Opinion TRENDS in Genetics Vol.23 No.10

Polycomb Group proteins: an evolutionary perspective

Sarah J. Whitcomb*, Amrita Basu*, C. David Allis and Emily Bernstein

Lab of Chromatin Biology, The Rockefeller University, 1230 York Ave, Box 78, New York, NY 10065, USA

The chromatin-associated Polycomb Group (PcG) proteins were first identified in genetic screens for homeotic transformations in *Drosophila melanogaster*. In addition to body patterning in metazoans, members of the PcG are now known to regulate epigenetic cellular memory, pluripotency and stem cell self-renewal. Here, we discuss the functional versatility of the PcG family and the evolutionary history of a subset of these proteins including *Drosophila* E(z), Pc, Psc, dRing and their homologs in plants and animals. We propose that PcG gene expansion and diversification contributed significantly to the complexity of heritable gene repression mechanisms in extant multicellular organisms.

PcG proteins regulate cellular memory

Epigenetic memory of gene expression profiles is believed to be crucial for the development of multicellular organisms. Polycomb Group (PcG) proteins were originally identified in *Drosophila melanogaster* as factors necessary to maintain cell-fate decisions throughout embryogenesis by repressing *Hox* genes in a body-segment-specific manner (Figure 1a,b) [1–3]. Now recognized as a large family of chromatin-associated proteins conserved from plants to humans, the PcG is involved in many cellular memory processes including body patterning (Figure 1a–f), X inactivation in female mammals [4] and vernalization in plants [5].

Here, we trace the evolutionary history of PcG proteins and speculate on their role(s) in the evolution of vertebrates (Figure 2). To underscore the significance and complexity of the PcG family, we first present the biochemical activities of PcG complexes and the breadth of PcG target genes in the context of development and cellular differentiation.

PcG complexes

PcG proteins are structurally and functionally diverse and form large multimeric complexes of two general types: Polycomb repressive complex 1 (PRC1) and PRC2 [6] (Figure 1g). These complexes post-translationally modify histone tails and are believed to cooperate in transcriptional repression of target genes by altering local, higher-order chromatin structure [7–9]. We describe *Drosophila* PRCs as representative of complexes in other organisms to be discussed below.

PRC2 contains four core proteins: Enhancer of zeste [E(z)], Extra sex combs (Esc), Suppressor of zeste 12 [Su(z)12], and Nucleosome remodeling factor 55-kDa subunit (Nurf55) (Figure 1g) [10]. E(z), a histone methyltransferase, catalyzes the trimethylation of histone H3 lysine 27 (H3K27me3) via its SET domain [10]. Interestingly, E(z) is catalytically inactive *in vitro* unless associated with other PRC2 complex members, which are responsible for either binding histones/nucleosomes or enhancing enzymatic activity [10–12]. Although H3K27me3 is associated with PcG-mediated transcriptional repression and PRC1 binding (see below), the targeting, readout and inheritance of this covalent modification remains unclear.

Core PRC1 is composed of Polycomb (Pc), dRing, Posterior sex combs (Psc) and Polyhomeotic (Ph) (Figure 1g) [6]. Pc has an N-terminal chromodomain (CD) and a C-terminal Pc box (Figure 3). CDs are found in many chromatin-associated proteins and are well-characterized methyllysine-binding modules [13]. Specifically, the CD of *Drosophila* Pc binds most strongly to H3K27me3, the modification generated by PRC2 [14,15]. The Pc box is a ~15-amino acid motif necessary for transcriptional repression of target genes and for interaction with dRing, the catalytically active subunit of PRC1 [16,17]. dRing, named for its RING-type zinc finger (Figure 3), is an E3 ubiquitin ligase that monoubiquitylates histone H2A at lysine 119 (H2AK119ub) [18]. This modification, along with H3K27me3, is important for PcG-mediated gene repression [18–19]. Psc is also a RING finger protein (Figure 3), and a murine Psc homolog enhances the ubiquitylation activity of dRing homologs *in vitro* and *in vivo* [20,21]. The precise function of Ph in PRC1 complexes remains to be characterized, but it has been speculated that Ph might promote the spreading of PcG complexes [22].

Purified PRCs also include other PcG proteins, different PcG isoforms, DNA-binding proteins, transcription factors and chromatin-modifying enzymes, such as histone

Glossary

Extant: still in existence, not extinct.
Homolog: a gene related to a second gene by descent from a common ancestral DNA sequence.
Key domains: domains with known catalytic activity, domains conserved between fly and mouse, and annotated domains from the SMART database.
Paralog: genes related by duplication within a genome.
Phylogenetics: the study of evolutionary history of a group of organisms.
Sequence identity: the percentage of identical matches between the two sequences over the reported aligned region.
Sequence similarity: the percentage of matches between the two sequences over the reported aligned region where a substitution scoring matrix, BLOSUM62, is used.

Corresponding author: Bernstein, E. (bernste@rockefeller.edu).
* These authors contributed equally.
Available online 7 September 2007.

