THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Biologie Santé

École doctorale : CBS2

Unité de recherche : Institut de Génétique Humaine_ UMR9002 CNRS-UM

Caractérisation de la diversité des sites de fixation des protéines du groupe Polycomb chez la Drosophile

Présentée par Marianne ENTREVAN
Le 29 Septembre 2017

Sous la direction de Giacomo CAVALLI
et Bernd SCHUTTENGRUBER

Devant le jury

composé de

[Robert FEIL, Directeur de recherche, IGMM]	Président du jury
[Frédérique PERONNET, Directeur de recherche, IBPS]	Rapporteur
[Slimane AIT-SI-ALI, Directeur de recherche, BioSPC]	Rapporteur
[Jacques VAN HELDEN, Directeur de recherche, Laboratoire TAGC]	Examinateur

Caractérisation de la diversité des sites de fixation des protéines du groupe Polycomb chez la Drosophile

Characterization of the diversity of the Polycomb group complexes Binding sites in Drosophila

Marianne Entrevan

2017

Remerciements

Je tiens à remercier tout particulièrement Giacomo Cavalli pour avoir tout d'abord accepté de me recevoir en stage de master 2, puis pour m'avoir soutenue pendant les quatre années qu'a durée ma thèse.

Sous la responsabilité de Bernd Schuettengruber, j'ai appris la rigueur qu'imposait le travail en science. En quatre années de thèse, j'ai énormément appris à ces côtés et je me rends compte aujourd'hui que je n'ai appris que très peu de choses en comparaison de toutes les connaissances et les richesses qu'apporte l'univers de la Recherche. Je tenais donc tout particulièrement à le remercier pour son soutien et pour son aide au quotidien. Je le remercie notamment pour son aide précieuse lors de la rédaction de ma revue et les corrections qu'il a apportées à ce manuscrit.

Je tenais également à remercier le reste des membres de l'équipe dont Aubin Thomas pour son aide en Bio-informatique durant la première partie de ma thèse. Ensuite, je tiens à remercier tout particulièrement Vincent Loubière pour son aide dans les analyses de Bio-informatique ainsi que pour son soutien ; mais également Yuki Ogiyama, pour son aide (toujours en bioinfo) ponctuelle mais extrêmement précieuse.

Merci à tous.

Ce manuscrit est dédié à ma mère, pour la remercier de son soutien sans faille tout au long de ma thèse.

Résumé

Les protéines du groupe Polycomb (PcG) ont initialement été identifiées chez la drosophile comme répresseurs transcriptionnels des gènes homéotiques. Aujourd'hui, nous savons que ces protéines jouent un rôle bien plus large puisqu'elles régulent des gènes dont les produits sont impliqués dans de nombreux processus biologiques (régulation des gènes HOX, maintien de la plasticité des cellules souches, la différenciation cellulaire, l'inactivation du chromosome X, la régulation des gènes soumis à empreintes). Leur dérégulation est source de nombreux cancers chez l'homme. Hautement conservées, elles forment deux principaux complexes : PRC 1 et 2 (*Polycomb repressive complex 1 and 2*), dont l'activité est respectivement reflétée par la mono-ubiquitinylation de la lysine 118 l'histone H2A (H2AK118Ub) et la tri-méthylation de la lysine 27 de l'histone H3 (H3K27me3). Chez la Drosophile, les sites de fixation de ces complexes sont appelés PRE (*Polycomb Responsive Elements*) où ils sont recrutés via des facteurs de transcription (FT).

La complexité du recrutement des complexes du PcG, chez la Drosophile comme chez les mammifères, est visible à différents niveaux : au niveau de la séquence même de leurs sites de fixations, au niveau des facteurs de transcription qui les recrutent, au niveau de l'interface entre les deux complexes PRC1 et PRC2 et enfin au niveau global, part le présence de ces complexes au niveau de sites transcriptionnellement actifs. L'ensemble de ces résultats démontre clairement la nature hétérogène des PRE. Ces derniers diffèrent non seulement par leur séquence, mais également par les FT qui les recrutent et enfin par la manière dont les complexes PcG sont recrutés (PRC2 recrute PRC1 ou le contraire).

Mon projet de thèse s'est donc dessiné autour d'une hypothèse : il existe différentes classes de PRE chez la Drosophile. Mon travail a donc consisté à définir ces différentes classes et à les caractériser pour en déduire des rôles spécifiques à l'échelle génomique. En effet, l'implication des complexes du PcG dans l'apparition de cancer chez l'Homme requière que l'on comprenne comment ces protéines sont recrutées à la chromatine.

Mes travaux de thèse ont permis d'identifier six classes différentes de sites de fixation aux protéines du PcG. Nous avons retrouvé une classe correspondant aux sites de fixations canoniques fixés par les protéines du PcG et présents au sein de larges domaines répressifs marqués par H3K27me3. Une seconde classe correspond à des éléments de régulation marqués par un état de pause transcriptionnelle. De façon surprenante, nous avons démontré qu'une grande partie des sites de fixation des complexes du PcG était localisée au niveau de régions transcriptionnellement actives. Ces classes de PRE diffèrent en particulier en éléments génomiques qui les composent. Deux classes correspondent à des enhancers développementaux. Une classe correspond à des promoteurs actifs pouvant réguler des gènes de ménage. Enfin, une dernière classe correspond à des bordures de TAD. Les sites actifs et réprimés fixés par le PcG fixent également des combinaisons différentes de FT. Des analyses *in vivo* associées à un transcriptome réalisé à partir de cellules mutantes pour une protéine du

PcG révèlent que les complexes du PcG jouent également un rôle de répresseur transcriptionnel au niveau des sites actifs.

L'ensemble de ces résultats suggère une hétérogénéité inattendue des sites de fixation des complexes du PcG et permettra de mieux comprendre les caractéristiques liées à ces protéines dont la dérégulation mène à l'apparition de cancers chez l'Homme marqués par leurs agressivités.

Abstract

Polycomb group (PcG) complexes were initially discovered in Drosophila as transcriptionnal repressors of homeotic genes. To date, we know that they are involves in a large pleithora of biological processes including the maintenance of stem cells plasticity, differentiation, X chromosome inactivation and imprinting. PcG complexes are highly conserved from Drosophila to Humans and can be divided into two main complexes: PRC1 and PRC2 (Polycomb repressive complex 1 and 2). Both complexes have a histone modifying activity: PRC1 catalyses the mono-ubiquitination of the lysine 118 on histone H2A (H2AK118Ub) and PRC2 catalyses the tri-methylation of the lysine 27 on histone H3 (H3K27me3).

In Drosophila, these complexes are recruited to cis regulatory elements named Polycomb Responsive Elements (PREs) that drive the epigenetic inheritance of silent chromatin states throughout development. Importantly, PcG complexes do not contain DNA-binding activity but are recruited to PREs via their interaction with Transcription Factors (TF) recognizing DNA motifs clustered at PREs. However the mechanism how PREs target PcG complexes is still not well understood due to the complexity of PcG recruitment, which is reflected at different levels: The DNA signature between PREs can differ significantly and several TF are implicated in PcG recruitment, but none of them is sufficient to recruit PcG complexes to PREs. Moreover PcG complexes can cooperate in different ways to stabilize each other's binding. Finally, another layer of complexity is found at a more global level since PcG complexes do not only bind repressed sites, but they are also found at active regions.

Therefore, our working hypothesis is that different classes of PREs exist in Drosophila. My PhD work was thus to define these different classes of PREs on a genome-wide scale and to functionally characterize them in order to get a complete molecular description of PRE function. Understanding how PcG complexes are recruited is of high importance, since deregulation of both, PcG complexes and their recruiting factors can led to cancer and diseases. My work led to the identification of six different classes of PREs that are characterized by different chromatin and genomic features. Interestingly the majority of PREs are associated with active genes that can be divided into housekeeping regulatory regions and developmental enhancers. In addition another class comprises bona fide chromatin domain boundaries. On the other hand PREs associated with repressed chromatin states shows features of previously described PREs and associate with repressed genes and PcG-associated histone marks. Finally another class comprises PREs that are likely in a poised chromatin state. We further demonstrated that PREs located at repressed and active regions differ in their combination of TF. In vivo analyses along with a transcriptomic analysis performed in cell lines mutated for a member of PcG complexes revealed that PcG complexes play a repressive role at both, active and repressed PREs.

Taken together, our result suggest an unexpected heterogeneity of PREs and contributes to the better understanding of their characteristics and function.

INTRODUCTION 19
Chapitre I : Chromatine et régulation de l'expression génique 21
I. De l'ADN à la chromatine 21
A. La molécule d'ADN 21
B. Le nucléosome, unité de base de la chromatine 22
C. Les structures d'ordre supérieur 24
D. Rôle de la chromatine dans l'expression génique 26
II. Régulation de la transcription génique 27
A. Les éléments régulateurs de l'ADN 27
B. Complexes de remodelage des nucléosomes 29
C. Les variants d'histones 31
D. Les modifications post-traductionnelles des histones 33
E. Communication entre les marques d'histones 38
F. Les domaines chromatiniens 40
G. Les domaines chromatiniens à l'échelle nucléaire 42
Chapitre II : Présentation des complexes du PcG et du TrxG 43
I. Les gènes homéotiques 43
II. Les protéines du TrxG (*Trithorax Group*) 46
III. Les complexes du PcG (*Polycomb Group*) 48
A. Le complexe PRC1 (*Polycomb Repressive Complex 1*) 49
B. Le complexe dRAF (*dRING Associated Factors*) 54
C. Le complexe Pho-RC (*Pho-Repressive Complex*) 54
D. Le complexe PR-DUB (*Polycomb repressive deubiquitinase*) 55
E. Le complexe PRC2 (*Polycomb Repressive Complex 2*) 55
IV. Mécanismes de régulation de la transcription liés à l'activité des protéines du PcG 60
A. Mécanismes impliquant les protéines du PcG dans la répression transcriptionnelle 60
B. Mécanismes impliquant les protéines du PcG dans l'activation transcriptionnelle (ENTREVAN *et al.* 2016) 65
Chapitre III : Fonctions biologiques du PcG 67

I. Rôle différentiel et dynamique des complexes du PcG lors de la différentiation 67
II. La dérégulation des gènes cibles du PcG mène à l'apparition de cancers ... 69
Chapitre IV : Complexité du recrutement de Polycomb à l'ADN 71
I. Les PRE/TRE et leurs caractéristiques 71
II. Complexité du recrutement des complexes du PcG à différents niveaux ... 73
A. Au niveau de la séquence des PRE 73
B. Les Facteurs de Transcription (FT) associés au recrutement des protéines du PcG chez la Drosophile 74
C. A l'interface des complexes eux-mêmes 79
D. Existence de sites de fixations des protéines du PcG au niveau de sites actifs 81
III. Le recrutement des protéines du PcG chez les mammifères est aussi complexe que chez la Drosophile (ENTREVAN *et al.* 2016) 82
CONCLUSION ET INTRODUCTION DU PROJET DE THESE 85
RESULTATS 87
I. CLUSTERING ANALYSIS REVEALED THE EXISTENCE OF SIX CLASSES OF PREs , CHARACTERIZED BY DISTINCS GENOMIC AND CHROMATIN FEATURES 93
A. Data set presentation 93
B. Six different classes of PREs are defined by distinct chromatin signatures 95
C. PREs associate with active, poised and repressed DNA elements 103
D. Distinct PRE classes are specified by specific combination of Insulator proteins and 3D interactions 109
E. PRE classes are defined by a distinct combination of sequence specific DNA binding proteins 115
II. FUNCTIONAL VALIDATION OF THE SIX CLASSES OF PREs 121
III. FUNCTIONAL ROLE OF POLYCOMB AT ACTIVE SITES 125
A. Creation of *Pc* mutant S2 cell lines using the CRISPR/Cas9 system 125
B. Transcriptome analyses of *Pc* mutant cell lines 127
C. Chromatin changes and PcG binding in Pc mutant cell lines 131
D. PRC1-dependent binding of PRC2 133
DISCUSSION 135

A. PREs are heterogeneous cis regulatory elements, present at repressed and active genomic regions (Figure 39)..... 142
B. Redefinition of PREs 143
C. Co-occupancy of PcG and TrxG proteins at PREs/TREs 144
D. Enhancer function of PREs..... 145
E. Distribution of histone marks around PRE classes 145
F. PRC1 and PRC2 only sites and Gene ontologies (GO) associated with PRE classes..... 146
G. Role of 3D chromatin interactions and insulator proteins in PRE function 147
H. PRE classes are specified by distinct Transcription factor (TF) signatures 148
I. Different functional roles of canonical and non-canonical PREs?..... 150
J. Intra-variability of canonical PREs 151
K. What is the role of PRC1 at active regions? 152
MATERIEL 154
REFERENCES..... 169
ANNEXES..... 193

Les publications scientifiques auxquelles j'ai participées sont données dans les Annexes. Il s'agit d'une revue scientifique et d'un article.

Figure 1 : Le nucléosome est l'unité structurale de la chromatine. 22
Figure 2: Histones et Formation du nucléosome 23
Figure 3 : Compaction de l'ADN en chromosome mitotique 25
Figure 4 : Eléments régulateurs des gènes 27
Figure 5 : Régulation de l'expression génique par les éléments de type enhancer 28
Figure 6 : Les complexes de remodelage de la chromatine 29
Figure 7 : Modes d'actions des complexes de remodelage de la chromatine . 30
Figure 8 : Modifications des histones 33
Figure 9 : Profils typiques des modifications d'histones le long d'un gène..... 36
Figure 10 : Différentes « couleurs » de la chromatine définies par des combinaisons distinctes en protéines et marques d'histones. 41
Figure 11 : Etablissement et maintien du profil d'expression des gènes homéotiques (HOX) de l'embryon à la mouche adulte. 44
Figure 12 : Complexes associés aux protéines du groupe Trithorax. 46
Figure 13 : Phénotypes associés aux mutations des protéines du PcG chez la Drosophile 48
Figure 14 : Présentation des complexes PRC1 chez la drosophile et les mammifères et d'autres complexes relatifs à l'activité PRC1 50
Figure 15 : Domaines fonctionnels de chaque protéine du PRC1 et de la protéine Scm 52
Figure 16 : Le complexe PRC2 et ses protéines accessoires. 56
Figure 17 : Les complexes du PcG forment de large domaines répressifs marqués par H3K27me3..... 61
Figure 18 : Régulation de la transcription par les protéines du PcG à différentes échelles. 62
Figure 19 : Dynamique de l'organisation chromatinienne au cours du développement..... 65
Figure 20 : Rôle des protéines du PcG lors de la différentiation cellulaire 68
Figure 21 : Test fonctionnel de l'activité PRE. 72
Figure 22 : Complexité du recrutement des complexes du PcG à l'interface même de ces protéines..... 80
Figure 23 : Complexité du recrutement du PcG à la chromatine à différents niveaux 84
Figure 24 : Quality of data used to analyze the diversity of PREs in Drosophila S2 cells. 95
Figure 25 : Clustering analysis of PREs revealed six different classes based on histones PTM..... 97
Figure 26 : Mean enrichment of histones PTM associated with H3K27 over a 10kb window. 98
Figure 27 : Defining the classes of PREs 100

Figure 28 : PREs are localized at active, poised or repressed DNA elements 103
Figure 29 : Mean enrichment of various proteins associated to TrxG and H3K4me2 over a 10kb window. 104
Figure 30 : The 6 classes of PREs differ by distincts signatures in insulator proteins 108
Figure 31 : The 6 classes of PREs are specified by different chromatin architectural characteristics 110
Figure 32 : Characterization of PcG recruiters profiles at the six classes of PREs 115
Figure 33 : The 6 classes of PREs are specified by distinct combinations in PcG recruiters 117
Figure 34 : Functional characterization of canonical and non-canonical PREs in flies 121
Figure 35 : Mutation of *Pc* gene in S2 cells using the CRISPR/Cas9 system 125
Figure 36 : Transcriptomic results of RNA-seq performed on *Pc* mutant cell lines 127
Figure 37 : Chromatin changes and PcG binding in *Pc* mutant cell lines 131
Figure 38 : Pc mutant cell lines show a decrease in H3K27me3 and an increase of H3K27ac. 132
Figure 39 : Different classes of PREs exist in the Drosophila genome 141
Figure 40 : Modified SD10 vector to test PRE activity. 165

INTRODUCTION

Chapitre I : Chromatine et régulation de l'expression génique

I. De l'ADN à la chromatine

A. La molécule d'ADN

Dans le monde du vivant, l'acide désoxyribonucléique (ADN) est la molécule qui détient l'information nécessaire pour le développement, la survie et la reproduction d'un organisme. Cette molécule est présente dans chacune des cellules d'un organisme et l'information qu'elle contient peut être convertie en message nécessaire à la production d'ARN et de protéines. Par ailleurs, pour le développement d'un organisme, les cellules doivent se diviser et l'ADN est alors capable d'être recopié afin d'être transmis. Chaque espèce possède un génome constitué par une séquence d'ADN qui lui est propre, porté par un certain nombre de chromosomes, et une composition génique particulière qui constitue son génome.

Il aura fallu près d'un siècle pour isoler cette molécule, déterminer sa composition, sa structure et pouvoir affirmer que l'ADN est le support physique de l'hérédité décrite par Johann Gregor Mendel. L'ADN est formé de deux chaines nucléotidiques enroulées l'une autour de l'autre pour former une structure en double hélice. Chaque nucléotide se compose, d'une base azotée (cytosine (C), guanine (G), thymine (T) ou adénine (A)), d'un sucre, le désoxyribose, et d'un groupement phosphate liant chaque nucléotide d'une même chaine. L'association des deux brins d'ADN s'effectue par des liaisons hydrogènes entre bases complémentaires. La cytosine s'associant uniquement à la guanine et la thymine avec l'adénine.

Dans les cellules eucaryotes, la molécule d'ADN est contenue dans un compartiment particulier qui est le noyau. De dimension micrométrique, chaque noyau contient cependant environ deux mètres d'ADN chez l'homme. L'ADN est donc extrêmement compacté dans le noyau. Pour se faire l'ADN, s'associe à des protéines afin de former des structures pouvant adopter différents niveaux de compactions, le plus élevé correspond au chromosome mitotique. Au sein du noyau, l'ADN n'est donc pas « nu » mais forme une structure appelée chromatine (du grec *chromos*, couleur, en référence aux travaux de W. Flemming qui remarqua des structures dans le noyau cellulaire absorbant fortement l'aniline, un colorant basique). L'organisation de la chromatine dans le noyau peut être décrite à différentes échelles qui sont abordées ci-après.

B. Le nucléosome, unité de base de la chromatine

La digestion partielle de l'ADN par des endonucléases génère des fragments d'ADN ayant une taille multiple de 200 bp (**Figure** 1A). De plus, l'observation d'ADN non purifié par microscopie électronique montre un chapelet de particules reliées les unes aux autres par des filaments d'ADN (**Figure** 1B). Ceci a permis de mettre en évidence l'existence d'une unité structurale de base au sein du noyau. Cette particule élémentaire a été appelée nucléosome.

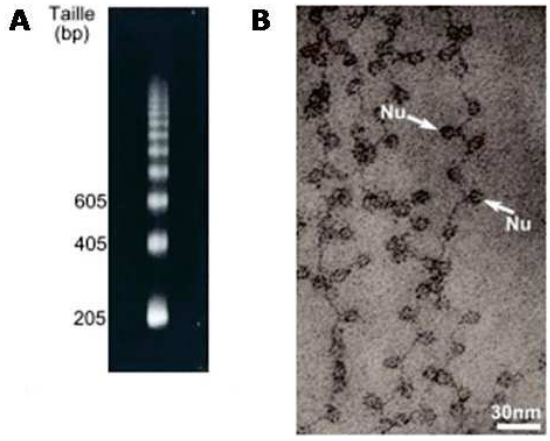


Figure 1 : Le nucléosome est l'unité structurale de la chromatine.

A- Digestion ménagée de la chromatine à la nucléase micrococcale (LEWIN. 2008 ; Genes IX book).

B- Chromatine isolée observée par microscopie électronique à transmission (OLINS and OLINS 2003)

Le nucléosome est formé de huit histones de cœur ([H2A, H2B, H3 et H4]x2) autour desquelles s'enroule 146 bp d'ADN (**Figure** 2B). Ces histones de cœur forment une famille de petites protéines basiques très conservées. Par ailleurs, leurs charges globales positives leur permettent une interaction forte avec les groupements phosphate de la molécule d'ADN chargés négativement. Les histones sont formées d'une queue N-terminale non structurée et d'un domaine structuré appelé *histone-fold* (**Figure** 2A). Ce dernier comprend trois hélices α qui permettent l'association de dimères d'histones selon un motif dit « en poignée de main » (**Figure** 2B)(ARENTS *et al.* 1991). La partie N-terminale des histones ne participent pas directement à la structure du nucléosome. Cependant, elles peuvent être modifiées après leur synthèse (on parle alors de modifications post-traductionnelles comme la méthylation ou l'acétylation de certains acides aminés) et ainsi impacter sur la compaction de niveau supérieur de la chromatine. Ces modifications et leurs impacts sur la transcription seront détaillés ultérieurement.

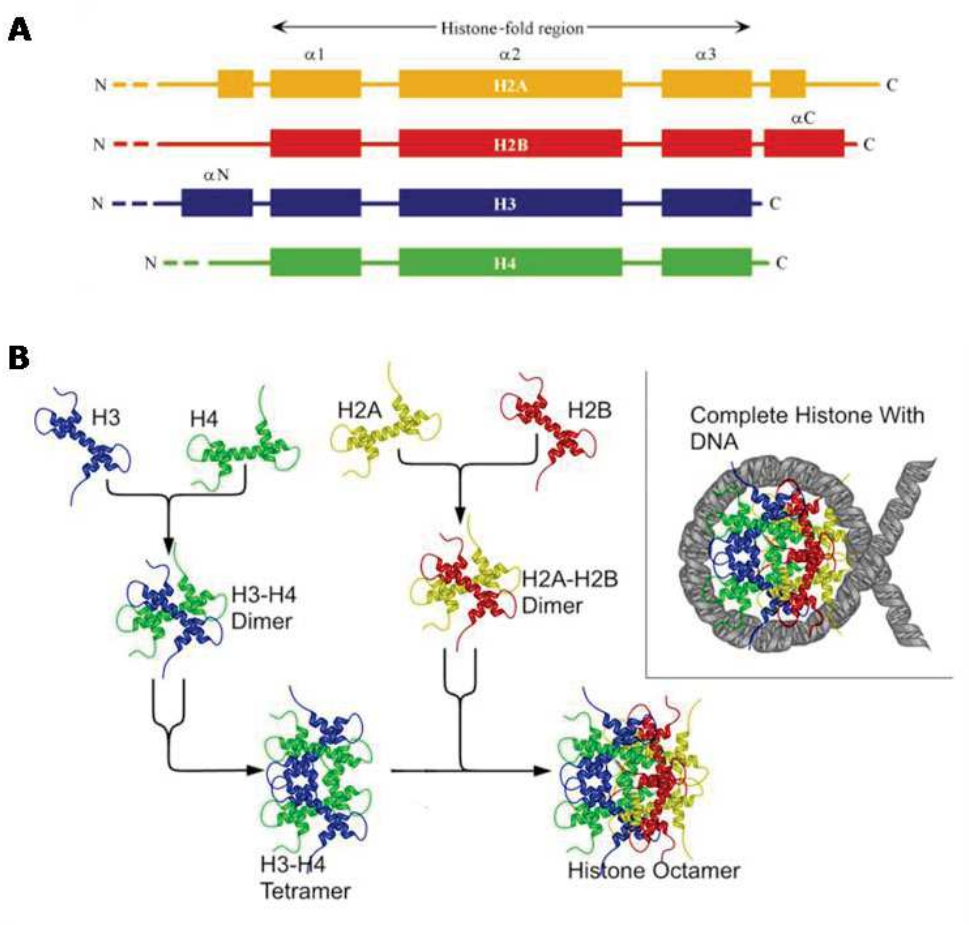


Figure 2: Histones et Formation du nucléosome

A- Alignement des 4 histones de cœur. De part et autre du motif *histone-fold* central se trouvent les queues N-terminales et les domaines C-terminaux (Adapté de (DUTNALL and RAMAKRISHNAN 1997)).

B- Les histones H3 et H4 forment un tétramère sur lequel s'ajoute 2 dimères H2A-H2B réalisant ainsi l'octamère d'histones autour duquel 1,75 tour d'ADN est enroulé pour former le nucléosome. Figure tirée de : Richard Wheeler (Zephyris) (English Wikipedia) [GFDL (http://www.gnu.org/copyleft/fdl.html)].

Pour conclure, la répétition de nucléosomes régulièrement espacés conduit à la formation de la fibre de 11 nm ou fibre en « collier de perle » (**Figure** 1B), où les nucléosomes sont séparés par l'ADN dit « de liaison ». Cette organisation correspond à l'état le plus décompacté de la chromatine (OUDET *et al.* 1975; LUGER *et al.* 1997).

C. Les structures d'ordre supérieur

Un état plus compact de la chromatine peut également être observé par microscopie électronique. Il s'agit de la fibre de 30 nm de diamètre dont la stabilité semble dépendre de l'association de l'histone H1 aux nucléosomes (FINCH and KLUG 1976; ROBINSON *et al.* 2006). Cette histone se lie aux nucléosomes au niveau de l'entrée et de la sortie de l'ADN ce qui influence la conformation de l'ADN de liaison. L'assemblage des huit protéines d'histones de cœur, de l'ADN et de l'H1 forme un complexe appelé chromatosome.

Plusieurs modèles non exclusifs ont été proposés concernant l'agencement des chromatosomes pour former la fibre de 30 nm (**Figure** 3A) (GRIGORYEV *et al.* 2009). Le premier modèle dit « solénoïde » propose un enroulement des nucléosomes selon un axe virtuel formant une hélice simple (**Figure** 3B) (FINCH and KLUG 1976; ROBINSON *et al.* 2006). Le second modèle dit « en zigzag » suggère que les ADN de liaisons traversent l'axe d'enroulement formant une hélice double (**Figure** 3B) (WOODCOCK *et al.* 1984; DORIGO *et al.* 2004).

Les niveaux d'organisation et de compaction supérieurs pour obtenir un chromosome métaphasique sont encore mal identifiés. Cependant, il a été décrit que la fibre de 30 nm pouvait se replier et former des boucles géantes de 100 Kb d'ADN maintenues par des liaisons entre la base des boucles (SAR=*Scaffold Attachment Region*) (RAZIN 2001; LIEBICH *et al.* 2002; FIORINI *et al.* 2006) (**Figure** 3A). Le repliement de ces boucles géantes de chromatine pour former une fibre de 700 nm correspondant à un bras chromosomique reste flou. Cependant, le taux de compaction du génome atteint son maximum lors de la formation du chromosome métaphasique (**Figure** 3A) où la compaction de la molécule d'ADN est d'environ 10.000x par rapport à une molécule d'ADN linéaire (BELMONT 2002).

Pour conclure, la chromatine n'est pas une structure statique mais est en perpétuel mouvement. Elle peut donc répondre rapidement aux processus biologiques auxquels elle est soumise.

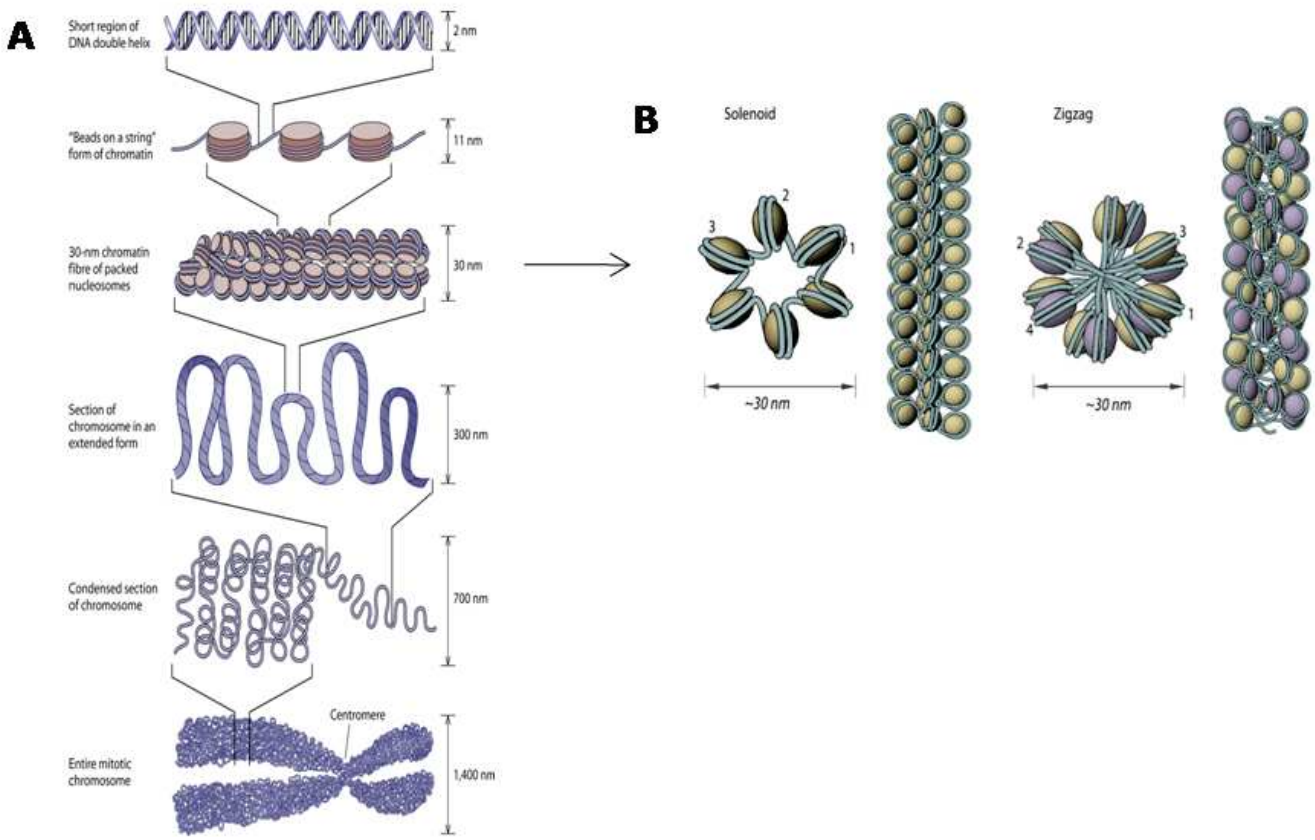


Figure 3 : Compaction de l'ADN en chromosome mitotique

A- De haut en bas : l'ADN s'enroule autour des octamères d'histones pour former la fibre de 11 nm (structure en « collier de perle ». Par l'association avec l'histone H1, la fibre adopte une structure de 30 nm de diamètre selon deux structures possibles (voir B). L'arrangement de la fibre de 30 nm en fibre de 700 nm semblerait passer par la formation de boucles géantes. Finalement, le chromosome mitotique représente la structure la plus compacte de l'ADN (illustration de (JANSEN and VERSTREPEN 2011)).

B- Deux modèles sont proposés pour la formation de la fibre de 30 nm. Celui de type « solénoïde » à gauche et celui dit en « zigzag » à droite (illustration de Creative Commons ; Used by permission from MBInfo: www.mechanobio.info; Mechanobiology Institute, National University of Singapore]).

D. Rôle de la chromatine dans l'expression génique

Dans les parties précédentes, nous avons vu le rôle du nucléosome dans l'organisation structurale de la chromatine. Cependant, la fonction de ces protéines ne se limite pas à établir l'architecture du génome mais contribue également à la régulation de l'expression génique.

En effet, l'ADN enroulé autour des protéines histones est moins accessible à la machinerie transcriptionnelle ce qui limite l'activation des gènes. Cependant, divers études montrent que la distribution des nucléosomes le long d'un gène actif n'est pas aléatoire. Les régions promotrices et terminales des gènes sont pauvres en nucléosomes par rapport aux parties codantes. La densité et la position des nucléosomes contribuent donc à permettre/inhiber l'expression des gènes.

Par ailleurs, nous verrons que les modifications post-traductionnelles des histones, tout comme la nature biochimique même des histones qui composent les nucléosomes jouent également un rôle direct sur la compaction/décompaction de la chromatine. L'ensemble des éléments régulant l'expression génique sont décrits ci-après.

II. Régulation de la transcription génique

A. Les éléments régulateurs de l'ADN

La séquence d'ADN détient un ensemble de signaux nécessaires à la régulation des gènes. Ces signaux ou éléments régulateurs sont de courtes régions génomiques qui correspondent à des sites de fixations spécifiques de protéines qui catalysent ou régulent la transcription des gènes. Selon leur position vis à vis du gène qu'ils régulent, on distingue différents types d'éléments : les éléments du promoteur (minimal ou proximal) et les éléments distaux (**Figure** 4). Dépendamment des protéines qui se fixent sur les éléments régulateurs distaux et de la fonction qu'elles exercent sur la transcription du gène qu'elles régulent, on parlera d'élément « enhancer » (effet positif sur la transcription) ou d'élément « silencer » (effet négatif sur la transcription). La distinction entre « enhancer » et « silencer » réside dans la nature même des protéines qui se fixent sur ces éléments. Enfin, les « insulateurs » permettent de restreindre l'effet d'un enhancer ou d'un silencer à un gène ou à un groupe de gènes. Ces régions régulatrices peuvent agir sur plusieurs gènes à la fois ou inversement un même gène peut être régulé par plusieurs de ces éléments régulateurs. Enfin la plupart de ces éléments ne correspondent pas à des régions codantes.

L'effet « enhancer » ou « silencer » des régions régulatrices permet de moduler l'expression d'un gène de manière spatio-temporelle. Ces régions peuvent être localisées au niveau du promoteur proximal mais aussi à des distances pouvant atteindre les 10kb chez la Drosophile (ZHOU *et al.* 2001).

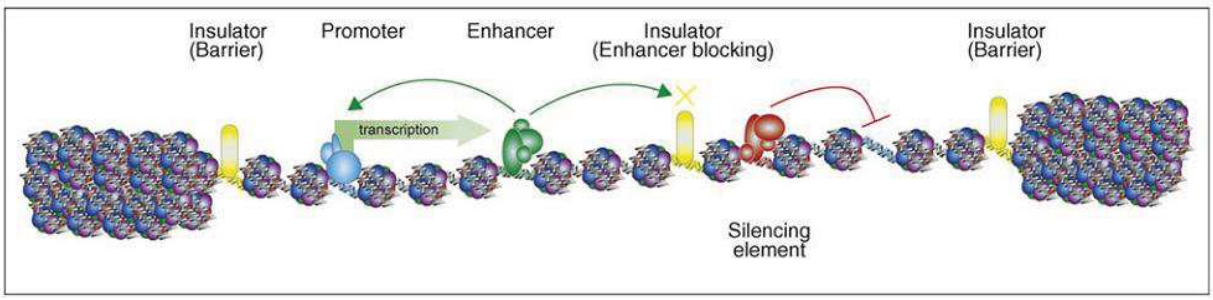


Figure 4 : Eléments régulateurs des gènes
Les éléments régulateurs des gènes, de type enhancer (vert) ou silencer (rouge), permettent la régulation de leurs gènes cibles à des distances variables. Les régions insulatrices (jaune) fixent des protéines qui permettent de limiter l'action des éléments enhancer/silencer à un certain nombre de gènes (Illustration de (HEINTZMAN and REN 2009)).

Ils existent différents modèles expliquant comment ces éléments régulateurs interagissent avec les promoteurs, en particulier quand ils sont séparés par de longues distances (**Figure** 5) (KOLOVOS *et al.* 2012). (i) Le modèle de *tracking* suggère que les protéines chargées au niveau de l'enhancer glissent le long de la chromatine jusqu'à atteindre le promoteur et stimuler son activation. (ii) Le modèle de *linking* suggère une polymérisation entre les protéines se fixant à l'enhancer et celles se fixant au promoteur permettant une liaison qui active le promoteur. (iii) Enfin, le dernier modèle dit de *looping* implique une interaction directe entre l'enhancer et le promoteur formant ainsi une boucle. Cette interaction est effectuée par le biais de protéines se liant à l'ADN (DBP = *DNA-binding protein*).

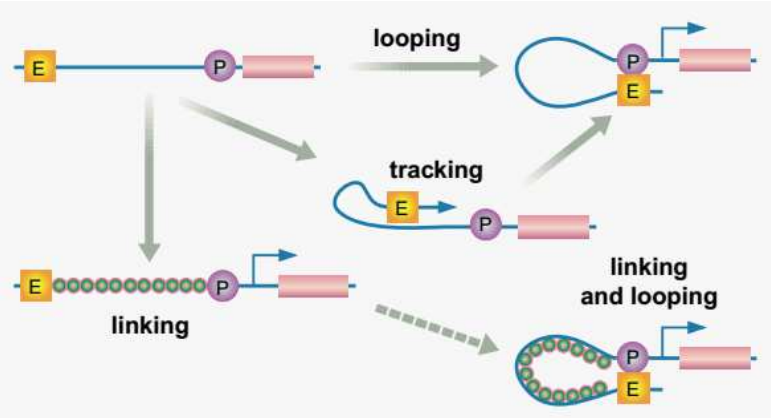


Figure 5 : Régulation de l'expression génique par les éléments de type enhancer

Modèles possibles permettant l'action d'une région régulatrice (E pour enhancer) avec le promoteur (P) d'un gène. En haut, le modèle de *looping* permet une interaction directe entre l'élément régulateur et le promoteur par l'intermédiaire de DBP permettant la formation d'une boucle. Au centre, le modèle de *tracking* propose que les protéines fixées à l'élément régulateur scannent l'ADN de manière active jusqu'à atteindre le promoteur. En bas à gauche, le modèle de *linking* propose la formation d'une chaine de protéines (en vert) qui s'étend le long de la chromatine jusqu'au promoteur pour permettre la régulation de ce dernier (en bas à droite) (Image tirée de (BULGER and GROUDINE 2002)).

Le cas du complexe Mediator (complexe de co-activation de la transcription) est sans doute celui qui illustre le mieux le modèle de *looping* (KAGEY *et al.* 2010). Cependant, les techniques de Capture de la Conformation des Chromosomes (3C et ses variantes) ont largement mis en évidence que de telles interactions sont courantes dans la régulation de la transcription (SEXTON *et al.* 2009).

B. Complexes de remodelage des nucléosomes

Comme décrit dans la partie précédente, l'ADN est extrêmement compacté dans le noyau. Cette organisation rend les régions régulatrices des gènes inaccessibles par défaut. D'une part parce que les DBP peuvent difficilement atteindre les séquences d'ADN en interaction avec les nucléosomes ; et d'autre part parce que les motifs d'ADN reconnus par ces DBP sont déformés et méconnaissables. Cette inévitable inaccessibilité de l'ADN au sein de la chromatine pourrait apparaitre comme un problème à première vue, mais il existe en réalité des enzymes qui peuvent remodeler les nucléosomes. Ces enzymes permettent donc de rendre localement accessible certaines parties de l'ADN. Ces enzymes utilisent l'énergie issue de l'hydrolyse de l'ATP (Adénosine-Tri-Phosphate) pour remodeler les nucléosomes.

Il existe quatre grandes familles de complexes de remodelage : SWI/SNF, ISWI, CHD et INO80 (**Figure** 6). Chacun des membres possède un domaine ATPase mais ils diffèrent par l'existence d'autres domaines protéiques leur permettant d'exécuter divers fonctions.

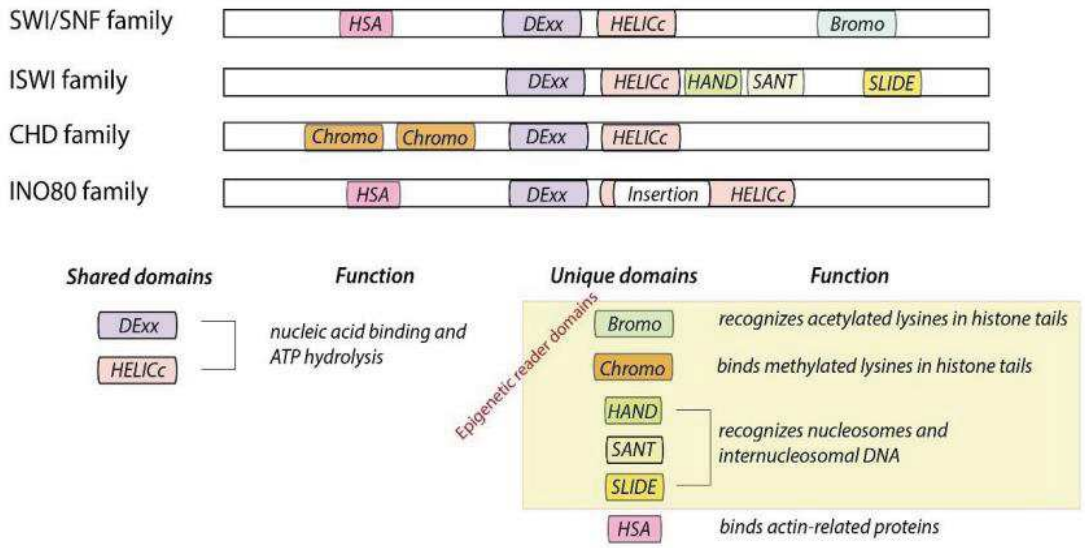


Figure 6 : Les complexes de remodelage de la chromatine
Les membres de chacune des familles possèdent un domaine ATPase. Les familles diffèrent par la présence d'autres domaines protéiques qui leur permettent d'effectuer des fonctions plus spécifiques (voir le tableau « function ») (Figure tirée de (LANGST and MANELYTE 2015)).

De façon simplifiée, les membres de la famille SWI/SNF (*SWItch/Sucrose Non-Fermentable*) sont définis par la présence d'un bromodomaine pouvant se fixer aux lysines acétylées des histones (voir **II.D : Modifications post-traductionnelles des histones**). Cette famille peut remodeler la chromatine en faisant glisser ou en éjectant les nucléosomes de l'ADN (**Figure** 7). Les membres de cette famille forment de larges complexes protéiques généralement formés autour de BRM (Brahma) chez la Drosophile ou Brg1 chez les Humains.

Les membres de la famille ISWI (*Imitation SWitch*) arborent trois domaines en C-terminale qui leur permettent de se fixer à l'ADN et aux queues d'histones H4 non-modifiées. Chez la Drosophile, l'enzyme ISWI est retrouvée avec d'autres complexes de remodelage comme NURF, CHRAC, ACF et RSF. La plupart de ces complexes catalyse l'espacement des nucléosomes mais sont également impliqués dans la compaction de la chromatine permettant la formation de structures d'ordre supérieur (**Figure** 7).

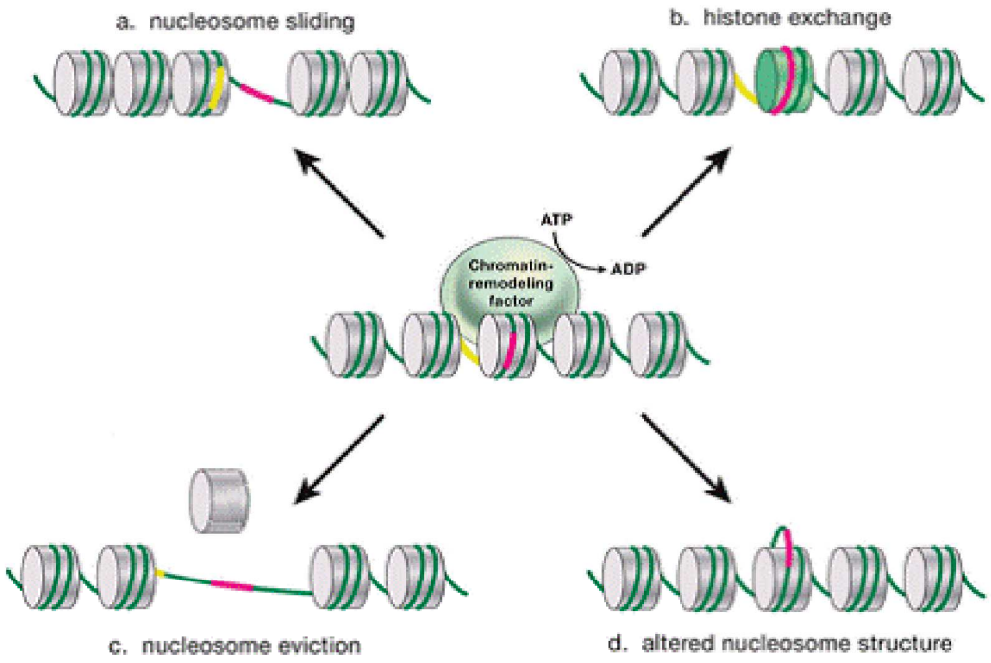


Figure 7 : Modes d'actions des complexes de remodelage de la chromatine

Le schéma (a) montre le glissement des nucléosomes le long de l'ADN ce qui permet l'accessibilité du motif rose aux DBP. (b) indique la substitution d'une histone par son variant au sein de l'octamère. L'éviction d'un nucléosome est montrée en (c) et permet également de libérer un site de reconnaissance aux DBP. Enfin, (d) indique l'altération de l'ADN autour d'un nucléosome ce qui conduit en la formation d'un espace entre l'ADN et le nucléosome. La région d'ADN est plus souple et permet de recruter des DBP. Figure tirée de : http://epigeneticss.blogspot.fr/

Les membres de la famille CHD (*Chromodomain-Helicase-DNA binding*) sont définis par la présence de deux chromodomaines en N-terminale. Les protéines de cette famille ont de nombreuses propriétés biologiques de par leur forte hétérogénéité (elles peuvent en effet être divisées en sous-groupes en fonction de la présence d'autres domaines protéiques). Le complexe le plus étudié est le complexe NURD (*Nucleosome remodeling and deacetylase*) qui est impliqué dans la répression transcriptionnelle de gènes cibles au cours du développement de la Drosophile et des mammifères.

Enfin, la caractéristique des enzymes de remodelage appartenant à la famille INO80 (*inositol requiring 80*) est la présence d'une insertion dans le domaine ATPase des enzymes. Leur fonction enzymatique n'est cependant pas affectée. Cette famille de protéines a la particularité de contrôler, de manière *genome-wide*, la déposition du variant d'histone H2A.Z (Voir **II.C : Les variants d'histones**) (PAPAMICHOS-CHRONAKIS *et al.* 2011). Nous verrons dans la partie suivante les caractéristiques associées à ce variant de l'histone H2A.

Ainsi, les complexes de remodelage de la chromatine permettent un accès dynamique des DBP au niveau de la séquence d'ADN en modifiant la distribution des nucléosomes. Ceci permet d'établir ou de modifier l'expression d'un gène. Cependant, la dynamique de la chromatine est aussi affectée par la substitution d'histones par leurs variants et par les modifications post-traductionnelles des histones au cœur des nucléosomes.

C. Les variants d'histones

Le remodelage des nucléosomes n'est pas la seule stratégie utilisée par les cellules eucaryotes pour réguler l'expression des gènes. En effet, les histones de cœur, dites canoniques (H2A, H2B, H3 et H4), possèdent des variants présentant des différences plus ou moins importantes en acides aminés par rapport aux histones canoniques. Ces différences ont un impact sur la structure et la stabilité des nucléosomes. De plus, ces variants peuvent également subir des modifications post-traductionnelles différentes des formes canoniques (ZILBERMAN *et al.* 2008). La différence majeure entre les histones canoniques et leurs variants est leur transcription. Les histones canoniques sont exclusivement transcrites durant la réplication de l'ADN alors que les formes variantes sont exprimées de manière constitutive. Les variants d'histones les plus communément étudiés sont décrits ci-après.

H2A possède trois variants notables. H2A.X est impliqué dans les mécanismes de réparation de l'ADN notamment sous sa forme phosphorylée (nommée ϒ-H2A) (FERNANDEZ-CAPETILLO *et al.* 2003; VAN ATTIKUM and GASSER 2009). L'incorporation du variant H2A.Z est associée à divers états transcriptionnels, inactifs, actifs ou en pause (ADAM *et al.* 2001; RAISNER *et al.* 2005; CREYGHTON *et al.* 2008; HARDY *et al.* 2009). En

effet, chez la Drosophile, la présence de H2Av aux promoteurs, l'homologue de H2A.Z, est corrélée avec l'état de pause de l'ARN polymérase II (Pol II) (MAVRICH *et al.* 2008). Cependant, chez la Drosophile et les mammifères, le variant H2A.Z peut également être retrouvé à la fois sur des promoteurs de gènes transcriptionnellement actifs marqués par la présence des protéines du groupe Trithorax (TrxG) (SARCINELLA *et al.* 2007), mais également au niveau de promoteurs de gènes réprimés, marqués par la présence des protéines du groupe Polycomb (PcG) (SWAMINATHAN *et al.* 2005; CREYGHTON *et al.* 2008). En effet, nous avons vu dans la partie précédente que le complexe de remodelage INO80 était responsable du dépôt de ce variant à la chromatine. Par ailleurs, la protéine INO80 interagit avec Pho chez la Drosophile, un membre du complexe du PcG. Ceci pourrait expliquer le recrutement de INO80 et donc la déposition de H2A.Z au niveau des gènes cibles du PcG (KLYMENKO *et al.* 2006; CAI *et al.* 2007). Dans ce dernier cas, H2A.Z est ubiquitiné par le complexe du PcG. Enfin, un dernier variant nommé macroH2A est principalement retrouvé au niveau du chromosome X inactivé chez les mammifères femelles (COSTANZI and PEHRSON 1998).

L'histone H3 possède deux variants universels. CENP-A est crucial lors de la mitose. Il prend place au niveau des centromères et permet le recrutement de protéines comme la Cohésine dont le rôle est important dans le maintien des chromatides sœurs lors de la mitose (SANTAGUIDA and MUSACCHIO 2009). Le deuxième variant est H3.3. Tout comme H2A.X, il joue un rôle à la fois dans l'activation des gènes mais également dans leur répression. Initialement, un certain nombre d'études ont attribué un rôle à ce variant dans l'activation transcriptionnelle étant donné son enrichissement sur le corps et le promoteur des gènes actifs (HENIKOFF 2008). De plus, les travaux de l'équipe de Gurdon réalisés chez le Xénope montrent que H3.3 joue un rôle dans le maintien de la mémoire d'activation. En effet, la présence de ce variant serait suffisante pour permettre le maintien de l'activation des gènes à travers les divisions cellulaires, sans besoin de facteurs de transcription pour réactiver les gènes après chaque division (NG and GURDON 2008). Ce rôle de « mémoire » ne semble cependant pas être conservé chez les mammifères et la Drosophile. Cependant, H3.3 est retrouvé enrichi aux promoteurs de gènes cibles des complexes du TrxG et du PcG (MITO *et al.* 2007) et est impliqué dans le recrutement de ce dernier dans les cellules souches embryonnaires (cellules ES) (BANASZYNSKI *et al.* 2013) ce qui a pour conséquence de réprimer les gènes cibles. Par ailleurs, H3.3 est également retrouvé aux niveaux de régions transcriptionnellement inactives telles que les centromères et télomères des chromosomes (SZENKER *et al.* 2012).

Pour conclure, nous avons vu que les principaux variants des histones canoniques sont impliqués dans des processus biologiques diverses et permettent une régulation de la transcription de manière locale. Cependant, comme suggéré dans les parties

précédentes, les histones peuvent être modifiées de manière post-traductionnelle ce qui impacte la régulation des gènes de manière encore plus spécifique.

D. Les modifications post-traductionnelles des histones

Les modifications post-traductionnelles des histones jouent un rôle fondamental dans la plupart des processus biologiques tels que la réplication, la réparation, la régulation de l'expression génique et la mitose. Ces modifications permettent de moduler les contacts chromatiniens en permettant de recruter des protéines spécifiques ou des complexes porteurs d'une activité enzymatique. Les acides aminés les plus souvent modifiés sont la lysine (K), l'arginine (R), la sérine (S) et la thréonine (T). Les modifications les plus étudiées à ce jour sont l'acétylation, la méthylation, l'ubiquitination et la phosphorylation (décrites ci-dessous et **Figure** 8), mais d'autres modifications ont également été d'écrites comme l'ADP-ribosylation, la sumoylation, la glycosylation ou la biotinylation (BANNISTER and KOUZARIDES 2011). Ces modifications covalentes sont principalement catalysées au niveau des queues N-terminales des histones ou de leur variants. Par ailleurs, ces modifications sont réversibles, ce qui permet une régulation fine de l'expression génique.

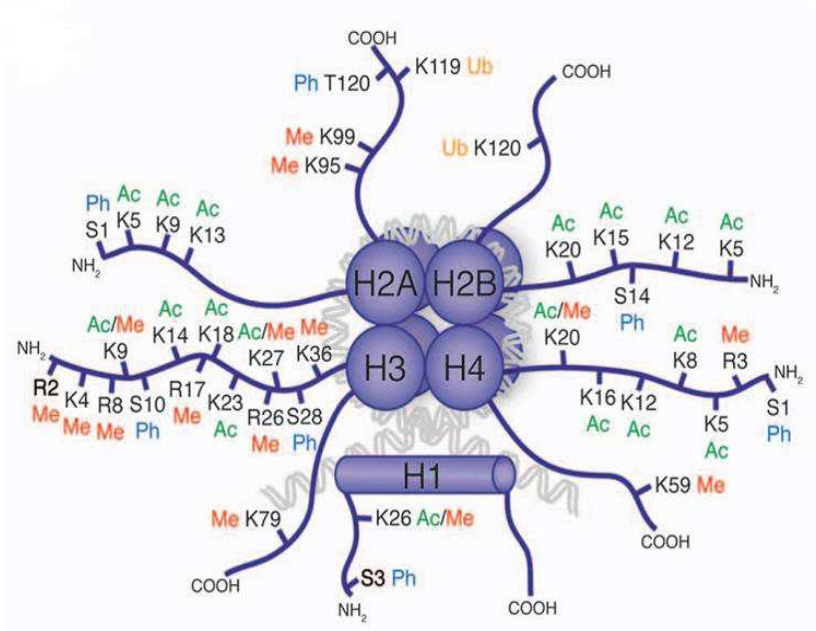


Figure 8 : Modifications des histones

Les extrémités N et C-terminales (indiquées NH_2 et COOH, respectivement) sortent du nucléosomes et subissent des modifications telles que la méthylation (rouge), l'acétylation (vert), la phosphorylation (bleu) et l'ubiquitination (orange) sur des résidus particuliers (R=Arginine ; K=Lysine ; S=Sérine). Les résidus sont numérotés selon leur position par rapport à l'extrémité N-terminale. Illustration de (TOLLERVEY and LUNYAK 2012).

i) L'acétylation

L'acétylation des histones est le plus souvent associée à l'activation transcriptionnelle. Les HAT (Histone acétyl-transférases) catalysent l'acétylation en transférant un groupement acétyle de l'acétyl Coenzyme A au niveau du groupement ammonium des lysines (abréviation = K). Cette modification a pour effet de neutraliser la charge positive des lysines, ce qui conduit en la diminution de l'interaction entre l'ADN et les histones. L'acétylation est donc associée à un état plus décondensé de la chromatine et est donc plus propice à l'activation des gènes (IMHOF and WOLFFE 1998; KUO *et al.* 1998).

Il existe trois grandes familles de HAT : les GNAT (*Gnc5-related Acetyl transferase*), les MYST et P300/CBP (*CREB binding protein*). Ces enzymes sont le plus souvent retrouvées au sein de larges complexes multimériques. Par exemple, la HAT dCBP a été purifiée à partir d'embryons de drosophile au sein du complexe activateur de la transcription du TrxG (PETRUK *et al.* 2001). Le complexe du TrxG associé à dCBP permet à la HAT de déposer un groupent acétyle au niveau de la lysine 27 de l'histone H3 (H3K27ac), une marque dont la présence aux TSS marque l'activation des gènes (WANG *et al.* 2008b; TIE *et al.* 2009). L'association des homologues de dCBP chez les mammifères (CBP et P300) avec le complexe du TrxG et l'activité sur H3K27 restent conservées.

L'acétylation des histones est principalement reconnue par un domaine protéique de type bromodomaine porté le plus souvent part des protéines issus des complexes HAT ou des complexes de remodelage de la chromatine, comme les membres de la famille SWI/SNF énoncés plus haut.

Selon les conditions physiologiques ou environnementales perçues par les cellules, le profil d'expression génique doit pouvoir être rapidement modifié. Par conséquent, l'acétylation des lysines doit également pouvoir être reversée pour permettre la répression des gènes. Cette action est effectuée par les HDAC (histone désacétylase) qui désacétylent les histones. La première HDAC découverte Rpd3 (*Reduced potassium dependency-3*) fût identifiée chez la levure en 1991 (VIDAL and GABER 1991). C'est la purification de son orthologue chez les mammifères qui a permis de démontrer son activité de désacétylase. Depuis, de nombreuses autres HDAC ont été identifiées. On en compte aujourd'hui 18 chez l'homme qui sont réparties en quatre classes en fonction de leur homologie de séquence avec l'enzyme Rpd3 et de leurs domaines fonctionnels (SETO and YOSHIDA 2014). Cinq de ces membres sont retrouvés chez la Drosophile (YANG and SETO 2008). Les HDAC montrent une faible spécificité pour leur substrat, une même enzyme pouvant désacétyler plusieurs résidus différents sur les histones. Par ailleurs, le fait que ces enzymes soient retrouvées au sein de complexes multimériques rend leur spécificité difficile à percer. Par exemple, les désacétylases HDAC1 et HDAC2 seules montrent une très faible activité enzymatique. Cependant, *in*

vivo, elles sont retrouvées au sein de trois grands complexes nommés Sin3, NuRD et CoREST. De manière générale, les gènes cibles de ces complexes sont impliqués dans divers mécanismes tels que les mécanismes développementaux, de divisions cellulaires et de migration cellulaire (HAYAKAWA and NAKAYAMA 2011).

ii) La méthylation

La méthylation peut être catalysée sur deux résidus, les lysines (K) (qui peuvent être mono-, di- ou tri-méthylées) et les arginines (R) (qui peuvent être mono-, di-méthylées symétriquement ou assymétriquement). Les enzymes catalysant ces méthylations sont donc très spécifiques et sont nommées HMT (Histone méthyl-transférases). Elles transfèrent un groupement méthyle provenant de la S-Adenosyl-Methionine aux résidus. Contrairement à l'acétylation, la méthylation ne modifie par la charge des résidus. Elle n'a donc pas d'effet en soi sur la structure de la chromatine. Cependant, la méthylation est associée à la régulation de la transcription et peut être associée soit à l'activation soit à la répression génique.

La méthylation des lysines a été plus étudiée que celle des arginines (ces dernières ne seront pas décrites dans ce manuscrit). La première HKMT (HMT ciblant les lysines (K)) a avoir été découverte est SUV39H1 chez les mammifères qui cible la lysine 9 de l'histone H3 (H3K9). Cette protéine, conservée chez la Drosophile, possède un domaine SET (Su(var)3-9, Enhancer-of-zeste and Trithorax) qui lui confère son activité (REA *et al.* 2000). Par la suite, un grand nombre de HKMT ont été découvertes dont les protéines Enhancer of zeste (E(z)) membre du complexe du PcG qui cible la lysine 27 de l'histone H3 (H3K27) et Trithorax (Trx), membre du complexe du TrxG qui cible la lysine 4 de l'histone H3 (H3K4). Cependant, il existe une autre classe de HKMT dépourvue du domaine SET comme l'enzyme Dot1 (cible H3K79).

De façon simplifiée, trois sites de méthylation sont impliqués dans l'activation de la transcription : H3K4, H3K36 et H3K79 et à l'inverse H3K9 et H3K27 sont quant à eux associés à la répression des gènes. Les études à grande échelle effectuées grâce aux technologies de ChIP (*Chromatin Immuno-Precipitation*) associées au séquençage (ChIP-seq ou ChIP-Chip) ont permis d'établir des cartes de profils de ces modifications le long des gènes (**Figure** 9). Ainsi, H3K4me3 est retrouvée enrichie dans la partie 5' des gènes et colocalise avec la Pol II sous forme initiée (phosphorylée sur la sérine 5). H3K36me3 est retrouvée dans la partie 3' et marque l'élongation de la transcription (Pol II phosphorylée sur la sérine 2). H3K4me2 s'étale autour des TSS actifs mais forme des piques autour des TSS des gènes réprimés (RICKELS *et al.* 2016). Enfin, H3K79me peut être retrouvée sur différentes positions des gènes selon son degré de méthylation et peut également être retrouvée au niveau de gènes réprimés. Ces phénomènes de

positions sont également retrouvés pour les marques associées à la répression. H3K9me2/3 associée aux TSS marque leur état répressif (VAKOC *et al.* 2005). Enfin, la marque H3K27me3, associée à l'activité du complexe du PcG, est retrouvée principalement en 5' des gènes chez les mammifères. Chez la Drosophile, cette marque peut cependant s'étendre et former de larges domaines répressifs contenant plusieurs gènes (BOYER *et al.* 2006; SCHWARTZ *et al.* 2006; SCHUETTENGRUBER *et al.* 2009).

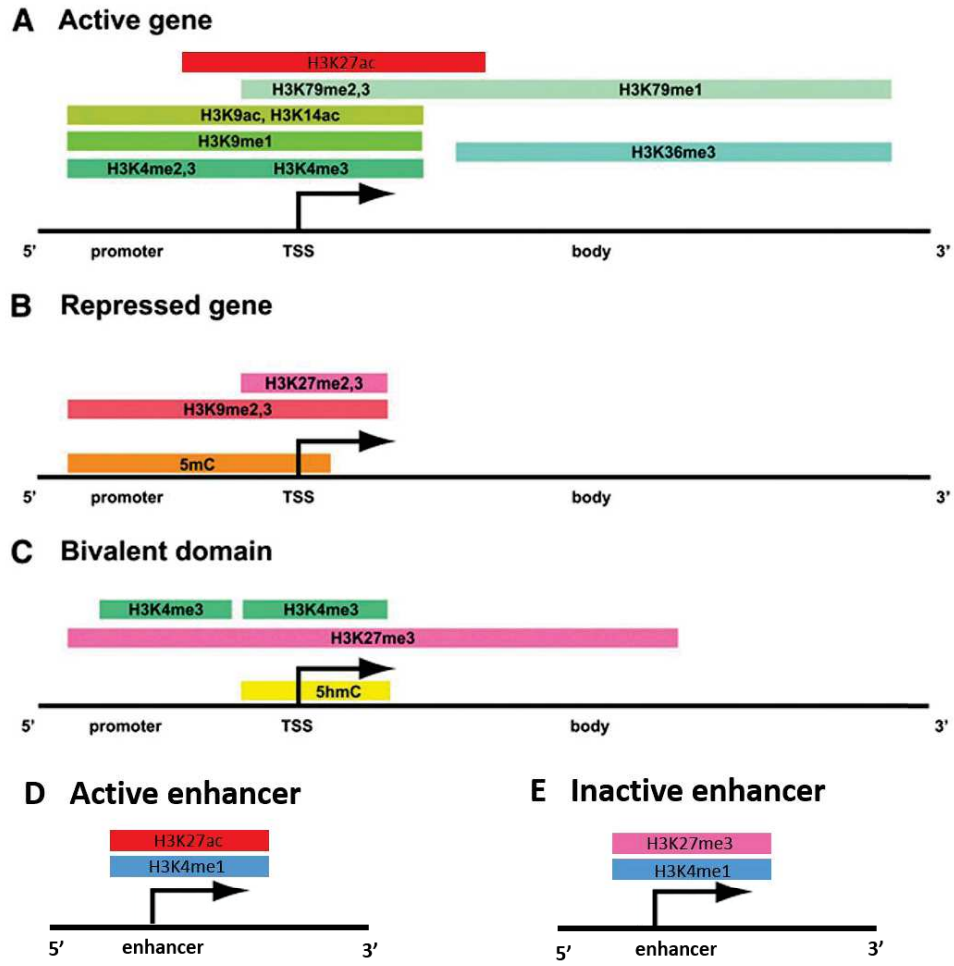


Figure 9 : Profils typiques des modifications d'histones le long d'un gène
Profils de différentes modifications post-traductionnelles d'histones établis chez les mammifères. 5mC correspond à la méthylation des ilots CpG de l'ADN. Figure adaptée de (BUTLER and DENT 2013).

Les protéines pouvant reconnaitre ces modifications d'histones possèdent soit un domaine Tudor, soit un Chromodomaine, soit un domaine PHD. Ces protéines peuvent appartenir aux complexes de remodelage de la chromatine (comme la protéine BPTF du complexe NURF qui reconnait spécifiquement la marque H3K4me3) ou à des complexes impliqués directement dans la répression et la compaction de la chromatine (comme la protéine Pc du complexe Polycomb qui reconnait spécifiquement H3K27me3 ou encore HP1 qui reconnait spécifiquement H3K9me2/3. Toutes deux possèdent un Chromodomaine). Enfin, ces modifications peuvent également être reconnues par des déméthylases.

Bien que la méthylation soit une modification extrêmement stable, elle peut cependant être retirée par les familles de protéines de type KDM1 (*Lysine (K) Demethylase 1*) ou JmjC (*Jumonji C*) (PEDERSEN and HELIN 2010). Par ailleurs, c'est au sein de cette famille que l'on retrouve la déméthylase JARID1 qui retire la marque activatrice H3K4me2/3. H3K27me2/3 est retirée par UTX, membre de la famille KDM1 (PEDERSEN and HELIN 2010).

iii) L'ubiquitination

L'ubiquitination des lysines consiste en la fixation d'une ubiquitine, un peptide de 8kDa. Cette modification est donc beaucoup plus imposante que les deux précédemment énoncées. L'ubiquitination se fait en trois étapes nécessitant l'action successive d'enzymes ; une enzyme d'activation E1, une enzyme de conjugaison E2 et pour finir une enzyme de liaison E3 (PICKART and EDDINS 2004). L'ubiquitine est transférée de E1 à E2 et est ensuite fixée au niveau des histones par l'action de E3. De manière générale, la poly-ubiquitination d'une protéine est associée au processus de dégradation de celle-ci. En revanche, les mono-ubiquitinations associées aux histones H2A et H2B sont associées à la régulation de la transcription.

Chez l'homme, la mono-ubiquitination de la lysine 120 de l'histone H2**B** est catalysée par Rad6A/B et est associée à l'activation de la transcription (KAO *et al.* 2004; XIAO *et al.* 2005). Par ailleurs, il existe un complexe de remodelage nommé Osa chez la Drosophile et BAF250 qui agit également comme ubiquitine-ligase de H2B120 (LI *et al.* 2010b). L'ubiquitination et la modulation de la chromatine sont donc liées. En revanche, la mono-ubiquitination de la lysine 119 de l'histone H2**A** (ou K118 chez la Drosophile) est associée à la répression des gènes (WANG *et al.* 2004a; ZHOU *et al.* 2008) et est déposée par la protéine RING1A/B (chez les mammifères) ou dRING (chez la drosophile) qui sont toutes deux associées au complexe Polycomb (discuté dans le Chapitre II).

Chez la Drosophile, H2BUb est retirée par dUSP36 (BUSZCZAK *et al.* 2009). Quant à H2AUb, elle peut être retirée par le complexe PR-DUB via l'enzyme Calypso (BAP1 chez les mammifères). Chez la Drosophile, ce complexe colocalise avec les protéines du PcG. De plus, H2AUb est également reconnue par la protéine ZRF1 (*zuotin-related factor 1*) qui provoque le retrait du complexe du PcG de manière spécifique (RICHLY *et al.* 2010). Cette protéine joue donc un rôle clef dans l'activation des gènes spécifiquement pré-ciblés par les protéines du PcG.

iv) La phosphorylation

La phosphorylation est catalysée par des kinases qui transfèrent un groupement phosphate de l'ATP sur des résidus sérines (S) ou thréonines (T). Comme pour l'acétylation, la phosphorylation altère la charge des histones par l'ajout de ce groupement négatif (MUSSELMAN *et al.* 2012b).

Les phosphorylations les plus étudiées sont celles des sérines 10 et 28 de l'H3 (H3S10P et H3S28P), toutes deux catalysées par des kinases de la famille Aurora. Initialement découverte comme impliquées dans la condensation et la ségrégation des chromosomes au cours de la mitose (WEI *et al.* 1999; GOTO *et al.* 2002), ces modifications sont également impliquées dans l'activation de la transcription. De façon surprenante, la mutation du résidu 28 de l'H3 (S28A) n'affecte pas la mitose mais décroit le niveau de H3K27me3 déposé par les protéines du PcG (YUNG *et al.* 2015). Ceci suggère que l'état non-phosphorylé de certain résidu est également important dans le recrutement à la chromatine de complexes comme celui du PcG.

La protéine 14-3-3ζ reconnait spécifiquement H3S10P et H3S28P et pourrait permettre le recrutement spécifique de facteurs impliqués dans l'activation transcriptionnelle (MACDONALD *et al.* 2005).

E. Communication entre les marques d'histones

Nous avons vu dans la partie précédente que les histones pouvaient être modifiées de manière post-traductionnelle. Ces modifications influent sur l'état transcriptionnel des gènes ciblés. Par ailleurs, les histones proches des gènes subissent des modifications concomitantes qui n'apparaissent pas indépendamment les unes des autres.

i) Combinaisons et communication entre les marques d'histones

Comme vu précédemment, les lysines peuvent subir trois modifications différentes : l'acétylation, la méthylation et l'ubiquitination. Les marques H3K27ac et H3K27me3 sont antagonistes et sont respectivement associées à l'activation ou la répression des gènes. Ces deux modifications sont donc exclusives.

Il existe cependant des modifications qui influent sur le recrutement d'autres protéines à la chromatine. Par exemple, le recrutement de HP1 par la marque H3K9me2/3 est perdu en présence d'une phosphorylation sur la sérine 10 voisine (FISCHLE *et al.* 2005). De la même manière, la phosphorylation de la sérine 28 de l'histone H3 provoque une diminution du recrutement des complexes du PcG à la chromatine ce qui a pour effet de diminuer le niveau de H3K27me3 (YUNG *et al.* 2015).

IL existe également des combinaisons en *trans* (entre résidus de différentes histones) comme par exemple l'association de H3K27me3 et de H2AUb, toutes deux déposées par le complexe du PcG.

Enfin, certaines marques d'histones sont plus spécifiquement associées à des éléments régulateurs. En effet, il a été montré que les enhancers, sont plus enrichis en H3K4me1 et H3K27ac que les promoteurs. (**Figure** 9) (KHARCHENKO *et al.* 2011).

Cependant, il est également possible de retrouver aux promoteurs de certains gènes des marques associées à la répression et à l'activation. Ces promoteurs sont dits « bivalents »

ii) Les promoteurs bivalents

Les promoteurs bivalents ont la particularité d'être enrichis en marque répressives (H3K27me3) et actives (H3K4me3) (VOIGT *et al.* 2013; HARIKUMAR and MESHORER 2015). Les gènes sont ainsi en état de pause transcriptionnelle (Pol II est phosphorylé sur la sérine 5), prêt à être soit réprimés (perte de H3K4me3) soit activés (perte de H3K27me3) selon le contexte cellulaire. Ces promoteurs sont particulièrement retrouvés dans les cellules ES où ils correspondent à des gènes de différenciation cellulaire. Ces gènes sont donc réprimés dans les cellules souches. Au cours du développement, certains deviennent actifs de manière tissu spécifique ce qui a pour conséquences de spécifier différents lignages cellulaires. De tels promoteurs n'ont pas été décrits chez la Drosophile (SCHUETTENGRUBER *et al.* 2009; SCHWARTZ *et al.* 2010).

Pour conclure, nous avons vu que l'état transcriptionnel des gènes était encodé par différents facteurs incluant les histones canoniques et leurs variants, mais également par leurs modifications post-traductionnelles. De manière remarquable, une

corrélation entre ces modifications d'histones et leurs localisations au sein du noyau est observée.

F. Les domaines chromatiniens

Dans la partie précédente, nous avons vu que les modifications post-traductionnelles des histones pouvaient impacter directement sur l'état de la chromatine. Ces modifications sont déposées par des complexes spécifiques et reconnues par des protéines également spécifiques.

Par ailleurs, ces modifications d'histones peuvent définir des domaines chromatiniens particuliers au sein du noyau. Traditionnellement, on divise le génome eucaryote en deux catégories. L'euchromatine constitue un environnement relâché contenant des gènes actifs. A l'inverse, l'hétérochromatine forme une structure plus compacte enrichie en gènes dont la transcription est réprimée. Cependant, les études à grande échelle des modifications d'histones révèlent des sous catégories de ces domaines chromatiniens. Chez la Drosophile par exemple, il existe cinq types distincts d'états chromatiniens, respectivement associés à une combinaison de modifications d'histones et de facteurs particuliers associés à la chromatine (**Figure** 10) (FILION *et al.* 2010). Par respect du mot *chroma* qui signifie couleur (grec), les auteurs ont donc proposé de les associer à des couleurs. Ils distinguent trois classes de chromatine répressives. L'hétérochromatine constitutive, verte, est marquée par la méthylation de H3K9 et est association aux protéines SU(VAR)3-9 et HP1. La chromatine associée au système du PcG (bleue) est caractérisée par la méthylation de H3K27 et par la présence des protéines du complexe du PcG. Cette chromatine est également considérée comme euchromatine facultative puisque les gènes réprimés par le PcG dans certain tissus peuvent être actifs (sans la présence du PcG) dans d'autres tissus. Enfin, la couleur noire a été donnée pour un type de chromatine pour laquelle aucune protéine spécifique n'a pu être définie. C'est par ailleurs la chromatine la plus abondante dans le génome. De plus, elle possède une propriété intrinsèque de répression génique.

Il existe enfin deux types de chromatines actives (jaune et rouge) qui diffèrent par leur composition biochimique et par les classes de gènes qu'elles contiennent. D'autres études n'ont pas retrouvé ces deux états chromatiniens. Cependant, la chromatine active a pu être classée en fonction des régions régulatrices associées comme « promoteurs actifs » ou « enhancers actifs » (ERNST *et al.* 2011; KHARCHENKO *et al.* 2011). De façon importante, ces domaines chromatiniens sont conservés chez les mammifères.

De manière importante, l'euchromatine et l'hétérochromatine ne sont pas retrouvées au même endroit au sein du noyau. De manière simplifiée, l'euchromatine est

retrouvée au centre du noyau alors que l'hétérochromatine est retrouvée en périphérie. De la même manière, les différentes classes d'hétérochromatines forment des territoires chromosomiques distincts des territoires transcriptionnellement actifs (CAVALLI and MISTELI 2013).

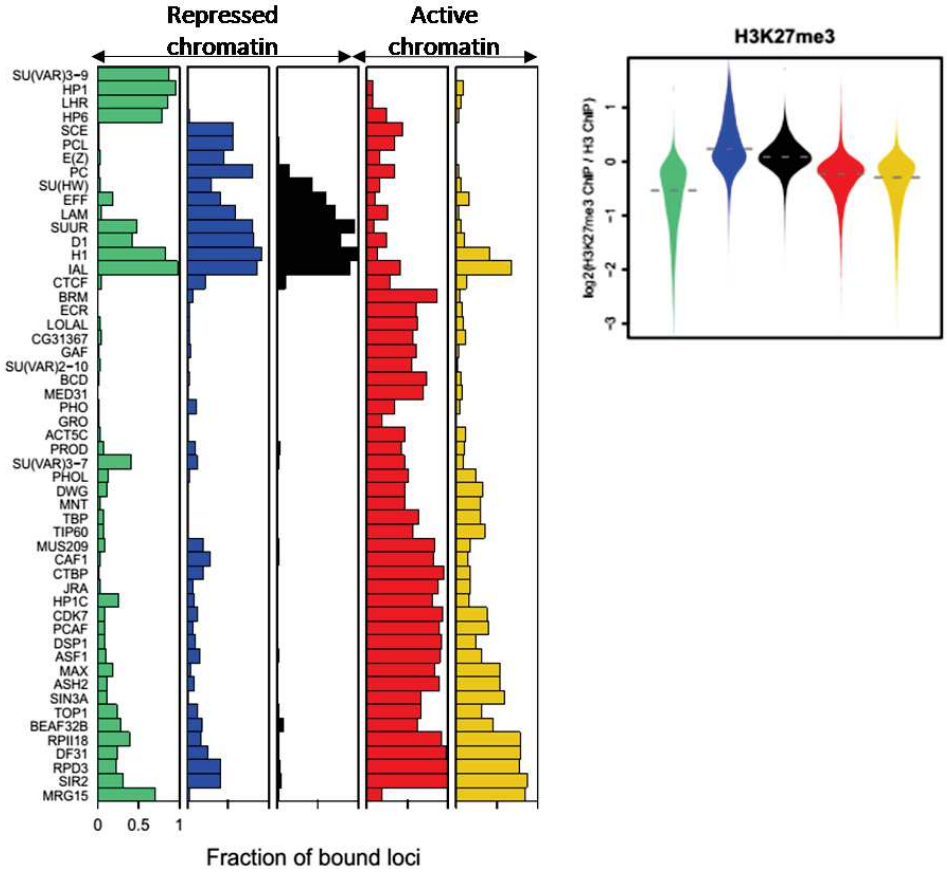


Figure 10 : Différentes « couleurs » de la chromatine définies par des combinaisons distinctes en protéines et marques d'histones.
La figure indique les différents états chromatiniens associés à la répression ou à l'activation des gènes. La figure de gauche indique les protéines testées dans les cellules Kc de Drosophile et la fraction de sites fixés pour chacune d'entre elles dans les différents états de la chromatine. La figure de droite indique le taux d'enrichissement en H3K27me3 dans chacun des cinq états chromatiniens. D'après cette figure, on note que la chromatine bleue est plus fortement enrichie en H3K27me3 que les quatre autres. Figure d'après (FILION *et al.* 2010).

Pour conclure, la chromatine est organisée en modules héritables de cellules en cellules lors des divisons cellulaires. Bien que l'ADN, support de l'information génétique, soit identique dans chacune des cellules d'un organisme, la complexité de cet organisme

requière une régulation fine de manière gène spécifique permis par les modifications de la chromatine.

G. Les domaines chromatiniens à l'échelle nucléaire

Comme vu précédemment, il existe différents types de chromatine suivant les modifications d'histones qui leurs sont associées. Ces différents types de chromatine ségrégent ensemble dans le noyau et conduisent à la formation de TAD (*Topologically Associating Domain*).

Le génome des métazoaires est organisé en TAD, qui correspondent à des régions linéaires de la chromatine se repliant en trois dimensions. Ces repliements forment des structures définies par la présence de fortes interactions. Les TAD sont hautement conservés à travers les espèces (DIXON *et al.* 2012; VIETRI RUDAN *et al.* 2015) et ségrégent de façon différentielle selon le profil des marques d'histones associées à ces régions, le niveau transcriptionnel des gènes au sein du TAD ou encore du timing de réplication des gènes. Les TAD peuvent donc être divisés en différentes catégories ; les TAD actifs ou encore les TAD réprimés (eux-mêmes sous-divisés en fonction de leur association avec les protéines du PcG, de l'hétérochromatine constitutive ou de la chromatine nulle (ou noire voir **Figure** 10). Les régions frontières entre les TAD sont marquées par la présence de gènes hautement transcrits et par la présence de protéines insulatrices telles que CTCF et CP190 (SEXTON *et al.* 2012; DOWEN *et al.* 2014).

Chapitre II : Présentation des complexes du PcG et du TrxG

La capacité de générer un organisme multicellulaire complexe à partir d'une cellule fécondée repose sur l'établissement de programmes transcriptionnels propres à chaque tissu. Ces programmes sont maintenus tout au long de la vie de l'organisme, bien après la disparition des FT (Facteurs de transcription) qui les ont initiés, par des mécanismes épigénétiques. Ces modifications épigénétiques correspondent à des modifications de l'état transcriptionnel de gènes promus par des modifications de la chromatine. De fait, la séquence d'ADN n'est pas altérée. De façon importante, ces changements de la chromatine, telles que les modifications d'histones ou les modifications covalentes de la séquences d'ADN (méthylation des ilots CpG, voir partie recrutement, Voir **Chapitre IV. III**), peuvent être maintenus au fils des divisions cellulaires. Contrairement aux modifications de la séquence d'ADN, les modifications épigénétiques sont réversibles.

Deux familles de protéines jouent un rôle important dans le maintien de l'état transcriptionnel des gènes aux cours des divisions cellulaires. Les complexes du PcG maintiennent l'état répressif des gènes. Les complexes du TrxG maintiennent l'état actif des gènes. Ces deux familles participent donc au maintien de la mémoire de programmes génétiques préalablement établis.

I. Les gènes homéotiques

Les protéines du PcG et du TrxG ont initialement été découvertes chez la Drosophile comme régulateurs transcriptionnels des gènes homéotiques ou gènes HOX. De façon importante, ces protéines sont hautement conservées chez tous les métazoaires.

Les gènes HOX sont un groupe de gènes régulant le développement de l'axe antéropostérieur des embryons. Ils ont initialement été découverts chez la Drosophile où ils se divisent en deux clusters distincts: les complexes Antennapedia et Bithorax (ANT-C et BX-C, respectivement) (LEWIS 1978; KAUFMAN *et al.* 1980; LEWIS *et al.* 1980) (**Figure** 11). Des mutations ou dérégulations de ces gènes conduisent à des transformations homéotiques (du terme *homeosis,* défini par William Bateson comme la conversion d'une structure en une autre structure) dans laquelle l'identité d'un segment donné est transformée en celui d'un autre segment.

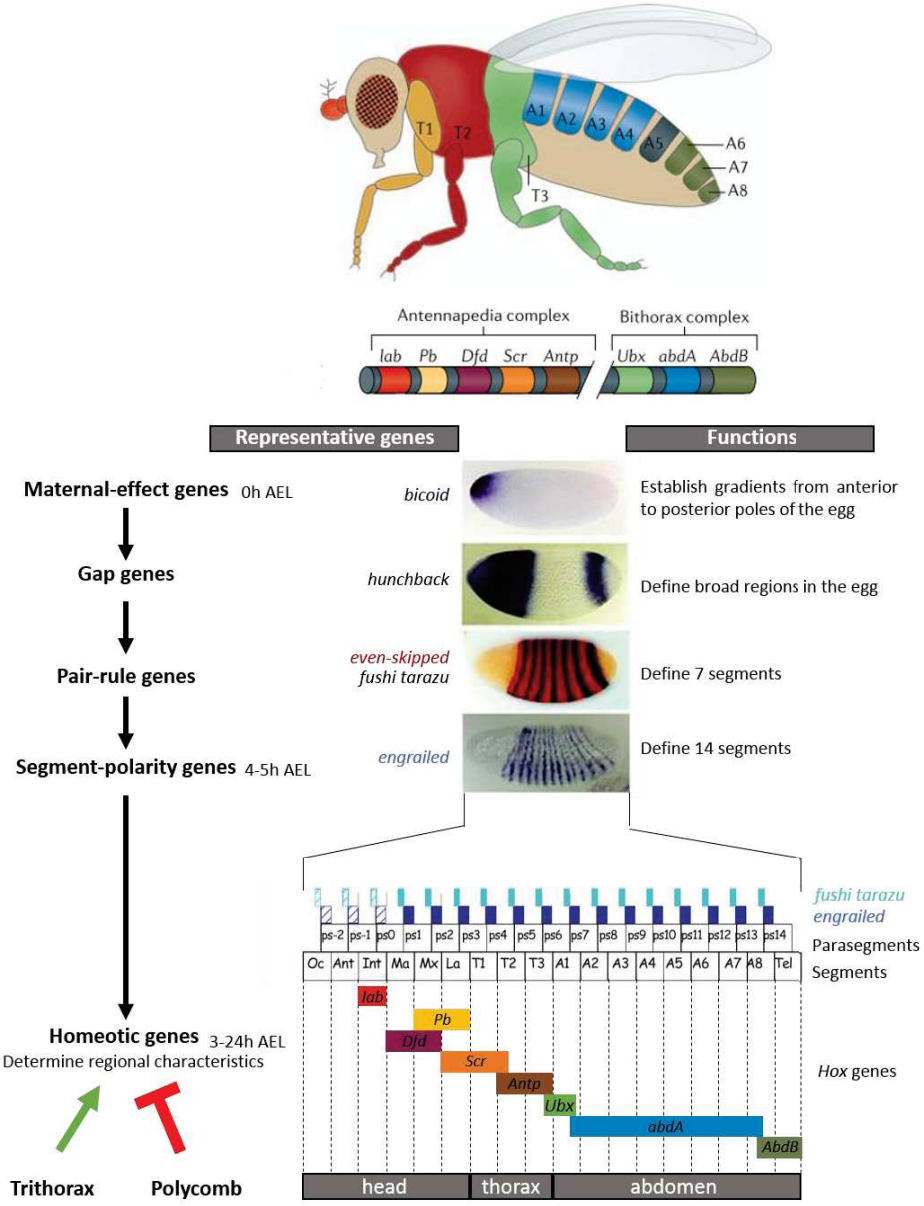


Figure 11 : Etablissement et maintien du profil d'expression des gènes homéotiques (HOX) de l'embryon à la mouche adulte.

