, W., Wang, Y., Jacobs, S.A., Kim, Y., Allis, C.D., and Khorasanizadeh, S. (2003). Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. Genes Dev. *17*, 1870–1881.

Fitzgerald, D.P., and Bender, W. (2001). Polycomb group repression reduces DNA accessibility. Mol. Cell. Biol. *21*, 6585–6597.

Francis, N.J., Saurin, A.J., Shao, Z., and Kingston, R.E. (2001). Reconstitution of a functional core Polycomb repressive complex. Mol. Cell *8*, 545–556.

Francis, N.J., Kingston, R.E., and Woodcock, C.L. (2004). Chromatin compaction by a Polycomb group protein complex. Science (80-.). *306*, 1574–1577.

Gearhart, M.D., Corcoran, C.M., Wamstad, J.A., and Bardwell, V.J. (2006). Polycomb group and SCF ubiquitin ligases are found in a novel BCOR complex that is recruited to BCL6 targets. Mol. Cell. Biol. *26*, 6880–6889.

Grau, D.J., Chapman, B.A., Garlick, J.D., Borowsky, M., Francis, N.J., and Kingston, R.E. (2011). Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes Dev. *25*, 2210–2221.

Gross, D.S., and Garrard, W.T. (1988). Nuclease hypersensitive sites in chromatin. Annu. Rev. Biochem. *57*, 159–197.

Hahn, S. (2004). Structure and mechanism of the RNA polymerase II transcription machinery. Nat. Struct. Mol. Biol. *11*, 394–403.

Hansen, K.H., Bracken, A.P., Pasini, D., Dietrich, N., Gehani, S.S., Monrad, A., Rappsilber, J., Lerdrup, M., and Helin, K. (2008). A model for transmission of the H3K27me3 epigenetic mark. Nat. Cell Biol. *10*, 1291–1300.

He, J., Shen, L., Wan, M., Taranova, O., Wu, H., and Zhang, Y. (2013). Kdm2b maintains murine embryonic stem cell status by recruiting PRC1 complex to CpG islands of developmental genes. Nat. Cell Biol.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., *et al.* (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Herz, H.-M., Mohan, M., Garruss, A.S., Liang, K., Takahashi, Y.-H., Mickey, K., Voets, O., Verrijzer, C.P., and Shilatifard, A. (2012). Enhancer-associated H3K4

monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev. *26*, 2604–2620.

Ho, C.K., Sriskanda, V., McCracken, S., Bentley, D., Schwer, B., and Shuman, S. (1998). The guanylyltransferase domain of mammalian mRNA capping enzyme binds to the phosphorylated carboxyl-terminal domain of RNA polymerase II. J. Biol. Chem. *273*, 9577–9585.

Hodgson, J.W., Argiropoulos, B., and Brock, H.W. (2001). Site-specific recognition of a 70-base-pair element containing d(GA)(n) repeats mediates bithoraxoid Polycomb group response element-dependent silencing. Mol. Cell. Biol. *21*, 4528–4543.

Horard, B., Tatout, C., Poux, S., and Pirrotta, V. (2000). Structure of a Polycomb Response Element and In Vitro Binding of Polycomb Group Complexes Containing GAGA Factor. Mol. Cell. Biol. *20*, 3187–3197.

Houalla, R., Devaux, F., Fatica, A., Kufel, J., Barrass, D., Torchet, C., and Tollervey, D. (2006). Microarray detection of novel nuclear RNA substrates for the exosome. Yeast *23*, 439–454.

Hsieh, J.J.-D., Cheng, E.H.-Y., and Korsmeyer, S.J. (2003). Taspase1: a threonine aspartase required for cleavage of MLL and proper HOX gene expression. Cell *115*, 293–303.

Iovino, N., Ciabrelli, F., and Cavalli, G. (2013). PRC2 Controls Drosophila Oocyte Cell Fate by Repressing Cell Cycle Genes. Dev. Cell *26*, 431–439.

