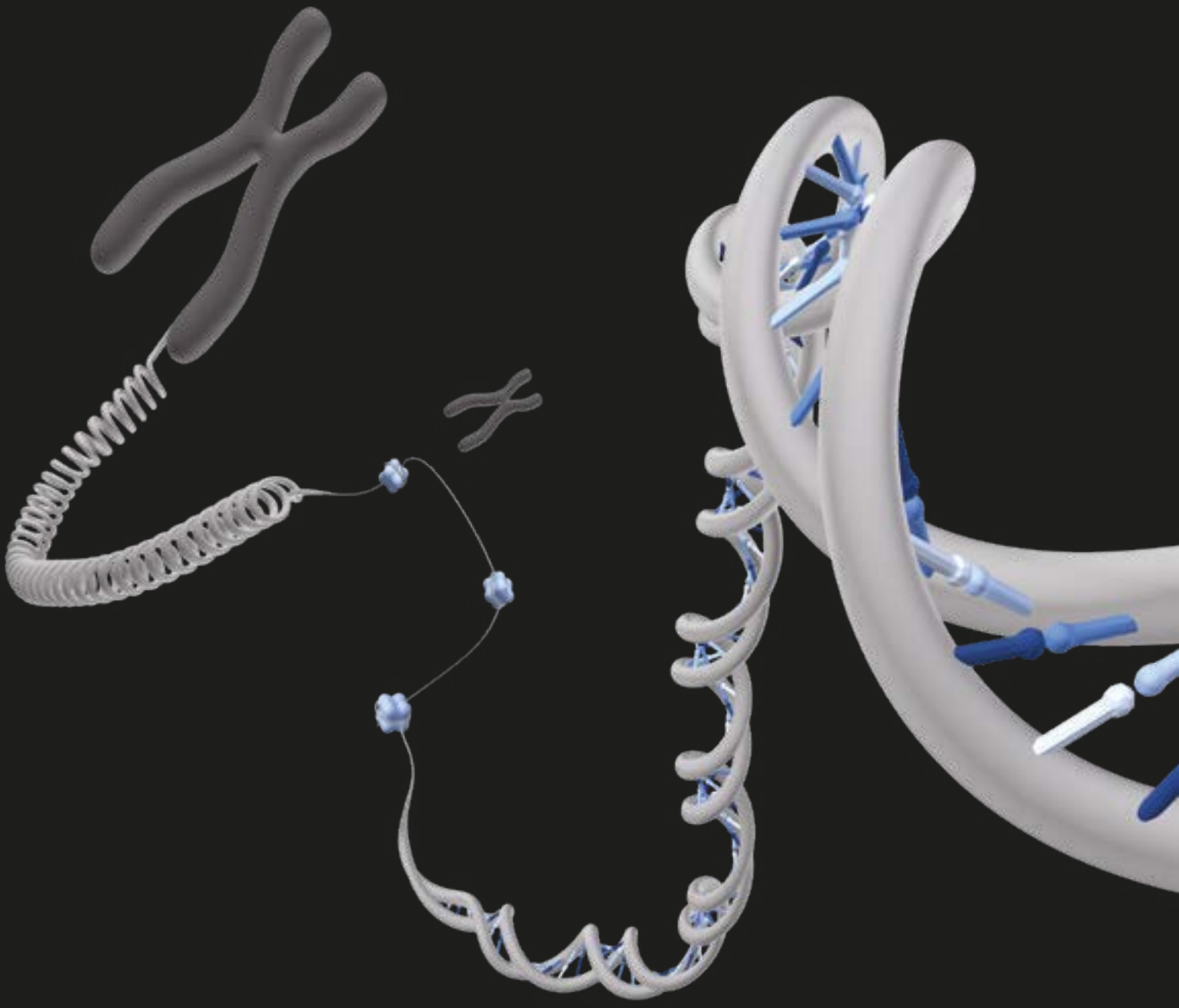
abcam

Polycomb and Trithorax group proteins in transcription and cancer



Contents

Introduction to Polycomb and Trithorax group proteins

The chromatin environment is crucial for maintaining cell-type-specific gene expression patterns and thereby cell identity. Different chromatin-modifying protein complexes can influence the accessibility of chromatin and local patterns of chromatin modifications. Two such protein families, the Polycomb group (PcG) and Trithorax group (TrxG), have been widely studied for their roles in modulating transcription and influencing cell fate. PcG and TrxG proteins were originally classified based on their distinct developmental phenotypes in genetic studies in *Drosophila*. Molecular mechanisms of the encoded proteins have since been studied, and homologs have been identified in many metazoan species. Both classes of proteins associate with several large, multimeric protein complexes, encompassing diverse molecular mechanisms and catalytic activities. PcG proteins are involved in maintaining transcriptional repression, while TrxG proteins modulate chromatin contributing to active transcription (Figure 1)[1].