La figure du haut représente les gènes HOX qui sont répartis en deux clusters, Antennapedia (Ant-C) et Bithorax (Bx-C) au niveau du chromosome III de la Drosophile (zones colorées en cylindre). L'ordre de ces gènes sur le génome est colinéaire à leur domaine d'expression le long de l'axe antéropostérieur de l'animal (code couleur des gènes HOX qui indique leur domaine d'expression dans la mouche adulte). La figure du bas montre le profil d'expression des gènes représentatifs de chaque classe de gènes indiquées à gauche (maternellement déposés ; Gap ; Pair-rule ; de segments et enfin homéotiques) puis le maintien du profil d'expression des gènes HOX par les protéines Trithorax et Polycomb. La figure du haut est tirée de (SPARMANN and VAN LOHUIZEN 2006). La figure du bas est adaptée de (ANGELINI *et al.* 2005).

Par ailleurs, l'ordre d'expression de chaque gène HOX le long de l'axe antéro-postérieure de l'embryon reflète l'ordre physique des gènes HOX au sein des clusters (HARDING *et al.* 1985; AKAM 1987) (**Figure** 11). Cette caractéristique reflète une colinéarité spatio-temporelle de l'expression et de la position des gènes HOX. Ainsi, chez la Drosophile, cette colinéarité s'explique par la présence du cluster ANT-C qui régule et détermine l'identité de la partie antérieure de la mouche, alors que le cluster BX-C est responsable de l'identité de la partie postérieure de la mouche (**Figure** 11). De manière remarquable, cette colinéarité spatio-temporelle est hautement conservée dans le phylum animal où les gènes HOX ont évolué par duplication puis divergence reflétant l'importance cruciale de ces gènes dans l'évolution des animaux. L'analyse rigoureuse de l'expression des gènes HOX au cours de l'embryogénèse fût récompensée d'un Prix Nobel en 1995 (Médecine et de Physiologie ; Edward B. Lewis, Christiane Nüsslein-Volhard, et Eric F. Wieschaus).

Les protéines du PcG et du TrxG ont initialement été découvertes par leur rôle essentiel dans le maintien de l'état transcriptionnel des gènes HOX (**Figure** 11) (GRIMAUD *et al.* 2006; SCHWARTZ and PIRROTTA 2007; SCHWARTZ and PIRROTTA 2008). En effet, la formation de l'axe antéropostérieur se fait dès les premières heures du développement grâce à la transcription d'ARNm déposés maternellement (**Figure** 11). Leur traduction déclenche une cascade de régulation entrainant l'expression séquentielle de gènes de segmentation (Gènes *GAP*, *Pair-rule* puis *segment-polarity*, Figure) qui vont établir un profil d'expression des gènes HOX (**Figure** 11). Cependant, les produits issus de l'expression des gènes de segmentation ne sont que transitoires. Ainsi, une fois le profil d'expression des gènes homéotiques établi, il doit être maintenu par des mécanismes qui vont altérer la conformation chromatinienne de ces gènes. Le maintien de la répression des gènes HOX se fait grâce à la relève des produits des gènes de segmentation par les protéines du PcG (**Figure** 11). Ces derniers vont maintenir la répression des gènes homéotiques en dehors de leur domaine d'expression. Il existe également un groupe de protéines dont l'activité est antagoniste à celle du PcG. Il s'agit des protéines du TrxG dont le rôle est de maintenir l'expression des gènes homéotiques à l'intérieur de leur domaine d'expression (**Figure** 11).

II. Les protéines du TrxG (*Trithorax Group*)

Les protéines du TrxG ont initialement été découvertes chez la Drosophile sur la base des phénotypes liés à leurs mutants (GRIMAUD *et al.* 2006). Les protéines du TrxG forment de larges complexes multimériques pouvant être classés en deux groupes. D'une part les complexes intervenant dans le remodelage de la chromatine ; d'autre part les complexes modifiant les histones (**Figure** 12).

Complexes	Drosophila	Mammals	Functions
Histone modifying complexes			
SET1/COMPASS	dSET1	SET1A/B	SET domain - Maintain the bulk level of H3K4me2/3 in the genome
MLL1/2 COMPASS-like	Trx	MLL1/2	SET domain Trx: - Maintain the memory of actively transcribed HOX genes - Deposition of H3K4me3 -Deposition of H3K4me2 at PREs MLL2: -Deposition of H3K4me3 at bivalent promoters in ES cells MLL1: - Deposition of H3K4me3 at only a subset of genes including HOX genes
MLL3/4 COMPASS-like	Trr dUTX	MLL3/4 UTX	SET domain of Trr deposits the H3K4me1 at enhancers UTX protein demethylates H3K27me3
ASH1	Ash1 dCBP	ASH1L CBP	SET domain of Ash1 deposits the H3K36me3 The HAT domain of dCBP is responsible for the acetylation of H3K27
ATP-dependent chromatin-remodelling complexes			
SWI/SNF	BRD	BRM/BRG1	The bromodomain binds to acetylated histones The two mammalian proteins are found in two distincts complexes BAF and PBAF which regulate chromatin structure of a large number of genes implicated in cell cycle, signaling and proliferation

Figure 12 : Complexes associés aux protéines du groupe Trithorax.
Les protéines du TrxG sont retrouvées au sein de deux types de complexes. Ceux intervenant dans les modifications d'histones et ceux intervenant dans le remodelage de la chromatine. Les fonctions majeures sont indiquées pour chaque complexe ainsi que les protéines portant l'activité enzymatique au sein des complexes. Le nom de ces enzymes est indiqué pour la Drosophile comme pour leur homologue mammalien. Adapté de (SCHUETTENGRUBER *et al.* 2011).

Les complexes impliqués dans le remodelage de la chromatine ont été décrits précédemment. Ci-dessous seront donc décrits les complexes du TrxG impliqués dans les modifications d'histones.

Les protéines du TrxG sont impliquées dans l'acétylation (PETRUK *et al.* 2001) et la méthylation d'histones (MILNE *et al.* 2002; NAKAMURA *et al.* 2002). Tout d'abord, le complexe COMPASS contient l'enzyme dSET1 (*Drosophila SET domain-containing 1*). Cette protéine est la méthylase majoritairement responsable du dépôt global de la triméthylation de H3K4 (WU *et al.* 2008; ARDEHALI *et al.* 2011; MOHAN *et al.* 2011; SHILATIFARD 2012). On trouve également deux complexes COMPASS-like dont l'un contient la protéine Trx (*Trithorax*) principalement responsable du maintien de l'état actif des gènes HOX et catalysant également la triméthylation de H3K4 (WU *et al.* 2008). L'autre complexe COMPASS-like contient la protéine Trr (*Trithorax-related*), responsable du dépôt de H3K4me1 au niveau des enhancers (HERZ *et al.* 2012; HU *et al.* 2013). Par ailleurs, ce dernier complexe contient également la déméthylase UTX. Finalement, le dernier complexe ASH1 est caractérisé par la présence de dCBP (PETRUK *et al.* 2001; TIE *et al.* 2009), responsable de l'acétylation de H3K27 et Ash1 qui catalyse la triméthylation de H3K36 (BANTIGNIES *et al.* 2000; YUAN *et al.* 2011).

III. Les complexes du PcG (*Polycomb Group*)

Les protéines du PcG ont été initialement identifiées chez la Drosophile sur la base de mutations associées à la dérépression des gènes HOX. En effet, les mouches mâles possèdent des peignes sexuels (ou soies ; *combs*) au niveau de la première paire de pattes. Chez les mouches hétérozygotes pour un gène du groupe Polycomb, on observe une apparition ectopique de soies au niveau de la seconde et de la troisième paire de pattes (**Figure** 13). Le rôle des gènes du PcG est d'autant plus important que les mutations à l'état homozygote sont létales. On observe alors une mort précoce des drosophiles au stade embryonnaire associée à une sévère transformation homéotique.

Chez la Drosophile comme chez les mammifères, il existe deux principaux complexes associés au PcG : PRC1 et PRC2, tous deux associés à des activités biochimiques différentes. De manière générale, PRC2 est responsable du dépôt de H3K27me3 aux sites de fixations des complexes du PcG. Cette marque est reconnue par le complexe PRC1 qui à son tour catalyse la mono-ubiquitination de la lysine 118 de l'histone H2A. Cependant, d'autres complexes, moins étudiés existent et seront décrits ci-dessous. Par ailleurs, ces complexes sont très conservés chez les mammifères. On observe cependant bien plus d'homologues fonctionnels chez ces derniers que chez la Drosophile.

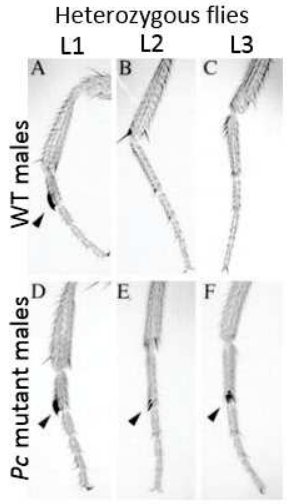


Figure 13 : Phénotypes associés aux mutations des protéines du PcG chez la Drosophile

En haut, les images correspondent respectivement aux trois paires de pattes (L1, L2 et L3 ; L pour *Legs*) d'une drosophile male. La première paire de pattes (L1) est caractérisé par des soies indiquées par la flèche. Ces soies n'existent pas sur les autres paires de pattes des mouches sauvages. Une mutation hétérozygote pour un gène du PcG provoque l'apparition ectopique de soies sur les deuxième et troisième paires de pattes (en bas) (Photo tirée de (HODGSON *et al.* 2001)).

A. Le complexe PRC1 (*Polycomb Repressive Complex 1*)

i) Chez la Drosophile

Le complexe PRC1 a été initialement purifié à partir d'embryons de Drosophile (SHAO *et al.* 1999; SAURIN *et al.* 2001). Il contient de manière stœchiométrique quatre sous-unités (**Figure** 14). Pc (*Polycomb*), Ph (*Polyhomeotic*), Psc/Su(z)2 (*Posterior sex combs*/*Suppressor of zeste 2*) et dRING (*Really Interesting New Gene* initialement nommé Sce pour *Sexcombs extra*) (**Figure** 14. A). Chacune de ces sous-unités possède des propriétés biochimiques et fonctionnelles qui leurs sont propres.

La sous-unité Ph est codée par deux gènes dupliqués *ph-d* et *ph-p*. Cette protéine possède des domaines fonctionnels conservés dont le domaine SAM (*Sterile Alpha Motif*) (PETERSON *et al.* 1997) (**Figure** 15). Ce domaine a la particularité de pouvoir polymériser et donc de permettre la formation d'homodimères Ph-Ph favorisant la répression des gènes cibles (ROBINSON *et al.* 2012). D'autre part, ce domaine permet également la formation d'hétérodimères avec d'autres protéines de la même famille (KIM *et al.* 2002). En effet, le domaine SAM de Ph est retrouvé dans deux autres protéines : Sfmbt (*Scm [Sexcomb on midleg] with four MBT [Malignant Brain Tumor] domains)* qui fait partie du complexe Pho-RC (voir ci-après) et Scm (*Sexcomb on midleg*). De façon intéressante, les deux domaines SAM respectifs à Ph et Scm interagissent *in vitro* pouvant ainsi former de longs copolymères (PETERSON *et al.* 1997; KYBA and BROCK 1998; KIM *et al.* 2002; KIM *et al.* 2005). Cependant, bien que Scm et sa propriété d'interaction avec Ph soient conservées chez les mammifères (BONASIO *et al.* 2014; NANYES *et al.* 2014), Scm est retrouvée en quantité sous-stœchiométrique par rapport au complexe PRC1 (SHAO *et al.* 1999). Néanmoins, Scm et les autres protéines du complexe PcG colocalisent sur de nombreux sites sur les chromosomes polytènes (KANG *et al.* 2015).

La sous-unité Pc possède un chromodomaine en N-terminale lui permettant de reconnaitre spécifiquement H3K27me3 (**Figure** 15). Pc possède également un autre domaine conservé chez les mammifères. Ce domaine situé en C-terminale a été nommé PcR box (*Polycomb repressor box* ou CBOX) en raison de son importance dans la répression transcriptionnelle. En effet, la délétion de la partie C-terminale de Pc affecte la répression des gènes cibles (BUNKER and KINGSTON 1994; MULLER 1995; SATIJN *et al.* 1997; BREILING *et al.* 1999), une propriété qui ne semble pas être conservée chez les orthologues mammaliens (GRAU *et al.* 2011; VOLKEL *et al.* 2012).

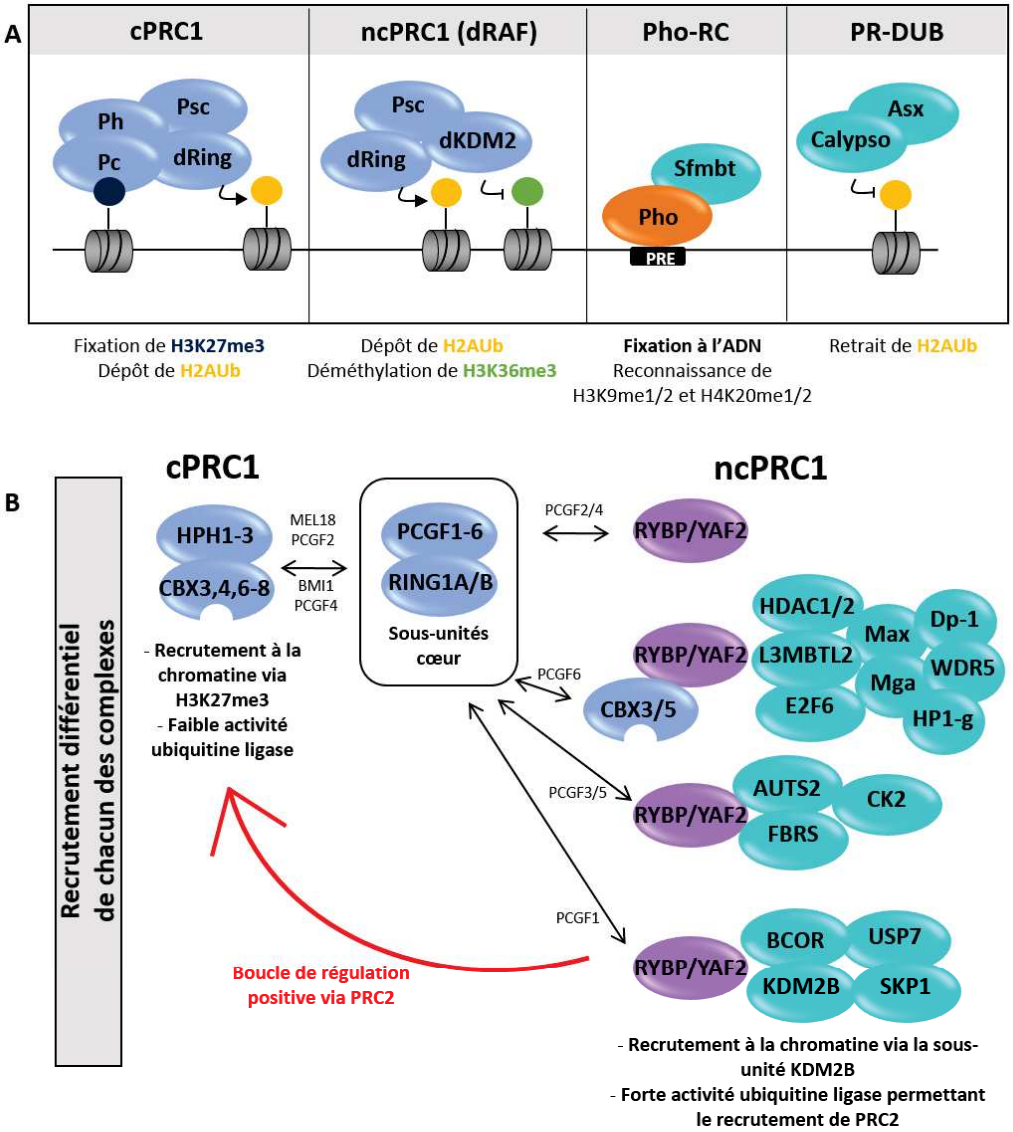


Figure 14 : Présentation des complexes PRC1 chez la drosophile et les mammifères et d'autres complexes relatifs à l'activité PRC1

A- Chez la Drosophile, on trouve le complexe PRC1 et le complexe dRAF. Tout comme PRC1, dRAF peut également déposer la modification H2AK118Ub. Cette marque est retirée par le complexe PR-DUB. Enfin, le complexe Pho-RC ne contient pas d'activité de modification de la chromatine mais peut être directement recruté aux PRE (qui sont les sites de fixation des complexes du PcG chez la Drosophile) par l'intervention du FT Pho (Voir **Chapitre IV.I**).

B- Chez les mammifères, on distingue deux grandes familles de protéines PRC1, les PRC1 canoniques et non-canoniques selon la présence de CBX ou RYBP/YAF2. Généralement, les complexes PRC1 canoniques sont assimilés au complexe PRC1 de la Drosophile. Le complexe non-canonique PRC1 contenant PCGF1 et BCOR est plus communément appelé complexe BCOR et apparait comme l'homologue fonctionnel de dRAF. Il est à noter que le complexe PR-DUB existe aussi chez les mammifères (voir texte).

La sous-unité Psc possède un homologue fonctionnel nommé Su(z)2 (LO *et al.* 2009). Ces deux protéines ont en commun un domaine RING (BRUNK *et al.* 1991; DESHAIES and JOAZEIRO 2009) leur permettant de s'associer avec dRING (**Figure** 15), une E3 ubiquitine ligase (voir plus bas). Cette association a été suggérée comme importante dans la stabilisation de l'enzyme E2 pour une mono-ubiquitination optimale de H2AK118 (BUCHWALD *et al.* 2006; LI *et al.* 2006). De plus, la partie C-terminale de Psc joue un rôle direct sur la structure chromatinienne (LO and FRANCIS 2010). En effet, Psc a la particularité de pouvoir inhiber la transcription *in vitro* (KING *et al.* 2002) et de compacter les nucléosomes (SHAO *et al.* 1999; FRANCIS *et al.* 2001; FRANCIS *et al.* 2004) (**Figure** 15, *domaine désordonné*). La région comprenant les acides aminés 456-1603 est prédite comme étant sans structure (EMMONS *et al.* 2009). Plus précisément, la région comprenant les acides aminés 456-909 permet la fixation aux nucléosomes mais ne peut pas les compacter. Cette activité est produite par la région située à l'extrémité C-terminale et comprenant les acides aminées 910-1603 (LO and FRANCIS 2010). De façon importante, la compaction de la chromatine par Psc est indépendante de dRING et de la marque H2AUb qui lui est associée (ESKELAND *et al.* 2010; LO *et al.* 2012). De façon intéressante, cette région n'est pas conservée entre Psc et son homologue fonctionnelle Su(z)2 (BRUNK *et al.* 1991), ce qui suggère que le rôle de compaction de la chromatine est propre à Psc. En effet, d'autres analyses révèlent que lors de la réplication de l'ADN, la protéine Psc reste maintenue à certains sites correspondant aux frontières des TAD (FOLLMER *et al.* 2012). De par ses propriétés d'assemblage à la chromatine (LO and FRANCIS 2010; LO *et al.* 2012), Psc semblerait contribuer au ré-établissement des domaines du PcG marqués par H3K27me3 après la mitose (FOLLMER *et al.* 2012). Enfin, la partie N-terminale de Psc comprend le domaine RAWUL (*Ring finger And WD40 Ubiquitin-Like*) (SANCHEZ-PULIDO *et al.* 2008; JUNCO *et al.* 2013), un domaine retrouvé dans la sous-unité dRING (**Figure** 15). Plus que toutes autres interactions non covalentes, le domaine RAWUL est certainement le plus important pour la formation du complexe PRC1 (**Figure** 15).

La sous-unité dRING est une E3 ubiquitine-ligase (FRITSCH *et al.* 2003; DE NAPOLES *et al.* 2004; WANG *et al.* 2004a). Chez la drosophile, elle est responsable de la mono-ubiquitination de la lysine 118 de l'histone H2A (Lysine 119 chez les mammifères). Nous avons vu plus haut que le domaine RING de dRING pouvait interagir avec Psc. Le domaine RAWUL de dRING est également capable de former des interactions avec Pc (WANG *et al.* 2008a; BEZSONOVA *et al.* 2009) pouvant générer des hétéro-dimères stables (**Figure** 15). Enfin, des mutants pour dRING montrent des phénotypes de dérégulation des gènes homéotiques typiques des mutants pour les protéines du groupe PcG (WANG *et al.* 2004a).

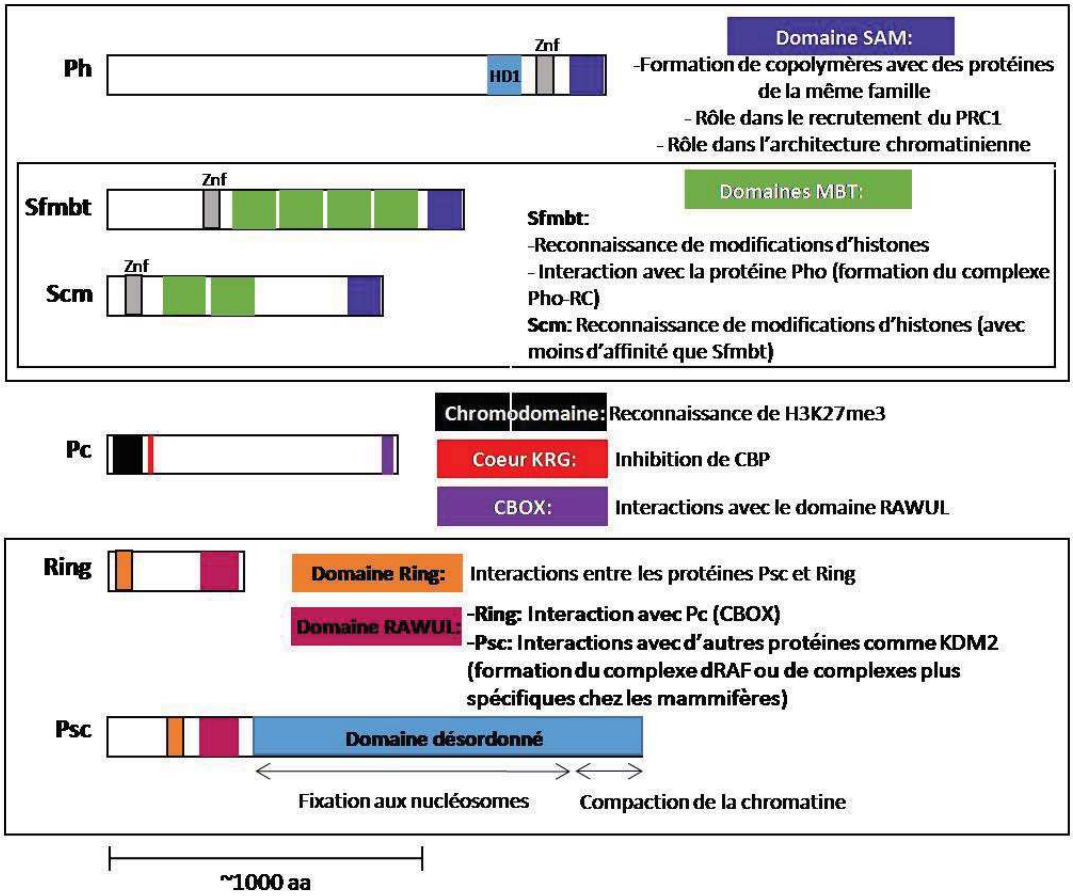


Figure 15 : Domaines fonctionnels de chaque protéine du PRC1 et de la protéine Scm

Les domaines fonctionnels de ces protéines sont conservés chez les mammifères. Leur fonction est décrite à droite du schéma.

ii) Chez les mammifères

La protéine Psc de la Drosophile est connue sous le nom de *Polycomb group finger* chez les vertébrés (PCGF) et est représentée par 6 homologues (PCGF 1-6 ; **Figure** 14B). La protéine Ph s'est diversifiée en 3 paralogues : PHC 1-3. La protéine dRing de la Drosophile a deux homologues chez les vertébrés RING1A et B. Pour finir, la protéine Pc, connue sous le nom de CBX chez les vertébrés, s'est diversifiée en 5 différentes protéines : CBX2, 4, 6, 7 et 8. Par ailleurs, les différents homologues de CBX montrent un profil de localisation nucléaire propre (SOWPATI *et al.* 2015) indiquant que chaque homologue possède son propre set de sites de fixations occupant des compartiments

nucléaires distincts. Ainsi, l'augmentation de la diversité des protéines du PRC1 reflète leurs implications dans des fonctions cellulaires différentes.

Gao *et al*, ont été les premiers à soumettre une analyse génomique et protéomique globale sur la nature de l'hétérogénéité des complexes PRC1 de mammifère (GAO *et al.* 2012). De cette étude ressort un point important : il existe deux types de complexes PRC1 chez les mammifères. Les complexes canoniques PRC1 contiennent RING1A/B associé soit à PCGF2 soit à PCGF4 (nommés respectivement cPRC1.2 et cPRC1.4) (**Figure** 15). Aux deux protéines s'ajoutent une sous-unité CBX et PHC dont la présence est essentielle au recrutement de ces complexes à la chromatine (via le chromodomaine conservé de CBX). Les complexes non-canoniques ne contiennent aucune sous-unité CBX ou PHC mais plutôt une sous-unité RYBP ou YAF2 (**Figure** 15). RYBP/YAF2 se fixent à la même poche d'interaction que CBX à la surface de RING1A/B (le domaine RAWUL évoqué dans la partie précédente). Ainsi, les interactions RING1A/B—RYBP/YAF2 et RING1A/B—CBX sont mutuellement exclusives (WANG *et al.* 2010; JUNCO *et al.* 2013; CHITTOCK *et al.* 2017). Une autre particularité des PRC1 non-canoniques est qu'ils peuvent contenir chacune des sous-unités PCGF (**Figure** 15). Le PRC1 non-canonique contenant PCGF6 (PRC1.6) est composé de différentes DBP (ex : E2F6, MAX and MGA), de protéines pouvant se fixer à des modifications d'histones (ex : CBX3), de protéines pouvant modifier les histones (ex : HDAC1 et 2) et d'autres sous-unités (TRIMARCHI *et al.* 2001; AKASAKA *et al.* 2002; TROJER *et al.* 2011; GAO *et al.* 2012; QIN *et al.* 2012). Les complexes PRC1 non-canoniques associés à PCGF3 ou PCGF5 partagent une grande partie de leurs sous-unités incluant FBRS, FBRSL1, CSNK2A et CSNK2B. Ils sont particulièrement intéressants du fait de la présence d'une autre sous-unité, AUTS2 qui est fréquemment mutée dans les maladies de troubles autistiques (GAO *et al.* 2012; GAO *et al.* 2014). De façon importante, le complexe non-canonique BCOR (PRC1.1 ; BCOR) associé à PCGF1 est l'homologue de dRAF chez la Drosophile (GEARHART *et al.* 2006; FARCAS *et al.* 2012; GAO *et al.* 2012; WU *et al.* 2013). Il comprend notamment la lysine déméthylase KDM2B ainsi que BCOR, BCORL1, SKP1 et USP7. Enfin, il est également possible que des complexes non-canoniques contenant PCGF2 ou PCFG4 existent puisque chacune de ces protéines peut immuno-précipiter RYBP et YAF2 (GAO *et al.* 2012; TAVARES *et al.* 2012). Cependant, la composition exacte de ces complexes reste à être élucidée.

Une telle diversité des complexes PRC1 suggère un rôle possible de chacun de ces complexes dans la régulation de réseaux transcriptionnels de manière type-cellulaire spécifique. De manière générale, il a été suggéré que les PRC1 non canoniques avaient un rôle particulier dans la régulation du cycle cellulaire et dans la régulation de gènes du métabolisme. En revanche, les PRC1 canoniques jouent un rôle plus prépondérant dans l'engagement des lignées cellulaires et dans la différenciation (MOREY *et al.* 2013; KLOET *et al.* 2016).

B. Le complexe dRAF (*dRING Associated Factors*)

Une autre étude biochimique réalisée à partir d'embryons de Drosophile a révélé que dRING pouvait s'associer à un autre complexe nommé dRAF (LAGAROU *et al.* 2008) (**Figure** 14A). Ce complexe est formé par l'association de dRING avec Psc et la déméthylase dKDM2 spécifique de H3K36me2. Par ailleurs, une reconstitution biochimique des complexes PRC1 et dRAF montre que dRING seule a une activité ubiquitinase très limitée. Cette activité est modérément augmentée par l'ajout de Psc et largement augmentée par l'ajout de dKDM2. Ceci suggère que l'essentiel de H2AUb déposé par le système Polycomb est associée à l'activité dRAF plutôt que par celle du PRC1. De façon intéressante, ce complexe et ses propriétés sont également conservés chez les mammifères à travers le complexe PRC1 non-canonique BCOR (**Figure** 14B).

Enfin, une étude récente réalisée par l'équipe de V.Pirrotta indique que l'activité ubiquitinase de dRING n'est pas une composante essentielle à la répression du système Polycomb (LEE *et al.* 2015). Etant donné que l'activité ubiquitine transférase de dRING requière son interaction avec un partenaire tel que Psc, Lee et al., ont remarqué que la perte de Psc dans des cellules de Drosophile n'affecte qu'une fraction de la totalité de H2AUb. En revanche, un *knock-down* de l(3)73Ah, un homologue de PCGF3 résulte en une perte de 70% de la totalité de H2AUb suggérant que près des trois quarts de H2AUb du génome de la Drosophile n'est pas liée avec l'activité dRING du système Polycomb. De plus, le profil de H2AUb est relativement étendu dans le génome et sa présence ne corrèle pas forcément avec les gènes cibles du PcG.

C. Le complexe Pho-RC (*Pho-Repressive Complex*)

Pho est une protéine importante dans le maintien de la répression par le système Polycomb (BROWN *et al.* 2003; WANG *et al.* 2004b). Contrairement aux autres protéines du PcG, elle possède quatre motifs *zinc-finger* lui permettant de se fixer à l'ADN (BROWN *et al.* 1998). Ainsi, le laboratoire de J. Muller, souhaitant en savoir d'avantage sur le rôle de Pho, a purifié les complexes protéiques dans lesquels se trouvait la protéine (KLYMENKO *et al.* 2006). Cette analyse a révélé la présence de Pho dans un complexe contenant également la protéine Sfmbt. Le complexe composé de ces deux protéines a été nommé Pho-RC (**Figure** 14A). De façon intéressante, la protéine Sfmbt est conservée chez les mammifères et possède deux domaines intéressants (KLYMENKO *et al.* 2006) (**Figure** 15). (1) Le premier domaine consiste en une répétition de 4 MBT (*Malignant Brain Tumor*) pouvant se fixer aux modifications d'histones suivantes : H3K9 et H4K20 mono et di-méthylées (GRIMM *et al.* 2009). Le domaine 4MBT contient deux sites distincts : un site de fixation aux modifications d'histones et un site permettant son interaction avec la protéine Pho (ALFIERI *et al.* 2013). Par ailleurs, une

mutation de ce dernier conduit en l'abolition de l'interaction Sfmbt-Pho et à la dérépression de certains gènes cibles du PcG (ALFIERI *et al.* 2013) indiquant un lien entre le complexe Pho-RC et PRC1. (2) Le second domaine intéressant est le domaine SAM, retrouvé notamment dans les protéines Ph et Scm (voir plus haut). Par ailleurs, les domaines respectifs de Ph et Sfmbt peuvent interagir *in vitro* (GRIMM *et al.* 2009), suggérant que Pho-RC pourrait recruter PRC1 via l'interaction Sfmbt-Ph (FREY *et al.* 2016). De façon intéressante, une étude récente a caractérisé un complexe similaire à Pho-RC chez les mammifères composé de la protéine SFMBT2 et d'une protéine contenant un domaine *zinc finger* comparable à celui de Pho : CASZ1. Cependant, le rôle de ce complexe chez les mammifères n'a pas encore été étudié (LECONA *et al.* 2013).

D. Le complexe PR-DUB (*Polycomb repressive deubiquitinase*)

Comme évoqué plus haut, la E3-ubiquitine-ligase dRING est retrouvée dans deux complexes PRC1 ; le complexe PRC1 et dRAF. Le résultat d'un screen génétique a identifié le complexe PR-DUB (**Figure** 14A), impliqué dans la dé-ubiquitination de H2A, fait partie du système Polycomb (GAYTAN DE AYALA ALONSO *et al.* 2007). Ce complexe est composé de l'ubiquitine hydrolase Calypso et de la protéine Asx (*Additional Sex Combs*). Par ailleurs, ce complexe est également conservé chez les mammifères (SCHEUERMANN *et al.* 2010; HAURI *et al.* 2016). De façon surprenante, l'absence du site catalytique de Calypso ou l'absence de Asx conduit en une augmentation globale de H2AUb, associée à une dérepression des gènes HOX (SCHEUERMANN *et al.* 2010). Ces résultats suggèrent qu'une régulation fine de la présence de H2AUb (activateur dRING et répresseur PR-DUB) est nécessaire à la répression par le système PcG (SCHUETTENGRUBER and CAVALLI 2010; SCHEUERMANN *et al.* 2012).

E. Le complexe PRC2 (*Polycomb Repressive Complex 2*)

i) Les protéines cœur du PRC2

Le complexe PRC2 a été purifié et identifié chez la Drosophile et les mammifères en 2002 (Pour la Drosophile : (KUZMICHEV *et al.* 2002; MULLER *et al.* 2002); Pour les mammifères : (CAO *et al.* 2002). Il est composé de l'histone méthyl-transférase E(z) (*Enhancer of zeste*) (**Figure** 16), dont l'activité Polycomb a été découverte plus tôt à

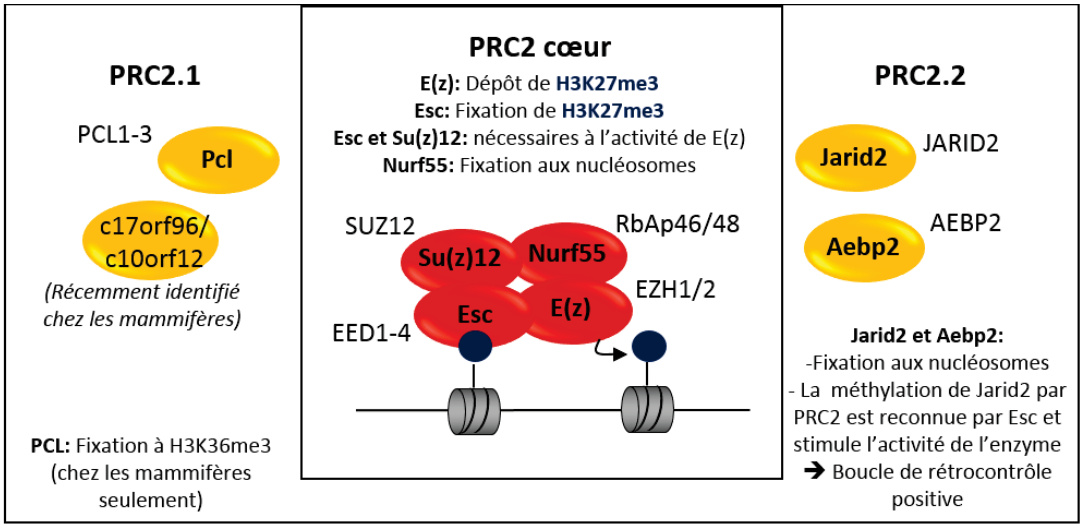


Figure 16 : Le complexe PRC2 et ses protéines accessoires.
Les protéines retrouvées chez la Drosophile sont indiquées dans les cercles. Leurs homologues chez les mammifères sont indiqués pour chacune des sous-unités. L'association des protéines cœur avec certaines protéines accessoires spécifie les sous-complexes PRC2.1 et PRC2.2.

partir d'études génétiques (LEWIS 1978). En effet, la perte de fonction de la protéine provoque un phénotype de dérépression des gènes homéotiques (JONES and GELBART 1990), caractéristique des mutants des protéines du PcG. Chez les mammifères, E(z) possède deux homologues : EZH1 et EZH2. Ces derniers sont mutuellement exclusifs et exprimés respectivement dans les cellules en prolifération (EZH2) et les cellules quiescentes (EZH1) (MARGUERON *et al.* 2008; SHEN *et al.* 2008). Par ailleurs, E(z) contient un domaine SET conservé (*Su(var)3-9, enhancer of zeste (EZ) and Trithorax*) (JONES and GELBART 1990; LAIBLE *et al.* 1997; DILLON *et al.* 2005) permettant la méthylation de la lysine 27 de l'histone H3 (H3K27me3) (CAO *et al.* 2002; KUZMICHEV *et al.* 2002). Plus important encore, E(z) seule ne possède pas d'activité HMT (CAO *et al.* 2002; KETEL *et al.* 2005; NEKRASOV *et al.* 2005). Ainsi, l'intégrité du complexe PRC2 est cruciale pour cette activité (**Figure** 17). L'activité de E(z) requière son association avec deux autres protéines dont Esc et Su(z)12.

Chez la Drosophile, la protéine Esc (*Extra Sex Combs*) (**Figure** 16) possède un homologue fonctionnel Escl (*Esc-like*). Tout comme Psc et Su(z)2 pour PRC1, les doubles mutants pour Esc/Escl montrent une sévère transformation phénotypique, propre aux mutants du PcG (KETEL *et al.* 2005; WANG *et al.* 2006b; KURZHALS *et al.* 2008; OHNO *et al.* 2008). Cependant, les deux protéines ne semblent pas être exprimées au même stade du développement. Le pique d'expression de Esc se produit lors de l'embryogénèse alors que celui de Escl se situe au cours du stade post-embryonnaire et persiste chez l'adulte (WANG *et al.* 2006b; KURZHALS *et al.* 2008). De façon importante, Esc et ses

orthologues mammaliens EED (*Embryonic Ectoderm Development*) possèdent un domaine WD40 leur permettant de reconnaitre la marque H3K27me3 (HANSEN *et al.* 2008; MARGUERON *et al.* 2009), ce qui permet le maintien de la marque lors de la réplication.

La troisième sous-unité importante de PRC2 est Su(z)12 (*Suppressor of zeste 12*), et son homologue mammalien SUZ12 (**Figure** 16). Chez la Drosophile, les mutants Su(z)12 montrent également des transformations phénotypiques propres aux mutants du PcG (BIRVE *et al.* 2001).

Su(z)12 travaille en tandem avec Nurf55 (encore appelé p55 ou CAF-1), la dernière sous-unité faisant partie du cœur de PRC2 (CAO *et al.* 2002; CZERMIN *et al.* 2002; KUZMICHEV *et al.* 2002) (**Figure** 16). Nurf55 est également conservée chez les mammifères sous le nom de RbAp46/48 (KUZMICHEV *et al.* 2002; ANDERSON *et al.* 2011).

ii) Les protéines accessoires du PRC2

Les protéines cœur du PRC2 peuvent s'associer à d'autres sous-unités avec une stœchiométrie moindre. Ces cofacteurs sont sensibles à la structure chromatinienne et contribuent donc à la régulation fine du complexe PRC2.

Pcl (*Polycomb-like*) (**Figure** 16) a été identifiée comme une protéine membre du système Polycomb sur la base d'observations phénotypiques chez des mouches mutantes (DUNCAN 1982) puis par sa colocalisation avec les autres protéines du Polycomb sur des chromosomes polytènes (LONIE *et al.* 1994). De façon importante, cette protéine semblerait former une forme spécifique du complexe PRC2 (O'CONNELL *et al.* 2001; TIE *et al.* 2003; NEKRASOV *et al.* 2007; SAVLA *et al.* 2008) nommé Pcl-PRC2 chez la Drosophile (NEKRASOV *et al.* 2007) et PRC2.1 chez les mammifères (HAURI *et al.* 2016) (**Figure** 16). Pcl possède deux domaines intéressants qui sont conservés chez ses trois orthologues mammaliens : PCL1-3 (COULSON *et al.* 1998). Le premier est un tandem de deux PHD (*plant homeodomain*) *fingers* qui lui permettent d'interagir directement avec E(z) (O'CONNELL *et al.* 2001). Cette association permet l'action de Pcl comme un cofacteur qui stimule spécifiquement la tri-méthylation par E(z) (NEKRASOV *et al.* 2007; CAO *et al.* 2008; SARMA *et al.* 2008). Par ailleurs, il a été suggéré que Pcl jouerait un rôle dans le recrutement de PRC2 chez la Drosophile (SAVLA *et al.* 2008). Cependant, comme toutes les protéines du système PcG, cette protéine ne possède pas de domaine de liaison à l'ADN suggérant que l'association à la chromatine est

indirecte. L'étude mécanistique du recrutement de PRC2 via Pcl a été mieux caractérisée chez les mammifères (ABED and JONES 2012; BALLARE *et al.* 2012; MUSSELMAN *et al.* 2012a; BRIEN *et al.* 2015). En effet, le second domaine important de la protéine PCL est le domaine TUDOR. Ce dernier forme une cage aromatique lui permettant de reconnaitre spécifiquement la marque H3K36me3 associée à l'activité transcriptionnelle (MUSSELMAN *et al.* 2012a). Ainsi, PCL permet le recrutement de PRC2 aux sites actifs ainsi que le recrutement de déméthylases, en particulier NO66 dans les cellules ES (BRIEN *et al.* 2012). Ainsi, bien que PCL ne soit pas nécessaire au renouvellement des cellules souches, son action est nécessaire à la répression *de novo* des gènes embryonnaires pour permettre la différenciation cellulaire (BALLARE *et al.* 2012; BRIEN *et al.* 2012). Contrairement à ses homologues mammalien, Pcl des Drosophiles ne possède pas les résidus aromatiques permettant la reconnaissance de H3K36me3 (BALLARE *et al.* 2012; MUSSELMAN *et al.* 2012a). Ainsi, le recrutement de PRC2 par Pcl chez la Drosophile est probablement accompli par un mécanisme différent (FRIBERG *et al.* 2010). De façon intéressante, une étude récente de spectrométrie de masse révèle que PRC2 forme deux complexes bien distincts selon la nature des protéines accessoires avec qui il est associé (HAURI *et al.* 2016). Ainsi, le PRC2.1 est exclusivement associé à PCL mais également à deux protéines C10ORF12 et C17ORF96 impliquées dans la régulation de l'activité PRC2 de manière site-spécifique (GRIJZENHOUT *et al.* 2016; HAURI *et al.* 2016) (**Figure** 16).

De la même manière, Hauri et al., ont défini un complexe PRC2.2 caractérisé par la présence des protéines accessoires JARID2 et AEBP2 (HAURI *et al.* 2016) (**Figure** 16). La protéine JARID2 appartient à la famille de protéines déméthylases Jumonji (KLOSE *et al.* 2006; TSUKADA *et al.* 2006). Contrairement aux autres membres de la famille Jumonji, le domaine catalytique de JARID2 est inactif (PENG *et al.* 2009; SHEN *et al.* 2009; LANDEIRA *et al.* 2010; LI *et al.* 2010a; PASINI *et al.* 2010). Cependant, JARID2 et PRC2 ciblent les mêmes sites de la chromatine des cellules ES. En effet, il a été démontré que JARID2 pouvait interagir avec EZ (Li, 2010) et SUZ12 (PENG *et al.* 2009). Par ailleurs, JARID2 possède un domaine de fixation aux nucléosomes qui permet la stabilisation du complexe PRC2 à la chromatine (SON *et al.* 2013). Dans ce sens, la déplétion de JARID2 résulte en la diminution de l'enrichissement en PRC2 à la chromatine (PASINI *et al.* 2010), soutenant l'idée que JARID2 serait impliqué dans le recrutement de PRC2. Néanmoins, le rôle de JARID2 sur l'activité HMTase de PRC2 a longtemps été ambigu. Alors que certaines études montrent une activité positive de JARID2 sur PRC2 (LI *et al.* 2010a; PASINI *et al.* 2010), d'autres montrent que la déplétion de JARID2 réduit (quoique faiblement) l'enrichissement de H3K27me3 (PENG *et al.* 2009; SHEN *et al.* 2009; ZHANG *et al.* 2011; KANEKO *et al.* 2014) suggérant que JARID2 pourrait inhiber l'activité enzymatique de PRC2. Une étude récente a cependant apporté un élément de réponse à cette ambiguïté. Sanulli et al., ont identifié que la lysine située en position 116 de JARID2 (et qui est par ailleurs conservée chez la Drosophile) est méthylée par PRC2.

Cette modification est reconnue par EED ce qui confère un changement allostérique de PRC2 et une stimulation de son activité enzymatique (SANULLI *et al.* 2015). De façon importante, cette étude démontre également le rôle crucial de cette modification dans la déposition de H3K27me3 durant la différenciation cellulaire. La deuxième protéine accessoire propre à PRC2.2 est AEBP2 (Jing chez la Drosophile) (CAO and ZHANG 2004; HAURI *et al.* 2016) (**Figure** 16). Tout comme son partenaire JARID2, AEBP2 est aussi un activateur allostérique de l'activité HMTase de PRC2 (CAO and ZHANG 2004; CIFERRI *et al.* 2012). De façon importante, AEBP2 contient un domaine *zinc finger* lui permettant de se fixer à l'ADN (HE *et al.* 1999). De plus, une étude de sa cartographie montre une importante colocalisation entre cette protéine et PRC2 chez le murin. Finalement, il a récemment été montré que la présence de la marque H2AUb déposée par PRC1 formait un substrat pour la fixation du complexe PRC2 quand ce dernier est associé à JARID2/AEBP2 (KALB *et al.* 2014; COOPER *et al.* 2016). De plus, la stimulation de l'activité HMTase de PRC2 est d'autant plus stimulée par AEBP2 en présence de JARID2 et de H2AUb (CAO and ZHANG 2004). Ceci caractérise une boucle de régulation positive permettant l'établissement des larges domaines répressifs marqués par H3K27me3.

IV. Mécanismes de régulation de la transcription liés à l'activité des protéines du PcG

Les protéines du PcG interviennent dans la répression de leurs gènes cibles à différentes échelles. Soit de manière linéaire le long de l'ADN, soit de manière tridimensionnelle, créant ainsi des foyers nucléaires marqués par la présence des complexes du PcG. De façon importante, des études récentes révèlent un rôle inattendu de ces protéines dans l'activation transcriptionnelle

A. Mécanismes impliquant les protéines du PcG dans la répression transcriptionnelle

i) Répression des gènes cibles de manière linéaire

La répression par les protéines du PcG implique la formation de larges domaines marqués par H3K27me3 (SCHUETTENGRUBER *et al.* 2009) (**Figure** 17).

En effet, cette marque peut directement empêcher la déposition de sa marque antagoniste H3K27ac et interférer dans le recrutement de la Pol II (CHOPRA *et al.* 2011). La déposition de H2AUb par PRC1 peut également interférer avec les processus de régulation de la transcription. En effet, il a été proposé que cette marque contribuait au maintien de la répression transcriptionnelle en empêchant le recrutement de la Pol II (STOCK *et al.* 2007; ENDOH *et al.* 2008; ENDOH *et al.* 2012) ou en bloquant son activité (ZHOU *et al.* 2008). Sa présence inhiberait également la tri-méthylation de l'histone H3K4 (DE NAPOLES *et al.* 2004; ENDOH *et al.* 2012). Cependant, le rôle de H2AUb dans la répression transcriptionnelle par le système Polycomb est encore discuté. En effet, des études plus récentes indiquent que chez la Drosophile, seule une partie des gènes cibles du PcG requière en réalité la présence de la marque H2AUb (GUTIERREZ *et al.* 2012). La répression de cette classe de gènes semble être étroitement liée avec la présence de PR-DUB, suggérant que cette répression requière une fine balance du niveau de H2AUb.

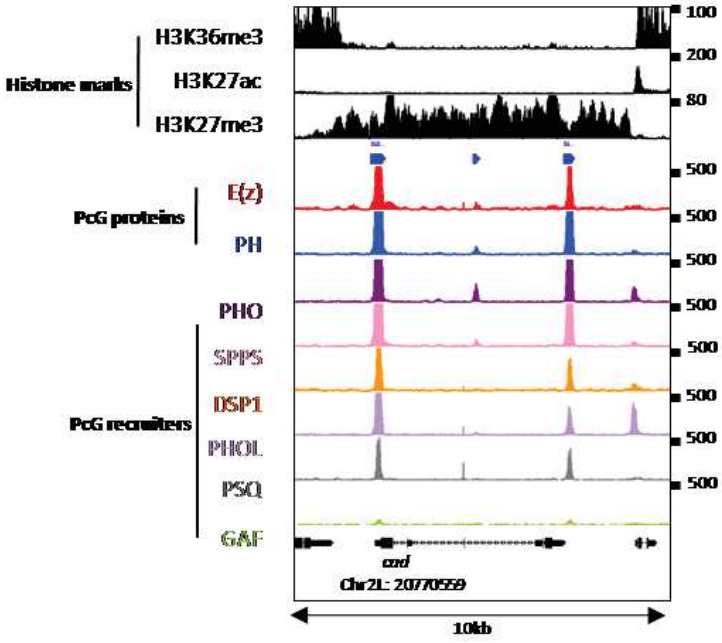


Figure 17 : Les complexes du PcG forment de large domaines répressifs marqués par H3K27me3

Capture d'écran réalisée à partir de données de ChIP-seq de différentes protéines et de modifications d'histones dans les cellules S2 de Drosophile (Voir Résultats). La figure indique le domaine PcG formé autour du gène *cad* (*cadum*) qui possède trois sites de fixations aux protéines du PcG (indiqués par des flèches bleues). On remarque une forte corrélation entre les protéines du PcG et leurs recruteurs.

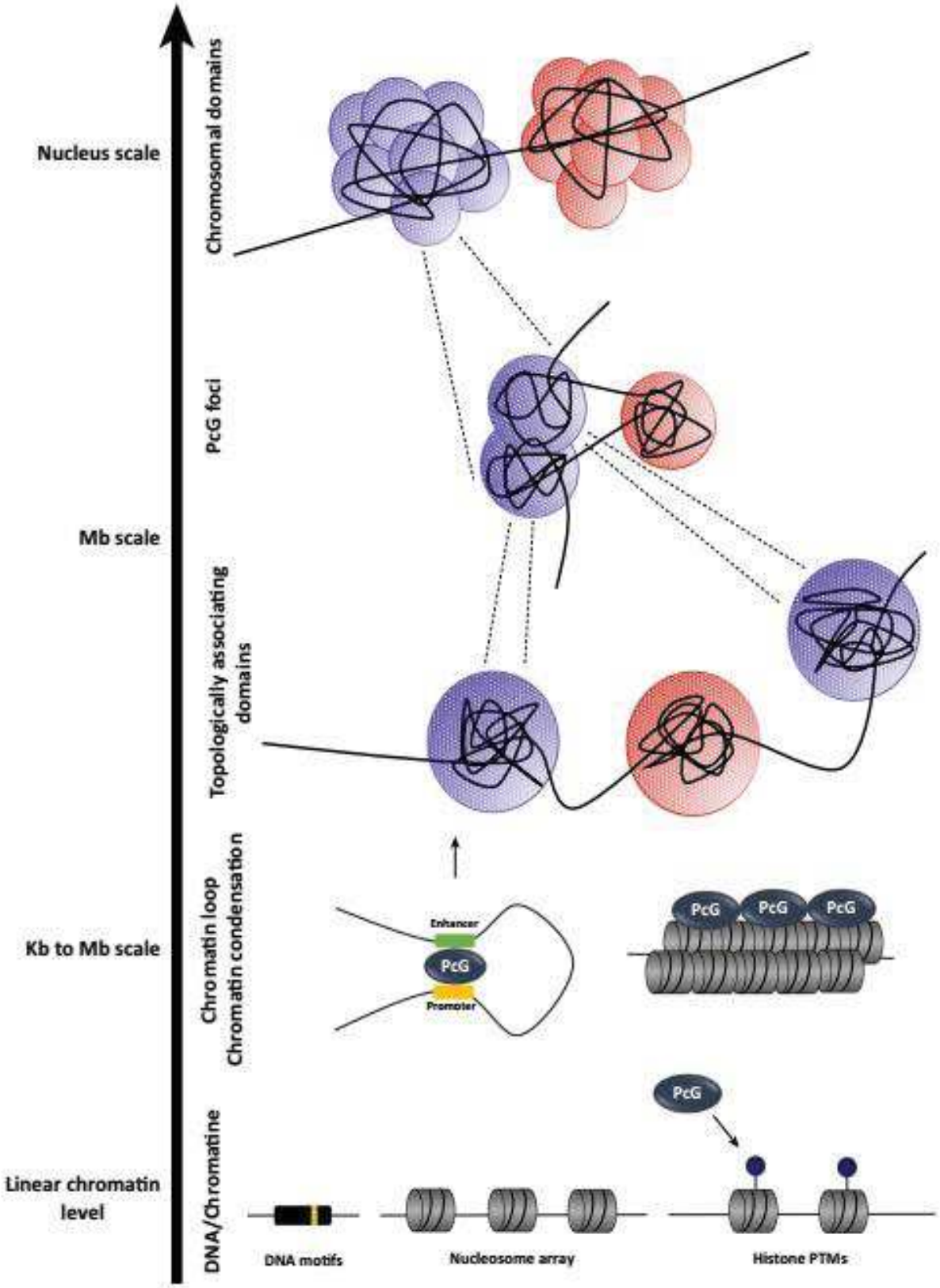
De plus, la cartographie de RING1B et de H2AUb dans des cellules ES de mammifères montre que seule la sous-unité RING1B est nécessaire à la répression des gènes HOX (probablement dû à une activité de la compaction de la chromatine (ESKELAND *et al.* 2010)) ; alors que H2AUb n'est pas nécessaire à la répression des gènes HOX (ENDOH *et al.* 2012). Ce dernier point est également appuyé par un papier récent visant à discriminer entre le rôle de dRING et l'ubiquitination de H2A chez la Drosophile (PENGELLY *et al.* 2015). Cette étude montre que seule H2AUb est essentielle pour la viabilité. En effet, les Drosophiles mutantes pour les histones H2A (H2A et H2Av) dont les lysines pouvant être ubiquitinées sont remplacées par des arginines ne sont pas viables. Cependant, les gènes HOX restent réprimés (maintien de H3K27me3) et les embryons ne montrent pas de phénotypes homéotiques comme les mutants du PcG. Ainsi, dRING est important pour le maintien de l'activité tri-méthylase de PRC2 mais pas H2AUb.

Récemment, une autre propriété liée à la protéine Pc, a été identifiée. Il s'agit de l'inhibition du dépôt de la marque antagoniste du PcG : H3K27ac (TIE *et al.* 2016). En effet, cette dernière est déposée par la protéine CBP (*CREB-binding protein*) chez la Drosophile et p300 chez les mammifères (TIE *et al.* 2009; TIE *et al.* 2016). L'activité optimale de CBP requière une auto-acétylation en *trans* du fragment AIL inclut dans le domaine catalytique HAT de l'enzyme (THOMPSON *et al.* 2004; ARIF *et al.* 2007; DELVECCHIO *et al.* 2013). La protéine Pc contient une petite région conservée, chez ses orthologues mammalien, dont le cœur KRG (**Figure** 15) peut interagir avec le domaine AIL de CBP et empêcher son auto-acétylation (TIE *et al.* 2016) ce qui inhibe l'acétylation de H3K27. Un autre rôle de PRC1 dans la répression transcriptionnelle a déjà été évoqué lors de la description de ce complexe. En effet, la protéine Psc possède une capacité intrinsèque à compacter les nucléosomes (FRANCIS *et al.* 2001) (**Figure** 18). De la même manière, CBX2 chez les mammifères possède une activité intrinsèque de compaction des nucléosomes (LAU *et al.* 2017). L'ensemble de ces résultats suggère que PRC1 peut initier la compaction de la chromatine de manière locale pour induire l'établissement et la propagation de H3K27me3 afin de permettre la formation des domaines répressifs.

Enfin, il est à noter que les protéines du PcG jouent également un rôle plus global dans la répression des gènes. En effet, la principale activité du PRC2 est de di-méthyler H3K27 (FERRARI *et al.* 2014). La marque H3K27me2 représente environ 70% des modifications de H3K27 alors que H3K27me3, est bien moins abondante et dépend fortement du type cellulaire (PETERS *et al.* 2003; EBERT *et al.* 2004; JUNG *et al.* 2010; FERRARI *et al.* 2014). Ainsi, le rôle répressif du PRC2 est bien plus global qu'attendu puisque H3K27me2 prévient de manière globale l'expression ectopique des gènes mais limite également l'expression de gènes dont l'activité transcriptionnelle est forte (LEE *et al.* 2015).

Figure 18 : Régulation de la transcription par les protéines du PcG à différentes échelles.

(Ci-après) Au niveau linéaire, les protéines du PcG sont recrutées au niveau de régions spécifiques et induisent une compaction de la chromatine ainsi que la déposition de la marque répressive H3K27me3. Les protéines du PcG peuvent également créer des boucles d'interactions entre des éléments distaux (ici enhancer en vert et promoteur en jaune). A plus grande échelle, les régions génomiques fixées par les protéines du PcG interagissent fortement et créent des TAD (*Topologycally Associating Domains*). Ces TAD peuvent également interagir entre eux pour former des foyers nucléaires visibles au microscope. En rouge des TAD actifs qui ségrégent de façon différentielle des TAD associés à l'activité du PcG. Figure tirée de (ENTREVAN *et al.* 2016).


Nucleus scale
Mb scale
Kb to Mb scale
Linear chromatin level
Chromosomal domains
PcG foci
Topologically associating domains
Chromatin loop
Chromatin condensation
DNA/Chromatine
Enhancer
PcG
Promoter
PcG
PcG
PcG
PcG
DNA motifs
Nucleosome array
Histone PTMs

ii) Répression des gènes cibles de manière tridimensionnelle (ENTREVAN *et al.* 2016)

En plus de leur effet local sur la chromatine, les protéines du PcG peuvent également créer des interactions à distances avec des régions régulatrices de l'ADN créant ainsi des boucles (entre leurs sites de fixations et les promoteurs, enhancers et insulateurs (LANZUOLO *et al.* 2007; COMET *et al.* 2011)). Ces boucles d'interactions sont visibles en microscopie où l'on observe une accumulation des protéines du PcG au niveau de foyers nucléaires appelés *PcG foci* (HERNANDEZ-MUNOZ et al. 2005; GONZALEZ et al. 2014) (**Figure** 18). En plus d'une interaction en *cis* entre les sites de fixation des protéines du PcG (à l'intérieur d'un même domaine), on observe également des interactions à plus longues distances en *trans* ce qui a pour effet d'accroitre la répression médiée par ces protéines (BANTIGNIES *et al.* 2003; BANTIGNIES and CAVALLI 2011; LI *et al.* 2013; SCHUETTENGRUBER and CAVALLI 2013). Une étude récente réalisée sur les cellules ES a révélé que RING1B jouait un rôle majeur dans l'interaction tridimensionnelle formant un réseau entre les promoteurs de certains gènes HOX (SCHOENFELDER *et al.* 2015). Le retrait de RING1B provoque une perte d'interactions promoteur-promoteur accompagnée d'une activation de ces gènes.

iii) Rôle des protéines du PcG dans l'organisation nucléaire (ENTREVAN *et al.* 2016)

Comme nous l'avons vu précédemment, les protéines du PcG induisent un niveau de complexité supplémentaire qui émerge à plus grande échelle, par la formation de TAD.

Les TAD ne sont pas des structures rigides. L'étude de la régulation des gènes HOX chez les mammifères montre que ces structures sont très dynamiques (NOORDERMEER *et al.* 2011; MALLO and ALONSO 2013; NOORDERMEER *et al.* 2014; VIEUX-ROCHAS *et al.* 2015) (**Figure** 19). Au cours du développement, les gènes HOX sont maintenus réprimés et sont regroupés dans le noyau au sein de TAD marqués par H3K27me3. Au fur et à mesure du développement, les gènes passent des TAD réprimés à des TAD actifs marqués par H3K4me3. La réorganisation des TAD corrèle donc avec le changement des modifications d'histones. Il est cependant impossible de distinguer si la perte de H3K27me3 induit cette réorganisation ou si le changement d'organisation des TAD induit la perte de H3K27me3.

Le complexe PRC1 et plus particulièrement la protéine Ph joue probablement un rôle clef dans la formation des TAD, notamment par son domaine SAM. Ce domaine a préalablement été décrit dans la présentation des protéines du PcG. Ce domaine permet l'oligomérisation de Ph (ISONO *et al.* 2013; WANI *et al.* 2016). Deux études récentes mettent en évidence le rôle fonctionnel de Ph dans la modulation de l'architecture chromatinienne. Boettinger *et al*., ont montré que les TAD associés à la répression par le PcG étaient plus compacts que ceux associés à la chromatine active

ou nulle (BOETTIGER *et al.* 2016). Cette propriété est perdue en absence de Ph. Une étude parallèle révèle que la perturbation des propriétés de polymérisation du domaine SAM de Ph corrèle avec une dispersion des foyers nucléaires et une activation de l'expression génique (WANI *et al.* 2016). Ces résultats suggèrent que l'organisation des protéines du PcG en foyers nucléaires est médiée par le domaine SAM de Ph.

.

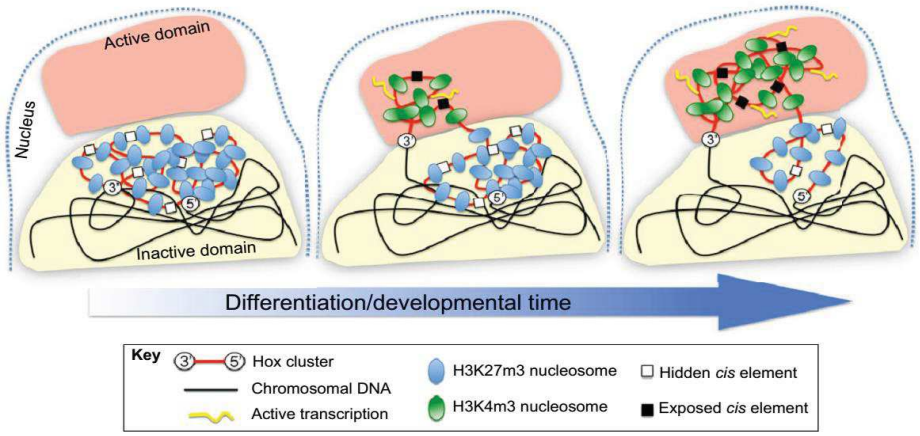


Figure 19 : Dynamique de l'organisation chromatinienne au cours du développement

Ce schéma, tiré de (MALLO and ALONSO 2013), montre le passage des gènes d'un TAD réprimé marqué par H3K27me3 à un TAD actif marqué par H3K4me3.

B. Mécanismes impliquant les protéines du PcG dans l'activation transcriptionnelle (ENTREVAN *et al.* 2016)

Bien que les complexes du PcG régulent leurs gènes cibles en les réprimant, des études récentes ont révélées un rôle inattendu de ces complexes dans l'activation transcriptionnelle. Les études de ChIP-seq des protéines du PcG ont tout d'abord révélées leur présence sur des promoteurs actifs (BROOKES *et al.* 2012; MOUSAVI *et al.* 2012; FRANGINI *et al.* 2013; KANEKO *et al.* 2013; VAN DEN BOOM *et al.* 2016). Par la suite, différents mécanismes conduisant en un changement de l'état transcriptionnel ont été identifiés. Premièrement, les complexes du PcG peuvent intervertir des sous-unités chez les mammifères de manière spécifique pour induire l'activation ou la répression transcriptionnelle selon les cas. Un exemple concret montre qu'au cours de la différenciation des cellules ES en cellules ectodermiques, la substitution de la sous-unité CBX7 par CBX8 de PRC1 est requise pour initier l'activation de certains gènes développementaux (CREPPE *et al.* 2014). De la même manière au cours de la

différentiation des cellules ES en précurseurs cardiaques, le complexe PRC1-Mel12 échange sa sous-unité CBX7 par CBX2 pour induire l'activation de gènes impliqués dans le développement du système cardio-vasculaire (MOREY *et al.* 2015). Deuxièmement, l'association des complexes du PcG avec d'autres protéines régulatrices peut également permettre l'activation des gènes. Notamment, l'association du complexe PRC1-5 à la protéine Auts2 conduit à l'activation transcriptionnelle des gènes cibles via le recrutement de l'HAT p300 (GAO *et al.* 2014). Une étude récente démontre également que la modification post-traductionnelle des protéines du PcG peut modifier leur activité. En effet, le complexe PRC1-5 agit comme un activateur transcriptionnel après la phosphorylation de RING1B par la protéine CK2 (GAO *et al.* 2014). Enfin, les protéines du PcG peuvent contribuer à l'activation transcriptionnelle en régulant la conformation topologique de séquences régulatrices en *cis*. L'activation spécifique de *Meis2* au cours du développement du cerveau requière une étape d'interaction transitoire entre le promoteur de *Meis2* et un enhancer de façon dépendante de RING1B (KONDO *et al.* 2016). En absence de RING1B cette interaction n'a pas lieu résultant en l'altération de l'activation génique.

Chez la Drosophile, une étude récente montre que les sites de fixations des protéines du PcG peuvent également agir comme enhancers *in vivo*, activant la transcription dans certaines cellules (rôle enhancer) alors que dans d'autres cellules, la transcription est maintenue réprimée (activité Polycomb) (ERCEG *et al.* 2017). L'utilisation du même élément chromatinien pour l'activation/répression spécifique permettrait une régulation fine de l'expression des gènes de manière spatio-temporelle.

Chapitre III : Fonctions biologiques du PcG

Les chapitres précédents ont évoqué le rôle des complexes du PcG dans le maintien de la répression des gènes HOX au cours du développement. Les mécanismes permettant le maintien de l'état transcriptionnel de leurs gènes cibles ont été décrits ainsi que le rôle très spécifique de ces complexes dans l'activation de certains gènes. Le rôle de ces complexes est donc très large et ils jouent un rôle crucial au cours de la différentiation cellulaire. Leur dérégulation propre ou celle de leurs gènes cibles conduit en l'apparition de cancers.

I. Rôle différentiel et dynamique des complexes du PcG lors de la différentiation

Chez les mammifères, les études de ChIP-seq réalisées sur des cellules ES et différenciées montrent que les complexes du PcG jouent un rôle clef dans l'identité cellulaire et la maintenance des programmes de différenciation au cours du développement (Morey *et al.* 2012; Morey *et al.* 2015). En particulier, dans les cellules ES, les promoteurs cibles du PcG sont marqués par H3K27me3 mais également par la marque active H3K4me3 déposée par le TrxG. Ces gènes sont dits « bivalents ». Ils sont en effet en état de pause transcriptionnelle. Suivant les signaux développementaux, ils peuvent soit être activés soit réprimés, et la présence des deux marques antagonistes rend cette transition plus rapide (Bernstein *et al.* 2006). Par ailleurs, les gènes cibles de PcG dans les cellules ES correspondent principalement à des gènes de différenciation (**Figure** 20). A contrario, ce sont les gènes de renouvellement cellulaires qui deviennent réprimés dans les cellules différenciées (**Figure** 20). Ceci indique qu'une redistribution des protéines du PcG a lieu au cours du développement. Ce phénomène a récemment été démontré aussi chez la Drosophile (Loubiere *et al.* 2016). Chez les mammifères, le ciblage dynamique des gènes cibles est notamment permis grâce aux protéines CBX. Par exemple, la protéine CBX7 est la plus abondante des protéines CBX dans les cellules ES. Pour que les cellules se différencient, CBX7 doit être remplacée par d'autres protéines de la famille comme CBX 2/4 ou 8. L'échec de ce remplacement résulte en une prolifération anormale des cellules souches hématopoïétiques et donc en leucémie (Klauke *et al.* 2013). De la même manière, les protéines PCGF confèrent une fonction biochimique spécifique aux complexes PRC1 associés qui contribue à la régulation de l'identité et de la destinée cellulaire (Zdzieblo *et al.* 2014; Morey *et al.* 2015; Kloet *et al.* 2016; Endoh *et al.* 2017). Par exemple, la présence de Mel18 (PCGF2) au sein du complexe PRC1 permet la permutation des autres sous-unités du complexe de manière

développement-spécifique pour promouvoir spécifiquement la différentiation des cellules cardiaques (MOREY *et al.* 2015).

La dynamique du complexe PRC2 lors du développement est relativement plus simple puisque seules les sous-unités EZH1 et EZH2 sont interverties lors de la différenciation (SHEN *et al.* 2008). C'est plutôt l'association du PRC2 avec les protéines accessoires qui permet un recrutement différentiel du PRC2 (GRIJZENHOUT *et al.* 2016; HAURI *et al.* 2016; KLOET *et al.* 2016; OLIVIERO *et al.* 2016). JARID2 est en effet plus fortement exprimé dans les cellules ES alors AEBP2 est plus présent dans les cellules différenciées.

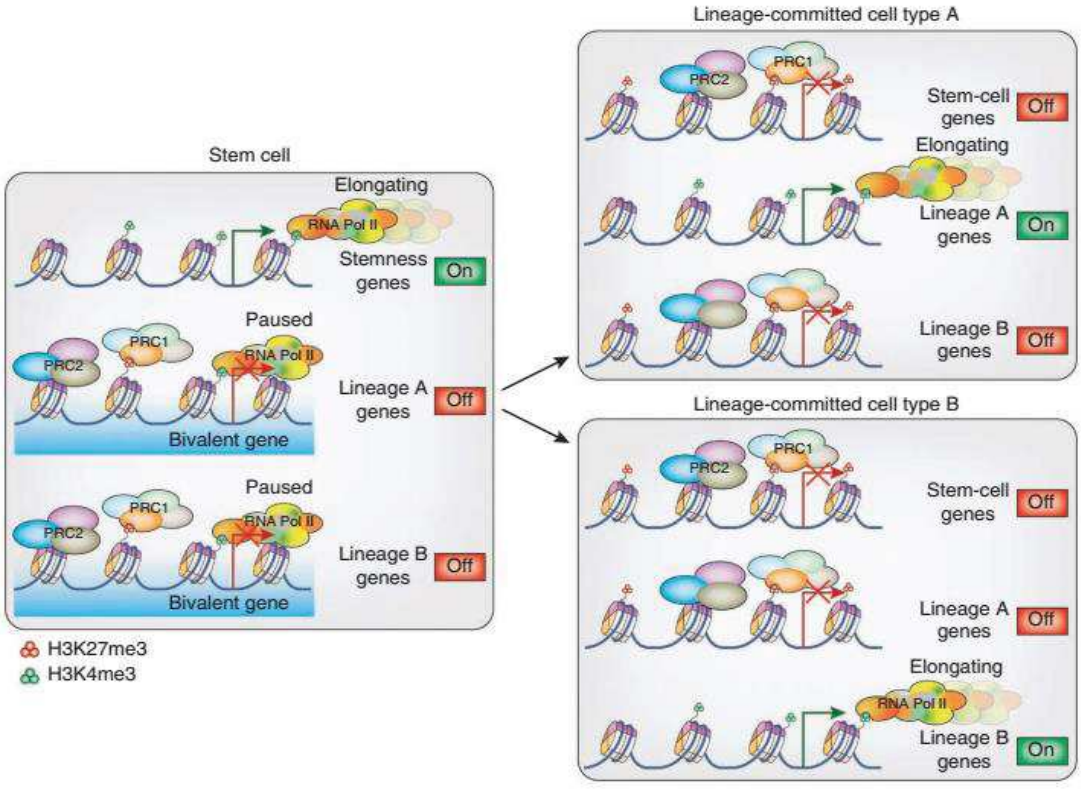


Figure 20 : Rôle des protéines du PcG lors de la différentiation cellulaire
La figure illustre le rôle des complexes du PcG dans la spéciation cellulaire. Les cellules souches ont la capacité de se différencier en tous les types cellulaires formant un organisme (simplifié ici par les types cellulaires A et B). Avant la différentiation, les gènes impliqués dans le maintien de la pluripotence sont activés alors que ceux spécifiques de lignées développementales sont inactifs. Durant la différentiation, les complexes du PcG sont expulsés d'un set de gènes propre à une lignée développementale. Illustration de (DI CROCE and HELIN 2013).

II. La dérégulation des gènes cibles du PcG mène à l'apparition de cancers

Le rôle des protéines du PRC1 dans les cancers reflète leur activité *in vivo* et en particulier dans les cellules ES. En effet, CBX7 joue un rôle important dans la prolifération des cellules souches. De façon corrélée, la surexpression de CBX7 dans les cellules différenciées est impliquée dans le développement de divers cancers dont la leucémie. CBX7 agit comme oncogène en réprimant le locus *INK4A/ARF* prévenant ainsi la sénescence et l'apoptose cellulaire. De façon similaire, BMI1 est hautement exprimée dans les cellules souches hématopoïétiques et est ensuite remplacée par son homologue Mel18 lors de la différenciation (MOREY *et al.* 2015). BMI1 agit également comme oncogène en réprimant CDKN2a et PTEN et en régulant la cascade de signalisation AKT (JACOBS *et al.* 1999). A contrario, Mel18 agit comme suppresseur de tumeur en régulant l'expression de nombreux oncogènes comme *c-myc* (GUO *et al.* 2007). Chez la Drosophile, PRC1 semble agir comme suppresseur de tumeur. En effet, la délétion de PRC1 conduit en l'apparition de tumeur alors que ce n'est pas le cas pour PRC2 (LOUBIERE *et al.* 2016).

Le rôle de PRC2 dans la formation des cancers est plutôt ambigu. En effet, EZH2 a été initialement décrit comme oncogène. En effet, la surexpression de EZH2 associée à une augmentation de H3K27me3 a été décrit comme marqueur de cancer de la prostate métastatique dont le diagnostic est défavorable pour la vie des patients (VARAMBALLY *et al.* 2002). De la même manière, des mutations du domaine catalytique de EZH2 sont impliquées dans l'apparition de lymphomes (MORIN *et al.* 2010; SNEERINGER *et al.* 2010). Basés sur ces observations, plusieurs molécules visant à inhiber l'activité de EZH2 ont été développées et sont testées cliniquement ou déjà utilisées (MORERA *et al.* 2016). Cependant, d'autres études montrent que PRC2 agit également comme suppresseur de tumeur. En effet, les mutations perte de fonction de EZH2, EED ou SUZ12 sont impliquées dans la formation de tumeurs cérébrales et corrèlent avec une forte diminution de H3K27me3 (LEE *et al.* 2014). Par ailleurs, une mutation hétérozygote ponctuelle de l'histone H3 substituant la lysine 27 en méthionine suffit à diminuer de manière globale le taux de H3K27me3 en expulsant PRC2 de ses gènes cibles (LEWIS *et al.* 2013; PIUNTI *et al.* 2017). Cette mutation est impliquée dans 80% de cancers cérébraux caractérisés par leur importante agressivité.

Comme nous l'avons vu dans ce chapitre, le rôle des complexes du PcG est extrêmement spécifique et dépend fortement du type cellulaire. Une fine balance de leurs activités est requise pour maintenir l'homéostasie des cellules. Il n'est donc pas surprenant de voir que d'une part la dérégulation/mutation d'une des sous-unités conduit en la formation de cancers, et d'autre part qu'un déséquilibre cellulaire en H3K27me3 peut avoir des effets dévastateurs conduisant en la perte de l'identité des

cellules. La compréhension des mécanismes de recrutement et d'action de ces complexes est donc cruciale pour permettre la mise en place de thérapies adaptées à chaque type de cancers liés aux complexes PcG.

Chapitre IV : Complexité du recrutement de Polycomb à l'ADN

Dans la partie précédente, nous avons vu l'importance fonctionnelle des complexes du PcG. Notamment, leur dérégulation, celle de leurs protéines recruteuses ou celle de leurs gènes cibles conduit en l'apparition de cancers agressifs chez l'Homme. Ainsi, il est crucial de comprendre comment ces complexes sont recrutés à la chromatine ainsi que leurs différents modes d'actions.

Chez la Drosophile comme chez les mammifères, les complexes PRC1 et PRC2 ne possèdent pas de sites de fixation à l'ADN. Ces complexes sont donc recrutés à la chromatine via leur interaction avec des protéines recruteuses. Il peut s'agir de facteurs de transcription (ou TF) ou de protéines accessoires (comme c'est le cas pour PRC2). Toutes ces protéines recruteuses sont relativement bien conservées. Cependant, les sites de fixations des complexes du PcG diffèrent entre la Drosophile et les mammifères. En effet, chez la Drosophile, les complexes du PcG sont recrutés au niveau de régions nommées PRE (*Polycomb Response Elements*). Chez les mammifères, seulement cinq PRE ont été identifiés à ce jour, les protéines du PcG étant majoritairement retrouvées aux ilots CpG (décrit plus bas ; il s'agit de séquences d'ADN enrichies en C (Cytosines) et G (Guanines) tous deux séparés par un groupement phosphate). Cependant, malgré une divergence de la nature des sites de fixation des complexes entre la Drosophile et les mammifères, la complexité du recrutement de ces complexes semble être conservée.

I. Les PRE/TRE et leurs caractéristiques

Chez la Drosophile, les protéines du PcG sont recrutées au niveau de régions génomiques nommées PRE (*Polycomb Response Elements*), identifiées en 1993 (SIMON *et al.* 1993). Les PRE sont nécessaires au maintien de la marque H3K27me3 lors de la réplication. En effet, la tri-méthylation de H3K27 des nucléosomes nouvellement incorporés se fait de manière séquence spécifique par le recrutement du PRC2 aux PRE (COLEMAN and STRUHL 2017; LAPRELL *et al.* 2017).

La fonctionnalité de ces séquences peut être testée par une analyse transgénique dont le but est l'introduction, dans le génome de la Drosophile, d'un vecteur contenant le PRE en amont d'un gène rapporteur comme le gène *white* donnant la couleur rouge aux yeux de mouches (dans un fond génétique mutant pour le gène *white* endogène) (**Figure** 22 ; (KASSIS and BROWN 2013)). Un PRE est validité comme tel quand il peut

recruter les complexes du PcG de manière ectopique au sein du transgène ce qui conduit en la répression du gène rapporteur et un éclaircissement des yeux de mouches (**Figure** 22). De plus, nous avons vu précédemment que les PRE pouvaient interagir en *trans*. Ainsi, dans des mouches homozygotes pour un transgène contenant un PRE, les PRE transgéniques interagissent ce qui conduit en une répression plus importante du gène rapporteur qu'à l'état hétérozygote (**Figure** 22). Cet effet est appelé PSS (*Pairing-sensitive silencing*).

De façon importante, les sites de fixation des protéines du PcG et du TrxG sont chevauchants (KOCH 2014; GEISLER and PARO 2015). Ils sont nommés PRE quand ils recrutent les protéines du PcG et participent à la répression des gènes cibles. En revanche, on parlera de TRE (*Trithorax Response Elements*) lorsqu'ils recrutent les protéines du TrxG et participent au maintien de l'activation des gènes. Le *switch* entre l'activité PRE/TRE n'est pas très bien connu mais implique une balance fine entre la présence des complexes du PcG et du TrxG.

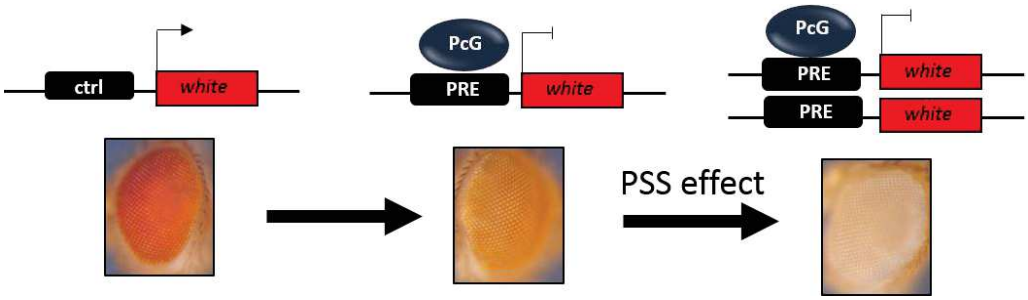


Figure 21 : Test fonctionnel de l'activité PRE.
L'insertion d'un PRE dans un transgène conduit en la répression du gène rapporteur *mini-white* (comparer photo de gauche et celle du milieu). A l'état homozygote pour le PRE transgénique, la répression médiée par les protéines du PcG est plus forte du fait de l'interaction entre les deux PRE transgéniques conduisant à une couleur de l'œil encore plus pâle qu'à l'état hétérozygote (comparer photo de droite avec celle du milieu). Cet effet est appelé PSS pour *Pairing Sensitive Silencing*.

II. Complexité du recrutement des complexes du PcG à différents niveaux

Comme vu dans la partie précédente, la marque H3K27me3 sert de site de fixation au complexe PRC1 via la sous-unité Pc (ou Cbx chez les mammifères). En absence du PRC2, cette modification d'histone est perdue, ainsi que le recrutement de PRC1 (CAO *et al.* 2002; WANG *et al.* 2004b; BOYER *et al.* 2006). Par ailleurs, nous avons vu que seul le FT Pho possédait un site de fixation à l'ADN et qu'il pouvait interagir avec le PRC2. Ces observations ont donné place à un modèle de recrutement hiérarchique (**Figure** 21) pour les complexes du PcG suggérant que Pho se liait à l'ADN permettant le recrutement de PRC2. Dans un deuxième temps, le PRC2 catalyse la tri-méthylation de l'H3K27. PRC1 est ensuite recruté via son interaction avec la modification d'histone. Cependant, les études plus récentes ne sont pas toujours en accord avec ce modèle suggérant l'existence de modes de recrutements alternatifs.

Par ailleurs, tous les PRE identifiés et testés chez la Drosophile ne sont pas identiques au niveau de leur séquence d'ADN. De la même manière, Pho n'est pas le seul FT à recruter le PcG. D'autre FT ont été identifiés comme importants dans le recrutement du PcG. L'ensemble de ces observations nous permet de définir quatre niveaux de complexité dans le recrutement des complexes du PcG.

A. Au niveau de la séquence des PRE

Quelles sont les caractéristiques séquentielles qui spécifient un PRE d'un autre site génomique ?

Deux études bio-informatiques ont été réalisées dans le but de prédire les PRE sur la base de leurs motifs d'ADN (voir la section suivante pour les motifs) (RINGROSE *et al.* 2003; FIEDLER and REHMSMEIER 2006). Cependant, ces analyses ne prédisent que 15-20% des sites génomiques fixés par les protéines du PcG et moins de 20% de ces sites prédits correspondent réellement à des sites fixés par le PcG chez l'embryon (SCHUETTENGRUBER *et al.* 2009). Ceci indique que les PRE ne peuvent pas uniquement être définis sur la base de cluster de motifs et/ou que d'autres informations nous manquent pour pouvoir les prédire. De plus, ces analyses bio-informatiques étaient basées sur des motifs principalement identifiés *in vitro* de manière biochimique.