Irminger-Finger, I., and Nöthiger, R. (1995). The Drosophila nielanogaster gene lethal(3) 73Ah encodes a ring finger protein homologous to the oncoproteins MEL-18 and BMI-1. Gene *163*, 203–208.

Jensen, T.H., Jacquier, A., and Libri, D. (2013). Dealing with Pervasive Transcription. Mol. Cell *52*, 473–484.

Jones, R.S., and Gelbart, W.M. (1990). Genetic analysis of the enhancer of zeste locus and its role in gene regulation in Drosophila melanogaster. Genetics *126*, 185–199.

Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the Drosophila ubx gene. J. Biol. Chem. *281*, 29064–29075.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., *et al.* (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484–1488.

Karakuzu, O., Wang, D.P., and Cameron, S. (2009). MIG-32 and SPAT-3A are PRC1 homologs that control neuronal migration in Caenorhabditis elegans. Development *136*, 943–953.

Ketel, C.S., Andersen, E.F., Vargas, M.L., Suh, J., Strome, S., and Simon, J.A. (2005). Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. Mol. Cell. Biol. *25*, 6857–6868.

Kim, H., Erickson, B., Luo, W., Seward, D., Graber, J.H., Pollock, D.D., Megee, P.C., and Bentley, D.L. (2010). Gene-specific RNA polymerase II phosphorylation and the CTD code. Nat. Struct. Mol. Biol. *17*, 1279–1286.

King, I.F., Francis, N.J., and Kingston, R.E. (2002). Native and recombinant Polycomb group complexes establish a selective block to template accessibility to repress transcription in vitro. Mol. Cell. Biol. *22*, 7919–7928.

Klymenko, T., Papp, B., Fischle, W., Köcher, T., Schelder, M., Fritsch, C., Wild, B., Wilm, M., and Müller, J. (2006). A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev. *20*, 1110–1122.

Kouzine, F., Wojtowicz, D., Yamane, A., Resch, W., Kieffer-Kwon, K.-R., Bandle, R., Nelson, S., Nakahashi, H., Awasthi, P., Feigenbaum, L., *et al.* (2013). Global Regulation of Promoter Melting in Naive Lymphocytes. Cell *153*, 988–999.

Ku, M., Koche, R.P., Rheinbay, E., Mendenhall, E.M., Endoh, M., Mikkelsen, T.S., Presser, A., Nusbaum, C., Xie, X., Chi, A.S., *et al.* (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet. *4*, e1000242.

Lagarou, A., Mohd-Sarip, A., Moshkin, Y.M., Chalkley, G.E., Bezstarosti, K., Demmers, J.A.A., and Verrijzer, C.P. (2008). dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev. *22*, 2799–2810.

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, H., Ohno, K., Voskoboynik, Y., Ragusano, L., Martinez, A., and Dimova, D.K. (2010). Drosophila RB proteins repress differentiation-specific genes via two different mechanisms. Mol. Cell. Biol. *30*, 2563–2577.

Lee, T.I., Jenner, R.G., Boyer, L.A., Guenther, M.G., Levine, S.S., Kumar, R.M., Chevalier, B., Johnstone, S.E., Cole, M.F., Isono, K., *et al.* (2006). Control of developmental regulators by Polycomb in human embryonic stem cells. Cell *125*, 301–313.

Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev. *24*, 265–276.

Lehmann, L., Ferrari, R., Vashisht, A.A., Wohlschlegel, J.A., Kurdistani, S.K., and Carey, M. (2012). Polycomb repressive complex 1 (PRC1) disassembles RNA polymerase II preinitiation complexes. J. Biol. Chem. *287*, 35784–35794.