 0168-9525/$ – see front matter doi:10.1016/j.tig.2007.08.006

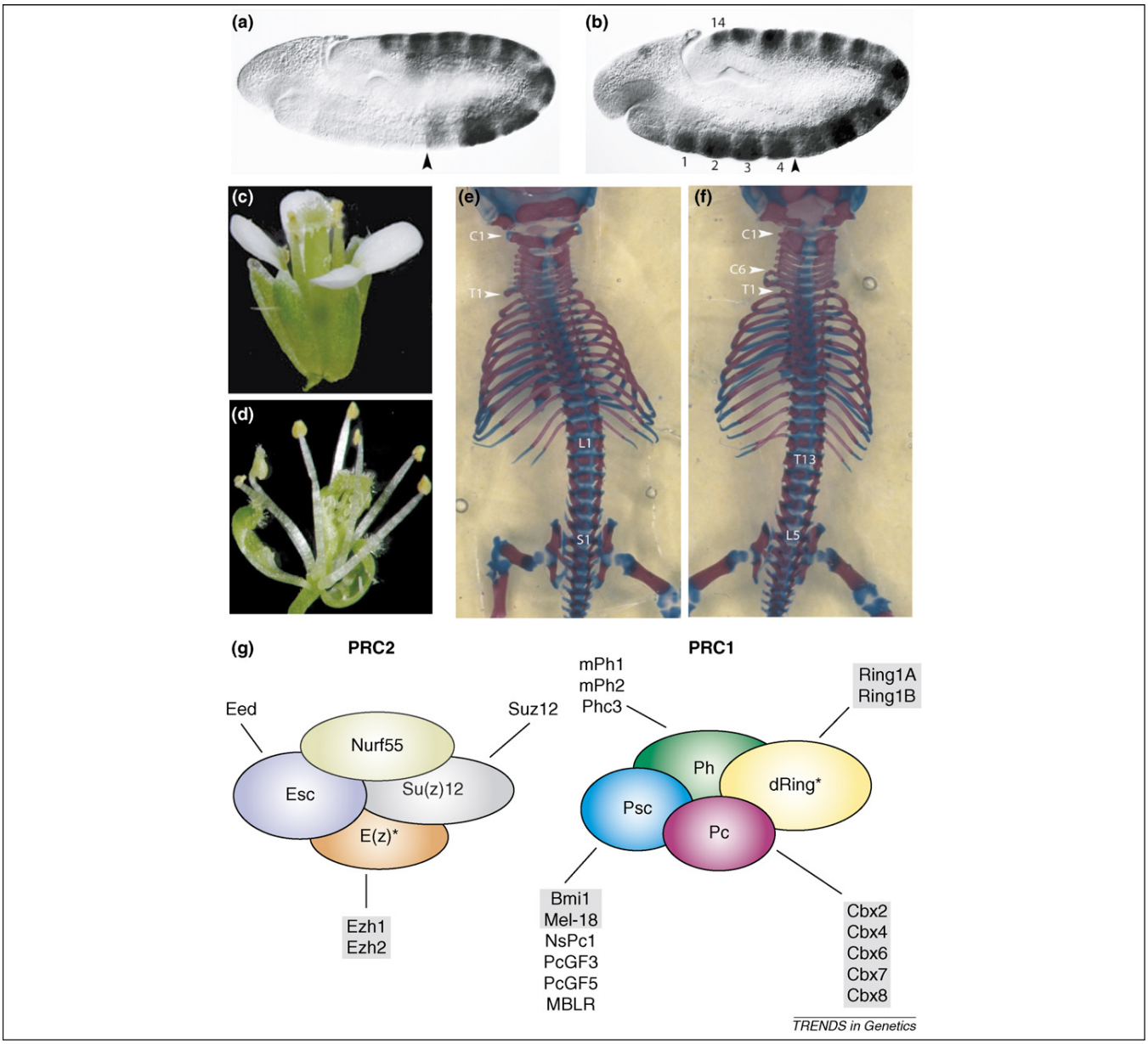

Figure 1. Homeotic transformations in PcG mutants. **(a,b)** *Drosophila* embryos, **(c,d)** *Arabidopsis* flowers and **(e,f)** mouse skeletons. Staining for the *Drosophila* Hox protein Ubx in **(a)** wild-type and **(b)** *Su(z)12* mutant embryos. **(a,b)** Arrowheads mark parasegment 5, the anterior limit of Ubx expression in wild-type embryos. **(b)** Numbers indicate de-repression of Ubx expression in parasegments 1–4 and 14 in *Su(z)12* mutant embryos. *Arabidopsis* flowers from **(c)** wild-type and **(d)** *FIE* (plant *Esc* homolog)-deficient plants. **(c)** Four distinct flower organs (sepal, petal, stamen, and carpel) are arranged in concentric circles in the wild-type flower. **(d)** The terminal flower on the stem of the *FIE*-deficient plant lacks petals and the sepals are transformed into carpeloid-like organs. Distal view of axial skeleton from **(e)** wild-type and **(f)** *Cbx2*-deficient mice. Note the following axial–skeletal posterior transformations in *Cbx2*-deficient mice: the last cervical vertebra C6 in *Cbx2*-deficient mice to the first thoracic vertebra T1 in wild-type; the last thoracic vertebra T13 to the first lumbar vertebra L1; and the lumbar vertebra L5 to the first sacral vertebra S1. **(g)** Schematic representation of known core members of PRC1 and PRC2 complexes. Asterisks represent catalytically active PRC subunits. *Drosophila* proteins are shown as colored ovals; mouse homologs of these proteins are listed adjacently. Gray boxes denote the mammalian homologs discussed in detail throughout the text. *Drosophila* images reproduced with permission from Ref. [72]; *Arabidopsis* images reproduced with permission from Ref. [43]; mouse images reproduced with permission from Ref. [73].

deacetylases [23–27]. Such factors probably contribute to transcriptional repression by modifying chromatin structure or by stabilizing PRC complexes at their target genes. For example, in *Drosophila*, sequence-specific DNA-binding proteins, such as Pleiohomeotic (Pho), interact with PRC subunits and can induce PRC binding to DNA regulatory elements known as PREs (PcG response elements) [6,28]. Although PREs are necessary and sufficient for PcG recruitment in *Drosophila*, such elements have yet to be identified in vertebrates or plants.

PcG target genes

Evidence is accumulating that many PRC subcomplexes exist *in vivo* [29]. For example, *Drosophila* polytene chromosome stainings for PcG proteins show non-identical patterns, and chromatin immunoprecipitation (ChIP) experiments support the hypothesis that PRC composition varies at different *Hox* genes [30,31]. Moreover, although *Drosophila* PcG mutants and mice deficient for individual PRC1 members share similar homeotic defects, each also has unique phenotypes, suggesting that different PRC