L'échec de la prédiction des PRE s'explique également par le fait que bien que ces motifs d'ADN recrutant les TF se trouvent effectivement enrichis aux PRE, ils sont également retrouvés au niveau de promoteurs où les protéines du PcG sont absentes

(SCHUETTENGRUBER *et al.* 2009). Par ailleurs, les PRE diffèrent dans le nombre de ces motifs mais également dans leur qualité. En effet, il a été démontré que certains PRE n'étaient enrichis pour aucun motif spécifique (en d'autres mots, ils ne contiennent aucun motif permettant le recrutement de FT) ce qui explique pourquoi les prédictions bio-informatiques ont échoué à les identifier (SCHUETTENGRUBER *et al.* 2009). Ainsi, la manière dont les FT sont recrutés spécifiquement aux PRE pour recruter les complexes du PcG est encore inconnue.

B. Les Facteurs de Transcription (FT) associés au recrutement des protéines du PcG chez la Drosophile

Comme vu précédemment, les PRE sont composés de multiples sites de fixations associés à différents facteurs. En dépit d'un grand nombre d'études visant à mieux comprendre la nature des PRE, nous ne savons pas, à ce jour, ce qui caractérise vraiment un PRE, ni les combinaisons de FT requis pour l'activité d'un PRE. Les études de différents PRE chez la Drosophile ont permis d'identifier différents facteurs importants dans l'activité PRE comme Pho/Phol, Spps, Gaf/Psq, Dsp1, Zeste, Grainyhead et Combgap (KASSIS and BROWN 2013). La cartographie de ces protéines dans différents tissus et à différents stades de développement a permis de mettre en lumière des points importants concernant la distribution de ces protéines (SCHWARTZ *et al.* 2006; SCHUETTENGRUBER *et al.* 2009): 1) Ces protéines ne se fixent pas à tous les PRE et peuvent se fixer à des sites autres que les PRE ; 2) Ces protéines peuvent être retrouvées sur des sites transcriptionnellement actifs. D'autre part, le phénotype des mutants de certains de ces facteurs montre leur rôle dans l'activation par le TrxG plutôt que dans la répression par le PcG suggérant un lien étroit entre l'activité de ces deux complexes antagonistes. De manière générale, il est admis qu'aucun de ces facteurs seul ne peut recruter les complexes du PcG et que le recrutement de ces complexes requière la contribution coopérative de ces différents facteurs. La combinaison de facteur recrutant les complexes du PcG peut varier selon les PRE.

i) Pho (Pleiohomeotic) et Phol (Pleiohomeotic-like)

Le premier FT identifié comme ayant un rôle dans le recrutement des complexes du PcG à la chromatine est la protéine Pho (BROWN *et al.* 1998). De plus, il est le seul membre de cette famille à pouvoir se lier directement à l'ADN. Les mutants pour cette protéine montrent un phénotype classique associé à l'activité du PcG (voir **Figure** 13). Le génome de la Drosophile code une protéine homologue Phol, dont le domaine de fixation à l'ADN est très conservé (BROWN *et al.* 2003). De façon intéressante, les mutants homozygotes pour *Phol* sont viables et ne montrent aucune dérégulation

homéotique. Cependant, des études génétiques indiquent que les protéines Pho et Phol interagissent ensemble dans la répression des gènes HOX (BROWN *et al.* 2003). Cependant, Phol semble avoir un rôle plus large que la répression médiée par les protéines du PcG. En effet, les œufs produit par des femelles mutantes pour *Phol* sont fertiles mais ne se développent pas ; une caractéristique qui n'est pas retrouvée pour les œufs issus de femelles mutantes pour *Pho* (BREEN and DUNCAN 1986).

Pho possède deux domaines conservés dont les domaines *spacer* et *Zinc finger*. Le domaine *Zinc finger* situé en C-terminale lui permet la fixation à l'ADN et lui permet également une interaction avec la protéine Esc du PRC2 (SATIJN *et al.* 2001; WANG *et al.* 2004b). D'autre part, le domaine *spacer* permet l'interaction avec E(z). Bien que les parties C-terminales de Pho et Phol soient conservées (80% d'identité), le domaine *spacer* de Phol ne contient que 45% d'identité avec celui de Pho (WANG *et al.* 2004b) ce qui explique pourquoi Phol peut interagir avec Esc mais pas avec E(z) (WANG *et al.* 2004b). Pho et Phol possèdent des homologues fonctionnels chez les mammifères qui sont respectivement YY1 et YY2 (BROWN *et al.* 1998; BROWN *et al.* 2003; NGUYEN *et al.* 2004; DREWS *et al.* 2009). YY1 a été identifiée comme impliquée à la fois dans l'activation transcriptionnelle que dans la répression médiée par le PcG de manière dépendante du contexte cellulaire (ATCHISON *et al.* 2003; GORDON *et al.* 2006; WANG *et al.* 2006a).

Bien que Pho et Phol jouent un rôle partiellement redondant dans le recrutement des complexes du PcG (WANG *et al.* 2004b), la cartographie de ces deux protéines révèle une divergence dans leurs sites de fixations (SCHUETTENGRUBER *et al.* 2009). En effet, de tous les sites fixés par Ph et Pc, 96% correspondent également à des sites fixés par Pho contre 21% pour Phol. Ceci suggère que contrairement à Phol, Pho pourrait être absolument requis pour le recrutement du complexe PRC1. Par ailleurs, des études *in vivo* et *in vitro* ont identifié un motif d'ADN commun pour Pho et Phol ; le motif GCCAT (FRITSCH *et al.* 1999; BUSTURIA *et al.* 2001; MISHRA *et al.* 2001; WANG *et al.* 2004b). Les résultats des études de cartographie révèlent cependant que Phol se fixe préférentiellement à ce motif d'ADN généralement associé aux TSS, alors que Pho se fixe à un motif plus dégénéré : TxxGCCAT généralement retrouvé aux PRE (KAHN *et al.* 2014; SCHUETTENGRUBER *et al.* 2014). Ainsi, le maintien de Pho aux motifs dégénérés des PRE serait plutôt dû à l'action combinatoire entre Pho et les complexes du PcG qu'à son interaction séquence-spécifique (KAHN *et al.* 2014; SCHUETTENGRUBER *et al.* 2014).

Quel est le rôle précis de Pho aux PRE ? En effet, Pho est retrouvé dans deux autres complexes : Pho-INO80, un complexe de remodelage des nucléosomes et le complexe Pho-RC où il s'associe avec Sfmbt. Alors que seule la protéine Pho est retrouvée associée avec INO80, Pho et Phol peuvent former des complexes stables avec Sfmbt (KLYMENKO *et al.* 2006). Ainsi, un des rôles de Pho et Phol serait de permettre la fixation de Sfmbt aux PRE.

Pour conclure, tous les PRE testés *in vivo r*equièrent la fixation de Pho pour l'activité PRE. Cependant, il a été suggéré que d'autres protéines telles que Gaf et Dsp1 étaient nécessaires pour permettre la fixation de Pho aux PREs. Enfin, bien que le rôle majeur de Pho soit la répression médiée par le PcG, Pho peut également agir comme activateur transcriptionnel dans certaines cellules du système nerveux de la Drosophile (FUJIOKA *et al.* 2008). Il reste donc à savoir si le rôle de Pho dans l'activation transcriptionnelle est spécifique à ces cellules ou s'il peut être généralisé aux PRE/TRE.

ii) Spps (*Sp1 factor for PSS*) et les protéines de la famille Sp1/KLF

En 2005, Brown *et al.*, identifient un motif GGGGCG au niveau du PRE du gène *engrailed* permettant le recrutement *in vitro* de protéines de la famille Sp1/KLF (BROWN *et al.* 2005). Il existe neuf membres de cette famille de protéines chez la Drosophile et huit d'entre elles sont capables de se fixer au PRE de *engrailed in vitro* (BROWN *et al.* 2005). Ces protéines ont la particularité de se fixer à l'ADN via trois domaines *Zinc finger* situés en C-terminale et de posséder en N-terminale, un domaine btd (*buttonhead*) dont le rôle est inconnu (BROWN *et al.* 2005). Cinq ans plus tard, l'équipe publie un travail centré sur SPPS (*Sp1 factor for PSS*), le membre de cette famille qui a le plus d'homologie avec ses orthologues mammaliens (BROWN and KASSIS 2010). Ils montrent, entre autre, que SPPS corrèle encore plus que Pho avec la protéine Psc de PRC1 sur les chromosomes polytènes. Cependant, aucune cartographie de la protéine Spps n'a encore été réalisée (la première a en réalité été faite dans le cadre de mes travaux de thèse ; voir la partie Résultats).

De manière surprenante, bien que les mutants *Spps* soient létaux, ils ne montrent cependant aucune dérégulation des gènes homéotiques suggérant l'existence potentielle de protéines redondantes (BROWN and KASSIS 2010). Par ailleurs, Spps et Pho interagissent dans la répression médiée par le PcG puisque les doubles mutants pour *Spps* et *Pho* montrent une dérégulation des gènes HOX encore plus importante que pour les mutants *Pho* seuls. Cette interaction est conservée chez les mammifères entre YY1 et Sp1, l'homologue de Spps (LEE *et al.* 1993; SETO *et al.* 1993).

Le rôle des autres membres de la famille Sp1/KLF chez la Drosophile dans la répression par les protéines du PcG n'a pas été étudié. Cependant, chez les mammifères, les protéines de cette famille sont impliquées dans plusieurs aspects de l'expression des gènes (KACZYNSKI *et al.* 2003; MCCONNELL and YANG 2010).

iii) GAF (GAGA factor) et Psq (Pipsqueak)

La séquence GAGAG est requise pour l'activité de beaucoup de PRE (KASSIS 2002; FUJIOKA *et al.* 2008). Ce motif est reconnu par deux membres de la famille de protéines BTB-POZ (BTB pour *BR-C, ttk and bab* et POZ pour *Pox virus and Zinc finger*) : GAF et Psq.

GAF est codé par le gène *Trithorax-like*. Le gène fût ainsi nommé en raison du rôle initialement identifié pour GAF à savoir une composante du complexe du TrxG, l'antagoniste du PcG. Cependant, la mutation de son motifs de fixation dans des PRE transgéniques abolie le recrutement du PcG (BUSTURIA *et al.* 2001; MISHRA *et al.* 2001). Par ailleurs, GAF permet de remodeler les nucléosomes aux PRE (en particulier de les retirer) (MAHMOUDI *et al.* 2003; FUDA *et al.* 2015) . De ce fait, il a été suggéré que le recrutement initial de GAF facilitait la fixation de Pho (MAHMOUDI *et al.* 2003). Le lien entre Psq et les protéines du PcG a tout d'abord été mis en lumière par des expériences biochimiques (HUANG and CHANG 2004). Par la suite, une interaction génétique entre *Psq* et *Pc* a été identifiée sur la base du phénotype obtenu dont les gènes HOX sont plus dérégulés qu'un mutant hétérozygote pour *Pc* (HUANG *et al.* 2002).

Bien que les deux protéines colocalisent sur des chromosomes polytènes (SCHWENDEMANN and LEHMANN 2002) leur domaine de liaison à l'ADN est différent. GAF contient un domaine *Zinc finger* (LEHMANN *et al.* 1998) alors que Psq contient un domaine Psq (4 répétitions d'un motif de 50 acides-aminés (HOROWITZ and BERG 1996)). Ainsi, il est possible que le recrutement de Psq requière un motif légèrement plus long (SIEGMUND and LEHMANN 2002)

Le rôle exact de ces deux protéines dans le recrutement des complexes du PcG est cependant difficile à définir car elles sont toutes deux impliquées dans d'autres processus biologiques (SIEGEL *et al.* 1993; BHAT *et al.* 1996; HUANG and CHANG 2004). D'autre part, la cartographie de la protéine GAF montre qu'elle n'est présente que sur 50% des sites occupés par les protéines du PcG (NEGRE *et al.* 2006; SCHUETTENGRUBER and CAVALLI 2009). La particularité des membres de la famille BTB-POZ est le domaine BTB qui permet aux protéines d'oligomériser en solution (LEHMANN *et al.* 1998). Par ailleurs, ce domaine est également retrouvé dans quelques protéines insulatrices comme Mod(mdg4) et CP190 (KYRCHANOVA *et al.* 2016). Il n'est donc pas surprenant de voir que GAF joue également un rôle dans l'organisation chromatinienne (MELNIKOVA *et al.* 2004; PETRASCHECK *et al.* 2005; FUDA *et al.* 2015), une fonction qui est par ailleurs conservée chez les mammifères (SRIVASTAVA *et al.* 2013).

iv) DPS1 (*Dorsal switch protein 1*)

La protéine DSP1 fait partie d'une famille de protéines nommée ETP (*Enhancers of trithorax and Polycomb*) dont les mutants ont des phénotypes incluant des dérégulations de gènes cibles du TrxG et du PcG (DECOVILLE *et al.* 2001; SALVAING *et al.* 2006). D'un point de vue biochimique, DSP1 fait partie des protéines HMG de type II. Les protéines HMG sont caractérisées par leur domaine HMG constitué d'environ 70 acides aminés leur permettant de se fixer à l'ADN. Contrairement au type I, les HMG de type II contiennent plusieurs domaines HMG et ont la particularité d'être moins spécifique quant aux motifs d'ADN qu'elles reconnaissent (DECOVILLE *et al.* 2001). Cependant, DSP1 a été identifiée comme recruteur du complexe PcG via un motif G(A) retrouvé dans différents PRE (DEJARDIN *et al.* 2005). En effet, une mutation de ce motif dans deux PRE transgéniques suffit à abolir le recrutement des protéines du PcG mais change également les PRE en TRE. Par ailleurs, le recrutement de GAF semblerait être indépendant de DSP1 (DEJARDIN *et al.* 2005). Cependant, la cartographie de DSP1 dans les embryons de Drosophile montre que, comme GAF, seuls 50% des sites fixés par le PcG sont aussi fixés par DSP1 (SCHUETTENGRUBER *et al.* 2009; SCHUETTENGRUBER *et al.* 2014). De plus, la séquence préalablement déterminée G(A) n'est pas enrichie au niveau de ces PRE, ce qui soutient le caractère non-spécifique des protéines HMG-II. Par ailleurs, DSP1 interagit avec Corto, une autre protéine ETP (SALVAING *et al.* 2006). Cette dernière peut interagir avec GAF et des protéines du PRC1 (SALVAING *et al.* 2003). Le rôle global des interactions DSP1-CORTO-GAF n'est pas encore élucidé. En particulier, le rôle de Corto dans le recrutement de PcG n'a pas été étudié.

v) Zeste

La protéine Zeste fait également partie de la famille ETP (KAL *et al.* 2000; SAURIN *et al.* 2001; HUR *et al.* 2002; DEJARDIN and CAVALLI 2004). Les mutants sont viables et fertiles (GOLDBERG *et al.* 1989). Il a été suggéré que Zeste pouvait servir de pont entre les PRE et les promoteurs (SAURIN *et al.* 2001). Cependant, la cartographie de la protéine Zeste montre une très faible colocalisation entre cette protéine et les membres du PcG (OKTABA *et al.* 2008). Ainsi, le rôle exact des protéines de la famille ETP comme DSP1 et Zeste dans le recrutement du PcG reste encore obscur.

vi) Grh (Grainyhead)

Tout comme les protéines évoquées ci-dessus, la protéine Grh a également été identifiée comme ayant un rôle dans le recrutement de PcG (BLASTYAK *et al.* 2006). *Grh* interagit avec *Pho in vitro* et génétiquement (BLASTYAK *et al.* 2006). De plus, Grh

interagit physiquement avec dRING et cette interaction est conservée chez les mammifères (TUCKFIELD *et al.* 2002). Malgré cela, le consensus de Grh (TGTTTTTT (BLASTYAK *et al.* 2006) ne semble pas être présent sur tous les PREs et la colocalisation entre cette protéine et le PcG est inconnue. En effet, aucun marquage sur chromosome polytène n'a été réalisé. Par ailleurs, tout comme GAF, DSP1 et ZESTE, Grh a initialement été découverte comme activateur transcriptionnel (BRAY *et al.* 1989). Il a été suggéré que son rôle en tant que répresseur/activateur était dépendant du contexte cellulaire (KASSIS and BROWN 2013).

vii) Cg (Combgap)

Par ailleurs, l'étude du PRE d'*engrailed* révèle également l'existence d'un motif GTGT présent à de nombreux PRE (RINGROSE *et al.* 2003; SCHUETTENGRUBER and CAVALLI 2009; SCHUETTENGRUBER *et al.* 2014). Récemment, la protéine Cg a été identifiée comme impliquée dans le recrutement des protéines PcG via ce motif (RAY *et al.* 2016). De façon intéressante, Cg interagit génétiquement avec Pho et Spps et colocalise fortement avec Spps sur les chromosomes polytènes. Malgré une forte colocalisation entre Cg et la protéine du PcG Ph, l'analyse des mutants Cg montre que de nombreux sites Ph ne sont pas affectés par la perte de Cg.

C. A l'interface des complexes eux-mêmes

Nous avons vu dans les sections précédentes la complexité du recrutement des complexes du PcG au niveau de la séquence même des PRE et des FT qui permettaient de les recruter. Cependant, le recrutement hiérarchique impliquant que ces FT servent au recrutement de PRC2 qui ensuite recrute PRC1 est aujourd'hui largement discuté.

En effet, le recrutement hiérarchique prédit la co-occurrence des complexes PRC1 ET PRC2 au niveau de leurs gènes cibles. Cependant, la cartographie de certaines sous-unités de ces complexes montrent l'existence de sites génomiques où PRC1 et PRC2 colocalisent, alors que d'autres sites sont uniquement fixés soit par l'un, soit par l'autre des complexes (BOYER *et al.* 2006; KU *et al.* 2008; BLACKLEDGE *et al.* 2014; KAHN *et al.* 2016).

La découverte de différentes formes de PRC1 chez les mammifères a permis de démontrer que seulement certains complexes PRC1 étaient associés à la marque H3K27me3 (GAO *et al.* 2012). Par ailleurs, les complexes PRC1 non-canoniques peuvent être recrutés indépendamment de PRC2 et de sa marque dans les cellules ES (TAVARES *et al.* 2012) (**Figure** 21). Ainsi, bien que PRC2 puisse recruter PRC1 à certains sites

génomiques où les deux complexes colocalisent, les études récentes montrent que PRC1 peut également recruter PRC2 ce qui indique que le recrutement hiérarchique proposé jusqu'alors peut être complètement réversé (BLACKLEDGE *et al.* 2014; COOPER *et al.* 2014; KALB *et al.* 2014; KAHN *et al.* 2016). Chez les mammifères, c'est plus particulièrement le complexe PRC1 non-canonique BCOR qui est impliqué dans le recrutement de PRC2. BCOR se fixe à la chromatine via sa sous-unité KDM2B (FARCAS *et al.* 2012; BLACKLEDGE *et al.* 2014) (**Figure** 21). La déposition de H2AUb par BCOR induit le recrutement de PRC2 et le dépôt de H3K27me3 (BLACKLEDGE *et al.* 2014).

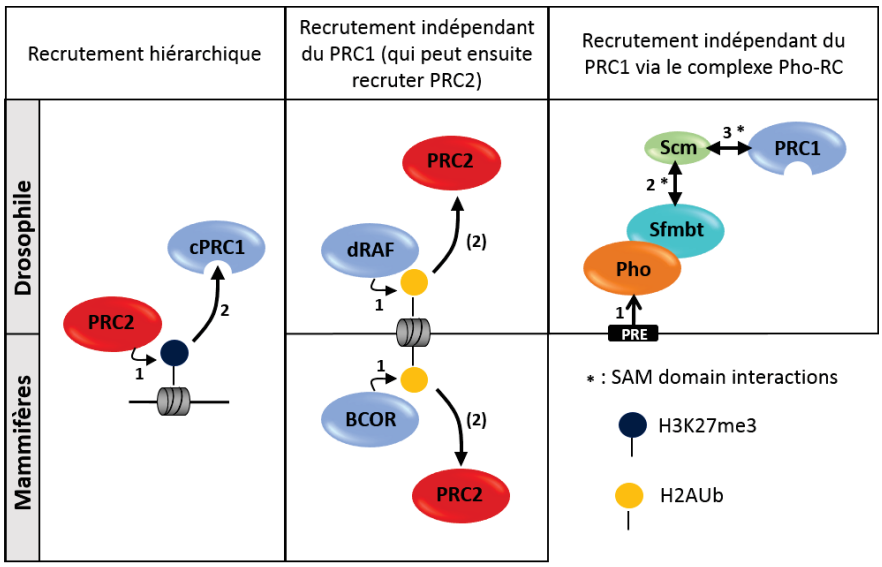


Figure 22 : Complexité du recrutement des complexes du PcG à l'interface même de ces protéines

Différents mécanismes de recrutement des complexes du PcG à la chromatine existent. Chez la Drosophile comme chez les mammifères, le recrutement hiérarchique proposant que PRC2 recrute PRC1 via H3H27me3 peut être réversé. PRC1 peut recruter PRC2 via H2AUb. Enfin, chez la Drosophile, Pho-RC composé de Pho et Sfmbt peut recruter PRC1 via la protéine Scm.

De la même manière chez la Drosophile, le complexe orthologue à BCOR, dRAF pourrait également être impliqué dans le recrutement de PRC2 (**Figure** 21) mais son rôle précis dans le recrutement des complexes PcG n'a pas encore été clairement établie (KAHN *et al.* 2016). En revanche, il a été récemment montré que le complexe Pho-RC était nécessaire au recrutement du PRC1 (**Figure** 21). En effet, Frey *et al.*, ont démontré l'importance de l'interaction entre les domaines SAM de la sous-unité Sfmbt et de la protéine Scm dans le recrutement du PRC1 via le domaine SAM de Ph. Ainsi, par l'intermédiaire de Scm, Pho-RC recruterait PRC1. De plus, le recrutement initial de PRC1

semble être important pour celui de PRC2 (FREY *et al.* 2016). Il est à noter que malgré l'importance de Pho-RC dans le recrutement de PRC1, il a été préalablement suggéré qu'en retour, PRC1 permettrait de stabiliser Pho-RC (KAHN *et al.* 2014). Il est donc fort probable que des interactions combinatoires entre les différents complexes permettent leur fixation de manière stable à la chromatine (SCHUETTENGRUBER *et al.* 2014).

D. Existence de sites de fixations des protéines du PcG au niveau de sites actifs

Le dernier niveau de complexité repose sur le fait que les protéines du PcG sont également retrouvées au niveau de gènes transcriptionnellement actifs marqués par la marque antagoniste à l'activité PcG : H3K27ac. En effet, Schaff *et al.*, ont tout d'abord montré que chez la larve de Drosophile, la cohésine permettait le recrutement de PRC1 au niveau de gènes actifs (SCHAAF *et al.* 2013). La perte du PRC1 sur ces sites est corrélée avec une diminution de la transcription. Par la suite, Loubière *et al.,* ont révélé l'importance fonctionnelle du PRC1 aux sites actifs de la larve en montrant que ce complexe régulait des gènes impliqués entre autres dans la prolifération cellulaire (LOUBIERE *et al.* 2016). La perte du PRC1 provoque l'apparition de cancers. De façon importante, les sites actifs fixés par le PcG ne sont pas visibles chez l'embryon de Drosophile (SCHUETTENGRUBER *et al.* 2009) indiquant que le recrutement du PcG aux régions actives pourrait se faire plus tard au cours du développement.

III. Le recrutement des protéines du PcG chez les mammifères est aussi complexe que chez la Drosophile (ENTREVAN *et al.* 2016)

Chez les mammifères, cinq PRE ont été identifiés (SING *et al.* 2009; WOO *et al.* 2010; CUDDAPAH *et al.* 2012; VASANTHI *et al.* 2013; BASU *et al.* 2014). De façon intéressante, ces PRE testés dans des transgènes peuvent également recruter le PcG chez la Drosophile indiquant que ces séquences possèdent des motifs conservés. Cependant, la très faible quantité de PRE retrouvée chez les mammifères indique que les sites de fixations des protéines du PcG sont différents (ENTREVAN *et al.* 2016).

Premièrement, les modifications post-traductionnelles des histones peuvent permettre le recrutement du PcG à certain locus. En effet, la déméthylase KDM2B et sa reconnaissance de la marque H3K36me2/3 permet le recrutement des PRC1 non-canoniques indépendamment de PRC2. C'est également le cas pour les PRC1 non-canoniques incluant E2F6 (recrutant les PRC1 non-canoniques aux *E-boxes* = *Enhancer boxes*). PCL permet le recrutement du PRC2 grâce au domaine TUDOR reconnaissant H3K36me3. Ainsi, comme vu précédemment, les interactions entre des protéines du PcG et des FT tels que E2F6, MGA/MAX ou encore JARID2 et AEBP2 permettent également leur recrutement à certain de leur gènes cibles.

De plus, les études *genome-wide* chez les mammifères montrent que les protéines du PcG sont plutôt associées aux îlots CpG hypo-méthylés (CGI) de promoteurs (BOYER *et al.* 2006; BRACKEN *et al.* 2006; LEE *et al.* 2006; TANAY *et al.* 2007; KU *et al.* 2008; LYNCH *et al.* 2012). De plus, des éléments artificiels enrichis en CGI et dépourvus de sites de fixations pour des FT peuvent recruter les protéines du PcG de manière ectopique (MENDENHALL *et al.* 2010; FARCAS *et al.* 2012; RIISING *et al.* 2014). Ceci laisse suggérer que les CGI chez les mammifères sont les équivalents des PRE chez la Drosophile.

Enfin, les longs ARN non-codants (lncRNAs) sont également impliqués dans le recrutement des protéines du PcG. Après la découverte que l'ARN XIST contribuait au recrutement de PcG pour permettre la répression du chromosome X, d'autres études révélèrent l'implication d'autres lncRNAs tels que HOTAIR ou KCNQ1 dans le recrutement de PcG aux gènes HOX ou aux gènes soumis à empreintes (voir (DAVIDOVICH and CECH 2015) pour une revue). Cependant, le rôle des lncRNAs dans le recrutement du PcG est encore discuté puisque certaines études montrent qu'au contraire, l'association ARN-PCR2 permettent de séquestrer PRC2 et donc de prévenir sa fixation à la chromatine (HERZOG *et al.* 2014; BELTRAN *et al.* 2016).

En résumé, le recrutement des complexes du PcG chez les mammifères est aussi complexe que chez la Drosophile. En effet, bien que les séquences de recrutement diffèrent, il est possible que différentes classes de sites de fixation aux protéines du

PcG existent en fonctions des facteurs (modifications post-traductionnelles d'histones, CGI, TF, ARN) qui les recrutent.

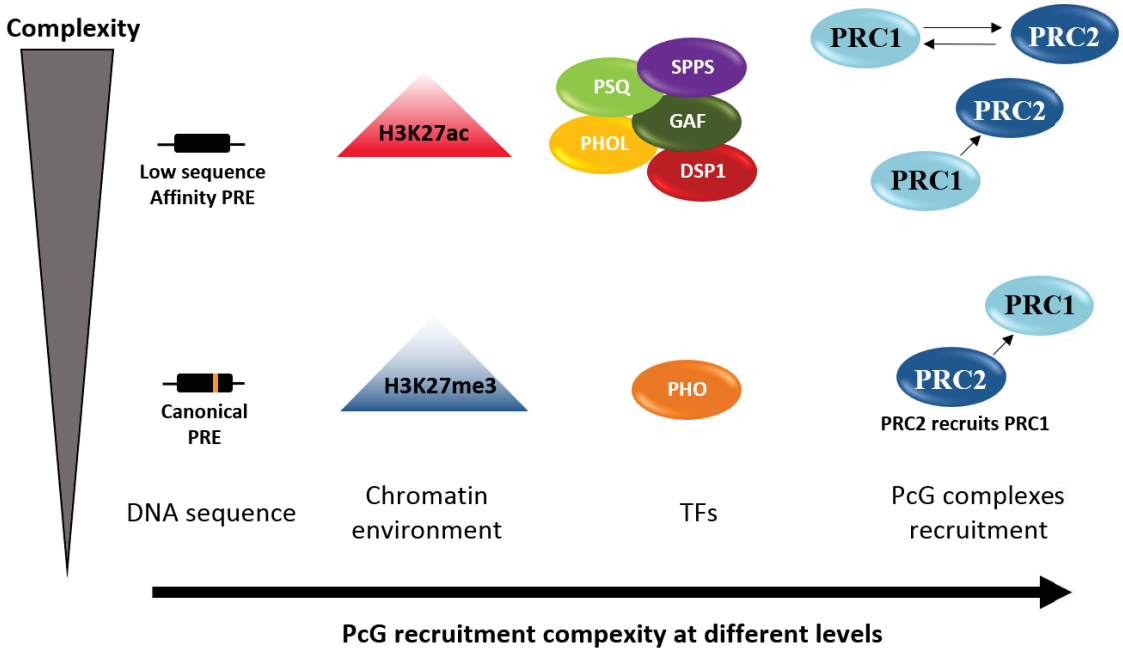


Figure 23 : Complexité du recrutement du PcG à la chromatine à différents niveaux

De gauche à droite : il est impossible de définir les PRE sur la base de leur séquence puisqu'il existe des PRE qui ne sont pas enrichis en motifs spécifiques. Ensuite, bien que les PRE soient définis par la présence des complexes du PcG au niveau de domaines répressifs marqués par H3K27me3, il existe des sites de fixation des protéines du PcG au sein de régions transcriptionnellement actives marquées par H3K27ac. De plus, bien que Pho ait été le premier FT identifié comme important dans le recrutement des complexes du PcG, d'autres FT ont été identifiés et il semblerait que ce soit l'action combinatoire de tous ces facteurs qui soit importante pour le recrutement des complexes du PcG. Enfin, bien qu'un modèle de recrutement hiérarchique ait été proposé suggérant que PRC2 recrutait PRC1, nous savons aujourd'hui que les complexes du PcG peuvent interagir de façon coopérative et que le recrutement hiérarchique peut être complétement réversé (PRC1 recrute PRC2).

CONCLUSION ET INTRODUCTION DU PROJET DE THESE

Les protéines du groupe Polycomb (PcG) ont initialement été identifiées chez la drosophile comme répresseurs transcriptionnels des gènes homéotiques. Aujourd'hui, nous savons que ces protéines jouent un rôle bien plus large puisqu'elles régulent des gènes dont les produits sont impliqués dans de nombreux processus biologiques (régulation des gènes HOX, maintien de la plasticité des cellules souches, la différenciation cellulaire, l'inactivation du chromosome X, la régulation des gènes soumis à empreintes). Leur dérégulation est source de nombreux cancers chez l'homme.

Hautement conservées, elles forment deux principaux complexes : PRC 1 et 2 (*Polycomb repressive complex 1 and 2*), dont l'activité est respectivement reflétée par la mono-ubiquitinylation de la lysine 118 l'histone H2A (H2AK118Ub) et la tri-méthylation de la lysine 27 de l'histone H3 (H3K27me3). Chez la Drosophile, les sites de fixation de ces complexes sont appelés PRE (*Polycomb Responsive Elements*) où ils sont recrutés via des facteurs de transcription (FT).

Nous avons vu dans le chapitre précédent la complexité du recrutement des protéines du PcG à quatre niveaux : au niveau de la séquence même de leurs sites de fixations, au niveau des FT qui les recrutent, au niveau de l'interface entre les deux complexes PRC1 et PRC2 et enfin au niveau global, part le présence de ces complexes au niveau de sites transcriptionnellement actifs (le tout est résumé dans la **Figure 23**).

L'ensemble de ces résultats démontre clairement la nature hétérogène des PRE.

Mon projet de thèse s'est donc dessiné autour d'une hypothèse : il existe différentes classes de PRE chez la Drosophile. Mon travail a donc consisté à définir ces différentes classes et à les caractériser pour en déduire des rôles spécifiques à l'échelle génomique. En effet, l'implication des complexes du PcG dans l'apparition de cancer chez l'Homme requière que l'on comprenne comment ces protéines sont recrutées à la chromatine.

RESULTATS

Page blanche

Résumé en français

Dans le but d'identifier différentes classes de PRE chez la Drosophile, nous avons cartographié les protéines E(z) et Ph (de PRC2 et de PRC1, respectivement) par Chip-seq à partir de cellules S2. Nous avons identifié 1543 PRE à partir desquels nous avons effectué une analyse de clustering basée sur l'environnement chromatinien. Six différentes classes de PRE ont été identifiées. De manière générale, seulement 1/3 des PRE sont retrouvés au sein de régions marquées par H3K27me3, les 2/3 colocalisant avec la marque H3K27ac. La caractérisation des différentes classes de PRE nous a permis de mettre en évidence une signature particulière, en DBP et en FT recrutant les protéines du PcG, propre à chacune des classes de PRE. Les résultats indiquent que les classes de PRE 1 et 4 sont caractérisées par une forte proportion d'enhancers actifs marqués par la présence de H3K27ac et H3K4me1. La classe 2 se caractérise par des PRE colocalisant avec les TSS de gènes hautement transcrits marqués par H3K27ac, H3K36me3 et H3K4me3. La classe 3 correspond à des PRE situés en bordure de TAD séparant des domaines marqués par H3K27me3 et H3K27ac. Par ailleurs, cette classe est enrichie en protéines insulatrices et montre une forte insulation d'après les résultats de Hi-C. La classe 5 se caractérise par des PRE localisés dans des régions faiblement enrichies en H3K27me3 et H3K27ac. La particularité des PRE de cette classe vient du fait que les TSS des gènes associés colocalisent avec l'ARN Pol II alors que le niveau d'expression de ces gènes est globalement faible. De fait, nous avons caractérisé cette classe comme correspondant à des éléments génomiques en état de pause transcriptionnelle. Enfin, la classe 6 correspond aux PRE canoniques caractérisés par un fort enrichissement en protéines E(z) et Ph retrouvées au sein de larges domaines marqués par H3K27me3. Les gènes cibles des PRE canoniques sont transcriptionnellement inactifs.

Afin de tester l'activité des PRE issus de chaque classe identifiée précédemment, nous avons conçu un vecteur contenant deux gènes rapporteurs nous permettant de tester l'activité des PRE à deux stades développementaux différents. De manière générale, les PRE canoniques montrent une activité de PRE aux deux stades de développement testés, activité qui est reflétée par la répression de l'expression des gènes rapporteurs. Cette propriété est retrouvée pour les PRE issus de la classe 5. Cependant, nos résultats suggèrent que certains PRE de la classe 5 semblent agir comme tel de manière développement spécifique. Les PRE issus des classes 1 à 4 ne montrent pas d'activité PRE.

Enfin, dans le but de mieux comprendre le rôle des protéines du PcG au niveau des régions transcriptionnellement actives (classes 1 à 4), deux lignées de cellules S2 KO pour le gène *Pc* ont été créées en utilisant le système CRISPR/Cas9. Les ChIP quantitatifs

effectués pour certaines sous-unités des complexes du PcG indiquent que le complexe dRAF est principalement présent au niveau des PRE des classes 1 à 4, alors que le complexe PRC1 est principalement retrouvé au niveau des PRE des classes 5 et 6. Par ailleurs, un transcriptome réalisé à partir des cellules mutantes nous permet de conclure un rôle différentiel entre les protéines du PcG situées aux PRE canoniques ou non-canoniques (classes 1 à 4). En effet, aux PRE canoniques, les protéines du PcG répriment totalement l'expression de leurs gènes cibles alors qu'aux PRE non-canoniques, ces protéines modèrent l'expression de leurs gènes cibles sans la réprimer complètement.

Pour conclure, les résultats obtenus révèlent une diversité des PRE bien plus grande que décrite initialement dans l'embryon où les gènes cibles des protéines du PcG sont finement régulés.

Page blanche

Tableau 1 : ChIP-seq data from Drosophila S2 cells that we used in this study.

	Protein/Mark	Mapping by: (Ref.)	Comment
PcG proteins	Ph	This study	PRC1 component
	E(z)	This study	PRC2 component
TrxG proteins	dSET1	Rickels et al., 2016	SET1/COMPASS component
	Trx	Rickels et al., 2016	MLL1/2 COMPASS-like component
	Trr	Rickels et al., 2016	MLL3/4 COMPASS-like component
Histone marks	H3K27me3	Herz et al., 2012	PRC2-mediated mark
	H3K27me2	Rickels et al., 2016	PRC2-mediated mark
	H3K27ac	Herz et al., 2012	Antagonistic mark to H3K27me3, also associated with enhancers
	H2A118ub	This study	PRC1-mediated mark
	H3K4me3	Herz et al., 2012	Mark associated with active promoters
	H3K4me1	Herz et al., 2012	Mark associated with enhancers
	H3K4me2	Rickels et al., 2016	Conserved Mark associated with PcG proteins
DNA binding factors involved in PcG recruitment	PHO	This study	PcG recruiter
	PHOL	This study	PcG recruiter
	DSP1	This study	PcG recruiter
	SPPS	This study	PcG recruiter
	PSQ	This study	PcG recruiter
	GAF	This study	PcG recruiter
Insulators	Beaf32	Liang et al., 2014	Insulator protein
	CTCF	Ong et al., 2014	Insulator protein
	CP190	Ong et al., 2014	Insulator protein
	Su(H)w	Ong et al., 2014	Insulator protein
	Mod(mdg4)2.2	Ong et al., 2014	Insulator protein
	Ibf 1 and Ibf2	Cuartero et al., 2014	Insulator protein
	ZIPIC	Maksimenko et al., 2015	Insulator protein
	Pita	Maksimenko et al., 2015	Insulator protein
Other	RNA Polymerase II	Smith et al., 2011	na
	DnaseI	Karchenko et al., 2011	na

I. CLUSTERING ANALYSIS REVEALED THE EXISTENCE OF SIX CLASSES OF PREs , CHARACTERIZED BY DISTINCS GENOMIC AND CHROMATIN FEATURES

A. Data set presentation

To verify our working hypothesis, we performed ChIP-seq to map the two main PcG complexes PRC2 and PRC1 (E(z) for PRC2 and Ph for PRC1), their associated histone marks H2AK118ub (later named H2AUb) and H3K27me3, and the majority of their recruiting factors (Pho, Phol, Spps, Dsp1, Psq and Gaf). All data were obtained in duplicates with good correlation, comprising 6-30 million aligned reads (**Figure 24 A and B**). The use of S2 cells shows many advantages: not only it represents a homogeneous population of cells, but it has been also used by the modENCODE project and other teams in the field to map a large plethora of chromatin-associated factors and histone marks. Thus, many chromatin data in addition to RNA-seq data are already available that we used for the subsequent analysis (**Table 1**).

We used MACS 2 for peak calling to define Ph and E(z) binding sites and identified 3129 significant sites for Ph and 2771 significant sites for E(z) (**Figure 24 C**). Peak calling of PcG recruiters identified 5944 sites for Pho, 4992 sites for Phol, 3915 sites for Spps, 3583 sites for Dsp1, 5953 sites for Psq and 7245 sites for Gaf.

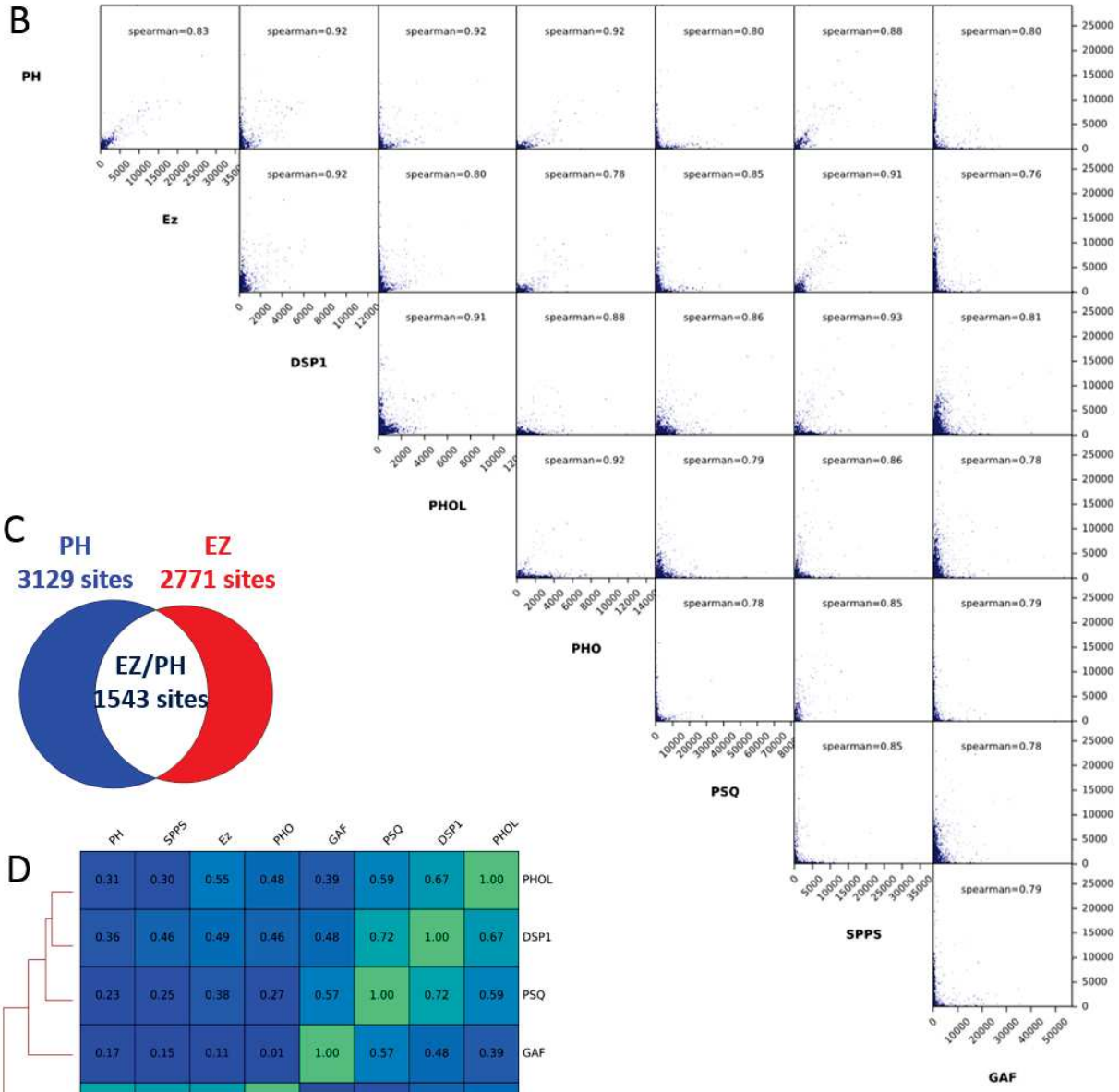
We then defined a PRE as being the intersection region between a E(z) and Ph bound site. We thus obtained 1543 high-confidence PcG binding sites with a peak length mean of 711bp (**Figure 24 C**). Although the correlation between PcG proteins and sequence specific DNA binding proteins was generally high on genome-wide scale (**Figure 24 B**), we noticed two different clusters of proteins at the 1543 PREs (**Figure 24 D**). PcG proteins E(z) and Ph were more correlated with Pho and Spps (correlation coefficient between 0.66 and 0.77) than with Phol, Dsp1, Psq and Gaf (correlation coefficient between 0.11 and 0.55). This gave us a first indication for the variability between the PREs as previously observed (SCHUETTENGRUBER *et al.* 2009; SCHUETTENGRUBER *et al.* 2014).

Figure 24

A

	Ph	E(z)	DSP1	PHOL	PHO	SPPS	PSQ	GAF
Spearman (Bin=500bp)	0.93	0.94	0.89	0.79	0.94	0.82	0.78	0.79

B



C

D

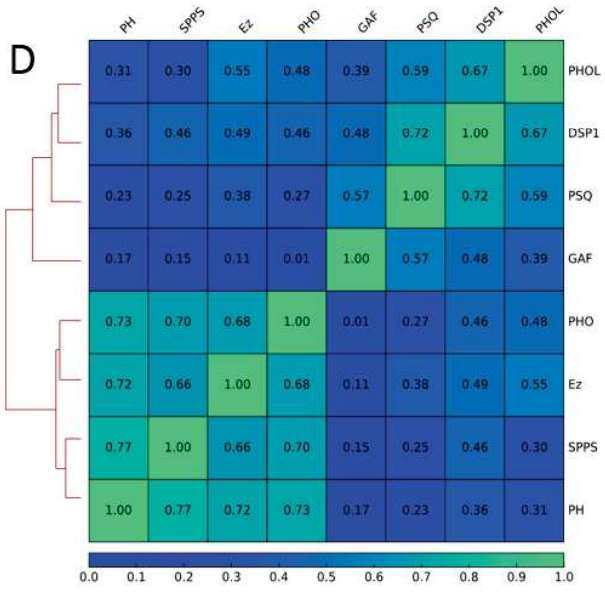


Figure 24 : Quality of data used to analyze the diversity of PREs in Drosophila S2 cells.

A. Table indicating the Spearman correlation calculated between each duplicate of protein (bin=500bp).
B. Spearman correlation was calculated genome-wide (bin=500bp) between each protein performed in ChIP-seq (only one replicate is shown in the figure).
C. Venn Diagram indicating the number of significant sites for E(z) and Ph PcG proteins. The intersection between both data gave a list of 1543 E(z)/Ph sites that were used for further analysis.
D. Spearman correlation between PcG proteins and PcG recruiters calculated on the 1543 selected sites.

B. Six different classes of PREs are defined by distinct chromatin signatures

As a first attempt to classify PREs, we clustered them based on their spatial environment. We used the list of 1543 E(z)/Ph sites (from now on called "PREs") as input for a clustering analysis using seqMINER (YE *et al.* 2011). Three histone marks associated with PcG-mediated repression or TrxG-mediated gene activation were used as variables: H3K27me3, the hallmark of PcG-dependent gene silencing; it's counteracting and mutually exclusive histone mark H3K27ac mediated by the histone acetyltransferase CBP, and the H3K36me3 mark that is generally found in the coding region of transcribed genes and deposited by the TrxG protein Ash1 (*Absent small and homeotic disc 1*).

Based on these histone marks we identified six different clusters characterized by a particular chromatin environment (**Figure 25 A**) and validated them by visual inspection using a genome browser (**Figure 25 G**, **Figure 27 A** and data not shown). The six classes can be divided into two superclusters: An active supercluster (I) comprising classes 1-4, specified by high levels of the active mark H3K27ac, and a repressive supercluster (II) comprising classes 5-6, associated with the repressive histone mark H3K27me3. The fact that most PREs associate with an active chromatin environment seems surprising at the first sight, but corroborates recent findings that PcG proteins are targeted to sites of active transcription (reviewed in (ENTREVAN *et al.* 2016)).

To further characterize the six clusters of PREs, we then analyzed the enrichments of other post translational modifications (PTMs) of histones including H3K4me3 (mark of active TSS) and H2AUb (PTM associated with PRC1 activity) (**Figure 25 A**) and performed statistical analyzes (**Figure 25 B-F**) by comparing the enrichment level of histone PTMs for the six classes of PREs with a random control class (RGR) composed of 1544 sites. Statistical analyses were also performed on two other classes of protein binding sites: the 1255 E(z) sites devoid of Ph protein (later named E(z)-only sites) and the 1615 Ph sites devoid of E(z) protein (later named Ph-only sites) (**Figure 24 C**).

Figure 25

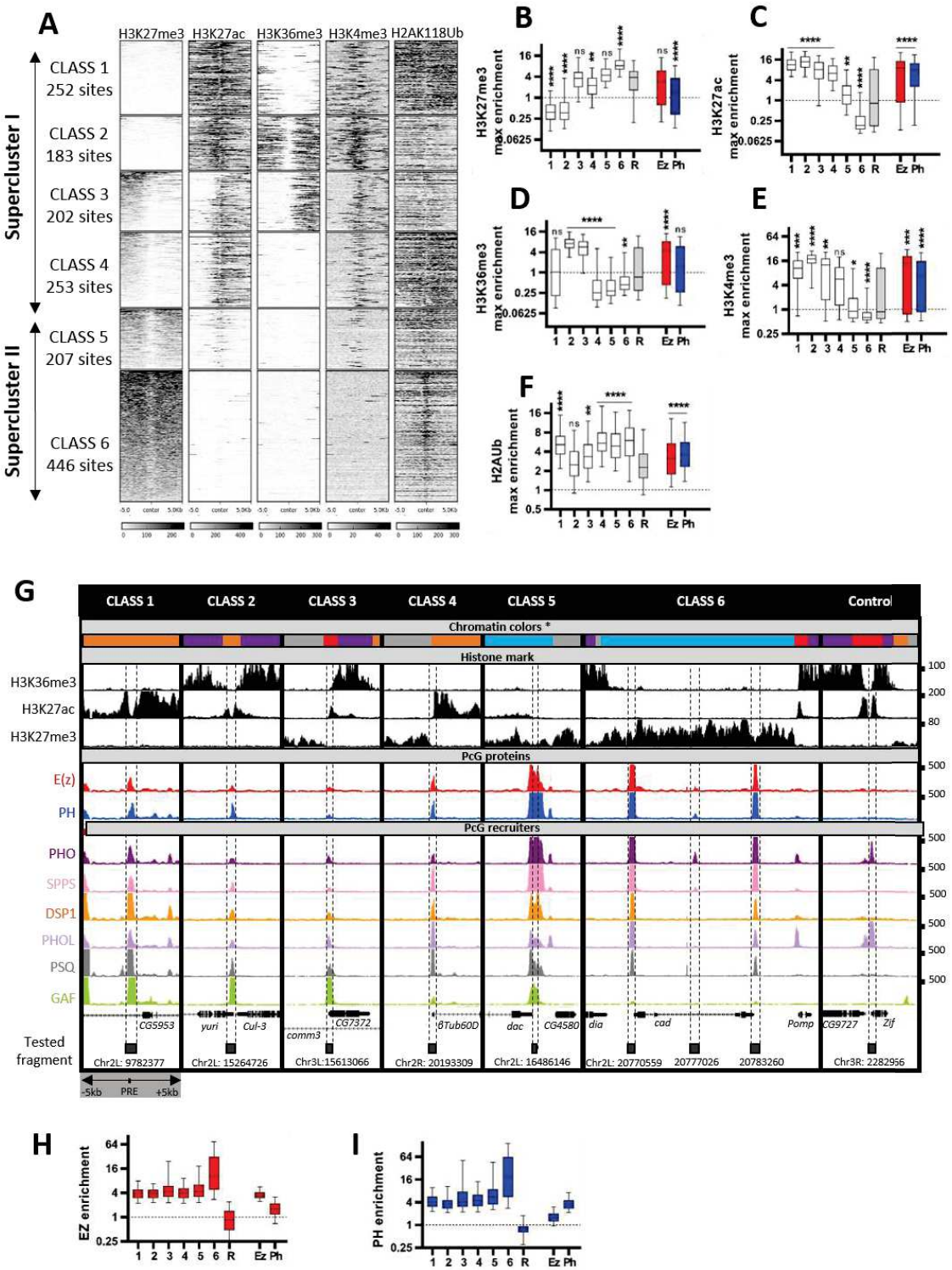

A
H3K27me3
H3K27ac
H3K36me3
H3K4me3
H2AK118Ub
Supercluster I
Supercluster II
CLASS 1
252 sites
CLASS 2
183 sites
CLASS 3
202 sites
CLASS 4
253 sites
CLASS 5
207 sites
CLASS 6
446 sites
B
H3K27me3
max enrichment
C
H3K27ac
max enrichment
D
H3K36me3
max enrichment
E
H3K4me3
max enrichment
F
H2AUb
max enrichment
1 2 3 4 5 6 R Ez Ph
G
CLASS 1
CLASS 2
CLASS 3
CLASS 4
CLASS 5
CLASS 6
Control
Chromatin colors *
Histone mark
H3K36me3
H3K27ac
H3K27me3
PcG proteins
E(z)
PH
PcG recruiters
PHO
SPPS
DSP1
PHOL
PSQ
GAF
CG5953
yuri
Cul-3
comm3
CG7372
βTub60D
dac
CG4580
dia
cad
Pomp
CG9727
Zif
Tested fragment
Chr2L: 9782377
Chr2L: 15264726
Chr3L:15613066
Chr2R: 20193309
Chr2L: 16486146
Chr2L: 20770559
20777026
20783260
Chr3R: 2282956
-5kb
PRE
+5kb
H
EZ enrichment
I
PH enrichment

Figure 25 : Clustering analysis of PREs revealed six different classes based on histones PTM

A:Clustering analysis of PREs was performed using three histones PTM (H3K27me3; H3K27ac and H3K36me3). Six classes were obtained. Then, H3K4me3 and H2AUb enrichments were mapped with respect of the predetermined classes.
B-F:Boxplots show the 5-95 percentile of max enrichments of various histones PTM obtained from 10kb windows centered on PREs. The results are shown for the six classes of PREs, a set of random regions (RGR) and E(z)- and Ph-only sites (see main text).
G:Snapshots of histone modifications and protein binding profiles of PREs corresponding to classes 1-6. Chromatin colors are based on (KHARCHENKO *et al.* 2011) (Red: Active TSS; Purple: Transcriptional elongation; Coral: Active enhancers; Blue: PcG-associated chromatin; Grey; Heterochromatin and null chromatin).
H-I:Boxplots show the 5-95 percentile of max enrichment of E(z) protein and Ph protein over a 10kb window centered on PREs. All boxplots are scaled in Log2.

Supercluster I

Classes 1 and 2 are both highly enriched in H3K27 acetylation and depleted of H3K27 methylation (**Figure 25 A, B and C**). They can be discriminated by the presence of high levels of H3K36me3 at class 2 PREs, which is absent from class 1 PREs (**Figure 25 A and D**). H3K4me3 levels are also highly enriched at class 2 PREs, whereas levels of H3K4me3 at class 1 PREs are still significant compared to the random class, but lower compared to class 1 (**Figure 25 A and E**). Intriguingly class 1 PREs are located within regions of significant H2A ubiquitination, whereas this histone mark is absent in class 2 PREs (**Figure 25 A and F**).

Classes 3 and 4 share the common feature that they are located at the borders of regions marked by repressive H3K27me3 mark and active H3K27ac mark (**Figure 25 A**). Thus, these two classes seem to be at the boundaries of active and repressive chromatin domains. Whereas active chromatin domains flanking class 3 PREs are in addition characterized by the presence of H3K36me3, this mark is absent in active chromatin regions flanking class 4 PREs (**Figure 25 A and D**). In contrast, class 4 PREs are located within regions marked by H2AK118ub which is not significantly enriched around class 3 PREs (**Figure 25 A and F**). H3K4me3 is occasionally associated with both classes of PREs, however only lowly or insignificantly enriched compared to the control regions (**Figure 25 A and E**).

Figure 26

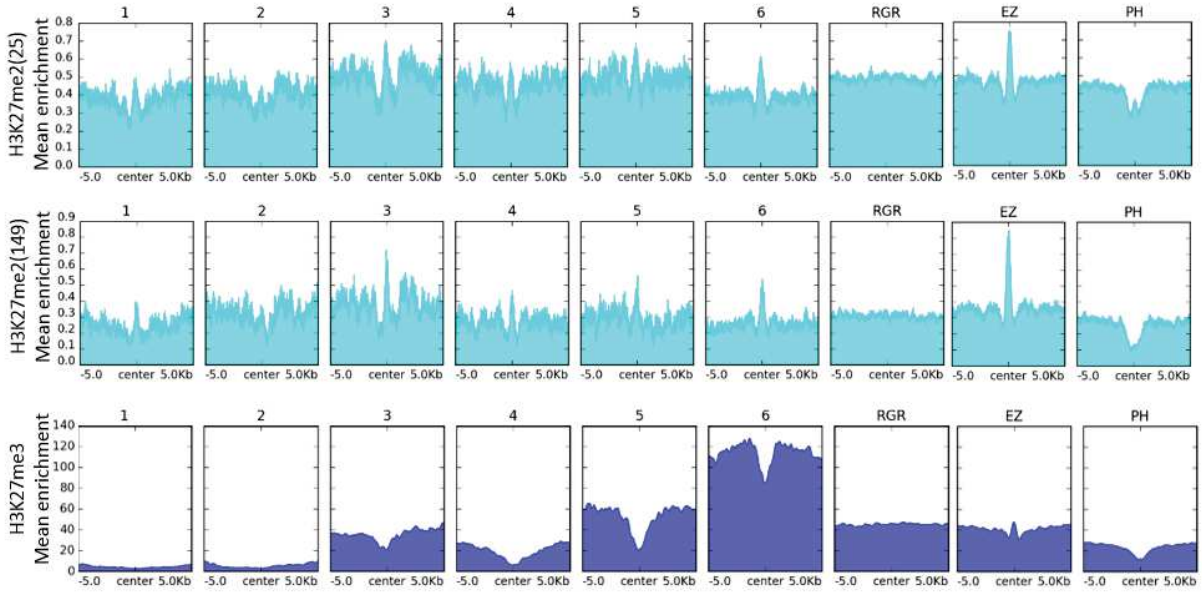


Figure 26 : Mean enrichment of histones PTM associated with H3K27 over a 10kb window.
Mean enrichment of H3K27me2 (aqua; lines 25 and 149 show biological replicates) and H3K27me3 (blue).

Supercluster II

Classes 5 and 6 are characterized by a repressed chromatin environment marked by the presence of H3K27me3 and the absence of high levels of H3K27ac (**Figure 25 A, B and C**). Class 6 is particularly enriched in the PRC2-associated mark H3K27me3 and depleted of active histone marks H3K27ac, H3K36me3 and H3K4me3 (**Figure 25 A-E**). In addition, this class is enriched in H2AK118ub around the peak centers (**Figure 25 A**).

Although H3K27me3 is present around class 5, it is not significantly enriched compared to the control class (**Figure 25 B**). Moreover, class 5 is specified by a low but significant enrichment in H3K27ac around the PRE center (**Figure 25 A and C**). Another difference between class 6 and class 5 PREs is the absence of H2AK118ub from class 5 and the low but insignificant enrichment of H3K4me3 levels in class 5 PREs. Thus, class 5 PREs show the typical characteristics of a bivalent chromatin structure (VOIGT *et al.* 2013; HARIKUMAR and MESHORER 2015). We therefore raised the hypothesis that this class represents a poised transcriptional state (discussed below). However at this stage our data do not allow to determine whether these opposing histone modifications are located on the same nucleosome or not. Moreover, their colocalization could also reflect a heterogeneity of S2 cells, where a subset of genes is repressed and associated with H3K27me3 in a subpopulation of the cells, while those genes that are active in another subpopulation of the cells are associated with H3K27ac.

The histone methyltransferase E(z) not only mediates tri-methylation of lysine 27 on histone H3, but is also responsible for mono- and di-methylation of H3K27. In particular, the H3K27me2 mark covers the large majority (70%) of the euchromatic genome (FERRARI *et al.* 2014; LEE *et al.* 2015) and it has been suggested that weak or transient interactions of PRC2 with the genome leads to the di-methylation of the genome which can suppress pervasive chromatin opening and transcriptional activation. The presence of significant level of E(z) without the presence of H3K27me3 mark therefore prompted us to look at the enrichment and distribution of H3K27me2 for each class of PREs (**Figure 26**). We noted a slight decrease in H3K27me2 levels in classes 1, 2 and 6. For classes 1 and 2, this decrease anti-correlates with the high level of H3K27ac, while the decreased level of H3K27me2 anti-correlates with the presence of H3K27me3 for class 6. More strikingly, the distribution of the H3K27me2 around PREs of the different classes is significantly different from the distribution of H3K27me3: Importantly, PREs were previously described as nucleosomes depleted regions (or regions with a low nucleosomal density). This can be visualized by the drop in H3K27me3 at the center of PREs (**Figure 26**). In strong contrast, a sharp peak of H3K27me2 is found at E(z)-only sites and at PREs of class 6, and to a lesser extent at PRE classes 3 and 5 (**Figure 26**). This observation, suggests that the few nucleosomes remaining at PREs (in particular of class 6) are marked by H3K27me2, whereas the surrounding chromatin is marked by H3K27me3. The biological importance of this

Figure 27

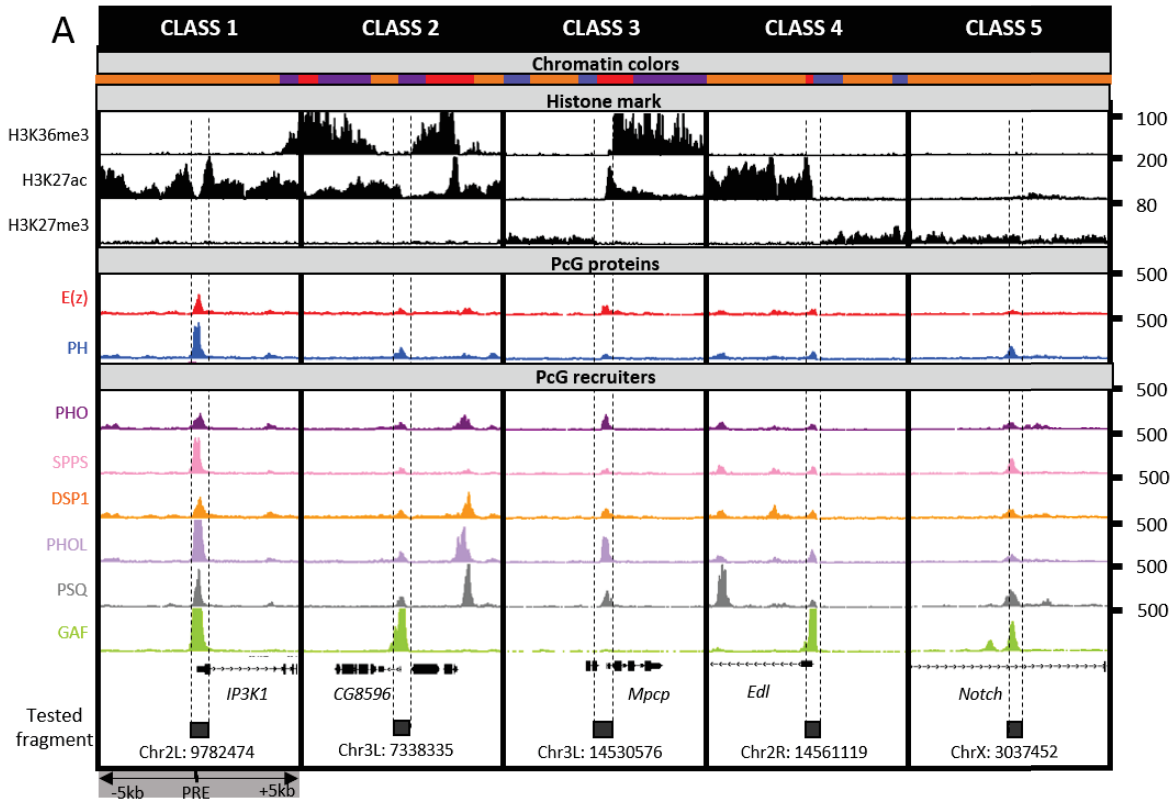


Figure 27 : Defining the classes of PREs

A. Snapshots of PREs selected for in vivo assays (see main text).
B. Proportion of sites overlapping with quartile of chromatin overture.
C. Proportion of sites overlapping with larval canonical PREs or Neo-PRC1 sites (see main text).

observation remains to be determined, however the presence of H3K27me2 at all E(z) bound sites confirms the biological activity of E(z) at genomic regions lacking H3K27me3.

Another significant difference between the two superclusters (active and repressed) are the enrichment levels of PcG proteins themselves. Therefore we assessed enrichment levels of PcG subunits for each class of PREs and we observed that, both E(z) and Ph enrichments are particularly high in class 6 (**Figure 25 H and I**, respectively) (SCHUETTENGRUBER *et al.* 2009). For the five other classes, the enrichment levels of PcG subunits are lower, but still significantly higher than that of the control class (RGR).

We next looked at the overlap of our six PRE classes with PREs recently mapped in larvae (LOUBIERE *et al.* 2016) (**Figure 27 C**). We found that many of class 6 PREs are conserved in Drosophila eye discs, while a high proportion of PREs from active classes 1-4 corresponds to so called "neo-PRC1 sites", which are PRC1 sites devoid of H3K27me3 that are absent in embryos but are acquired later in Drosophila development. Overall these results support the idea that active PRE classes are likely to be biologically relevant rather than being false positive in ChIP-seq.

Finally, we analyzed the chromatin accessibility of the 6 PRE classes and Ph- and E(z)-only sites (**Figure 27 B**). PREs of classes 1-4 are specified by a more opened chromatin structure compared to RGR, E(z)- and Ph-only sites, emphasizing the active properties of these sites. Contrarily, class 6 PREs is characterized by a more compacted environment, which correlates with the compaction activities of PcG proteins. Finally, class 5 PREs shows an intermediate chromatin opening state, which might reflect their poised/bivalent status.

In summary, our clustering analysis revealed the existence of 6 PRE classes that are characterized by distinct chromatin signatures: PREs of class 6 are characterized by high binding levels of PcG proteins, within regions marked my H3K27me3. In addition, these PREs are marked by H2AUb and are depleted of active chromatin marks. These PREs therefore resemble most to the characteristics of previously described PREs in drosophila embryos and we called them therefore "**canonical PREs**". PREs of class 5 have the characteristics usually associated with bivalent chromatin (co-occurrence of active and repressive chromatin marks), and might represent a poised state of PREs. PREs of classes 1-4 are associated with active chromatin environment. Classes 3 and 4 PREs correlate with the characteristics of chromatin borders, because they are located between repressive and active chromatin domains, whereas class 1 and 2 PREs correlate with chromatin features usually associated with active genes (**Figure 25 A**).

Figure 28

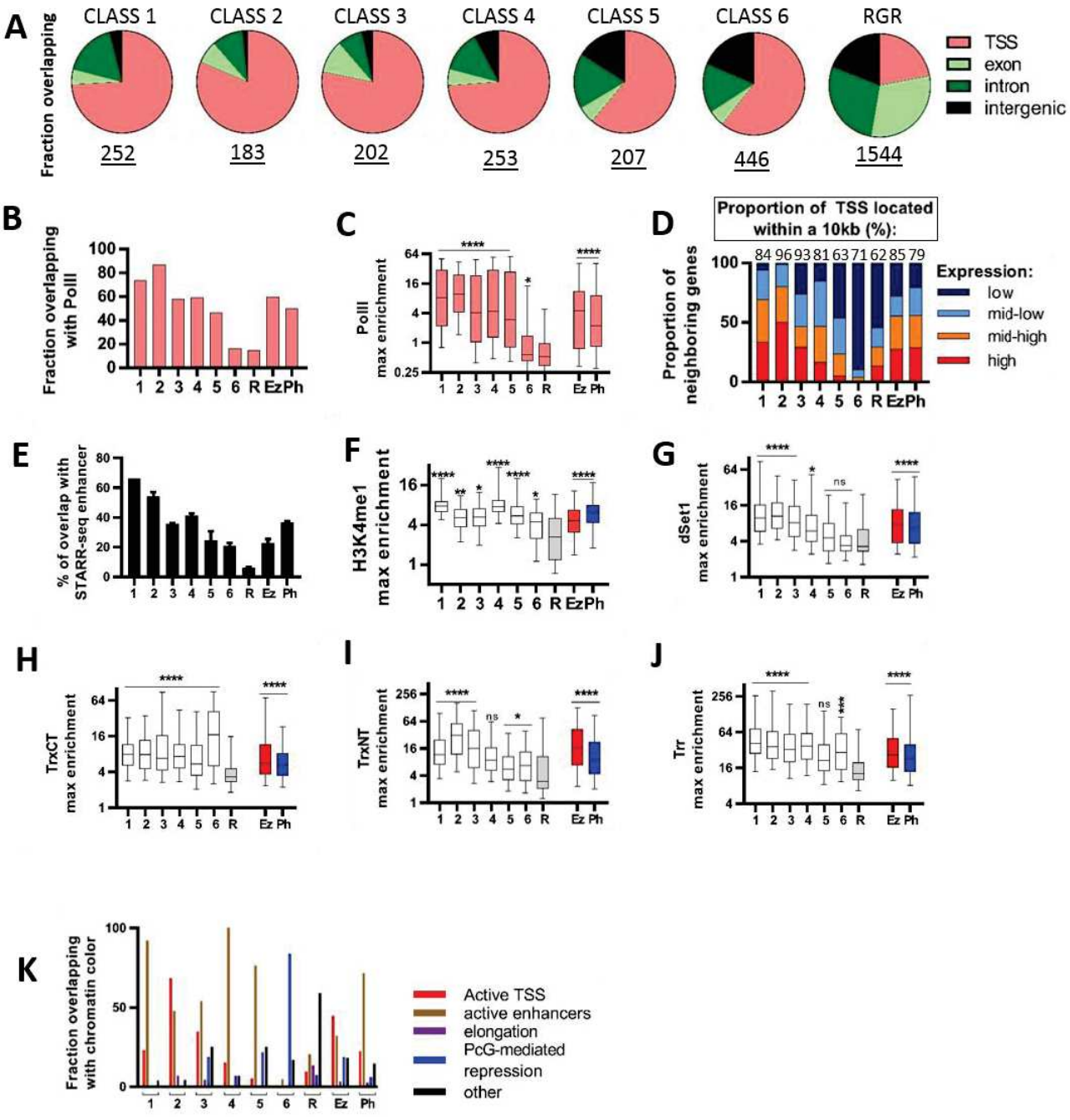

A
CLASS 1
CLASS 2
CLASS 3
CLASS 4
CLASS 5
CLASS 6
RGR
Fraction overlapping
TSS
exon
intron
intergenic
252
183
202
253
207
446
1544
B
Fraction overlapping with PolII
C
PolII max enrichment
D
Proportion of TSS located within a 10kb (%):
84 96 93 81 63 71 62 85 79
Proportion of neighboring genes
Expression:
low
mid-low
mid-high
high
E
% of overlap with STARR-seq enhancer
F
H3K4me1 max enrichment
G
dSet1 max enrichment
H
TrxCT max enrichment
I
TrxNT max enrichment
J
Trr max enrichment
K
Fraction overlapping with chromatin color
Active TSS
active enhancers
elongation
PcG-mediated repression
other

Figure 28 : PREs are localized at active, poised or repressed DNA elements

A. Fraction of PREs from the six classes overlapping with different genomic features.
B. Percentage of PREs from each class overlapping with PolII. PREs associated to TSS were selected if the distance between the center of PREs and the first TSS is comprised within a window of +/- 500bp.
C. Boxplots show the max enrichment of PolII from a 10kb window calculated on previous selected PREs associated with TSS (B).
D. Transcriptional level of genes associated to each class. The transcriptional level was divided into four. The proportion of TSS located within a 10kb window centered on the mid of PREs is given.
E. Proportion of PREs colocalizing with STARR-seq enhancer peaks.
F. Max enrichment in H3K4me1 over a 10kb window.
G. Mean enrichment in dSET1 protein over a 10kb window.
H. Mean enrichment in TrxCT protein over a 10kb window.
I. Mean enrichment in TrxNT protein over a 10kb window.
J. Mean enrichment in Trr protein over a 10kb window.
K. Proportion of PREs colocalizing with various chromatin colors as defined by (KHARCHENKO *et al.* 2011)

C. PREs associate with active, poised and repressed DNA elements

After having classified PREs based on their associated chromatin marks we next wanted to characterize them in more detail by analyzing their associated genomic features. Therefore, we first mapped their positions relative to several genomic features including transcription start sites (TSS), exons, introns and intergenic regions (**Figure 28 A**).

Genome wide mapping studies of PREs in Drosophila embryos (SCHUETTENGRUBER *et al.* 2009) showed that around 50% of the PREs overlap with TSSs, whereas the other half of the PREs can be located far away from gene promoters. Similarly, around 60% of Class 6 PREs (canonical PREs) are localized at TSS, whereas a high fraction of PREs fall outside TSSs and are found within intergenic regions. Class 5 PREs shows a similar distribution around genes than class 6 PREs. In contrast, PREs corresponding to active classes (1-4) show a higher enrichment in TSSs (around 75%) than classes 5 and 6 and they are significantly depleted in intergenic regions (max 5%) (**Figure 28 A**).

We next focused on these TSSs to look at their overlap with RNA Polymerase II (Pol II) (**Figure 28 B**). We found that at least 60% of TSS from active classes 1-4 along with TSS from E(z)- and Ph-only sites colocalize with Pol II. Most interestingly class 5 PREs also

Figure 29

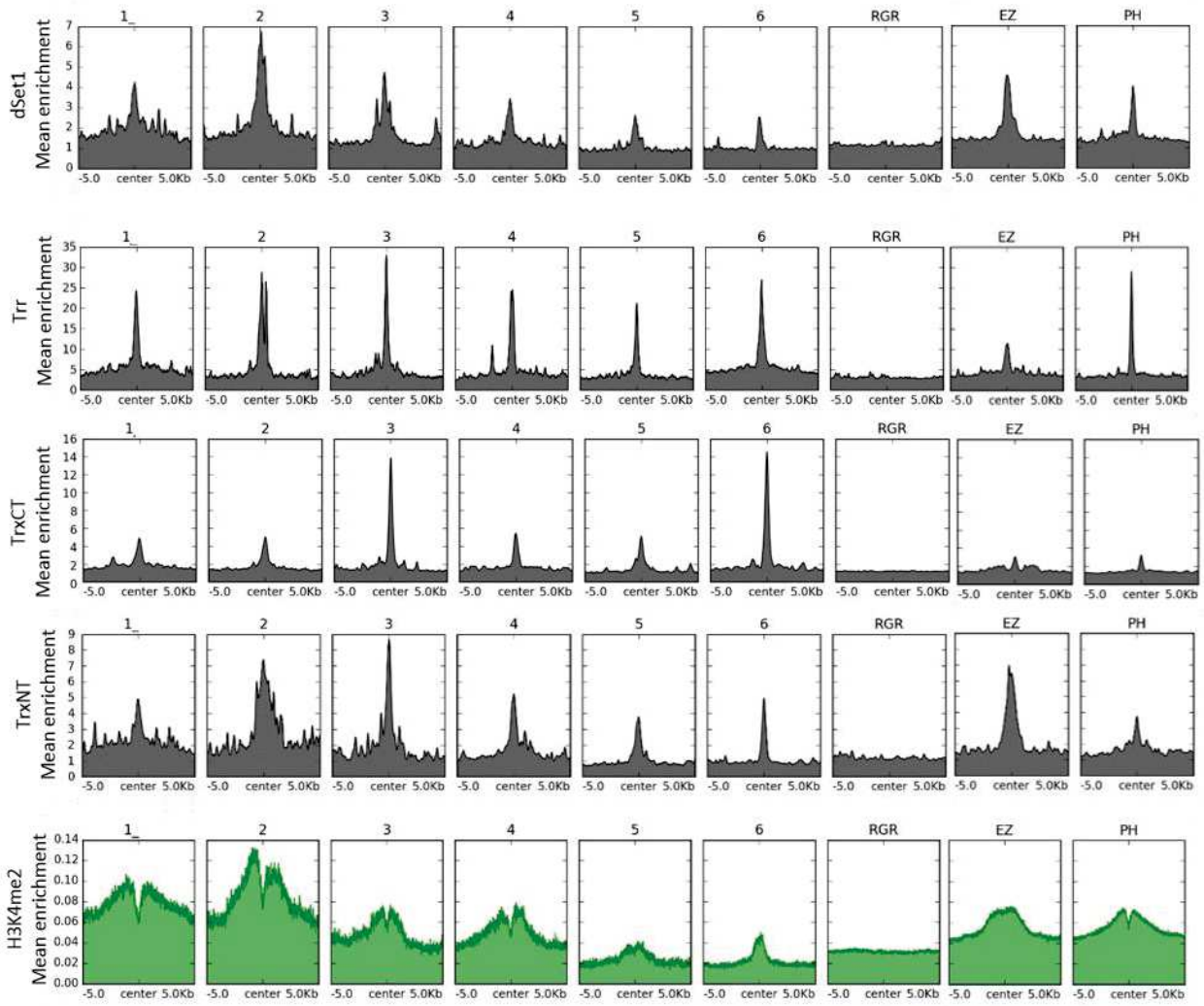


Figure 29 : Mean enrichment of various proteins associated to TrxG and H3K4me2 over a 10kb window.

Mean enrichments of dSET, Trr, TrxC and TrxN over a 10kb window around the mid of PREs. The mean enrichment for H3H4me2 is also given (green).

shows a significant enrichment of Pol II binding, whereas class 6 PREs does not show significant Pol II binding (**Figure 28 B**). This result is in concordance with a complementary analysis, measuring the levels of Pol II at these classes (**Figure 28 C**).

When looking at the expression level of neighboring genes of PREs from the six classes, we observed that classes 1-4 are strongly enriched in highly active genes and depleted in lowly expressed genes, which is in concordance with high levels of Pol II binding at these PRE classes (**Figure 28 D**). The same observation was done for E(z)- and Ph-only sites. In agreement with the absence of Pol II binding, class 6 PREs is associated with repressed genes. Most interestingly, class 5 PREs associates mainly with low or mid-lowly expressed genes, although Pol II levels are similar compared to active classes (compare **Figure 28 C and D**), supporting our hypothesis that the Pol II present at these sites could be in a poised state.

Although PREs of active classes 1-4 mainly correspond to TSSs, we noticed a high proportion of these PREs colocalizing with intra- or intergenic regions (**Figure 28 A**). We thus looked at the overlap with enhancers identified by the STARR-seq method (ARNOLD *et al.* 2013) (**Figure 28 E**). Overall, all classes of PREs along with both E(z)- and Ph-only sites have significant overlap with enhancers. However, we noticed that PREs of class 1 are most strongly enriched in enhancers (more than 60% of overlap with PREs from class 1). To complement this analysis we next analyzed levels of H3K4me1, a histone mark that is specifically associated with enhancer regions. The analyses of H3K4me1 showed a more defined classification (**Figure 28 F**). Indeed, both PRE classes 1 and 4 show a strong enrichment in the enhancer-associated mark along with H3K27ac (**Figure 25 C**), suggesting that those two classes are particularly enriched in active enhancers.

The fact that the majority of PREs associate with gene expression and active histone marks, prompted us to next analyze the binding of TrxG proteins mediating histone methyltransferase activity. In particular, we analyzed dSET1, which is known to maintain the bulk of H3K4me3 at active genes, Trithorax (Trx-N and Trx-C) which has been previously shown to mediate H3K4me2 at PREs contributing to the maintenance of the active state, and Trithorax-related (Trr,) an enzyme known to deposits H3K4me1 at enhancers (RICKELS *et al.* 2016). dSet1 binding levels (**Figure 28 G** and **Figure 29**) correlate well with levels of H3K4me3 (**Figure 25 E**) and is found most strongly enriched at PRE classes 1-3 and to a less extend at class 4. In contrast, its enrichment at classes 5 and 6 is weak, correlating with low levels of H3K4me3. The Trx protein is proteolytically cleaved by Taspase1, generating an N-terminal and a C-terminal fragment which can heterodimerize in vitro (HSIEH *et al.* 2003a; HSIEH *et al.* 2003b). Previous ChIP-on-chip studies showed that the C-terminal fragment, which contains the SET domain is more strongly enriched at embryonic PREs, whereas the N-terminal fragment preferentially associates with active promoters in embryos (SCHUETTENGRUBER *et al.* 2009). In agreement with these previous observations, we found that Trx-C is

strongly enriched at canonical PREs (class 6), although it is found also significantly enriched at active PRE classes (**Figure 28 H** and **Figure 29**). In contrast, the levels of Trx-N correlates better with the levels of H3K4me2 and is found predominantly enriched at active PRE classes (**Figure 28 I** and **Figure 29**). The presence of dSet1, Trx-N and their associated histone mark H3K4me2/3 at classes 1-4 suggests that they could correspond to PREs that mediate gene activation and are therefore in a TRE-like state. Finally, we analyzed enrichment level of Trr, an enzyme known to deposits H3K4me1 at enhancers. Similar to dSet1 and Trx-N, Trr is most strongly enriched at classes 1-4 (**Figure 28 J** and **Figure 29**), which show the highest overlap with STARR-seq enhancers (**Figure 28 E**).

Next, we analyzed the distribution of PRE classes in relation to the different chromatin colors defining different active or repressive chromatin environments and regulatory regions (KHARCHENKO *et al.* 2011). In agreement with the observations described above all classes except class 6 show a high overlap with the chromatin colours defined by active enhancers. Class 2 and to a less extend also class 1, 3 and 4 are enriched in red chromatin defined by active TSSs, whereas class 6 and to a less extent class 5 overlap with blue chromatin colour representing PcG-repressed chromatin states (**Figure 28 K**).

In summary, the analysis of the genomic features associated with the different PRE classes revealed potential differences in the functional properties of PREs and allowed us to redefine several PRE classes as follows:

REPRESSED PRE CLASSES:

Canonical PREs (class 6) fit the characteristics of previously described PREs in Drosophila embryos: They are associated with both TSSs and intergenic regions of repressed genes that are devoid of Pol II. Class 5 PREs show a similar genomic distribution than canonical PREs, and are generally associated with Pol II but lowly transcribed, providing further evidence that these PREs may correspond to poised DNA regulatory elements (TSS and enhancers).

ACTIVE PRE CLASSES:

Both classes (1 and 2) are particularly enriched in TSSs and enhancers. However, class 1 shows higher enrichment of H3K4me1 and lower enrichment of H3K4me3, whereas class 2 shows higher levels of H3K4me3 and lower levels of H3K4me1. *Vice versa*, class 1 shows the highest overlap with enhancer regions and a slightly lower overlap with TSSs, whereas class 2 shows the highest overlap with TSSs, and a slightly lower overlap with enhancer regions. We therefore named PRE class 1 as “active enhancer class” and PRE class 2 as “active TSS class”.

CHROMATIN BORDER PRE CLASSES:

Similar to active PRE classes, class 3 and 4 PREs are also associated with regulatory regions (TSSs and enhancers) of actively transcribed genes. To better characterize these two classes we analyzed the distribution of insulator proteins, usually associated with domain borders, which is described in the following paragraph.

Figure 30

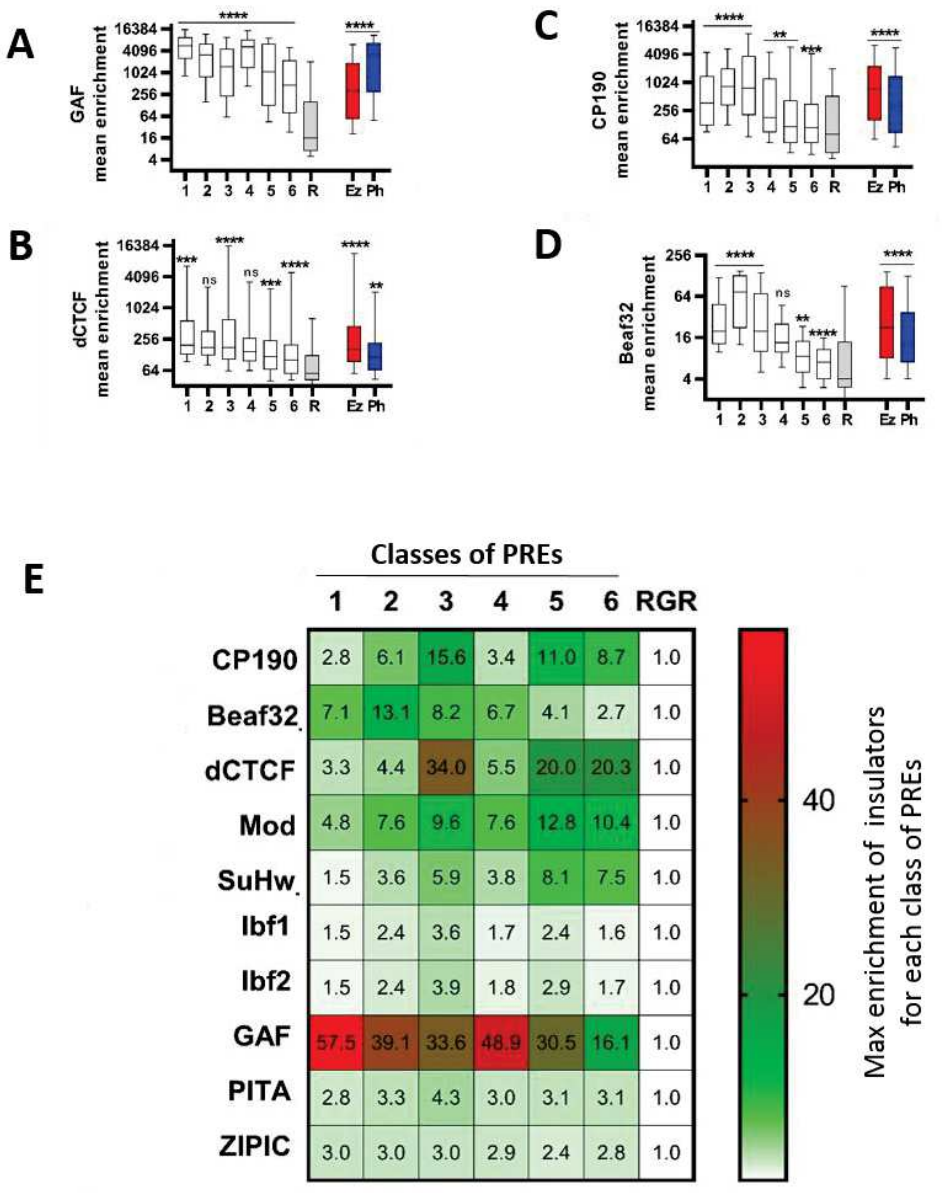


Figure 30 : The 6 classes of PREs differ by distincts signatures in insulator proteins

A-D: Boxplots showing the mean enrichment of various insulator proteins within a 10kb window around the mid of PREs.