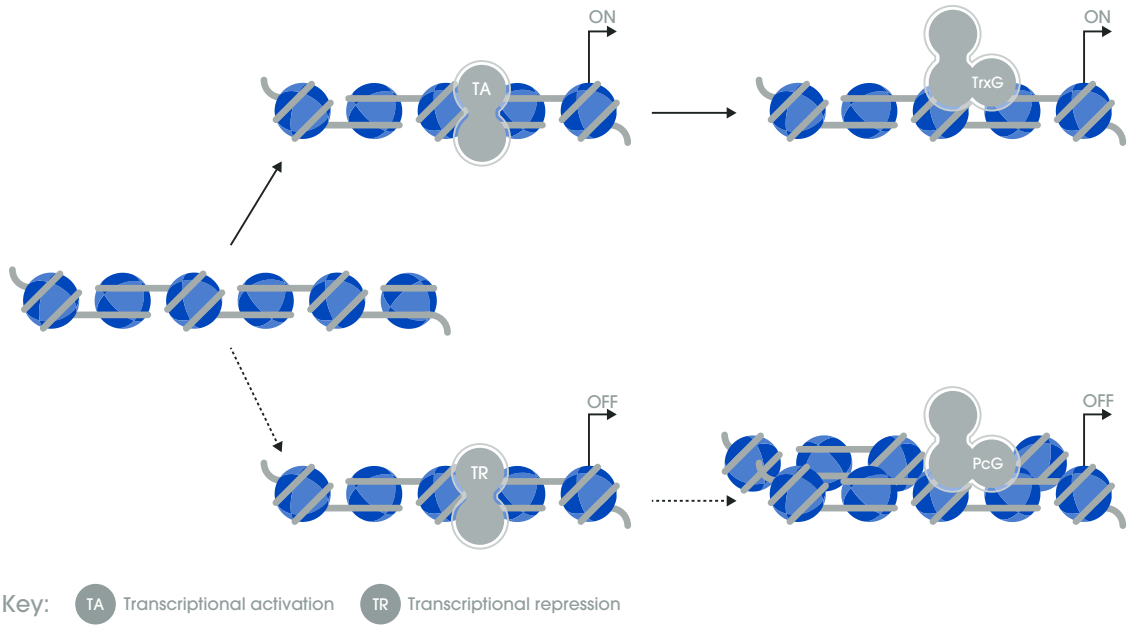


Figure 1. Polycomb and Trithorax group proteins modulate gene expression through the establishment of distinct chromatin states. Transcription initiation leads to the recruitment of TrxG proteins, facilitating an open chromatin environment permissive for active transcription. Termination of transcription leads to PcG recruitment to facilitate a more compact, repressive chromatin environment.

Polycomb repressive complexes: an overview

PcG proteins form the Polycomb repressive complexes 1 and 2 (PRC1/PRC2) as well as the Polycomb repressive deubiquitinase complex (PR-DUB) (Figure 2). Genomic binding sites of PRC1 and PRC2 overlap, and both complexes are primarily enriched at CpG-rich promoters of non-transcribed genes, where they contribute to the correct spatiotemporal expression pattern of developmentally regulated genes. Loss of PRC function during development leads to embryonic lethality, and PcG proteins are often found mutated or deregulated in human diseases, including cancer.