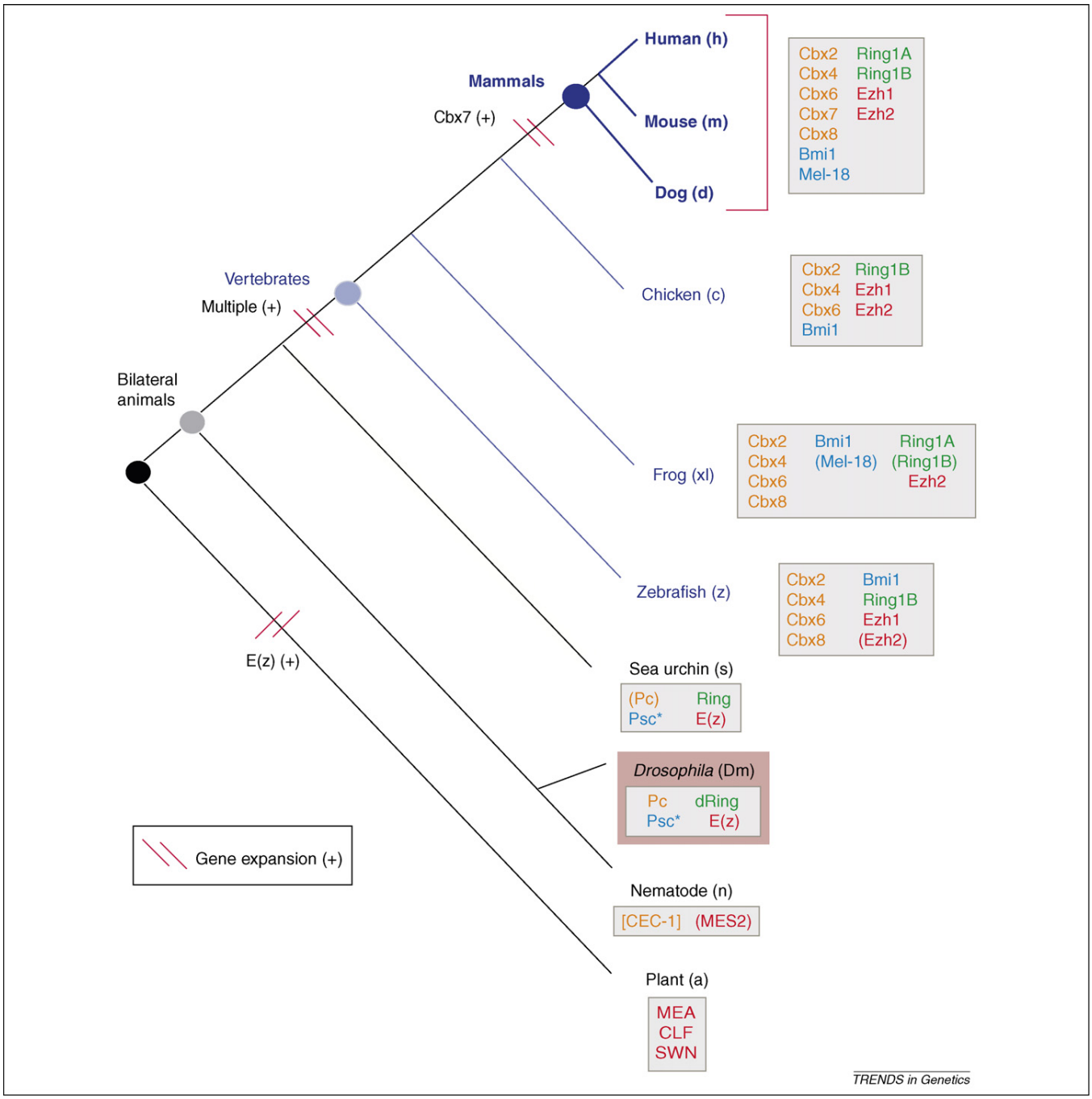

Figure 2. Phylogenetic representation of selected organisms and their PcG homologs. A phylogenetic tree of selected model organisms from plants to humans is shown (adapted from http://www.tolweb.org/tree/). This tree illustrates that PcG-encoding genes have undergone multiple duplication events through evolution; the most dynamic period appears to be during the evolution of vertebrates from invertebrates. PRC1 components seem to have been lost in *C. elegans*. However, CEC-1 might be a functional PRC1 homolog (see text for details). *Drosophila* proteins, used (here and in the text) as our PcG reference set, are highlighted in the red box. Shaded boxes next to each organism display homologs of E(z) (red), Pc (orange), Psc (blue) and dRing (green) proteins in each organism. Red slash marks represent probable gene expansion events. The black and grey nodes represent the common ancestor of all selected model organisms and extant bilateral animals, respectively. The light blue and dark blue nodes denote the common ancestor of extant vertebrate and mammalian species, respectively. Parentheses denote proteins with biochemically or genetically defined PcG activity but lacking sufficient sequence conservation with the *Drosophila*, mouse or human proteins to be predicted as homologs by our methods. Brackets indicate putative PcG proteins that were identified by sequence similarity but that need to be confirmed functionally. Asterisks represent proteins that might have multiple (putative) paralogs within a given organism. Note that branch lengths do not represent evolutionary distance between organisms.

subcomplexes exist and have at least some non-redundant target genes [30,32–34].

Although *Hox* genes are the most intensely studied PcG target genes, recent genome-wide ChIP-chip studies have reinforced the role of PcG proteins in development and highlighted their role in pluripotency and cell-fate decisions [35–39]. In mouse and human embryonic stem (ES) cells, PcG proteins are bound to promoters of hundreds of transcriptionally silent genes, including transcription factors, morphogens, receptors and signaling proteins [35,36]. Notably, PcG proteins are also bound to a small subset of actively transcribed genes, perhaps priming them

for future repression [37,40]. Such studies allow evolutionary analysis of PcG target genes and might aid in the identification of DNA sequences that serve as PcG-binding sites in vertebrates [28,37].

Evolutionary history of PcG proteins

To examine the evolutionary history of PcG proteins, we bioinformatically searched for homologs of *Drosophila* PcG proteins in diverse multicellular organisms (Figures 2–4). Our PcG reference set was composed of well-characterized representatives of PRC1 and PRC2: E(z) Pc, Psc and dRing. A sequence-similarity method between domains/motifs and full-length sequence was used to identify PcG homologs (see the supplementary material online and Glossary) that were mapped onto a phylogenetic tree of representative organisms from plant and bilateral animal kingdoms (Figure 2). This tree offers intriguing insights into the evolution of PcG proteins and allows us to speculate about the potential role(s) of PcG proteins in animal evolution.

A common mechanism of evolution is gene duplication and subsequent divergence of coding sequences or regulatory elements. Based on our analysis, PcG genes are likely to have undergone multiple duplication events in their evolutionary history (Figure 2). Perhaps the most dynamic period was during the evolution of vertebrates from invertebrate ancestors. The extant invertebrates *Drosophila* and sea urchin have single copies of the PcG proteins in our reference set, with the exception of Psc (Figure 2; see the supplementary material online). By contrast, vertebrate species have multiple paralogs of most PcG members (Figures 2–4). One striking example of PcG expansion is the Pc family. Represented by a single gene in invertebrates, there are up to five Pc homologs in vertebrates with differences in domain structure and biochemical properties (see below; Figures 2,3).

Remarkably, *Hox* genes also expanded in the evolution of vertebrates. *Hox* genes are typically organized into genomic clusters; sea urchin (like *Drosophila*) has only one *Hox* gene cluster, whereas vertebrates have multiple clusters found on separate chromosomes [41]. Since *Hox* genes are conserved PcG targets and key determinants of body-plan specification in bilateral animals, it is possible that after genomic duplication of *Hox* and PcG genes, they diversified and co-evolved to fulfill new roles in the development of vertebrate body plans. Furthermore, although plant genomes do not encode *Hox* genes, PcG proteins repress other key developmental regulators in *Arabidopsis thaliana*, suggesting that the role of PcG in development is conserved from plants to humans [42,43].

Of the four proteins in our PcG set, the *Arabidopsis* genome only has E(z) homologs (Figures 2,4) [28,42]. In *Arabidopsis*, PcG-mediated transcriptional repression requires H3K27me3 [44], however, homologs of PRC1 have not been found in plants. What protein(s) interprets H3K27me? Recently a CD-containing protein, LIKE HETEROCHROMATIN PROTEIN 1 (LHP1), was shown to bind H3K27me3 *in vitro* and associate with H3K27me3 chromatin domains *in vivo* [45,46]. Interestingly, LHP1 is required for the epigenetic maintenance of vernalization ('memory' of winter that promotes flowering in spring), a process mediated in part by PcG [47,48]. Although the sequence of LHP1 is more homologous to HP1-like proteins (also transcriptional repressors) than to Pc, these data suggest that LHP1 cooperates with plant PRC2 complexes. Whether LHP1 exists in a PRC1-like complex is of particular interest.