E: The colored heatmap summarizes the max enrichment level of insulator proteins of each class of PREs. The maximum value of enrichment was identified for each insulator within a 10kb window centered on PREs. These values were divided by the maximum values obtained for each insulator for RGR class. An example of interpretation is that the enrichment in Beaf32 is approximately two fold higher in class 2 compared to other classes.

D. Distinct PRE classes are specified by specific combination of Insulator proteins and 3D interactions

Insulators are architectural DNA-binding proteins that control the interactions among genomic regulatory elements and act as chromatin boundaries. Genome wide mapping of insulator proteins and the identification of topologically associating chromatin domains (TADs) using Hi-C approaches indicated an important role of insulator proteins, in particular CTCF, in the definition of TAD borders (PHILLIPS-CREMINS and CORCES 2013; VAN BORTLE *et al.* 2014; BONEV and CAVALLI 2016). We mapped the mean enrichment of ten proteins known to have insulator properties (Beaf32; CP190; dCTCF; Su(Hw); Mod(mdg4), Ibf1, Ibf2, Gaf, Pita and Zipic) in a window centered on PREs (data not shown) and calculated the enrichment levels for each class of PREs (**Figure 30 A-D**). **Figure 30 E** summarizes the enrichment levels for each insulator protein in each class of PREs. Interestingly, each class is marked by a specific signature of insulator proteins. All active PRE classes (in particular class 1 and 4) are highly enriched in GAF binding (**Figure 30 A and E**), a PcG recruiters known for its insulator properties (MELNIKOVA *et al.* 2004; PETRASCHECK *et al.* 2005; FUDA *et al.* 2015). In contrast, GAF binding is also observed at PRE classes 5 and 6, however the mean enrichment levels are lower compared to the active class. CTCF is also significantly enriched at repressed PRE classes 5 and 6 (**Figure 30 B and E**), which is in agreement with previous findings in Drosophila embryos showing that PREs at the Hox gene clusters frequently associate with insulator elements bound by CTCF (HOLOHAN *et al.* 2007). However, the highest enrichment levels of CTCF and its cofactor CP190 was observed for PREs of class 3 (**Figure 30 B, C and E**). Since this class is associated with domain borders separating repressive H3K27me3 domains and active domains (marked by H3K27ac and H3K36me3), this observations suggest that these regions correspond to *bona fide* TAD boundaries. Intriguingly, in contrast to class 3 PREs, class 4 PREs, which are also associated with domain borders between active and repressed chromatin domains are not specifically enriched in insulator proteins (except GAF) similar to the other active PRE classes 1 and 2. Finally, class 2 is characterized by a stronger enrichment in Beaf32 compared to other classes (**Figure 30 D and E**). The other insulators studied here (Ibf1, Ibf2, Pita and Zipic) are lowly enriched in all classes of PREs. We did not find any signature in insulator proteins for E(z)- and Ph-only sites.

Figure 31

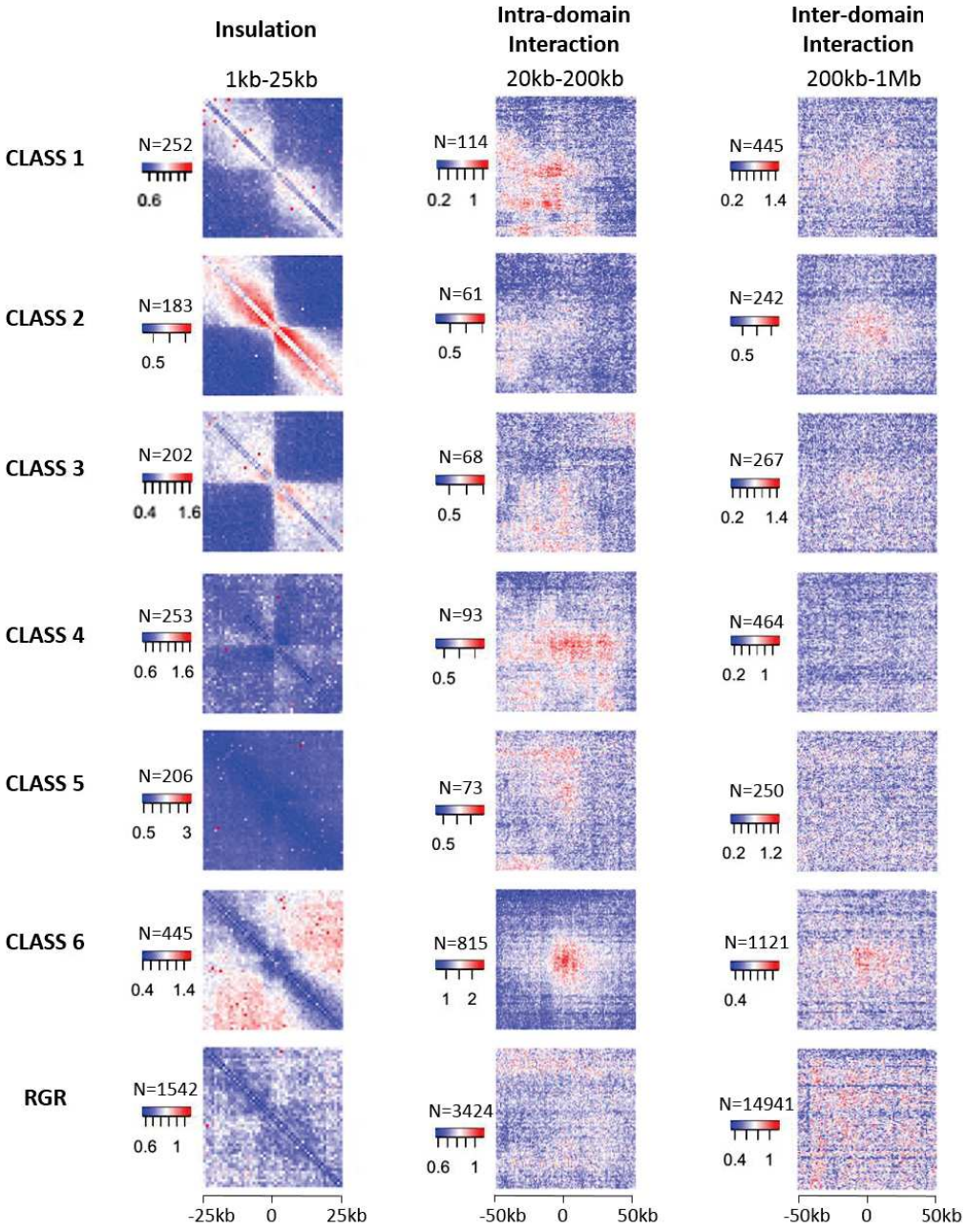


Figure 31 : The 6 classes of PREs are specified by different chromatin architectural characteristics

Paired-end spatial chromatin analysis (PE-SCAn) at various distances. From 1kb-25kb, PE scans indicate the insulation properties of each class. From 20kb-200kb, PE scans indicate the short-range interactions among PREs of each class. Finally, from 200kb-1Mb, PE scans indicate the long-range interactions among PREs of each class. Hi-C were obtained from embryos (Ogiyama Yuki et al., unpublished data).

PREs have been previously shown to interact with each other and with gene regulatory regions within Polycomb domains (LANZUOLO *et al.* 2007) and these interactions might be important for PcG target gene regulation. TAD boundaries are generally cell type invariant and stable throughout the lifespan of an organism (DIXON *et al.* 2012). We therefore took advantage of a high-resolution Hi-C data set from late embryos (unpublished data; Ogiyama Yuki et al.) and analyzed chromatin interaction of the different PRE classes.

Since class 3 is characterized by PREs located at domain borders between repressed and active regions (**Figure 25 A**), and is enriched in insulator proteins (**Figure 30**), we first sought for characteristics in relation to TAD boundaries. To look for insulation properties, we mapped very short-range interactions between PREs of each class (**Figure 31**; 1-25kb). Indeed, we observed insulation properties for PRE class 3, as it is the case for the active PRE classes 2 and to a weaker extent for class 1. In contrast, we did not see any insulation properties for repressed PRE classes 5 and 6. Surprisingly, we also could not detect insulation properties for class 4, indicating that these sites are not associated with real (*bona-fide*) domain borders.

We then analyzed short-range (intra-domain) chromatin interactions (**Figure 31**; 20-200kb) among PREs of each individual PRE class: Comparison of the interaction maps around PREs from class 6 revealed a significant accumulation of interactions, indicating that canonical PREs preferentially interact with each other as previously reported for PREs of Drosophila embryos. Some intra-domain interactions could also be observed for PRE classes 1 and 4, although the accumulation of interactions is less clear. At larger distances (inter-domain) (**Figure 31**; 200kb to 1Mb), the interaction frequencies between canonical PREs decrease, whereas we started to see some interactions at long distances for active PRE class 2.

Taken together, the analysis of insulator association and 3D chromatin interactions of the different PRE classes shed some more light on the different PRE properties that are summarized below and in **Figure 39**.

Class 6 (Canonical PREs): These PREs correspond to the previously described PREs of Drosophila embryos, located within repressive Polycomb domains marked by H3K27me3, associated with repressed genes lacking Pol II and characterized by a significant accumulation of interactions between PREs at short ranges within these domains. As it is the case for PREs in Drosophila embryos, these PREs can be found either at TSSs or at intergenic regions and are frequently colocalize with the insulator proteins CTCF and to a lesser extend Mod(mdg4) (**Figure 30**).

Page blanche

Class 5 (Poised PREs): These PREs associated with active and repressed histone marks are likely in a poised state. They are bound by Pol II but correspond to lowly transcribed genes. They show similar distribution around genes and insulator-binding characteristics compared to canonical PREs, but do not show significant chromatin interactions.

Class 4 (Active Enhancer class): This class is highly enriched in active enhancers (presence of H3K4me1 and H3K27ac). Although they are located between repressive and active chromatin marks, these PREs do not associate particularly with any insulator protein, and they do not show insulation activity, arguing that these sites do not correspond to bona fide domain borders. Similar to the active enhancer class 1, class 4 PREs show some signs of intra-domain interactions.

Class 3 (Domain border PREs): This PRE class preferentially associates with CTCF and CP190, has insulation properties and correspond to regulatory regions of active genes. Thus, these PREs show all features of typical chromatin domain boundaries. No significant chromatin interactions within these PRE class are detected.

Class 2 (Active TSS class): This class is characterized by a higher ratio of H3K4me3 vs H3K4me1. PREs are associated with TTSs (which can also act as enhancers in many cases) of highly active genes. They are preferentially bound by Beaf32 and shows strong insulation activity as it has been previously reported for highly active genes (DIXON *et al.* 2012). In addition, a weak accumulation of long range contacts between these PREs can be detected.

Class 1 (Active Enhancer class): This class is characterized by a higher ratio of H3K4me1 vs H3K4me3. PREs associates with active genes and is strongly associated with active enhancers. This PRE class is preferentially bound by GAF, shows weak insulation activity and shows some accumulation of short range contacts as detected for PRE class 4.

Figure 32

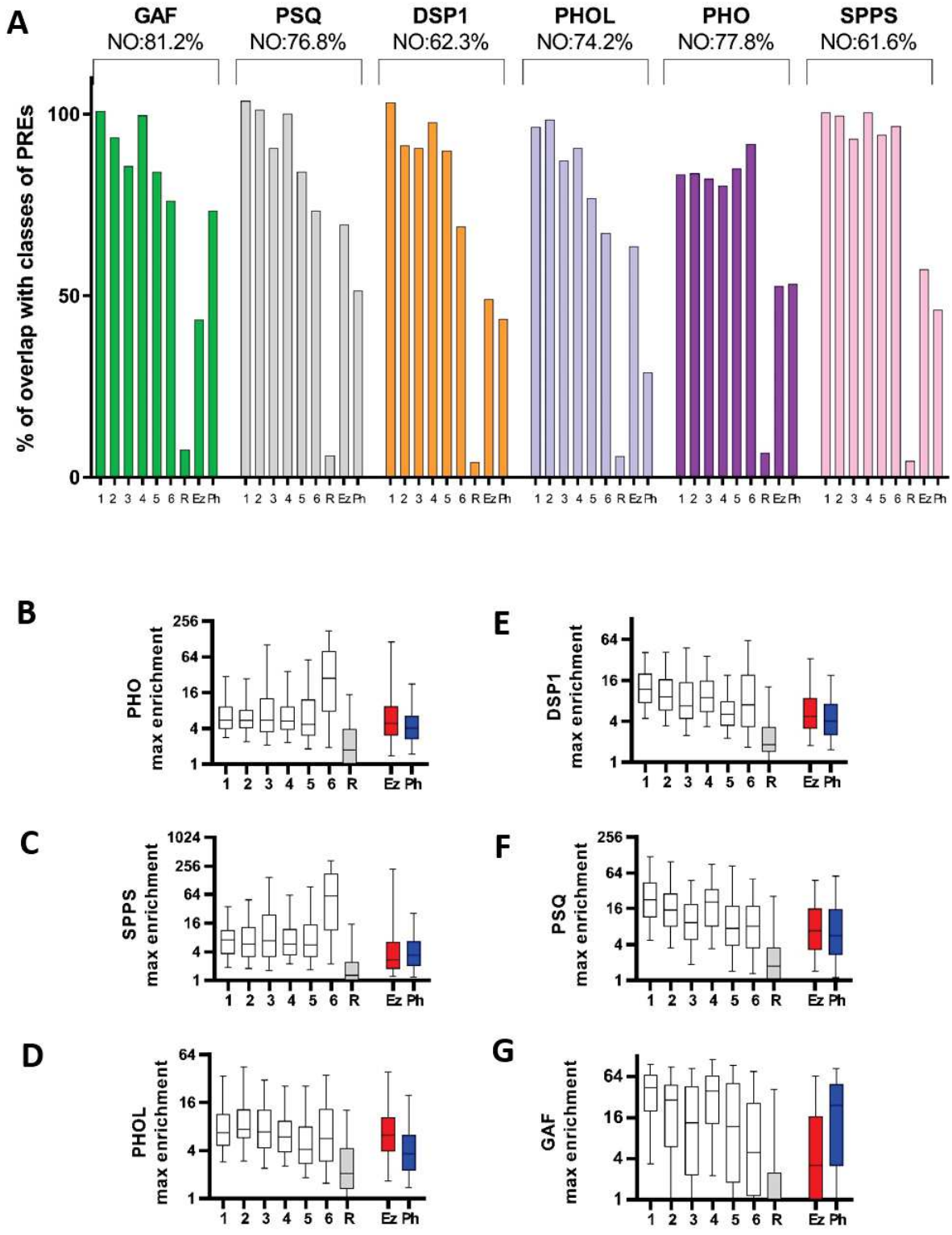

A
GAF
NO:81.2%
PSQ
NO:76.8%
DSP1
NO:62.3%
PHOL
NO:74.2%
PHO
NO:77.8%
SPPS
NO:61.6%
% of overlap with classes of PREs
100
50
0
1 2 3 4 5 6 R Ez Ph
B
PHO
max enrichment
256
64
16
4
1
1 2 3 4 5 6 R Ez Ph
E
DSP1
max enrichment
64
16
4
1
1 2 3 4 5 6 R Ez Ph
C
SPPS
max enrichment
1024
256
64
16
4
1
1 2 3 4 5 6 R Ez Ph
F
PSQ
max enrichment
256
64
16
4
1
1 2 3 4 5 6 R Ez Ph
D
PHOL
max enrichment
64
16
4
1
1 2 3 4 5 6 R Ez Ph
G
GAF
max enrichment
64
16
4
1
1 2 3 4 5 6 R Ez Ph

Figure 32 : Characterization of PcG recruiters profiles at the six classes of PREs

A.Proportion of each class of PREs colocalizing with various PcG recruiters (Gaf, Psq, Dsp1, Phol, Pho and Spps). NO indicates the percentage of TF peaks that do not overlap with the 1543 E(z)/Ph sites.

B-G. Boxplots show the max enrichment of the TF for each class within a 10kb window.

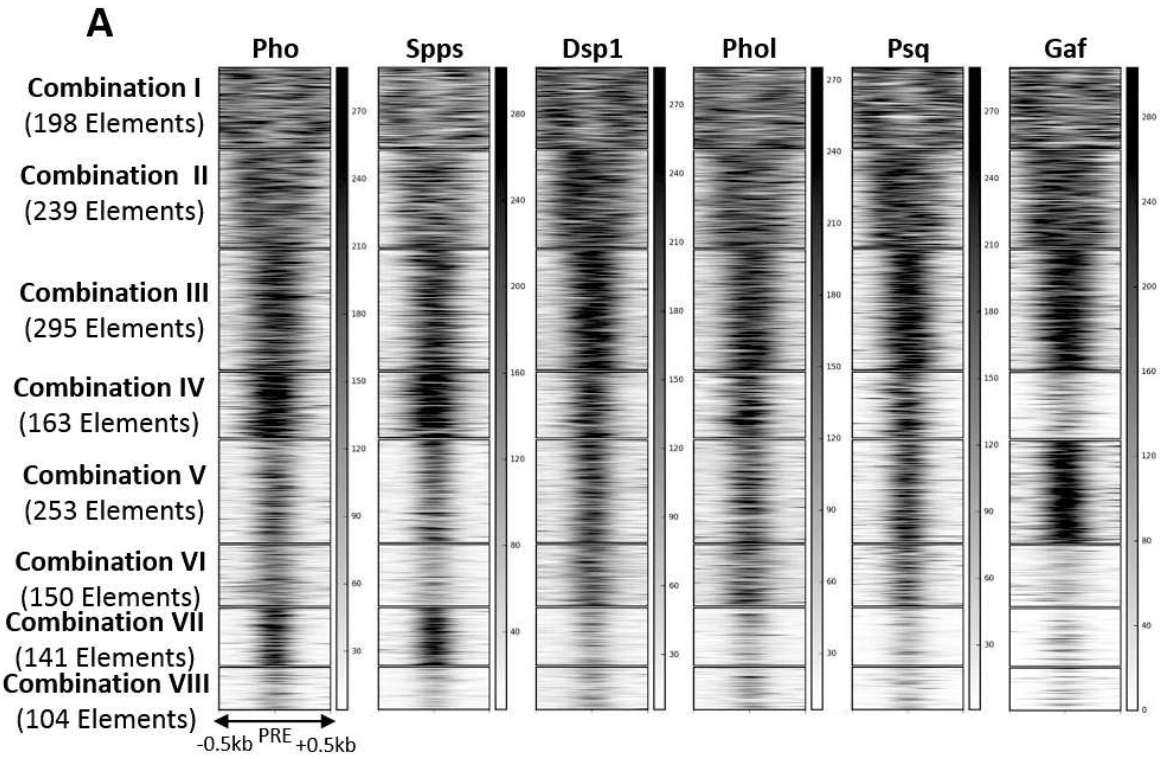
E. PRE classes are defined by a distinct combination of sequence specific DNA binding proteins

As described in the introduction section, recruitment of PcG complexes is mediated by a combinatorial action of several sequence-specific DNA-binding proteins such as Pleiohomeotic (PHO), its homologue Pleiohomeotic-like (PHOL), GAGA factor (GAF), Pipsqueak (PSQ), Dorsal switch protein (DSP1) and SPPS (SP1/KLF). Importantly, none of these PcG recruiters on their own are sufficient to recruit PcG complexes to PREs and the majority of these transcription factors (TFs) is involved in transcriptional activation as well as repression.

To address the question, whether different classes of PREs uses a different combinations of transcription factors (TFs) to recruit PcG complexes, we first analyzed the binding of TFs at the PRE classes 1-6. The correlation of TF binding and PcG proteins PH and E(z) (**Figure 24 D**) gave a first hint for different behavior of TF binding at PcG binding sites, since Pho and Spps are more correlated with PcG proteins than GAF, PSQ, DSP1 and PHOL (**Figure 24 D**). When looking at the proportion of overlap between PcG recruiters and the six classes of PREs, we noticed an extremely high overlap between TF binding and all six PRE classes (**Figure 32 A**). Gaf, Psq, Phol and Dsp1 show a very similar pattern: they are present at 80% to 100% of active PRE classes 1 to 4 and class 5, whereas the overlap of these three TFs with class 6 PREs is slightly reduced (60-70%). Further, we noticed that neither Pho nor Spps do specify a particular class of PREs, but they both are similarly distributed in all six classes (around 80% for PHO and 90-100% for SPPS for all classes).

To determine whether quantitative differences of TF binding can discriminate the six PRE classes, we next analyzed the enrichment levels of TFs at different classes (**Figure 32 B-G**). This analysis confirmed that Pho and Spps share a similar enrichment pattern. Although they are found at the same number of PREs in each class, both proteins are more strongly enriched at canonical PREs compared to PRE classes 1-5 (**Figure 32 B and C**).

Figure 33


A
Pho
Spps
Dsp1
Phol
Psq
Gaf
Combination I
(198 Elements)
Combination II
(239 Elements)
Combination III
(295 Elements)
Combination IV
(163 Elements)
Combination V
(253 Elements)
Combination VI
(150 Elements)
Combination VII
(141 Elements)
Combination VIII
(104 Elements)
-0.5kb PRE +0.5kb

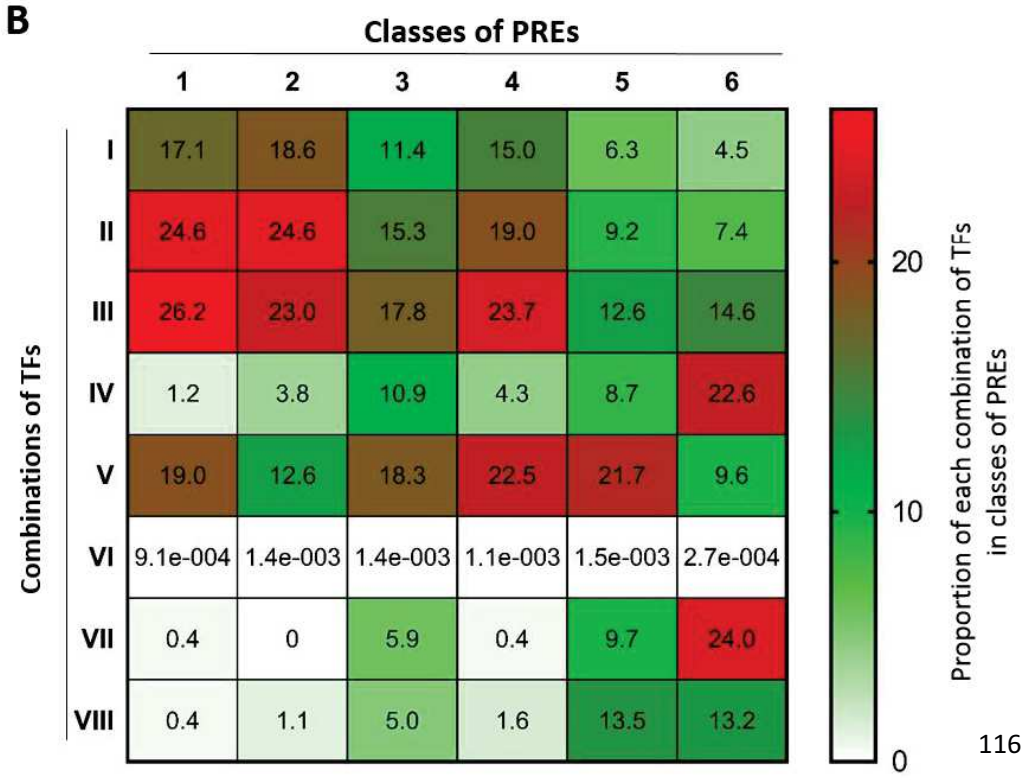

B
Classes of PREs
1 2 3 4 5 6
Combinations of TFs
I 17.1 18.6 11.4 15.0 6.3 4.5
II 24.6 24.6 15.3 19.0 9.2 7.4
III 26.2 23.0 17.8 23.7 12.6 14.6
IV 1.2 3.8 10.9 4.3 8.7 22.6
V 19.0 12.6 18.3 22.5 21.7 9.6
VI 9.1e-004 1.4e-003 1.4e-003 1.1e-003 1.5e-003 2.7e-004
VII 0.4 0 5.9 0.4 9.7 24.0
VIII 0.4 1.1 5.0 1.6 13.5 13.2
20
10
0
Proportion of each combination of TFs in classes of PREs

Figure 33 : The 6 classes of PREs are specified by distinct combinations in PcG recruiters

A. Heatmap showing the eight combinations of TFs found in the 1543 E(z)/Ph sites. Two superclusters can be distinguished (see main text).
B. The colored heatmap indicates the proportion of each combination of TFs found in the clusters of PREs (red is a high proportion, green indicates a mild proportion and white signifies a low proportion). The percentages of a given combination of TFs within each class of PREs is indicated.

These results confirm the initial observation of a high correlation found for these two TFs and PcG proteins (**Figure 24 D**) and suggest that PHO and SPPS play a major role in the recruitment of PcG proteins to the canonical PRE class 6. Psq, Dsp1 and Phol are similarly enriched in all classes of PREs (**Figure 32 D, E and F**). In contrast, Gaf binding to the canonical PRE class 6 is generally lower compared to the other classes (**Figure 32 G**). Overall, these results indicate that high levels of, both, Pho and Spps specify canonical PREs. *Vice versa* higher levels of GAF mark the active PRE classes. However, this analysis does not allow to define particular combinations of other TFs for each class of PREs.

In order to characterize particular combinations of PcG recruiter proteins enriched at the different classes of PREs that might specify each classes of PREs, we used the 1543 E(z)/Ph sites as input for another clustering analysis, using Pho, Spps, Phol, Dsp1, Psq and Gaf enrichments as variables. Eight clusters were identified (Combinations I-VIII, **Figure 33 A**), which can be summarized into two "superclusters": Supercluster A comprises clusters I, II, III and V all characterized by high levels of GAF. Combinations I, II and III are similarly enriched in the six TFs and differ by the peak lengths (from broad, Combination I to sharp, Combination III), whereas Combination V shows a stronger enrichment in GAF protein compared to the other TFs. Supercluster B comprises the clusters IV, VI, VII and VIII and is defined by low levels of GAF. Combinations IV and VII are characterized by higher levels of PHO and SSPS compared to the other recruiters, whereas combination VI is particularly enriched for DSP1, PHOL and PSQ. Finally, combination VIII is characterized by a low enrichment for all TFs.

In order to assign particular combinations of TFs to each class of PREs, we intersected the two clustering analyses (**Figure 33 B**). Active PRE classes 1-4 share a strong enrichment for combinations I, II, III and V, representing supercluster A. The combination of TFs of class 5 resembles mostly that of class 3 with a higher enrichment for combination V. Very unlike the five other PRE classes, the canonical PRE class 6 is characterized by a high enrichment for combinations IV and VII including a high ratio of both Pho and Spps proteins compared to other recruiters. Finally, we noticed that the combination VI distinguished by a high ratio of Dsp1, Phol and Psq compared to other proteins does not define any classes of PREs by itself.

Page blanche

Overall, these results indicate that the 6 classes of PREs are characterized by two distinct combinations of PcG recruiters. Active PRE classes 1-4 are characterized by a similar combination of PcG recruiters and are generally marked by the presence of all TFs including Pho, Spps, Dsp1, Phol, Psq and Gaf. At these PREs TF can exhibit various shapes, forming broad to sharp peaks. Although classes 5 PREs associate with lowly repressed genes and represent a putative poised state, it shares similar TF binding characteristics than the active PRE classes. Finally, the class 6 corresponding to canonical PREs differs from the others by a strong enrichment in both Pho and Spps proteins and by lower levels of GAF.

Figure 34

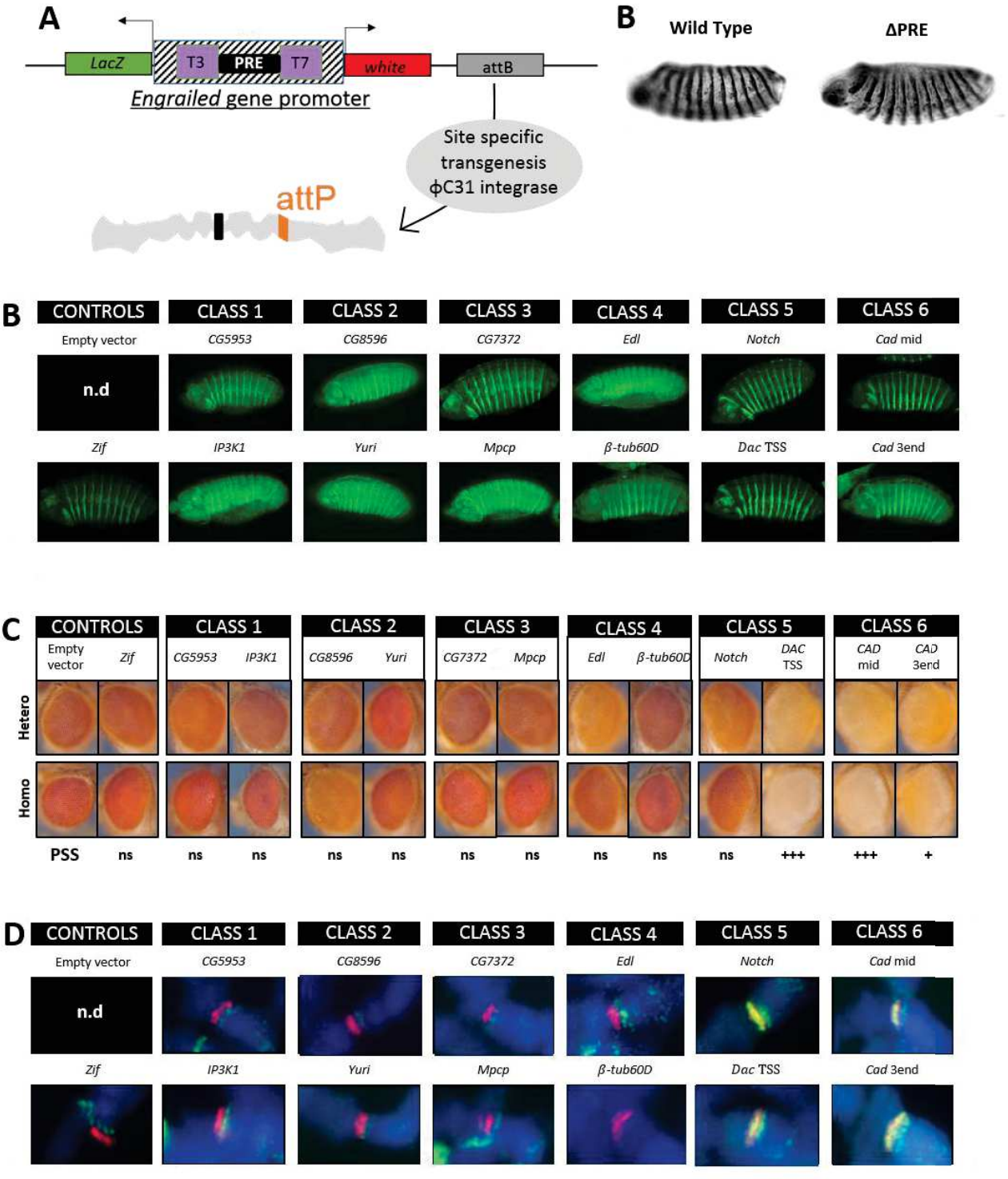

A
LacZ
T3
PRE
T7
white
attB
Engrailed gene promoter
Site specific transgenesis ϕC31 integrase
attP
B
Wild Type
ΔPRE
B
CONTROLS
CLASS 1
CLASS 2
CLASS 3
CLASS 4
CLASS 5
CLASS 6
Empty vector
CG5953
CG8596
CG7372
Edl
Notch
Cad mid
n.d
Zif
IP3K1
Yuri
Mpcp
β-tub60D
Dac TSS
Cad 3end
C
CONTROLS
CLASS 1
CLASS 2
CLASS 3
CLASS 4
CLASS 5
CLASS 6
Empty vector
Zif
CG5953
IP3K1
CG8596
Yuri
CG7372
Mpcp
Edl
β-tub60D
Notch
DAC TSS
CAD mid
CAD 3end
Hetero
Homo
PSS
ns
ns
ns
ns
ns
ns
ns
ns
ns
ns
ns
+++
+++
+
D
CONTROLS
CLASS 1
CLASS 2
CLASS 3
CLASS 4
CLASS 5
CLASS 6
Empty vector
CG5953
CG8596
CG7372
Edl
Notch
Cad mid
n.d
Zif
IP3K1
Yuri
Mpcp
β-tub60D
Dac TSS
Cad 3end

Figure 34 : Functional characterization of canonical and non-canonical PREs in flies

A. Scheme of the SD10 vector used to test PRE activity. Each region to test was inserted into the *engrailed* regulatory region upstream of two reporter genes: the *LacZ* gene allows to test PRE activity in embryos and the *mini-white* gene in adults. Importantly, two exogenic sequences T3 and T7 were inserted at either side of tested regions to allow q-ChIP experiments at the transgenic sites (see text). Finally, the vector contains an *attB* site allowing the site specific integration of the constructs at *attP2* site (see M&M).

B. Pattern of *LacZ* expression in WT condition (with PRE) or in absence of PRE (staining with β-Galactosidase antibody from (CUNNINGHAM *et al.* 2010)).

C. Functional PRE test in homozygous embryos (stage 14). Two representative PREs of each class were tested along with one negative controls (vector in which the *zif* promoter was inserted). PRE activity can be assessed by comparing the *en*-like-pattern of *LacZ* in stripes between controls and transgenic lines (see main text).

D. The comparison of eye pigmentation between heterozygous and homozygous transgenic lines allows the calculation of the PSS (Pairing Sensitive Silencing) (see main text).

E. Immuno-FISH on polytene chromosomes. Ph is stained in green while the transgene is visualized in red. A recruitment of Ph by the transgene gives a yellow color as seen for PREs from classes 5 and 6. Note that an endogenous Ph band is detected close (but not overlapping) to the transgene insertion site in the control (ZIF) line.

II. FUNCTIONAL VALIDATION OF THE SIX CLASSES OF PREs

As described above, we identified six different classes of PREs characterized by distinct chromatin signatures, genomic characteristics and specific DNA-binding protein signatures. In order to validate PRE activity *in vivo* and to analyze if they have different functional properties, we tested these PRE classes in transgenic reporter gene assays in flies using the site specific integration system based on the bacteriophage ΦC31 integrase. ΦC31 integrase mediated transgenesis uses a single *attP* docking site (acceptor site) to integrate in the fly genome, which can be used to site-specifically integrate an *attB*-containing plasmid (reviewed in (VENKEN and BELLEN 2007)). Therefore we generated a customized plasmid based on the so called "SD10 plasmid", previously used to test PRE activity (DEVIDO *et al.* 2008) (**Figure 34 A**). This plasmid contains an *engrailed* enhancer and promoter regulatory region upstream of the *LacZ* reporter gene in one direction and a *mini-white* reporter gene in the other direction (**Figure 34 A**). The *engrailed* regulatory region drives *LacZ* expression within 14 stripes in the drosophila embryo that resembles expression of the endogenous *engrailed* gene. Importantly the endogenous *engrailed* regulatory region contains a PRE required for maintaining repression between the stripes, which has been deleted in the plasmid. Therefore, the engrailed sequence in the plasmid should not have PRE activity resulting in an ectopic expression of *LacZ* between stripes (as previously shown by (DEVIDO *et al.* 2008)) (**Figure 34 B**). In contrast, integration of another sequence containing PRE

activity results in the maintenance of the restricted *en*-like expression pattern of the *LacZ* gene (CUNNINGHAM *et al.* 2010; ABED *et al.* 2013).

We modified this plasmid by inserting an *attB* site to be able to site specifically integrate the plasmid to exclude positional effects and to better compare different PREs with each other. In addition, we inserted a barcode composed of T3 and T7 primer sequences up and downstream of a multiple cloning site (MCS), where we inserted the different classes of PREs. This barcode should allow us to analyze PcG binding at the transgenic PRE, without detecting the endogenous PRE copy. Two candidates of each PRE class were selected based on representative criteria (**Figure 25 G and 27 A**). Each of them was inserted into the MCS of the modified SD10 vector and the plasmids were injected into fly embryos containing the *attp2* docking site located at chromosome 3L to generate transgenic fly lines carrying the different PRE classes inserted at the same genomic position. In addition we used a sequence from the *ZIF* gene promoter as control, which is known not to be bound by PcG proteins at any developmental stage (SCHUETTENGRUBER *et al.* 2009; SCHUETTENGRUBER *et al.* 2014). To analyze expression of the *LacZ* reporter gene we performed immuno-staining experiments of stage 14 embryos to assess for b-galactosidase expression (**Figure 34 C**). Insertion of the *Zif* control sequence resulted in a slight misexpression of *LacZ* between the strips (**Figure 34 C**). However, by comparing the *en*-like patterns between *Zif*-containing transgene and the PRE-inserted transgenes, we noticed various differences. First, we observed that all PREs from active classes (1-4) showed a global activation of *LacZ* expression in the regions between the stripes compared to the control line (**Figure 34 C**). Contrarily, PREs derived from classes 5 and 6 generally showed a more restricted *en*-like pattern (PREs from *Notch*, *DAC* genes and PRE named *Cad*-mid), except the PRE located at the 3'end of the *Cad* gene. We concluded that PREs derived from classes 1-4 mediate gene activation rather than repression at the embryonic stage, whereas PREs from classes 5 and 6 generally have classical PRE activity in embryos. The differences in PRE activities observed for class 6 PREs suggest that some canonical PREs are more efficient than others to recruit PcG complexes (in our case, the PRE located at the middle of *Cad* gene seems to be more efficient than that of 3'end).

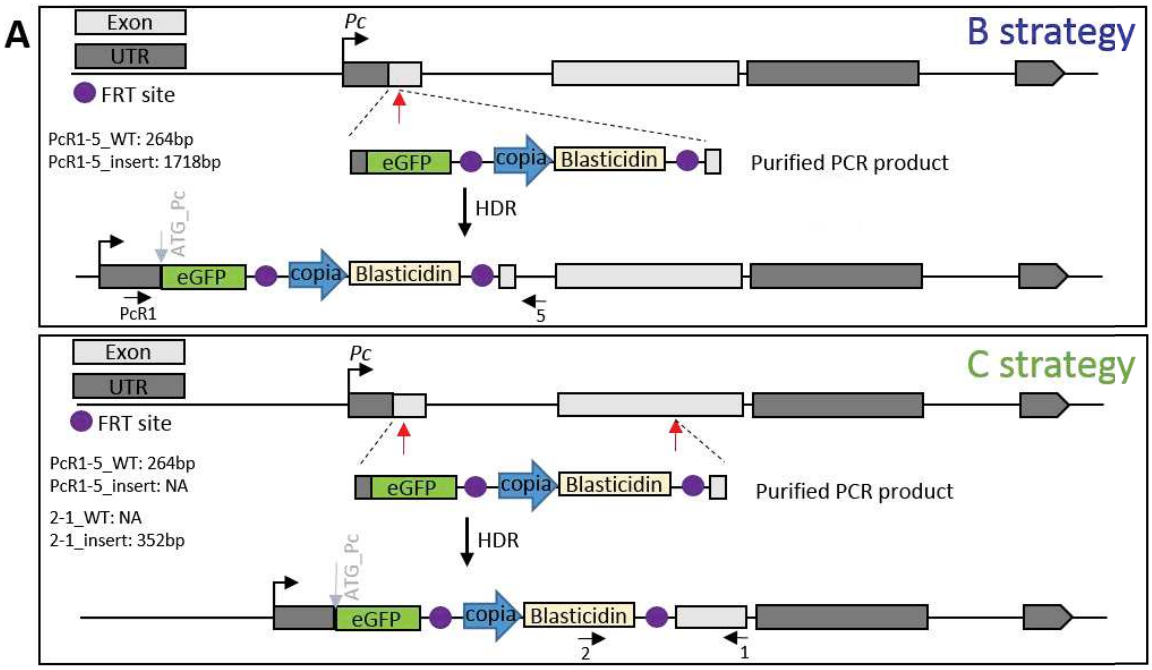
To test whether PREs from active classes could mediate repressive states at later development stages, we looked at the PRE activity in the adult fly. Indeed, as mentioned above the SD10 vector also contains the *mini-white* reporter gene (**Figure 34 A**). It was previously demonstrated that the insertion of a PRE upstream of the *mini-white* gene leads to a repression of the reporter gene, which is reflected by a lightning of the eye color of the fly. Furthermore, PREs exhibit a so called "pairing sensitive silencing" effect (PSS), meaning that repression of the *mini-white* reporter gene is even stronger when the transgene is present in a homozygous state compared to the heterozygous state. We thus compared and measured the eye pigmentation between

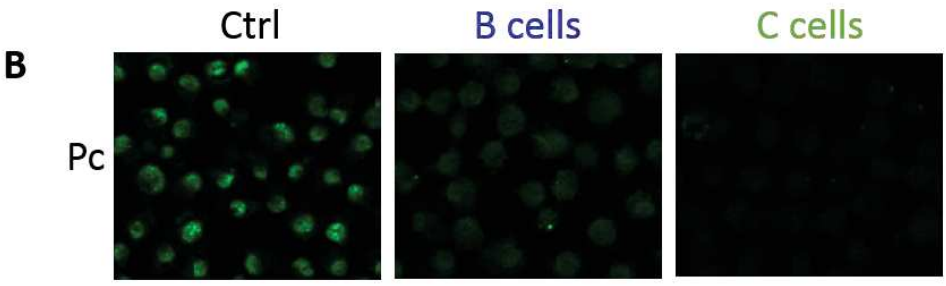
heterozygous and homozygous transgenic flies (**Figure 34 C**). Consistent with results obtained in embryos, none of the PREs from active classes 1-4 exhibit PSS activity, contrarily to most of the PREs from classes 5 and 6, except the PRE from the *Notch* gene locus, which was less efficient in mediating repression of the *mini-white* reporter gene (**Figure 34 C).**

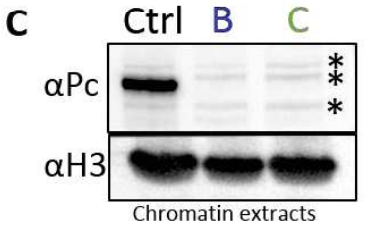
Finally, we tested the recruitment of PcG proteins to the transgenic PREs by performing Immuno-FISH experiments in Drosophila polytene chromosomes of 3rd instar larvae. Therefore, we used a DNA probe corresponding to the transgene insertion site (*attp2*) to detect the location of the transgene, combined with an immunostaining using Ph antibody. In agreement with observations made in embryos or adult flies, PREs from classes 1-4 are not associated with Polycomb in larval polytene chromosomes, whereas PREs from classes 5 and 6 recruit Polycomb to the transgene insertion site (**Figure 34 D).**

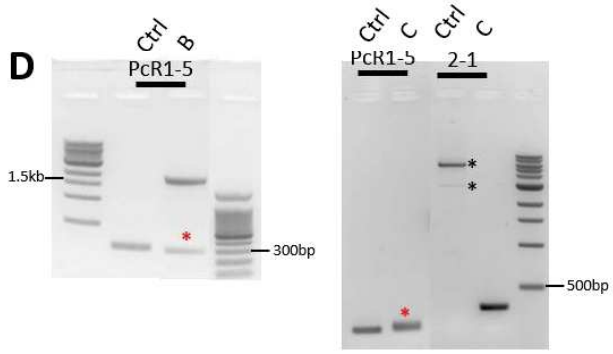
In conclusion, testing our six PRE classes *in vivo* using transgenic reporter genes, showed that non-canonical PREs from active classes 1-4 do not mediate gene silencing in the tested assays but rather mediate active states, indicating that they do not have classical PRE activity. PREs of class 5 and 6 do have PRE activity that seems to be maintained throughout the lifespan of the fly. However, our results suggest that some PREs (in particular class 5, e.g *Notch* PRE) could be timing-specific. In addition, PREs from active classes could act in a tissue specific manner, and mediate repression in only a small subset of cells at a certain developmental stage.

Figure 35


A
Exon
UTR
FRT site
Pc
B strategy
PcR1-5_WT: 264bp
PcR1-5_insert: 1718bp
eGFP
copia
Blasticidin
Purified PCR product
HDR
ATG_Pc
PcR1
5
C strategy
PcR1-5_WT: 264bp
PcR1-5_insert: NA
2-1_WT: NA
2-1_insert: 352bp
2
1


B
Ctrl
B cells
C cells
Pc


C
Ctrl
B
C
αPc
αH3
Chromatin extracts


D
Ctrl
B
PcR1-5
1.5kb
300bp
Ctrl
C
Ctrl
C
PcR1-5
2-1
500bp


E
C PcR1-5
5
primer
Ctrl
C cells
1981 GCTTTTGGATGATTTTCTCAGCCGCGTACACTAGATc--gACTGGATCGTCGGTCGCATt 2038
17 GCTTTTGGATGATTTTCTCAGCCGCGTACACTAGATCTAGACTGGATCGTC--------- 67

Figure 35 : Mutation of *Pc* gene in S2 cells using the CRISPR/Cas9 system

A. Two strategies were used to mutate *Pc* gene (see M&M and main text)
B. *Pc* mutant clones were tested by immunos-staining of the Pc protein. Cell lines B and C were selected for further experiments.
C. Western blotting of chromatin extracts from B and C cell lines. 10µg were loaded (n=2). * indicates non-specific bands.
D. Test of marker gene cassette insertion by PCR. In both mutant cell lines, a band with the WT allele size was amplified indicated that at least one allele of each cell line did not integrate the marker gene cassette. In order to verify the mutation of these WT allele size bands, they were sequenced (red*). Black* indicate non-specific bands.
E. Sequencing result for the WT allele size band of the C cell line. Blue indicates the exon 1 sequence. Red indicates the downstream gRNA sequence with the cg (in letter case) indicating the cutting site. We thus confirm that the WT allele size band amplified from C cell line contains a deletion.

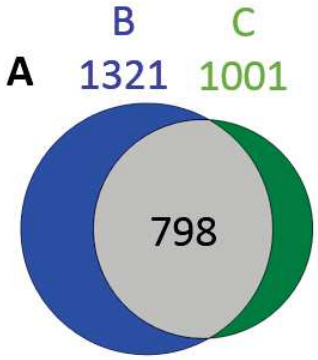
III. FUNCTIONAL ROLE OF POLYCOMB AT ACTIVE SITES

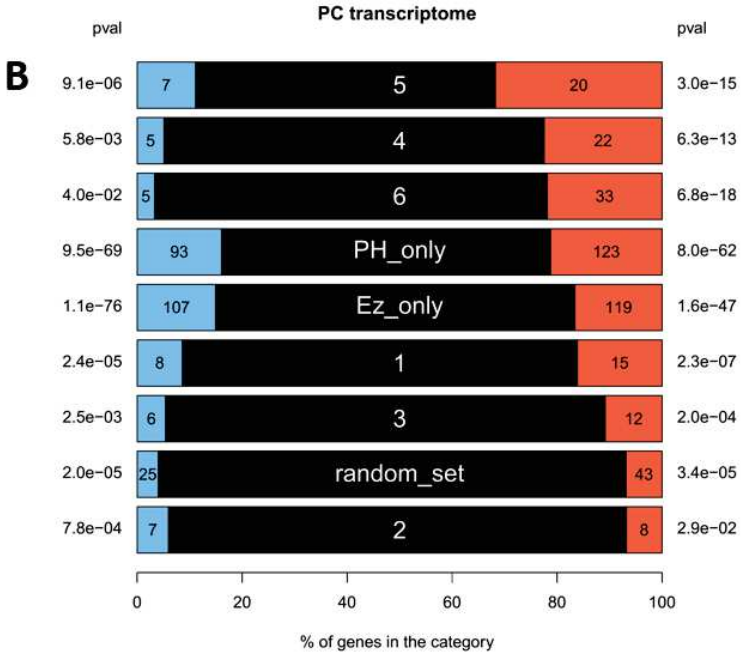
A. Creation of *Pc* mutant S2 cell lines using the CRISPR/Cas9 system

The observations described above that PREs from classes 1-4 are associated with gene activity, led us to test the role of PcG proteins at these active sites. To do so, we generated stable S2 cell lines (which are derived from embryos) carrying loss of function mutation of the *Polycomb* (*Pc*) gene using the CRISPR/Cas9 system. We decided to knock out *Pc* rather than Ph, because the Polyhometic (Ph) gene is duplicated, making it more difficult to delete both copies. In addition, because of the polyploidy of S2 cells, complete protein knock out by the genome engineering is more challenging. We thus started from S2 cells already expressing the Cas9 enzyme (BOTTCHER *et al.* 2014) and co-transfected them with a template DNA fragment and different guide RNAs depending on strategy B or C (see **Figure 35 A** and M&M). Briefly, both strategies lead to the insertion of the DNA template downstream of the *Pc* START codon disrupting the open reading frame, but the strategy C also leads to an excision of a part of exon 1 and 2 of *Pc* gene (**Figure 35 A**).

Isolated clones were validated for Pc depletion by several assays. First, immuno-staining of Pc protein revealed that 8 clones over the 89 tested, showed a strong depletion in Pc protein, which only showed background staining in the cytoplasm (**Figure 35 B** and data not shown).

Figure 36


A
B
C
1321
1001
798


B
PC transcriptome
pval
pval
9.1e−06
5.8e−03
4.0e−02
9.5e−69
1.1e−76
2.4e−05
2.5e−03
2.0e−05
7.8e−04
3.0e−15
6.3e−13
6.8e−18
8.0e−62
1.6e−47
2.3e−07
2.0e−04
3.4e−05
2.9e−02
7
5
20
5
4
22
5
6
33
93
PH_only
123
107
Ez_only
119
8
1
15
6
3
12
25
random_set
43
7
2
8
0
20
40
60
80
100
% of genes in the category

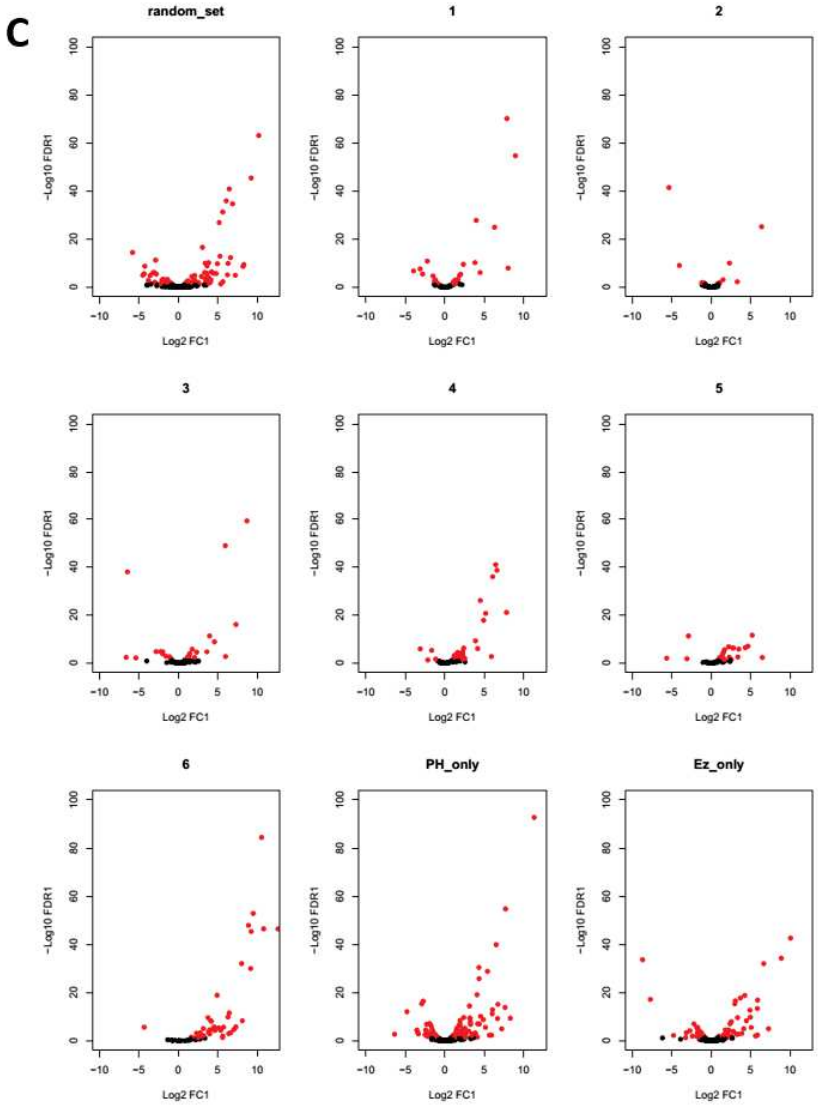

C
random_set
1
2
3
4
5
6
PH_only
Ez_only
−Log10 FDR1
Log2 FC1

Figure 36 : Transcriptomic results of RNA-seq performed on *Pc* mutant cell lines

A. Venn diagram showing the number of misexpressed genes in both cell lines. Further analyses were performed on the 798 commonly misexpressed in both cell lines.
B. Stacked bar plot showing the down-regulated (in blue) and up-regulated (in red) genes expressed as % (X axis) of all the genes in the given categories of PREs (Y axis). The number of up and down-regulated genes are shown in the colored bars, and the p-values corresponding to the hypergeometric test preformed on down- and up-regulated genes are shown on the left and right, respectively. Classes of PREs are indicated in the black region in the center of the bars.
C. Volcano plots performed from the B cell line indicate the log2 Fold change (X axis) and the log10 FRD (Y axis) of each misexpressed gene per class of PREs. Black and red dots indicate unsignificant or significant misexpressed genes, respectively.

Two clones were selected (clone B and C corresponding to the different strategy used to obtain them) and complete Pc protein depletion was further confirmed by Western-blot on chromatin extracts (**Figure 35 C**).

We also amplified the mutated region by PCR (**Figure 35 D**). Strikingly, although both cell lines show a correct insertion of the marker gene cassette, a fragment with a size similar to that of WT allele could also be amplified. Sequencing of these fragments confirmed that both *Pc* mutated cell lines B and C do not have the WT allele anymore, but carry small INDELs (**Figure 35 E** and data not showed). We thus concluded that the insertion of the marker gene cassette is not efficient at 100%, since some alleles are only cut and mutated by the action of gRNAs during the non-homologous end joining repair mechanism. Most importantly, INDELs result in the appearance of a STOP codon, leading to the disruption of the open reading frame and to the loss of Pc protein.

B. Transcriptome analyses of *Pc* mutant cell lines

In order to assess the role of Pc on the transcriptional regulation of genes associated with our six PRE classes, we performed RNA-seq analysis of our two *Pc* mutant cell lines and the control line (S2 cells expressing Cas9). Experiments were performed in triplicate and genes significantly deregulated in each mutant cell line compared to the control line were identified: A total of 1321 genes were misexpressed in mutant line B, whereas 1001 genes were deregulated in mutant line C. The vast majority of genes (798) were commonly found misexpressed in both mutant cell lines (**Figure 36 A**).

Page blanche

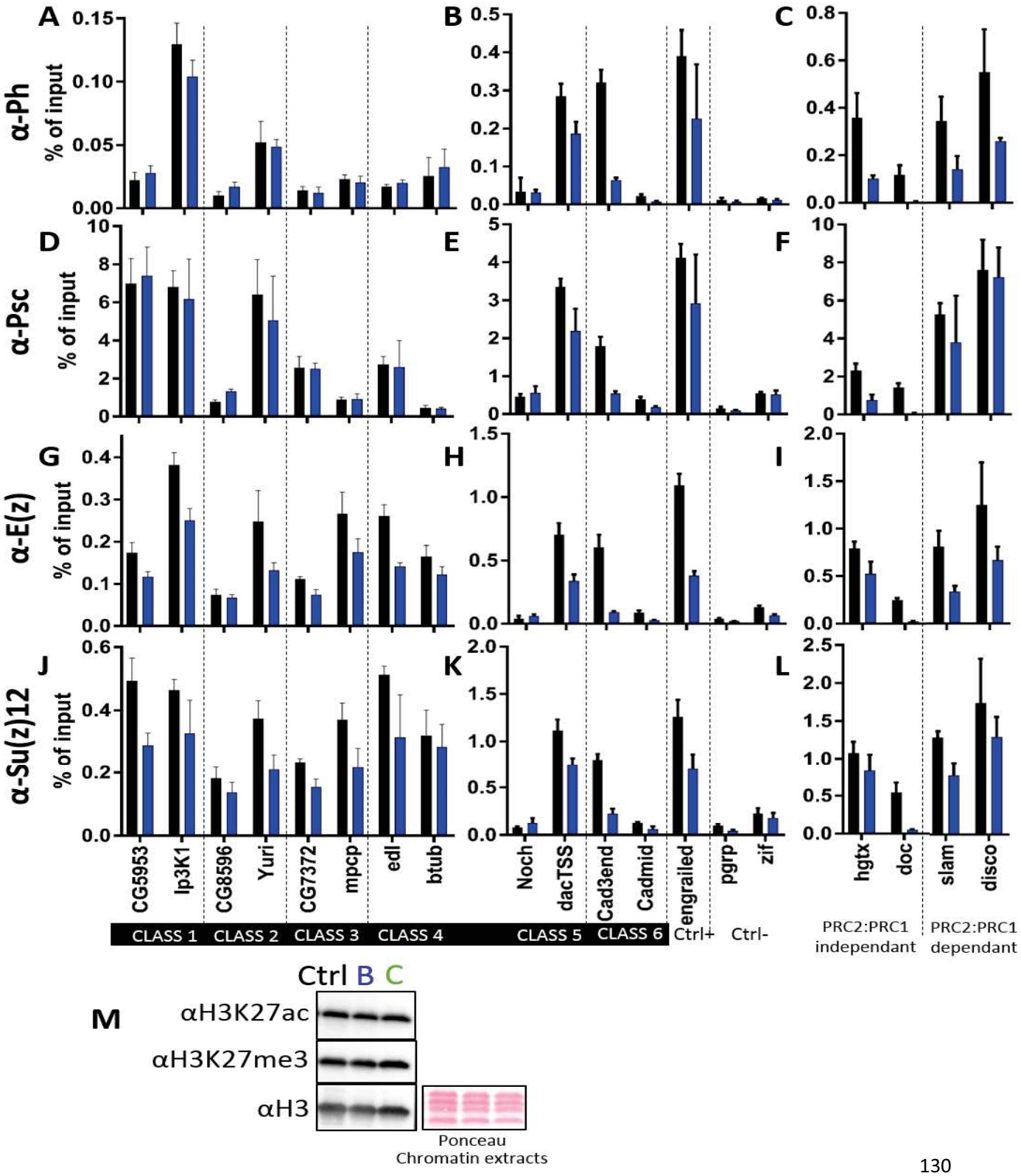
The 798 misexpressed genes were classified according to our PRE classification and a set of Random genes was selected as control. The majority of deregulated genes upon *Pc* depletion correspond to upregulated genes (**Figure 36 B**; red in bar plots). More specifically, more than 20% of the genes associated with class 6 were found to be upregulated with a high log-fold change (superior to 5) upon loss of Pc function (**Figure 36 B and C**). Genes associated with class 5 show to be the most impacted in term of number of genes (**Figure 36 B**). Indeed, more than 30% of the genes belonging to this class are upregulated in Pc mutant cells (**Figure 36 B**). However, the level of this upregulation is mild compared to genes upregulated in class 6 (**Figure 36 C**). This could be explained by the fact that genes in class 6 are generally repressed, whereas genes in class 5 already show low or moderate expression in the presence of Pc. Therefore, depletion of Pc induces stronger gene activation of repressed class 6 genes compared to lowly expressed class 5 genes.

Genes of active PRE classes 1-4 show upregulation of 5% (class 2), 10% (class 3), 15 (class 1) and 20% (class 4) of their target genes (**Figure 36 B**). The number of upregulated genes from class 2 is not significant (**Figure 36 B**). Class 2 comprises the genes belonging to the most highly transcribed genes in the genome (**Figure 28 D**). Thus, it is not surprising to see that loss of Pc has little impact on this class. The fact that a small fraction of upregulated genes of the control (random set) is significant (**Figure 36 B and C**) reflects the bias of the genes toward upregulation in the context of Pc depletion.

A small subset of genes within each PRE class also shows downregulation upon loss of Pc, however their number is not significantly enriched compared to the control gene set. Moreover a significant fraction of genes within the PH-only and E(z)-only classes are downregulated. However, the fact that E(z)-only targeted genes are affected in the same manner than Ph targeted genes argues that the misregulation of these genes is more likely due to indirect effects than to a direct effect of loss of Pc.

Thus together this results indicates that Pc does not play a role in gene activation at active PRE classes, but rather in gene repression. The impact of Pc depletion at non-canonical PREs is discrete but still significant suggesting that even at active genes, PcG proteins do play a role in avoiding a full level of transcription.

Figure 37


A
B
C
D
E
F
G
H
I
J
K
L
α-Ph
α-Psc
α-E(z)
α-Su(z)12
% of input
CG5953
lp3K1
CG8596
Yuri
CG7372
mpcp
edl
btub
Noch
dacTSS
Cad3end
Cadmid
engrailed
pgrp
zif
hgtx
doc
slam
disco
CLASS 1
CLASS 2
CLASS 3
CLASS 4
CLASS 5
CLASS 6
Ctrl+
Ctrl-
PRC2:PRC1 independant
PRC2:PRC1 dependant

Ctrl B C
M
αH3K27ac
αH3K27me3
αH3
Ponceau
Chromatin extracts

Figure 37 : Chromatin changes and PcG binding in *Pc* mutant cell lines

A-L: q-ChIP of PcG proteins from B cell line (in blue) and control cell line (black) at different classes of PREs. The experiment was also performed on PREs previously identified as being PRC2:PRC1 independent in term of PcG complexes recruitment and PRC2:PRC1 dependent (C, F, I and L; see main text and (KAHN *et al.* 2016)). (n=3).

M: Western-blotting of chromatin extracts (1µg) from both B and C cell lines using H3K27me3 and H3K27ac antibodies (n=1).

C. Chromatin changes and PcG binding in Pc mutant cell lines

Since changes in the transcriptome of the two different mutant cell lines are very similar, we decided to use only one mutant cell line (line B) for subsequent analysis, aiming to study changes in PcG binding and their associated histone marks upon loss of Pc function. Therefore, we performed q-ChIP on canonical and non-canonical PREs (**Figure 37 A-L**). q-Chip were performed using antibody against Ph that is part of PRC1, Psc, a component part of PRC1 and the non-canonical complex dRAF (described in the Introduction) and E(z) and Su(z)12, both subunits of PRC2.

As observed previously by ChIP-seq experiments, non-canonical PREs as well as the canonical PRE "*CAD*-mid" have a low enrichment for Ph and E(z) proteins compared to negative regions (see *Pgrp* and *Zif* promoters as negative controls) (**Figure 37 A, B, G and H**). Surprisingly, binding of Ph to non-canonical PREs is not affected suggesting that Ph subunit of PRC1 is able to bind these sites in the absence of Pc. However, given the small enrichment of PcG proteins at non-canonical PREs, it is difficult to assess a significant effect of loss of Ph at these sites.

The overall effect is more visible for E(z) and Su(z)12 proteins that globally decrease at active sites and canonical PREs upon loss of Pc **(Figure 37 G and H**). The global decrease of E(z) and Su(z)12 at all tested PRE classes, correlates with a decrease of H3K27me3 and a concomitant increase in the antagonistic mark H3K27ac as observed in immunostaining experiments (**Figure 38**). However, the changes in histone modification levels seem to be moderate, since no significant changes could be detected by Western blot analysis **(Figure 37 M**).

Strikingly, while PcG proteins such as Pc, Ph, E(z) and Su(z)12 are weakly enriched at the majority of the non-canonical PREs, Psc protein is highly enriched at most of these sites compared to PRC1 and PRC2 **(Figure 37 D and E**). This might reflect the presence of the non-canonical complex dRAF at active PRE classes. Interestingly, Psc binding is not significantly affected at active PRE classes upon Pc depletion **(Figure 37 D**). In contrast, Psc binding decreases at canonical PREs **(Figure 37 E**). This finding suggests that the majority of Psc bound at active sites is part of the dRAF complex, whereas Psc bound at canonical PREs is part of the PRC1 complex.

Figure 38

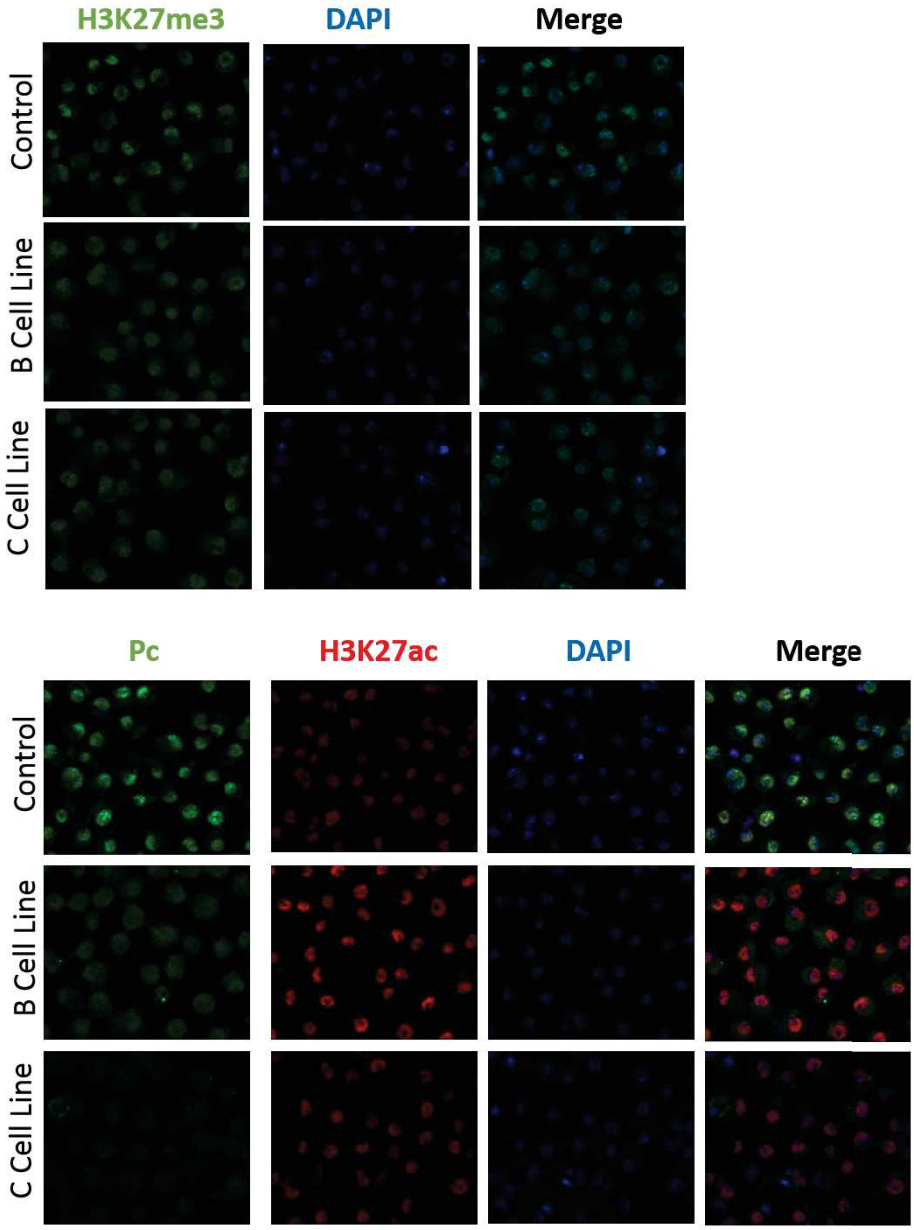


Figure 38 : Pc mutant cell lines show a decrease in H3K27me3 and an increase of H3K27ac.

Immunostaining was performed using H3K27me3 antibody (green). To assess the levels of H3K27ac and Pc, co-immunostaining was performed using both Pc (green) and H3K27ac (red) antibodies. DNA was stained with DAPI (blue).

Overall, these results suggest that loss of Pc results in a global decrease of PRC2 at PREs, correlating with a decrease in H3K27me3 and an increase in the counteracting H3K27ac mark, which might be key for the activation of PcG target genes in *Pc*-mutant cells. Moreover, the ratio between dRAF and PRC1 seems to be different at active PREs and canonical PREs. Whereas at active PREs dRAF is the predominant PcG complex, canonical PREs are predominantly bound by PRC1. Finally, binding of dRAF to PREs is independent to Pc.

D. PRC1-dependent binding of PRC2

The long-standing dogma of a hierarchical recruitment model, in which TFs recruit PRC2, which subsequently recruits PRC1 via the interaction of Polycomb with the PRC2-deposited H3K27me3 mark has been challenged by many reports, and research in the last decade suggested that recruitment of PcG complexes is rather cooperative than hierarchical (KAHN *et al.* 2014; SCHUETTENGRUBER *et al.* 2014) (see introduction for more details). In addition, a recent report (KAHN *et al.* 2016) provided evidence that at the majority of PREs the hierarchy is inversed and PRC2 binding is dependent on PRC1. This study used Psc-mutant cells to propose the existence of two PRE classes, one that recruits PRC2 in a PRC1-dependent manner, and another class where the two PcG complexes are recruited independently (KAHN *et al.* 2016).

Our observation of a global decrease of PRC2 in Pc-mutant cells, prompted us to review the existence of these two classes of PREs. Therefore, we analyzed PRC1 and PRC2 binding in our Pc-mutant cells at 4 additional PREs that has been reported to be PRC1/PRC2 independent (*hgtx, doc*) or PRC1-dependent (*slam, disco*) (**Figure 37 C, F, I and L**). Pc depletion resulted in the reduction of PH binding at all 4 tested PREs **(Figure 37 C**), indicating that PRC1 complex is lost at these PREs. Intriguingly, PSC binding is only reduced at the *hgtx* and *doc* PREs, whereas at the *slam* and *disco* PREs Psc levels are not significantly changed, suggesting that binding of dRAF at the two latter PREs is not affected **(Figure 37 F**). In contrast, dRAF is either absent or its binding is reduced upon Pc depletion at the *hgtx* and *doc* PREs. Most importantly both E(z) and Su(z)12 levels are significantly decreased at the *disco*, *slam* and *doc* PREs, while their levels also decrease at the *hgtx* PRE, however the reduction is not significant **(Figure 37 I and L**). Thus, in contrast to Psc mutant cells, in our Pc-mutant cells PRC2 seems to be globally dependent on PRC1, both at canonical PREs mediating repression and at non-canonical PREs associated with active genes.

Page blanche

DISCUSSION

Résumé en français

Initialement, les PRE ont été identifiés fonctionnellement par leur capacité à recruter de manière ectopique les protéines du PcG dans un transgène. Ces derniers répriment alors un gène rapporteur de manière stable au travers des divisions cellulaires. Les études réalisées à grande échelle à partir d'embryons de Drosophile ont ensuite défini les PRE comme des éléments génomiques situés au sein de larges domaines répressifs marqués par H3K27me3 où se fixent les protéines du PcG. Cependant, des études plus récentes et réalisées à partir de tissus larvaires montrent la présence des protéines du PcG au sein de régions transcriptionnellement actives marquées par H3K27ac. Notre analyse de clustering réalisée à partir de 1543 PRE identifiés dans les cellules S2 de Drosophile corrobore la présence de ces complexes au sein de régions actives. Seulement 1/3 des PRE sont retrouvés au sein de régions marquées par H3K27me3. Nos résultats suggèrent que l'acquisition des protéines du PcG au niveau des gènes actifs se fait de manière dynamique au cours du développement.

Nous avons également réussi à définir différentes classes de PRE. Outre les PRE «canoniques » (présents au sein de domaines marqués par H3K27me3 et montrant une forte interaction intra-domaines) retrouvés dans l'embryon, nous avons identifié deux classes correspondant à des enhancers actifs (Classes 1 et 4). Ce résultat est en adéquation avec une étude récente qui a démontré que chez la Drosophile, certains PRE pouvaient agir comme enhancer. Ces éléments doubles permettent entre autre l'expression et la répression d'un gène de manière cellule spécifique. Nous avons également identifié deux classes de PRE situés au niveau de frontières de TAD : La classe de PRE 2 correspond à des TSS de gènes hautement transcrits qui pourraient correspondre à des gènes de ménages (ces derniers sont enrichis aux frontières de TAD). La classe 3 montre un fort enrichissement en protéines insulatrices. Les classes 1 à 4 sont donc caractérisées par la présence des protéines du PcG au niveau de régions transcriptionnellement actives et marquées par H3K27ac. Nous avons cependant remarqué la présence des protéines du TrxG au niveau de ces sites, dont l'activité est antagoniste à celle des protéines du PcG. Nos résultats suggèrent qu'une fine balance dans le ratio entre ces deux complexes pourrait permettre une régulation fine de l'output transcriptionnelle des gènes cibles. Enfin, la dernière classe de PRE identifiée correspond à des PRE associés à des éléments génomiques en état de pause transcriptionnelle (Classe 5).

Pour déterminer si ces différentes classes de PRE requièrent des combinaisons particulières en FT permettant le recrutement des protéines du PcG, nous avons analysé ces FT pour chacune des classes. Nos résultats indiquent que tous les FT que nous avons étudiés colocalisent avec les protéines du PcG. Cependant, leurs niveaux

d'enrichissements varient d'une classe à l'autre. De manière générale, nous avons identifié les protéines Pho et Spps comme marqueurs des PRE canoniques suggérant que ces deux protéines pourraient jouer un rôle majeur dans le recrutement du PcG au niveau de ces sites. Les PRE localisés au niveau de régions actives montrent des enrichissements relativement similaires pour tous les TF. Ces régions sont notamment caractérisées par un fort enrichissement en protéine GAF. La présence de GAF pourrait expliquer la présence d'interactions retrouvées au sein des classes 1 et 4 d'enhancers (GAF pourrait permettre les interactions entre les enhancers et les TSS) et permettre de manière générale le recrutement de la Pol II au niveau des gènes actifs. Ainsi, parce que les différentes classes de PRE diffèrent, non pas par la présence ou non de certain TF mais plutôt par leurs niveaux d'enrichissements, ceci pourrait expliquer pourquoi les prédictions bio-informatiques des PRE sur la base de leurs motifs ont jusqu'alors échoué.

Historiquement, les PRE ont été fonctionnellement caractérisés par des rapporteurs transgéniques. Cependant, seulement une trentaine de PRE ont été testés, ces derniers correspondant tous à des PRE « canoniques ». Ainsi, les propriétés fonctionnelles générales des PRE restent floues. Nous avons testé deux PRE issus de chacune de nos classes. Les résultats montrent que seuls les PRE issus des classes 6 (canonique) et 5 (éléments en pause) montrent une activité « classique » de PRE, à savoir la répression du gène rapporteur. Les autres PRE montrent au contraire une activation du gène rapporteur. Nous avons par la suite identifié une différence majeure entre les PRE canoniques et non-canoniques : ils ne sont pas ciblés par les mêmes complexes du PcG. En effet, le complexe PRC1 est majoritairement présent aux PRE canoniques alors que c'est le complexe dRAF qui semble être majoritairement présent au niveau des PRE non-canoniques. Cependant, l'importance biologique de cette différence reste à être déterminée.

Enfin, pour mieux comprendre le rôle des protéines du PcG au niveau des PRE non-canoniques, nous avons réalisé des cellules S2 mutantes pour le gène *Pc* et réalisé un transcriptome de ces cellules. Contrairement à certaines publications suggérant un rôle positif des protéines du PcG dans l'activité transcriptionnelle, nos résultats montrent que même au niveau des PRE non-canoniques, les protéines du PcG jouent un rôle de répresseur transcriptionnel. En revanche, contrairement aux gènes associés aux PRE canoniques, ceux associés aux PRE non-canoniques ne sont pas transcriptionnellement éteints. Nos résultats montrent donc qu'au niveau des PRE non-canoniques, les protéines du PcG sont impliquées dans une régulation très fine de l'expression de leurs gènes cibles. Ce résultat concorde avec une publication récente qui montre le rôle important des protéines du PcG dans la diminution transcriptionnelle de gènes associés à la prolifération cellulaire chez la larve. Une perte de Pc conduit en une expression encore plus importante de ces gènes et en l'apparition de cancers. Par ailleurs, il a été

démontré que les gènes actifs associés aux protéines du PcG étaient ceux qui montraient le plus de bruit transcriptionnel. Ainsi, la présence de ces protéines pourrait permettre de réduire ce bruit.

Pour conclure, nos résultats montrent une bien plus grande diversité des sites de fixation aux protéines du PcG que précédemment identifiée. Cette diversité de PRE reflète la diversité des fonctions biologiques de ces protéines aux sites transcriptionnellement actifs et réprimés. Nos travaux centrés sur le rôle des protéines du PcG aux sites actifs sont d'une grande importance et les études futures permettront de mieux comprendre le rôle des protéines du PcG au niveau de ces sites.

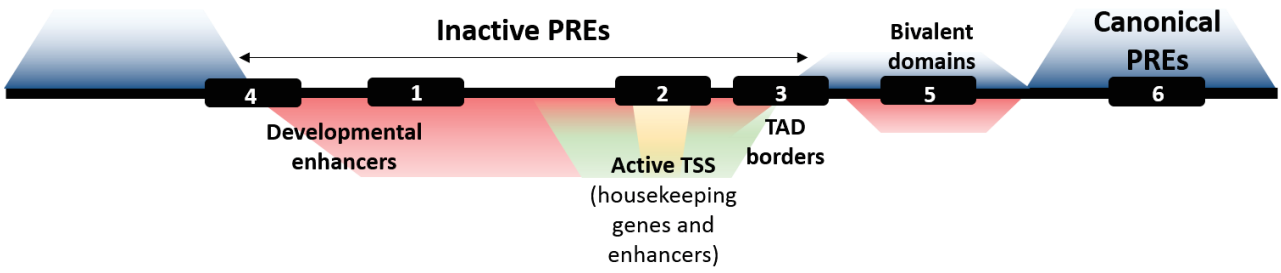


Figure 39 : Different classes of PREs exist in the Drosophila genome

Overall, our study reveals the existence of different classes of PREs characterized by their chromatin environment (active or repress), their genomic characteristics, their recruiting factors (not shown here but see main text). Overall, we showed that canonical PREs marked by H3K27me3 represent only 1/3 of all PREs (class 6). Class 5 of PREs are in poised state marked by both active and repressed histone modifications. Class 2 represents active TSS and could represents a class composed of PREs associated to housekeeping regulatory regions (CUBENAS-POTTS *et al.* 2017). Class 3 of PREs is associated to TAD borders. Finally, both classes 1 and 4 correlate with developmental enhancers (CUBENAS-POTTS *et al.* 2017) but differ in their chromatin environment. We characterized PREs at active regions as inactive because they do not act as canonical PREs when tested in transgenic rapporter assays. They rather correspond to TREs.
H3K27me3 domains are marked in blue. H3K27ac domains are marked in red. H3K36me3 domain is marked in green while H3K4me3 peak is mark in yellow.

A. PREs are heterogeneous cis regulatory elements, present at repressed and active genomic regions (Figure 39)

Initially, PREs have been functionally characterized using transgenic reporter gene assays in Drosophila, demonstrating that these DNA regulatory elements recruit PcG factors and mediate epigenetic inheritance of silent or active chromatin states throughout development (KASSIS and BROWN 2013). Subsequent genome-wide mapping studies performed in Drosophila embryos, defined PREs as genomic elements marked by the presence of PcG proteins within repressive chromatin domains marked by H3K27me3 (SCHWARTZ *et al.* 2006; SCHUETTENGRUBER *et al.* 2009). Importantly these initial studies (based on the ChIP-on-chip technology) did not detect PcG proteins at active sites. However, they already revealed that not all PREs are alike, but they can differ in their underlying sequences or their composition of sequence specific DNA binding proteins. In addition, more recent data indicated a role of PcG complexes at active genes. First, Schaaf *et al*., showed that in Drosophila imaginal discs, the cohesin protein facilitates the binding of PRC1 at many active genes, but their binding at repressed genes are mutually antagonistic (SCHAAF *et al.* 2013). Second, a subsequent study demonstrated the functional importance of PRC1 at active sites in Drosophila eye disc in regulating genes implicated in cell proliferation, signaling and polarity (LOUBIERE *et al.* 2016). Together, these observations revealed the dynamic binding of PcG proteins during development and the appearance of new PcG target sites during drosophila development marked by the antagonistic TrxG mark: H3K27ac. Moreover, several genome wide mapping studies reported the association of PcG proteins with actively transcribed genes in different species (BROOKES *et al.* 2012; MOUSAVI *et al.* 2012; FRANGINI *et al.* 2013; KANEKO *et al.* 2013; VAN DEN BOOM *et al.* 2016). Altogether, these studies clearly demonstrate the heterogeneous nature of PREs: PREs can associated with repressed and active sites; they do differ in their sequence characteristics, in their composition of bound sequence specific DNA binding proteins or in their composition of bound PcG complexes and/or associated histone marks. It is still unknown how PcG complexes are recruited to PREs lacking H3K27 methylation.

Our working hypothesis was therefore that different classes of PREs exist that use different protein factors, chromatin features and sequence features to target PcG proteins. In addition, PcG protein might have different functional properties depending on the PRE class. This could also explain why attempts to bioinformatically prediction of PREs failed so far (RINGROSE *et al.* 2003; FIEDLER and REHMSMEIER 2006). Consequently, one has first to define different classes of PREs and subsequently determine their characteristics in order to understand their mode of function. Therefore, we used Drosophila S2 cells to assess the functional diversity of PcG binding-sites. S2 cells are embryonic-derived cells that are thought to have macrophages-like properties. In contrast to Drosophila embryos or larval tissue, S2 cells are comprised of a more or less

homogenous population of cells. This should allow us to compare directly binding intensities of PRE-associated proteins and define classes of PREs based on their different chromatin characteristics.

We found a good correlation between PREs marked by H3K27me3 from S2 cells to larval PREs marked by H3K27me3. Importantly, we identified many PcG binding-sites devoid of H3K27me3 mark that are localized within active chromatin environment and we found that many of these active PcG-binding sites are observed, both in S2 cells and larvae. Therefore, our results support the binding of PcG complexes to active genomic loci in S2 cells. The fact that these active PcG binding sites were not identified in embryos can be due to several reasons: 1) PcG complexes simply do not bind to active regions in embryonic tissue. 2) PcG complexes are bound to active regions in only few cells of the embryo and ChIP-seq assays are not sensitive enough to detect them. Indeed, Drosophila embryos are composed of many different cell types, and genome wide mapping of PREs in this heterogeneous tissue might only detect PcG binding sites that are bound in the majority of cells. The latter explanation is also supported by the fact that more recent mapping studies in late Drosophila embryos using very deep sequencing, detected PH binding sites outside H3K27me3 domains at weak intensities (SCHUETTENGRUBER *et al.* 2014). This observation supports the hypothesis that PREs at active regions (outside H3K27me3 domains) are tissue and/or developmental specific PcG binding sites present in only a small population of embryos.