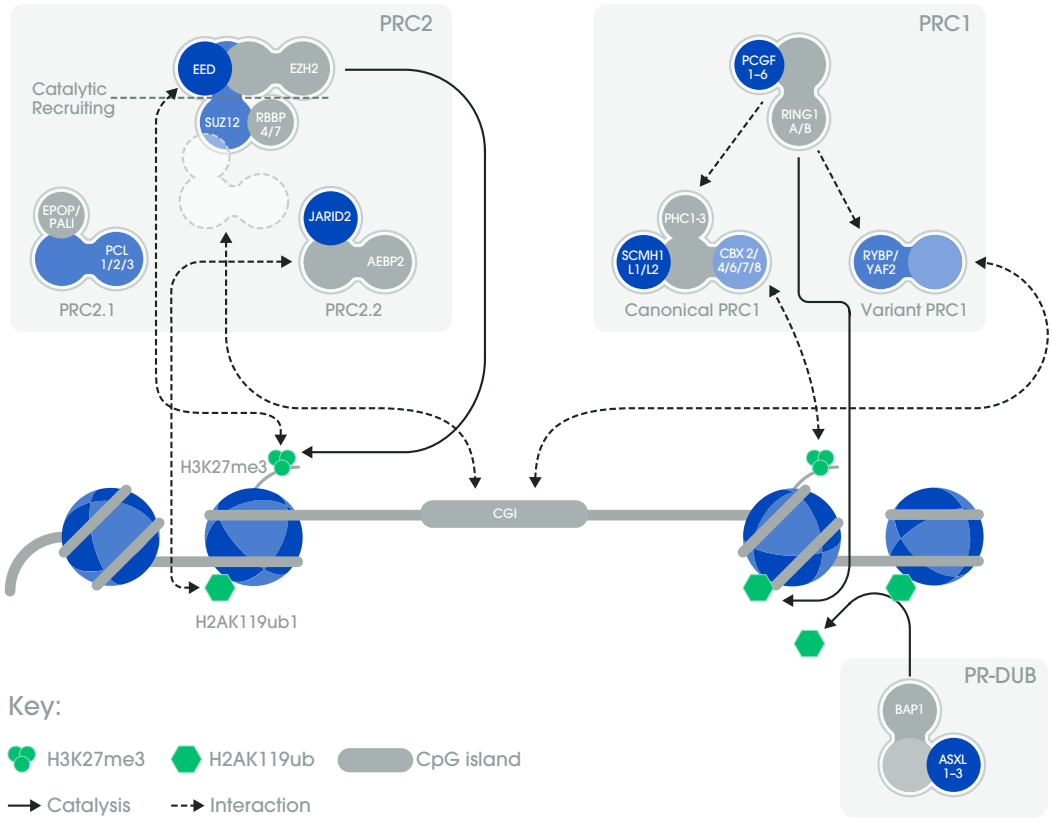


Figure 2. Transcriptional regulation by Polycomb repressive complexes. PRCs are recruited to CpG islands (CGIs) of non-transcribed genes, contributing to their maintained repression. PRC2 catalyzes H3K27 methylation, which is recognized by EED of PRC2 and CBX of PRC1. PRC1 catalyzes H2AK119 ubiquitination, which can be bound by JARID2/AEBP2-containing PRC2.2 complexes. H2AK119ub1 can be removed by PR-DUB.

PRC1 complex

PRC1 complexes contain an E3 ubiquitin ligase (RING1A or RING1B) with activity towards H2AK119, closely associated with a PCGF component (PCGF1-6), which is required for the catalytic activity of the complex to form H2AK119ub1 (Figure 2). Based on biochemical purification experiments, several different 'classes' of PRC1 complexes have been described, termed PRC1.1-PRC1.6 according to the incorporated PCGF homolog.