As in plants, only the presence of PRC2 homologs has been verified in *Caenorhabditis elegans*. However, the common ancestor of bilateral animals had PRC1-like proteins (Figure 2). Our bioinformatic analysis uncovered a putative Pc homolog in *C. elegans*, previously identified as *C. elegans* chromobox 1 (CEC-1) [49]. Little is known about CEC-1 except that it localizes exclusively to somatic nuclei and dissociates from chromosomes at mitosis [49]. The domain structure of CEC-1 supports its classification as a Pc homolog: an N-terminal CD and a C-terminal Pc box (as well as a second putative Pc box after the CD) (Figures 3,5a; see the supplementary material online). Although the sequence similarity between full-length CEC-1 and dPc or dHP1 are equal (27%), CEC-1 lacks two important sequence characteristics of HP1 proteins: a chromo-shadow domain and a stretch of glutamic acid residues N-terminal of the CD (Figure 3). The possibility that CEC-1 might regulate H3K27me-dependent gene repression in the worm soma is intriguing but requires further investigation.

Conservation and diversification of PcG homologs

Key domains of PcG homologs are highly conserved between evolutionarily distant organisms and among paralogs in a given organism (typically >75% amino acid similarity) (Figures 3–5). However, outside of key domains, PcG proteins have diverged significantly from their *Drosophila* counterparts and from their paralogs (Figures 3,4,5b). Accumulating evidence suggests that PcG paralogs have specialized expression patterns and functions (see below) [49–51]. An important challenge is to understand the functional significance of developmentally regulated expression of paralogs and how this impacts PRC composition, genomic targeting and/or mechanism of transcriptional repression. To illustrate the functional diversification of PcG paralogs, we continue to focus on homologs of *Drosophila* Pc, Psc, dRing and E(z).

Pc homologs

Vertebrate model organisms have between three and five Pc homologs [known as Chromobox (Cbx)], which all have highly conserved CDs and Pc boxes (Figures 2,3,5b). However, paralogs differ greatly in length and in the presence of other domains and motifs; these factors might contribute to differential function (Figure 3,5b).

Mammalian Cbx proteins differentially effect cell-cycle regulation. The overexpression of Cbx7 or Cbx8 but not Cbx4 (Cbx2 and Cbx6 were not tested) bypasses replicative senescence in human and mouse fibroblasts owing in part to the repression of the *INK4a-ARF* locus [52,53]. However, they do so in the context of distinct PRC1 complexes; Cbx8 depends on Bmi1 (Psc homolog, see below) to bind the *INK4a-ARF* locus and to extend lifespan, whereas Cbx7-mediated bypass of senescence is Bmi1-independent [52,53]. It remains to be shown if Cbx7-containing PRC1 complexes preferentially contain another Psc homolog

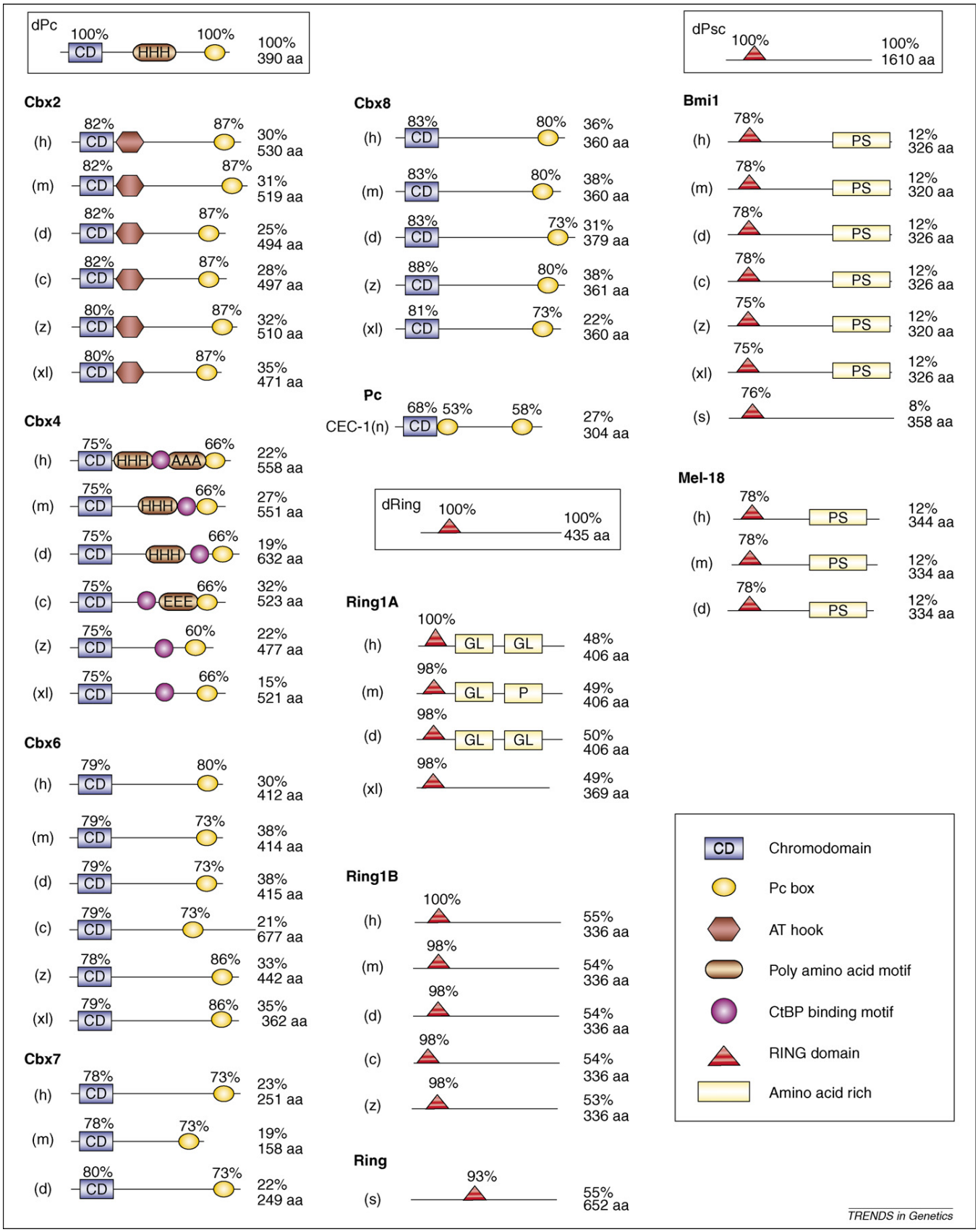

Figure 3. Domain and motif structure of selected PRC1 proteins. Comparison of domain and motif structure of selected PRC1 proteins in *Drosophila* (boxed), human, mouse, dog, chicken, zebrafish, frog, nematode and sea urchin. Protein lengths are scaled exclusively within homolog groups, not among paralog groups, and are represented by a black line. Numbers shown above domains are percentage similarity to the domain in the *Drosophila* homolog. Numbers to the right of proteins represent percentage similarity to the full-length *Drosophila* sequence and the number of amino acids in the protein. Note the high percentage similarity between domains, but low