B. Redefinition of PREs

Our clustering analysis of PREs detected in S2 cells demonstrated that only 1/3 of PcG binding sites are localized within H3K27me3 marked chromatin regions corresponding to the so called "canonical PREs" (Class 6) previously described by (SCHWARTZ *et al.* 2006; SCHUETTENGRUBER *et al.* 2009). This class shows all characteristics of previously described Drosophila PREs: They associates with repressive Polycomb domains marked by H3K27me3 and repressed genes lacking Pol II. Moreover, these sites can be located at TSSs, but are also frequently found far away from their target genes.

In contrast, 2/3 of the PREs are found within an active chromatin environment, mainly marked by the presence of H3K27ac. We called these PREs "non-canonical PREs" (class 1-4) as they do not fit the initial characteristics of Drosophila PREs. They usually show a high enrichment for active histone marks (H3K27ac, H3K36me3 and H3K4me3), associate with active genes bound by Pol II and mainly associated with TTSs of genes.

Finally, we identified another non-canonical class that is particularly intriguing and that we defined as "poised PREs" (class 5). This class is marked by low levels of both H3K27me3 and H3K27ac, two histone marks that are mutually exclusives. We did not

assess whether these two modifications were deposited on the same nucleosomes. Therefore, we cannot exclude the formal possibility that these sites reflect a transcriptional heterogeneity of S2 cells, where a subset of genes is repressed and is associated with H3K27me3 in a subpopulation of the cells, while those genes that are active in another subpopulation of the cells are associated with H3K27ac. However several observations argues that these sites are indeed in a "bivalent", or poised state and could correspond to canonical PREs being poised for an activation later in drosophila development: 1) They are bound by Pol II but correspond to lowly transcribed genes. 2) They show similar distribution around genes and insulator-binding characteristics compared to canonical PREs. 3) Transcriptome analysis of Pc-mutant cells revealed that genes associated with this class are highly responsive to the loss of function of Pc. Therefore, PcG proteins at these regions might keep Pol II in an inactive/poised state. This resembles the repression of bivalent genes in embryonic stem cells (ES cells) marked by H3K27me3 and H3K4me3, which are poised to get expressed or repressed during ES cell differentiation (DI CROCE and HELIN 2013).

C. Co-occupancy of PcG and TrxG proteins at PREs/TREs

PcG-mediated repression has been considered to be the default state, with gene activation mediated by the counteracting activity of TrxG proteins (KLYMENKO and MULLER 2004). Intriguingly, both in flies and mammals, PcG and TrxG complexes extensively colocalize at chromatin regardless of the activity state of the target gene (PAPP and MULLER 2006; BEISEL *et al.* 2007; ENDERLE *et al.* 2011). In agreement with these previous observations, we find that members of TrxG proteins which are part of COMPASS complexes mediating different states of H3K4 methylation (Trr, Trx and dSET1) strongly associates with active classes of PREs, although they are also present at lower levels at repressed PREs (class 5 and 6). In contrast, PcG binding at repressed PRE classes is significantly higher compared to active classes. This suggests that the ratio between local binding of TrxG and PcG proteins could switch the balance between PcG-mediated repression and TrxG-dependent gene activation. In analogy, it has been recently demonstrated that the SWI/SNF chromatin remodeling complex is constantly involved in a dynamic competition with PRC1. The outcome of this competition at individual loci might depend on the amount of PRC1 being recruited, which would drive each locus into an open or closed chromatin conformation (KADOCH *et al.* 2017; STANTON *et al.* 2017).

D. Enhancer function of PREs

Importantly, a recent report demonstrated a dual functionality of some cis regulatory elements as transcriptional silencing elements (PREs) and as transcriptional activators (enhancers) (ERCEG *et al.* 2017). This study thus showed that a subset of Drosophila PREs can function as developmental enhancers *in vivo*, activating transcription in specific spatial domains. *Vice versa*, 50 % of the enhancers can act as PREs in a certain tissue (ERCEG *et al.* 2017). In agreement with this observation, we found that all PRE classes show a high and significant overlap with enhancer regions defined by STARR seq (ARNOLD *et al.* 2013). In particular, classes 1 and 4 are highly enriched in enhancers regions. Moreover, both show a high enrichment in both H3K27ac and H3K4me1, two histone marks that are associated with active enhancer activities. PcG proteins can have two possible functions at active PRE/enhancer elements (class 1-4): They could repress the activity of the associated enhancer, keeping the target gene available for another enhancer element. Thus in this model PcG proteins would exert a repressive function, although the associated target gene is active. Another possibility for the function of PcG proteins at active PRE/enhancer elements is that they mediate repression in one tissue or cell context, but not in another. In both scenarios, PcG proteins stay associated with the PRE, but activation is mediated via the recruitment of (tissue specific) activating factors (e.g TFs, TrxG proteins).

E. Distribution of histone marks around PRE classes

As mentioned above, canonical PREs (class 6) and to a less extent, poised PREs (class 5) are located within a repressed chromatin environment marked by H3K27me3. As previously observed by other groups (MOHD-SARIP *et al.* 2006; SCHWARTZ *et al.* 2006; MITO *et al.* 2007) we also detected a drop in H3K27me3 signal around these PREs, suggesting that PREs are nucleosome free regions. The same is true for active PRE classes 1-4, when looking at the distribution of H3K4me2/3, showing a drop at the PRE center. Thus, all PRE classes seem to be hypersensitive sites having a less dense nucleosomal occupancy, however the relative distribution of histone marks around PRE centers can be quite different between PRE classes. Surprisingly, we noticed a strong peak of H3K27me2 in particular at the center of canonical PREs, suggesting that the few nucleosomes remaining at PREs are rather di- than tri-methylated on H3K27. The presence of H3K27me2 at PREs is surprising and the biological importance of this observation remains to be determined. One hypothesis is the presence of the demethylase UTX at PREs locally removing the tri-methylation mark at PREs, whereas this enzyme is unable to demethylate histones in the surrounding region. Similar H3K27me2 peaks were also identified at E(z) sites that do no overlap with Ph (class E(z)-only). This supports the existence of a PRC2 activity outside of canonical PREs. Indeed

the H3K27me2 mark covers the large majority (70%) of the euchromatic genome (FERRARI *et al.* 2014; LEE *et al.* 2015) and it has been suggested that weak or transient interactions of PRC2 with the genome leads to the di-methylation of the genome which can suppress pervasive chromatin opening and transcriptional activation.

In addition to a peak of H3K27me2, nucleosomes at the PRE center of canonical PREs are also di-methylated on H3K4 (H3K4me2). This confirms a recent report, showing that Trx-dependent H3K4 dimethylation (H3K4me2) marks Drosophila PREs and maintains the developmental expression pattern of nearby genes (RICKELS *et al.* 2016). Interestingly the relative distribution of H3K4me2 around active PRE classes is different: Surrounding nucleosomes show high levels of H3K4 methylation, whereas at the PRE center methylation levels are reduced. In summary, active and repressed PRE classes are not only defined by different chromatin modifications but also by a different spatial distribution of histone marks.

F. PRC1 and PRC2 only sites and Gene ontologies (GO) associated with PRE classes

Although PRC1 and PRC2 generally co-regulate their target genes, we also identified E(z) and Ph-only sites that target actively transcribed genes. Genomic sites bound by either PRC1 or PRC2 were previously identified in both Drosophila and mammals (BOYER *et al.* 2006; KU *et al.* 2008). However, we did not go too far in seeking the particular characteristics of these sites. To better understand the differential role of PRC1 and PRC2, an analysis of the ontologies of their target genes should be done. We are currently performing the analysis of the ontology of targeted genes for each class of PREs. Our preliminary results (data not shown) revealed that canonical class 6 is highly enriched for transcription factors involved in gene regulation as it was previously shown for targeted genes of canonical PREs in embryos (SCHWARTZ *et al.* 2006; SCHUETTENGRUBER *et al.* 2009). Gene functions associated with developmental processes were enriched in classes 1, 3 and 4, which could reflect their enhancer function (data not shown). No particular GO was associated to class 2 (data not shown). As a short-term analysis to do, we plan to cross our different classes of PREs with a published list of housekeeping genes (HUG *et al.* 2017). We expect to find an enrichment of housekeeping genes for class 2, since this PRE class targets the most highly expressed genes in the genome and is not associated to any particular GO. Moreover, class 2 PREs shows a stronger enrichment of the Beaf32 insulator protein (discussed below). Importantly, over 85% of the genes with a TSS associated with BEAF32 were shown to be on a list of housekeeping gene (JIANG *et al.* 2009; FELLER *et al.* 2012; LAM *et al.* 2012), supporting the hypothesis that class 2 mainly comprises housekeeping genes.

G. Role of 3D chromatin interactions and insulator proteins in PRE function

Metazoan genomes are organized into distinct modules called topologically associating domains (TADs), which correspond to linear chromatin regions that fold into specific 3D structures defined by strong genomic interactions within a domain, whereas interactions between TAD borders are strongly reduced. TAD borders are formed in regions containing actively expressed genes and clusters of insulator binding sites (SANBORN *et al.* 2015; KONDO *et al.* 2016). As already mentioned above, PcG proteins contribute to the formation of a specific type of repressive TADs, called Polycomb domains that are characterized by the presence of H3K27me3 (reviewed in (ENTREVAN *et al.* 2016)). Within these domains, PREs (corresponding to our group of "canonical PREs") have been shown to form looping interactions, leading to the clustering of PcG target genes and their regulatory regions (like PREs, promoters, enhancers, insulators) (COMET *et al.* 2011). The functional importance of PRE interactions is still mysterious, but it has been suggested that these 3D chromatin interactions might be important for efficient silencing of their target genes (BANTIGNIES *et al.* 2011).

By analyzing the interaction maps around PREs with mid-to-high level of H3K27me3 (class 5 and 6) we showed that they do not correlate with TAD boundaries (since they do not show insulation activity), but are rather localized within a TADs. Most importantly, as previously shown for PREs in Drosophila embryos, canonical PREs in S2 cells (class 6 PREs) showed a strong accumulation of short-range interactions, corresponding to inter-domain interactions. In contrast, long-range interactions among PREs are reduced but still detectable, indicating that these canonical PRE interactions are mainly restricted to the same TAD. However, PREs can also interact over long distances, as it has been shown for PREs of the two Hox gene clusters (BANTIGNIES *et al.* 2011). Although, inter-domain looping contacts of PREs seems to be a prominent feature of canonical PREs, we could also detect some signs of short-range contacts between PREs of classes 1 and 4, which are particularly enriched in enhancers. These PRE classes do not show insulation activity (class 4) or show weak insulation activity (class 1). In contrast, for class 2 PREs we detected a weak enrichment of long-range contacts corresponding to inter-domain interactions. Class 2 PREs are associated with highly active genes, most likely housekeeping genes (see above) and shows very strong insulation activity. Interestingly, it has been recently reported that housekeeping gene loci, acts to separate topological domains (EL-SHARNOUBY *et al.* 2017) and that enhancers for housekeeping genes, which are associated with H3K4me3 form multi-TSS interaction networks and are related to the position of boundaries in Drosophila Kc cells (CUBENAS-POTTS *et al.* 2017). Since our class 2 PREs show the same chromatin features, it is tempting to speculate that they actually correspond to these housekeeping enhancers described in Kc cells. In contrast to housekeeping enhancers, developmental enhancers are marked by H3K4me1 and are more likely to generate

single looping contacts within TADs (CUBENAS-POTTS *et al.* 2017). So again, these features correspond precisely to our PRE classes 1 and 4, suggesting that these two classes correspond to developmental enhancers. Finally, we observed for class 3 PREs a strong association with the insulator protein CTFC and its cofactor CP190. Since this class also shows strong insulation activity, we conclude that class 3 PREs corresponds to TAD borders. However, in contrast to TAD borders of class 1 and 2 these borders do not involve chromatin interactions.

So what are the factors mediating 3D chromatin interactions of the different PRE classes, and could PcG proteins play a role in the 3D architecture of the different PRE classes? Canonical PREs are enriched in the two insulator proteins CTCF and GAF. Both proteins have been shown to be able to mediate chromatin interactions (MELNIKOVA *et al.* 2004; PETRASCHECK *et al.* 2005; CUBENAS-POTTS *et al.* 2017), and are therefore good candidates to mediate PRE interactions. However recent work in our laboratory showed that the looping anchor points of PREs within H3K27me3 domains in embryos are not associated with CTCF, whereas they correlate very well with GAF binding, suggesting an important role for GAF in mediating canonical PRE looping interactions within Polycomb domains (Ogiyama *et al*., in preparation). On the other hand, Polyhomeotic (Ph) itself might contribute to higher order chromatin organization by forming long-range contacts between distant Ph-bound chromosomal sites. Ph contains a sterile alpha motif (SAM)-domain, which is capable to form homo- and hetero-oligomers through its self-association with its SAM domain (for a review, see (ENTREVAN *et al.* 2016)).

H. PRE classes are specified by distinct Transcription factor (TF) signatures

As exposed in the introduction in more detail, several transcription factors, including Pleiohomeotic (PHO), its homolog Pleiohomeotic-like (PHOL), GAGA factor (GAF), Pipsqueak (PSQ), Dorsal switch protein (DSP1), Zeste, Grainyhead (GH) and SPPS (SP1 for PSS) bind to specific DNA motifs, which are frequently found enriched at PREs (reviewed in (KASSIS and BROWN 2013)). Among them, PHO protein is thought to play a key role in building a binding platform for the recruitment of other PcG complexes. However, none of these TFs alone is sufficient for PcG recruitment and all factors seem to be involved in repression as well as in gene activation.

To understand whether different PRE classes use different combination of TFs to recruit PcG proteins, we analyzed TF binding at the different classes. Although globally, binding of all examined TFs correlate well with all PRE classes and most PREs are co-bound by all TFs, we detected two key quantitative differences between active (class 1-4) and canonical PREs (class 6): High levels of both Pho and Spps were markers of canonical

PREs, suggesting that these two TFs are main PcG recruiters at these sites. This is in agreement with the important genetic interaction of Pho and Spps proteins, leading to an enhanced deregulation of Hox genes (BROWN and KASSIS 2010). The role of Spps in PcG recruitment is less well characterized: Spps mutant flies do not show typical homeotic phenotypes usually associated with PcG proteins and no genome wide binding profile has been reported so far. However, our results suggest a high importance of Spps in PcG recruitment to canonical PREs. All active PREs are characterized by a similar combination of PcG recruiters and are marked by high levels of Gaf. Within those classes, the shape of TF peaks differ, forming broad to sharp peaks. The fact that all PcG recruiters are found also at active PRE classes supports their role in gene activation beside repression. In particular, Gaf was shown previously to play a role in Pol II recruitment (FUDA *et al.* 2015). Indeed, it interacts with chromatin remodelers such as NERF allowing Pol II recruitment at many TSS in S2 cells (FUDA *et al.* 2015). In summary, the strong co-localization of all examined TFs at the different PRE classes reflects their redundant and cooperative function in PcG recruitment, which in turn explains why the loss of a single PcG recruiter generally does not have major effect on PcG recruitment on a genome wide scale, and why a single PcG recruiter protein on it's own is inefficient in targeting PcG proteins to chromatin.

Previous studies tried to bioinformatically predict PREs based on a specific combination of DNA motifs of known PcG recruiters (RINGROSE *et al.* 2003; FIEDLER and REHMSMEIER 2006). However, only a weak overlap was found between bioinformatical predicted PREs and PREs identified in embryos (SCHUETTENGRUBER *et al.* 2009). Two more recent studies provided a possible explanation, by revealed that although DNA motifs are important to some extent, it is rather the combinatorial interaction of PcG proteins and their recruiters that are important to stabilize the proteins at PREs. In addition, PcG proteins and 3D chromatin interactions might stabilize TF binding (KAHN *et al.* 2014; SCHUETTENGRUBER *et al.* 2014). Of particular interest is the Pho protein, which was shown to bind to strong motifs outside of H3K27me3 while it can bind to weaker motifs at canonical PREs. Therefore, with the aim to discriminate differences between canonical and non-canonical PREs at the DNA level, we are currently performing a motif discovery analysis. Since these analyses are still at a preliminary state, results are not included in the current manuscript. We are aiming to identify DNA motifs enriched in each single TF binding profile to define or redefine in vivo binding motifs for each TF. The second step of this analysis is to look at the enrichment of these DNA-binding motifs within active versus repressed PRE classes. We hope to find several TF-binding motifs specifically enriched at non-canonical PREs and others specifically enriched at canonical PREs.

I. Different functional roles of canonical and non-canonical PREs?

PREs have been functionally characterized and defined by using transgenic reporter gene assays as genomic elements capable of recruiting PcG complexes to ectopic genomic regions and capable of mediating gene silencing of a transgene reporter gene. To date, only about 30 PREs have been functionally validated in transgenic lines, and all these PREs belong to the canonical PRE class 6. Thus, the functional requirements and general properties of PREs remain poorly characterized and non-canonical PREs (class 1-4) have not been tested in functional reporter gene assays yet.

Indeed our reporter gene assays, designed to be able to directly compare PRE activity of genomic fragments due to the site-specific integration, indicate that non-canonical PREs have a different functional behavior than that of canonical PREs. Canonical PREs showed the expected characteristics of previously characterized PREs: They ectopically recruit PcG proteins to the transgene insertion site in Drosophila polytene chromosomes, they maintain the repressed state of the *LacZ* reporter gene in the appropriate segment of the fly embryo and they repress the *mini-white* reporter gene. Interestingly, poised PREs identified in S2 cells (class 5), seem to have similar functional properties than the canonical PREs, suggesting that in most of the fly tissues examined these PREs are in a repressed state similar to canonical PREs. The class 5 PRE from the *Notch* gene locus showed a PRE activity in embryos but lack PRE activity in adult. This could indicate that it mediates PRE activity in a stage specific manner.

The non-canonical PREs of active class 1-4 tested in this study do not show classical PRE properties. On the contrary, the vast majority of active PREs from S2 cells mediate activation in Drosophila embryos leading to the ectopic expression of the *LacZ* reporter gene. We could not detect PcG binding at the transgene locus containing active PREs in Drosophila polytene chromosomes. Notably, this assay is not very sensitive, and qChIP experiments at different developmental stages will allow us to determine whether this class of PREs associates with PcG proteins at a particular developmental stage in flies. In addition, more non-canonical PREs should have been tested to reliably conclude that non-canonical PREs are developmental-stages specific.

Another difference between canonical PREs and active PRE classes is the ratio of PRC1 and dRAF complexes bound at these sites. Our qChIP analysis in Pc mutant cells indicates that active PREs are bound predominantly by dRAF, whereas canonical PREs are bound predominantly by PRC1. The biological significance of this differential association of PcG complexes remains to be determined.

J. Intra-variability of canonical PREs

The functional analysis of canonical PRE in embryos revealed that one of the canonical PREs localized at the 3'end of the *Caudal* (*Cad*) gene (CAD 3end) does not have strong PRE activity at the embryonic stage, since it is not able to maintain the repressed state of the *LacZ* reporter gene between the stripes of the *engrailed* expression domains. Contrarily at the adult stage, the PRE is efficiently repressing the *mini-white* reporter gene. Another PRE from the same Polycomb domain (named *CAD*-mid) showed strong PRE activity at both stages raising the question of the importance of these PREs in the formation of Polycomb domains during the development (**Figure** 17). Indeed, this result suggests that PREs within the same Polycomb domain might have different functions at specific developmental stages.

What could be the importance of such a functional divergence for the formation of the PcG domain? To answer this question, a long-term project would be to mutate each site independently within their endogenous genomic context using the CRISPR/Cas9 system in order to assess their respective importance for the formation of Polycomb domains. When looking at the PcG recruiter binding, we noticed a difference in their signature. *CAD*-mid is lowly but significantly enriched for both Pho and Spps, but not bound by the other recruiters, while *CAD*-3'end is highly enriched for all PcG recruiters except Gaf. Thus, clearly canonical PREs do not behave as a single phenomenon and it would be interesting to refine the classification of PREs by doing another clustering analysis using the canonical PREs as input and the six PcG recruiters as variables to determine whether different sub-classes of canonical PREs exist.

Another aspect of the diversity of canonical PREs is linked to the results we obtained by performing q-ChIP in S2 cells lacking Pc. In a recent study, Kahn *at al.* identified two classes of canonical PREs that diverge in the way they recruit PRC2 (KAHN *et al.* 2016). One class recruits PRC2 independently of PRC1, whereas at the other class PRC2 recruitment is dependent on PRC1. Importantly, since they used *Psc* knock out cells, they could not discriminate whether loss of PRC2 binding is the consequence of loss of PRC1 function or due to disruption of dRAF, which also contains Psc (see Introduction). Our results using *Pc* knock out cells that disrupt PRC1 but not dRAF (as suggested by the fact that Psc binding is generally not or only weakly affected upon *Pc* knock out) suggest that the recruitment of PRC2 is globally dependent on PRC1 at all canonical PREs as it is the case for active PRE classes, arguing for a general dependence of PRC2 on PRC1. In addition, our data suggest that dRAF is not sufficient to recruit PRC2 to their targets.

K. What is the role of PRC1 at active regions?

It has been previously reported that RNAi-mediated knock-down of Pc in Drosophila BG3 cells does not automatically lead to the activation of all Pc-bound genes, but induces a very specific genomic response (SCHWARTZ *et al.* 2010). So in other words, upon RNAi-mediated Pc knock down, only a small subset of PcG-target genes is activated. Two possible scenarios could explain this observation: 1) Pc knock down by RNAi is not efficient enough (RNAi mediated knock down only reduces Pc levels to 20%), and the remaining Pc levels are sufficient to maintain the repressed state of PcG target genes. 2) Transcriptional activators necessary for the activation of PcG target genes are absent for the majority of PcG repressed genes, and only present for the subset of genes activated upon loss of function of Pc. At these genes Pc knock down lowers the threshold for the amount of activator needed to switch target gene repression. Our Pc-null mutant cell line derived by using CRISPR/Cas9 mediated mutation, allowed us to re-evaluate these hypothesis, and, more importantly allowed us to determine the effect of loss of PcG function on target genes of active PRE classes 1-4. Our transcriptomic analysis revealed a global bias toward upregulation of genes upon loss of Pc function. However, only a small but significant subset of canonical PREs is activated upon Pc knock down. This supports the hypothesis that transcriptional activators for the majority of PcG target genes is not available, or that other silencing mechanism maintain genes in a repressed state.

In addition recent reports, suggested a role of PcG proteins in transcriptional activation (see Introduction). How PcG complexes activate their target genes is largely unknown. In mammals, it was suggested that substitution of some PcG isoforms is implicated in the switch from a repressor to an activator function. Another study revealed a structural role of PcG proteins in chromatin folding to mediate gene activation by bridging a TSS to their enhancers (reviewed in (ENTREVAN *et al.* 2016)). The fact that PcG proteins are associated with active genes in Drosophila S2 cells therefore raised the possibility that they might have a positive role in the regulation of these genes. However, our transcriptomic analysis showed that not only genes from canonical PRE classes were upregulated but also genes of active PRE classes 1-4. With the exception of class 2 (active TSS class) all classes show a significant upregulation of the number of their target genes. In contrast, the number of downregulated genes in active PRE classes is low and mostly unsignificant. This result suggests that at non-canonical PREs, PcG proteins act also as transcriptional repressors, rather than activators. Our results are in concordance with a recent study that revealed a functional role of PRC1 at active sites (LOUBIERE *et al.* 2016). Importantly, regulation of PcG target genes seems not to be an ON-OFF system, but rather involves the “fine tuning” of gene expression levels, supported by the fact that at active PcG target genes, PcG proteins rather diminish the transcription than abolishing it.

One mechanism explaining how PcG proteins might regulate gene expression at actively transcribed genes is by regulating local chromatin structure that could affect Pol II elongation or processivity. For example, Pc has been shown to directly interact with the histone acetyltransferase CBP and inhibit its HAT activity to lower H3K27ac levels and possibly dampen gene expression levels. An alternative hypothesis is that high levels of dRAF at active PRE target genes results in reduced levels of the elongation mark H3K36me2/3 via its demethylase activity mediated by dKdm2, therefore negatively regulating Pol II processivity. Furthermore, a recent study demonstrated another functional role of PcG protein at active genes in mammalian cells. Indeed, active PcG-targeted genes show greater cell-to-cell variation in their expression compared to active genes not targeted by PcG complexes (KAR *et al.* 2017). Thus, PcG complexes might modulate transcriptional noise in gene expression of their active target genes. How this process is regulated remains to be determined.

Taken together, this study revealed an unexpected degree of diversity of PREs in Drosophila. The diversity of PREs goes hand in hand with the diversity of PcG function at repressed as well as at actively transcribed genes. Focusing on non-canonical PREs and the specific role of PcG proteins in regulating active genes as well as the newly discovered role of PcG complexes in gene activation is of particular importance, and future studies will contribute to better understand the molecular mechanism of PcG function at active genes

MATERIEL & METHODES

S2 cell culture and CRISPR/Cas9 editing of *Pc* gene

D. melanogaster S2 cells (obtained from the Drosophila Genomics Resource Center) were cultured in Schneider's Drosophila medium (Invitrogen) supplemented with 10% FCS (Hyclone) and 1:100 Penicillin/Streptomycin (ThermoFisher_15140122).
D. melanogaster S2 cells already expressing Cas9 enzyme were obtained from Klaus Foerstemann (clone 5-3; (BOTTCHER *et al.* 2014)). For *Pc* KO, two strategies were designed: B and C (see main text).

1. First, **templates for homologous recombination** (HR) was obtained by PCR on the plasmid pMH3 (provided by Klaus Foerstemann) with primers specific to each strategy. The expected size is 1850 nt for both strategy.

Method: Primers for B and C strategies (specific to the *Pc* gene in upper case):
Pc_B_S: 5'-TACGCAGAATTGTAAACCAGAAGTTAATTGCAAATAAAACGAATAATAAAACGT TCCGAGAAGATTATTAATTAAAAATGgtgagcaagggcgaggagct-3';
Pc_B_AS: 5'CGCTCATCTCACCTTCTTAACGCGCTTTTGGATGATTTTCTCAGCCGCGTACACT AGATCgaagttcctattctctagaaagtataggaacttccatatg-3';
Pc_C_S: identical to Pc_B_S;
Pc_C_AS: 5'TTTCTGGTTGATAGCCAGGTTGTTGCACGGTTTTGGTATGTTATTGTTCTCGGAA GGAAcgaagttcctattctctagaaagtataggaacttccatatg-3'

Réactif	Quantité (µL)
Plasmid pMH3 [100pg/µL]	10
Tampon 5x	10
dNTP [10mM each]	1
Primers [10µM each]	2
GoTaq G2 Flexi DNA Polymerase Promega_M7805	0,25
MgCl2 [50mM]	2
H2O	qsp 50

Cycle:		
Step 1	94°C-2'	
2	94°C-20''	
3	55°C-20''	
4	72°C-45''	Back to step 2 - X35
5	72°C-45''	
6	4°C-∞	

2. Then, we **produced sgRNAs by PCR**. Strategy B requires only one sgRNA targeting exon 1 whereas strategy C requires 2 sgRNAs targeting both exon 1 and 2. sgRNAs were obtained by overlap-extension PCR using a sgRNA scaffold primer serving as template during PCR; a unique primer sense and sgRNA antisens primers specific to each exon. The expected size is 110 nt for both strategy.

Method: Primers (Region specific of *Pc* gene of antisens primer is indicated in upper case; the cutting site in red):

Scaffold primer: 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTAT CAACTTGAAAAAGTGGCACCGAGTCGGTGC-3'
Antisens primer: 5'-GCACCGACTCGGTGCCACT-3'
Sens primer targeting exon1: 5'-taatacgactcactataGGCGCGTACACTAGAT cgACgttttagagct-3'
Sens primer targeting exon2: 5'-taatacgactcactataGGTTCTCGGAAGGAA ccTGgttttagagct-3'

Réactif	Quantité (µL)
Oligos scaffold [0,1µM]	5
Targeting primer [10µM]	1
Primer AS scaffold [10µM]	1
Tampon 5x	10
dNTP [10mM each]	1
Phusion DNA Polymerase NEB_M0530	0,5
MgCl2 [50mM]	2
H2O	qsp 50

Cycle:		
Step	1	94°C-2'
	2	94°C-20''
	3	50°C-20''
	4	72°C-20''
	5	4°C-∞

Back to step 2 - X35

3. **sgRNAs were generated by IVT** (In Vitro Transcription) using the previous obtained PCR products as templates. 1µg of these PCR templates was used for IVT using T7 Megascript kit (LifeTechnology_AM1333) with the specification of the manual. sgRNAs were finally purified using Phenol:Chloroform:Isoamyl Alcohol (25:24:1; v/v) and resuspended in 100µL of water.

4. Finally, for **cell transfection**, a 96 well-plate was seeded with 50.000 cells/well. Effectene (Qiagen_301427) was used to transfect cells with 50ng of HR template and 100ng of sgRNA. Four days later, cells were splitted 1:5 and submitted to selection using blasticidin (final concentration: 25µg/mL). Clonal dilutions were performed by seeding plates with three different cell concentrations: 8000 cells/mL, 1600 cells/mL and 320 cells/mL. Clonal cell lines were attested for *Pc* KO but IF using Pc Ab, western-blotting and whole genome sequencing to ensure no aspecific integration of the HR template.

Immunofluorescence experiments of S2 cells

Up to three millions of cells/mL in PBS1X were plated on a Poly-L-lysine coverslip and let attached for an hour at RT. Coverslip was rinsed into PBS1X and placed into a 6-well plate. Cells on slide were then fixed for 10 min using PBS1X, 4% Paraformaldehyde (3mL per well) and washed 3 times in PBS1X (1min, 3min and 5min). Cells were then permeabilized 15 min with PBS1X, 0.1% Triton and washed 3 times in PBT (PBS1X, 0.1% Triton). Cells were blocked with PBT, 2% BSA for 30 min. Finally, appropriate primary antibody diluted in PBT (1:250), 2% BSA was added on slides overnight (4°C). The day after, the slide was washed 3 times in PBT and the secondary antibody diluted in PBT, 2% BSA (1:200) was applied on the slide for 1 hour. After 3 washes, slide was incubated with DAPI (0.5µg/mL) for 10 min. Slide was washed twice in PBS and mounted on drop of Prolong reagent (LifeTechnologies_P36930).

A Buffer	Final Conc
HEPES 1M	20mM
EDTA 0.5M	0.15mM
EGTA 0.5M	0.15mM
KCl 1M	10mM
Spermidine 63.7mM	0.15mM
Spermine 28.7mM	0.15mM
DTT 0.1M	1mM
Complete, EDTA-free Protease Inhibitor Cocktail Tablets (Sigma-Aldrich_04693132001)	1

High Salt Buffer	Final Conc
Tris (pH7.65) 1M	20mM
EDTA 0.5M	0.2mM
Glycerol	25% (v/v)
NaCl 58M	900mM
MgCl2 1M	1.5mM

Sucrose Buffer	Final Conc
Tris (pH7.65) 1M	20mM
NaCl 5M	60mM
KCl 1M	15mM
Sucrose	0.34M
Spermidine 63.7mM	0.15mM
Spermine 28.7mM	0.15mM

SR Buffer	Final Conc
HEPES 1M	50mM
EDTA 0.5M	0.25mM
Sucrose	70% (m/v)
KCl 1M	10mM
Spermidine 63.7mM	0.15mM
Spermine 28.7mM	0.15mM
DTT 0.1M	1mM
Complete, EDTA-free Protease Inhibitor Cocktail Tablets (Sigma-Aldrich_04693132001)	1x

Cell Fractionation and Western blotting

The compositions of buffer used for cell fractionation are listed on the left page.
1.To lyse cells, the cell pellet was first estimated (Volume=V). Then, 3V of Buffer A was added to cells. Rapidly, 4V/9 NP40, 10% was added and mixed by inverting tube. Immediately, 8V/9 of SR buffer was added and mixed by inverting tube. The mix was then centrifuged for 5 min at 2.000g. The supernatant corresponding to cytoplasmic fraction was isolated.
2.The volume of the pellet (corresponding to nuclei) was estimated (Volume=V') and resuspended into V' of sucrose Buffer. Then, 0.4V of high salt buffer was added drop by drop while vortexing and the mix was kept on ice for 30 min. The total volume was estimated (Vtot) and 1/3Vtot of sucrose buffer was added. The mix was then centrifuged for 10 min at 10.000rpm. The supernatant corresponding to soluble nuclear fraction was isolated.
3. The volume of the pellet (corresponding to chromatin extract) was estimated (Volume=V'') and resuspended into 3V'' of sucrose Buffer. Then, MNase was added (0.0025U/µL) along with $CaCl_2$ (1mM final concentration) and incubated 10 min at 37°C. A sonication step was finally performed using Bioruptor (Diagenode; HIGH power; five cycles; 30sec ON; 30sec OFF). After centrifugation at 13.000rpm for 15 min, the supernatant corresponding to chromatin-associated proteins was isolated and quantified using BCA kit (Pierce). Western-blots were performed using antibodies against Pc (SCHUETTENGRUBER *et al.* 2009), H3K27me3 (Active Motif_39155; rabbit), H3K27ac (Active Motif_39685; mouse). All primary antibodies were diluted 1:1000 and secondary at 1:5000.

Immunofluorescence on whole mount embryos

Embryos were dechorianated in bleach (3.6%) for 3-4 min and placed into a 2mL tube. 900µL of 16% Formaldehyde was added followed by 900µl of Heptane. After mixing, tubes were placed on rotating wheel for 7 min for fixing. Formaldehyde was removed and substituted by 900 mL of Methanol (to remove the vitellin membrane) and tubes were immediately vortexed for 1 min. Embryos were then washed 3 times in Methanol. In a second step, fixed embryos were transferred stepwise to PBS1X, Tween20 0.1% using the following solutions: MeOH:PBT: 70:30%; 50:50%; 30:70%; 100% PBS1X, Tween20 0.1%. Embryos were blocked for 2 hours in PBT (PBS1X, Triton 0.3%) with 2% BSA. Primary antibody (against β-Gal) was diluted in PBT, 2% BSA (Promega_Z3781; 1:400) and incubated with embryos overnight at 4°C on a rotating wheel. The day after, embryos were washed in PBT (3 short washes of 5 min followed by 3 long washes of 20 min). Incubation with the secondary antibody was performed for 1 hour (α-mouse ALEXA 488, 1:500 in PBT). After washed in PBT, embryos were stained with DAPI diluted

in PBS1X, Tween20 0.1% (0.5µg/mL) for 10 min. After two washes in PBS1X, Tween20 0.1%, embryos where mounted with Prolong reagent (LifeTechnologies_P36930).

Preparation of chromatin

ChIPs were performed on Drosophila S2 cells. Cells were grown at 25°C in a standard medium until confluence. Then, cells were scrapped and transferred to 15-mL tubes (approximately 1T75 for four IPs). Cells were cross-linked by adding formaldehyde directly to culture media to a final concentration of 1.8% followed by an incubation of 10 min at room temperature. Cross-linking was stopped by adding 225mM glycine followed by incubation for 5 min. Cells were rinsed twice with 10 mL of cold PBS 1x + PMSF (1mM). Cells were subsequently washed in 10 mL of Buffer I (0.25% Triton X-100, 10mM EDTA, 0.5M EGTA, 10mM HEPES (pH 7.6) and protease inhibitors) and 10 mL of Buffer II (0.2M NaCl, 1mM EDTA, 0.5M EGTA, 10mM HEPES (pH 7.6) and protease inhibitors). Nuclei were resuspended in 1 mL of Lysis Buffer (140mM NaCl, 15mM HEPES (pH 7.6), 1mM EDTA, 0.5mM EGTA, 1% Triton X-100, 0.5 mM DTT, 0.1% sodium deoxycholate, 0.1% SDS, 0.5% N-lauroylsarcosine and protease inhibitors) and incubated for 30 min on a rotating wheel at 4°C. Chromatin was sonicated using a Bioruptor (Diagenode) for 10 min (settings 30 sec on, 30 sec off, high power). Sheared chromatin had an average length of 300 to 800 bp. After sonication and 10 min high speed centrifugation, fragmented chromatin was recovered in the supernatant.

Chromatin ImmunoPrecipitation

Chromatin was pre-cleared by addition of 100µL of Dynabeads (Invitrogen) followed by overnight incubation at 4°C. Beads were removed and chromatin was splitted into appropriate number of samples. Antibodies were added to the supernatant (a control in the absence of antibody called 'Mock IP' was performed at the same time) and samples were incubated for 4 hours at 4°C in a rotating wheel. 40µL of dynabeads was added and incubation was continued overnight at 4°C. Antibody-protein complexes were subsequently washed in three different buffers (each wash 5 min at 4°C): the FAT buffer (38.75mM TrisHCl (pH8), 7mM EDTA, 0.2% SDS, 56.25mM NaCl, 0.375% Triton X-100 and protease inhibitors), the FA buffer (100mM HEPES (pH 7.6), 2mM EDTA, 1M NaCl, 1% Triton X-100, 0.1% sodium deoxycholate and protease inhibitors) and LiCl Buffer (100mM HEPES (pH 7.6), 500mM LiCl, 1% Np40-Nonidet, 0.1% sodium deoxycholate and protease inhibitors). Two final washes were performed in 1mM EDTA, 10mM Tris (pH 8) (TE) buffer.
Chromatin was eluted from dynabeads in two steps; first in 60 µL of 10mM EDTA, 1% SDS, 50mM Tris (pH 8) at 65°C for 15 min followed by a chromatin recovery. Beads were re-incubated in 150ul of TE, 0.67% SDS, 15 min at 65°C. The combined eluate (150ul)

was incubated overnight at 65°C to reverse cross-links and treated with Proteinase K for 3h at 50°C.
DNA was purified on MicroChIP DiaPure columns (Diagenode) and resuspended in 20µL H_2O for sequencing or 100-200µL H_2O for qPCR.

q-Chip were performed using antibodies against Pc and Ph (SCHUETTENGRUBER *et al.* 2009) along with Psc (this study) for PRC1 and E(z) and Su(z)12 (LOUBIERE *et al.* 2016) for PRC2. Primers used to analyze the six classes of PRE in S2 cells are given in **Table 2** (see below).

Chip-seq were performed using antibodies against Ph and E(z) PcG components; H2AK118Ub (Ozyme, 8240S); and the following PcG recruiters: Pho (Brown et al., 2003), Phol (Wang et al., 2004), Spps (Brown et al., 2010), Gaf (SCHUETTENGRUBER *et al.* 2009), Psq (HUANG *et al.* 2002) and Dsp1 (MOSRIN-HUAMAN *et al.* 1998).
For ChIP-seq, 20 ng of the ChIP reaction and input DNA were used for the library preparation. ChIP-seq was performed by the Montpellier GenomiX facility (MGX, http://www.mgx.cnrs.fr/index.php).

Tableau 2 : List of primers used for q-ChIP in S2 cells

PRE_group	Oligoname	Sequence	length (bp)
1	IP3K1_S1	TCCAATTCCCAGAGAACACTTC	117
	IP3K1_AS1	GGATGGCGACTAACTGGTTTC	
	CG5953_S1	ACACCACTGCGCATTTCTAAG	109
	CG5953_AS1	TAGTGTTAGTGTGCTCGGCTTG	
2	yuri_S1	CTTGCCACGATCCAAATACC	152
	yuri_AS1	TGAGAGTCCGTCGAGTTTTTC	
	CG8596_S1	GACCACTGGAGATGAACGAAC	126
	CG8596_AS1	TTGGCTCCTCTCTCTCTTCG	
3	Mpcp_S1	GTCCATAGCCGAAGATTTTGAC	112
	Mpcp_AS1	GCAGCGATTTCTCACGAATAC	
	CG7372_S1	CTCACCAACACACATGTACGC	157
	CG7372_AS1	CGTTCAGAACAAATCCAAATCC	
4	edl_S1	TATCTGTGTCTGCGTTCTGCTC	90
	edl_AS1	GAGAGCGCAGTCAACAAAAAC	
	Btub60D_S1	AGCTCGATAACTCCGCATTG	164
	Btub60D_AS1	ATGTTCTGGTGACTTTGAGTCG	
5	DAC_TSS_S1	TCTCGCATTACACCAATACCAG	169
	DAC_TSS_AS1	GGGAGGATTTGAAGAACACAAC	
	Notch_S1	TGGGTTGGAAAGAGAGAGAGTG	137
	Notch_AS1	ACGAATAAACGAGTGTGGGAAC	
6	cad_mid_S1	CACATCGGCCTCTCTCTCTC	128
	cad_mid_AS1	AAACCTCCAGCTTTAGTCCAAG	
	cad_3end_S1	CTTCTCTTTACCCAATCGTTCG	101
	cad_3end_AS1	GACTGACTCGAATGTCGATTTG	
Controls	ZIF promoter_S1	cgctttacacaacactttagcaac	119
	ZIF promoter_AS1	cacgtctgtggggtaagctc	
	engrailed_S1	ggcttgttaggcagcaatatgac	148
	engrailed_AS1	tgaacagtgccgctatatgacc	

ChIP-seq data processing

ChIP-seq experiments were performed in duplicates and DNA samples were sequenced on HiSeq2000, filtered and aligned on D. melanogaster genome (Dm3, Release 5) with Bowtie 2. All ChIPs for DNA-binding proteins were analyzed using MACS2 with standard parameters and used the input as control. Only the peaks from MACS with a minimum enrichment of 3 fold and a maximum FDR of 10% were considered as enriched. Considering the high correlation between duplicates, the list of enriched sites were

obtained by intersecting the duplicates. This gave us a list of enriched sites for each DNA-binding protein. The 1543 E(z)/Ph sites were obtained by doing the intersection between the list of enriched sites of E(z) and that of Ph. Because Ph forms larger peaks than that of E(z), we used the option –wa of bedtools intersect to keep the size of Ph peaks. The H2AUb data was analyzed with SICER using standard parameters.
Data were visualized using IGB (Integrated Genome Browser) (http://bioviz.org/igb/).

Clustering analyses were performed using seqMINER (1.3.3) using the KMeans ranked as clustering normalization of ChIP-seq signals. The clustering results were plotted using deepTools package (https://github.com/fidelram/deepTools/) with a bin size of 10bp. To create boxplots with mean enrichments of DNA-binding proteins, the computeMatrix tool of deepTools was used. This allows to create matrices taking the mean enrichments of each bin (10bp) over the appropriate window. The maximum value over the window was then taken for each sites.

ChIP-seq data used in this study are listed in **Table 1**.

Hi-C seq analysis

Hi-C data were obtained from 16-18h old embryos (Ogiyama Yuki et al., unpublished data). Merged data were processed by juicer pipeline. Matrices were taken as Observed/Expected KR nomalised value at 1kb resolution by using juice box command line tools. 2D information from bed files corresponding to classes of PREs, E(z)- or Ph-only sites and RGR were obtained by intersection of two peaks within the appropriate size regions (1kb-25kb to see the insulation; 20kb-200kb to look at intra-domain interactions and 200kb-1Mb to look at inter-domain interactions). To get Hi-C matrices from 2D information, we used 1kb resolution Hi-C data and +/- 50 bins from 2D points were taken. Cumulative Observed/Expected KR values are shown as a heat map.

RNA-seq

We performed RNA-seq of two different cell lines mutant for *Pc* gene (cell lines B and C) and of S2 cell line expressing Cas9. RNA-seq was performed in triplicate for each cell line. Cells were grown until confluence. Then RNA was isolated using Trizol and Chloroform. RNA was purified using the RNeasy MiniKit (Quiagen). 4µg of RNA was used for RNA-seq that was performed by the Montpellier GenomiX facility (MGX, http://www.mgx.cnrs.fr/index.php).

Data were aligned on D.melanogaster genome (Dm6, Release 6.13) using Bowtie 2.2.9. Differentially expressed genes were then obtained using EdgeR 3.16.5. Genes with

adjusted p-value less than 5% (according to the FDR method from Benjamini-Hochberg) were declared differentially expressed.

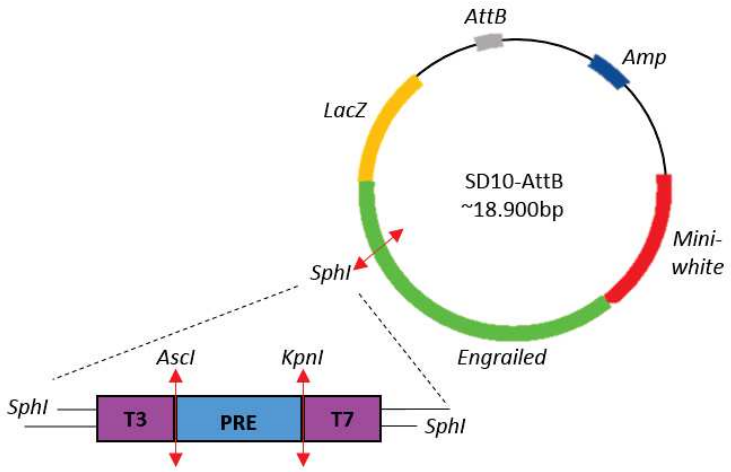


Figure 40 : Modified SD10 vector to test PRE activity.

The *attB* site containing SD10 vector contains the engrailed promoter along with regulatory elements downstream of the *mini-white* gene and upstream of the *LacZ* gene. The *SphI* site allows the insertion of sequences that contain the PRE to test. Each PRE was inserted to the sequence thanks to *AscI* and *KpnI* sites. The sequence contains T3 and T7 sequences that flanked the PRE to allow specific q-ChIP of the transgenic PRE without considering the endogenous PRE.

SD10 vector modifications

Original SD10 vector was obtained from (Cunningham, 2010). First, a *attB* sequence was inserted using *BamHI* restriction enzyme. The addition of that sequence allows for site-specific integration in flies using the ϕC31-integrase system into attp2 (chromosome 3L, cytological position 68A4). The resulting vector is shown in **Figure** 40. Then, a construct composed of a PRE flanked by restriction sites for *AscI* and *KpnI* and flanked by T3/T7 sequences was synthetized. In particular, the fragment was bordered by *SphI* restriction sites to allow its integration into SD10_AttB plasmid (**Figure** 40). The addition of T3/T7 sequences allows the analyses of the ectopic PREs by q-ChIP without considering endogenous copies. The fragment was inserted into SD10_AttB using *SphI* restriction site giving rise to the final plasmid named SD10_AttP_PRE. Other PRE sequences were inserted into the plasmid using the two restricted sites: *AscI* and *KpnI.* The PRE sequences flanked by *AscI* and *KpnI* restriction sites were obtained by PCR with primers listed below (**Table 3**). Finally, all PRE-containing plamids were sent to BestGene Inc for injection in flies.

Eye pigmentation assay

For the eye pigmentation assay, 5 heads of 4-day-old male flies were collected and homogenized in EPE buffer (30%EtOH-HCL [pH 2]) and incubated for 1 hour at 25°C in the dark. After centrifugation, eye pigmentation was quantified by spectrometry at 480 nm. The PSS score was calculated from mean pigment levels as heterozygote/homozygote for each PRE. Transgenic flies were photographed with a standard light microscope, always using the same time of exposure for each diverged PRE and controls

Polytene chromosomes

Polytenes were fixed in PBS-3.7% formaldehyde for 3 min at RT and washed in PBS for 15 min. A standard FISH protocol (LAVROV *et al.* 2004) was then applied to the slides, except that chromosomes were incubated in a 2XSSC bath at 70°C for 45min prior to alkali denaturation using NaOH 0.07M for 10 min at RT. FISH probes (with Attp2 templates) were labeled using the FISH Tag DNA Multicolor kit (Invitrogen) according to the manufacturer's instructions. Immunostaining of polytene chromosomes was performed using goat and rabbit polyclonal antibodies specific for PH (1:500) (o/n, 4°C). After a washing step, an appropriate secondary antibodies were used (1:200, 1h, RT), diluted in blocking serum with 2% NDS. Chromosomes were then counterstained with DAPI and mounted for fluorescence microscopy in Vectashield (Vector laboratories).

Tableau 3: PCR primers used to sequence the PREs.

Each primer contains a flanked site corresponding to either *AscI* or *KpnI*. The primers were designed in order for the integrated PRE to be in the same orientation according to its endogenous target gene.

PRE_group	Oligoname	Sequence	PRE length (bp)
1	IP3K1_S1	AAAGGCGCGCCCATCCCGAGATATCCAGAAGAC	652
	IP3K1_AS1	AGTGGTACCTGGAGCACCGTAGAAGTCATAG	
	CG5953_S1	AGTGGTACCGGGGACTCAAAATAACAAACAG	796
	CG5953_AS1	ACTGGCGCGCCTGTGACCAATTTTTCGGCTTAG	
2	yuri_S2	AGTGGTACCCAAATCGCAAATCACAAATCAC	689
	yuri_AS1	ATCGGCGCGCCATGCGACCAAACAAACTTATCC	
	CG8596_S1	AATGGTACCGCTTCCTCTCACTCGATTTTTG	673
	CG8596_AS1	ATTGGCGCGCCACGCAGGTAGAAGTTTGGTAGG	
3	Mpcp_S1	ATTGGCGCGCCCATGTGAAATCATGGCAGAAC	741
	Mpcp_AS1	AGAGGTACCCGGACTTACGACACCAAATTTAC	
	CG7372_S1	ATTGGCGCGCCGAGTTTTGGCCGCTATGTTG	651
	CG7372_AS1	GGCGGTACCCTCCTCGTCTTCCTCATCCTC	
4	edl_S1	GGCGGTACCGCTTCGTTTTTCACTTAACACG	688
	edl_AS1	TATGGCGCGCCGTTGGCTCAATAAGCAGGAAAG	
	Btub60D_S1	AATGGCGCGCCAAGTTCAATGGTGTAAGGGAGTG	644
	Btub60D_AS2	ATTGGTACCCGGCCTGCAGGTTCACGATTTC	
5	DAC_TSS_S1	CTTGGTACCTTCTCTTGCGAATGACAAATTC	573
	DAC_TSS_AS1	GAAGGCGCGCCGACTGGCTATGCAAATATGCTG	
	Notch_S1	ATCGGCGCGCCACATACTTCCTCCCCCATGTTG	650
	Notch_AS1	GGCGGTACCTGGCTGTAAGTGTAGTGGCAAG	
6	cad_mid_S1	ATTGGCGCGCCTTTCCTTTGTTGCCTCTAATGC	774
	cad_mid_AS1	AGCGGTACCATATAAATTCCGGTCCCTCCTC	
	cad_3end_S1	ATCGGCGCGCCACGTATTCACCCGCATCTTTAC	676
	cad_3end_AS1	GGCGGTACCTGCAACTGCTCAAAATACTCAG	

REFERENCES

Abed, J. A., C. L. Cheng, C. R. Crowell, L. L. Madigan, E. Onwuegbuchu *et al.*, 2013 Mapping polycomb response elements at the Drosophilla melanogaster giant locus. G3 (Bethesda) 3**:** 2297-2304.

Abed, J. A., and R. S. Jones, 2012 H3K36me3 key to Polycomb-mediated gene silencing in lineage specification. Nat Struct Mol Biol 19**:** 1214-1215.

Adam, M., F. Robert, M. Larochelle and L. Gaudreau, 2001 H2A.Z is required for global chromatin integrity and for recruitment of RNA polymerase II under specific conditions. Mol Cell Biol 21**:** 6270-6279.

Akam, M., 1987 The molecular basis for metameric pattern in the Drosophila embryo. Development 101**:** 1-22.

Akasaka, T., N. Takahashi, M. Suzuki, H. Koseki, R. Bodmer *et al.*, 2002 MBLR, a new RING finger protein resembling mammalian Polycomb gene products, is regulated by cell cycle-dependent phosphorylation. Genes Cells 7**:** 835-850.

Alfieri, C., M. C. Gambetta, R. Matos, S. Glatt, P. Sehr *et al.*, 2013 Structural basis for targeting the chromatin repressor Sfmbt to Polycomb response elements. Genes Dev 27**:** 2367-2379.

Anderson, A. E., U. C. Karandikar, K. L. Pepple, Z. Chen, A. Bergmann *et al.*, 2011 The enhancer of trithorax and polycomb gene Caf1/p55 is essential for cell survival and patterning in Drosophila development. Development 138**:** 1957-1966.

Angelini, D. R., P. Z. Liu, C. L. Hughes and T. C. Kaufman, 2005 Hox gene function and interaction in the milkweed bug Oncopeltus fasciatus (Hemiptera). Dev Biol 287**:** 440-455.

Ardehali, M. B., A. Mei, K. L. Zobeck, M. Caron, J. T. Lis *et al.*, 2011 Drosophila Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. EMBO J.

Arents, G., R. W. Burlingame, B. C. Wang, W. E. Love and E. N. Moudrianakis, 1991 The nucleosomal core histone octamer at 3.1 A resolution: a tripartite protein assembly and a left-handed superhelix. Proc Natl Acad Sci U S A 88**:** 10148-10152.

Arif, M., G. V. Kumar, C. Narayana and T. K. Kundu, 2007 Autoacetylation induced specific structural changes in histone acetyltransferase domain of p300: probed by surface enhanced Raman spectroscopy. J Phys Chem B 111**:** 11877-11879.

Arnold, C. D., D. Gerlach, C. Stelzer, L. M. Boryn, M. Rath *et al.*, 2013 Genome-wide quantitative enhancer activity maps identified by STARR-seq. Science 339**:** 1074-1077.

Atchison, L., A. Ghias, F. Wilkinson, N. Bonini and M. L. Atchison, 2003 Transcription factor YY1 functions as a PcG protein in vivo. EMBO J 22**:** 1347-1358.

Ballare, C., M. Lange, A. Lapinaite, G. M. Martin, L. Morey *et al.*, 2012 Phf19 links methylated Lys36 of histone H3 to regulation of Polycomb activity. Nat Struct Mol Biol 19**:** 1257-1265.

Banaszynski, L. A., D. Wen, S. Dewell, S. J. Whitcomb, M. Lin *et al.*, 2013 Hira-dependent histone H3.3 deposition facilitates PRC2 recruitment at developmental loci in ES cells. Cell 155**:** 107-120.

Bannister, A. J., and T. Kouzarides, 2011 Regulation of chromatin by histone modifications. Cell Res 21**:** 381-395.

Bantignies, F., and G. Cavalli, 2011 Polycomb group proteins: repression in 3D. Trends Genet 27**:** 454-464.

Bantignies, F., R. H. Goodman and S. M. Smolik, 2000 Functional interaction between the coactivator Drosophila CREB-binding protein and ASH1, a member of the trithorax group of chromatin modifiers. Mol Cell Biol 20**:** 9317-9330.

Bantignies, F., C. Grimaud, S. Lavrov, M. Gabut and G. Cavalli, 2003 Inheritance of Polycomb-dependent chromosomal interactions in Drosophila. Genes Dev 17**:** 2406-2420.

Bantignies, F., V. Roure, I. Comet, B. Leblanc, B. Schuettengruber *et al.*, 2011 Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell 144**:** 214-226.

Basu, A., F. H. Wilkinson, K. Colavita, C. Fennelly and M. L. Atchison, 2014 YY1 DNA binding and interaction with YAF2 is essential for Polycomb recruitment. Nucleic Acids Res 42**:** 2208-2223.

Beisel, C., A. Buness, I. M. Roustan-Espinosa, B. Koch, S. Schmitt *et al.*, 2007 Comparing active and repressed expression states of genes controlled by the Polycomb/Trithorax group proteins. Proc Natl Acad Sci U S A 104**:** 16615-16620.

Belmont, A. S., 2002 Mitotic chromosome scaffold structure: new approaches to an old controversy. Proc Natl Acad Sci U S A 99**:** 15855-15857.

Beltran, M., C. M. Yates, L. Skalska, M. Dawson, F. P. Reis *et al.*, 2016 The interaction of PRC2 with RNA or chromatin is mutually antagonistic. Genome Res 26**:** 896-907.

Bernstein, B. E., T. S. Mikkelsen, X. Xie, M. Kamal, D. J. Huebert *et al.*, 2006 A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125**:** 315-326.

Bezsonova, I., J. R. Walker, J. P. Bacik, S. Duan, S. Dhe-Paganon *et al.*, 2009 Ring1B contains a ubiquitin-like docking module for interaction with Cbx proteins. Biochemistry 48**:** 10542-10548.

Bhat, M. A., A. V. Philp, D. M. Glover and H. J. Bellen, 1996 Chromatid segregation at anaphase requires the barren product, a novel chromosome-associated protein that interacts with Topoisomerase II. Cell 87**:** 1103-1114.

Birve, A., A. K. Sengupta, D. Beuchle, J. Larsson, J. A. Kennison *et al.*, 2001 Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. Development 128**:** 3371-3379.

Blackledge, N. P., A. M. Farcas, T. Kondo, H. W. King, J. F. McGouran *et al.*, 2014 Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell 157**:** 1445-1459.

Blastyak, A., R. K. Mishra, F. Karch and H. Gyurkovics, 2006 Efficient and specific targeting of Polycomb group proteins requires cooperative interaction between Grainyhead and Pleiohomeotic. Mol Cell Biol 26**:** 1434-1444.

Boettiger, A. N., B. Bintu, J. R. Moffitt, S. Wang, B. J. Beliveau *et al.*, 2016 Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. Nature 529**:** 418-422.

Bonasio, R., E. Lecona, V. Narendra, P. Voigt, F. Parisi *et al.*, 2014 Interactions with RNA direct the Polycomb group protein SCML2 to chromatin where it represses target genes. Elife 3**:** e02637.

Bonev, B., and G. Cavalli, 2016 Organization and function of the 3D genome. Nat Rev Genet 17**:** 661-678.

Bottcher, R., M. Hollmann, K. Merk, V. Nitschko, C. Obermaier *et al.*, 2014 Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells. Nucleic Acids Res 42**:** e89.

Boyer, L. A., K. Plath, J. Zeitlinger, T. Brambrink, L. A. Medeiros *et al.*, 2006 Polycomb complexes repress developmental regulators in murine embryonic stem cells. Nature. 441**:** 349-353. Epub 2006 Apr 2019.

Bracken, A. P., N. Dietrich, D. Pasini, K. H. Hansen and K. Helin, 2006 Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev. 20**:** 1123-1136. Epub 2006 Apr 1117.

Bray, S. J., B. Burke, N. H. Brown and J. Hirsh, 1989 Embryonic expression pattern of a family of Drosophila proteins that interact with a central nervous system regulatory element. Genes Dev 3**:** 1130-1145.

Breen, T. R., and I. M. Duncan, 1986 Maternal expression of genes that regulate the bithorax complex of Drosophila melanogaster. Dev Biol 118**:** 442-456.

Breiling, A., E. Bonte, S. Ferrari, P. B. Becker and R. Paro, 1999 The Drosophila polycomb protein interacts with nucleosomal core particles In vitro via its repression domain. Mol Cell Biol 19**:** 8451-8460.

Brien, G. L., G. Gambero, D. J. O'Connell, E. Jerman, S. A. Turner *et al.*, 2012 Polycomb PHF19 binds H3K36me3 and recruits PRC2 and demethylase NO66 to embryonic stem cell genes during differentiation. Nat Struct Mol Biol 19**:** 1273-1281.

Brien, G. L., E. Healy, E. Jerman, E. Conway, E. Fadda *et al.*, 2015 A chromatin-independent role of Polycomb-like 1 to stabilize p53 and promote cellular quiescence. Genes Dev 29**:** 2231-2243.

Brookes, E., I. de Santiago, D. Hebenstreit, K. J. Morris, T. Carroll *et al.*, 2012 Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. Cell Stem Cell 10**:** 157-170.

Brown, J. L., C. Fritsch, J. Mueller and J. A. Kassis, 2003 The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. Development 130**:** 285-294.

Brown, J. L., D. J. Grau, S. K. DeVido and J. A. Kassis, 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.

Brown, J. L., and J. A. Kassis, 2010 Spps, a Drosophila Sp1/KLF family member, binds to PREs and is required for PRE activity late in development. Development 137**:** 2597-2602.

Brown, J. L., D. Mucci, M. Whiteley, M. L. Dirksen and J. A. Kassis, 1998 The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. Mol Cell. 1**:** 1057-1064.

Brunk, B. P., E. C. Martin and P. N. Adler, 1991 Drosophila genes Posterior Sex Combs and Suppressor two of zeste encode proteins with homology to the murine bmi-1 oncogene. Nature 353**:** 351-353.

Buchwald, G., P. van der Stoop, O. Weichenrieder, A. Perrakis, M. van Lohuizen *et al.*, 2006 Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. EMBO J 25**:** 2465-2474.

Bulger, M., and M. Groudine, 2002 TRAPping enhancer function. Nat Genet 32**:** 555-556.

Bunker, C. A., and R. E. Kingston, 1994 Transcriptional repression by Drosophila and mammalian Polycomb group proteins in transfected mammalian cells. Mol Cell Biol 14**:** 1721-1732.

Busturia, A., A. Lloyd, F. Bejarano, M. Zavortink, H. Xin *et al.*, 2001 The MCP silencer of the Drosophila Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. Development. 128**:** 2163-2173.

Buszczak, M., S. Paterno and A. C. Spradling, 2009 Drosophila stem cells share a common requirement for the histone H2B ubiquitin protease scrawny. Science 323**:** 248-251.

Butler, J. S., and S. Y. Dent, 2013 The role of chromatin modifiers in normal and malignant hematopoiesis. Blood 121**:** 3076-3084.

Cai, Y., J. Jin, T. Yao, A. J. Gottschalk, S. K. Swanson *et al.*, 2007 YY1 functions with INO80 to activate transcription. Nat Struct Mol Biol 14**:** 872-874.

Cao, R., H. Wang, J. He, H. Erdjument-Bromage, P. Tempst *et al.*, 2008 Role of hPHF1 in H3K27 methylation and Hox gene silencing. Mol Cell Biol 28**:** 1862-1872.

Cao, R., L. Wang, H. Wang, L. Xia, H. Erdjument-Bromage *et al.*, 2002 Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298**:** 1039-1043.

Cao, R., and Y. Zhang, 2004 SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. Mol Cell 15**:** 57-67.

Cavalli, G., and T. Misteli, 2013 Functional implications of genome topology. Nat Struct Mol Biol 20**:** 290-299.

Chittock, E. C., S. Latwiel, T. C. Miller and C. W. Muller, 2017 Molecular architecture of polycomb repressive complexes. Biochem Soc Trans 45**:** 193-205.

Chopra, V. S., D. A. Hendrix, L. J. Core, C. Tsui, J. T. Lis *et al.*, 2011 The polycomb group mutant esc leads to augmented levels of paused Pol II in the Drosophila embryo. Mol Cell 42**:** 837-844.

Ciferri, C., G. C. Lander, A. Maiolica, F. Herzog, R. Aebersold *et al.*, 2012 Molecular architecture of human polycomb repressive complex 2. Elife 1**:** e00005.

Coleman, R. T., and G. Struhl, 2017 Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. Science 356.

Comet, I., B. Schuettengruber, T. Sexton and G. Cavalli, 2011 A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. Proc Natl Acad Sci U S A 108**:** 2294-2299.

Cooper, S., M. Dienstbier, R. Hassan, L. Schermelleh, J. Sharif *et al.*, 2014 Targeting polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. Cell Rep 7**:** 1456-1470.

Cooper, S., A. Grijzenhout, E. Underwood, K. Ancelin, T. Zhang *et al.*, 2016 Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. Nat Commun 7**:** 13661.

Costanzi, C., and J. R. Pehrson, 1998 Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. Nature 393**:** 599-601.

Coulson, M., S. Robert, H. J. Eyre and R. Saint, 1998 The identification and localization of a human gene with sequence similarity to Polycomblike of Drosophila melanogaster. Genomics 48**:** 381-383.

Creppe, C., A. Palau, R. Malinverni, V. Valero and M. Buschbeck, 2014 A Cbx8-containing polycomb complex facilitates the transition to gene activation during ES cell differentiation. PLoS Genet 10**:** e1004851.

Creyghton, M. P., S. Markoulaki, S. S. Levine, J. Hanna, M. A. Lodato *et al.*, 2008 H2AZ is enriched at polycomb complex target genes in ES cells and is necessary for lineage commitment. Cell 135**:** 649-661.

Cubenas-Potts, C., M. J. Rowley, X. Lyu, G. Li, E. P. Lei *et al.*, 2017 Different enhancer classes in Drosophila bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. Nucleic Acids Res 45**:** 1714-1730.

Cuddapah, S., T. Y. Roh, K. Cui, C. C. Jose, M. T. Fuller *et al.*, 2012 A novel human polycomb binding site acts as a functional polycomb response element in Drosophila. PLoS One 7**:** e36365.

Cunningham, M. D., J. L. Brown and J. A. Kassis, 2010 Characterization of the polycomb group response elements of the Drosophila melanogaster invected Locus. Mol Cell Biol 30**:** 820-828.

Czermin, B., R. Melfi, D. McCabe, V. Seitz, A. Imhof *et al.*, 2002 Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111**:** 185-196.

Davidovich, C., and T. R. Cech, 2015 The recruitment of chromatin modifiers by long noncoding RNAs: lessons from PRC2. RNA 21**:** 2007-2022.

de Napoles, M., J. E. Mermoud, R. Wakao, Y. A. Tang, M. Endoh *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.

Decoville, M., E. Giacomello, M. Leng and D. Locker, 2001 DSP1, an HMG-like protein, is involved in the regulation of homeotic genes. Genetics 157**:** 237-244.

Dejardin, J., and G. Cavalli, 2004 Chromatin inheritance upon Zeste-mediated Brahma recruitment at a minimal cellular memory module. Embo J. 23**:** 857-868. Epub 2004 Feb 2012.

Dejardin, J., A. Rappailles, O. Cuvier, C. Grimaud, M. Decoville *et al.*, 2005 Recruitment of Drosophila Polycomb group proteins to chromatin by DSP1. Nature 434**:** 533-538.

Delvecchio, M., J. Gaucher, C. Aguilar-Gurrieri, E. Ortega and D. Panne, 2013 Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. Nat Struct Mol Biol 20**:** 1040-1046.

Deshaies, R. J., and C. A. Joazeiro, 2009 RING domain E3 ubiquitin ligases. Annu Rev Biochem 78**:** 399-434.

DeVido, S. K., D. Kwon, J. L. Brown and J. A. Kassis, 2008 The role of Polycomb-group response elements in regulation of engrailed transcription in Drosophila. Development 135**:** 669-676.

Di Croce, L., and K. Helin, 2013 Transcriptional regulation by Polycomb group proteins. Nat Struct Mol Biol 20**:** 1147-1155.

Dillon, S. C., X. Zhang, R. C. Trievel and X. Cheng, 2005 The SET-domain protein superfamily: protein lysine methyltransferases. Genome Biol 6**:** 227.

Dixon, J. R., S. Selvaraj, F. Yue, A. Kim, Y. Li *et al.*, 2012 Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485**:** 376-380.

Dorigo, B., T. Schalch, A. Kulangara, S. Duda, R. R. Schroeder *et al.*, 2004 Nucleosome arrays reveal the two-start organization of the chromatin fiber. Science 306**:** 1571-1573.

Dowen, J. M., Z. P. Fan, D. Hnisz, G. Ren, B. J. Abraham *et al.*, 2014 Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell 159**:** 374-387.

Drews, D., M. Klar, C. Dame and A. U. Brauer, 2009 Developmental expression profile of the YY2 gene in mice. BMC Dev Biol 9**:** 45.

Duncan, I. M., 1982 Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. Genetics 102**:** 49-70.

Dutnall, R. N., and V. Ramakrishnan, 1997 Twists and turns of the nucleosome: tails without ends. Structure 5**:** 1255-1259.

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

El-Sharnouby, S., B. Fischer, J. P. Magbanua, B. Umans, R. Flower *et al.*, 2017 Regions of very low H3K27me3 partition the Drosophila genome into topological domains. PLoS One 12**:** e0172725.

Emmons, R. B., H. Genetti, S. Filandrinos, J. Lokere and C. T. Wu, 2009 Molecular genetic analysis of Suppressor 2 of zeste identifies key functional domains. Genetics 182**:** 999-1013.

Enderle, D., C. Beisel, M. B. Stadler, M. Gerstung, P. Athri *et al.*, 2011 Polycomb preferentially targets stalled promoters of coding and noncoding transcripts. Genome Res 21**:** 216-226.

Endoh, M., T. A. Endo, T. Endoh, Y. Fujimura, O. Ohara *et al.*, 2008 Polycomb group proteins Ring1A/B are functionally linked to the core transcriptional regulatory circuitry to maintain ES cell identity. Development 135**:** 1513-1524.

Endoh, M., T. A. Endo, T. Endoh, K. Isono, J. Sharif *et al.*, 2012 Histone H2A mono-ubiquitination is a crucial step to mediate PRC1-dependent repression of developmental genes to maintain ES cell identity. PLoS Genet 8**:** e1002774.

Endoh, M., T. A. Endo, J. Shinga, K. Hayashi, A. Farcas *et al.*, 2017 PCGF6-PRC1 suppresses premature differentiation of mouse embryonic stem cells by regulating germ cell-related genes. Elife 6.

Entrevan, M., B. Schuettengruber and G. Cavalli, 2016 Regulation of Genome Architecture and Function by Polycomb Proteins. Trends Cell Biol 26**:** 511-525.

Erceg, J., T. Pakozdi, R. Marco-Ferreres, Y. Ghavi-Helm, C. Girardot *et al.*, 2017 Dual functionality of cis-regulatory elements as developmental enhancers and Polycomb response elements. Genes Dev.

Ernst, J., P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward *et al.*, 2011 Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473**:** 43-49.

Eskeland, R., M. Leeb, G. R. Grimes, C. Kress, S. Boyle *et al.*, 2010 Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. Mol Cell 38**:** 452-464.

Farcas, A. M., N. P. Blackledge, I. Sudbery, H. K. Long, J. F. McGouran *et al.*, 2012 KDM2B links the Polycomb Repressive Complex 1 (PRC1) to recognition of CpG islands. Elife 1**:** e00205.

Feller, C., M. Prestel, H. Hartmann, T. Straub, J. Soding *et al.*, 2012 The MOF-containing NSL complex associates globally with housekeeping genes, but activates only a defined subset. Nucleic Acids Res 40**:** 1509-1522.

Fernandez-Capetillo, O., A. Celeste and A. Nussenzweig, 2003 Focusing on foci: H2AX and the recruitment of DNA-damage response factors. Cell Cycle 2**:** 426-427.

Ferrari, K. J., A. Scelfo, S. Jammula, A. Cuomo, I. Barozzi *et al.*, 2014 Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. Mol Cell 53**:** 49-62.

Fiedler, T., and M. Rehmsmeier, 2006 jPREdictor: a versatile tool for the prediction of cis-regulatory elements. Nucleic Acids Res. 34**:** W546-550.

Filion, G. J., J. G. van Bemmel, U. Braunschweig, W. Talhout, J. Kind *et al.*, 2010 Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell 143**:** 212-224.

Finch, J. T., and A. Klug, 1976 Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A 73**:** 1897-1901.

Fiorini, A., S. Gouveia Fde and M. A. Fernandez, 2006 Scaffold/Matrix Attachment Regions and intrinsic DNA curvature. Biochemistry (Mosc) 71**:** 481-488.

Fischle, W., B. S. Tseng, H. L. Dormann, B. M. Ueberheide, B. A. Garcia *et al.*, 2005 Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. Nature 438**:** 1116-1122.

Follmer, N. E., A. H. Wani and N. J. Francis, 2012 A polycomb group protein is retained at specific sites on chromatin in mitosis. PLoS Genet 8**:** e1003135.

Francis, N. J., R. E. Kingston and C. L. Woodcock, 2004 Chromatin compaction by a polycomb group protein complex. Science 306**:** 1574-1577.

Francis, N. J., A. J. Saurin, Z. Shao and R. E. Kingston, 2001 Reconstitution of a functional core polycomb repressive complex. Mol Cell 8**:** 545-556.

Frangini, A., M. Sjoberg, M. Roman-Trufero, G. Dharmalingam, V. Haberle *et al.*, 2013 The aurora B kinase and the polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. Mol Cell 51**:** 647-661.

Frey, F., T. Sheahan, K. Finkl, G. Stoehr, M. Mann *et al.*, 2016 Molecular basis of PRC1 targeting to Polycomb response elements by PhoRC. Genes Dev 30**:** 1116-1127.

Friberg, A., A. Oddone, T. Klymenko, J. Muller and M. Sattler, 2010 Structure of an atypical Tudor domain in the Drosophila Polycomblike protein. Protein Sci 19**:** 1906-1916.

Fritsch, C., D. Beuchle and J. Muller, 2003 Molecular and genetic analysis of the Polycomb group gene Sex combs extra/Ring in Drosophila. Mech Dev 120**:** 949-954.

Fritsch, C., J. L. Brown, J. A. Kassis and J. Muller, 1999 The DNA-binding polycomb group protein pleiohomeotic mediates silencing of a Drosophila homeotic gene. Development. 126**:** 3905-3913.

Fuda, N. J., M. J. Guertin, S. Sharma, C. G. Danko, A. L. Martins *et al.*, 2015 GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. PLoS Genet 11**:** e1005108.

Fujioka, M., G. L. Yusibova, J. Zhou and J. B. Jaynes, 2008 The DNA-binding Polycomb-group protein Pleiohomeotic maintains both active and repressed transcriptional states through a single site. Development 135**:** 4131-4139.

Gao, Z., P. Lee, J. M. Stafford, M. von Schimmelmann, A. Schaefer *et al.*, 2014 An AUTS2-Polycomb complex activates gene expression in the CNS. Nature 516**:** 349-354.

Gao, Z., J. Zhang, R. Bonasio, F. Strino, A. Sawai *et al.*, 2012 PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. Mol Cell 45**:** 344-356.

Gaytan de Ayala Alonso, A., L. Gutierrez, C. Fritsch, B. Papp, D. Beuchle *et al.*, 2007 A genetic screen identifies novel polycomb group genes in Drosophila. Genetics 176**:** 2099-2108.

Gearhart, M. D., C. M. Corcoran, J. A. Wamstad and V. J. Bardwell, 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.

Geisler, S. J., and R. Paro, 2015 Trithorax and Polycomb group-dependent regulation: a tale of opposing activities. Development 142**:** 2876-2887.

Goldberg, M. L., R. A. Colvin and A. F. Mellin, 1989 The Drosophila zeste locus is nonessential. Genetics 123**:** 145-155.

Gonzalez, I., J. Mateos-Langerak, A. Thomas, T. Cheutin and G. Cavalli, 2014 Identification of regulators of the three-dimensional polycomb organization by a microscopy-based genome-wide RNAi screen. Mol Cell 54**:** 485-499.

Gordon, S., G. Akopyan, H. Garban and B. Bonavida, 2006 Transcription factor YY1: structure, function, and therapeutic implications in cancer biology. Oncogene 25**:** 1125-1142.

Goto, H., Y. Yasui, E. A. Nigg and M. Inagaki, 2002 Aurora-B phosphorylates Histone H3 at serine28 with regard to the mitotic chromosome condensation. Genes Cells 7**:** 11-17.

Grau, D. J., B. A. Chapman, J. D. Garlick, M. Borowsky, N. J. Francis *et al.*, 2011 Compaction of chromatin by diverse Polycomb group proteins requires localized regions of high charge. Genes Dev 25**:** 2210-2221.

Grigoryev, S. A., G. Arya, S. Correll, C. L. Woodcock and T. Schlick, 2009 Evidence for heteromorphic chromatin fibers from analysis of nucleosome interactions. Proc Natl Acad Sci U S A 106**:** 13317-13322.

Grijzenhout, A., J. Godwin, H. Koseki, M. R. Gdula, D. Szumska *et al.*, 2016 Functional analysis of AEBP2, a PRC2 Polycomb protein, reveals a Trithorax phenotype in embryonic development and in ESCs. Development 143**:** 2716-2723.

Grimaud, C., N. Negre and G. Cavalli, 2006 From genetics to epigenetics: the tale of Polycomb group and trithorax group genes. Chromosome Res 14**:** 363-375.

Grimm, C., R. Matos, N. Ly-Hartig, U. Steuerwald, D. Lindner *et al.*, 2009 Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. EMBO J.

Guo, W. J., S. Datta, V. Band and G. P. Dimri, 2007 Mel-18, a polycomb group protein, regulates cell proliferation and senescence via transcriptional repression of Bmi-1 and c-Myc oncoproteins. Mol Biol Cell 18**:** 536-546.

Gutierrez, L., K. Oktaba, J. C. Scheuermann, M. C. Gambetta, N. Ly-Hartig *et al.*, 2012 The role of the histone H2A ubiquitinase Sce in Polycomb repression. Development 139**:** 117-127.

Hansen, K. H., A. P. Bracken, D. Pasini, N. Dietrich, S. S. Gehani *et al.*, 2008 A model for transmission of the H3K27me3 epigenetic mark. Nat Cell Biol 10**:** 1291-1300.

Harding, K., C. Wedeen, W. McGinnis and M. Levine, 1985 Spatially regulated expression of homeotic genes in Drosophila. Science 229**:** 1236-1242.

Hardy, S., P. E. Jacques, N. Gevry, A. Forest, M. E. Fortin *et al.*, 2009 The euchromatic and heterochromatic landscapes are shaped by antagonizing effects of transcription on H2A.Z deposition. PLoS Genet 5**:** e1000687.

Harikumar, A., and E. Meshorer, 2015 Chromatin remodeling and bivalent histone modifications in embryonic stem cells. EMBO Rep 16**:** 1609-1619.

Hauri, S., F. Comoglio, M. Seimiya, M. Gerstung, T. Glatter *et al.*, 2016 A High-Density Map for Navigating the Human Polycomb Complexome. Cell Rep 17**:** 583-595.

Hayakawa, T., and J. Nakayama, 2011 Physiological roles of class I HDAC complex and histone demethylase. J Biomed Biotechnol 2011**:** 129383.

He, G. P., S. Kim and H. S. Ro, 1999 Cloning and characterization of a novel zinc finger transcriptional repressor. A direct role of the zinc finger motif in repression. J Biol Chem 274**:** 14678-14684.

Heintzman, N. D., and B. Ren, 2009 Finding distal regulatory elements in the human genome. Curr Opin Genet Dev 19**:** 541-549.

Henikoff, S., 2008 Nucleosome destabilization in the epigenetic regulation of gene expression. Nat Rev Genet 9**:** 15-26.

Hernandez-Munoz, I., P. Taghavi, C. Kuijl, J. Neefjes and M. van Lohuizen, 2005 Association of BMI1 with polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. Mol Cell Biol 25**:** 11047-11058.

Herz, H. M., M. Mohan, A. S. Garruss, K. Liang, Y. H. Takahashi *et al.*, 2012 Enhancer-associated H3K4 monomethylation by Trithorax-related, the Drosophila homolog of mammalian Mll3/Mll4. Genes Dev 26**:** 2604-2620.

Herzog, V. A., A. Lempradl, J. Trupke, H. Okulski, C. Altmutter *et al.*, 2014 A strand-specific switch in noncoding transcription switches the function of a Polycomb/Trithorax response element. Nat Genet 46**:** 973-981.

Hodgson, J. W., B. Argiropoulos and H. W. Brock, 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.

Holohan, E. E., C. Kwong, B. Adryan, M. Bartkuhn, M. Herold *et al.*, 2007 CTCF genomic binding sites in Drosophila and the organisation of the bithorax complex. PLoS Genet. 3**:** e112.

Horowitz, H., and C. A. Berg, 1996 The Drosophila pipsqueak gene encodes a nuclear BTB-domain-containing protein required early in oogenesis. Development 122**:** 1859-1871.

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

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

Hu, D., X. Gao, M. A. Morgan, H. M. Herz, E. R. Smith *et al.*, 2013 The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. Mol Cell Biol 33**:** 4745-4754.

Huang, D. H., and Y. L. Chang, 2004 Isolation and characterization of CHRASCH, a polycomb-containing silencing complex. Methods Enzymol 377**:** 267-282.

Huang, D. H., Y. L. Chang, C. C. Yang, I. C. Pan and B. King, 2002 pipsqueak encodes a factor essential for sequence-specific targeting of a polycomb group protein complex. Mol Cell Biol 22**:** 6261-6271.

Hug, C. B., A. G. Grimaldi, K. Kruse and J. M. Vaquerizas, 2017 Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription. Cell 169**:** 216-228 e219.

Hur, M. W., J. D. Laney, S. H. Jeon, J. Ali and M. D. Biggin, 2002 Zeste maintains repression of Ubx transgenes: support for a new model of Polycomb repression. Development 129**:** 1339-1343.

Imhof, A., and A. P. Wolffe, 1998 Transcription: gene control by targeted histone acetylation. Curr Biol 8**:** R422-424.

Isono, K., T. A. Endo, M. Ku, D. Yamada, R. Suzuki *et al.*, 2013 SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. Dev Cell 26**:** 565-577.

Jacobs, J. J., B. Scheijen, J. W. Voncken, K. Kieboom, A. Berns *et al.*, 1999 Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. Genes Dev 13**:** 2678-2690.

Jansen, A., and K. J. Verstrepen, 2011 Nucleosome positioning in Saccharomyces cerevisiae. Microbiol Mol Biol Rev 75**:** 301-320.

Jiang, N., E. Emberly, O. Cuvier and C. M. Hart, 2009 Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. Mol Cell Biol 29**:** 3556-3568.

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

Junco, S. E., R. Wang, J. C. Gaipa, A. B. Taylor, V. Schirf *et al.*, 2013 Structure of the polycomb group protein PCGF1 in complex with BCOR reveals basis for binding selectivity of PCGF homologs. Structure 21**:** 665-671.

Jung, H. R., D. Pasini, K. Helin and O. N. Jensen, 2010 Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. Mol Cell Proteomics 9**:** 838-850.

Kaczynski, J., T. Cook and R. Urrutia, 2003 Sp1- and Kruppel-like transcription factors. Genome Biol 4**:** 206.

Kadoch, C., R. T. Williams, J. P. Calarco, E. L. Miller, C. M. Weber *et al.*, 2017 Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. Nat Genet 49**:** 213-222.

Kagey, M. H., J. J. Newman, S. Bilodeau, Y. Zhan, D. A. Orlando *et al.*, 2010 Mediator and cohesin connect gene expression and chromatin architecture. Nature 467**:** 430-435.

Kahn, T. G., E. Dorafshan, D. Schultheis, A. Zare, P. Stenberg *et al.*, 2016 Interdependence of PRC1 and PRC2 for recruitment to Polycomb Response Elements. Nucleic Acids Res.

Kahn, T. G., P. Stenberg, V. Pirrotta and Y. B. Schwartz, 2014 Combinatorial interactions are required for the efficient recruitment of pho repressive complex (PhoRC) to polycomb response elements. PLoS Genet 10**:** e1004495.

Kal, A. J., T. Mahmoudi, N. B. Zak and C. P. Verrijzer, 2000 The Drosophila brahma complex is an essential coactivator for the trithorax group protein zeste. Genes Dev 14**:** 1058-1071.

Kalb, R., S. Latwiel, H. I. Baymaz, P. W. Jansen, C. W. Muller *et al.*, 2014 Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. Nat Struct Mol Biol 21**:** 569-571.

Kaneko, S., R. Bonasio, R. Saldana-Meyer, T. Yoshida, J. Son *et al.*, 2014 Interactions between JARID2 and noncoding RNAs regulate PRC2 recruitment to chromatin. Mol Cell 53**:** 290-300.

Kaneko, S., J. Son, S. S. Shen, D. Reinberg and R. Bonasio, 2013 PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. Nat Struct Mol Biol 20**:** 1258-1264.

Kang, H., K. A. McElroy, Y. L. Jung, A. A. Alekseyenko, B. M. Zee *et al.*, 2015 Sex comb on midleg (Scm) is a functional link between PcG-repressive complexes in Drosophila. Genes Dev 29**:** 1136-1150.

Kao, C. F., C. Hillyer, T. Tsukuda, K. Henry, S. Berger *et al.*, 2004 Rad6 plays a role in transcriptional activation through ubiquitylation of histone H2B. Genes Dev 18**:** 184-195.

Kar, G., J. K. Kim, A. A. Kolodziejczyk, K. N. Natarajan, E. Torlai Triglia *et al.*, 2017 Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. Nat Commun 8**:** 36.