Levine, S.S., Weiss, A., Erdjument-Bromage, H., Shao, Z., Tempst, P., and Kingston, R.E. (2002). The core of the polycomb repressive complex is compositionally and functionally conserved in flies and humans. Mol. Cell. Biol. *22*, 6070–6078.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, B., Carey, M., and Workman, J.L. (2007). The Role of Chromatin during Transcription. Cell *128*, 707–719.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, L., Wang, X., Stolc, V., Li, X., Zhang, D., Su, N., Tongprasit, W., Li, S., Cheng, Z., Wang, J., *et al.* (2006a). Genome-wide transcription analyses in rice using tiling microarrays. Nat. Genet. *38*, 124–129.

Li, Z., Cao, R., Myers, M.P., Zhang, Y., and Xu, R.-M. (2006b). structure of a Bmi-1-Ring1B Polycomb group ubiquitin ligase complex. J. Biol. Chem. *281*, 20643–20649.

Manak, J.R., Dike, S., Sementchenko, V., Kapranov, P., Biemar, F., Long, J., Cheng, J., Bell, I., Ghosh, S., Piccolboni, A., *et al.* (2006). Biological function of unannotated transcription during the early development of Drosophila melanogaster. Nat. Genet. *38*, 1151–1158.

Margueron, R., Justin, N., Ohno, K., Sharpe, M.L., Son, J., Drury, W.J., Voigt, P., Martin, S.R., Taylor, W.R., De Marco, V., *et al.* (2009). Role of the polycomb protein EED in the propagation of repressive histone marks. Nature *461*, 762–767.

Marshall, N.F., and Price, D.H. (1995). Purification of P-TEFb, a transcription factor required for the transition into productive elongation. J. Biol. Chem. *270*, 12335–12338.

McCall, K., and Bender, W. (1996). Probes of chromatin accessibility in the Drosophila bithorax complex respond differently to Polycomb-mediated repression. EMBO J. *15*, 569–580.

Mito, Y., Henikoff, J.G., and Henikoff, S. (2007). Histone replacement marks the boundaries of cis-regulatory domains. Science *315*, 1408–1411.

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

Mohd-Sarip, A., Venturini, F., Chalkley, G.E., and Verrijzer, C.P. (2002). Pleiohomeotic can link polycomb to DNA and mediate transcriptional repression. Mol. Cell. Biol. *22*, 7473–7483.

Mohd-Sarip, A., van der Knaap, J.A., Wyman, C., Kanaar, R., Schedl, P., and Verrijzer, C.P. (2006). Architecture of a Polycomb Nucleoprotein Complex. Mol. Cell *24*, 91–100.

Mohrmann, L., and Verrijzer, C.P. (2005). Composition and functional specificity of SWI2/SNF2 class chromatin remodeling complexes. Biochim. Biophys. Acta *1681*, 59–73.

Muller, J., and Kassis, J.A. (2006). Polycomb response elements and targeting of Polycomb group proteins in Drosophila. Curr. Opin. Genet. Dev. *16*, 476–484.

Müller, J., Hart, C.M., Francis, N.J., Vargas, M.L., Sengupta, A., Wild, B., Miller, E.L., O'Connor, M.B., Kingston, R.E., and Simon, J.A. (2002). Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell *111*, 197–208.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344–1349.

De Napoles, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., *et al.* (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev. Cell *7*, 663–676.

Nechaev, S., and Adelman, K. (2011). Pol II waiting in the starting gates: Regulating the transition from transcription initiation into productive elongation. Biochim. Biophys. Acta *1809*, 34–45.

Neil, H., Malabat, C., d'Aubenton-Carafa, Y., Xu, Z., Steinmetz, L.M., and Jacquier, A. (2009). Widespread bidirectional promoters are the major source of cryptic transcripts in yeast. Nature *457*, 1038–1042.

Nekrasov, M., Klymenko, T., Fraterman, S., Papp, B., Oktaba, K., Köcher, T., Cohen, A., Stunnenberg, H.G., Wilm, M., and Müller, J. (2007). Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J. *26*, 4078–4088.