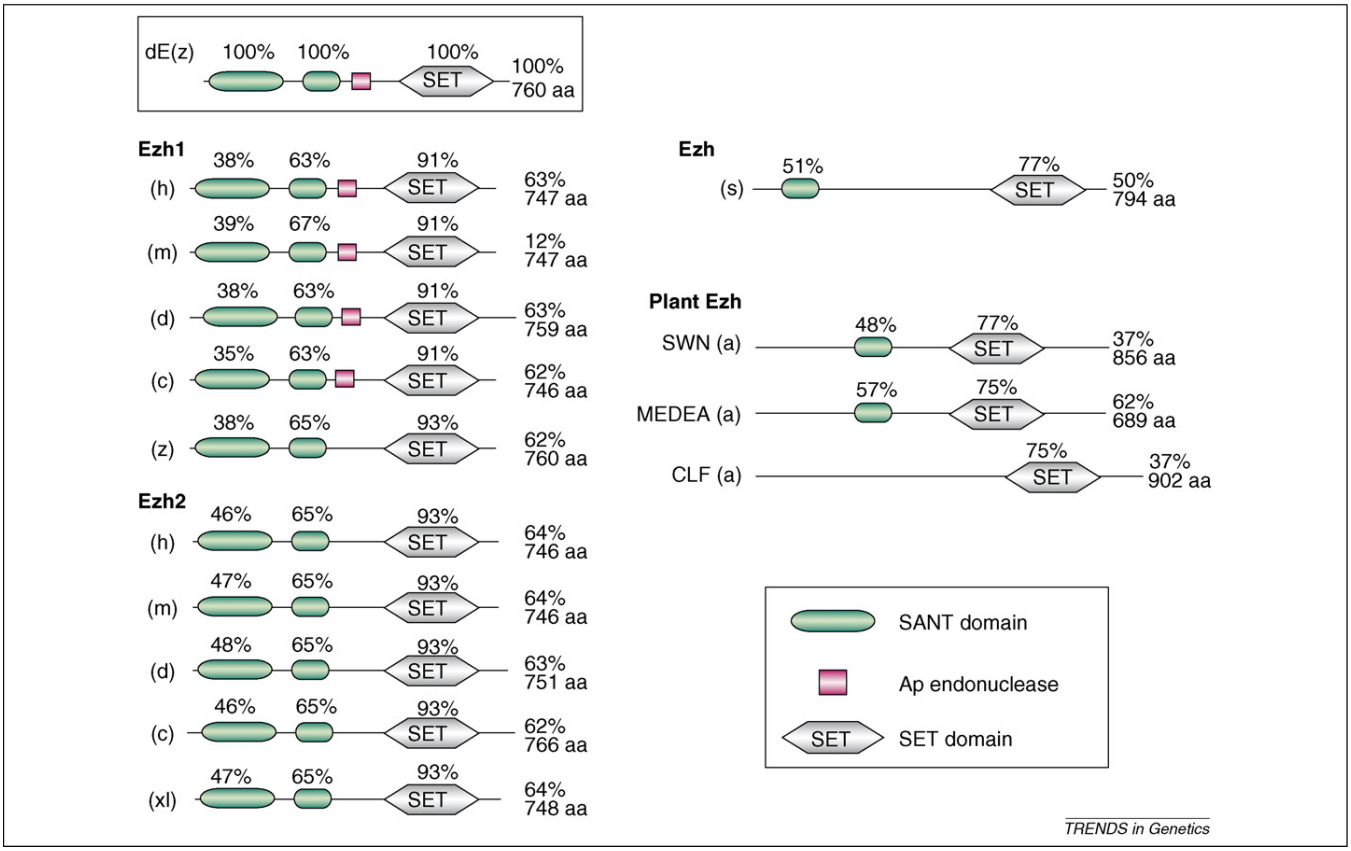

Figure 4. Domain and motif structure of selected PRC2 proteins. Comparison of domain and motif structure of selected PRC2; see Figure 3 for details. E(z) homologs are grouped based on their sequence similarity to mouse Ezh1 and Ezh2. *Arabidopsis* E(z) homologs are listed separately on the right. Abbreviations: a, plant; c, chicken; d, dog; h, human; m, mouse; s, sea urchin; xl, frog; z, zebrafish.

required for Cbx7-mediated repression of this tumor suppressor locus.

Cbx proteins also specifically interact with non-PcG proteins. Cbx4 is the only member of the family that binds the transcriptional co-repressor C-terminal binding protein (CtBP) [54]. CtBP interacts with transcription factors that might target Cbx4-containing complexes to specific DNA sequences. Cbx4 is also unique among Pc homologs as an E3 SUMO ligase [55]. The full range of Cbx4 SUMO targets is unknown, but the sumoylation of several transcriptional regulators, including CtBP, is enhanced by Cbx4 [55,56]. Additionally, recent biochemical data suggest that the five mammalian Cbx proteins have different histone-binding preferences: the Cbx CDs bind differentially to H3K27me3 and H3K9me3, unlike *Drosophila* Pc CD, which prefers H3K27me3 [57].

Psc homologs

Mel-18 and Bmi1 (two of six Psc homologs in mammals; Figure 1g) are also likely to be non-redundant paralogs, despite their 63% amino acid sequence identity. Bmi1- and Mel-18-deficient mice display similar but unique phenotypes [33,34], and only ~30% of Bmi1-regulated genes were found to be co-regulated by Mel-18 and vice-versa [58]. Additionally, in some cases, Mel-18 and Bmi1 have opposite effects on cell-cycle regulation. Retroviral insertion of *Bmi1* into Eμ-*myc* transgenic mice accelerated lymphomagenesis (oncogenic) [59], whereas cells from transgenic mice overexpressing Mel-18 arrest in G1/S of the cell cycle (tumor-suppressor function) [7]. Moreover, stable Mel-18-knockdown fibroblasts induce tumor formation when injected into mice [60,61].

However, the relationship between these two paralogs is complex. A recent study using stable knockdown (RNAi) of Bmi1 and Mel-18 in cancer cell lines revealed similar, not opposing consequences on cell growth [58]. Furthermore, Mel-18 can act as a tumor suppressor by repressing Bmi1 in cancer cells [62]. Resolving these discrepancies is particularly important for drug therapy.