Kassis, J. A., 2002 Pairing-sensitive silencing, polycomb group response elements, and transposon homing in Drosophila. Adv Genet 46**:** 421-438.

Kassis, J. A., and J. L. Brown, 2013 Polycomb group response elements in Drosophila and vertebrates. Adv Genet 81**:** 83-118.

Kaufman, T. C., R. Lewis and B. Wakimoto, 1980 Cytogenetic Analysis of Chromosome 3 in DROSOPHILA MELANOGASTER: The Homoeotic Gene Complex in Polytene Chromosome Interval 84a-B. Genetics 94**:** 115-133.

Ketel, C. S., E. F. Andersen, M. L. Vargas, J. Suh, S. Strome *et al.*, 2005 Subunit contributions to histone methyltransferase activities of fly and worm polycomb group complexes. Mol Cell Biol 25**:** 6857-6868.

Kharchenko, P. V., A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle *et al.*, 2011 Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471**:** 480-485.

Kim, C. A., M. Gingery, R. M. Pilpa and J. U. Bowie, 2002 The SAM domain of polyhomeotic forms a helical polymer. Nat Struct Biol 9**:** 453-457.

Kim, C. A., M. R. Sawaya, D. Cascio, W. Kim and J. U. Bowie, 2005 Structural organization of a Sex-comb-on-midleg/polyhomeotic copolymer. J Biol Chem 280**:** 27769-27775.

King, I. F., N. J. Francis and R. E. Kingston, 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.

Klauke, K., V. Radulovic, M. Broekhuis, E. Weersing, E. Zwart *et al.*, 2013 Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. Nat Cell Biol 15**:** 353-362.

Kloet, S. L., M. M. Makowski, H. I. Baymaz, L. van Voorthuijsen, I. D. Karemaker *et al.*, 2016 The dynamic interactome and genomic targets of Polycomb complexes during stem-cell differentiation. Nat Struct Mol Biol 23**:** 682-690.

Klose, R. J., E. M. Kallin and Y. Zhang, 2006 JmjC-domain-containing proteins and histone demethylation. Nat Rev Genet 7**:** 715-727.

Klymenko, T., and J. Muller, 2004 The histone methyltransferases Trithorax and Ash1 prevent transcriptional silencing by Polycomb group proteins. EMBO Rep 5**:** 373-377.

Klymenko, T., B. Papp, W. Fischle, T. Kocher, M. Schelder *et al.*, 2006 A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. Genes Dev 20**:** 1110-1122.

Koch, L., 2014 Gene regulation: Yin and Yang of Polycomb/Trithorax response elements. Nat Rev Mol Cell Biol 15**:** 630.

Kolovos, P., T. A. Knoch, F. G. Grosveld, P. R. Cook and A. Papantonis, 2012 Enhancers and silencers: an integrated and simple model for their function. Epigenetics Chromatin 5**:** 1.

Kondo, T., S. Ito and H. Koseki, 2016 Polycomb in Transcriptional Phase Transition of Developmental Genes. Trends Biochem Sci 41**:** 9-19.

Ku, M., R. P. Koche, E. Rheinbay, E. M. Mendenhall, M. Endoh *et al.*, 2008 Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. PLoS Genet 4**:** e1000242.

Kuo, M. H., J. Zhou, P. Jambeck, M. E. Churchill and C. D. Allis, 1998 Histone acetyltransferase activity of yeast Gcn5p is required for the activation of target genes in vivo. Genes Dev 12**:** 627-639.

Kurzhals, R. L., F. Tie, C. A. Stratton and P. J. Harte, 2008 Drosophila ESC-like can substitute for ESC and becomes required for Polycomb silencing if ESC is absent. Dev Biol 313**:** 293-306.

Kuzmichev, A., K. Nishioka, H. Erdjument-Bromage, P. Tempst and D. Reinberg, 2002 Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. Genes Dev 16**:** 2893-2905.

Kyba, M., and H. W. Brock, 1998 The Drosophila polycomb group protein Psc contacts ph and Pc through specific conserved domains. Mol Cell Biol 18**:** 2712-2720.

Kyrchanova, O., V. Mogila, D. Wolle, G. Deshpande, A. Parshikov *et al.*, 2016 Functional Dissection of the Blocking and Bypass Activities of the Fab-8 Boundary in the Drosophila Bithorax Complex. PLoS Genet 12**:** e1006188.

Lagarou, A., A. Mohd-Sarip, Y. M. Moshkin, G. E. Chalkley, K. Bezstarosti *et al.*, 2008 dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. Genes Dev 22**:** 2799-2810.

Laible, G., A. Wolf, R. Dorn, G. Reuter, C. Nislow *et al.*, 1997 Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at S. cerevisiae telomeres. EMBO J 16**:** 3219-3232.

Lam, K. C., F. Muhlpfordt, J. M. Vaquerizas, S. J. Raja, H. Holz *et al.*, 2012 The NSL complex regulates housekeeping genes in Drosophila. PLoS Genet 8**:** e1002736.

Landeira, D., S. Sauer, R. Poot, M. Dvorkina, L. Mazzarella *et al.*, 2010 Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. Nat Cell Biol 12**:** 618-624.

Langst, G., and L. Manelyte, 2015 Chromatin Remodelers: From Function to Dysfunction. Genes (Basel) 6**:** 299-324.

Lanzuolo, C., V. Roure, J. Dekker, F. Bantignies and V. Orlando, 2007 Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat Cell Biol 9**:** 1167-1174.

Laprell, F., K. Finkl and J. Muller, 2017 Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. Science 356**:** 85-88.

Lau, M. S., M. G. Schwartz, S. Kundu, A. J. Savol, P. I. Wang *et al.*, 2017 Mutation of a nucleosome compaction region disrupts Polycomb-mediated axial patterning. Science 355**:** 1081-1084.

Lavrov, S., J. Dejardin and G. Cavalli, 2004 Combined immunostaining and FISH analysis of polytene chromosomes. Methods Mol Biol 247**:** 289-303.

Lecona, E., L. A. Rojas, R. Bonasio, A. Johnston, O. Fernandez-Capetillo *et al.*, 2013 Polycomb protein SCML2 regulates the cell cycle by binding and modulating CDK/CYCLIN/p21 complexes. PLoS Biol 11**:** e1001737.

Lee, H. G., T. G. Kahn, A. Simcox, Y. B. Schwartz and V. Pirrotta, 2015 Genome-wide activities of Polycomb complexes control pervasive transcription. Genome Res 25**:** 1170-1181.

Lee, J. S., K. M. Galvin and Y. Shi, 1993 Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. Proc Natl Acad Sci U S A 90**:** 6145-6149.

Lee, T. I., R. G. Jenner, L. A. Boyer, M. G. Guenther, S. S. Levine *et al.*, 2006 Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 125**:** 301-313.

Lee, W., S. Teckie, T. Wiesner, L. Ran, C. N. Prieto Granada *et al.*, 2014 PRC2 is recurrently inactivated through EED or SUZ12 loss in malignant peripheral nerve sheath tumors. Nat Genet 46**:** 1227-1232.

Lehmann, M., T. Siegmund, K. G. Lintermann and G. Korge, 1998 The pipsqueak protein of Drosophila melanogaster binds to GAGA sequences through a novel DNA-binding domain. J Biol Chem 273**:** 28504-28509.

Lewis, E. B., 1978 A gene complex controlling segmentation in Drosophila. Nature 276: 565-570.

Lewis, P. W., M. M. Muller, M. S. Koletsky, F. Cordero, S. Lin *et al.*, 2013 Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science 340: 857-861.

Lewis, R. A., T. C. Kaufman, R. E. Denell and P. Tallerico, 1980 Genetic Analysis of the Antennapedia Gene Complex (Ant-C) and Adjacent Chromosomal Regions of DROSOPHILA MELANOGASTER. I. Polytene Chromosome Segments 84b-D. Genetics 95: 367-381.

Li, G., R. Margueron, M. Ku, P. Chambon, B. E. Bernstein *et al.*, 2010a Jarid2 and PRC2, partners in regulating gene expression. Genes Dev 24: 368-380.

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

Li, X. S., P. Trojer, T. Matsumura, J. E. Treisman and N. Tanese, 2010b Mammalian SWI/SNF--a subunit BAF250/ARID1 is an E3 ubiquitin ligase that targets histone H2B. Mol Cell Biol 30: 1673-1688.

Li, Z., R. Cao, M. Wang, M. P. Myers, Y. Zhang *et al.*, 2006 Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. J Biol Chem 281: 20643-20649.

Liebich, I., J. Bode, M. Frisch and E. Wingender, 2002 S/MARt DB: a database on scaffold/matrix attached regions. Nucleic Acids Res 30: 372-374.

Lo, S. M., N. K. Ahuja and N. J. Francis, 2009 Polycomb group protein Suppressor 2 of zeste is a functional homolog of Posterior Sex Combs. Mol Cell Biol 29: 515-525.

Lo, S. M., and N. J. Francis, 2010 Inhibition of chromatin remodeling by polycomb group protein posterior sex combs is mechanistically distinct from nucleosome binding. Biochemistry 49: 9438-9448.

Lo, S. M., K. A. McElroy and N. J. Francis, 2012 Chromatin modification by PSC occurs at one PSC per nucleosome and does not require the acidic patch of histone H2A. PLoS One 7: e47162.

Lonie, A., R. D'Andrea, R. Paro and R. Saint, 1994 Molecular characterisation of the Polycomblike gene of Drosophila melanogaster, a trans-acting negative regulator of homeotic gene expression. Development 120: 2629-2636.

Loubiere, V., A. Delest, A. Thomas, B. Bonev, B. Schuettengruber *et al.*, 2016 Coordinate redeployment of PRC1 proteins suppresses tumor formation during Drosophila development. Nat Genet 48: 1436-1442.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent and T. J. Richmond, 1997 Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature 389: 251-260.

Lynch, M. D., A. J. Smith, M. De Gobbi, M. Flenley, J. R. Hughes *et al.*, 2012 An interspecies analysis reveals a key role for unmethylated CpG dinucleotides in vertebrate Polycomb complex recruitment. EMBO J 31: 317-329.

Macdonald, N., J. P. Welburn, M. E. Noble, A. Nguyen, M. B. Yaffe *et al.*, 2005 Molecular basis for the recognition of phosphorylated and phosphoacetylated histone h3 by 14-3-3. Mol Cell 20: 199-211.

Mahmoudi, T., L. M. Zuijderduijn, A. Mohd-Sarip and C. P. Verrijzer, 2003 GAGA facilitates binding of Pleiohomeotic to a chromatinized Polycomb response element. Nucleic Acids Res 31**:** 4147-4156.

Mallo, M., and C. R. Alonso, 2013 The regulation of Hox gene expression during animal development. Development 140**:** 3951-3963.

Margueron, R., N. Justin, K. Ohno, M. L. Sharpe, J. Son *et al.*, 2009 Role of the polycomb protein EED in the propagation of repressive histone marks. Nature 461**:** 762-767.

Margueron, R., G. Li, K. Sarma, A. Blais, J. Zavadil *et al.*, 2008 Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. Mol Cell 32**:** 503-518.

Mavrich, T. N., C. Jiang, I. P. Ioshikhes, X. Li, B. J. Venters *et al.*, 2008 Nucleosome organization in the Drosophila genome. Nature 453**:** 358-362.

McConnell, B. B., and V. W. Yang, 2010 Mammalian Kruppel-like factors in health and diseases. Physiol Rev 90**:** 1337-1381.

Melnikova, L., F. Juge, N. Gruzdeva, A. Mazur, G. Cavalli *et al.*, 2004 Interaction between the GAGA factor and Mod(mdg4) proteins promotes insulator bypass in Drosophila. Proc Natl Acad Sci U S A 101**:** 14806-14811.

Mendenhall, E. M., R. P. Koche, T. Truong, V. W. Zhou, B. Issac *et al.*, 2010 GC-rich sequence elements recruit PRC2 in mammalian ES cells. PLoS Genet 6**:** e1001244.

Milne, T. A., S. D. Briggs, H. W. Brock, M. E. Martin, D. Gibbs *et al.*, 2002 MLL targets SET domain methyltransferase activity to Hox gene promoters. Mol Cell 10**:** 1107-1117.

Mishra, R. K., J. Mihaly, S. Barges, A. Spierer, F. Karch *et al.*, 2001 The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. Mol Cell Biol 21**:** 1311-1318.

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

Mohan, M., H. M. Herz, E. R. Smith, Y. Zhang, J. Jackson *et al.*, 2011 The COMPASS family of H3K4 methylases in Drosophila. Mol Cell Biol 31**:** 4310-4318.

Mohd-Sarip, A., J. A. van der Knaap, C. Wyman, R. Kanaar, P. Schedl *et al.*, 2006 Architecture of a polycomb nucleoprotein complex. Mol Cell 24**:** 91-100.

Morera, L., M. Lubbert and M. Jung, 2016 Targeting histone methyltransferases and demethylases in clinical trials for cancer therapy. Clin Epigenetics 8**:** 57.

Morey, L., L. Aloia, L. Cozzuto, S. A. Benitah and L. Di Croce, 2013 RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. Cell Rep 3**:** 60-69.

Morey, L., G. Pascual, L. Cozzuto, G. Roma, A. Wutz *et al.*, 2012 Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. Cell Stem Cell 10**:** 47-62.

Morey, L., A. Santanach, E. Blanco, L. Aloia, E. P. Nora *et al.*, 2015 Polycomb Regulates Mesoderm Cell Fate-Specification in Embryonic Stem Cells through Activation and Repression Mechanisms. Cell Stem Cell 17**:** 300-315.

Morin, R. D., N. A. Johnson, T. M. Severson, A. J. Mungall, J. An *et al.*, 2010 Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. Nat Genet 42**:** 181-185.

Mosrin-Huaman, C., L. Canaple, D. Locker and M. Decoville, 1998 DSP1 gene of Drosophila melanogaster encodes an HMG-domain protein that plays multiple roles in development. Dev Genet 23**:** 324-334.

Mousavi, K., H. Zare, A. H. Wang and V. Sartorelli, 2012 Polycomb protein Ezh1 promotes RNA polymerase II elongation. Mol Cell 45**:** 255-262.

Muller, J., 1995 Transcriptional silencing by the Polycomb protein in Drosophila embryos. EMBO J 14**:** 1209-1220.

Muller, J., C. M. Hart, N. J. Francis, M. L. Vargas, A. Sengupta *et al.*, 2002 Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. Cell 111**:** 197-208.

Musselman, C. A., N. Avvakumov, R. Watanabe, C. G. Abraham, M. E. Lalonde *et al.*, 2012a Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nat Struct Mol Biol 19**:** 1266-1272.

Musselman, C. A., M. E. Lalonde, J. Cote and T. G. Kutateladze, 2012b Perceiving the epigenetic landscape through histone readers. Nat Struct Mol Biol 19**:** 1218-1227.

Nakamura, T., T. Mori, S. Tada, W. Krajewski, T. Rozovskaia *et al.*, 2002 ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation. Mol Cell 10**:** 1119-1128.

Nanyes, D. R., S. E. Junco, A. B. Taylor, A. K. Robinson, N. L. Patterson *et al.*, 2014 Multiple polymer architectures of human polyhomeotic homolog 3 sterile alpha motif. Proteins 82**:** 2823-2830.

Negre, N., J. Hennetin, L. V. Sun, S. Lavrov, M. Bellis *et al.*, 2006 Chromosomal distribution of PcG proteins during Drosophila development. PLoS Biol. 4**:** e170. Epub 2006 Apr 2020.

Nekrasov, M., T. Klymenko, S. Fraterman, B. Papp, K. Oktaba *et al.*, 2007 Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. EMBO J 26**:** 4078-4088.

Nekrasov, M., B. Wild and J. Muller, 2005 Nucleosome binding and histone methyltransferase activity of Drosophila PRC2. EMBO Rep 6**:** 348-353.

Ng, R. K., and J. B. Gurdon, 2008 Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. Nat Cell Biol 10**:** 102-109.

Nguyen, N., X. Zhang, N. Olashaw and E. Seto, 2004 Molecular cloning and functional characterization of the transcription factor YY2. J Biol Chem 279**:** 25927-25934.

Noordermeer, D., M. Leleu, P. Schorderet, E. Joye, F. Chabaud *et al.*, 2014 Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. Elife 3**:** e02557.

Noordermeer, D., M. Leleu, E. Splinter, J. Rougemont, W. De Laat *et al.*, 2011 The dynamic architecture of Hox gene clusters. Science 334**:** 222-225.

O'Connell, S., L. Wang, S. Robert, C. A. Jones, R. Saint *et al.*, 2001 Polycomblike PHD fingers mediate conserved interaction with enhancer of zeste protein. J Biol Chem 276**:** 43065-43073.

Ohno, K., D. McCabe, B. Czermin, A. Imhof and V. Pirrotta, 2008 ESC, ESCL and their roles in Polycomb Group mechanisms. Mech Dev 125**:** 527-541.

Oktaba, K., L. Gutierrez, J. Gagneur, C. Girardot, A. K. Sengupta *et al.*, 2008 Dynamic regulation by polycomb group protein complexes controls pattern formation and the cell cycle in Drosophila. Dev Cell 15**:** 877-889.

Olins, D. E., and A. L. Olins, 2003 Chromatin history: our view from the bridge. Nat Rev Mol Cell Biol 4**:** 809-814.

Oliviero, G., G. L. Brien, A. Watson, G. Streubel, E. Jerman *et al.*, 2016 Dynamic protein interactions of the Polycomb Repressive Complex 2 during differentiation of pluripotent cells. Mol Cell Proteomics.

Oudet, P., M. Gross-Bellard and P. Chambon, 1975 Electron microscopic and biochemical evidence that chromatin structure is a repeating unit. Cell 4**:** 281-300.

Papamichos-Chronakis, M., S. Watanabe, O. J. Rando and C. L. Peterson, 2011 Global regulation of H2A.Z localization by the INO80 chromatin-remodeling enzyme is essential for genome integrity. Cell 144**:** 200-213.

Papp, B., and J. Muller, 2006 Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. Genes Dev. 20**:** 2041-2054.

Pasini, D., P. A. Cloos, J. Walfridsson, L. Olsson, J. P. Bukowski *et al.*, 2010 JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature 464**:** 306-310.

Pedersen, M. T., and K. Helin, 2010 Histone demethylases in development and disease. Trends Cell Biol 20**:** 662-671.

Peng, J. C., A. Valouev, T. Swigut, J. Zhang, Y. Zhao *et al.*, 2009 Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. Cell 139**:** 1290-1302.

Pengelly, A. R., R. Kalb, K. Finkl and J. Muller, 2015 Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. Genes Dev 29**:** 1487-1492.

Peters, A. H., S. Kubicek, K. Mechtler, R. J. O'Sullivan, A. A. Derijck *et al.*, 2003 Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell 12**:** 1577-1589.

Peterson, A. J., M. Kyba, D. Bornemann, K. Morgan, H. W. Brock *et al.*, 1997 A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. Mol Cell Biol 17**:** 6683-6692.

Petrascheck, M., D. Escher, T. Mahmoudi, C. P. Verrijzer, W. Schaffner *et al.*, 2005 DNA looping induced by a transcriptional enhancer in vivo. Nucleic Acids Res 33**:** 3743-3750.

Petruk, S., Y. Sedkov, S. Smith, S. Tillib, V. Kraevski *et al.*, 2001 Trithorax and dCBP acting in a complex to maintain expression of a homeotic gene. Science 294**:** 1331-1334.

Phillips-Cremins, J. E., and V. G. Corces, 2013 Chromatin insulators: linking genome organization to cellular function. Mol Cell 50**:** 461-474.

Pickart, C. M., and M. J. Eddins, 2004 Ubiquitin: structures, functions, mechanisms. Biochim Biophys Acta 1695**:** 55-72.

Piunti, A., R. Hashizume, M. A. Morgan, E. T. Bartom, C. M. Horbinski *et al.*, 2017 Therapeutic targeting of polycomb and BET bromodomain proteins in diffuse intrinsic pontine gliomas. Nat Med 23**:** 493-500.

Qin, J., W. A. Whyte, E. Anderssen, E. Apostolou, H. H. Chen *et al.*, 2012 The polycomb group protein L3mbtl2 assembles an atypical PRC1-family complex that is essential in pluripotent stem cells and early development. Cell Stem Cell 11**:** 319-332.

Raisner, R. M., P. D. Hartley, M. D. Meneghini, M. Z. Bao, C. L. Liu *et al.*, 2005 Histone variant H2A.Z marks the 5' ends of both active and inactive genes in euchromatin. Cell 123**:** 233-248.

Ray, P., S. De, A. Mitra, K. Bezstarosti, J. A. Demmers *et al.*, 2016 Combgap contributes to recruitment of Polycomb group proteins in Drosophila. Proc Natl Acad Sci U S A 113**:** 3826-3831.

Razin, S. V., 2001 The nuclear matrix and chromosomal DNA loops: is their any correlation between partitioning of the genome into loops and functional domains? Cell Mol Biol Lett 6**:** 59-69.

Rea, S., F. Eisenhaber, D. O'Carroll, B. D. Strahl, Z. W. Sun *et al.*, 2000 Regulation of chromatin structure by site-specific histone H3 methyltransferases. Nature 406**:** 593-599.

Richly, H., L. Rocha-Viegas, J. D. Ribeiro, S. Demajo, G. Gundem *et al.*, 2010 Transcriptional activation of polycomb-repressed genes by ZRF1. Nature 468**:** 1124-1128.

Rickels, R., D. Hu, C. K. Collings, A. R. Woodfin, A. Piunti *et al.*, 2016 An Evolutionary Conserved Epigenetic Mark of Polycomb Response Elements Implemented by Trx/MLL/COMPASS. Mol Cell 63**:** 318-328.

Riising, E. M., I. Comet, B. Leblanc, X. Wu, J. V. Johansen *et al.*, 2014 Gene silencing triggers polycomb repressive complex 2 recruitment to CpG islands genome wide. Mol Cell 55**:** 347-360.

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

Robinson, A. K., B. Z. Leal, L. V. Chadwell, R. Wang, U. Ilangovan *et al.*, 2012 The growth-suppressive function of the polycomb group protein polyhomeotic is mediated by polymerization of its sterile alpha motif (SAM) domain. J Biol Chem 287**:** 8702-8713.

Robinson, P. J., L. Fairall, V. A. Huynh and D. Rhodes, 2006 EM measurements define the dimensions of the "30-nm" chromatin fiber: evidence for a compact, interdigitated structure. Proc Natl Acad Sci U S A 103**:** 6506-6511.

Salvaing, J., M. Decoville, E. Mouchel-Vielh, M. Bussiere, A. Daulny *et al.*, 2006 Corto and DSP1 interact and bind to a maintenance element of the Scr Hox gene: understanding the role of Enhancers of trithorax and Polycomb. BMC Biol 4**:** 9.

Salvaing, J., A. Lopez, A. Boivin, J. S. Deutsch and F. Peronnet, 2003 The Drosophila Corto protein interacts with Polycomb-group proteins and the GAGA factor. Nucleic Acids Res 31**:** 2873-2882.

Sanborn, A. L., S. S. Rao, S. C. Huang, N. C. Durand, M. H. Huntley *et al.*, 2015 Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. Proc Natl Acad Sci U S A 112**:** E6456-6465.

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

Santaguida, S., and A. Musacchio, 2009 The life and miracles of kinetochores. EMBO J 28: 2511-2531.

Sanulli, S., N. Justin, A. Teissandier, K. Ancelin, M. Portoso *et al.*, 2015 Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. Mol Cell 57: 769-783.

Sarcinella, E., P. C. Zuzarte, P. N. Lau, R. Draker and P. Cheung, 2007 Monoubiquitylation of H2A.Z distinguishes its association with euchromatin or facultative heterochromatin. Mol Cell Biol 27: 6457-6468.

Sarma, K., R. Margueron, A. Ivanov, V. Pirrotta and D. Reinberg, 2008 Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. Mol Cell Biol 28: 2718-2731.

Satijn, D. P., M. J. Gunster, J. van der Vlag, K. M. Hamer, W. Schul *et al.*, 1997 RING1 is associated with the polycomb group protein complex and acts as a transcriptional repressor. Mol Cell Biol 17: 4105-4113.

Satijn, D. P., K. M. Hamer, J. den Blaauwen and A. P. Otte, 2001 The polycomb group protein EED interacts with YY1, and both proteins induce neural tissue in Xenopus embryos. Mol Cell Biol 21: 1360-1369.

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

Savla, U., J. Benes, J. Zhang and R. S. Jones, 2008 Recruitment of Drosophila Polycomb-group proteins by Polycomblike, a component of a novel protein complex in larvae. Development 135: 813-817.

Schaaf, C. A., Z. Misulovin, M. Gause, A. Koenig, D. W. Gohara *et al.*, 2013 Cohesin and polycomb proteins functionally interact to control transcription at silenced and active genes. PLoS Genet 9: e1003560.

Scheuermann, J. C., A. G. de Ayala Alonso, K. Oktaba, N. Ly-Hartig, R. K. McGinty *et al.*, 2010 Histone H2A deubiquitinase activity of the Polycomb repressive complex PR-DUB. Nature 465: 243-247.

Scheuermann, J. C., L. Gutierrez and J. Muller, 2012 Histone H2A monoubiquitination and Polycomb repression: the missing pieces of the puzzle. Fly (Austin) 6: 162-168.

Schoenfelder, S., R. Sugar, A. Dimond, B. M. Javierre, H. Armstrong *et al.*, 2015 Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. Nat Genet 47: 1179-1186.

Schuettengruber, B., and G. Cavalli, 2009 Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development 136: 3531-3542.

Schuettengruber, B., and G. Cavalli, 2010 The DUBle life of polycomb complexes. Dev Cell 18: 878-880.

Schuettengruber, B., and G. Cavalli, 2013 Polycomb domain formation depends on short and long distance regulatory cues. PLoS One 8: e56531.

Schuettengruber, B., M. Ganapathi, B. Leblanc, M. Portoso, R. Jaschek *et al.*, 2009 Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol 7: e13.

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

Schuettengruber, B., N. Oded Elkayam, T. Sexton, M. Entrevan, S. Stern *et al.*, 2014 Cooperativity, specificity, and evolutionary stability of Polycomb targeting in Drosophila. Cell Rep 9**:** 219-233.

Schwartz, Y. B., T. G. Kahn, D. A. Nix, X. Y. Li, R. Bourgon *et al.*, 2006 Genome-wide analysis of Polycomb targets in Drosophila melanogaster. Nat Genet 38**:** 700-705.

Schwartz, Y. B., T. G. Kahn, P. Stenberg, K. Ohno, R. Bourgon *et al.*, 2010 Alternative epigenetic chromatin states of polycomb target genes. PLoS Genet 6**:** e1000805.

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

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

Schwendemann, A., and M. Lehmann, 2002 Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. Proc Natl Acad Sci U S A 99**:** 12883-12888.

Seto, E., B. Lewis and T. Shenk, 1993 Interaction between transcription factors Sp1 and YY1. Nature 365**:** 462-464.

Seto, E., and M. Yoshida, 2014 Erasers of histone acetylation: the histone deacetylase enzymes. Cold Spring Harb Perspect Biol 6**:** a018713.

Sexton, T., F. Bantignies and G. Cavalli, 2009 Genomic interactions: chromatin loops and gene meeting points in transcriptional regulation. Semin Cell Dev Biol 20**:** 849-855.

Sexton, T., E. Yaffe, E. Kenigsberg, F. Bantignies, B. Leblanc *et al.*, 2012 Three-dimensional folding and functional organization principles of the Drosophila genome. Cell 148**:** 458-472.

Shao, Z., F. Raible, R. Mollaaghababa, J. R. Guyon, C. T. Wu *et al.*, 1999 Stabilization of chromatin structure by PRC1, a Polycomb complex. Cell 98**:** 37-46.

Shen, X., W. Kim, Y. Fujiwara, M. D. Simon, Y. Liu *et al.*, 2009 Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. Cell 139**:** 1303-1314.

Shen, X., Y. Liu, Y. J. Hsu, Y. Fujiwara, J. Kim *et al.*, 2008 EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. Mol Cell 32**:** 491-502.

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

Siegel, V., T. A. Jongens, L. Y. Jan and Y. N. Jan, 1993 pipsqueak, an early acting member of the posterior group of genes, affects vasa level and germ cell-somatic cell interaction in the developing egg chamber. Development 119**:** 1187-1202.

Siegmund, T., and M. Lehmann, 2002 The Drosophila Pipsqueak protein defines a new family of helix-turn-helix DNA-binding proteins. Dev Genes Evol 212**:** 152-157.

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

Sing, A., D. Pannell, A. Karaiskakis, K. Sturgeon, M. Djabali *et al.*, 2009 A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. Cell 138**:** 885-897.

Sneeringer, C. J., M. P. Scott, K. W. Kuntz, S. K. Knutson, R. M. Pollock *et al.*, 2010 Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 (H3K27) in human B-cell lymphomas. Proc Natl Acad Sci U S A 107**:** 20980-20985.

Son, J., S. S. Shen, R. Margueron and D. Reinberg, 2013 Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. Genes Dev 27**:** 2663-2677.

Sowpati, D. T., S. Ramamoorthy and R. K. Mishra, 2015 Expansion of the polycomb system and evolution of complexity. Mech Dev 138 Pt 2**:** 97-112.

Sparmann, A., and M. van Lohuizen, 2006 Polycomb silencers control cell fate, development and cancer. Nat Rev Cancer 6**:** 846-856.

Srivastava, S., D. Puri, H. S. Garapati, J. Dhawan and R. K. Mishra, 2013 Vertebrate GAGA factor associated insulator elements demarcate homeotic genes in the HOX clusters. Epigenetics Chromatin 6**:** 8.

Stanton, B. Z., C. Hodges, J. P. Calarco, S. M. Braun, W. L. Ku *et al.*, 2017 Smarca4 ATPase mutations disrupt direct eviction of PRC1 from chromatin. Nat Genet 49**:** 282-288.

Stock, J. K., S. Giadrossi, M. Casanova, E. Brookes, M. Vidal *et al.*, 2007 Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. Nat Cell Biol 9**:** 1428-1435.

Swaminathan, V., A. H. Kishore, K. K. Febitha and T. K. Kundu, 2005 Human histone chaperone nucleophosmin enhances acetylation-dependent chromatin transcription. Mol Cell Biol 25**:** 7534-7545.

Szenker, E., N. Lacoste and G. Almouzni, 2012 A developmental requirement for HIRA-dependent H3.3 deposition revealed at gastrulation in Xenopus. Cell Rep 1**:** 730-740.

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

Tavares, L., E. Dimitrova, D. Oxley, J. Webster, R. Poot *et al.*, 2012 RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. Cell 148**:** 664-678.

Thompson, P. R., D. Wang, L. Wang, M. Fulco, N. Pediconi *et al.*, 2004 Regulation of the p300 HAT domain via a novel activation loop. Nat Struct Mol Biol 11**:** 308-315.

Tie, F., R. Banerjee, C. Fu, C. A. Stratton, M. Fang *et al.*, 2016 Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. Proc Natl Acad Sci U S A 113**:** E744-753.

Tie, F., R. Banerjee, C. A. Stratton, J. Prasad-Sinha, V. Stepanik *et al.*, 2009 CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development 136**:** 3131-3141.

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

Tollervey, J. R., and V. V. Lunyak, 2012 Epigenetics: judge, jury and executioner of stem cell fate. Epigenetics 7**:** 823-840.

Trimarchi, J. M., B. Fairchild, J. Wen and J. A. Lees, 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.

Trojer, P., A. R. Cao, Z. Gao, Y. Li, J. Zhang *et al.*, 2011 L3MBTL2 protein acts in concert with PcG protein-mediated monoubiquitination of H2A to establish a repressive chromatin structure. Mol Cell 42**:** 438-450.

Tsukada, Y., J. Fang, H. Erdjument-Bromage, M. E. Warren, C. H. Borchers *et al.*, 2006 Histone demethylation by a family of JmjC domain-containing proteins. Nature 439**:** 811-816.

Tuckfield, A., D. R. Clouston, T. M. Wilanowski, L. L. Zhao, J. M. Cunningham *et al.*, 2002 Binding of the RING polycomb proteins to specific target genes in complex with the grainyhead-like family of developmental transcription factors. Mol Cell Biol 22**:** 1936-1946.

Vakoc, C. R., S. A. Mandat, B. A. Olenchock and G. A. Blobel, 2005 Histone H3 lysine 9 methylation and HP1gamma are associated with transcription elongation through mammalian chromatin. Mol Cell 19**:** 381-391.

van Attikum, H., and S. M. Gasser, 2009 Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol 19**:** 207-217.

Van Bortle, K., M. H. Nichols, L. Li, C. T. Ong, N. Takenaka *et al.*, 2014 Insulator function and topological domain border strength scale with architectural protein occupancy. Genome Biol 15**:** R82.

van den Boom, V., H. Maat, M. Geugien, A. Rodriguez Lopez, A. M. Sotoca *et al.*, 2016 Non-canonical PRC1.1 Targets Active Genes Independent of H3K27me3 and Is Essential for Leukemogenesis. Cell Rep 14**:** 332-346.

Varambally, S., S. M. Dhanasekaran, M. Zhou, T. R. Barrette, C. Kumar-Sinha *et al.*, 2002 The polycomb group protein EZH2 is involved in progression of prostate cancer. Nature 419**:** 624-629.

Vasanthi, D., A. Nagabhushan, N. K. Matharu and R. K. Mishra, 2013 A functionally conserved Polycomb response element from mouse HoxD complex responds to heterochromatin factors. Sci Rep 3**:** 3011.

Venken, K. J., and H. J. Bellen, 2007 Transgenesis upgrades for Drosophila melanogaster. Development 134**:** 3571-3584.

Vidal, M., and R. F. Gaber, 1991 RPD3 encodes a second factor required to achieve maximum positive and negative transcriptional states in Saccharomyces cerevisiae. Mol Cell Biol 11**:** 6317-6327.

Vietri Rudan, M., C. Barrington, S. Henderson, C. Ernst, D. T. Odom *et al.*, 2015 Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. Cell Rep 10**:** 1297-1309.

Vieux-Rochas, M., P. J. Fabre, M. Leleu, D. Duboule and D. Noordermeer, 2015 Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. Proc Natl Acad Sci U S A 112**:** 4672-4677.

Voigt, P., W. W. Tee and D. Reinberg, 2013 A double take on bivalent promoters. Genes Dev 27: 1318-1338.

Volkel, P., P. Le Faou, J. Vandamme, D. Pira and P. O. Angrand, 2012 A human Polycomb isoform lacking the Pc box does not participate to PRC1 complexes but forms protein assemblies and represses transcription. Epigenetics 7: 482-491.

Wang, C. C., J. J. Chen and P. C. Yang, 2006a Multifunctional transcription factor YY1: a therapeutic target in human cancer? Expert Opin Ther Targets 10: 253-266.

Wang, H., L. Wang, H. Erdjument-Bromage, M. Vidal, P. Tempst *et al.*, 2004a Role of histone H2A ubiquitination in Polycomb silencing. Nature 431: 873-878.

Wang, L., J. L. Brown, R. Cao, Y. Zhang, J. A. Kassis *et al.*, 2004b Hierarchical recruitment of polycomb group silencing complexes. Mol Cell 14: 637-646.

Wang, L., N. Jahren, M. L. Vargas, E. F. Andersen, J. Benes *et al.*, 2006b Alternative ESC and ESC-like subunits of a polycomb group histone methyltransferase complex are differentially deployed during Drosophila development. Mol Cell Biol 26: 2637-2647.

Wang, R., U. Ilangovan, A. K. Robinson, V. Schirf, P. M. Schwarz *et al.*, 2008a Structural transitions of the RING1B C-terminal region upon binding the polycomb cbox domain. Biochemistry 47: 8007-8015.

Wang, R., A. B. Taylor, B. Z. Leal, L. V. Chadwell, U. Ilangovan *et al.*, 2010 Polycomb group targeting through different binding partners of RING1B C-terminal domain. Structure 18: 966-975.

Wang, Z., C. Zang, J. A. Rosenfeld, D. E. Schones, A. Barski *et al.*, 2008b Combinatorial patterns of histone acetylations and methylations in the human genome. Nat Genet 40: 897-903.

Wani, A. H., A. N. Boettiger, P. Schorderet, A. Ergun, C. Munger *et al.*, 2016 Chromatin topology is coupled to Polycomb group protein subnuclear organization. Nat Commun 7: 10291.

Wei, Y., L. Yu, J. Bowen, M. A. Gorovsky and C. D. Allis, 1999 Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. Cell 97: 99-109.

Woo, C. J., P. V. Kharchenko, L. Daheron, P. J. Park and R. E. Kingston, 2010 A region of the human HOXD cluster that confers polycomb-group responsiveness. Cell 140: 99-110.

Woodcock, C. L., L. L. Frado and J. B. Rattner, 1984 The higher-order structure of chromatin: evidence for a helical ribbon arrangement. J Cell Biol 99: 42-52.

Wu, M., P. F. Wang, J. S. Lee, S. Martin-Brown, L. Florens *et al.*, 2008 Molecular regulation of H3K4 trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol 28: 7337-7344.

Wu, X., J. V. Johansen and K. Helin, 2013 Fbxl10/Kdm2b recruits polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. Mol Cell 49: 1134-1146.

Xiao, T., C. F. Kao, N. J. Krogan, Z. W. Sun, J. F. Greenblatt *et al.*, 2005 Histone H2B ubiquitylation is associated with elongating RNA polymerase II. Mol Cell Biol 25: 637-651.

Yang, X. J., and E. Seto, 2008 The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. Nat Rev Mol Cell Biol 9: 206-218.

Ye, T., A. R. Krebs, M. A. Choukrallah, C. Keime, F. Plewniak *et al.*, 2011 seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res 39**:** e35.

Yuan, W., M. Xu, C. Huang, N. Liu, S. Chen *et al.*, 2011 H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem 286**:** 7983-7989.

Yung, P. Y., A. Stuetzer, W. Fischle, A. M. Martinez and G. Cavalli, 2015 Histone H3 Serine 28 Is Essential for Efficient Polycomb-Mediated Gene Repression in Drosophila. Cell Rep 11**:** 1437-1445.

Zdzieblo, D., X. Li, Q. Lin, M. Zenke, D. J. Illich *et al.*, 2014 Pcgf6, a polycomb group protein, regulates mesodermal lineage differentiation in murine ESCs and functions in iPS reprogramming. Stem Cells 32**:** 3112-3125.

Zhang, Z., A. Jones, C. W. Sun, C. Li, C. W. Chang *et al.*, 2011 PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. Stem Cells 29**:** 229-240.

Zhou, B., A. Bagri and S. K. Beckendorf, 2001 Salivary gland determination in Drosophila: a salivary-specific, fork head enhancer integrates spatial pattern and allows fork head autoregulation. Dev Biol 237**:** 54-67.

Zhou, W., P. Zhu, J. Wang, G. Pascual, K. A. Ohgi *et al.*, 2008 Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. Mol Cell 29**:** 69-80.

Zilberman, D., D. Coleman-Derr, T. Ballinger and S. Henikoff, 2008 Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. Nature 456**:** 125-129.

ANNEXES

Trends in Cell Biology

Special Issue: 25 Years of Trends in Cell Biology

Review

Regulation of Genome Architecture and Function by Polycomb Proteins

Marianne Entrevan,[1] Bernd Schuettengruber,[1,*] and Giacomo Cavalli[1,*]

Polycomb group (PcG) proteins dynamically define cellular identities through the epigenetic repression of key developmental regulatory genes. PcG proteins are recruited to specific regulatory elements to modify the chromatin surrounding them. In addition, they regulate the organization of their target genes in the 3D space of the nucleus, and this regulatory function of the 3D genome architecture is involved in cell differentiation and the maintenance of cellular memory. In this review we discuss recent advances in our understanding of how PcG proteins are recruited to chromatin to induce local and global changes in chromosome conformation and regulate their target genes.

PcG Proteins in Cell Identity, Epigenetic Gene Regulation, and Chromatin Architecture

The capacity to generate different cell identities from an identical genome sequence, such as that of the single-cell zygote, relies on the establishment of cell type-specific expression programs that are maintained during development, even in the absence of the original initiating transcription factors (TFs), by so-called **epigenetic** (see Glossary) mechanisms [1]. Various modes of epigenetic gene regulation are used to fix transcriptional programs in time and space. These include DNA methylation, post-translational modifications of histone tails, noncoding RNAs (ncRNAs), and mitotic bookmarking factors. Another layer of complexity is provided by the fact that eukaryotic genomes are tightly folded and packaged into chromosomes to fit within the cell nucleus. The development of **chromosome conformation capture (3C)**-based experimental assays like 4C, 5C, ChIA-PET, and Hi-C have revealed that metazoan genomes fold hierarchically into **topologically associating domains (TADs)**, which appear to be a common pattern in higher-order chromosomal folding and contribute to the regulation of nuclear architecture in many species [2]. Changes in nuclear organization are an important component among epigenetic mechanisms contributing to robust and stable gene silencing. However, how these multiple layers interconnect mechanistically to reinforce each other's activity remains unclear. The evolutionarily conserved **PcG** proteins are one of the most prominent epigenetic silencing systems, playing a central role in cell differentiation and the maintenance of cell identity (reviewed in [3]). PcG proteins are chromatin-associated factors that locally modify chromatin through their histone-modifying activities to regulate their target genes. Furthermore, they participate in chromatin looping and long-range interactions between TADs marked by trimethylation of histone H3 lysine 27 (H3K27me3), thereby regulating global genome architecture (Figure 1, Key Figure). In this review we discuss recent advances in our understanding of how PcG proteins are recruited to their target genes and how they orchestrate genome function.

Trends

Nuclear architecture is not only important for the efficient compaction and decompaction of the genome during cell division, but has important functions in coordinating gene regulatory networks and orchestrating cellular identity.

Changes in nuclear organization are considered an important complement to epigenetic mechanisms contributing to robust and stable gene silencing.

Recruitment of Polycomb group (PcG) proteins to their target sites not only modulates local chromatin structure but also mediates looping interactions between regulatory elements and shapes global nuclear architecture, thereby regulating gene expression at multiple scales.

The evolutionarily conserved PcG proteins regulate cell identity and cell differentiation by orchestrating 3D genome architecture.

[1]Institute of Human Genetics, CNRS UPR1142 and University of Montpellier, 141 Rue de la Cardonille, 34396, Montpellier Cedex 5, France

*Correspondence:
Bernd.Schuttengruber@igh.cnrs.fr (B. Schuettengruber) and
Giacomo.Cavalli@igh.cnrs.fr (G. Cavalli).

© 2016 Elsevier Ltd. All rights reserved.

Key Figure

Polycomb Group (PcG) Proteins Are Implicated in Multiple Layers of Chromatin Organization

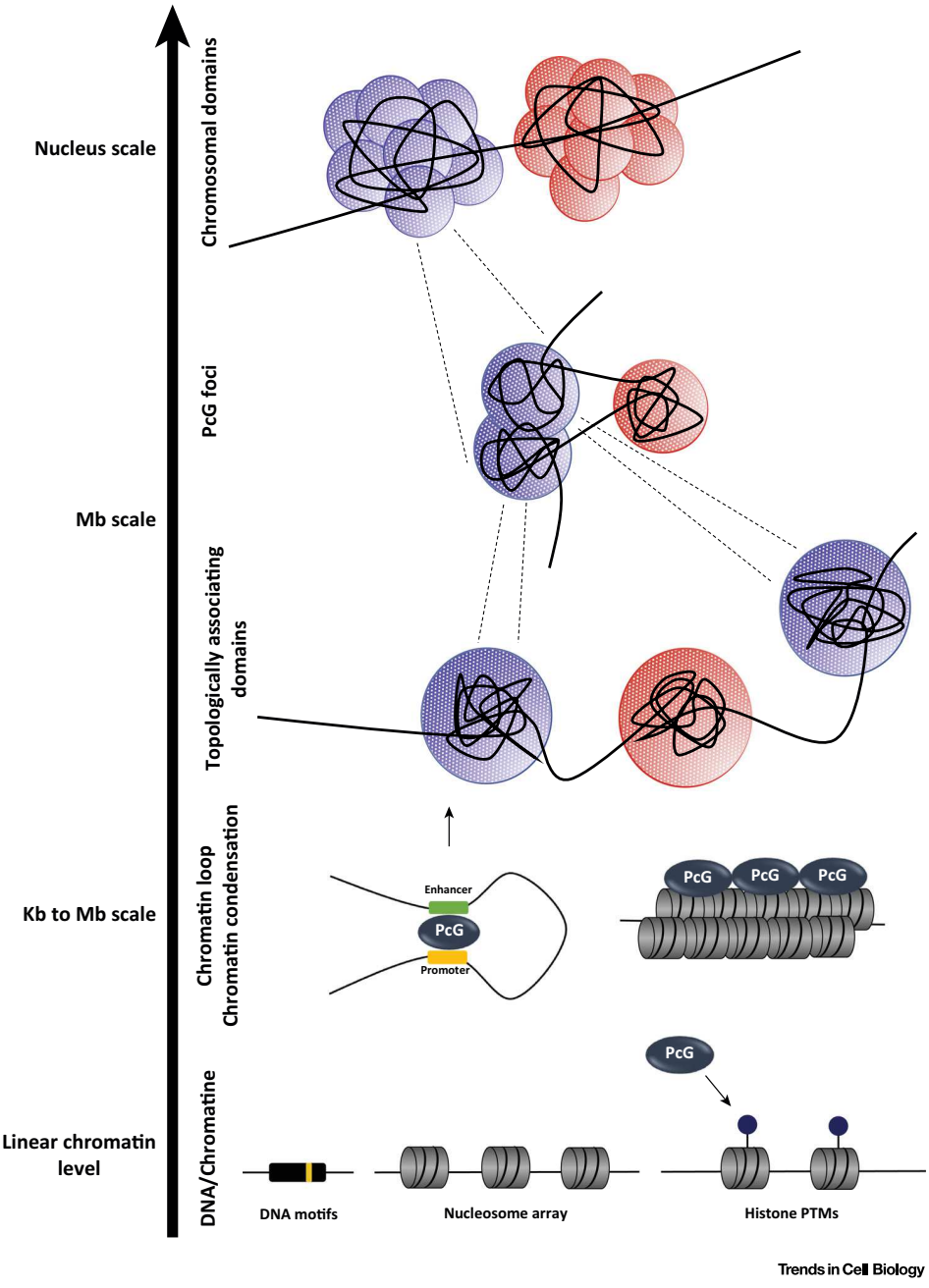


Figure 1. At the linear chromatin level, DNA is wrapped around histone proteins to form nucleosomes. PcG proteins are recruited to the DNA by specific DNA elements to enable post-translational modification (PTM) of surrounding nucleosomes and to regulate nucleosomal structure. Binding of PcG proteins to chromatin can also induce its compaction and PcG proteins are involved in mediating looping interactions between *cis*-regulatory elements like

(Figure legend continued on the bottom of the next page.)

Glossary

Chromosome conformation capture (3C)-derived techniques: 3C is a high-throughput molecular biology technique used to study chromatin structure. 3C provides information about the interaction frequencies of DNA sequences therefore providing information on 3D genome structure.

CpG islands (CGIs): short interspersed DNA sequences that have a high density of CpG dinucleotides and are predominantly non-methylated. In vertebrates, most CGIs map to promoters.

Embryonic stem cells (ESCs): pluripotent stem cells derived from the inner cell mass of blastocysts; have the ability to differentiate into all three germ layers.

Epigenetic: heritable changes in gene expression that do not involve changes to the underlying DNA sequence and are maintained even in the absence of the initial stimulus.

Homeotic (*HOX*) genes: a highly conserved group of genes controlling, in time and space, the body plan formation of an organism. *HOX* genes are expressed in a spatially collinear fashion.

Polycomb group (PcG) proteins: a family of proteins involved in the silencing of key developmental genes.

Polycomb Response Elements (PREs): *cis*-regulatory elements that recruit PcG proteins and mediate epigenetic gene silencing.

Sterile Alpha Motifs (SAM) domain: putative protein interaction module present in many chromatin-associated factors.

Topologically associating domains (TAD): linear units of chromatin that fold as discrete 3D structures tending to favor internal, intra-TAD rather than external, inter-TAD chromatin interactions. They are delimited by sharp boundaries containing housekeeping genes and insulators sites. TADs are detected by 3C-related methods. TADs are observed in many animal species whereas their presence in non-animal species is less clear, suggesting that alternative mechanisms of genome folding are possible.

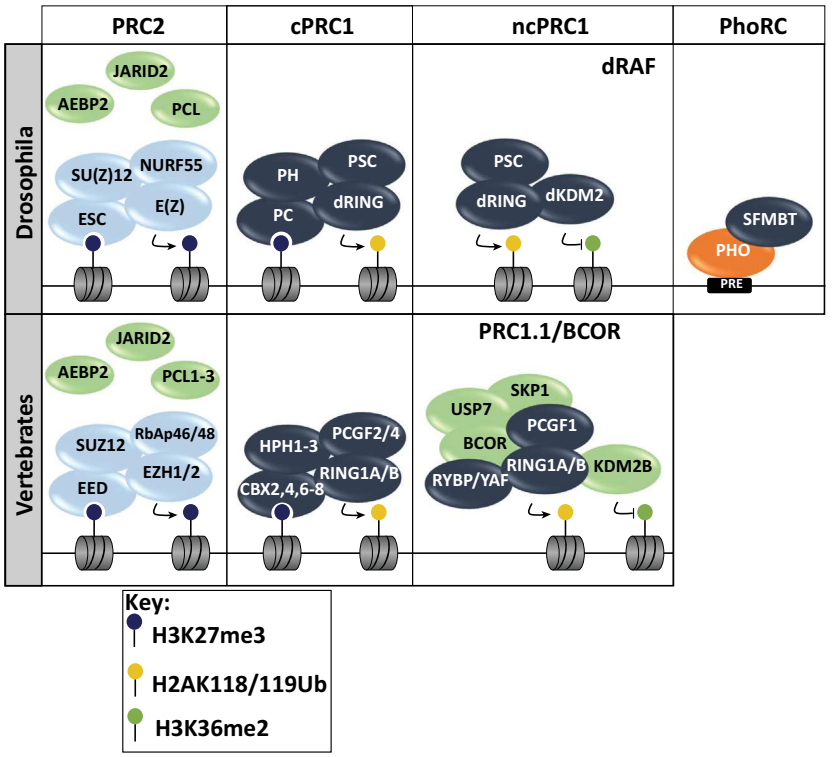


Figure 2. Polycomb Group (PcG) Complex Diversity. Composition of PcG core complexes and their associating proteins in vertebrates and flies. The complete composition of each complex is achieved by interactions between core proteins [light and dark blue for polycomb repressive complex 2 (PRC2) and PRC1, respectively] and accessory proteins (in green). Pho-repressive complex (PhoRC) is the only PcG complex that contains a DNA-binding activity mediated by Pleiohomeotic (PHO) (in orange). The enzymatic activity and histone-binding properties of each complex are indicated.

Structural and Functional Diversity of PcG Proteins

PcG genes were originally identified from mutations that induce homeotic transformations dependent on ectopic expression of **homeotic (*HOX*) genes** [4,5]. Since then, a large number of studies have shown that PcG proteins can dynamically regulate many genes that play key roles in the regulation of cellular processes including cell identity, cell fate choices, cell cycle control, cellular senescence, genomic imprinting, stem cell plasticity, and cellular transformation leading to cancer [6]. This functional diversity is achieved by the assembly of PcG proteins into chromatin-associated multiprotein complexes containing subunits with enzymatic activities that modify histones and regulatory subunits that can modulate enzymatic activities or their chromosomal distribution. The core components of PcG complexes are highly conserved between flies and vertebrates; however, vertebrate PcG complexes have a more diverse composition that arises from alternative versions of subunits, paralogs, or protein isoforms. The three main PcG complexes in vertebrates and *Drosophila* (reviewed in more detail in [7]) are briefly described below (Figure 2).

enhancers and promoters. Specific genomic domains exhibit strong genomic interactions mediated by PcG proteins to form topologically associating domains (TADs). In addition, Polycomb-repressed TADs show long-range interactions with each other mediated by PcG proteins. PcG proteins accumulate in the nucleus to form PcG foci, which are the nuclear counterparts of genomic domains silenced by PcG proteins and may contain individual PcG TADs or multiple TADs engaged in long-range interactions.

Polycomb Repressive Complex 1 (PRC1)

PRC1 complexes can be subdivided into canonical (cPRC1) and noncanonical (ncPRC1) complexes. Both contain two core subunits: a Ring1 protein (RING1A and RING1B) and one of the six PcG Ring Finger proteins (PCGF1–6). RING1A/B contains E3 ubiquitin ligase activity, which monoubiquitinates histone H2A on lysine 119 (H2AK119ub) [8,9]. The main difference between cPRC1 and ncPRC1 is their composition of accessory proteins, which can modulate their recruitment to specific chromatin sites or catalytic activity. cPRC1 is specified by the presence of one chromobox protein (CBX2, 4, 6–8), which binds the H3K27me3 mark with varying affinity [10,11], and one Polyhomeotic (PH) homolog protein (HPHC1–3). By contrast, ncPRC1 possesses YY1-binding protein (RYBP) [12] or its homolog YAF2 [13], which can be found associated with particular sets of accessory proteins defining specific subcomplexes. In *Drosophila*, cPRC1 comprises the *Drosophila* homologs for RING1B, PCGF, CBX, and HPH; namely, dRING (or SCE), PSC (or its functional homolog SU(Z)2 [14]), PC, and PH [15]. One of the mammalian ncPRC1 complexes, PRC1-1 (also named BCOR), has a *Drosophila* homolog called dRING-associated factor complex (dRAF) containing the proteins dRING and PSC and the histone demethylase dKDM2 [16]. dKDM2 demethylates H3K36me2 but also strongly stimulates histone H2A ubiquitylation by dRING. The mammalian BCOR complex comprises RING1B, a PCGF1, BCOR, SK1, USP7, and KDM2B [17,18].

PRC2

In both mammals and *Drosophila*, the PRC2 core complex comprises three proteins: the SET domain-containing histone methyltransferase enhancer of zeste (EZH2 or EZH1), embryonic ectoderm development (EED), and suppressor of zeste (SUZ12). This core complex can catalyze mono-, di-, and trimethylation of histone H3 on lysine 27 (H3K27me1/2/3) [19–21]. EZH1 and EZH2 are mutually exclusive and differentially expressed in proliferating or non-dividing tissues [22,23]. EED binds the H3K27me3 mark and might therefore contribute to the self-propagation of the repressive H3K27me3 mark [24,25]. Finally, the RbAp46/48 proteins were shown to stabilize the complex [26,27]. The *Drosophila* PRC2 core complex comprises E(Z), ESC, SUZ(12), and NURF55 (also called p55 or CAF-1) [28,29]. Several accessory proteins can associate with core PRC2. Among them, PHD finger protein 1 (PHF1) [or its *Drosophila* homolog Polycomb-like (PCL)] stimulates the trimethylation activity of EZH2 towards the H3K27me2 substrate [30,31]. The mammalian zinc finger protein AEBP2 is required for optimal enzymatic activity of the complex [32] and can trigger PRC2 recruitment to chromatin [33]. Finally, a Jumonji family of transcriptional repressors lacking histone demethylase activity, JARID2 [34,35], has a double function for PRC2: it stabilizes PRC2 occupancy on chromatin and regulates H3K27me3 deposition [36–39].

Pho-Repressive Complex (PhoRC)

Drosophila PhoRC comprises the zinc finger protein Pleiohomeotic (PHO) and Scm-related gene containing four mbt domains (dSFMBT), which can bind to H3K9me1 and H3K20me2 through its MBT domain [40]. PhoRC is the only PcG complex containing DNA-binding activity provided by PHO and is likely to play an important role in PRC1 recruitment. In mammals, a recent study has identified a putative functional homolog of PhoRC containing the SFMBT2 protein and the zinc finger domain-containing protein CASZ1, but its role in transcriptional repression has been poorly characterized [41]. No enzymatic activity is associated with PhoRC.

Recruitment of PcG Complexes to Chromatin

Understanding how PcG complexes are recruited to DNA has been a hot topic for many years and many gaps remain in our knowledge (see Outstanding Questions).

PcG Recruitment in Flies

In *Drosophila*, PcG complexes are recruited to specific DNA elements termed **Polycomb Response Elements (PREs)**. Several TFs, including PHO, its homolog PHO-like (PHOL),

GAGA factor (GAF), Pipsqueak (PSQ), Dorsal switch protein (DSP1), Zeste (Z), Grainyhead (GH), SPPS (SP1/KLF), and Combgap (CG) [42], bind to their cognate DNA motifs, which are frequently found enriched at PREs (reviewed in [43]). PHO is thought to play a key role in building a binding platform for the recruitment of other PcG complexes [40]. A hierarchical recruiting mechanism was proposed [44] in which PHO recruits PRC2 via direct interaction with ESC and EZ subunits [44], although this interaction is likely to be transient [40]. E(Z) would then deposit H3K27me3 and this mark would be recognized by the chromodomain of the PC, tethering PRC1 to PREs. However, more recent data put this hierarchical model into question. First, PRC1 can bind to some of its targets even in the absence of H3K27me3 [45]. Direct recruitment of PRC1 by PhoRC is likely to be mediated by the PRC1 accessory protein Sex Combs on Midleg (SCM) [46], which might act as a 'molecular bridge' connecting the PhoRC and PRC1 complexes via the oligomerization ability of the common **Sterile Alpha Motifs (SAM) domain** of dSFMBT, SCM, and PH [46–49]. Second, at least in vertebrates, the hierarchy can be reversed, since PRC1-mediated H2A ubiquitination has been shown to lead to recruitment of PRC2 [50,51]. Similarly in flies, PRC1-mediated H2A ubiquitination has been shown to create a binding site for Aebp2–Jarid2-containing PRC2 to promote H3K27 trimethylation [52]. Third, in addition to PhoRC recruitment to PREs through sequence-specific binding of PHO to the DNA, PhoRC is stabilized by PRC1 [53,54], indicating a positive feedback loop of PRC1 on PhoRC binding.

PcG Recruitment in Vertebrates

Although the repertoire of PcG target genes is highly conserved from flies to vertebrates, the DNA sequences and recruitment mechanisms diverge significantly but also show some similarities. At the sequence level, **CpG islands (CGIs)** play a critical role in PcG recruitment [55,56] and have been suggested to be the mammalian counterparts of fly PREs. Recent studies demonstrated that PRC1 complexes are targeted to CGIs by the zinc finger domain-containing demethylase KDM2B to catalyze H2A ubiquitination, which in turn can recruit PRC2 [50,51,57,58]. However, most KDM2B-bound CGIs sites do not recruit PcG proteins (Box 1). Recent evidence suggests that the transcription status at CGIs is a key component of PcG recruitment, as global inhibition of transcription leads to ectopic recruitment of PcG proteins to silenced CGIs [56]. A 'chromatin sampling' model has been proposed whereby PcG proteins weakly interact with all potential binding sites (CGIs) but transcription can prevent stable PcG binding [59]. A similar scenario could be considered in flies for active promoters bound by PHO (Box 1). A recent study in *Xenopus* suggests that non-methylated DNA rather than GC richness is the key for PcG recruitment [60]. Furthermore, the possibility remains that a specific combination of sequence-specific DNA-binding proteins might direct PcG complexes to unmethylated sequences. Several DNA-binding proteins, such as AEBP2 [61], Rest [62,63], Runx1 [64], E2F6 [65–67], Snail [68], and Jarid2 [34,39,69], have been linked to PcG recruitment. Intriguingly, YY1, the mammalian homolog of the key PcG recruiter in flies, PHO, has been shown to recruit PcG proteins [13] to only a minority of PcG target genes [70,71]. The implication of the recently identified mammalian homolog of GAF [72] in recruiting PcG complexes has not yet been investigated. Finally, ncRNAs were suggested to play an important role in the recruitment of PcG complexes in mammals but not in flies (for reviews see [73,74]). In summary, multiple pathways and mechanisms contribute to recruit PcG proteins in flies and in vertebrates. Whether these different strategies reflect the recruitment of different PcG complexes containing specific subunits (canonical vs noncanonical) should be clarified in the future.

Gene Repression by PcG Proteins Involves Multiple Layers of Chromatin Organization

PcG complexes can repress their target genes via multiple non-mutually exclusive mechanisms and chromatin organization levels.

Box 1. PcG-Targeting Elements in Flies and Vertebrates

What are the sequence characteristics that specify a genomic locus to be a PRE and distinguish it from non-PcG-recruiting sites? In *Drosophila*, bioinformatic approaches looking for similarities between PREs based on the combination of pairs of sequence motifs for DNA-binding proteins have relatively low prediction potential [121,122]. This might be explained by the fact that DNA motifs for sequence-specific DNA-binding proteins are found not only at PREs but also at active promoters [123]. In agreement with this observation, a large portion of the sites bound by TFs involved in PcG recruitment (e.g., the zinc finger domain-containing protein PHO) are located not only at PcG-bound regions but also at many promoter regions of genes associated with active histone marks (H3K4me3 sites) that are devoid of PcG proteins (Figure I) [123]. In addition, knockdown of PHO does not alter global PcG binding [43]. In vertebrates, non-methylated CGIs are considered to be the functional homologs of fly PREs [55,56,70]. It is noteworthy that although KDM2B (which contains a zinc finger domain) was found to be associated with all CGIs genome wide [50,57,58], it colocalizes with PcG proteins at only around 30% of these sites, and most PcG-binding sites are not affected on loss of KDM2B. This is reminiscent of the situation in flies. In addition, one-third of vertebrate PcG-binding sites do not overlap with annotated CGIs (Figure I). Likewise, fly PREs differ in their sequence and a significant subset of PREs do not contain DNA motifs for PcG-recruiter TFs. Together, these observations suggest that PREs do not behave as a single class but that different types of PRE might exist, characterized by different sequence features and different mechanisms for targeting PcG proteins. The challenge ahead is to define these different classes of PREs and subsequently determine their features to understand their mode of function.

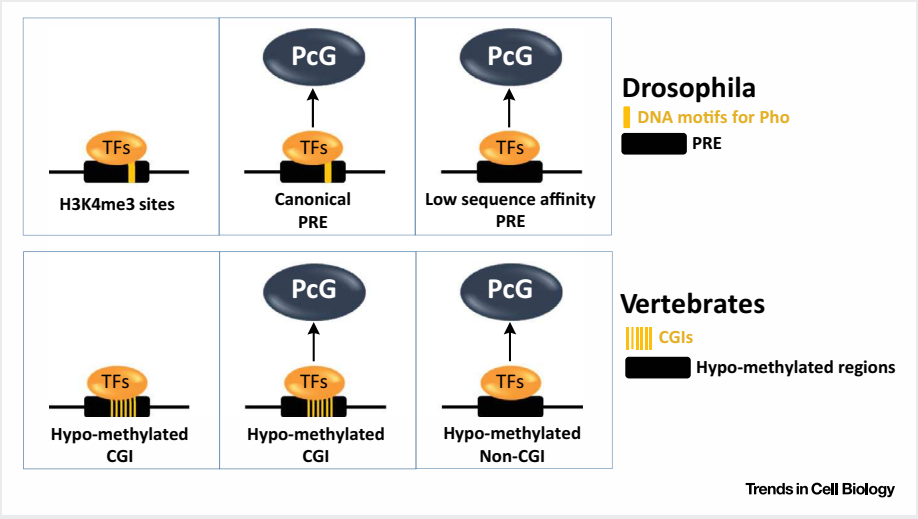


Figure I. Comparison of Fly and Vertebrate Polycomb Group (PcG) Recruiting Sequences.

PcG Function at the Linear Chromatin Level

PREs act as nucleation sites for PcG complexes to create large repressive Polycomb domains in which H3K27me3 is crucial for PRC2-mediated gene silencing [75]. H3K27me3 can directly block the deposition of the antagonistically activating acetylation mark on H3K27 (H3K27ac) and can interfere with the recruitment of RNA polymerase II (RNA Pol II) to target promoters [76]. PRC1-mediated H2A ubiquitination on lysine 119 (H2AK119ub) has also been suggested to interfere with multiple steps of the transcription process, including transcription initiation through blocking methylation of H3K4 [77], RNA Pol II release by interfering with the recruitment of the FACT complex [78], and transcriptional elongation [79]. However, the relevance of H2AK119ub in PcG-dependent gene silencing is disputed. In particular, flies deficient for H2AK119ub can fully maintain repression of PRC1 target genes and do not show characteristic PcG phenotypes [80]. A recent study discovered a new mechanism of PRC1-dependent gene silencing: Polycomb physically interacts with the acetyltransferase CREB-binding protein (CBP) and represses its catalytic activity towards H3K27. Moreover, Polycomb and CBP were shown to co-occupy many promoters together with paused RNA Pol II, suggesting a role for PRC1 in gene pausing [81]. Since H3K27Ac also marks active enhancers, preventing the deposition of H3K27Ac might be particularly important to prevent illegitimate enhancer activation in tissues where they must be silenced, a possibility that

remains to be tested. In addition, PSC has an intrinsic ability to compact nucleosome arrays and inhibit chromatin remodeling activity [15]. Chromatin compaction mediated by PRC1 seems to be independent of PRC2 [82] and H2A ubiquitination, since the catalytic activity of RING1B is dispensable for chromatin compaction and gene repression [83]. Local chromatin compaction may precede the *in vivo* deposition of H3K27me3 and may be important to provide a better substrate for PRC2 activity [84]. Together, this evidence suggests that PRC1 might initiate local chromatin compaction to induce the establishment and propagation of H3K27me3 to form repressive Polycomb domains.

PcG complexes can also play more global roles in gene regulation than originally anticipated: PRC2 is required for the dimethylation of H3K27, which accounts for up to 70% of all histone H3, whereas H3K27me3 is much less abundant and strongly depends on cell type [85–89]. Knockdown of either E(Z) or SU(Z)12 in *Drosophila* cells results in global transcriptional derepression, most prominently in silent or weakly transcribed regions, which correlates with a global increase in H3K27ac and H3K4me1. Therefore, H3K27me2 might prevent spurious transcription throughout the genome [90]. The link between H2AK119ub and PRC1-mediated repression is more mysterious. H2AK119ub was found to be surprisingly widespread but only very partially associated with PRC1 [90]. Finally, one should note that, in addition to a repressive function, several recent reports have raised the possibility of a role of PcG proteins in transcriptional activation (Box 2).

PcG Function in Chromatin Looping/Interactions

In addition to their local effects, PcG proteins mediate chromatin looping interactions between DNA regulatory elements (like PREs, promoters, enhancers, and insulators [91,92]), therefore providing an additional layer in the complexity of PcG-mediated gene regulation (Figure 1). Although the global 3D genome organization is markedly different between flies and vertebrates due to differences in genome size and the presence or absence of homologous chromosome pairing, PcG proteins seem to be key players in chromatin organization in both systems. On microscopy, PcG proteins accumulate in discrete subnuclear structures (so-called PcG foci) in flies [93] and vertebrates [94] that have been suggested to represent the locations of PcG-mediated gene silencing, where PcG target sites cluster via chromatin looping (reviewed in [95]). The first evidence for a physical interaction of PREs with each other and with target promoters was provided in the largest *Drosophila* Polycomb domain, the Bithorax complex (*BX-C*) [91]. Subsequently, a role for EZH2 in chromatin looping and gene silencing of the *GATA-4* locus was demonstrated in **embryonic stem cells (ESCs)** [96]. In addition to an action of PREs in *cis* (within the same Polycomb domain), PREs of different Polycomb domains can interact over long distances in *Drosophila*, thereby enhancing PcG-dependent silencing [97]. Although long-range interactions are far from being pancellular, functional PcG protein-dependent *trans*-interactions of a transgenic PRE with its homologous endogenous PRE have been described. They result in the relocalization of PREs within the same PcG foci, correlating with increased PcG binding at the interacting loci and with reinforced silencing of genes within the Polycomb domain [98–100].

In mouse ESCs, the PRC2 subunit EED has been shown to be required for the maintenance of interactions between Polycomb-regulated regions separated by tens of megabases or located on different chromosomes [101]. These interactions involve *HOX* and other genes [102]; they are not seen in ground-state pluripotent mESCs (cultured in 2i medium), where H3K27me3 is largely absent, but they appear in primed mESCs (cultured in serum), which gain H3K27me3 at PcG targets [103]. Another recent study showed a major role of RING1B in establishing long-range interactions between promoters of PcG-regulated genes in mESCs [104]. On RING1 depletion, the contact network of PcG target genes was lost, whereas gene networks of pluripotency factors that form independently of PcG proteins remained unaffected, indicating a causal and specific role of PRC1 in PcG target gene network formation. PcG-bound promoters also

Box 2. Gene Activation by PcG Proteins

Although PcG proteins mainly regulate their targets by silencing them, they have also been shown to mediate gene activation in specific cases. A first hint of this role came from genome-wide mapping studies reporting the association of PcG proteins with actively transcribed genes in various species [124–128]. Subsequently, several mechanisms for a switch from repressive to activating PcG complexes were revealed. First, PcG complexes can exchange subunits in a stage-specific manner to instruct sequential gene activation and repression programs: during ESC differentiation towards an ectodermal cell fate, a subunit exchange from CBX7 to CBX8 is required for the initial activation of a subset of developmental genes [129]. Similarly, on differentiation of ESCs into cardiac mesoderm precursors, the Mel18–PRC1 complex swaps its CBX7 subunit with CBX2 to mediate the activation of genes involved in cardiovascular system development [130]. Finally, during lineage commitment, disengagement of EZH2 from silenced genes correlates with their activation whereas EZH1-containing complexes are recruited to newly activated genes, positively influencing transcription [127]. The association of PcG complexes with additional regulatory proteins can also lead to gene activation: association of the PRC1–5 complex with the Auts2 protein triggers gene activation via recruitment of the histone acetyltransferase p300 [131]. Further, a PRC1 complex containing RING1B, BMI1, and CBX7 was found to interact with Aurora B kinase in quiescent B cells at active genes, where it is required for binding of RNA Pol II to active promoters [125]. Recently, Polycomb has been shown to function together with PR-SET7 to modulate H4K20me1, leading to gene activation [132]. Moreover, post-translational modifications of PcG complex subunits can lead to a switch in their activity. The PRC1–5 complex can act as a transcriptional activator after phosphorylation of the RING1B subunit by CK2 protein, which decreases the catalytic activity of RING1B [131]. In addition, PcG proteins can act as positive regulators of transcription after mitosis [133]. BMI1 and RING1A can stimulate the ubiquitination of chromatin at gene promoters during mitosis, a bookmark that might be required for the reactivation of marked genes at entry into G1 phase. Finally, PcG proteins can contribute to gene activation by regulating the topological conformation of *cis*-regulatory sequences (Figure I). Activation of *Meis2* during embryonic development of the midbrain requires a transient interaction step of multiple regulatory sequences mediated by RING1B. In the absence of RING1B enhancer–promoter interactions cannot occur, resulting in impaired gene activation [105].

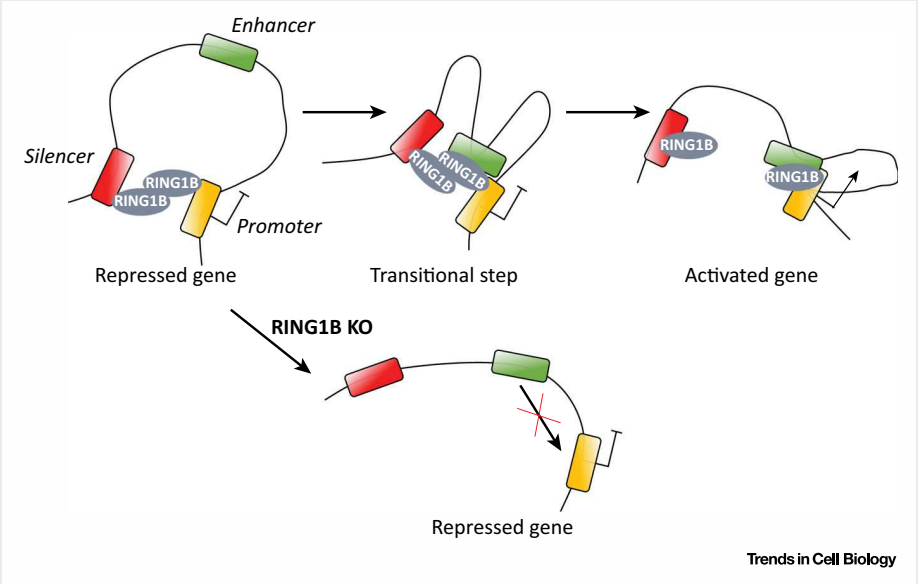


Figure I. Gene Activation by Polycomb Group (PcG) Proteins through Regulation of Genome Architecture.

interacted with poised enhancers marked by H3K4me1 and H3K27me3. Interestingly, poised enhancers became activated whereas enhancer–promoter contacts were maintained in RING1 knockout cells, indicating that PRC1 proteins are not essential for these contacts but inhibit inappropriate activation of target promoters. A different role of RING1B in mediating interactions of distal regulatory elements has been revealed by Kondo *et al.* during embryonic development of the mouse midbrain. In the repressed state, the *Meis2* gene promoter is associated with a silencing element bound by RING1B. Activation of the *Meis2* gene requires a topological switch mediated by RING1B resulting in an association between the *Meis2* gene promoter and an enhancer. In the absence of RING1B, the enhancer can no longer contact the promoter region, resulting in impaired *Meis2* gene expression (Box 2) [105]. Here, in contrast to the study of

Schoenfelder *et al.*, RING1B is thus required to mediate promoter–enhancer interactions to activate *Meis2* gene expression. In summary, both studies highlight the role of RING1B in regulating 3D chromatin structures involved in gene regulation, while future studies are needed to investigate how these contacts are created and to understand the diversity of their molecular roles in regulating gene expression.

PcG Function in Global Nuclear Architecture

An emerging additional layer in PcG-dependent gene regulation involves global nuclear architecture. Metazoan genomes are organized into TADs, which correspond to linear chromatin regions that fold into specific 3D structures defined by strong genomic interactions within a domain, whereas interactions spanning TAD borders are strongly reduced. Previously described chromatin looping interactions at the *Drosophila BX-C* [91] or the vertebrate *GATA-4* locus [96] do not suffice to specify TADs. Whether they are the driving forces in their formation remains to be studied, and possibly the combined action of cohesion and CCCTC-binding factor (CTCF) is an important component in shaping these architectural units of the genome [106].

TADs are highly conserved across different cell types and species [107,108] and segregate according to their association with the nuclear lamina, replication timing, patterns of histone modifications, and the transcriptional level of the genes within a TAD. Different TADs can be therefore divided into active and repressed (repressive TADs being further subdivided into heterochromatin-associated, Polycomb-associated and null, which contain repressed genes without a specific chromatin composition). TAD borders are formed in regions containing actively expressed genes and clusters of insulator-binding sites [109,110].

Since the discovery of TADs, much effort has been devoted to understand the link between the dynamics of these 3D structures and lineage-specific gene expression and cell identity. The *HOX* gene clusters are among the best-studied PcG topological domains. When *HOX* genes are transcriptionally inactive, they associate into a single TAD that is well separated from flanking active regions. CTCF seems to be a key protein in insulating active and repressed *HOX* clusters into spatially disjoint domains [111]. Vieux-Rochas *et al.* showed that repressed *HOX* clusters marked by H3K27me3 are part of a network of H3K27me3-marked long-range interactions that extend beyond a single TAD [102]. Importantly, these intra- and interchromosomal interactions occur in a globally active chromatin environment. On *HOX* gene activation, active genes progressively segregate into active TADs and the transition in spatial configuration coincides with the change of chromatin marks from a repressed to an active state [112,113]. Taken together, these data indicate that TAD reorganization correlates with changes in histone marks but do not distinguish whether loss of PcG-associated chromatin marks is a cause or a consequence of TAD reorganization.

The role of PRC2 in regulating global nuclear architecture is somewhat conflicting. Although mutation of the PRC2 subunit EED in mouse ESCs results in the loss of interactions between Polycomb-regulated regions [101,103], the loss of PRC2 was shown to have no effects on TAD boundaries or on the global contact pattern of a large region of the X chromosome [114], indicating that TAD formation is not due to domain-wide H3K27me3 enrichment. However, a recent study demonstrated that induced recruitment of EZH2 can reposition a subcompartment to a new nuclear compartment occupied by other PcG-bound regions. Importantly, switching between nuclear compartments can be uncoupled from transcriptional changes, suggesting that specific nuclear subcompartments form as a consequence of affinities between chromatin-associated proteins or modifications [115].

The PRC1 component PH might be a key player in higher-order chromatin organization by forming long-range contacts between distant PH-bound chromosomal sites. PH contains a

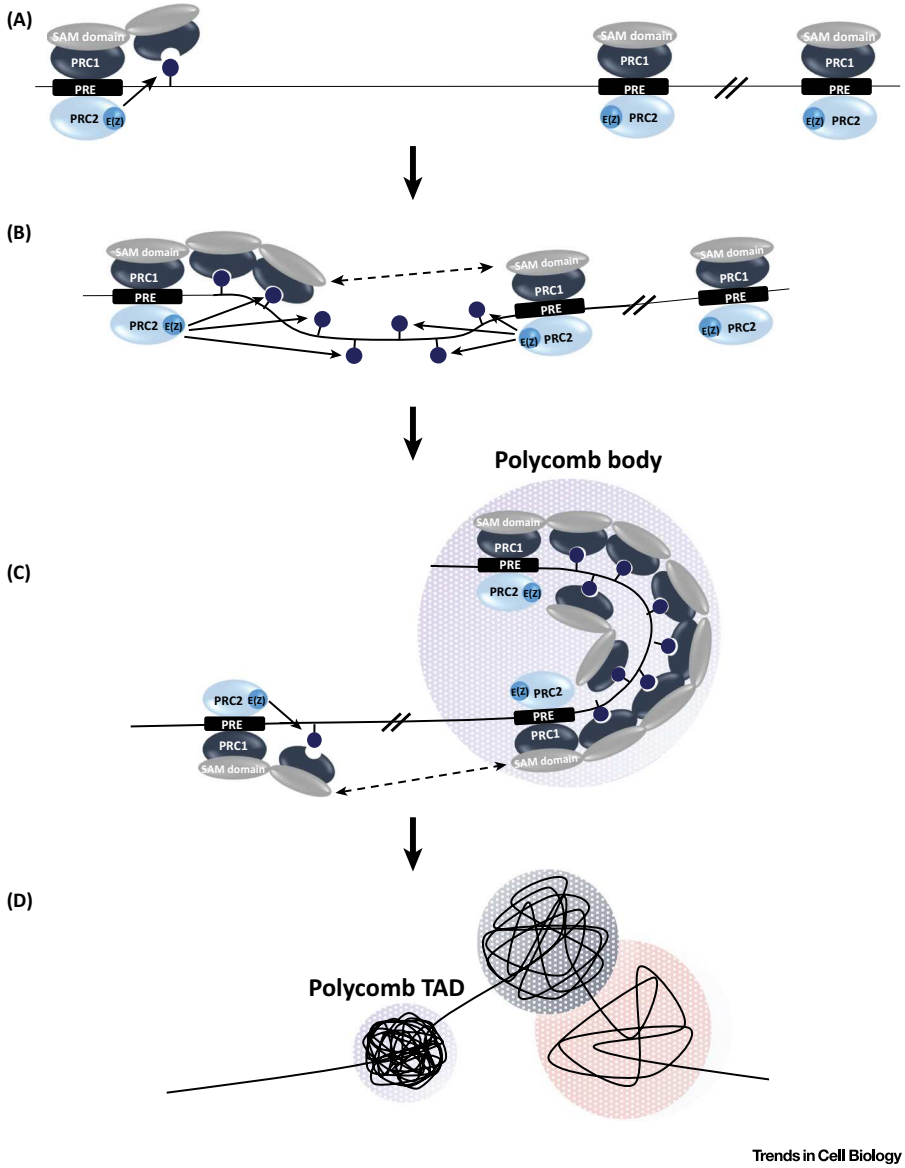


Figure 3. Role of the Sterile Alpha Motifs (SAM) Domain of Polyhomeotic (PH) Protein in the Formation of Polycomb Domains and Mediating Long-Range Interactions. (A) Polycomb group (PcG) complexes are recruited to their target sites [Polycomb Response Elements (PREs)] and binding is retained through SAM-mediated interactions of polycomb repressive complex 1 (PRC1). (B) Distal PREs interact with each other. PcG-mediated looping (indicated by the broken arrow) might be facilitated by the polymerization of PH SAM domains bound at different PREs. Interaction of PREs and chromatin looping helps to bring surrounding chromatin regions closer to the PRE, helping the PRC2 complex to trimethylate histone H3 (blue lollipops) at the surrounding region. (C) Trimethylation of histone H3 lysine 27 (H3K27me3) and PH SAM polymerization facilitate the propagation of PRC1 over extended chromatin regions, helping to robustly repress chromatin domains. SAM polymerization-dependent PRC1 clustering around PREs corresponds to microscopically visible subnuclear Polycomb domains (Polycomb body). (D) The chromosome domain counterpart of microscopically visible PcG foci is represented by Polycomb topologically associating domains (TADs). Polymerization of PRC1 complexes via the PH SAM domain stimulates the formation of Polycomb-repressed TADs (blue) with a significantly higher chromatin-packing density compared with active (red) or transcriptionally inactive (light grey) domains. This architecture prevents the intermingling of Polycomb domains with active domains. SAM-mediated PRC1 clustering might also facilitate long-range interactions of various PcG TADs, thereby enhancing gene silencing.

SAM domain, which can form homo- and hetero-oligomers [116,117]. Oligomerization of PH via its SAM domain is essential for PcG-mediated repression in *Drosophila* [118]. In addition, PRC1 oligomerization has been proposed to stabilize PcG complexes bound to their target sites and has been proposed to facilitate the binding of PcG complexes to extended chromatin regions by mediating long-range interactions between separated PRC1-binding sites in *Drosophila* [116] (Figure 3). Two recent super-resolution imaging studies in *Drosophila* provided further mechanistic evidence for the role of the PRC1 component PH in the modulation of chromatin architecture. Boettiger *et al.* revealed fundamentally different chromatin behavior of Polycomb-associated TADs compared with active or transcriptionally inactive TADs: Polycomb TADs are the most compact and show the highest inter-TAD interactions but minimal contact with other domains, whereas active and transcriptionally inactive (null) TADs show higher levels of intermixing (Figure 3). These particular properties of Polycomb TADs are lost on loss of function of PH [119]. A parallel study showed that disruption of the polymerization activity of the PH SAM domain leads to a marked dispersal of PcG clusters/foci and chromatin interactions, which correlates with changes in gene expression levels. This suggests that the organization of PcG proteins into nanometer-scale Polycomb clusters/foci mediated by the PH SAM domain may shape genome architecture through chromatin interactions in *Drosophila* [117]. Future studies will reveal whether the PH SAM domain has similar functions in vertebrates. The peculiarity of PH in regulating chromatin organization raises the intriguing hypothesis that canonical or noncanonical PcG complexes may play distinct roles in 3D genome organization, depending on the presence or absence of PH (see Outstanding Questions).

Box 3. 3D Genome Architecture Regulates TF Binding

Chromatin folding and establishment of 3D genome architecture is thought to occur downstream of the initial targeting of TFs and chromatin-modifying complexes [95]. A recent study challenges this dogma and suggests that the 3D genome architecture of Polycomb-associated topological domains can influence the binding of specific chromatin factors to the DNA [54]: a comparative genomics study in *Drosophila* species demonstrated that sequence-specific binding of the sequence-specific DNA-binding protein PHO outside a Polycomb context requires the presence of strong Pho consensus motifs. By contrast, within Polycomb domains PHO is able to bind to genomic sites containing far weaker motifs. Notably, these sites participate in frequent chromatin interactions, consistent with known looped interactions between PREs (Figure I) [91,97,110]. By contrast, similar genomic regions outside Polycomb domains show much lower contact frequencies and no Pho binding (Figure I). This suggests that the 3D association of genomic sites within Polycomb domains stabilizes the binding of a TF. Therefore, nuclear architecture can have a regulatory function in TF binding, similar to local chromatin structure (such as nucleosome positioning or chromatin compaction). Future work will show whether this finding reflects a specific feature of Polycomb domains or whether it might apply to other chromatin factors and TADs.