O'Sullivan, J.M., Tan-Wong, S.M., Morillon, A., Lee, B., Coles, J., Mellor, J., and Proudfoot, N.J. (2004). Gene loops juxtapose promoters and terminators in yeast. Nat. Genet. *36*, 1014–1018.

Ogawa, H., Ishiguro, K.-I., Gaubatz, S., Livingston, D.M., and Nakatani, Y. (2002). A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. Science *296*, 1132–1136.

Oktaba, K., Gutiérrez, L., Gagneur, J., Girardot, C., Sengupta, A.K., Furlong, E.E.M., and Müller, J. (2008). Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. Dev. Cell *15*, 877–889.

Ostuni, R., Piccolo, V., Barozzi, I., Polletti, S., Termanini, A., Bonifacio, S., Curina, A., Prosperini, E., Ghisletti, S., and Natoli, G. (2013). Latent enhancers activated by stimulation in differentiated cells. Cell *152*, 157–171.

Pandey, R.R., Mondal, T., Mohammad, F., Enroth, S., Redrup, L., Komorowski, J., Nagano, T., Mancini-Dinardo, D., and Kanduri, C. (2008). Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. Mol. Cell *32*, 232–246.

Pengelly, A.R., Copur, Ö., Jäckle, H., Herzig, A., and Müller, J. (2013). A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. Science *339*, 698–699.

Peters, A.H.F.M., Kubicek, S., Mechtler, K., O'Sullivan, R.J., Derijck, A.A.H.A., Perez-Burgos, L., Kohlmaier, A., Opravil, S., Tachibana, M., Shinkai, Y., *et al.* (2003). Partitioning and Plasticity of Repressive Histone Methylation States in Mammalian Chromatin. Mol. Cell *12*, 1577–1589.

Peterson, C.L. (2002). Chromatin remodeling enzymes: taming the machines. Third in review series on chromatin dynamics. EMBO Rep. *3*, 319–322.

Petruk, S., Sedkov, Y., Riley, K.M., Hodgson, J., Schweisguth, F., Hirose, S., Jaynes, J.B., Brock, H.W., and Mazo, A. (2006). Transcription of bxd noncoding RNAs promoted by trithorax represses Ubx in cis by transcriptional interference. Cell *127*, 1209–1221.

Preker, P., Almvig, K., Christensen, M.S., Valen, E., Mapendano, C.K., Sandelin, A., and Jensen, T.H. (2011). PROMoter uPstream Transcripts share characteristics with mRNAs and are produced upstream of all three major types of mammalian promoters. Nucleic Acids Res. *39*, 7179–7193.

Renner, D.B., Yamaguchi, Y., Wada, T., Handa, H., and Price, D.H. (2001). A highly purified RNA polymerase II elongation control system. J. Biol. Chem. *276*, 42601–42609.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., *et al.* (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Robin, P., Fritsch, L., Philipot, O., Svinarchuk, F., and Ait-Si-Ali, S. (2007). Post-translational modifications of histones H3 and H4 associated with the histone methyltransferases Suv39h1 and G9a. Genome Biol. *8*, R270.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett. *579*, 909–915.

Sánchez, C., Sánchez, I., Demmers, J.A.A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2007). Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. Mol. Cell. Proteomics *6*, 820–834.

Sanchez-Pulido, L., Devos, D., Sung, Z.R., and Calonje, M. (2008). RAWUL: a new ubiquitin-like domain in PRC1 ring finger proteins that unveils putative plant and worm PRC1 orthologs. BMC Genomics *9*, 308.

De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.-L., and Natoli, G. (2010). A large fraction of extragenic RNA pol II transcription sites overlap enhancers. PLoS Biol. *8*, e1000384.

Sarge, K.D., and Park-Sarge, O.-K. (2005). Gene bookmarking: keeping the pages open. Trends Biochem. Sci. *30*, 605–610.