Although Bmi1 and Mel-18 form stable PRC1 complexes with similar composition, only Bmi1 has been shown to positively regulate Ring1B ubiquitylation of H2AK119 [20]. A recent structural study suggested that Bmi1 enhances Ring1B activity by stabilizing the interaction

percentage similarity of full-length sequence, between the *Drosophila* protein and its homologs in other organisms. Also note different amino acid lengths of paralog groups (e.g. Cbx4 versus Cbx7) and different domain structure (e.g. Cbx2 versus Cbx4). See text for details. Pc homologs are grouped based on their sequence similarity to mouse Cbx2, Cbx4, Cbx6, Cbx7 and Cbx8. Putative *C. elegans* Pc protein is shown at bottom (CEC-1). Psc homologs are grouped based on their sequence similarity to mouse Bmi1 and Mel-18. Homologs of dRing are grouped based on their sequence similarity to mouse Ring1A and Ring1B. Sea urchin Ring is equally similar to mouse Ring1A and Ring1B, and is listed separately at the bottom. Abbreviations: c, chicken; d, dog; h, human; m, mouse; n, nematode; s, sea urchin; xl, frog; z, zebrafish.

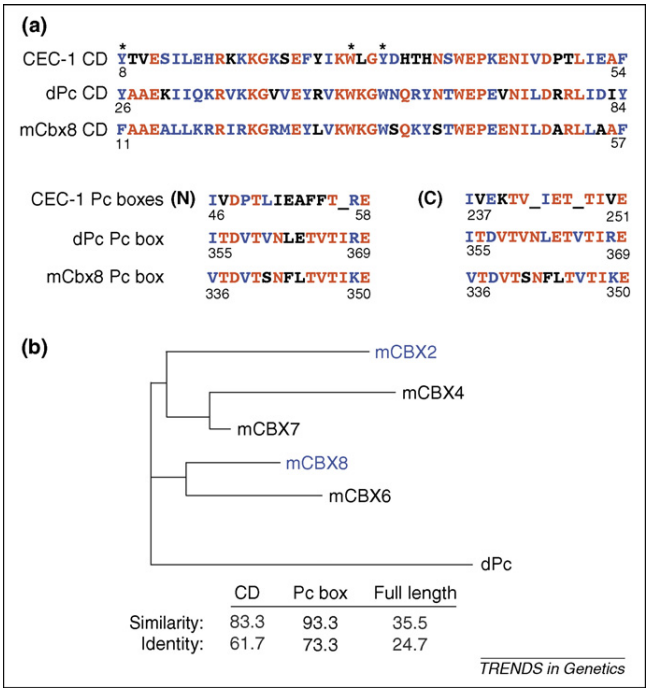

Figure 5. CEC-1 sequence alignment and phylogeny of mouse Cbx proteins. **(a)** Sequence alignment is shown between the chromodomains (CD) and Pc boxes of dPc, mouse Cbx8 and CEC-1 proteins. An asterisk represents aromatic residues within the chromodomains that are required for histone methyllysine binding. Pc-box-like features are found in both the N-terminal and C-terminal regions of CEC-1 and are aligned individually to the Pc boxes of dPc and mouse Cbx8. (N) and (C) represent the N-terminal and C-terminal Pc-box-like features of CEC-1, respectively. Amino acids highlighted in red represent residues that are identical to each other, and those highlighted in blue represent residues that are evolutionarily similar. Overall, our analysis suggests that CEC-1 might represent a Pc homolog, rather than a HP1 homolog. However, this remains to be rigorously tested. **(b)** Pairwise sequence similarities were calculated between all mouse Pc proteins (Cbx 2, Cbx4 and Cbx6–Cbx8) and an unrooted neighbor-joining tree was constructed using the PHYLIP software program (http://evolution.genetics.washington.edu/phylip.html). Evolutionary distance between paralogs is represented by the tree branches, which are drawn to scale. The table below shows percentage sequence identity and percentage sequence similarity between the chromodomains, Pc boxes and the full length sequences of Cbx2 and Cbx8 proteins. These proteins were selected for comparison because of their comparable length and evolutionary distance. Note the high similarity and identity between key domains, but significantly less similarity and identity in the full length sequences.

between the E2 ligase, Ring1B (the E3 ligase) and the substrate (H2A) [63]. The authors propose a potential H2A-binding surface on Bmi1 that contains residues not conserved in Mel-18 [63]. Perhaps Mel-18 does not enhance Ring1B activity because it lacks this binding surface. Additionally, Bmi1 directs self-ubiquitylation of Ring1B on specific lysines to generate atypical mixed chains that are necessary for H2A monoubiquitylation (H2Aub) [64]. Therefore, the presence of different Psc homologs in PRC1 might change its catalytic activity *in vivo*.

dRing homologs

Vertebrate homologs of dRing, Ring1A and Ring1B, also exhibit some functional divergence (Figures 1g,2,3). Although they share long stretches of high conservation, Ring1A- and Ring1B-deficient mice have drastically different phenotypes [65,66]. Mice heterozygous for Ring1A exhibit classic homeotic transformations and skeletal defects [65], whereas Ring1B heterozygous mice show no skeletal phenotype [66]. However, Ring1B is essential for normal gastrulation, and null embryos do not survive past embryo day 10.5 [66]. The differential severity of these phenotypes correlates with the extent of H2Aub depletion in these knockouts. Global H2Aub is drastically reduced in Ring1B- but not Ring1A-null ES cells [19].

Biochemical work supports the differential functions revealed in these mouse studies. Full-length recombinant Ring1B but not Ring1A has ubiquitin-ligase activity for H2A (unlike full-length protein, N-terminal Ring1A has E3 activity [21]), and Ring1B association with Ring1A enhances this activity [18,20]. However, the *in vivo* situation is probably more nuanced than the *in vitro* studies suggest. Although global H2Aub is drastically reduced in Ring1B-null cells, H2Aub staining is maintained on the inactive X chromosome [19]; only in Ring1B/Ring1A double-knockout cells is H2Aub lost from this structure [19], revealing functional redundancy of Ring1A and Ring1B in some contexts.

E(z) homologs

The mammalian organisms that we focused on have two E(z) homologs: Ezh1 and Ezh2 (Figures 1g,2,4). Little is known about the functional differences between these paralogs in mammals, but the ancestral *E(z)* gene also expanded in plant lineages (Figure 2). *Arabidopsis* has three *E(z)* homologs: *MEDEA* (*MEA*), *CURLY LEAF* (*CLF*), and *SWINGER* (*SWN*) with largely non-overlapping patterns of expression [67]. These paralog proteins regulate different developmental processes and non-identical sets of genes [68–70]. Although similar in domain structure (Figure 4), ectopic expression of *MEA* or *SWN* cannot rescue *CLF*-deficient plants, suggesting some functional divergence of paralogs [69]. Additionally, phylogenetic analysis of E(z) homolog SET domains in a variety of angiosperm species revealed that *Arabidopsis* CLF, SWN and MEA proteins cluster into three separate clades [69]. This suggests that, since an ancient gene duplication, E(z) homologs have diverged and become fixed in diverse plant species, presumably for specialized, beneficial functions [69].