PRC1 complexes with PCGF2/PCGF4 are traditionally termed 'canonical' PRC1 (cPRC1) complexes. These complexes can incorporate a CBX (CBX2,4,6-8), a PHC subunit (PHC1-3), and an SCM subunit (SCMH1/L1/L2) (Figure 2). The 'non-canonical' or 'variant' PRC1 (ncPRC/vPRC1) complexes have been shown to incorporate many different non-core subunits, contributing to target site-specificity and regulation of catalytic rate[2].

PRC2 complex

The core PRC2 complex consists of a histone methyltransferase subunit (EZH1 or EZH2), which depends on association with SUZ12 and EED for *in vitro* activity and the additional association with the histone-binding proteins RBBP4 or RBBP7 for *in vivo* activity. PRC2 catalyzes mono-, di- and trimethylation of H3K27 (Figure 2). H3K27me3 is highly enriched at PRC2-bound chromatin regions, while H3K27me2 is wide-spread, covering large intergenic regions, and H3K27me1 is enriched within coding regions of transcribed genes.

Non-core PRC2 components interact with core PRC2 to form two distinct subcomplexes (Figure 2):

- PRC2.1 containing a PCL component (PCL1-3) along with EPOP or PALI
- PRC2.2 containing JARID2 and AEBP2

The non-core subunits modulate PRC2 activity and provide target site specificity, thereby contributing to the correct spatiotemporal deposition of H3K27 methylation[3].

PR-DUB complex

The PR-DUB complex catalyzes the deubiquitination of PRC1-deposited H2AK119ub1. It consists of a deubiquitinase subunit (BAP1) and an ASXL homolog (ASXL1-3), which is required for the catalytic activity of the complex (Figure 2). The core complex is further associated with accessory proteins, including the FOXK1/2 transcription factors and chromatin-modifying proteins that contribute to recruitment and chromatin modulation. Given that PR-DUB counteracts PRC1 function through the deubiquitination of H2AK119ub1, its classification as a PRC may seem counterintuitive. However, mutants of the *Drosophila* homologs of BAP1 and ASXL1-3 yield Polycomb-like phenotypes of homeotic transformation[1].

Chromatin binding and transcriptional regulation by PRCs

Several molecular mechanisms contribute to the correct genomic targeting of PRCs. Despite the years of studies of transcription factors, non-coding RNAs, and specific DNA features, no clear cause for mammalian PRC recruitment has been identified, although PRCs have been found to bind CpG-rich promoters of non-transcribed genes. Non-core subunits of PRC1 and PRC2 have been shown to guide target site specificity through direct chromatin interactions, strengthened by chromatin-interactions from core subunits. Some studies suggest that non-coding RNAs may be involved in recruiting PRCs to specific loci. Conversely, several studies indicate that RNA transcripts counteract PRC binding by competing for DNA binding surface, thereby providing a 'sensing mechanism' to explain selective recruitment of PRCs to non-transcribed genes[3].

During recruitment, PRCs also influence each other directly through several mechanisms. Thus, the CBX components of cPRC1 have an affinity for H3K27me3, which led to the traditional, hierarchical model of PRC recruitment. In this model, PRC2 is recruited first, followed by cPRC1 through its binding to H3K27me3. The model has been complicated and expanded by the finding that PRC2 loss does not lead to the global loss of H2AK119ub1 and the discovery of vPRC1 complexes relying on PRC1 components, such as KDM2B and RYBP, for chromatin binding.

In recent years, the non-core PRC2.2 subunits, JARID2 and AEBP2, have been shown to bind H2AK119ub1. H2AK119ub1 has been suggested to contribute to PRC2 recruitment and/or modulation of its catalytic activity, thereby adding reciprocity to the PRC1/PRC2 relationship. This interdependency can represent a way of providing robustness to the repression system; however, it provides some technical challenges when detangling the mechanistic contributions of single PRC components to transcriptional regulation and cell identity[1,3,4].

PRC binding and histone modifications are suggested to confer transcriptional repression via various mechanisms, including prevention RNA polymerase II binding or release, chromatin compaction, or co-transcriptional RNA degradation. Repression depends on the catalytic activities of both PRC1 and PRC2, which may well vary in distinct cell types, developmental stages, and genomic contexts. However, the relative contributions of each complex and modification remain to be understood[5].