Saurin, A.J., Shao, Z., Erdjument-Bromage, H., Tempst, P., and Kingston, R.E. (2001). A Drosophila Polycomb group complex includes Zeste and dTAFII proteins. Nature *412*, 655–660.

Scheuermann, J.C., de Ayala Alonso, A.G., Oktaba, K., Ly-Hartig, N., McGinty, R.K., Fraterman, S., Wilm, M., Muir, T.W., and Müller, J. (2010). Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature *465*, 243–247.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. Cell *128*, 735–745.

Schwartz, Y.B., and Pirrotta, V. (2007). Polycomb silencing mechanisms and the management of genomic programmes. Nat. Rev. Genet. *8*, 9–22.

Schwartz, Y.B., and Pirrotta, V. (2008). Polycomb complexes and epigenetic states. Curr. Opin. Cell Biol. *20*, 266–273.

Schwartz, Y.B., Kahn, T.G., Nix, D.A., Li, X.-Y., Bourgon, R., Biggin, M., and Pirrotta, V. (2006). Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat. Genet. *38*, 700–705.

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

Seenundun, S., Rampalli, S., Liu, Q.-C., Aziz, A., Palii, C., Hong, S., Blais, A., Brand, M., Ge, K., and Dilworth, F.J. (2010). UTX mediates demethylation of H3K27me3 at muscle-specific genes during myogenesis. EMBO J. *29*, 1401–1411.

Shao, Z., Raible, F., Mollaaghababa, R., Guyon, J.R., Wu, C.T., Bender, W., and Kingston, R.E. (1999). Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell *98*, 37–46.

Simcox, A., Mitra, S., Truesdell, S., Paul, L., Chen, T., Butchar, J.P., and Justiniano, S. (2008). Efficient genetic method for establishing Drosophila cell lines unlocks the potential to create lines of specific genotypes. PLoS Genet. *4*, e1000142.

Simon, J., Chiang, A., Bender, W., Shimell, M.J., and O'Connor, M. (1993). Elements of the Drosophila Bithorax Complex That Mediate Repression by Polycomb Group Products. Dev. Biol. *158*, 131–144.

Singh, B.N., and Hampsey, M. (2007). A transcription-independent role for TFIIB in gene looping. Mol. Cell *27*, 806–816.

Singh, J., and Padgett, R.A. (2009). Rates of in situ transcription and splicing in large human genes. Nat. Struct. Mol. Biol. *16*, 1128–1133.

Smith, C.L., Horowitz-Scherer, R., Flanagan, J.F., Woodcock, C.L., and Peterson, C.L. (2003). Structural analysis of the yeast SWI/SNF chromatin remodeling complex. Nat. Struct. Biol. *10*, 141–145.

Smith, S.T., Petruk, S., Sedkov, Y., Cho, E., Tillib, S., Canaani, E., and Mazo, A. (2004). Modulation of heat shock gene expression by the TAC1 chromatin-modifying complex. Nat. Cell Biol. *6*, 162–167.

Steinmetz, E.J., Conrad, N.K., Brow, D.A., and Corden, J.L. (2001). RNA-binding protein Nrd1 directs poly(A)-independent 3'-end formation of RNA polymerase II transcripts. Nature *413*, 327–331.

Steinmetz, E.J., Warren, C.L., Kuehner, J.N., Panbehi, B., Ansari, A.Z., and Brow, D.A. (2006). Genome-wide distribution of yeast RNA polymerase II and its control by Sen1 helicase. Mol. Cell *24*, 735–746.

Stolc, V., Gauhar, Z., Mason, C., Halasz, G., van Batenburg, M.F., Rifkin, S.A., Hua, S., Herreman, T., Tongprasit, W., Barbano, P.E., *et al.* (2004). A gene expression map for the euchromatic genome of Drosophila melanogaster. Science *306*, 655–660.