Conclusions and extensions

The phylogenetic tree (Figure 2) and discussion presented here argue that PcG genes underwent multiple duplication events in the evolution of plants and animals. We favor the view that these extra genomic copies diverged in sequence resulting in differential functions or expression patterns conferring fitness advantages. Paralogs in extant species might not only differentially associate with other PcG and non-PcG proteins but might also regulate different target genes. However, this needs to be rigorously tested. Given that the evolution of complex traits (including body plan) are thought to be driven largely by changing the expression of developmental regulators, the expansion of PcG transcriptional repressors and their subsequent diversification might have contributed to the drastic adaptive radiation that occurred during the evolution of vertebrates.

Here, we focused on PcG gene expansion events, but PcG genes were probably also lost from genomes. Although fungi are expected to have E(z) homolog(s) given their common ancestor with plants and animals, PcG proteins

have yet to be identified in unicellular fungi, such as yeast. However, *Neurospora crassa*, a filamentous fungus, has an E(z) homolog [28]. Given that PcG proteins are critical for cellular differentiation, we speculate that the E(z) gene in *Neurospora* was retained because it contributes to multicellular developmental stages of this organism. PRC1 genes also seem to have been lost during some branches of animal evolution, in *C. elegans* for example (Figure 2). Others have suggested that this might be due to fragmentation and rearrangement of the *Hox* clusters, which occurred multiple times in evolution [28].

Although beyond the scope of this article, PcG-mediated transcriptional repression is balanced by an antagonistic group of chromatin-modifying complexes called the trithorax Group (trxG), which has also expanded and functionally diversified in mammals [71]. It is likely that PcG, trxG and their common gene targets (including *Hox* genes) have expanded and co-evolved, influencing the evolution of increasingly complex mechanisms of heritable gene repression. Finally, as a field, we have only scratched the surface of the functional consequences of PcG paralog divergence, and we look forward to studies aimed at addressing these questions.

Acknowledgements

We thank T. Milne and A. Tanay for critical reading and advice on this manuscript and Y. Katoh-Fukui (National Institute for Basic Biology, Aichi, Japan), J. Mueller (EMBL, Heidelberg, Germany) and N. Ohad (Tel Aviv University, Tel Aviv, Israel) for high-resolution images in Figure 1. We apologize to all those whose work we would not cite due to space constraints. Funding for S.J.W. is provided by the Graduate Research Fellowship Program of the NSF and NIH Merit Grant GM 53512 to C.D.A.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.tig.2007.08.006.

References

1 Kennison, J.A. (1995) The Polycomb and trithorax group proteins of *Drosophila*: trans-regulators of homeotic gene function. *Annu. Rev. Genet.* 29, 289–303
2 Jurgens, G. (1985) A group of genes controlling the spatial expression of the bithorax complex in *Drosophila*. *Nature* 316, 153–155
3 Schwartz, Y.B. and Pirrotta, V. (2007) Polycomb silencing mechanisms and the management of genomic programmes. *Nat. Rev. Genet.* 8, 9–22
4 Heard, E. (2004) Recent advances in X-chromosome inactivation. *Curr. Opin. Cell Biol.* 16, 247–255
5 Sung, S. and Amasino, R.M. (2004) Vernalization and epigenetics: how plants remember winter. *Curr. Opin. Plant Biol.* 7, 4–10
6 Ringrose, L. and Paro, R. (2004) Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins. *Annu. Rev. Genet.* 38, 413–443
7 Sparmann, A. and van Lohuizen, M. (2006) Polycomb silencers control cell fate, development and cancer. *Nat. Rev. Cancer* 6, 846–856
8 Francis, N.J. *et al.* (2004) Chromatin compaction by a polycomb group protein complex. *Science* 306, 1574–1577
9 Shao, Z. *et al.* (1999) Stabilization of chromatin structure by PRC1, a Polycomb complex. *Cell* 98, 37–46
10 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
11 Nekrasov, M. *et al.* (2005) Nucleosome binding and histone methyltransferase activity of *Drosophila* PRC2. *EMBO Rep.* 6, 348–353
12 Tie, F. *et al.* (2007) The N-terminus of *Drosophila* ESC binds directly to histone H3 and is required for E(Z)-dependent trimethylation of H3 lysine 27. *Mol. Cell. Biol.* 27, 2014–2026
13 Eissenberg, J.C. (2001) Molecular biology of the chromo domain: an ancient chromatin module comes of age. *Gene* 275, 19–29
14 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
15 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
16 Muller, J. (1995) Transcriptional silencing by the Polycomb protein in *Drosophila* embryos. *EMBO J.* 14, 1209–1220
17 Schoorlemmer, J. *et al.* (1997) Ring1A is a transcriptional repressor that interacts with the Polycomb-M33 protein and is expressed at rhombomere boundaries in the mouse hindbrain. *EMBO J.* 16, 5930–5942
18 Wang, H. *et al.* (2004) Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878
19 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
20 Cao, R. *et al.* (2005) Role of Bmi-1 and Ring1A in H2A ubiquitylation and *Hox* gene silencing. *Mol. Cell* 20, 845–854
21 Buchwald, G. *et al.* (2006) Structure and E3-ligase activity of the Ring-Ring complex of polycomb proteins Bmi1 and Ring1b. *EMBO J.* 25, 2465–2474
22 Kim, C.A. *et al.* (2002) The SAM domain of polyhomeotic forms a helical polymer. *Nat. Struct. Biol.* 9, 453–457
23 Tie, F. *et al.* (2001) The *Drosophila* Polycomb Group proteins ESC and E(Z) are present in a complex containing the histone-binding protein p55 and the histone deacetylase RPD3. *Development* 128, 275–286
24 Kim, S.Y. *et al.* (2006) Juxtaposed Polycomb complexes co-regulate vertebral identity. *Development* 133, 4957–4968
25 Saurin, A.J. *et al.* (2001) A *Drosophila* Polycomb group complex includes Zeste and dTAFII proteins. *Nature* 412, 655–660
26 Kuzmichev, A. *et al.* (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* 14, 183–193
27 Kuzmichev, A. *et al.* (2005) Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1859–1864
28 Schuettengruber, B. *et al.* (2007) Genome regulation by polycomb and trithorax proteins. *Cell* 128, 735–745
29 Jacobs, J.J. and van Lohuizen, M. (2002) Polycomb repression: from cellular memory to cellular proliferation and cancer. *Biochim. Biophys. Acta* 1602, 151–161
30 Santamaria, P. (1998) Genesis versus epigenesis: the odd jobs of the Polycomb group of genes. *Int. J. Dev. Biol.* 42, 463–469
31 Strutt, H. and Paro, R. (1997) The polycomb group protein complex of *Drosophila melanogaster* has different compositions at different target genes. *Mol. Cell. Biol.* 17, 6773–6783
32 Takihara, Y. *et al.* (1997) Targeted disruption of the mouse homologue of the *Drosophila* polyhomeotic gene leads to altered anteroposterior patterning and neural crest defects. *Development* 124, 3673–3682
33 Akasaka, T. *et al.* (1996) A role for mel-18, a Polycomb group-related vertebrate gene, during the anteroposterior specification of the axial skeleton. *Development* 122, 1513–1522
34 van der Lugt, N.M. *et al.* (1994) Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the bmi-1 proto-oncogene. *Genes Dev.* 8, 757–769
35 Boyer, L.A. *et al.* (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* 441, 349–353
36 Lee, T.I. *et al.* (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* 125, 301–313
37 Bracken, A.P. *et al.* (2006) Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev.* 20, 1123–1136
38 Tolhuis, B. *et al.* (2006) Genome-wide profiling of PRC1 and PRC2 Polycomb chromatin binding in *Drosophila melanogaster*. *Nat. Genet.* 38, 694–699
39 Schwartz, Y.B. *et al.* (2006) Genome-wide analysis of Polycomb targets in *Drosophila melanogaster*. *Nat. Genet.* 38, 700–705