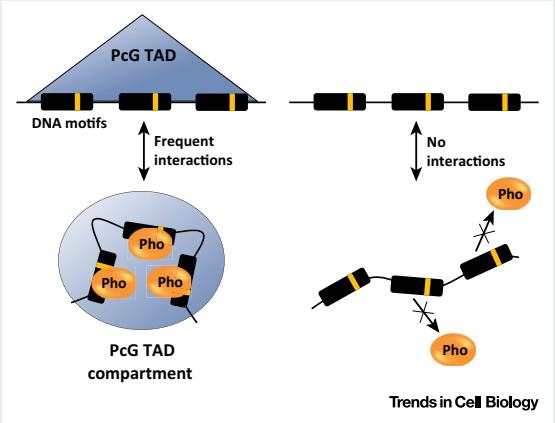


Figure I. Pleiohomeotic (PHO) Transcription Factor (TF) Binding Is Enhanced by Chromosome Folding.

In summary, recruitment of PcG proteins to their target sites does not only modulate local chromatin structure but also mediates looping interactions between regulatory elements and shapes global nuclear architecture, thereby regulating gene expression at multiple scales. Intriguingly, a recent study proposed that PcG topological domains are not only defined by the binding of chromatin factors to the DNA but also set the stage for cooperative binding of specific chromatin factors in designated chromatin hubs (Box 3) [54]. If this feature were to extend to other DNA-binding proteins, 3D chromosome architecture would play a global role in regulating gene expression by modulating protein–DNA interactions.

Concluding Remarks

Since the discovery of PcG proteins four decades ago, the spectrum of PcG function has not stopped growing. Consistent with the huge functional diversity of PcG proteins, biochemical studies have shown a greater diversity of PcG complexes than was originally anticipated. PcG complexes can employ various mechanisms to find their chromosomal targets. An intriguing, but yet to be tested, hypothesis is that different PcG complexes might use different target sequences and mechanisms to find their target DNA and exert specific functions. Once recruited to their targets, PcG proteins employ diverse mechanisms to regulate their target genes and may also activate some of their targets in specific circumstances.

PcG proteins regulate chromatin structure at multiple levels starting at the linear genome by modifying histones, local chromatin compaction, mediating looping interactions between regulatory regions, regulating long-distance interactions, and organizing TADs in the nuclear space. However, the cause–consequence relationship between Polycomb-mediated gene regulation and subnuclear chromatin organization remains unclear (see Outstanding Questions). Strategies to improve the resolution of genome-wide interaction maps will further help in understanding the precise role of PcG components in regulating nuclear architecture. In addition, the application of high-resolution imaging and new genome-engineering tools such as CRISPR/Cas9 will help not only to more precisely describe the genome architecture but also to modulate genome structure. Together these approaches should allow the uncoupling of loss of PcG binding to chromatin from the loss of long-range interactions and help us understand mechanistically the role of PcG proteins in regulating genome architecture. Finally, the intriguing hypothesis that nuclear architecture itself confers epigenetic and heritable gene control should be tested in the future (see Outstanding Questions). It has been previously shown that *Drosophila* PREs can mediate long-distance interchromosomal interactions and that these interactions may be heritable through mitotic and meiotic cell division [99]. Importantly, PSC is partially maintained on chromatin during mitosis at sites corresponding to boundaries of topological domains [120], suggesting a possible role of PSC bookmarking in the maintenance of chromatin compartments throughout the cell cycle.

Outstanding Questions

What are the differences in the recruitment mechanisms and sequence requirements for canonical and noncanonical PcG complexes and do different classes of PRE exist in flies and mammals?

How are PcG complexes recruited to active genes?

Do different PcG complexes play distinct roles in organizing the 3D genome?

What is the cause–consequence relationship between Polycomb-mediated gene regulation and subnuclear chromatin organization?

Does the nuclear architecture itself confer epigenetic and heritable gene control in a tissue- or cell type-dependent manner?

Acknowledgments

The authors thank Thomas Sexton (IGBMC, France) for critically reading the manuscript. Research in the laboratory of G.C. was supported by grants from the European Research Council (ERC-2008-AdG No 232947), the European H2020 EINFRA MuG grant, the CNRS, the INSERM, the European Network of Excellence EpiGeneSys, the Agence Nationale de la Recherche (EpiDevoMath), the MMTT grant of the ITMO Cancer and INSERM, and the Laboratory of Excellence EpiGenMed. M.E. was supported by the French Ministry of Research.

References

1. Ptashne, M. (2007) On the use of the word 'epigenetic'. *Curr. Biol.* 17, R233–R236
2. Sexton, T. and Cavalli, G. (2015) The role of chromosome domains in shaping the functional genome. *Cell* 160, 1049–1059
3. Di Croce, L. and Helin, K. (2013) Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* 20, 1147–1155
4. Lewis, E.B. (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565–570
5. Duncan, I.M. (1982) Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of *Drosophila melanogaster*. *Genetics* 102, 49–70

6. Schuettengruber, B. and Cavalli, G. (2009) Recruitment of Polycomb group complexes and their role in the dynamic regulation of cell fate choice. *Development* 136, 3531–3542
7. Aranda, S. *et al.* (2015) Regulation of gene transcription by Polycomb proteins. *Sci. Adv.* 1, e1500737
8. de Napoles, 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
9. Wang, H. *et al.* (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878
10. Fischle, W. *et al.* (2003) Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* 17, 1870–1881
11. Min, J. *et al.* (2003) Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* 17, 1823–1828
12. Tavares, L. *et al.* (2012) RYBP–PRC1 complexes mediate H2A ubiquitylation at Polycomb target sites independently of PRC2 and H3K27me3. *Cell* 148, 664–678
13. Basu, A. *et al.* (2014) YY1 DNA binding and interaction with YAF2 is essential for Polycomb recruitment. *Nucleic Acids Res.* 42, 2208–2223
14. Lo, S.M. *et al.* (2009) Polycomb group protein Suppressor 2 of zeste is a functional homolog of Posterior Sex Combs. *Mol. Cell Biol.* 29, 515–525
15. Francis, N.J. *et al.* (2001) Reconstitution of a functional core Polycomb repressive complex. *Mol. Cell* 8, 545–556
16. Lagarou, A. *et al.* (2008) dKDM2 couples histone H2A ubiquitylation to histone H3 demethylation during Polycomb group silencing. *Genes Dev.* 22, 2799–2810
17. Gearhart, M.D. *et al.* (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
18. Sanchez, C. *et al.* (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
19. Pasini, D. *et al.* (2004) Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO J.* 23, 4061–4071
20. Muller, J. *et al.* (2002) Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. *Cell* 111, 197–208
21. 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
22. Margueron, R. *et al.* (2008) Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Mol. Cell* 32, 503–518
23. Shen, X. *et al.* (2008) EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining stem cell identity and executing pluripotency. *Mol. Cell* 32, 491–502
24. Hansen, K.H. *et al.* (2008) A model for transmission of the H3K27me3 epigenetic mark. *Nat. Cell Biol.* 10, 1291–1300
25. Margueron, R. *et al.* (2009) Role of the Polycomb protein EED in the propagation of repressive histone marks. *Nature* 461, 762–767
26. Ciferri, C. *et al.* (2012) Molecular architecture of human Polycomb repressive complex 2. *Elife* 1, e00005
27. Satrimafitrah, P. *et al.* (2015) RbAp48 is essential for viability of vertebrate cells and plays a role in chromosome stability. *Chromosome Res.*
28. Anderson, A.E. *et al.* (2011) The enhancer of trithorax and Polycomb gene Caf1/p55 is essential for cell survival and patterning in *Drosophila* development. *Development* 138, 1957–1966
29. Nekrasov, M. *et al.* (2005) Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. *EMBO Rep.* 6, 348–353
30. Nekrasov, M. *et al.* (2007) Pcl-PRC2 is needed to generate high levels of H3-K27 trimethylation at Polycomb target genes. *EMBO J.* 26, 4078–4088
31. Sarma, K. *et al.* (2008) Ezh2 requires PHF1 to efficiently catalyze H3 lysine 27 trimethylation in vivo. *Mol. Cell Biol.* 28, 2718–2731
32. Cao, R. and Zhang, Y. (2004) SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED–EZH2 complex. *Mol. Cell* 15, 57–67
33. Kim, H. *et al.* (2009) AEBP2 as a potential targeting protein for Polycomb repression complex PRC2. *Nucleic Acids Res.* 37, 2940–2950
34. Peng, J.C. *et al.* (2009) Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139, 1290–1302
35. Zhang, Z. *et al.* (2011) PRC2 complexes with JARID2, MTF2, and esPRC2p48 in ES cells to modulate ES cell pluripotency and somatic cell reprogramming. *Stem Cells* 29, 229–240
36. da Rocha, S.T. *et al.* (2014) Jarid2 Is implicated in the initial xist-induced targeting of PRC2 to the inactive X chromosome. *Mol. Cell* 53, 301–316
37. Landeira, D. *et al.* (2010) Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA polymerase II to developmental regulators. *Nat. Cell Biol.* 12, 618–624
38. Pasini, D. *et al.* (2010) JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464, 306–310
39. Son, J. *et al.* (2013) Nucleosome-binding activities within JARID2 and EZH1 regulate the function of PRC2 on chromatin. *Genes Dev.* 27, 2663–2677
40. Klymenko, T. *et al.* (2006) A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev.* 20, 1110–1122
41. Lecona, E. *et al.* (2013) Polycomb protein SCML2 regulates the cell cycle by binding and modulating CDK/CYCLIN/p21 complexes. *PLoS Biol.* 11, e1001737
42. Ray, P. *et al.* (2016) Combgap contributes to recruitment of Polycomb group proteins in *Drosophila*. *Proc. Natl. Acad Sci. U.S.A.* 113, 3826–3831
43. Kassis, J.A. and Brown, J.L. (2013) Polycomb group response elements in *Drosophila* and vertebrates. *Adv. Genet.* 81, 83–118
44. Wang, L. *et al.* (2004) Hierarchical recruitment of Polycomb group silencing complexes. *Mol. Cell* 14, 637–646
45. Pasini, D. *et al.* (2007) The Polycomb group protein Suz12 is required for embryonic stem cell differentiation. *Mol. Cell Biol.* 27, 3769–3779
46. Grimm, C. *et al.* (2009) Molecular recognition of histone lysine methylation by the Polycomb group repressor dSfmbt. *EMBO J.* 28, 1965–1977
47. Kim, C.A. *et al.* (2005) Structural organization of a sex-comb-on-midleg/polyhomeotic copolymer. *J. Biol. Chem.* 280, 27769–27775
48. Peterson, A.J. *et al.* (1997) A domain shared by the Polycomb group proteins Scm and ph mediates heterotypic and homotypic interactions. *Mol. Cell Biol.* 17, 6683–6692
49. Peterson, A.J. *et al.* (2004) Requirement for sex comb on midleg protein interactions in *Drosophila* Polycomb group repression. *Genetics* 167, 1225–1239
50. Blackledge, N.P. *et al.* (2014) Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and Polycomb domain formation. *Cell* 157, 1445–1459
51. Cooper, S. *et al.* (2014) Targeting Polycomb to pericentric heterochromatin in embryonic stem cells reveals a role for H2AK119u1 in PRC2 recruitment. *Cell Rep.* 7, 1456–1470
52. Kalb, R. *et al.* (2014) Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nat. Struct. Mol. Biol.* 21, 569–571
53. Kahn, T.G. *et al.* (2014) Combinatorial interactions are required for the efficient recruitment of pho repressive complex (PhoRC) to Polycomb response elements. *PLoS Genet.* 10, e1004495
54. Schuettengruber, B. *et al.* (2014) Cooperativity, specificity, and evolutionary stability of Polycomb targeting in *Drosophila*. *Cell Rep.* 9, 219–233
55. Mendenhall, E.M. *et al.* (2010) GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* 6, e1001244

56. Riising, E.M. *et al.* (2014) Gene silencing triggers Polycomb repressive complex 2 recruitment to CpG islands genome wide. *Mol. Cell* 55, 347–360
57. Farcas, A.M. *et al.* (2012) KDM2B links the Polycomb repressive complex 1 (PRC1) to recognition of CpG islands. *Elife* 1, e00205
58. Wu, X. *et al.* (2013) Fbxl10/Kdm2b recruits Polycomb repressive complex 1 to CpG islands and regulates H2A ubiquitylation. *Mol. Cell* 49, 1134–1146
59. Klose, R.J. *et al.* (2013) Chromatin sampling – an emerging perspective on targeting Polycomb repressor proteins. *PLoS Genet.* 9, e1003717
60. van Heeringen, S.J. *et al.* (2014) Principles of nucleation of H3K27 methylation during embryonic development. *Genome Res.* 24, 401–410
61. Kim, H. *et al.* (2015) AEBP2 as a transcriptional activator and its role in cell migration. *Genomics* 105, 108–115
62. Dietrich, N. *et al.* (2012) REST-mediated recruitment of Polycomb repressor complexes in mammalian cells. *PLoS Genet.* 8, e1002494
63. Ren, X. and Kerppola, T.K. (2011) REST interacts with Cbx proteins and regulates Polycomb repressive complex 1 occupancy at RE1 elements. *Mol. Cell Biol.* 31, 2100–2110
64. Yu, M. *et al.* (2012) Direct recruitment of Polycomb repressive complex 1 to chromatin by core binding transcription factors. *Mol. Cell* 45, 330–343
65. Attwooll, C. *et al.* (2005) A novel repressive E2F6 complex containing the Polycomb group protein, EPC1, that interacts with EZH2 in a proliferation-specific manner. *J. Biol. Chem.* 280, 1199–1208
66. Ogawa, H. *et al.* (2002) A complex with chromatin modifiers that occupies E2F- and Myc-responsive genes in G0 cells. *Science* 296, 1132–1136
67. Trimarchi, J.M. *et al.* (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
68. Herranz, N. *et al.* (2008) Polycomb complex 2 is required for E-cadherin repression by the Snail1 transcription factor. *Mol. Cell Biol.* 28, 4772–4781
69. Londhe, P. and Davie, J.K. (2013) Interferon-gamma resets muscle cell fate by stimulating the sequential recruitment of JARID2 and PRC2 to promoters to repress myogenesis. *Sci. Signal* 6, ra107
70. Ku, M. *et al.* (2008) Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet.* 4, e1000242
71. Vella, P. *et al.* (2012) Yin Yang 1 extends the Myc-related transcription factors network in embryonic stem cells. *Nucleic Acids Res.* 40, 3403–3418
72. Matharu, N.K. *et al.* (2010) Vertebrate homologue of *Drosophila* GAGA factor. *J. Mol. Biol.* 400, 434–447
73. Bauer, M. *et al.* (2015) The quest for mammalian Polycomb response elements: are we there yet? *Chromosoma*
74. Brockdorff, N. (2013) Noncoding RNA and Polycomb recruitment. *RNA* 19, 429–442
75. Pengelly, A.R. *et al.* (2013) A histone mutant reproduces the phenotype caused by loss of histone-modifying factor Polycomb. *Science* 339, 698–699
76. Chopra, V.S. *et al.* (2011) The Polycomb group mutant esc leads to augmented levels of paused Pol II in the *Drosophila* embryo. *Mol. Cell* 42, 837–844
77. Nakagawa, T. *et al.* (2008) Deubiquitylation of histone H2A activates transcriptional initiation via trans-histone cross-talk with H3K4 di- and trimethylation. *Genes Dev.* 22, 37–49
78. Zhou, W. *et al.* (2008) Histone H2A monoubiquitination represses transcription by inhibiting RNA polymerase II transcriptional elongation. *Mol. Cell* 29, 69–80
79. Stock, J.K. *et al.* (2007) Ring1-mediated ubiquitination of H2A restrains poised RNA polymerase II at bivalent genes in mouse ES cells. *Nat. Cell Biol.* 9, 1428–1435
80. Pengelly, A.R. *et al.* (2015) Transcriptional repression by PRC1 in the absence of H2A monoubiquitylation. *Genes Dev.* 29, 1487–1492
81. Tie, F. *et al.* (2016) Polycomb inhibits histone acetylation by CBP by binding directly to its catalytic domain. *Proc. Natl. Acad Sci. U. S.A.* 113, E744–E753
82. Francis, N.J. *et al.* (2004) Chromatin compaction by a Polycomb group protein complex. *Science* 306, 1574–1577
83. Eskeland, R. *et al.* (2010) Histone acetylation and the maintenance of chromatin compaction by Polycomb repressive complexes. *Cold Spring Harb Symp. Quant. Biol.* 75, 71–78
84. Yuan, W. *et al.* (2012) Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* 337, 971–975
85. Ebert, A. *et al.* (2004) Su(var) genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. *Genes Dev.* 18, 2973–2983
86. Ferrari, K.J. *et al.* (2014) Polycomb-dependent H3K27me1 and H3K27me2 regulate active transcription and enhancer fidelity. *Mol. Cell* 53, 49–62
87. Jung, H.R. *et al.* (2010) Quantitative mass spectrometry of histones H3.2 and H3.3 in Suz12-deficient mouse embryonic stem cells reveals distinct, dynamic post-translational modifications at Lys-27 and Lys-36. *Mol. Cell Proteomics* 9, 838–850
88. Peters, A.H. *et al.* (2003) Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* 12, 1577–1589
89. Voigt, P. *et al.* (2012) Asymmetrically modified nucleosomes. *Cell* 151, 181–193
90. Lee, H.G. *et al.* (2015) Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res.* 25, 1170–1181
91. Lanzuolo, C. *et al.* (2007) Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* 9, 1167–1174
92. Comet, I. *et al.* (2011) A chromatin insulator driving three-dimensional Polycomb Response Element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2294–2299
93. Gonzalez, I. *et al.* (2014) Identification of regulators of the three-dimensional Polycomb organization by a microscopy-based genome-wide RNAi screen. *Mol. Cell* 54, 485–499
94. Hernandez-Munoz, I. *et al.* (2005) Association of BMI1 with Polycomb bodies is dynamic and requires PRC2/EZH2 and the maintenance DNA methyltransferase DNMT1. *Mol. Cell Biol.* 25, 11047–11058
95. Cheutin, T. and Cavalli, G. (2014) Polycomb silencing: from linear chromatin domains to 3D chromosome folding. *Curr. Opin. Genet. Dev.* 25C, 30–37
96. Tiwari, V.K. *et al.* (2008) PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.* 6, 2911–2927
97. Bantignies, F. *et al.* (2011) Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* 144, 214–226
98. Schuettengruber, B. and Cavalli, G. (2013) Polycomb domain formation depends on short and long distance regulatory cues. *PLoS ONE* 8, e56531
99. Bantignies, F. *et al.* (2003) Inheritance of Polycomb-dependent chromosomal interactions in *Drosophila*. *Genes Dev.* 17, 2406–2420
100. Li, H.B. *et al.* (2013) Insulators target active genes to transcription factories and Polycomb-repressed genes to Polycomb bodies. *PLoS Genet.* 9, e1003436
101. Denholtz, M. *et al.* (2013) Long-range chromatin contacts in embryonic stem cells reveal a role for pluripotency factors and Polycomb proteins in genome organization. *Cell Stem Cell* 13, 602–616
102. Vieux-Rochas, M. *et al.* (2015) Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain. *Proc. Natl. Acad Sci. U.S.A.* 112, 4672–4677
103. Joshi, O. *et al.* (2015) Dynamic reorganization of extremely long-range promoter–promoter interactions between two states of pluripotency. *Cell Stem Cell* 17, 748–757
104. Schoenfelder, S. *et al.* (2015) Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome. *Nat. Genet.* 47, 1179–1186

105. Kondo, T. *et al.* (2016) Polycomb in transcriptional phase transition of developmental genes. *Trends Biochem. Sci.* 41, 9–19

106. Sanborn, A.L. *et al.* (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad Sci. U.S.A.* 112, E6456–E6465

107. Dixon, J.R. *et al.* (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380

108. Vietri Rudan, M. *et al.* (2015) Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell Rep.* 10, 1297–1309

109. Dowen, J.M. *et al.* (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387

110. Sexton, T. *et al.* (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472

111. Narendra, V. *et al.* (2015) CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. *Science* 347, 1017–1021

112. Noordermeer, D. *et al.* (2014) Temporal dynamics and developmental memory of 3D chromatin architecture at Hox gene loci. *Elife* 3, e02557

113. Noordermeer, D. *et al.* (2011) The dynamic architecture of Hox gene clusters. *Science* 334, 222–225

114. Nora, E.P. *et al.* (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385

115. Wijchers, P.J. *et al.* (2016) Cause and consequence of tethering a subTAD to different nuclear compartments. *Mol. Cell* 61, 461–473

116. Isono, K. *et al.* (2013) SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* 26, 565–577

117. Wani, A.H. *et al.* (2016) Chromatin topology is coupled to Polycomb group protein subnuclear organization. *Nat. Commun.* 7, 10291

118. Gambetta, M.C. and Muller, J. (2014) O-GlcNAcylation prevents aggregation of the Polycomb group repressor Polyhomeotic. *Dev. Cell* 31, 629–639

119. Boettiger, A.N. *et al.* (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529, 418–422

120. Follmer, N.E. *et al.* (2012) A Polycomb group protein is retained at specific sites on chromatin in mitosis. *PLoS Genet.* 8, e1003135

121. Fiedler, T. and Rehmsmeier, M. (2006) jPREdictor: a versatile tool for the prediction of cis-regulatory elements. *Nucleic Acids Res.* 34, W546–W550

122. Ringrose, L. *et al.* (2003) Genome-wide prediction of Polycomb/Trithorax response elements in *Drosophila* melanogaster. *Dev. Cell* 5, 759–771

123. Schuettengruber, B. *et al.* (2009) Functional anatomy of Polycomb and Trithorax chromatin landscapes in *Drosophila* embryos. *PLoS Biol.* 7, e13

124. Brookes, E. *et al.* (2012) Polycomb associates genome-wide with a specific RNA polymerase II variant, and regulates metabolic genes in ESCs. *Cell Stem Cell* 10, 157–170

125. Frangini, A. *et al.* (2013) The Aurora B kinase and the Polycomb protein ring1B combine to regulate active promoters in quiescent lymphocytes. *Mol. Cell* 51, 647–661

126. Kaneko, S. *et al.* (2013) PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1258–1264

127. Mousavi, K. *et al.* (2012) Polycomb protein Ezh1 promotes RNA polymerase II elongation. *Mol. Cell* 45, 255–262

128. van den Boom, V. *et al.* (2016) Non-canonical PRC1.1 targets active genes independent of H3K27me3 and is essential for leukemogenesis. *Cell Rep.* 14, 332–346

129. Creppe, C. *et al.* (2014) A Cbx8-containing Polycomb complex facilitates the transition to gene activation during ES cell differentiation. *PLoS Genet.* 10, e1004851

130. Morey, L. *et al.* (2015) Polycomb regulates mesoderm cell fate-specification in embryonic stem cells through activation and repression mechanisms. *Cell Stem Cell* 17, 300–315

131. Gao, Z. *et al.* (2014) An AUTS2-Polycomb complex activates gene expression in the CNS. *Nature* 516, 349–354

132. Lv, X. *et al.* (2016) A positive role for Polycomb in transcriptional regulation via H4K20me1. *Cell Res.* 2016

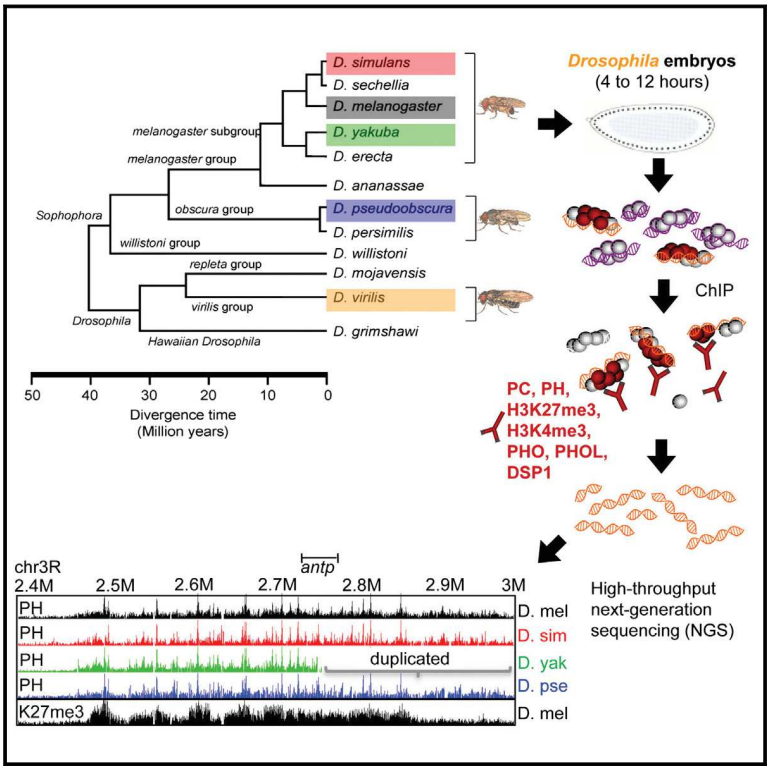
133. Arora, M. *et al.* (2016) RING1A and BMI1 bookmark active genes via ubiquitination of chromatin-associated proteins. *Nucleic Acids Res.* 44, 2136–2144

Article

Cell Reports

Cooperativity, Specificity, and Evolutionary Stability of Polycomb Targeting in *Drosophila*

Graphical Abstract



Highlights

Extremely high conservation of Polycomb repressive domains in five *Drosophila* species

cis-driven Polycomb response element (PRE) evolution is rare but possible

PHO recruitment to Polycomb domains is evolutionarily robust to motif changes

PRC1 stabilizes binding of PHO to PREs with low-quality motifs

Authors

Bernd Schuettengruber, Noa Oded Elkayam, ..., Amos Tanay, Giacomo Cavalli

Correspondence

amos.tanay@weizmann.ac.il (A.T.),
giacomo.cavalli@igh.cnrs.fr (G.C.)

In Brief

Schuettengruber et al. present an extensive comparative epigenomics data set, providing new insights into *cis*-driven versus buffered evolution of Polycomb recruitment and Polycomb domain specificity. Using chromatin immunoprecipitation sequencing and transgenic assays, they demonstrate an extremely high conservation of Polycomb repressive domains in five *Drosophila* species. Using Hi-C and knockout experiments, they challenge the standard hierarchical Polycomb recruitment model and demonstrate that cooperative rather than hierarchical interactions among DNA motifs, transcription factors, and Polycomb group complexes define Polycomb domains.

Accession Numbers

GSE60428
GSE61471

Schuettengruber et al., 2014, Cell Reports 9, 219–233
October 9, 2014 ©2014 The Authors
http://dx.doi.org/10.1016/j.celrep.2014.08.072

Cell Reports
Article

OPEN ACCESS
CellPress

Cooperativity, Specificity, and Evolutionary Stability of Polycomb Targeting in *Drosophila*

Bernd Schuettengruber,[1,4] Noa Oded Elkayam,[2,4] Tom Sexton,[1,5] Marianne Entrevan,[1] Shani Stern,[2] Aubin Thomas,[1] Eitan Yaffe,[2] Hugues Parrinello,[3] Amos Tanay,[2,*] and Giacomo Cavalli[1,*]
[1]Institute of Human Genetics, UPR 1142, CNRS, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France
[2]Department of Computer Science and Applied Mathematics and Department of Biological Regulation, The Weizmann Institute of Science, Rehovot 7610001, Israel
[3]Montpellier GenomiX IBiSA, 141 rue de la Cardonille, 34396 Montpellier Cedex 5, France
[4]Co-first author
[5]Present address: Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/UDS, 67400 Illkirch, France
*Correspondence: amos.tanay@weizmann.ac.il (A.T.), giacomo.cavalli@igh.cnrs.fr (G.C.)
http://dx.doi.org/10.1016/j.celrep.2014.08.072
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/3.0/).

SUMMARY

Metazoan genomes are partitioned into modular chromosomal domains containing active or repressive chromatin. In flies, Polycomb group (PcG) response elements (PREs) recruit PHO and other DNA-binding factors and act as nucleation sites for the formation of Polycomb repressive domains. The sequence specificity of PREs is not well understood. Here, we use comparative epigenomics and transgenic assays to show that *Drosophila* domain organization and PRE specification are evolutionarily conserved despite significant *cis*-element divergence within Polycomb domains, whereas *cis*-element evolution is strongly correlated with transcription factor binding divergence outside of Polycomb domains. Cooperative interactions of PcG complexes and their recruiting factor PHO stabilize PHO recruitment to low-specificity sequences. Consistently, PHO recruitment to sites within Polycomb domains is stabilized by PRC1. These data suggest that cooperative rather than hierarchical interactions among low-affinity sequences, DNA-binding factors, and the Polycomb machinery are giving rise to specific and strongly conserved 3D structures in *Drosophila*.

INTRODUCTION

The regulation of complex genomes in multicellular organisms requires both flexibility and stability. Genomes must be flexible enough to accommodate multiple cell-type-specific transcriptional programs. Simultaneously, genome regulation must be sufficiently stable to avoid aberrant gene activation in committed or differentiated cells. Genomes have adapted to this challenge by evolving a sparse dispersion of genes within vast genomic territories, which are dotted with hundreds to thousands of small regulatory elements. Importantly, the local sequence specificity of metazoan regulatory elements within such vast territories is not higher than that observed in much more compact genomes, leading to the spontaneous evolution of millions of possible spurious binding sites for a typical sequence-specific transcription factor (TF). Nevertheless, extensive mapping of binding sites for hundreds of DNA-binding factors has shown that only many thousands of enhancer elements, rather than millions of potential spurious binding sequences, are specifically identified and engaged by combinations of TFs and chromatin regulators. It was suggested that such specificity is facilitated by cooperative binding of TFs (Junion et al., 2012) and by epigenetic mechanisms that selectively provide access to a small subset of the genome. The specificity of epigenetic regulation itself, however, remains poorly understood. For example, simple hierarchical models postulating that sequence-specific "pioneer" factors dictate specificity in genome regulation during cell-fate commitment cannot explain much of the experimental data (Rothenberg, 2014). Recently, the discovery of topologically associating domains (TADs) in flies and mammals showed that chromosomes can compartmentalize genomes into relatively isolated building blocks (Dixon et al., 2012; Nora et al., 2012; Sexton et al., 2012), but how such compartmentalization affects regulatory specificity remains unclear.

Repressive Polycomb topological domains, characterized by the presence of the H3K27me3 mark, constitute a major subdivision of the eukaryotic genome and provide a paradigm for understanding epigenome regulation and chromosomal domain structure. Initial genetic studies in flies characterized the Polycomb group (PcG) system as being responsible for maintaining *HOX* gene repression following its initial setup during early embryonic development (Duncan, 1982; Lewis, 1985). Subsequent work demonstrated a more dynamic role of PcG proteins in defining cellular identities through the epigenetic repression of key developmental regulators (reviewed in Schuettengruber and Cavalli, 2009). Genomic analysis generated a multilayered view of PcG-mediated gene regulation involving a combination of Polycomb repressive complex 1 (PRC1) and 2 (PRC2) their associated histone marks, HA2 ubiquitylation of lysine 119 (H2AK119ub) or trimethylation of lysine 27 on histone H3 (H3K27me3), and chromosomal looping (Bantignies and Cavalli, 2011). In *Drosophila*, specific sequences termed Polycomb group response elements (PREs) are known to recruit PcG complexes to their target genes

A

Species	All Genome				H3K27me3 Sites			
	% of D. xxx not aligned to D. mel	% of D. mel not aligned to D. xxx	Not Mappable (within aligned)	Non-Exon Point Mutation (D. xxx - D.mel)	% of D. xxx not aligned to D. mel	% of D. mel not aligned to D. xxx	Not Mappable (within aligned)	Non-Exon Point Mutation (D. xxx - D.mel)
D. mel	NA	NA	10.70%	NA	NA	NA	NA	NA
D. sim	13.06%	11.02%	3.33%	9%	3.80%	0.01%	1.61%	5%
D. yak	10.27%	16.82%	7.45%	17%	0.92%	4.05%	8.11%	11%
D. pse	30.98%	29.74%	1.81%	45%	8.80%	16.35%	2.07%	36%
D. vir	32.66%	53.90%	4.22%	66%	18.62%	25.72%	0.30%	56%

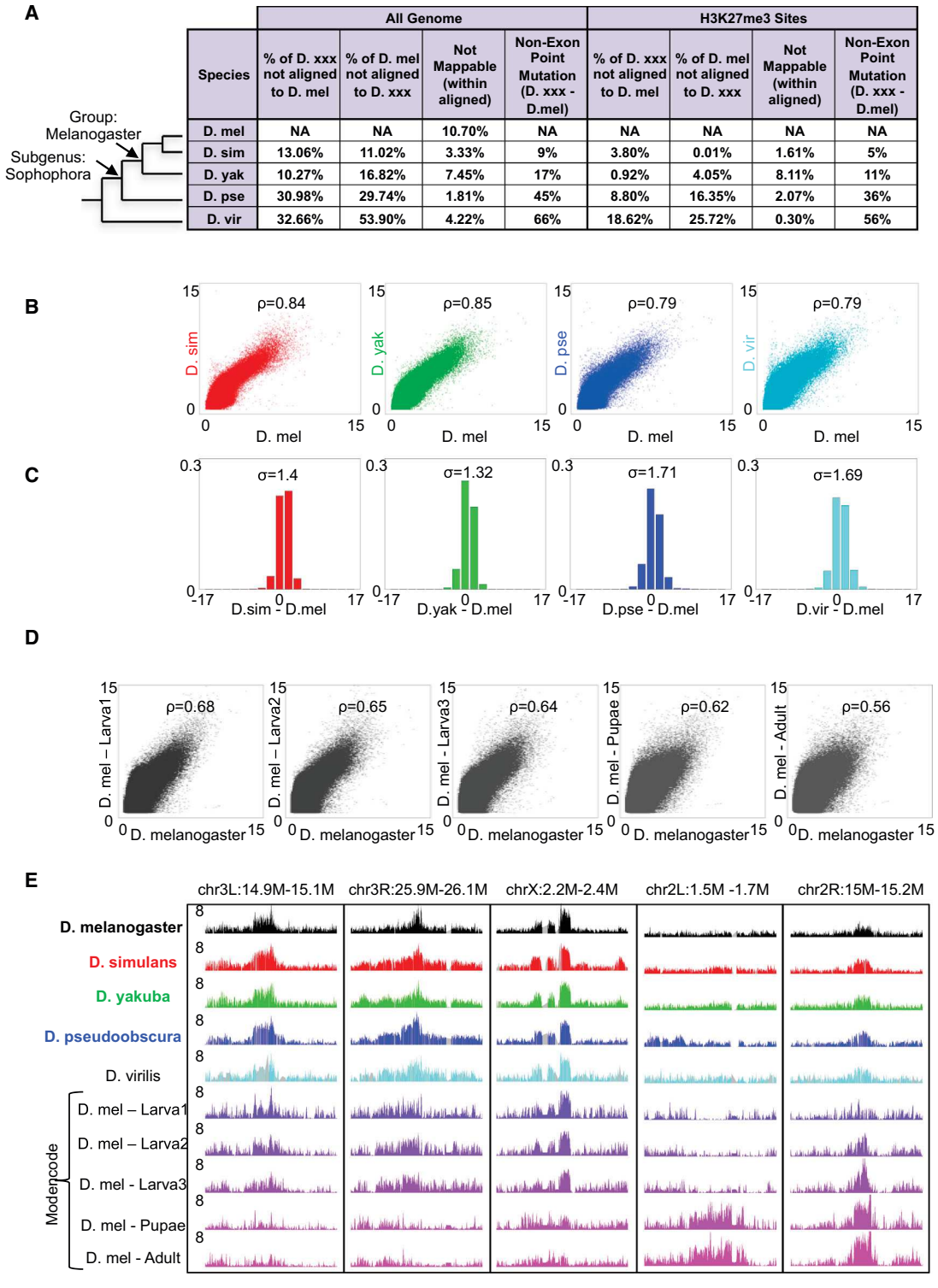


Figure 1. Comparative Epigenomics of Polycomb Domains

(A) Phylogenetic statistics of the *Drosophila* species used in this study. Data on divergence are provided with respect to the entire genome or regions that are annotated as H3K27me3-marked domains in D.mel.

(B) Scatterplots depicting the correlation between H3K27me3 ChIP-seq enrichment in syntenic loci for the indicated *Drosophila* species.

(legend continued on next page)

via their interaction with sequence-specific DNA-binding proteins defined as PcG recruiters (Müller and Kassis, 2006). Wang et al. (2004) proposed a recruitment model that suggests a hierarchy of binding events to PREs: first, sequence-specific binding of DNA-binding factors lead to the subsequent targeting of PRC2, which then helps to recruit the later-acting PRC1 complex via its interaction with the PRC2-specific histone mark H3K27me3. However, in mammals it was recently demonstrated that PRC1 variant complexes can be recruited to DNA independently of PRC2, but in turn the PRC1-associated mark H2AK119ub helps to recruit PRC2 (Blackledge et al., 2014; Cooper et al., 2014). The sequence requirements for targeting of mammalian PRC complexes are still unclear, but CpG islands seem to play a major role.

Isolation and perturbation analyses of *Drosophila* PREs have uncovered several sequence motifs that are required for the assembly of the PcG machinery on reporter constructs (Brown et al., 1998; Déjardin et al., 2005). Among these, the binding sites of Pleiohomeotic (PHO) and Dorsal Switch Protein 1 (DSP1) were particularly enriched in genomic catalogs of putative PREs (Schuettengruber et al., 2009). However, none of the motifs or their associated DNA-binding proteins are sufficient to recruit PcG proteins to their targets (Müller and Kassis, 2006). Whereas mutation of PHO induces homeotic phenotypes similar to those observed in loss-of-function PcG mutants, mutations of DSP1 exhibit a variety of phenotypes—some similar to PcG mutants and others more typical of mutants for a Trithorax group gene, a factor known to counteract PcG function. Hence, PcG-recruiting factors are associated with transcriptional activation as well as repression (Fujioka et al., 2008; Kwong et al., 2008; Schuettengruber et al., 2009). PHO, in particular, binds numerous putative promoter and enhancer elements outside a PRE context (Schuettengruber et al., 2009), and it was shown to interact with the INO80 chromatin-remodeling complex in addition to PcG complexes (Klymenko et al., 2006). Of note, the mammalian homolog of PHO, YY1, is able to rescue a *pho* zygotic mutant, but does not seem to play a major role in PcG recruitment in mammals (Atchison et al., 2003; Mendenhall et al., 2010). DSP1 was also shown to bind to many non-PRE loci in *Drosophila* embryos, which are marked strongly by GAGA motifs (Schuettengruber et al., 2009). Thus, despite the unambiguous genetic and genomic evidence for PREs, and for the involvement of PHO and DSP1 in their function, the specificity of the process remains difficult to understand based on studies of individual factors or loci. Thus far, any attempts to predict PREs from genomic sequence alone have only been partly successful (Fiedler and Rehmsmeier, 2006; Kassis and Brown, 2013; Ringrose et al., 2003; Schuettengruber et al., 2009; Zeng et al., 2012).

Here, we study the function and evolution of PRE sequences within the broader context of multigenic Polycomb topological domains (hereafter referred to as Polycomb domains). Comparative epigenomics shows that during the evolution of *Drosophila* species, and despite extensive sequence divergence, the structure of the Polycomb domain in syntenic genomic regions remained perfectly conserved. This remarkable stability was facilitated by the high conservation of PRC1 binding at putative PREs. In the relatively few cases in which PRC1 binding diverged, transgenic PRE assays indicate that *cis*-element divergence was the likely cause of functional divergence. In other cases, however, divergence in *cis* is not linked to functional PRE divergence in Polycomb domains. We explain this effect by showing that targeting of the PcG recruiter PHO to PREs within Polycomb domains can be driven by PHO-DNA interactions that occur over a wide spectrum of noncanonical, low-affinity binding sites. Such sites are hypothesized to buffer pronounced evolutionary divergence without significant loss of PRE function. Instead of high sequence affinity, our data indicate that PHO recruitment relies on cooperative effects with other TFs (e.g., DSP1) and on a positive feedback loop induced by PRC1 binding. Therefore, the sequence specificity of Polycomb domains in flies is established through cooperative rather than hierarchical interactions between sequence-specific factors and the Polycomb machinery, in a way that potentially also involves the formation of 3D chromatin hubs associating with several PREs. A similar interplay among sequence-specific genomic signals, epigenetic factors, and large-scale chromosomal structure may have general implications for regulatory genomics and epigenomics in metazoans.

RESULTS

Evolutionary Conservation of H3K27me3 Repressive Domains in *Drosophila* Embryos

To characterize the evolution of Polycomb domains in *Drosophila*, we performed chromatin immunoprecipitation sequencing (ChIP-seq) for the H3K27me3 mark on 4- to 12-hr-old embryos from *D. melanogaster* (D.mel), *D. simulans* (D.sim), *D. yakuba* (D.yak), *D. pseudoobscura* (D.pse), and *D. virilis* (D.vir) (see Supplemental Experimental Procedures). These species represent variable levels of evolutionary distance from D.mel, with overall local identity on alignable sequences varying between 92% (D.sim) and 39% (D.vir) of the D.mel genome. We note that sequences marked by H3K27me3 are evolving more slowly than other genomic regions, as indicated by both the point mutation level and the overall retained aligned sequences (Figure 1A). Nonetheless, the rate of sequence divergence is substantial and in principle could support changes in the definition of genomic Polycomb domains. Surprisingly, however, a comparative analysis of the species ChIP-seq that was projected onto the D.mel genome indicated a remarkable pointwise conservation of H3K27me3 occupancies (Figure 1B). The overall correlation between the orthologous profiles was higher than 0.79 even for the remote species D.vir. Moreover, the root-mean-square of pairwise differential ChIP-seq ranged between 1.32 and 1.7 (Figure 1C), maintaining levels comparable to those

(C) Summary of H3K27me3 divergence in each species compared with D.mel.
(D) Comparison of H3K27me3 enrichment levels in D.mel embryos and other developmental stages.
(E) Examples of evolutionarily conserved H3K27me3 domains that are developmentally plastic. The y axis represents normalized ChIP-seq values.
See also Figure S1.

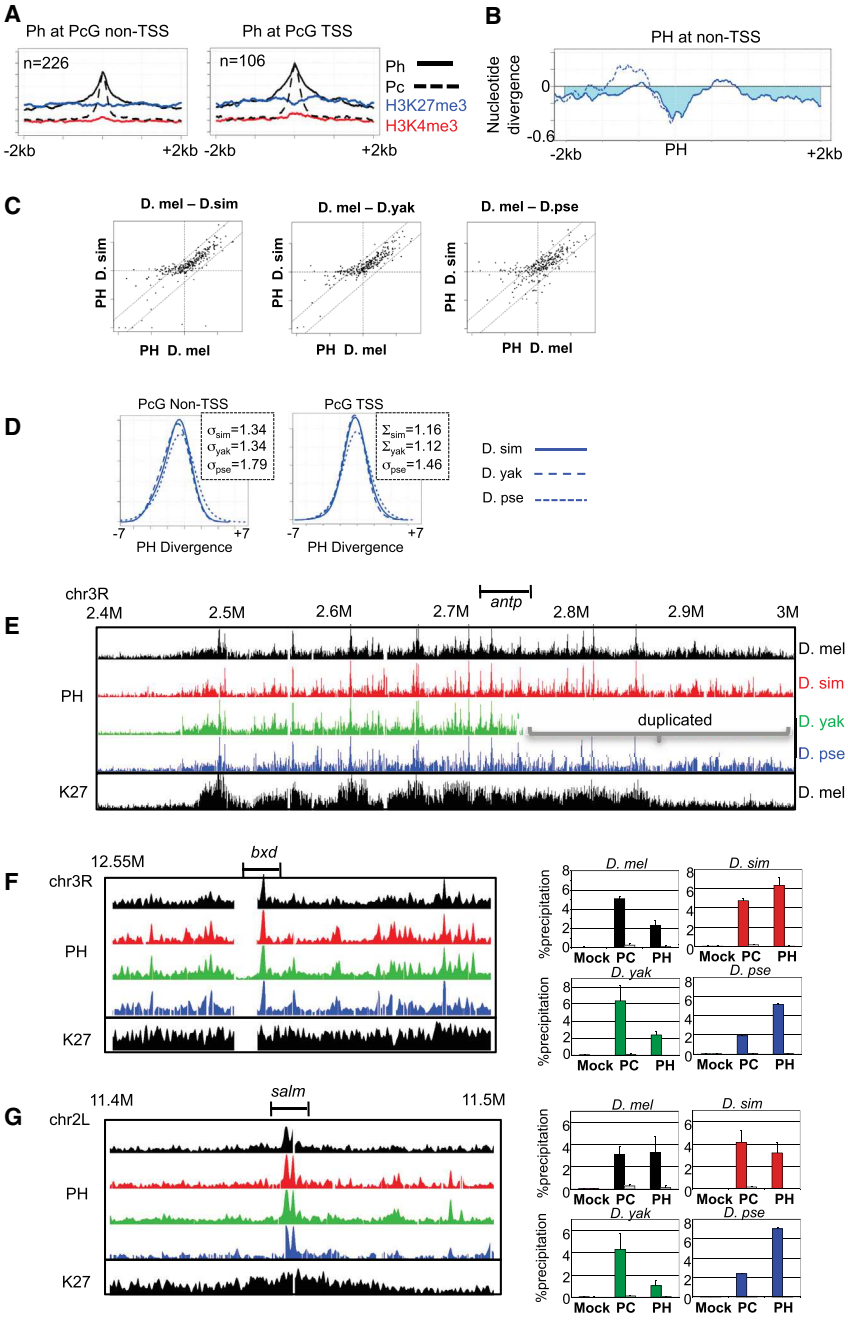


Figure 2. High Conservation of PRC1-Binding Sites

(A) Spatial distributions of D.mel ChIP-seq data for PH, PC, H3K27me3, and H3K4me3, pooling statistics around identified PH sites that are classified according to their proximity to a TSS (TSS: <500 bp from a known gene start site; non-TSS: >500 bp distant from TSSs).

(B) Spatial distributions of divergence statistics (log of ratios between observed and expected nucleotide substitutions) around PH sites within Polycomb domains (defined by Hi-C data). Sites in a non-TSS context (N, >500 bp from a TSS) are shown by solid lines, and sites in a TSS context are shown by dashed lines. All sites are oriented according to the strand of the nearest TSS.

(C) Genome-wide comparative analysis of PH ChIP-seq enrichment, showing pairwise data for all PH sites that were mapped on a syntenic, mappable locus.

(D) Summary of the divergence in PH ChIP-seq enrichment between D.mel and three other species, stratified according to TSS distance. The reported σ values are computed as the SD of the differential ChIP-seq data over all PH sites.

(E) Comparative ChIP-seq for PH in four *Drosophila* species, reflecting conservation of the epigenetic structure in the *Antp* complex. A region in D.yak that is duplicated (and therefore is not uniquely identifiable in ChIP-seq) is marked.

(F and G) Comparative ChIP-seq (left) and validation data by qChIP (right) for the conserved bxd and salm loci. Error bars represent the SD of the means of three independent experiments.

See also Figure S2 and Tables S1 and S2.

PRC1-Binding Sites Are Conserved amidst Dynamic Sequence Evolution

We expanded our ChIP-seq data set to study the evolution of PRC1-binding sites (via its core components Polyhomeotic [PH] and Polycomb [PC]). We also generated H3K4me3 ChIP-seq profiles to map transcription start sites (TSSs) in four *Drosophila* species. Using our previously constructed *Drosophila* Hi-C maps (Sexton et al., 2012), we identified Polycomb topological domains and studied the different localization of PRC1, H3K27me3, and H3K4me3 distributions around PH peaks within these Polycomb domains, classing sites according to their proximity to a TSS (defining TSS versus non-TSS sites; Figure 2A). There is some sequence conservation (mean 30% decrease in convergence rate) in the −200 to +200 bp range around non-TSS PRC1 sites, with accelerated evolution in the 400–800 bp around the site (Figure 2B), consistent with the broad evolutionary patterns observed around *Drosophila* enhancers (Kenigsberg and Tanay, 2013). Despite this relatively mild sequence conservation, a comparative PH ChIP-seq analysis (Figures S2A and S2B) indicated that observed between replicate experiments (Figures S1A and S1B). The extreme evolutionary conservation of H3K27me3 domains is nontrivial, since H3K27me3 domains are developmentally plastic (Nègre et al., 2011; Figures 1D, 1E, and S1C). Thus, although reprogramming of Polycomb organization is not observed evolutionarily, it occurs throughout *Drosophila*'s life cycle. In conclusion, nonduplicated and nondeleted Polycomb domains in the embryonic stage conserved their H3K27me3 association to near completeness despite substantial underlying sequence divergence.

PRC1-binding sites are very highly conserved and rarely diverge at all. Systematically, we estimated a high degree of conservation of PH-binding levels in all species, with conservation slightly higher in TSSs (Figures 2C and 2D). These data indicated that PRC1 occupancy tolerates significant *cis*-element divergence, but also provided us with specific cases of divergence for analysis at higher resolution.

Linking *cis*-Element and PRC1 Occupancy Divergence

When we screened the ChIP-seq profiles, we identified 379 sites within PcG domains that are conserved across the *Drosophila* species (Figures 2E–2G and S3; Table S1) and 32 potentially diverged elements (Figure S2C). For example, a putative diverged PRC1-binding site, located within the Antp locus, shows increased PH-binding levels in D.pse that correlates with the gain of one strong GAGA repeat and two PHO (GCCATTT) boxes (Figure 3A). Both of these motifs were previously suggested to be important for PRC1 recruitment in D.mel (Schuettengruber et al., 2009). In another case, an element within the *Sox21b* region loses PRC1 binding in D.pse concomitantly with divergence of the GAGA repeat sequence, but not of a PHO box (Figure 3B). To validate that these elements are indeed functionally divergent due to local sequence perturbation, we generated transgenic reporter D.mel flies carrying either the D.mel sequence or the orthologous D.pse sequence upstream of a mini-white reporter gene. As controls, fly lines carrying the empty vector, a promoter region (*Zif*) that is not associated with PcG proteins at any developmental stage, or the well-characterized D.mel bxd PRE were generated. The reporter constructs were integrated at the same genomic position to avoid position effects, and PRE activity was determined by analyzing repression of the reporter gene using the eye color as a readout (Figures 3C–3H) or by determining the ability of the transgene to recruit the repressive histone mark H3K27me3 by quantitative ChIP (qChIP) assays (Figures 3I–3K). For the analyzed *Antp* locus (3R282), the D.mel sequence did not show any PRE characteristics (i.e., no significant repression of the reporter gene compared with the control "vector only," and no association of the transgene with repressive histone marks), whereas the orthologous sequence derived from D.pse significantly repressed the reporter gene in a dosage-dependent manner (pairing-sensitive silencing [PSS]) and was associated with increased levels of the H3K27me3 mark (Figures 3C, 3F, and 3I). In contrast, the putative diverged PRE sequences from the *Sox21b* gene locus (3L141) only showed functional PRE features when derived from D.mel (Figures 3E, 3H, and 3K). In summary, our transgenic analysis shows that the divergence of specific sequence elements (PHO and GAGA) could underlie *cis*-driven divergence of PRC1 binding in a few cases.

Extension of the comparative sequence analysis to 12 *Drosophila* species suggested that an association between GAGA/PHO motifs and the orthologous PREs defined by the conserved recruitment of PRC1 in four species is more universally conserved (Figures S4A and S4B). In fact, PHO and GAGA motifs are somewhat more conserved in the context of these orthologous PRE sites than in the rest of the genome (Figures S4C and S4D), even when stratifying for regional sequence conservation (Figure S4E). However, this degree of conservation is far from sufficient to predict the very highly conserved recruitment of PRC1 alone (Figure S4F). A possible explanation for this discrepancy is that some buffering mechanism contributes to the stabilization of PRC1 recruitment in a way that can tolerate significant divergence of local sequence elements.

Combinatorial PHO and DSP1 Occupancy Marks PRC1 Sites

To explore the relationship between PHO/DSP1 binding conservation and PRC1 recruitment conservation in more detail, we performed ultradeep ChIP-seq for the PcG recruiter factors (Figure S5A), focusing first on the D.mel genome. We compared the sites for one or both factors with PRC1 recruitment and H3K27me3 labeling, classing the sites based on whether or not they are present at TSSs and/or within Polycomb domains. Figure 4A shows that joint PHO/DSP1 sites within Polycomb domains co-occur perfectly with PRC1-binding sites (n = 159 and n = 103 for non-TSS and TSS contexts, respectively). PHO-only binding sites can also be observed at non-TSSs within Polycomb domains (n = 206), although in these cases PRC1 enrichment is observed at weaker levels. In contrast to these patterns, outside of Polycomb domains, we observe all combinations of PHO- and DSP1-binding sites (joint binding and PHO- or DSP1-only sites), at non-TSSs or TSSs. Taken together, these results show that PHO and DSP1 cobinding distributions are globally correlated with PRC1 occupancy, whereas PHO binding and DSP1 binding are less correlated with each other in other genomic contexts where PRC1 binding is generally lacking.

PHO Binds Weak *cis* Elements in a Polycomb-Domain-Dependent Fashion

A motif enrichment analysis in PHO- and DSP1-binding sites gave the expected binding sequences for the two factors. Further analysis revealed that combined PHO/DSP1-binding sites are characterized by strong GAGA motif enrichment (up to 85-fold higher than background level) but little or no PHO motif enrichment (Figures 4B and 4C). On the other hand, PHO sites lacking DSP1 enrichment are enriched for PHO motifs (45- to 60-fold over background levels), but not GAGA motifs, and DSP1-only binding sites are enriched for GAGA motifs, but not PHO motifs. To better understand the corecruitment of PHO with DSP1 in the absence of canonical PHO motifs, we derived ChIP-seq distributions at genomic sequences based on their fit to the PHO consensus (from a completely randomized sequence to a perfect consensus match), further breaking down the sequences to their presence at TSSs and the epigenetic identity (Polycomb, active, or null; Sexton et al., 2012) of their topological domains (Figures 4D and 4E). Strikingly, PHO recruitment to non-TSS sequences within Polycomb domains is effective even when the sequence is far from the optimal consensus. PHO recruitment to such low-specificity sites is far less efficient within null domains and is only efficient at active domains when in a TSS context. DSP1 recruitment to weak GAGA motifs is generally weaker outside of a TSS context, and is equally sensitive in active and Polycomb domains, with very poor recruitment in null domains. From these analyses, we conclude that sequence information is interpreted in a context-dependent manner to

OPEN ACCESS
CellPress

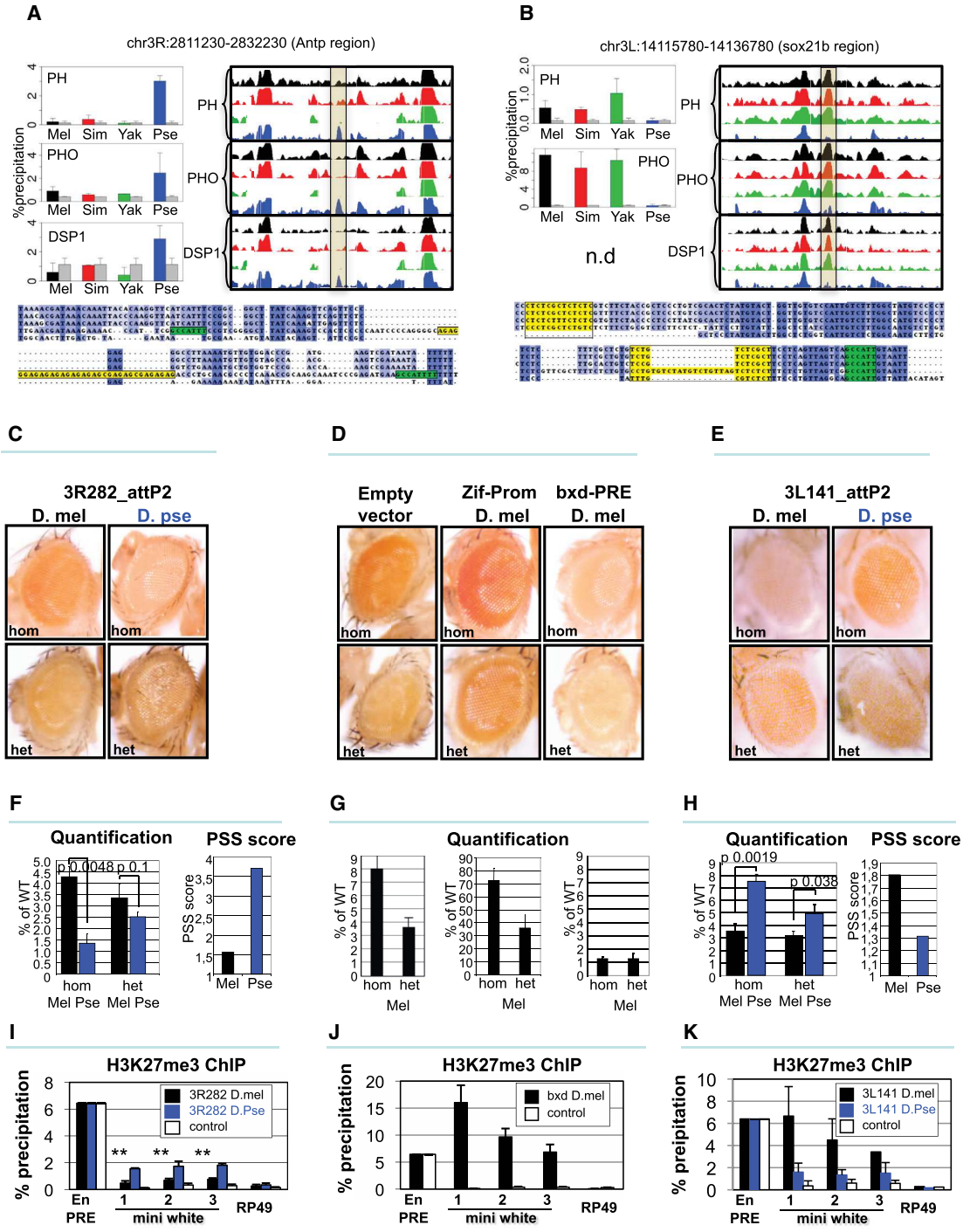


Figure 3. Scenarios for *cis*-Driven PRE Evolution

(A) qChIP (left), regional ChIP-seq profile (right), and annotated sequence alignment (bottom) for a region in the Antp locus in which a D.pse-specific PH site is detected. GAGA repeats are highlighted in yellow and PHO motifs are highlighted in green.

(B) Similar to (A), but showing data for PH binding that is specifically lost in D.pse at the *Sox21b* gene region compared with the other species.

(C–K) Results of transgenic reporter assays.

(C–E) Eye phenotype of transgenic fly lines. Four-day-old male flies, either heterozygous (bottom) or homozygous (top), are shown.

(F–H) Quantification of eye pigment in the heads of transgenic flies. Pigment levels are expressed as the percentage of WT pigment. Heads of 4-day-old male flies were used for each assay. The SD from three independent experiments is shown. The PSS score was calculated from mean pigment levels as heterozygote/homozygote for each diverged PRE region, normalized to the same ratio calculated for the "vector only" control line.

(legend continued on next page)

determine PHO or DSP1 binding, or their joint binding. Sequences within Polycomb domains, in particular, are capable of recruiting PHO through weak binding sites, possibly through cooperative mechanisms.

PHO Binding Evolutionary Divergence Is Correlated with *cis*-Element Divergence Only outside of Polycomb Domains

Analysis of overall sequence conservation around strong D.mel PHO and GAGA motifs (Figure 5A and 5B) shows that these motifs are moderately well conserved themselves (showing only a 30% decrease in divergence rate compared with the background). On the other hand, these motifs are typically located in a region of 200–400 bp, showing significant overall conservation. Regional conservation is observed, to some extent, even around motifs that lack PHO or DSP1 binding in embryos. Therefore, the evolutionary conservation of PHO-and DSP1-binding sites may involve not only the known *cis* elements analyzed above but also additional sequences that together define an element on a scale of a few hundred bases. We performed comparative ChIP-seq of PHO and DSP1 in four *Drosophila* species and showed that indeed, despite the limited conservation of the motifs, most of the factors' binding landscapes are well conserved (Figures 5C, 5D, and S5B–S5G), showing quantitatively higher stability than previously described for other factors that are not part of the Polycomb recruitment machinery (Figure S5H). Breaking down sites according to their context, we observed generally high PHO binding conservation ($\sigma = \sim 1.5$) with some preference for TSSs within Polycomb domains ($\sigma = 1.1$), and higher divergence of Dsp1 sites ($\sigma = \sim 2$), with more constrained evolution at TSSs within active domains ($\sigma = 1.6$). Importantly, the divergence in PHO binding within Polycomb domains is uncorrelated with the sequence divergence of PHO motifs (Figure 5E), whereas we observe a significant correlation between the two in active domains or null domains (spearman rho coefficient up to 0.35, $p << 10^{-10}$). On the other hand, the divergence of DSP1 binding is significantly correlated with GAGA motif divergence in all contexts (rho 0.19–0.42, $p << 10^{-10}$). Combined with the observations of low-affinity motif enrichment described above, our evolutionary analysis suggests that within Polycomb domains, PHO binding is buffered by cooperative factors and thus is capable of engaging low-affinity sites and being less sensitive to the evolutionary divergence of high-affinity binding sites. This buffering might contribute to a highly conserved evolutionary signature and in turn may underlie the conservation of H3K27me3 domains and PRC1 binding landscapes.

Recruiting the Recruiter: Predicting PHO Binding Intensity Given PHO Motifs and PRC1 Occupancy

According to the hierarchical PcG recruitment model (Wang et al., 2004), PHO should be a pioneer factor, binding its cognate DNA motifs specifically to promote the recruitment of PRC2 and PRC1. As noted above, however, sequence specificity alone cannot predict PHO and DSP1 binding accurately, so the high degree of correlation and predictability of PRC1 binding levels from the factors' binding profiles in all species (Figures S6A–S6C) reflects a strong, but not necessarily causative, correlation. Using classification of PHO-binding sites into Polycomb, active, and null domains, we observed a remarkably high correlation between the quantitative binding intensity of PH and PHO within the Polycomb domain context. A large range of PHO binding levels in the Polycomb domain context is also observed (Figure 6A). However, an analysis of the correlation between factor binding and sequence affinity (as predicted from the pho motif; Figure 6B) suggested that sequence-driven recruitment of PHO is strongly supported in active or null domains, but not in non-TSS contexts within Polycomb domains. To resolve this apparent contradiction, which goes against the hierarchical Polycomb recruitment model (Wang et al., 2004), we normalized PHO binding intensity by PH binding intensity to the same site, and recomputed the degree of sequence to binding correlation (Figure 6C). Modeling PHO as a function of PH binding and PHO motifs in this manner generated accurate quantitative predictions, explaining 73%–80% of the variance and significantly exceeding other hypothetical models that aim to predict PHO binding from combinations of its motifs and other factors (Figures S6D and S6E). The data therefore suggest that instead of a well-separated hierarchy of mechanisms, PRE specificity may involve a bidirectional interaction between recruiters and PRCs. An initial modest sequence specificity for PHO recruitment within Polycomb domains may be amplified through a PRC1-dependent positive feedback loop, making the ultimate PHO binding landscapes within Polycomb domains a complex function of sequence, DNA-binding factors, and PRC1 interactions.

Topological Clustering of PHO Sites within Polycomb Domains

In flies, Polycomb domains cluster in the nucleus to form specific compartments called Polycomb bodies (Cheutin and Cavalli, 2014). Within these nuclear compartments, PREs form specific spatial contacts (Lanzuolo et al., 2007; Lo Sardo et al., 2013), and it was suggested that PcG-mediated repression works in part by forming chromosomal loops that bring PREs into contact and antagonize transcriptional activation on repressed TSSs (Cheutin and Cavalli, 2014). A possible scenario for more robust binding of PHO within Polycomb domains is that looping brings PHO-binding sites into close spatial proximity, thereby increasing the local concentration of PHO as well as Polycomb proteins and facilitating PHO binding even at suboptimal DNA motifs. Indeed, analysis of the degree of PHO clustering and its conservation between D.mel and D.pse strongly suggested that the spatial organization of PHO-binding sites within Polycomb domains is more conserved compared with other genomic regions (Figure 6D). In order to better characterize the 3D environment of PHO-binding sites, we generated an

(I–K) qChIP assays performed on embryos (0–12 hr old) of the indicated transgenic fly lines using H3K27me3 antibodies. Primers 1 and 2 amplify the TSS region, whereas primer 3 amplifies the *white* coding region. Data were normalized to the positive control (*engrailed* PRE, "En"). As a negative control, the housekeeping gene *rp49* was used (**$p < 0.01$, *$p < 0.05$ as calculated from a two-tailed t test). Error bars represent the SD of the means of three independent experiments. See also Figure S4 and Table S2.

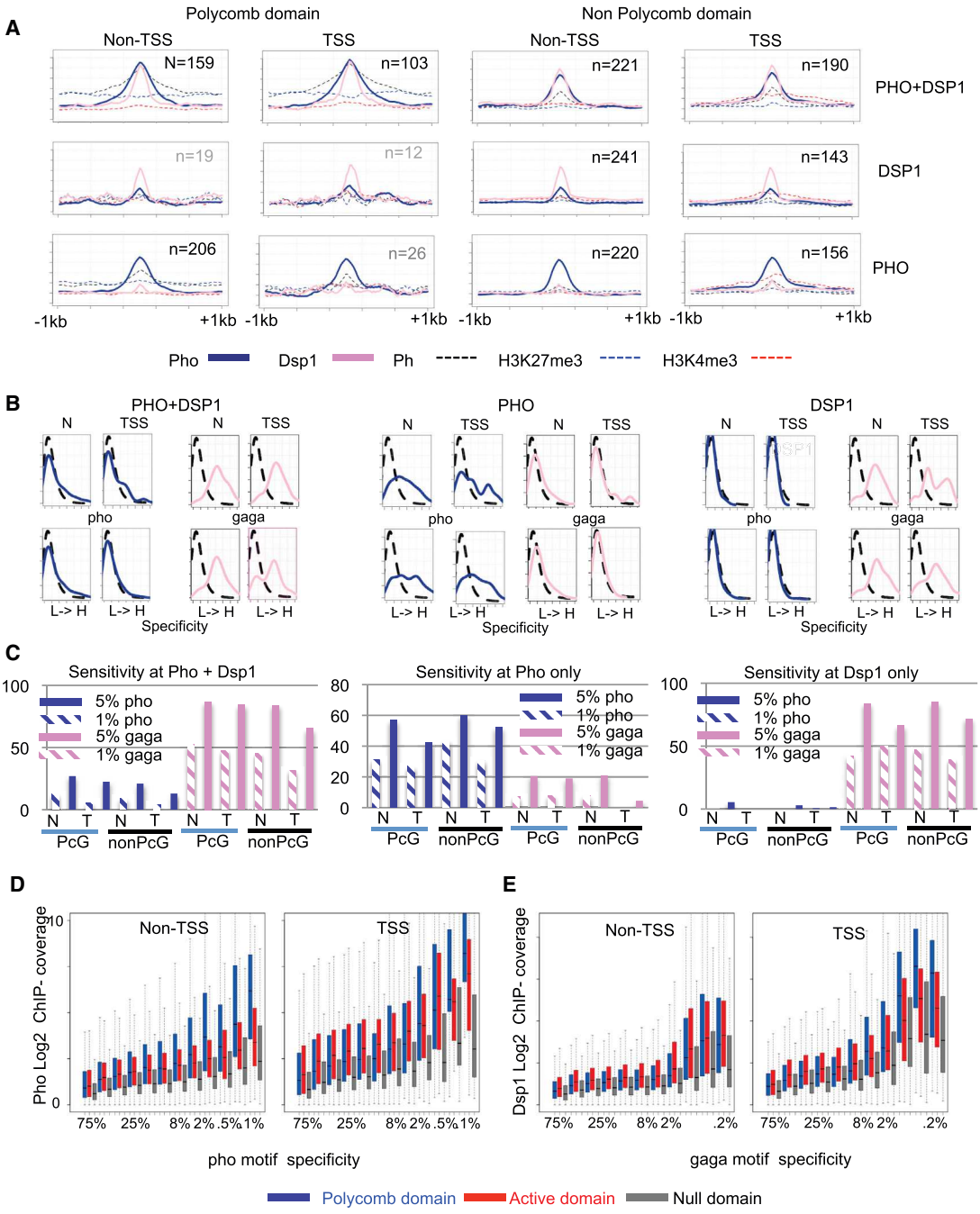


Figure 4. PcG Recruiter Sequence Specificity Is Dependent on Context

(A) Mean spatial enrichment statistics for PHO and DSP1 for groups of joint (top) and factor-specific (middle and bottom) sites. Sites are further stratified according to their association with a Hi-C Polycomb domain and their TSS proximity (TSS, within 500 bp of a TSS; N, >500 bp away). Data for H3K27me3 and H3K4me3, as well as for the PRC1 component PH, are added in dashed lines.

(B) Enrichment of sequence motifs is depicted for groups of joint and factor-specific sites as defined in (A). For each group, the distribution of motif scores is compared with the background distribution (black dashes) that is estimated from TSS-linked or unlinked sequences as appropriate.

(C) Summary of motif enrichment. Fold change is estimated from a comparison with the top first and fifth percentiles of the background distribution.

(D) Boxplot showing the genome-wide distribution of PHO ChIP-seq enrichments at sites with increasingly more specific motif scores (x axis, lower quartile on the left, upper 0.1 percentile on the right). The analysis is stratified to TSS and non-TSS contexts (right and left panels) and to the type of Hi-C domain in which each locus is contained (color-coded boxplots).

(E) Similar to (D), but for DSP1 ChIP-seq enrichments with respect to GAGA motif scores.

See also Figure S4.

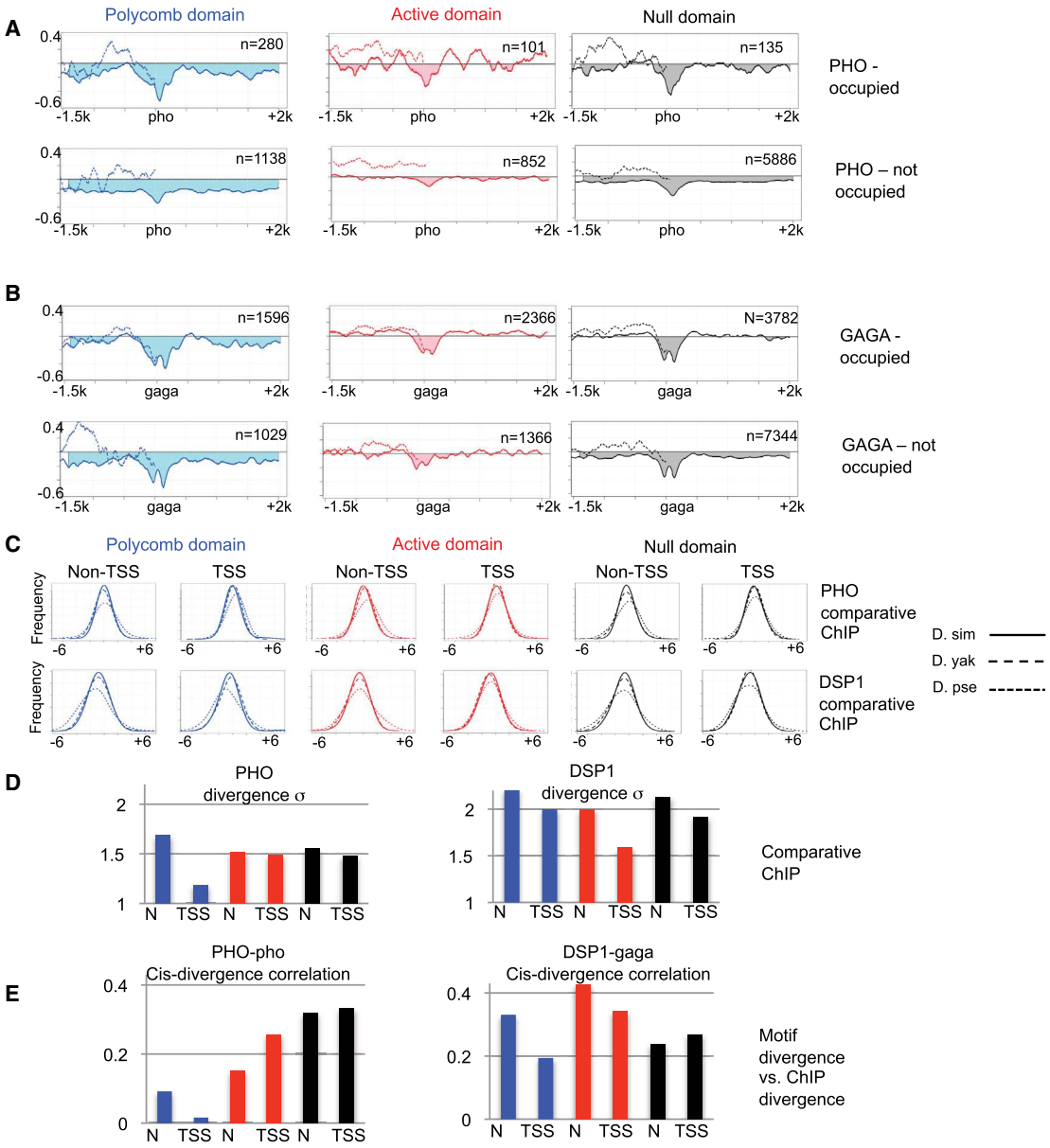


Figure 5. *cis*-Driven Evolution of PHO Is Buffered within Polycomb Domains

(A) Sequence divergence statistics around PHO motifs, estimated by the log of the ratio between the observed and expected numbers of substitutions in a 12-species *Drosophila* phylogeny. All occurrences of a PHO motif above some threshold were stratified according to their domain context (blue, Polycomb; red, active; black, null), their TSS proximity (>500 bp for a TSS [solid line], <500 bp from a TSS [dashed line in the left half of the graph]), and their PHO ChIP-seq occupancy.

(B) Similar to (A), but analyzing GAGA motifs and DSP1 sites.

(C) Divergence of PHO and DSP1 ChIP-seq on syntenic *Drosophila* sequences. Data for a pairwise comparison of D.mel and three other species are shown. For each pair, divergence is estimated on sites that are occupied by a factor in either of the species. Stratification according to domain type and TSS proximity was done as described above.

(D) Summary of divergence statistics shown in (C).

(E) Spearman correlation values testing the linkage between divergence in motif scores (following sequence divergence) and divergence in factor binding as estimated by comparative ChIP-seq. All values above 0.1 are statistically highly significant ($p << 10^{10}$).

See also Figure S5.

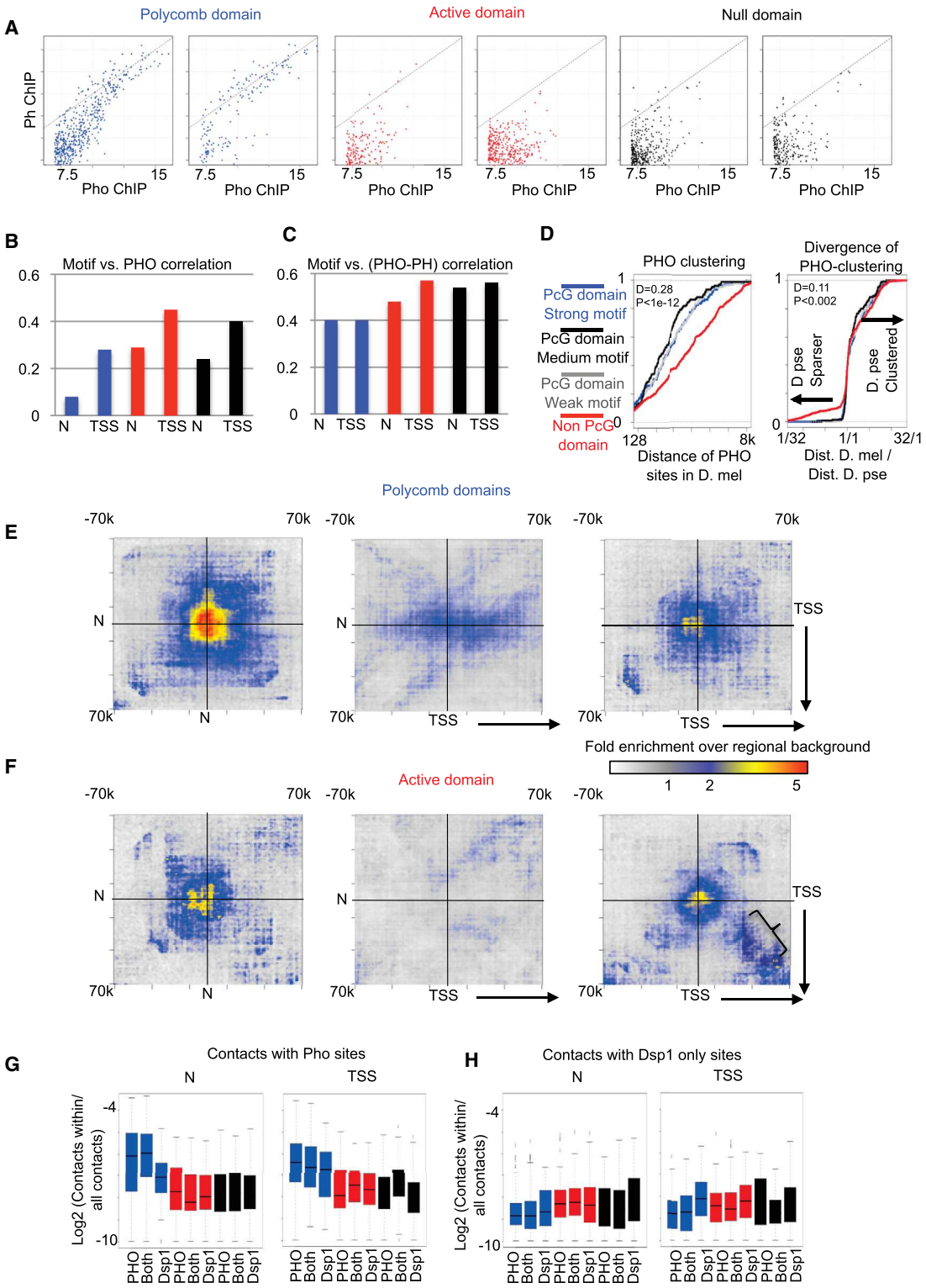


Figure 6. Potential Cooperative Factors underlying PHO Genomic Specificity

(A) Scatterplots comparing ChIP-seq enrichment values for PHO and PH in different epigenomic contexts.

(B) Spearman correlations between PHO motif scores and PHO ChIP-seq for different epigenomic contexts.

(C) Spearman correlations between pho motif scores and the PH-normalized PHO ChIP-seq binding intensity for different epigenomic contexts.

(legend continued on next page)

ultrahigh-coverage Hi-C map from D.mel embryos and explored the internal Polycomb domain structure at a higher resolution than was previously attainable. Hi-C was performed as previously described (Sexton et al., 2012) and sequencing was extended in order to obtain ~281 million mapped and filtered contacts. When we looked at the interaction maps around pairs of PHO non-TSS sites within Polycomb domains, we observed a high degree of enrichment of interactions within 10 kb of the sites compared with regions more distal (~50 kb) to the sites (Figure 6E). Contact enrichment was significantly weaker for non-TSS sites in active domains (Figure 6F). Furthermore, we observed enrichment of contacts between non-TSS and TSSs sites within Polycomb domains, but not in active domains. For TSSs within Polycomb domains, the contact enrichment was localized, whereas for TSSs in active domains, we observed preferential contacts on the TSSs but also some potential contacts over the gene body (the diagonal of increased contacts in the lower-right quadrant of the Hi-C submatrix). We also quantified the absolute (rather than regionally normalized) fraction of contacts that PHO-binding sites form with other PHO sites (with or without DSP1 cobinding; Figure 6G). We found that PHO sites preferentially contact each other in Polycomb domains, but not in active or null domains, showing that in addition to the spatial preferences identified in Figure 6E, genomic clustering of PHO sites and intradomain compaction give rise to a distinct topological environment around PHO sites in Polycomb domains. These data raise some questions about the role of PHO- and PRC1-bound chromatin hubs in combining the low specificities of several unlinked genomic loci into a more stable folded structure.

PRC1 Knockout Results in a Polycomb-Domain-Specific Reduction in PHO Binding

To further test the cooperative nature of PHO and PRC1 recruitment within Polycomb domains, we analyzed PHO binding within or outside of Polycomb domains in PH mutant embryos. Notably, in these mutants, PC recruitment was also strongly reduced (Figure 7D), suggesting that the stability of the whole PRC1 complex was affected. qChIP experiments revealed a reduced binding of PHO to PREs within Polycomb domains in PH mutant embryos, whereas binding of PHO to active promoters outside the Polycomb context was not significantly affected (Figures 7A and S7A). In order to extend this analysis to a genome-wide scale, we performed PHO ChIP-seq in PH mutant embryos. Analysis of differential PHO binding in different contexts showed a highly significant decrease in PHO recruitment specifically in Polycomb domains (Figures 7B and 7C). This was concomitant with a significant reduction in PRC1 binding, as expected (Figure 7D). Interestingly, in PH mutant embryos, we detected a significant correlation between pho motifs and PHO binding even within Polycomb domains (rho = 0.17, $p < 6 \times 10^{-5}$ in non-TSS loci, rho = 0.44 $p < 10^{-8}$ in TSSs), in contrast to the lack of such dependency in wild-type (WT) (rho = 0.08, $p < 0.07$ in non TSSs, rho = 0.28, $p < 3 \times 10^{-4}$ in TSSs). We next wished to test whether outside of Polycomb domains and PRC1 presence, PHO-binding sites are colocalized with some alternative chromosomal factor. Analysis of domain-type-dependent colocalization of PHO and DSP1 with insulator proteins (Nègre et al., 2010) showed that in non-PcG contexts, TSS PHO sites were strongly colocalized with CP190 and BEAF32 (Figure S7A), and possibly linked to active transcription (Figure S7B) and enrichment of long-range contacts (Figure S7C). We did not detect significant insulator enrichment on DSP1 sites, or on PHO sites in a non-TSS context. Together, these data indicate that PHO recruitment to DNA within Polycomb domains involves a feedback interaction with PRC1. This feedback is likely facilitated through local cooperativity, but also occurs within the context of long-range contacts formed in PcG hubs as described above. Even outside of Polycomb domains, PHO binding may be promoted by and/or promote the formation of long-range contacts via colocalization with the general, non-sequence-specific CP190 protein.

DISCUSSION

Multilayer Organization and the Evolutionary Buffering of Polycomb Domains

We used comparative epigenomics to demonstrate that Polycomb domains are an extremely well conserved feature of the genome during fly evolution. In fact, the evolutionary profile of epigenomic domain organization in embryos of five *Drosophila* species indicates a complete lack of divergence of H3K27me3-marked Polycomb domains in syntenic regions. A similar high conservation of the H3K27me3 pattern across *Drosophila* species was recently described (Arthur et al., 2014). Polycomb domains typically harbor several PH-marked PREs, and a comparative analysis showed that these are also highly conserved and the few loci that show a divergence of PRC1 occupancy patterns are not correlated with overall domain divergence. Likewise, the binding of PHO and DSP1 is highly conserved (to a degree at

(D) Cumulative distributions (left) of the distances between PHO sites and the nearest adjacent PHO site outside of PcG domains (red) and within PcG domains on three levels of motif energy (blue, high; gray, medium; black, low). Right: cumulative distribution of differential PHO-to-PHO distances for syntenic loci between D.mel and D.pse, again stratifying according to domain context. Kolmogorov-Smirnov statistics for the PcG versus non-PcG distributions are indicated.
(E) 2D submatrices derived from Hi-C data, centered according to pairs of PHO-binding sites within PcG domains. The submatrices are oriented according to the strand of the closest TSS. Site pairs are classified as non-TSS (left), TSS (right), and mixed contexts (middle). The total number of observed contacts in each spatial bin was divided by the number expected from a technical correction model. Ratios were then further normalized according to the mean ratio in the outer 7 kb frame of each matrix. A fixed color scale was then applied to visualize the matrices.
(F) Similar to (E), but using PHO-binding sites in active domains.
(G) The total (marginal) number of contacts observed for 2 kb elements centered on PHO sites were counted and normalized to the total number of contacts between PHO sites and other sites, classed according to sites binding only PHO, only DSP1, or both PHO and DSP1. This analysis is distinct from that shown in (E) because it omits the regional normalization and thus reflects the absolute frequency of recruiter-recruiter interactions within Polycomb domains.
(H) Similar to (G), but based on control sites that show DSP1 enrichment without PHO.
See also Figure S6.