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

Tanaka, Y., Katagiri, Z.-I., Kawahashi, K., Kioussis, D., and Kitajima, S. (2007). Trithorax-group protein ASH1 methylates histone H3 lysine 36. Gene *397*, 161–168.

Tan-Wong, S.M., Zaugg, J.B., Camblong, J., Xu, Z., Zhang, D.W., Mischo, H.E., Ansari, A.Z., Luscombe, N.M., Steinmetz, L.M., and Proudfoot, N.J. (2012). Gene loops enhance transcriptional directionality. Science *338*, 671–675.

Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., Bezstarosti, K., Taylor, S., Ura, H., Koide, H., *et al.* (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell *148*, 664–678.

Tie, F., Prasad-Sinha, J., Birve, A., Rasmuson-Lestander, A., and Harte, P.J. (2003). A 1-megadalton ESC/E(Z) complex from Drosophila that contains Polycomblike and RPD3. Mol. Cell. Biol. *23*, 3352–3362.

Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131–3141.

Tie, F., Banerjee, R., Conrad, P.A., Scacheri, P.C., and Harte, P.J. (2012). Histone demethylase UTX and chromatin remodeler BRM bind directly to CBP and modulate acetylation of histone H3 lysine 27. Mol. Cell. Biol. *32*, 2323–2334.

Trimarchi, J.M., Fairchild, B., Wen, J., and Lees, J.A. (2001). The E2F6 transcription factor is a component of the mammalian Bmi1-containing polycomb complex. Proc. Natl. Acad. Sci. U. S. A. *98*, 1519–1524.

Vignali, M., Hassan, A.H., Neely, K.E., and Workman, J.L. (2000). ATP-Dependent Chromatin-Remodeling Complexes. Mol. Cell. Biol. *20*, 1899–1910.

Vincenz, C., and Kerppola, T.K. (2008). Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. Proc. Natl. Acad. Sci. U. S. A. *105*, 16572–16577.

Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R.S., and Zhang, Y. (2004a). Role of histone H2A ubiquitination in Polycomb silencing. Nature *431*, 873–878.

Wang, L., Brown, J.L., Cao, R., Zhang, Y., Kassis, J.A., and Jones, R.S. (2004b). Hierarchical recruitment of polycomb group silencing complexes. Mol. Cell *14*, 637–646.

Wu, X., Johansen, J.V., and Helin, K. (2013). Fbxl10/Kdm2b Recruits Polycomb Repressive Complex 1 to CpG Islands and Regulates H2A Ubiquitylation. Mol. Cell *null*.

Wyers, F., Rougemaille, M., Badis, G., Rousselle, J.-C., Dufour, M.-E., Boulay, J., Régnault, B., Devaux, F., Namane, A., Séraphin, B., *et al.* (2005). Cryptic pol II transcripts are degraded by a nuclear quality control pathway involving a new poly(A) polymerase. Cell *121*, 725–737.

Xu, Z., Wei, W., Gagneur, J., Perocchi, F., Clauder-Münster, S., Camblong, J., Guffanti, E., Stutz, F., Huber, W., and Steinmetz, L.M. (2009). Bidirectional promoters generate pervasive transcription in yeast. Nature *457*, 1033–1037.

Yuan, W., Xu, M., Huang, C., Liu, N., Chen, S., and Zhu, B. (2011). H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J. Biol. Chem. *286*, 7983–7989.

Zhao, J., Sun, B.K., Erwin, J.A., Song, J.-J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. Science *322*, 750–756.

Zhou, W., Zhu, P., Wang, J., Pascual, G., Ohgi, K.A., Lozach, J., Glass, C.K., and Rosenfeld, M.G. (2008). Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol. Cell *29*, 69–80.

Zink, D., and Paro, R. (1995). Drosophila Polycomb-group regulated chromatin inhibits the accessibility of a trans-activator to its target DNA. EMBO J. *14*, 5660–5671.