Polycomb repressive complexes in development and disease

In addition to the developmental defects of PcG mutants in *Drosophila*, genetic studies in mice have shown that many PcG proteins are essential for normal development, with knockout phenotypes ranging from early embryonic lethality to homeotic transformation or perinatal lethality. Furthermore, mutations in genes encoding PcG proteins are associated with several human congenital disorders, underlining their key role during normal development[6,7].

Also, PRC function is often disrupted in many different types of human cancer through deregulation of expression levels or somatic mutations of genes encoding PcG proteins or histone mutants leading to a substitution of the PRC2 substrate lysine (H3K27M)[1,8]. Interestingly, both activating and inactivating mutations have been described, making it difficult to classify PcG proteins as traditional oncogenes or tumor suppressors. In this context, the oncogenic role of PRC deregulation has been proposed to stem from more general deregulation of the chromatin environment (as opposed or in addition to any direct impact on specific target genes)[9].

Therapeutic targeting of PcG proteins

The high frequency of cancer-associated mutations along with the promise of 'epigenetic therapies' has led to the development of various chemical compounds targeting PcG proteins. These include small-molecule inhibitors blocking the catalytic activity of EZH2 or preventing H3K27me3-binding by EED or CBX proteins or disrupting other protein-protein interfaces. Excitingly, the first EZH2 inhibitor, Tazemetostat, was recently FDA-approved for the treatment of follicular lymphoma and epithelioid sarcoma. Several other compounds are under preclinical and clinical development, including those targeting the catalytic activity of EZH2 or the integrity of PRC2 through degradation or obstruction of binding surfaces[10].

Interestingly, PcG and TrxG proteins show functional dependencies in tumorigenesis. For instance, some SWI/SNF-mutated cancers are sensitive to PRC2 inhibition and are among the cancers for which EZH2 inhibition has been approved[8,10]. Elucidating the mechanistic basis for these cancer vulnerabilities may help to devise optimized patient stratification and implement therapies targeting PcG and TrxG proteins in cancer.

TrxG proteins: an overview

TrxG proteins have been defined as factors suppressing the Polycomb phenotypes in *Drosophila*, indicating the opposing mechanistic function of the gene products. The TrxG protein complexes have diverse molecular functions with catalytic activities generally associated with active transcription. Many TrxG proteins are essential for preserving transcriptional patterns, and thereby cell fate, and the genes encoding TrxG proteins are often found mutated or deregulated in human cancers[12].

The TrxG proteins also segregate into large, multimeric protein complexes. These complexes can be largely sub-divided into two groups: COMPASS/COMPASS-like and SWI/SNF (mammalian homologs called BAF/PBAF).

COMPASS and COMPASS-like complexes

COMPASS and COMPASS-like complexes are SET-domain containing methyltransferases, catalyzing mono-, di- and trimethylation of H3K4 (Figure 3). The catalytic core of these complexes consists of WDR5, ASH2, RBBP5, and DPY30 (WARD). Each specific complex incorporates a SET-domain-containing methyltransferase (SET1, MLL1/2, or MLL3/4) as well as numerous additional subunits, yielding distinct chromatin binding patterns and additional catalytic activities to the complexes[1].