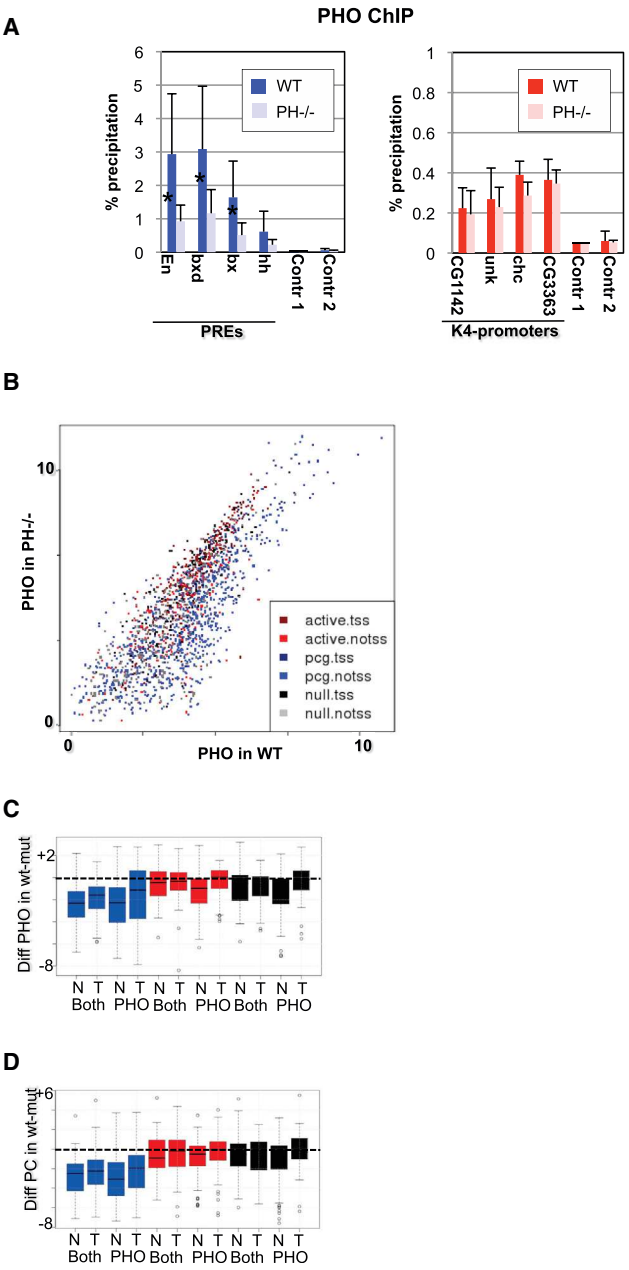


Figure 7. PRC1 Knockout Selectively Impairs PHO Recruitment to Polycomb Domains

(A) qChIP experiments of WT or PH mutant embryos ($PH^{-/-}$) using PHO antibodies. Primers specific for PREs (left) or active promoters bound by PHO outside the PcG context (right) were used. Results are represented as the percentage of input chromatin precipitated. The SD was calculated from at least four independent replicate experiments (*$p < 0.05$ as calculated from a two-tailed t test).

(B) PHO ChIP-seq intensities of WT and PH mutant embryos, color-coded according to context.

(C) Boxplot depicting differential PHO ChIP-seq binding in WT and PH mutant embryos, classified according to domain type (color), co-occurrence of DSP1 (both/PHO), and TSS context (N, non-TSS; T, TSS).

(D) Similar to (C), but showing differential PC ChIP-seq intensities.

See also Figure S7 and Table S2.

least as strongly, and possibly more strongly, than binding of individual factors; Bradley et al., 2010; He et al., 2011), but even cases of diverged factor occupancies are usually not correlated with overall PRE divergence. In marked contrast, the sequences underlying PREs and Polycomb domains are diverging extensively, and sequence-based prediction of PREs across *Drosophila* species suggested that divergence of PREs could occur frequently (Hauenschild et al., 2008). However, neither our ChIP-seq experiments nor our transgenic reporter assays support this dynamic behavior (Figure S3). Instead, we show that such sequence divergence is buffered by the epigenetic targeting mechanisms to maintain Polycomb domains. We suggest that the multilayered organization uses redundancy and cooperativity to facilitate the remarkable Polycomb domain conservation. This is occurring both in *cis*, where several TFs collaborate to define a regulatory element even when the underlying sequence is imperfect (see, e.g., Stefflova et al., 2013), and at the domain level, where several PREs participate to define the PcG domain structure and possibly stabilize each other.

The Sequence Specificities of PREs Are Not Fully Encoded in *cis*

Although PREs are associated with several known sequence features (such as GAGA- and PHO-binding motifs) in a statistically significant way, these features are not sufficient to distinguish many PREs from the genomic background and from other PHO- or DSP1-bound active chromatin elements (Schuettengruber et al., 2009). There are many possible explanations for this lack of specificity, including the existence of additional, yet-to-be-characterized sequence-specific recruiting factors; the involvement of nucleosome positioning; transcription of noncoding RNAs; or imperfect modeling of the sequence specificity of the known factors. The data presented here, however, introduce a new perspective that can help resolve this conundrum. In contrast to previous hypotheses, the data show that even when strong binding sites are lacking, PHO and DSP1 may bind PREs directly through weak (but highly nonrandom) motifs. Remarkably, sequence affinities that are completely nonspecific on a genomic scale (possibly defining millions of spurious sites) are still highly informative for predicting the binding intensity within the context of a PRE. The strong correlation of PHO binding with weak but nonrandom motifs makes it unlikely that binding to these sites represents indirect binding via interaction/looping with strong binding sites. The data show that in order to understand PRE sequence specificity, we must take into account multiple potential binding sites with variable affinities and fidelities, and consider their cooperative interaction in the context of the PRE chromosomal landscape. This idea is compatible with the evolutionary constraints on PRE sequences, which we have demonstrated here to affect a spectrum of binding affinities rather than to conserve classical binding sites alone.

Cooperative Rather than Hierarchical Interactions of PHO and PRC1 Contribute to PRE and Domain Specification

What might be the molecular mechanism that allows the specific binding of weak sites in the context of Polycomb domains? One possibility is that cooperative binding of TFs at PREs supports

their occupancy of weak motifs. Indeed, we found that PHO and DSP1 are bound jointly at PREs (with weak underlying sequence motifs), whereas at other regions of the genome where the factors bind alone, they are usually associated with strong sequence motifs. This observation is in agreement with the recently proposed "TF collective model," according to which combinatorial TF binding occurs with little or no apparent sequence motifs for at least a subset of the bound factors (Junion et al., 2012).

In addition, we show that transient interactions of DNA-binding proteins with weak affinity sites are stabilized by the presence of the PcG proteins themselves. A similar observation of a positive feedback of PRC1 on PHO binding was recently reported (Kahn et al., 2014) and is further supported by the fact that cooperative binding of PHO and Polycomb to PREs can occur even in vitro (Mohd-Sarip et al., 2005). In vivo, long-range contacts involving remote PREs within the same (or even a different) Polycomb domain may contribute to this process (Bantignies et al., 2011; Lanzuolo et al., 2007; Sexton et al., 2012). Clustering of multiple flanking PREs in the 3D space of the nucleus might generate Polycomb compartments characterized by high concentrations of PcG proteins as well as their recruiting DNA-binding proteins. In this scenario, loss of occupancy following the dissociation of any of these factors from DNA may be more easily replenished by the concentrated stock of factor within a Polycomb compartment compared with individual binding sites present elsewhere in the genome. This may push the equilibrium toward increased PHO and DSP1 binding to low-affinity sites and partially reduce the evolutionary pressure to maintain the nucleotidic sequence of recruiter motifs at PREs. Structural long-range effects may also inhibit PcG recruitment in cases where active enhancers and TSSs are in proximity to a candidate PRE sequence. Our analysis suggests that H3K4me3-marked loci are also highly conserved, but the low-affinity PHO- or DSP1-binding sites in them are completely uncorrelated with occupancy of these factors, further supporting a model of highly organized and cooperative epigenomic organization.

In conclusion, the data presented here indicate that sequence conservation collaborates with 3D chromatin architecture to maintain an exceptional evolutionary stability of Polycomb-regulated loci in fly genomes. This phenomenon highlights the contribution of chromosome domains and their particular looping structures to epigenomic specificity and genome evolution. Hi-C analysis in mammals has revealed that topological domains are a strikingly conserved feature between the mouse and human genomes (Dixon et al., 2012). Our data raise the possibility that, beyond combinatorial contributions by TF-binding sites in close proximity, the confinement of regulatory elements within TADs and their frequent DNA contacts constitute significant driving forces that also affect DNA sequence evolution in these and possibly many other species.

EXPERIMENTAL PROCEDURES

Fly Strains

Drosophila species were obtained from the *Drosophila* Stock Center (http://cis.arl.arizona.edu/stock.htm; stock numbers: D.sim 14021-0251.195, D.yak 14021-0261.01, D.pse 14011-0121.94, and D.vir 15010-1051.17). In addition, the Oregon-R w1118 line of D.mel (R. Paro, Center of Biosystems, Science and Engineering, Basel, Switzerland) was used.

Generation of Transgenic Fly Lines

Approximately 1.6 kb DNA fragments corresponding to conserved or diverged PRE regions from both D.mel and D.pse were obtained by PCR using specific primers. Each PCR fragment was cloned into a slightly modified attB-P[acman]-CmR-BW vector (BACPAC Resources Center) at the BamHI restriction site. Plasmids were sent to BestGene for site-specific integration using the PhiC31 integrase system into attp2 (chromosome 3L, cytological position 68A4) or VK7 sites (chromosome 3R, cytological position 82A1). Site-specific integration was verified by standard PCR analyses.

Eye Color Pigmentation Assay and Eye Imaging

For the eye pigmentation assay, 10 or 20 heads (depending on the eye color) of 4-day-old male flies were collected and homogenized in EPE buffer (30% EtOH-HCL [pH 2]) and incubated for 1 hr at 25°C in the dark. After centrifugation, eye pigmentation was quantified by spectrometry at 480 nm. The PSS score was calculated from mean pigment levels as heterozygote/homozygote for each diverged PRE region, normalized to the same ratio calculated for the "vector only" control line. Transgenic flies were photographed with a standard light microscope, always using the same time of exposure for each diverged PRE and controls (*$p < 0.01$ as calculated from a two-tailed t test).

ChIP Experiments

ChIP experiments on whole *Drosophila* embryos 4–12 hr after egg laying were performed essentially as described in Schuettengruber et al. (2009). Antibodies were diluted 1:100 for IP. For qChIP, after immunoprecipitation and DNA purification, enrichment of specific DNA fragments was analyzed by real-time PCR using Roche Light Cycler equipment and accessories as described in Comet et al. (2006). Data are expressed as the percentage of input chromatin precipitated for each region examined. As a negative control, *Rp49* was included in the PCR experiments. For primer sequences, see Table S2.

For ChIP-seq, 20 ng of the ChIP reaction and input DNA were used for the library preparation. ChIP-seq was performed by the Montpellier GenomiX facility (MGX, http://www.mgx.cnrs.fr/index.php).

Antibodies

Antibodies against modified histones were obtained from Millipore (H3K4me3 [#04-745] and H3K27me3 [#07-449]). Antibodies against the proteins PC, PH, and DSP1 are described in Schuettengruber et al. (2009). The antibody against PHO used for ChIP-seq in the different *Drosophila* species is described in Klymenko et al. (2006).

Sorting of PH Mutant Flies

Ph^{del} mutant flies (Feng et al., 2011) were crossed to a balancer line expressing GFP from the Krüppel promoter. Homozygous Ph^{del}/Ph^{del} embryos were collected from the heterozygous stock by selecting for GFP-negative embryos using an embryo sorter (COPAS SELECT; Union Biometrica). ChIP on WT and $Ph^{del/del}$ embryos was performed as described above. The PHO antibodies used are described in Schuettengruber et al. (2009). Note that this antibody gives essentially the same ChIP-seq profiles as the PHO antibody used in the species ChIP-seq (Klymenko et al., 2006), but seems to have reduced affinity in some non-melanogaster strains (data not shown).

Low-Level ChIP-Seq Analysis and Phylogenetic Projection

ChIP-seq mapping, normalization, and phylogenetic projection were performed as described in Supplemental Experimental Procedures. Briefly, we created binding profiles and identified nonmappable and nonalignable regions for each of the species independently. We then projected all syntenic and mappable regions onto the D.mel coordinate space to facilitate comparison.

Motif Finding

Discriminative motif finding was performed as previously described (Schuettengruber et al., 2009). The positive set in each species consisted of 400 bp elements around peaks of PH or H3K4me3. The background set was defined by shifting the positive set by 1,000 bp.

Sequence Affinities

For each motif and each species, we generated the sequence affinity in 20 bp bins with respect to the D.mel genome by summing over the position weight matrix (PWM) likelihoods across the orthologous sequence in the appropriate species. Each 20 bp bin was extended to the summation of the surrounded 400 bp window. Again, in order to control for variable sequence content in the different species, we transformed the sum of PWM likelihoods value to minus log2 of its (1 − quantile) value.

Phylogenetic Motif Tracing

To analyze the enrichment of the inferred motifs in PH or H3K4me3 sites along the *Drosophila* phylogeny, we used multiple alignment of 12 *Drosophila* species and computed the enrichments as defined above in sequences that are orthologous to the D.mel and D.pse elements. To ensure that the orthologous elements were of a uniform size, we always used 400 bp around the center of the projected locus.

Sequence Conservation Analysis

We used our previously described context-aware inferred substitution statistics (Chachick and Tanay, 2012; Kenigsberg and Tanay, 2013) to extract statistics on the observed and expected numbers of point substitutions for each alignable D.mel locus. We then pooled together loci around given landmarks (e.g., PH-binding sites or PHO/GAGA motifs) and summed up the total number of such substitutions at each relative offset to the landmark. The conservation statistics were then derived as log2(observed/expected).

PHO Syntenic Clustering

To create Figure 6D, we defined the PHO clustering score for each PHO site as the distance to the nearest adjacent PHO site (in logarithmic scale). To determine the extent to which this clustering property is conserved, we projected all D.mel PHO sites to the D.pse genome and computed the projected clustering score for adjacent sites that maintain their co-occurrence on the same chromosome or contig. We note that one must use caution when interpreting these data, since the analysis disregards all sites that were mapped to different chromosomes in D.pse, and the analysis is by definition asymmetric (e.g., performing a similar analysis starting from D.pse sites would not generate precisely the same distribution).

Hi-C Analysis

Hi-C raw sequence filtering, mapping, and normalization were done as previously described (Sexton et al., 2012). This provided statistics on the observed number of contacts for each pair of restriction fragments and the number expected from a technical background model. Given a set of sites, we characterized the spatial contact structure around interactions between them (Figures 6E and 6F) by extracting and pooling observed and expected statistics from 100 kb × 100 kb submatrices centered on the contact point of each pair. In cases of overlaps between such submatrices, we always assigned the contact to the pair of sites more proximal to it. We performed our analysis at restriction fragment resolution and then pooled data for 5 kb × 5 kb bins defined by distance to the contact point. Since the matrices were extracted around contacts with variable distances (and therefore with variable background contact intensity), we normalized the pooled matrix statistics using the average log(observed/expected) ratio of bins in the outer 5 kb frame. We also used an alternative approach (Figure 6G) in which we computed for windows of 2 kb around each site the total number of observed contacts with other 2 kb windows around sites of the same family, and divided this value by the total (marginal) number of contacts observed for restriction fragments within this window. When we used this normalization approach, we did not further normalize by the technical correction model.

ACCESSION NUMBERS

The raw sequencing data reported in this work have been deposited in the NCBI GEO under accession numbers GSE60428 and GSE61471.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and two tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2014.08.072.

AUTHOR CONTRIBUTIONS

B.S. and G.C. designed the experiments. N.O. and A.T. designed evolutionary and computational analyses. B.S., T.S., and M.E. performed ChIP experiments, Hi-C assays, and fly work. N.O. and T.A analyzed ChIP-seq data. N.O. implemented modeling and regression approaches with help from S.S. E.Y. helped with processing and normalizing raw Hi-C data. H.P. performed library preparation and sequencing. A.T., G.C., N.O., and B.S analyzed the data and wrote the manuscript.

ACKNOWLEDGMENTS

We are grateful to Jurg Müller and Judy Kassis for providing PHO antibodies, and D. Locker for providing DSP1 antibodies. We thank Mythily Ganapathi for helpful discussions. Research at G.C.'s lab was supported by grants from the European Research Council (ERC-2008-AdG No 232947), the CNRS, the INSERM, the European Network of Excellence EpiGeneSys, the Agence Nationale de la Recherche, and the Association pour la Recherche sur le Cancer. Research at A.T.'s lab was supported by the European Research Council, the European Network of Excellence EpiGeneSys, and the Israel Science Foundation.

Received: May 13, 2014
Revised: July 8, 2014
Accepted: August 27, 2014
Published: October 2, 2014

REFERENCES

Arthur, R.K., Ma, L., Slattery, M., Spokony, R.F., Ostapenko, A., Nègre, N., and White, K.P. (2014). Evolution of H3K27me3-marked chromatin is linked to gene expression evolution and to patterns of gene duplication and diversification. Genome Res. *24*, 1115–1124.

Atchison, L., Ghias, A., Wilkinson, F., Bonini, N., and Atchison, M.L. (2003). Transcription factor YY1 functions as a PcG protein in vivo. EMBO J. *22*, 1347–1358.

Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. Trends Genet. *27*, 454–464.

Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., Tixier, V., Mas, A., and Cavalli, G. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in Drosophila. Cell *144*, 214–226.

Blackledge, N.P., Farcas, A.M., Kondo, T., King, H.W., McGouran, J.F., Hanssen, L.L., Ito, S., Cooper, S., Kondo, K., Koseki, Y., et al. (2014). Variant PRC1 complex-dependent H2A ubiquitylation drives PRC2 recruitment and polycomb domain formation. Cell *157*, 1445–1459.

Bradley, R.K., Li, X.Y., Trapnell, C., Davidson, S., Pachter, L., Chu, H.C., Tonkin, L.A., Biggin, M.D., and Eisen, M.B. (2010). Binding site turnover produces pervasive quantitative changes in transcription factor binding between closely related Drosophila species. PLoS Biol. *8*, e1000343.

Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.L., and Kassis, J.A. (1998). The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. Mol. Cell *1*, 1057–1064.

Chachick, R., and Tanay, A. (2012). Inferring divergence of context-dependent substitution rates in Drosophila genomes with applications to comparative genomics. Mol. Biol. Evol. *29*, 1769–1780.

Cheutin, T., and Cavalli, G. (2014). Polycomb silencing: from linear chromatin domains to 3D chromosome folding. Curr. Opin. Genet. Dev. *25*, 30–37.

Comet, I., Savitskaya, E., Schuettengruber, B., Nègre, N., Lavrov, S., Parshikov, A., Juge, F., Gracheva, E., Georgiev, P., and Cavalli, G. (2006). PRE-mediated bypass of two Su(Hw) insulators targets PcG proteins to a downstream promoter. Dev. Cell *11*, 117–124.

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

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.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Duncan, I.M. (1982). Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. Genetics *102*, 49–70.

Feng, S., Huang, J., and Wang, J. (2011). Loss of the Polycomb group gene polyhomeotic induces non-autonomous cell overproliferation. EMBO Rep. *12*, 157–163.

Fiedler, T., and Rehmsmeier, M. (2006). jPREdictor: a versatile tool for the prediction of cis-regulatory elements. Nucleic Acids Res. *34*, W546–W550.

Fujioka, M., Yusibova, G.L., Zhou, J., and Jaynes, J.B. (2008). The DNA-binding Polycomb-group protein Pleiohomeotic maintains both active and repressed transcriptional states through a single site. Development *135*, 4131–4139.

Hauenschild, A., Ringrose, L., Altmutter, C., Paro, R., and Rehmsmeier, M. (2008). Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species. PLoS Biol. *6*, e261.

He, Q., Bardet, A.F., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A., and Zeitlinger, J. (2011). High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. Nat. Genet. *43*, 414–420.

Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. Cell *148*, 473–486.

Kahn, T.G., Stenberg, P., Pirrotta, V., and Schwartz, Y.B. (2014). Combinatorial interactions are required for the efficient recruitment of pho repressive complex (PhoRC) to polycomb response elements. PLoS Genet. *10*, e1004495.

Kassis, J.A., and Brown, J.L. (2013). Polycomb group response elements in Drosophila and vertebrates. Adv. Genet. *81*, 83–118.

Kenigsberg, E., and Tanay, A. (2013). Drosophila functional elements are embedded in structurally constrained sequences. PLoS Genet. *9*, e1003512.

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.

Kwong, C., Adryan, B., Bell, I., Meadows, L., Russell, S., Manak, J.R., and White, R. (2008). Stability and dynamics of polycomb target sites in Drosophila development. PLoS Genet. *4*, e1000178.

Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. Nat. Cell Biol. *9*, 1167–1174.

Lewis, E.B. (1985). Regulation of the genes of the bithorax complex in Drosophila. Cold Spring Harb. Symp. Quant. Biol. *50*, 155–164.

Lo Sardo, F., Lanzuolo, C., Comoglio, F., De Bardi, M., Paro, R., and Orlando, V. (2013). PcG-mediated higher-order chromatin structures modulate replication programs at the Drosophila BX-C. PLoS Genet. *9*, e1003283.

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

Mohd-Sarip, A., Cléard, F., Mishra, R.K., Karch, F., and Verrijzer, C.P. (2005). Synergistic recognition of an epigenetic DNA element by Pleiohomeotic and a Polycomb core complex. Genes Dev. *19*, 1755–1760.

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

Nègre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A., et al. (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet. *6*, e1000814.

Nègre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R., et al. (2011). A cis-regulatory map of the Drosophila genome. Nature *471*, 527–531.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

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

Rothenberg, E.V. (2014). The chromatin landscape and transcription factors in T cell programming. Trends Immunol. *35*, 195–204.

Schuettengruber, B., and Cavalli, G. (2009). Recruitment of polycomb group complexes and their role in the dynamic regulation of cell fate choice. Development *136*, 3531–3542.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol. *7*, e13.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458–472.

Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. Cell *154*, 530–540.

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

Zeng, J., Kirk, B.D., Gou, Y., Wang, Q., and Ma, J. (2012). Genome-wide polycomb target gene prediction in Drosophila melanogaster. Nucleic Acids Res. *40*, 5848–5863.

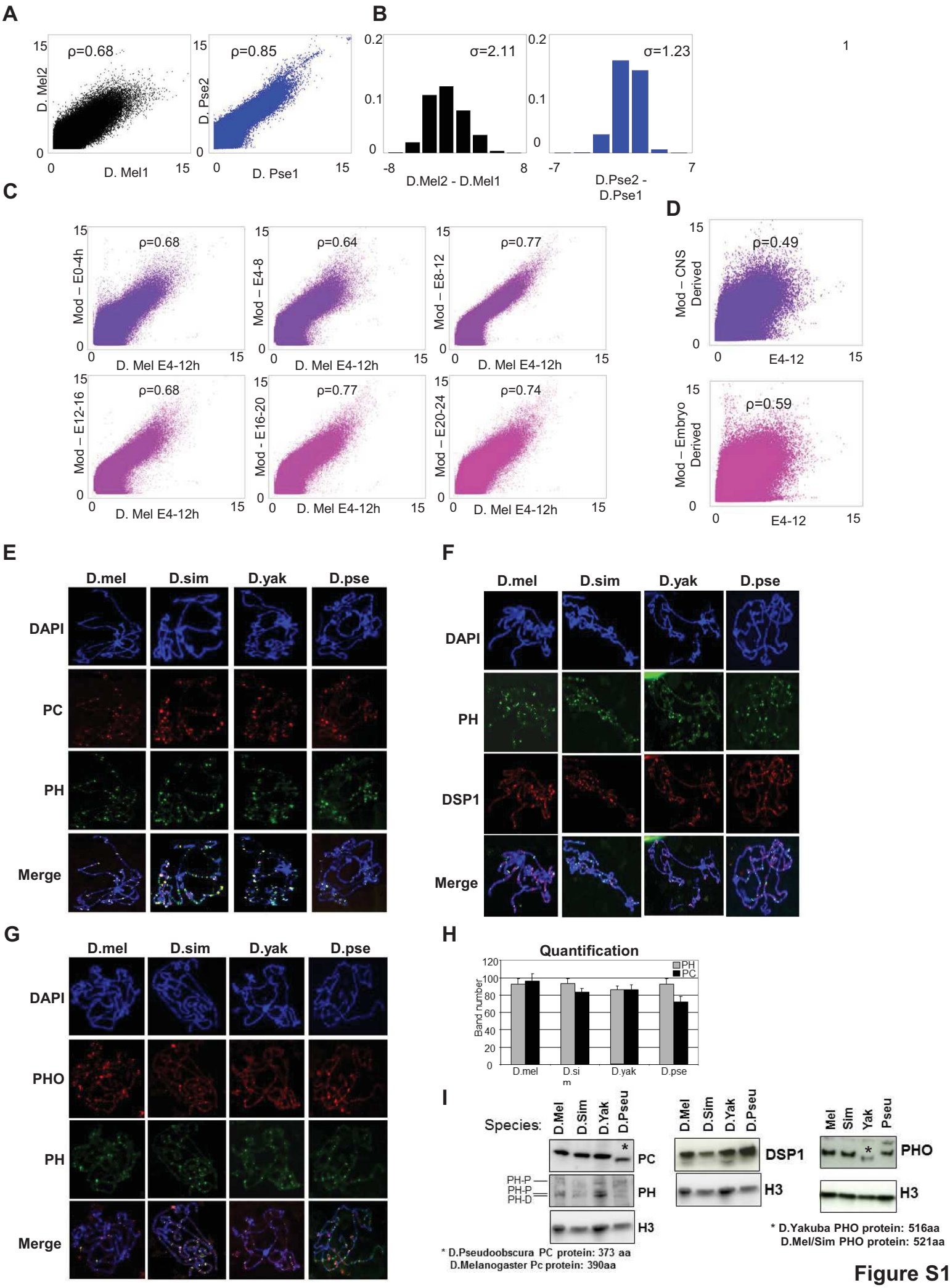
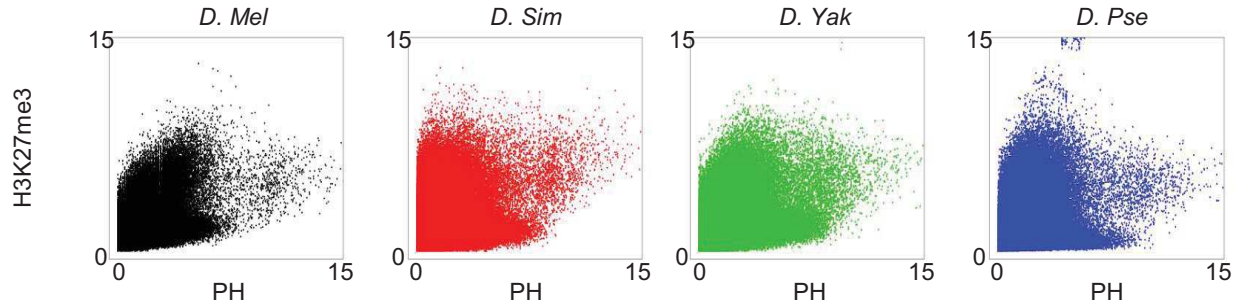

A
B
C
D
E
F
G
H
I
ρ=0.68
ρ=0.85
σ=2.11
σ=1.23
D. Mel2
D. Mel1
D. Pse2
D. Pse1
D.Mel2 - D.Mel1
D.Pse2 - D.Pse1
ρ=0.64
ρ=0.77
ρ=0.74
ρ=0.49
ρ=0.59
Mod – E0-4h
Mod – E4-8
Mod – E8-12
Mod – E12-16
Mod - E16-20
Mod – E20-24
D. Mel E4-12h
Mod – CNS Derived
Mod – Embryo Derived
E4-12
D.mel
D.sim
D.yak
D.pse
DAPI
PC
PH
DSP1
Merge
PHO
Quantification
Band number
D.mel
D.sim
D.yak
D.pse
Species:
D.Mel
D.Sim
D.Yak
D.Pseu
Mel
Sim
Yak
Pseu
H3
* D.Pseudoobscura PC protein: 373 aa
D.Melanogaster Pc protein: 390aa
* D.Yakuba PHO protein: 516aa
D.Mel/Sim PHO protein: 521aa

Figure S1

Figure S1, related to Figure 1.

Validation of ChIP-seq quality and antibody cross reactivity in Drosophila species: A, Scatterplots of H3K27me3 enrichment values between replicate ChIP-seq experiments. **B,** Distribution of differential ChIP-seq between different replicate experiments in D. mel and D. pse. **C,** Scatterplots of ChIP-seq H3K27me3 enrichment values compared between D. Melanogaster (4-12 hours) and different embryonic stages of D. Melanogaster from ModEncode (Negre et al.). **D,** Scatterplots of H3K27me3 enrichment values compared between D. Melanogaster (4-12 hours) and two different cell lines extracted from ModEncode (CNS derived and embryo derived)**. E-I,** test of cross reactivity of antibody specificity in Drosophila species: **E,** Immunostaining experiments with anti-PC and anti-PH, **F**, anti-DSP1 and anti-PH or **G**, anti-PHO and anti-PH antibodies on polytene chromosomes of 3rd instar larvae from four Drosophila species. **H,** Average band numbers of PC and PH from polytene chromosomes stained in panel A. 5-10 genomes were assessed per species and per antibody. Error bars show standard deviation. **I,** Western blot analysis of embryonic nuclear extracts of four Drosophila species with anti-PC and anti-PH (left), anti-DSP1 (middle), or anti-PHO (right) antibodies. Anti-histone H3 is shown as loading control. Note that the pseudoobscura PC protein is slightly smaller (373 aa) compared to the melanogaster subgroup PC protein (390 aa).

A



B

sim

mel \ sim	(0,4]	(4, 6]	(6, 8]	(8, inf]
(0,4]	5	4	2	13
(4, 6]	4	11	8	40
(6, 8]	1	4	15	105
(8, inf]	8	4	5	450

yak

sim \ yak	(0,4]	(4, 6]	(6, 8]	(8, inf]
(0,4]	5	3	1	9
(4, 6]	5	5	4	9
(6, 8]	0	2	11	17
(8, inf]	24	10	12	557

yak

mel \ yak	(0,4]	(4, 6]	(6, 8]	(8, inf]
(0,4]	8	2	4	12
(4, 6]	5	7	8	43
(6, 8]	2	4	8	111
(8, inf]	18	4	9	430

pse

sim \ pse	(0,4]	(4, 6]	(6, 8]	(8, inf]
(0,4]	1	1	1	11
(4, 6]	4	0	0	18
(6, 8]	1	3	4	22
(8, inf]	25	38	76	448

pse

mel \ pse	(0,4]	(4, 6]	(6, 8]	(8, inf]
(0,4]	8	3	0	15
(4, 6]	6	6	9	38
(6, 8]	3	12	21	87
(8, inf]	16	21	52	363

pse

yak \ pse	(0,4]	(4, 6]	(6, 8]	(8, inf]
(0,4]	5	2	1	23
(4, 6]	1	0	0	15
(6, 8]	0	2	5	22
(8, inf]	26	36	75	438

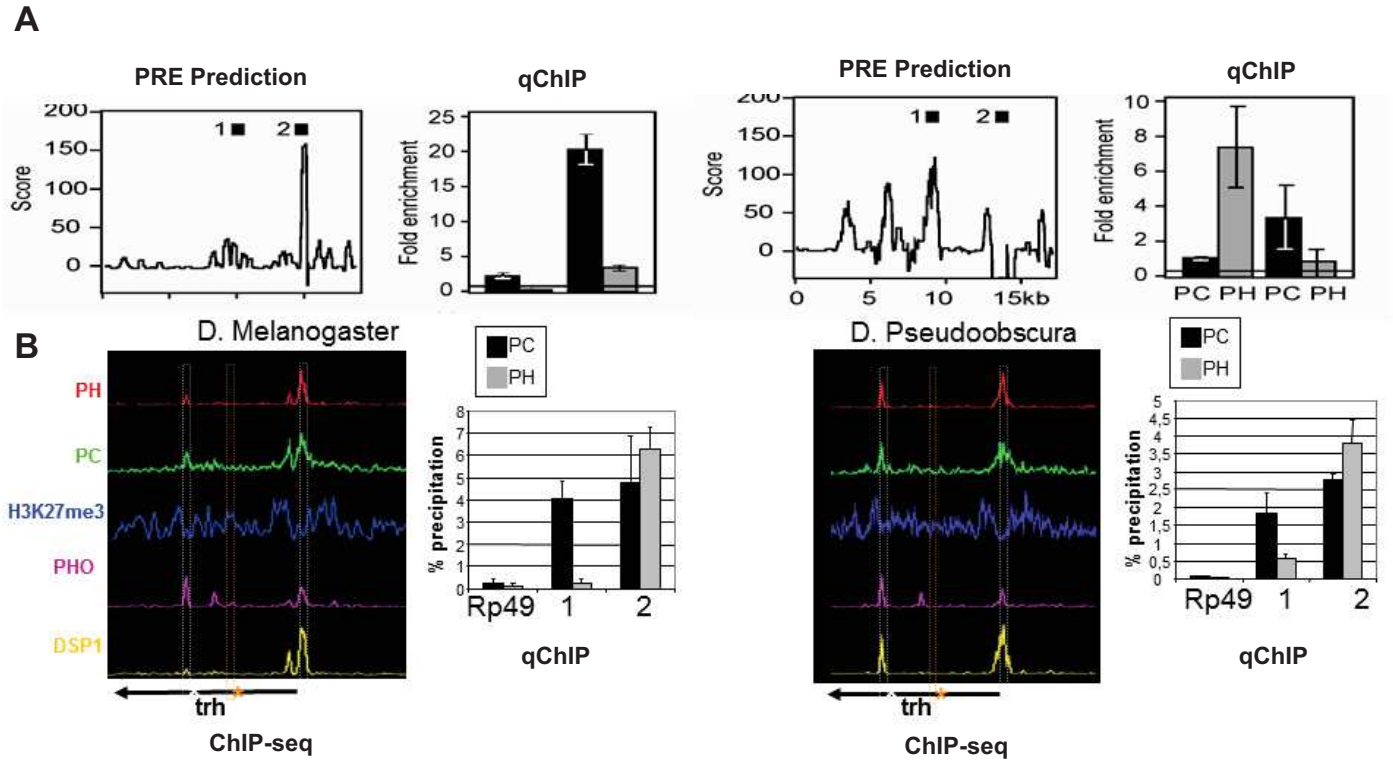
C

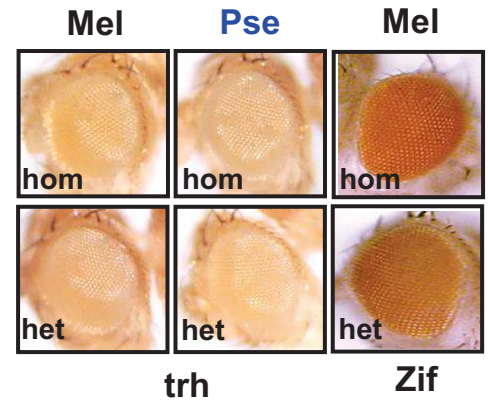
ID	chrom	start	end	mel_PH	sim_PH	yak_PH	pse_PH	mel_PHO	sim_PHO	yak_PHO	pse_PHO	mel_DSP1	sim_DSP1	yak_DSP1	pse_DSP1
1	chr2L	2454720	2454920	10.15	3.73	10.32	11.03	8.86	3.60	8.40	8.95	10.39	3.10	10.49	11.94
2	chr2L	3806520	3806720	4.48	3.45	5.89	10.05	3.97	5.71	6.72	7.79	5.75	5.24	7.09	5.26
3	chr2L	6824210	6824410	5.01	7.41	9.90	8.16	8.24	8.38	10.33	8.32	2.06	3.81	6.24	6.82
4	chr2L	12613100	12613300	5.22	9.75	10.06	6.77	8.03	9.90	8.84	7.09	2.76	5.65	5.29	2.70
5	chr2L	21792920	21793120	4.31	5.94	8.87	10.18	7.51	7.16	8.84	7.90	3.80	0.64	7.20	8.18
6	chr2L	21853960	21854160	3.99	4.32	3.42	10.07	2.39	6.21	5.70	11.75	2.90	1.94	4.02	7.14
7	chr2R	3939320	3939520	5.98	0.00	8.33	10.49	12.21	0.00	15.61	8.59	5.31	0.00	7.65	8.05
8	chr2R	5866670	5866870	14.39	4.15	12.59	11.97	14.84	2.94	11.50	11.85	9.84	0.64	8.93	9.29
9	chr2R	7460860	7461060	5.80	7.98	7.74	10.38	9.14	9.20	7.67	11.74	2.00	2.50	1.89	6.96
10	chr2R	16821330	16821530	6.37	9.26	10.19	6.75	8.01	8.27	8.28	6.44	3.39	5.67	8.59	4.22
11	chr2R	18937060	18937260	5.42	10.46	10.17	4.37	9.17	10.86	10.76	4.98	1.95	3.43	4.13	5.27
12	chr2R	19531870	19532070	4.13	9.59	10.10	4.35	9.62	9.85	10.56	4.11	2.62	3.15	5.93	2.90
13	chr3L	10272710	10272910	6.14	9.73	10.54	7.13	8.66	9.90	10.19	4.54	3.86	6.33	6.81	2.60
14	chr3L	10302640	10302840	5.23	9.70	10.50	8.07	9.69	10.43	10.47	8.14	4.10	5.56	7.03	4.57
15	chr3L	14126290	14126490	12.60	13.49	13.49	3.64	13.42	13.13	13.24	4.12	9.56	10.15	10.53	1.79
16	chr3L	18297570	18297770	4.90	9.31	9.90	9.10	3.94	6.53	2.88	2.76	5.00	6.65	6.90	6.18
17	chr3L	18847700	18847900	5.68	9.85	NA	9.94	8.15	10.02	NA	8.40	5.55	6.69	NA	7.19
18	chr3R	2821620	2821820	4.19	3.24	1.14	10.52	1.52	2.69	0.09	9.81	1.43	0.80	0.28	7.85
19	chr3R	8900400	8900600	9.73	10.76	10.20	4.40	9.74	10.01	9.63	1.24	8.30	7.60	7.68	3.38
20	chr3R	10481600	10481800	4.11	5.98	11.35	NA	5.50	7.26	13.02	NA	1.88	1.58	7.58	NA
21	chr3R	12489230	12489430	4.74	2.58	4.58	10.15	5.02	7.04	3.53	10.16	2.98	4.00	4.35	7.54
22	chr3R	12504930	12505130	2.05	6.60	7.94	10.60	3.67	5.86	1.07	9.67	0.69	2.73	3.21	7.23
23	chr3R	12515070	12515270	4.69	5.76	10.40	9.27	6.00	4.91	10.41	8.01	5.26	4.00	7.63	6.73
24	chr3R	12597870	12598070	6.97	0.32	9.80	10.09	9.35	0.00	10.29	10.29	6.33	0.64	7.54	7.68
25	chr3R	12645190	12645390	9.49	9.98	10.95	3.84	10.01	9.68	10.23	4.21	5.00	5.86	6.99	2.95
26	chr3R	12855660	12855860	8.59	10.01	9.90	5.68	8.97	10.02	8.97	3.81	4.56	5.22	7.63	4.31
27	chrX	581710	581910	3.06	3.07	5.17	10.56	5.29	2.94	6.17	3.71	8.62	8.83	9.74	7.83
28	chrX	5490480	5490680	5.19	4.78	2.76	12.95	4.82	7.33	1.04	9.31	2.12	4.26	1.03	13.71
29	chrX	8707520	8707720	4.27	5.88	1.78	10.81	8.12	8.20	0.16	9.26	5.47	4.35	0.17	11.23
30	chrX	12515710	12515910	4.19	9.88	10.82	9.99	2.45	8.36	7.76	5.39	8.36	7.60	10.28	10.29
31	chrX	13848510	13848710	2.19	9.75	9.71	10.51	6.81	9.07	8.37	9.15	1.62	6.95	7.54	8.94
32	chrX	17683590	17683790	12.06	NA	13.23	8.14	10.85	NA	10.14	6.21	12.47	NA	12.94	7.65

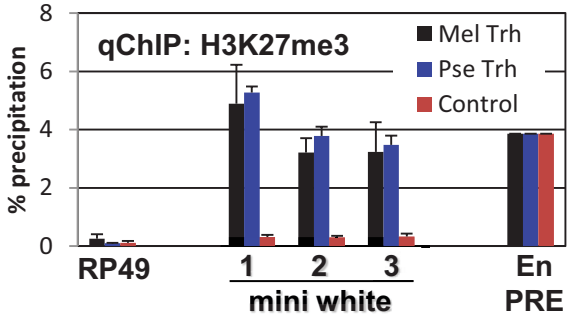
Figure S2

Figure S2, related to Figure 2.

Conservation of Polyhomeotic (PH) binding across Drosophila species: A, PH enrichments values compared to H3K27me3 enrichment values in all species excluding D. virilis**. B,** The numbers of conserved and diverged PH sites given variable threshold conditions, demonstrating very high, but not complete conservation. Counting PREs is problematic since it requires setting an arbitrary threshold on a quantitative phenomenon. ~30-40 PH sites have strong evidence for divergence (enrichment of over 8 in one species and below 4 in another species). **C,** Putative diverged PREs. Shown are the maximum ChIP-Seq values of PH, PHO and DSP1 (color-coded) in the four fly species within the regions of putative diverged PREs.


A
PRE Prediction
qChIP
PRE Prediction
qChIP
Score
Fold enrichment
PC PH PC PH
B
D. Melanogaster
PH
PC
H3K27me3
PHO
DSP1
trh
ChIP-seq
% precipitation
Rp49 1 2
qChIP
D. Pseudoobscura
trh
ChIP-seq
qChIP


C
Mel
Pse
Mel
hom
het
trh
Zif


D
qChIP: H3K27me3
Mel Trh
Pse Trh
Control
% precipitation
RP49
1
2
3
mini white
En
PRE

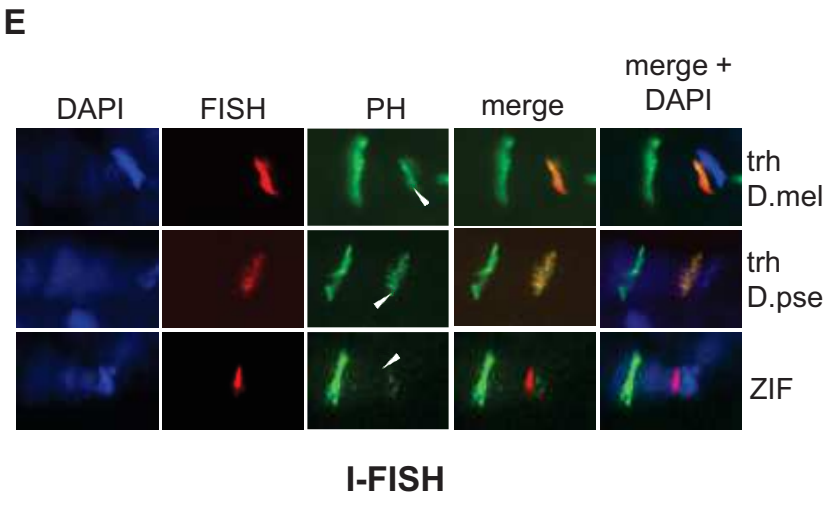
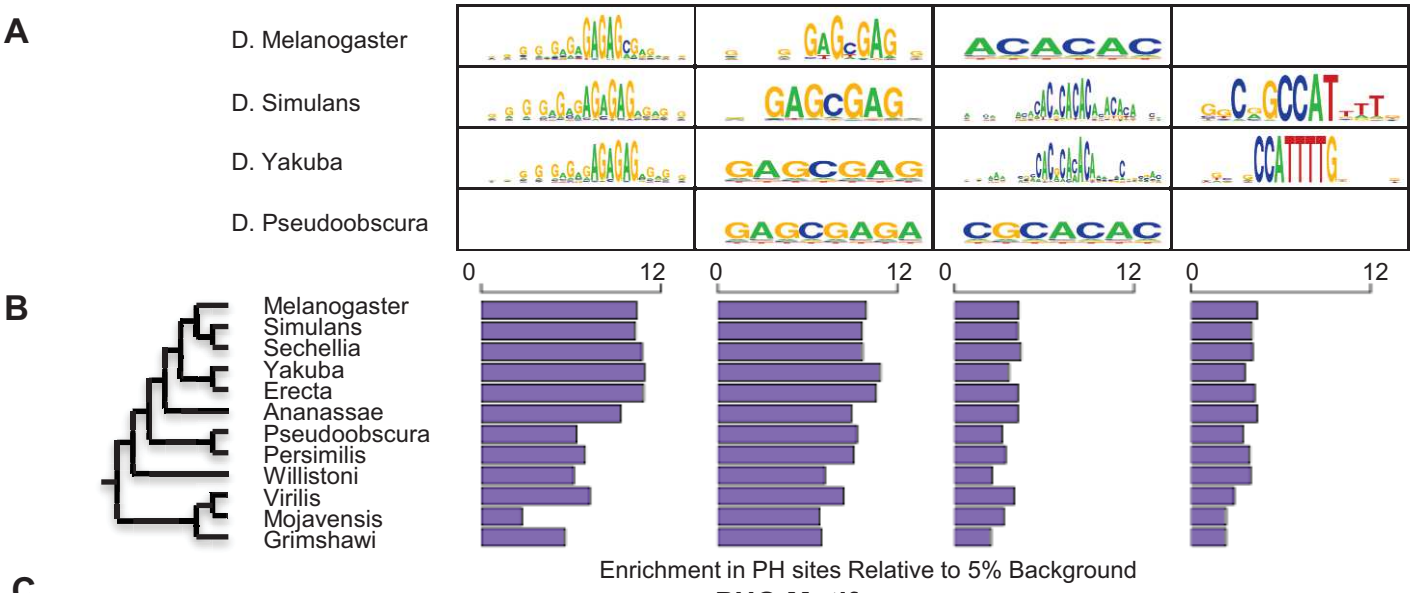
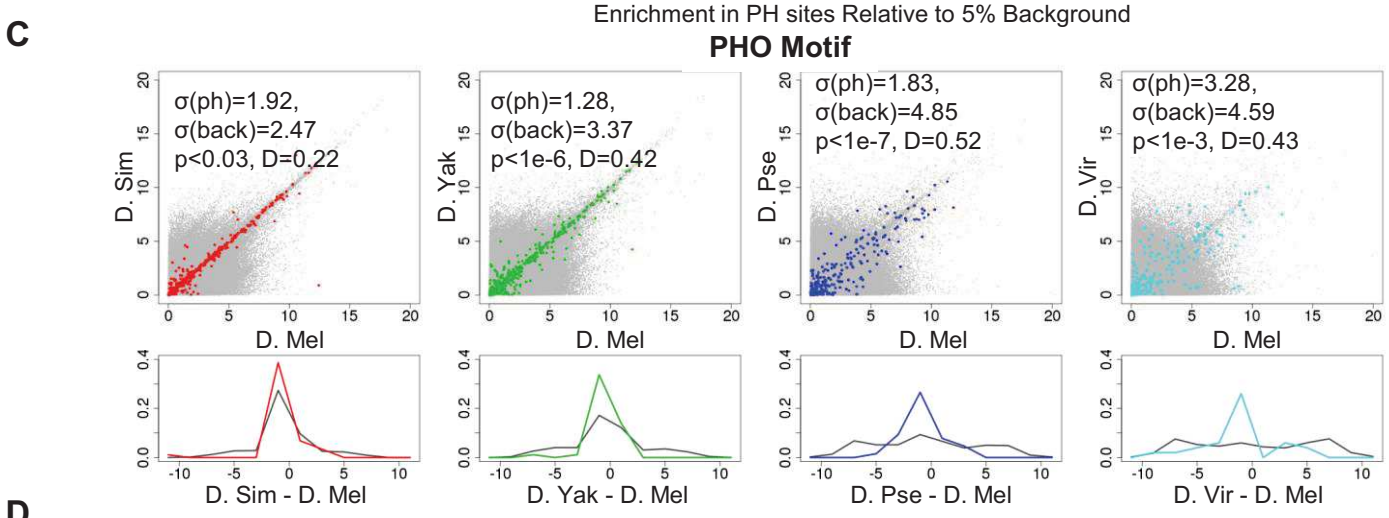

E
DAPI
FISH
PH
merge
merge + DAPI
trh D.mel
trh D.pse
ZIF
I-FISH

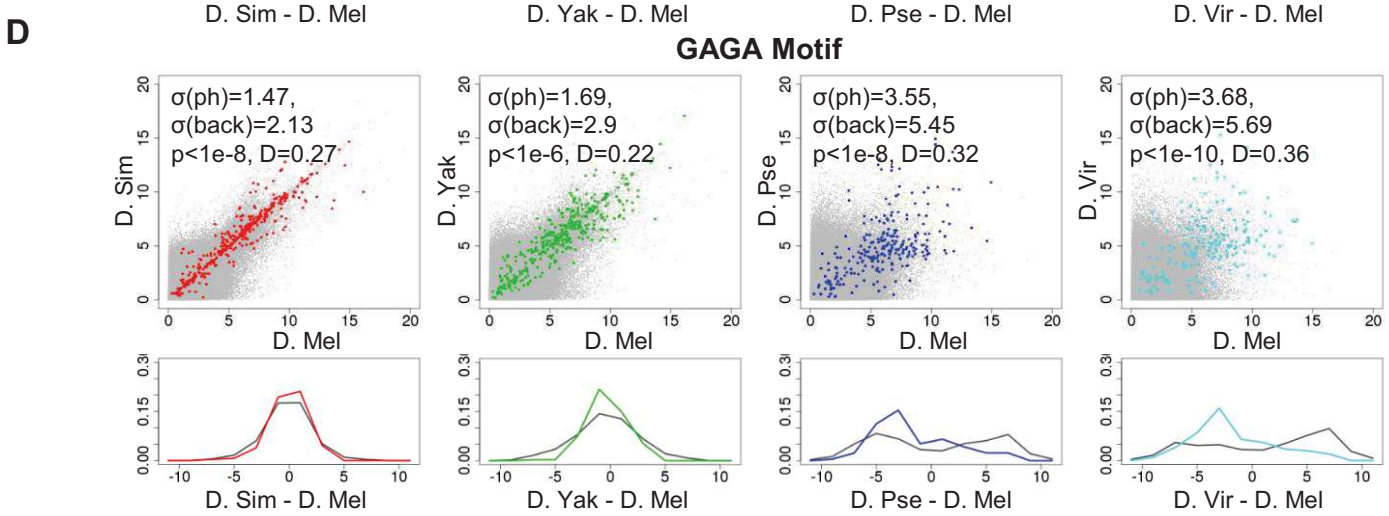
Figure S3

Figure S3, related to Figure 3.

Comparison of PRC1 binding at the trachealess (*trh*) gene locus with previously published binding profiles by (Hauenschild et al., 2008). A, PRE prediction scores for orthologous regions at the *trh* gene locus and binding of PC and PH at predicted PREs analysed by qChIP from (Hauenschild et al., 2008) are shown for D.melanogaster (left side) or D.pseudoobscura (right side). PRE prediction and qChIP experiments from (Hauenschild et al., 2008) suggests PRE divergence: No PRE was predicted at the promoter region of the *trh* gene in D. pseudoobscura and binding levels of PC and PH are weak (region 2), whereas a PRE is predicted in the coding region of the *trh* gene in D. pseudoobscura, which is absent in the melanogaster subgroup (region 1). **B,** Binding of PH and PC to the trachealess (trh) gene locus as determined by ChIP-seq (left) or qChIP experiments (right) in our study. White dashed boxes represent conserved PREs across Drosophila species, correlating with high levels of both PcG proteins. Dashed orange box indicates putative diverged PRE regions, previously identified by (Hauenschild et al., 2008). In agreement with our ChIP-seq results no diverged PcG binding can be observed and high levels of PC and PH can be detected at the promoter region of the trh gene locus (region 2) in all species by qChIP. Region 1, which is predicted to be a PRE in D. pseudoobscura (see panel A) is only strongly bound by PC, whereas the binding levels of PH are low in all tested species. Note that we observe broad binding of PC (but not PH) around PREs in qChIP experiments as we have reported previously in ChIP-on-chip experiments (Schuettengruber et al., 2009). Therefore the higher binding levels of PC to region 1 of the trh gene locus are unlikely to reflect intrinsic PRE activity. To further validate the conserved PRE activity of trh region 2, we tested this region in transgenic reporter assays: **C-E**, Results of transgenic reporter assays. **C**, Eye phenotype of transgenic fly lines: 4 day old male flies, either heterozygous (bottom) or homozygous (top) are shown on the left. **D,** qChIP assays performed on embryos (0-12 hours) of indicated transgenic fly lines using H3K27me3 antibodies (middle). Primers 1 and 2 amplify the TSS region whereas primer 3 amplifies the *white* coding region. Data were normalized to the positive control (*engrailed* PRE “En”). **E,** Immuno-FISH experiments were performed using PH antibodies on polytene chromosomes of the indicated transgenic lines. A DNA probe corresponding to the transgene insertion site was used to detect the location of the transgenic PRE (FISH, shown in red). PH immunostaining is shown in green. Note that an endogenous PH band is detected close (but not overlapping) to the transgene insertion site in the control (ZIF) line.







E

		KS (back VS ph (delta(yak-mel)) pVal / D (#cons data set)			KS (back VS ph (delta(pse-mel)) pVal / D (#cons data set)		
	Seq Conservation Bin	mel 0-5	mel 5-8	mel 8-Inf	mel 0-5	mel 5-8	mel 8-Inf
PHO motif	1 (least conserved)	<1e-3 / 0.37 (30)	0.05 / 0.79 (3)	0.49 / 0.49 (3)	0.08 / 0.27 (22)	-	0.4 / 0.64 (2)
	2	<1e-7 / 0.21 (186)	0.003 / 0.31 (35)	0.08 / 0.31 (17)	0.008 / 0.16 (104)	<1e-4 / 0.51 (23)	0.07 / 0.44 (9)
	3 (most conserved)	0.001 / 0.18 (117)	<1e-3 / 0.48 (18)	0.38/0.31 (9)	0.5 / 0.1 (68)	<1e-3 / 0.71 (9)	0.06 / 0.48 (7)
GAGA motif	1 (least conserved)	0.82 / 0.17 (14)	0.05 / 0.34 (16)	0.66 / 0.28 (6)	0.55 / 0..26 (9)	0.1 / 0.36 (12)	0.04 / 0.76 (3)
	2	0.26 / 0.1 (94)	<1e-3 / 0.22 (93)	<1e-3 / 0.29 (51)	0.35 / 0.15 (41)	<1e-3 / 0.28 (61)	<1e-3 / 0.36 (34)
	3 (most conserved)	0.002 / 0.3 (40)	<1e-3 / 0.24 (70)	0.16 / 0.2 (34)	0.16 / 0.22 (26)	0.002 / 0.31 (38)	0.003 / 0.4 (20)

F

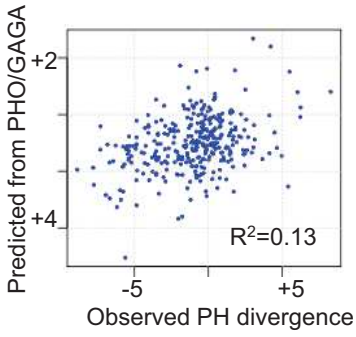


Figure S4

Figure S4, related to Figure 4.

Sequence composition of conserved PREs. A, Results of de-novo motif analysis in PH binding sites, compared to background sequences. PWM models were inferred separately for each species. **B,** Enrichment ratios for motif occurrence in PREs compared to background sequences. We used the same PWM in all species, and tested sequences that were orthologous to conserved PREs sequences given our four ChIP-seq profiles **C,D,** Conservation of PHO and DSP1 motifs in the genomic background (gray) and conserved PREs (colored). Each point represents one genomic locus, and the plotted values represent the sequence affinity as determined by the PHO or GAGA motif model, compared between species. In each of the lower panels we plot the distribution of sequence affinity difference between *D. melanogaster* and one of the other species. We computed these distributions separately for orthologous PRE sequences and the genomic background. The standard deviations of the distributions are noted, together with the Kolgorov-Smirnov statistics comparing them. **E**, Kolgorov-Smirnov statistics from conservation analysis that was stratified according to regional overall conservation of the locus (rows showing low, medium and high conservation), and the motif affinity in D.mel (low, medium and high in columns). For each stratum we show the D and P values of a KS analysis comparing PH sites to the background. This stratification is important since PH sites are biased toward higher than average conservation. **F**, The result of a linear model aiming to predict the divergence in PH binding intensity from the divergence in PHO and GAGA motif affinities between D. mel and D. pse. While the model reflects a highly significant association between the motifs and PH evolution, it explains only a small fraction of the variation (R^2 = 0.13), suggesting that PH evolution involves additional factors.

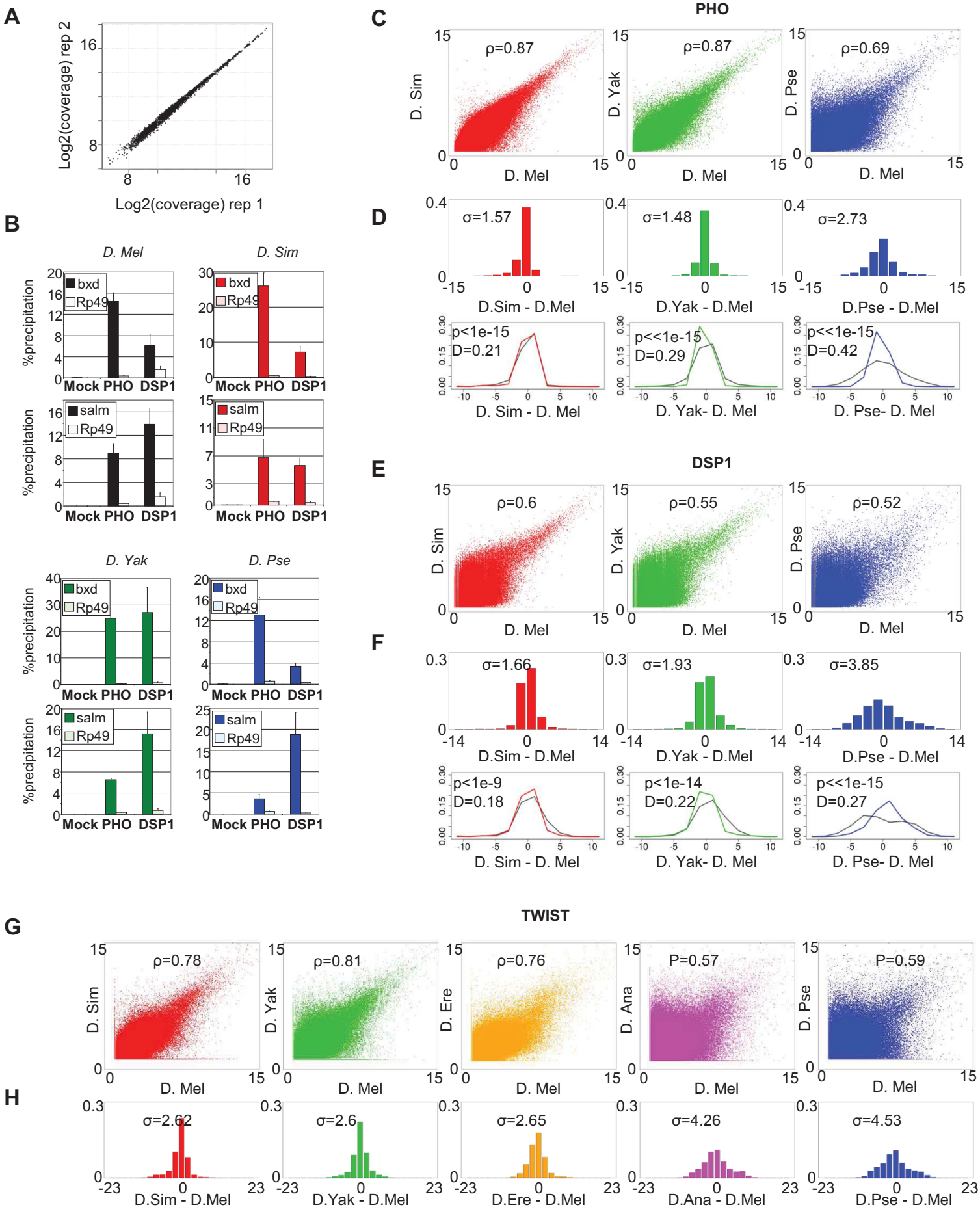
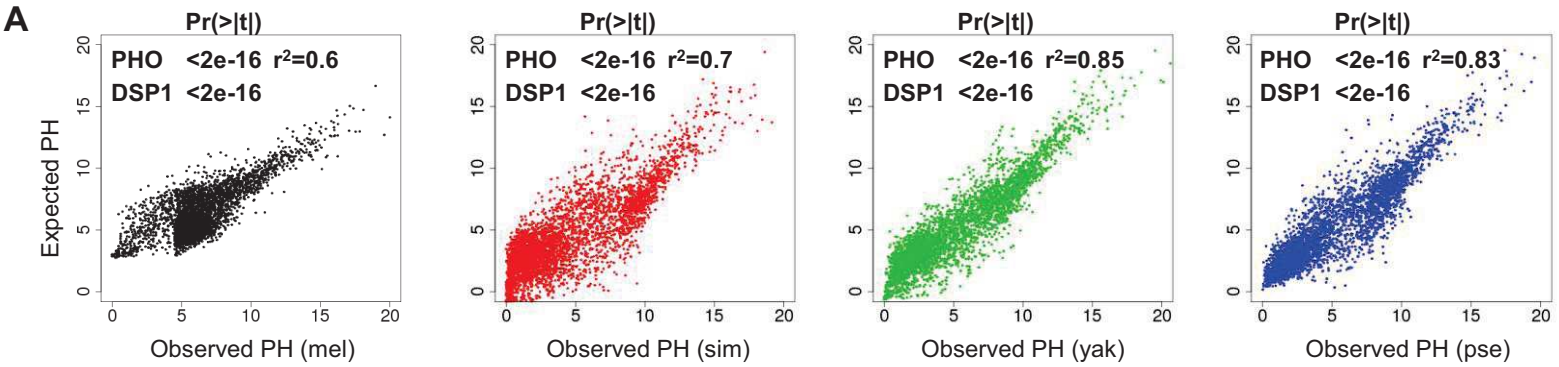
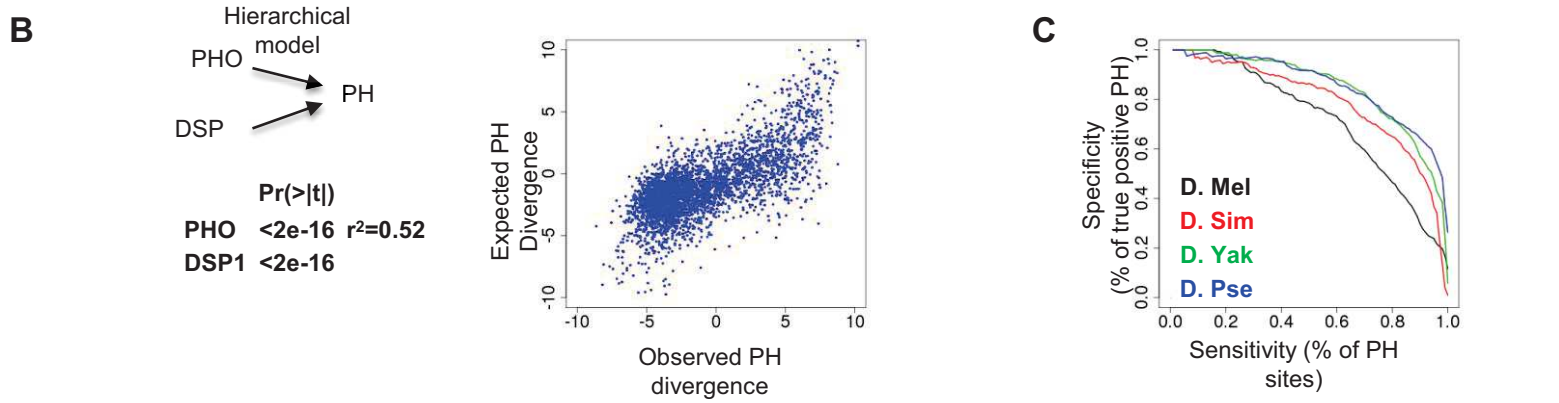

A
Log2(coverage) rep 2
Log2(coverage) rep 1
B
D. Mel
D. Sim
D. Yak
D. Pse
%precipitation
bxd
salm
Rp49
Mock PHO DSP1
C
PHO
ρ=0.87
ρ=0.87
ρ=0.69
D. Sim
D. Yak
D. Pse
D. Mel
D
σ=1.57
σ=1.48
σ=2.73
D.Sim - D.Mel
D.Yak - D.Mel
D.Pse - D.Mel
p<1e-15
D=0.21
p<<1e-15
D=0.29
p<<1e-15
D=0.42
D. Sim - D. Mel
D. Yak- D. Mel
D. Pse- D. Mel
E
DSP1
ρ=0.6
ρ=0.55
ρ=0.52
F
σ=1.66
σ=1.93
σ=3.85
p<1e-9
D=0.18
p<1e-14
D=0.22
p<<1e-15
D=0.27
G
TWIST
ρ=0.78
ρ=0.81
ρ=0.76
P=0.57
P=0.59
D. Ere
D. Ana
H
σ=2.62
σ=2.6
σ=2.65
σ=4.26
σ=4.53
D.Sim - D.Mel
D.Yak - D.Mel
D.Ere - D.Mel
D.Ana - D.Mel
D.Pse - D.Mel

Figure S5

Figure S5, related to Figure 5.

PcG recruiter binding across Drosophila species. A, comparing ultra-deep ChIP-seq enrichment values in replicate experiment. Shown are data from replicate experiment for PHO in D. mel, depicting normalized enrichment values in the X and Y axis. The data provides excellent dynamic range. **B,** ChIP-qPCR validation for PHO and DSP1 binding levels in the salm and bxd loci **C,** Comparative analysis of PHO ChIP-seq enrichment values in four fly species. **D,** Shown in the upper panels are the distributions of PHO binding divergences, computed as the differences in aligned ChIP-seq enrichment values between D. melanogaster and one of three other species. The distributions and standard deviations are estimated from loci with a PHO enrichment value in D. melanogaster >7. Shown in the lower panels are the divergence distributions within conserved PH sites (colored) compared to the distribution of PHO divergence in non-PH PHO binding sites (gray). **E,F,** Same as C-D, but with DSP1 ChIP-seq. **G,H,** Same as C-D, but analyzing TWIST ChIP-Seq data. Note that data on two additional species are shown (D. Erecta and D. Ananassae) (He et al.).


A
Pr(>|t|)
PHO <2e-16 r²=0.6
DSP1 <2e-16
Expected PH
Observed PH (mel)
Pr(>|t|)
PHO <2e-16 r²=0.7
DSP1 <2e-16
Observed PH (sim)
Pr(>|t|)
PHO <2e-16 r²=0.85
DSP1 <2e-16
Observed PH (yak)
Pr(>|t|)
PHO <2e-16 r²=0.83
DSP1 <2e-16
Observed PH (pse)


B
Hierarchical model
PHO
DSP
PH
Pr(>|t|)
PHO <2e-16 r²=0.52
DSP1 <2e-16
Expected PH Divergence
Observed PH divergence
C
Specificity (% of true positive PH)
Sensitivity (% of PH sites)
D. Mel
D. Sim
D. Yak
D. Pse

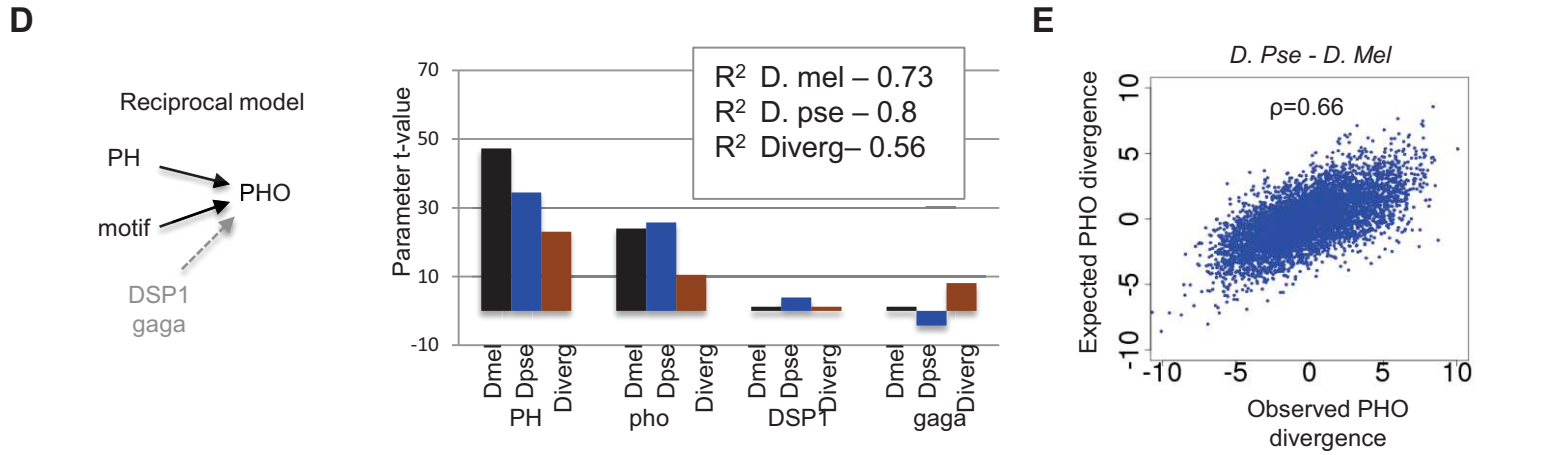

D
Reciprocal model
PH
PHO
motif
DSP1
gaga
Parameter t-value
R² D. mel – 0.73
R² D. pse – 0.8
R² Diverg– 0.56
Dmel
Dpse
Diverg
PH
pho
DSP1
gaga
E
D. Pse - D. Mel
ρ=0.66
Expected PHO divergence
Observed PHO divergence

Figure S6

Figure S6, related to Figure 6.

Prediction of protein binding levels: **A,** A linear model predicts quantitatively PH binding from PHO and DSP1 binding intensities. Shown are the R-squared values and the p-values of the two-sided t-tests indicating that both PHO and DSP1 contribute significantly to the prediction of PH binding. **B,** Same as **A** but predicting divergence in PH ChIP-seq enrichment given divergence in PHO and DSP1 ChIP-seq enrichments. **C,** Graph of the specificity obtained when predicting PH sites given PHO and DSP1 enrichment with decreasing thresholds. The X-axis depicts the sensitivity of the predictor (the fraction of PH sites that are predicted when setting the threshold). Each curve shows data for a different species. **D,** an alternative model predicting PHO levels from linear combinations of the pho motif score and PH levels was fitted. The R^2 values obtained in D. mel and D.pse, or when fitting the divergent PHO levels are shown. The bar graph shows the significance of the parameters used by the models indicating that trying to add the DSP1 Chip-seq levels or the GAGA motifs are not informative for predicting PHO levels, but that both PH binding and the PHO motif synergistically contribute to improve model fit. **E,** shown are the results of the model predicting PHO divergence from pho motif and PH ChIP-seq divergence between D. mel and D. pse.

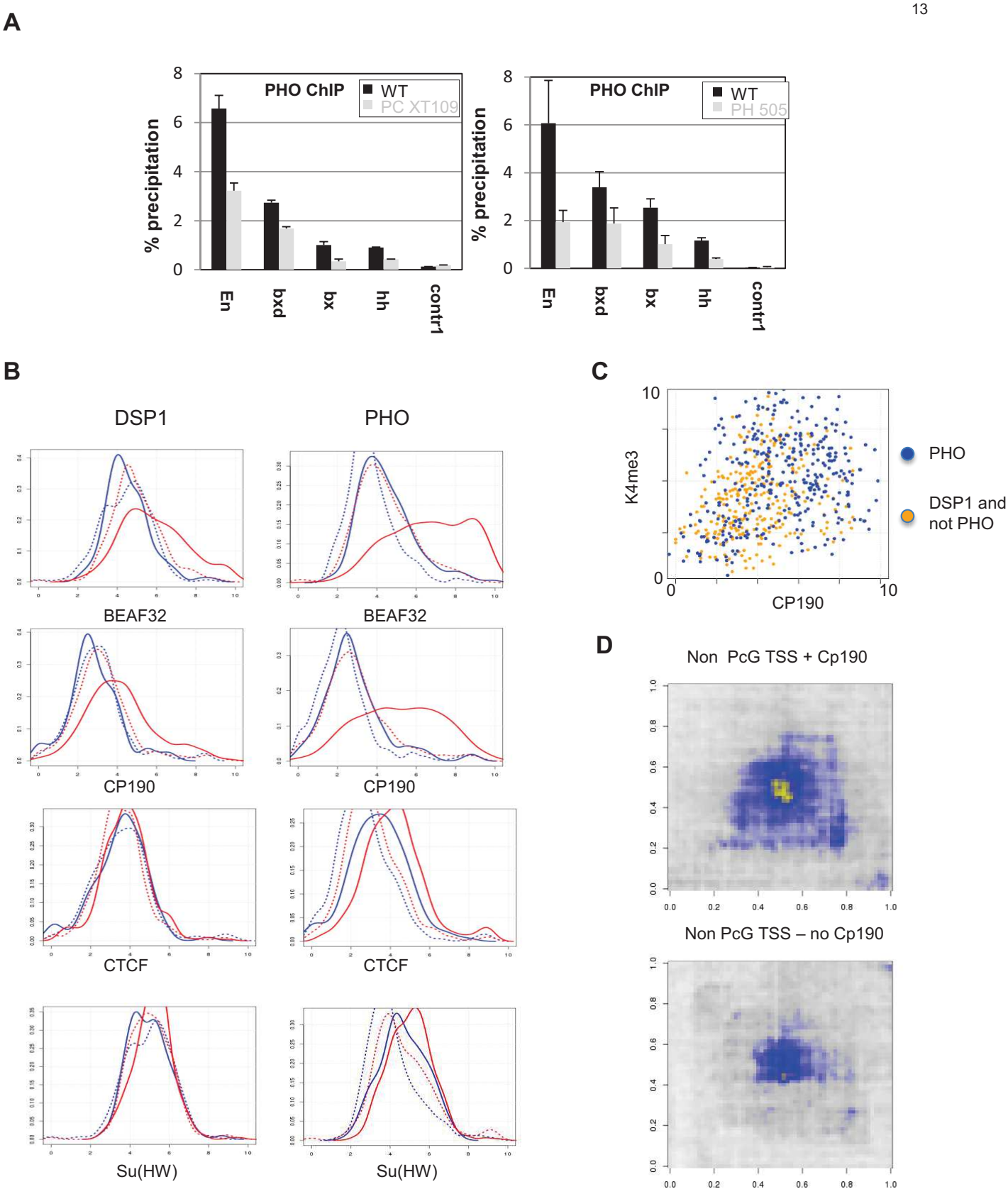

A
PHO ChIP
WT
PC XT109
% precipitation
En
bxd
bx
hh
contr1
PHO ChIP
WT
PH 505
% precipitation
B
DSP1
PHO
BEAF32
CP190
CTCF
Su(HW)
C
K4me3
CP190
PHO
DSP1 and not PHO
D
Non PcG TSS + Cp190
Non PcG TSS – no Cp190

Figure S7

Figure S7, related to Figure 7.

PHO binding in PRC1 mutant embryos: A, qChIP experiments of WT versus PC XT109 mutant embryos (left) or WT versus PH505 mutant embryos (right) using PHO antibodies. Primers specific for PREs or a control region were used. Results are represented as percentage of input chromatin precipitated. **B**, Distribution of insulator binding intensities on PHO or DSP1 sites, stratified according to context (Blue – PcG domain, Red – non PcG domain, solid – TSS, dashed – non TSS). **C,** Scatter plot depicting CP190 and H3K4me3 ChIP-seq values for PHO sites out of PcG context, and for DSP1 sites that are both out of PcG context and are not co-occupied by PHO. **D**, High coverage Hi-C data was used to reconstruct the distribution of contacts between TSS PHO sites that are out of PcG context and divided according to the presence of Cp190. Hi-C analysis strategy is as described for Fig 6D.

Table S1, related to Figure 2.

List of conserved PREs across Drosophila species.

Table S2, related to Figure 2, Figure 3, Figure 7, Figure S3, Figure S5 and Figure S7

List of primer sequences used in this study.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Embryonic development rates of Drosophila species

The embryonic development of the D. melanogaster subgroup (D.mel, D.sim and D.yak) is very similar, whereas the embryonic development of D. pseudoobscura and D.virils proceeds with a slightly slower rate. To be able to compare embryos of the different species at the same developmental stage we let embryos of D. pseudoobscura and D. virilis develop for one additional hour to minimize differences in sampling of developmental stages in the tested species. In addition to the fact that we look at a wide range of embryonic stages (4-12 hours) we exclude that differences in the observed ChIP enrichment levels are due to a different distribution of embryonic ages of the tested species.

Testing the cross reactivity of Antibodies in Drosophila species

All antibodies were raised against the D. melanogaster proteins. Protein sequence analysis revealed a high degree of conservation of Polycomb (PC), Polyhomeotic (PH), DSP1 and PHO proteins across the tested species (data not shown). To verify the cross reactivity and specificity of antibodies in different *Drosophila* species (except antibodies against modified histones), we performed immunostaining experiments of polytene chromosomes prepared from 3rd instar larvae in different *Drosophila* species and compared it to the pattern observed in D. melanogaster (**Fig. S1E-H**). In all tested species a similar number of bands with comparable intensities were detected for both PcG proteins. Even more importantly PC and PH perfectly colocalize in all species, indicating that these antibodies are not impaired in the recognition of its epitope in a given species and do not cross-react with other chromatin associated proteins. Furthermore, Western Blot analysis on embryonic nuclear extracts of the four species (**Fig. S1I**) indicated that antibodies against PC and PH recognize their substrates with comparable affinities. As previously reported (Dejardin et al., 2005) DSP1 binds to hundreds of bands on polytene chromosomes in D. melanogaster showing a moderate overlap with PcG proteins. Importantly, immunostaining experiments across different species revealed very similar staining patterns and number of DSP1 bands from D. melanogaster to D. pseudoobscura (**Fig. S1F**). In addition Western blot analysis on embryonic nuclear extracts of the four species using DSP1 antibody revealed nearly identical results (**Fig. S1I**). As observed for DSP1, PHO immunostaining experiments on polytene chromosomes revealed a very similar staining pattern in all 4 species (**Fig. S1G**). A moderate overlap of PHO with PcG proteins was observed as reported previously (Brown et al., 2003). Importantly, staining intensities in all species are very similar on polytene chromosomes and Western blot analysis. Finally, the fact that we do not see a general

decrease in enrichment levels in qChIP or ChIP-seq assays in any tested *Drosophila* species, strongly indicates that the affinity of the antibodies is similar in all species.

Raw ChIP-seq data analysis

Comparative ChIP-seq analysis was performed by mapping sequence reads to each of the species and projecting normalized coverage profiles over the D. melanogaster genome, taking into consideration regions with low mappability and/or alignability in each species as follows:

A. Mapping - For each ChIP-seq experiment, we mapped short reads (36/50 bps) to the appropriate reference genome using Bowtie-0.12.2 by specifying '-a –m 1 -best -strata' options. Our reference genomes included the heterochromatic fraction of chromosomes and non-assembled contigs, and any read that could be mapped to more than one genomic locus was filtered from further processing. We used genome assemblies downloaded from UCSC (dm3, droSim1, droYak2, dp4, and droVir3).

B. Pile up - Mapped tags were extended to 140bp and coverage statistics for 20bp bins over each of the genomes were computed. This was done following analysis of the correlation between coverage in the forward and reverse strands, which indicated peak correlation between the two at an offset of 140bp, suggesting the empirical fragment length assayed by ChIP-seq was 140bp.

C. Normalizing per-species mappability. Each species genome contains regions that are non-unique and therefore cannot be profiled using short read tags. We generated a control profile for each species by applying the mapping and binning procedure as described above for all possible 36/50bp subsequences in each of the genomes. We derived mappability tracks for each species, indicating the fraction of fragments covering each 20bp genomic bin that could be mapped uniquely to the respective genome. We then normalized the coverage score of each experiment by dividing it with this mappability value. Regions with mappability lower than 50% were filtered out and marked as Not Mappable in subsequent analysis. Note that unmappable regions were different between the species due to duplications and deletions.

D. Coverage normalization. To control for ChIP-seq coverage and variable ChIP-seq specificity, we transformed raw coverage values to minus log2 of its (1 - quantile) value For example, a normalized value of 9 indicates coverage is in the top $1\text{-}2^{-9}$ quantile (e.g. – in the top (1/512)th of the distribution).

Projecting ChIP-seq onto the melanogaster genome. To facilitate epigenomic comparisons, we projected all functional data over the syntenic fraction of the D. melanogaster genome. We used the reference multiple alignment obtained from the UCSC genome browser and created projected ChIP-seq tracks (in D. melanogaster genomic coordinates, using fixed 20bp bins) from each of the ChIP-seq coverage profiles in the other species. Our projection algorithm iterated over 20bp bins in D. melanogaster and identified, for each species, all 20bps that were aligned to it. We then performed the following calculations:

* If no bins were aligned to the melanogaster bin, the projected track value was defined as not alignable.

* If all bins aligned to the melanogaster bin were defined as not mappable in the other genome, the projected track value was defined as not mappable.

Both unalignable and unmappable bins were discarded from subsequent analysis.

* In all other cases, the projected bin value was defined as the weighted average of the values of all aligned bins (weighted according to the length of the aligned sequence).

Note that the projected ChIP-seq values represent log(1-quantiles) values in their respective original genome, not in the melanogaster, or the syntenic melanogaster genomes.

SUPPLEMENTAL REFERENCES

Brown, J.L., Fritsch, C., Mueller, J., and Kassis, J.A. (2003). The Drosophila pho-like gene encodes a YY1-related DNA binding protein that is redundant with pleiohomeotic in homeotic gene silencing. Development *130*, 285-294.

Dejardin, 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.

Hauenschild, A., Ringrose, L., Altmutter, C., Paro, R., and Rehmsmeier, M. (2008). Evolutionary plasticity of polycomb/trithorax response elements in Drosophila species. PLoS Biol *6*, e261.

He, Q., Bardet, A.F., Patton, B., Purvis, J., Johnston, J., Paulson, A., Gogol, M., Stark, A., and Zeitlinger, J. High conservation of transcription factor binding and evidence for combinatorial regulation across six Drosophila species. Nat Genet *43*, 414-420.

Negre, N., Brown, C.D., Ma, L., Bristow, C.A., Miller, S.W., Wagner, U., Kheradpour, P., Eaton, M.L., Loriaux, P., Sealfon, R.*, et al.* A cis-regulatory map of the Drosophila genome. Nature *471*, 527-531.

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., van Lohuizen, M., Tanay, A., and Cavalli, G. (2009). Functional anatomy of polycomb and trithorax chromatin landscapes in Drosophila embryos. PLoS Biol *7*, e13.