40 Papp, B. and Muller, J. (2006) Histone trimethylation and the maintenance of transcriptional ON and OFF states by trxG and PcG proteins. *Genes Dev.* 20, 2041–2054
41 Finnerty, J.R. and Martindale, M.Q. (1998) The evolution of the *Hox* cluster: insights from outgroups. *Curr. Opin. Genet. Dev.* 8, 681–687
42 Schubert, D. *et al.* (2005) Epigenetic control of plant development by Polycomb-group proteins. *Curr. Opin. Plant Biol.* 8, 553–561
43 Katz, A. *et al.* (2004) FIE and CURLY LEAF polycomb proteins interact in the regulation of homeobox gene expression during sporophyte development. *Plant J.* 37, 707–719
44 Schubert, D. *et al.* (2006) Silencing by plant Polycomb-group genes requires dispersed trimethylation of histone H3 at lysine 27. *EMBO J.* 25, 4638–4649
45 Turck, F. *et al.* (2007) *Arabidopsis* TFL2/LHP1 specifically associates with genes marked by trimethylation of histone H3 lysine 27. *PLoS Genet.* 3, e86
46 Zhang, X. *et al.* (2007) The *Arabidopsis* LHP1 protein colocalizes with histone H3 Lys27 trimethylation. *Nat. Struct. Mol. Biol.* DOI: 10.1038/nsmb1283
47 Sung, S. *et al.* (2006) Epigenetic maintenance of the vernalized state in *Arabidopsis thaliana* requires LIKE HETEROCHROMATIN PROTEIN 1. *Nat. Genet.* 38, 706–710
48 Mylne, J.S. *et al.* (2006) LHP1, the *Arabidopsis* homologue of HETEROCHROMATIN PROTEIN1, is required for epigenetic silencing of FLC. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5012–5017
49 Agostoni, E. *et al.* (1996) *cec-1*, a soma-specific chromobox-containing gene in *C. elegans*. *Dev. Biol.* 178, 316–326
50 Gunster, M.J. *et al.* (2001) Differential expression of human Polycomb group proteins in various tissues and cell types. *J. Cell. Biochem.* 81, 129–143
51 Lessard, J. *et al.* (1998) Stage-specific expression of polycomb group genes in human bone marrow cells. *Blood* 91, 1216–1224
52 Gil, J. *et al.* (2004) Polycomb CBX7 has a unifying role in cellular lifespan. *Nat. Cell Biol.* 6, 67–72
53 Dietrich, N. *et al.* (2007) Bypass of senescence by the polycomb group protein CBX8 through direct binding to the *INK4A-ARF* locus. *EMBO J.* 26, 1637–1648
54 Sewalt, R.G. *et al.* (1999) C-Terminal binding protein is a transcriptional repressor that interacts with a specific class of vertebrate Polycomb proteins. *Mol. Cell. Biol.* 19, 777–787
55 Kagey, M.H. *et al.* (2003) The polycomb protein Pc2 is a SUMO E3. *Cell* 113, 127–137
56 Roscic, A. *et al.* (2006) Phosphorylation-dependent control of Pc2 SUMO E3 ligase activity by its substrate protein HIPK2. *Mol. Cell* 24, 77–89
57 Bernstein, E. *et al.* (2006) Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.* 26, 2560–2569
58 Wiederschain, D. *et al.* (2007) Contribution of polycomb homologues bmi-1 and mel-18 to medulloblastoma pathogenesis. *Mol. Cell. Biol.* 27, 4968–4979
59 Haupt, Y. *et al.* (1991) Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. *Cell* 65, 753–763
60 Kanno, M. *et al.* (1995) mel-18, a Polycomb group-related mammalian gene, encodes a transcriptional negative regulator with tumor suppressive activity. *EMBO J.* 14, 5672–5678
61 Tetsu, O. *et al.* (1998) mel-18 negatively regulates cell cycle progression upon B cell antigen receptor stimulation through a cascade leading to c-myc/cdc25. *Immunity* 9, 439–448
62 Guo, W.J. *et al.* (2007) Mel-18 acts as a tumor suppressor by repressing Bmi-1 expression and down-regulating Akt activity in breast cancer cells. *Cancer Res.* 67, 5083–5089
63 Li, Z. *et al.* (2006) Structure of a Bmi-1-Ring1B polycomb group ubiquitin ligase complex. *J. Biol. Chem.* 281, 20643–20649
64 Ben-Saadon, R. *et al.* (2006) The polycomb protein Ring1B generates self atypical mixed ubiquitin chains required for its *in vitro* histone H2A ligase activity. *Mol. Cell* 24, 701–711
65 del Mar Lorente, M. *et al.* (2000) Loss- and gain-of-function mutations show a polycomb group function for Ring1A in mice. *Development* 127, 5093–5100
66 Voncken, J.W. *et al.* (2003) Rnf2 (Ring1b) deficiency causes gastrulation arrest and cell cycle inhibition. *Proc. Natl. Acad. Sci. U. S. A.* 100, 2468–2473
67 Goodrich, J. *et al.* (1997) A Polycomb-group gene regulates homeotic gene expression in *Arabidopsis*. *Nature* 386, 44–51
68 Makarevich, G. *et al.* (2006) Different Polycomb group complexes regulate common target genes in *Arabidopsis*. *EMBO Rep.* 7, 947–952
69 Chanvivattana, Y. *et al.* (2004) Interaction of Polycomb-group proteins controlling flowering in *Arabidopsis*. *Development* 131, 5263–5276
70 Kohler, C. *et al.* (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. *Genes Dev.* 17, 1540–1553
71 Ruthenburg, A.J. *et al.* (2007) Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Mol. Cell* 25, 15–30
72 Birve, A. *et al.* (2001) Su(z)12, a novel *Drosophila* Polycomb group gene that is conserved in vertebrates and plants. *Development* 128, 3371–3379
73 Katoh-Fukui, Y. *et al.* (1998) Male-to-female sex reversal in M33 mutant mice. *Nature* 393, 688–692