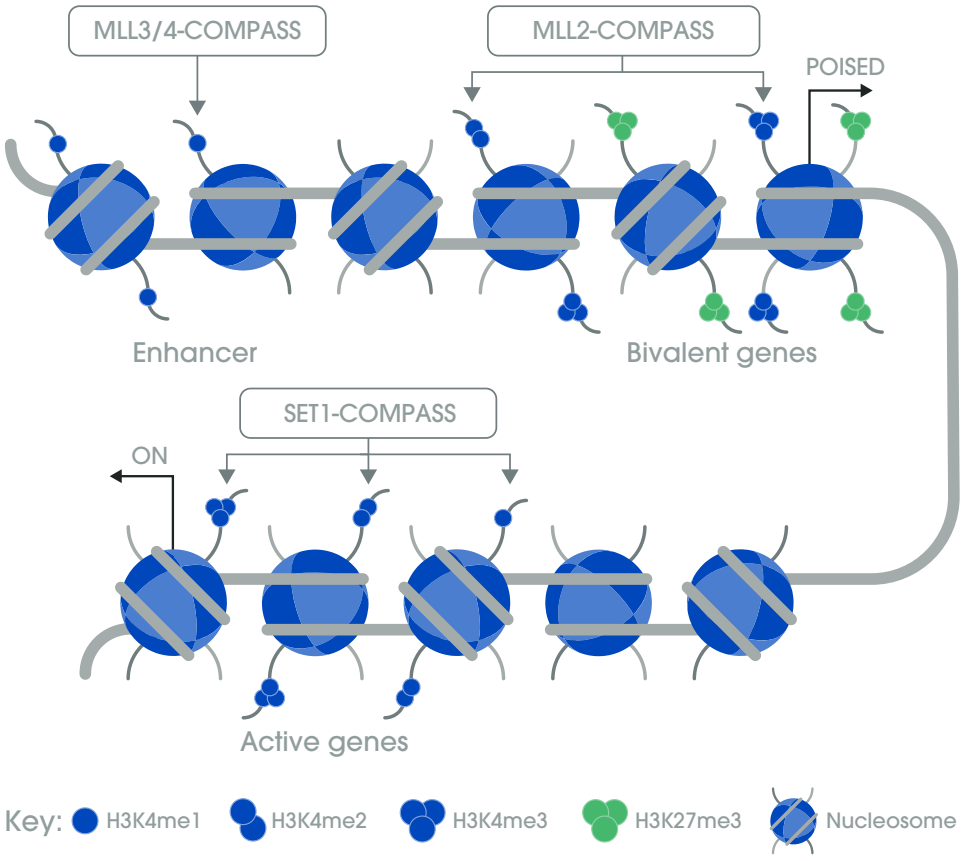


Figure 3. Different COMPASS complexes regulate H3K4 methylation in distinct genomic regions. SET1-COMPASS catalyzes H3K4me3 at the promoters of actively transcribed genes. MLL1/2-COMPASS-like complexes catalyze H3K4me3 at developmentally regulated and bivalent genes also marked by H3K27me3. MLL3/4 COMPASS-like complexes are responsible for the establishment of H3K4me1 at enhancer elements.

SET1-COMPASS complexes contain the methyltransferase SET1A or SET1B and several additional proteins, including WDR82, HCF1, and the DNA-binding protein CXXC1. SET1-COMPASS is the main H3K4 methyltransferase in mammalian cells, which catalyzes H3K4 methylation at the promoters of actively transcribed genes (Figure 3).

MLL1/2-COMPASS-like complexes contain either MLL1/KMT2A or MLL2/KMT2B along with HCF1 and MENIN. These complexes appear to direct H3K4 methylation to a more discrete set of target genes. MLL1-containing complexes target specific genes such as HOX genes, while MLL2-containing complexes catalyze H3K4me3 at developmentally regulated and bivalent genes also marked by H3K27me3 (Figure 3).

MLL3/4-COMPASS-like complexes contain the MLL3 or MLL4 methyltransferase along with NCOA6 and PA1 as well as the H3K27 demethylase UTX/KDM6, thus potentially directly counteracting Polycomb function at target genes. MLL3/4-COMPASS-like complexes act as the primary methyltransferase responsible for H3K4me1 at enhancer elements (Figure 3)[7].

SWI/SNF complexes

The SWI/SNF complexes are ATP-dependent chromatin-remodeling complexes mediating nucleosome sliding or eviction associated with active transcription (Figure 4). In mammalian cells, the two SWI/SNF ATPase homologs, SMARCA4/BRG1 or SMARCA2/BRM, each interact with about eight additional proteins to form the core complex, shared by both BAF and PBAF. Additional complex-specific subunits provide functional diversity to the complexes; this includes subunits ARID1A/B and DPF1/2/3 for BAF and PBRM1, BRD7, ARID2, and PHF10 for PBAF.

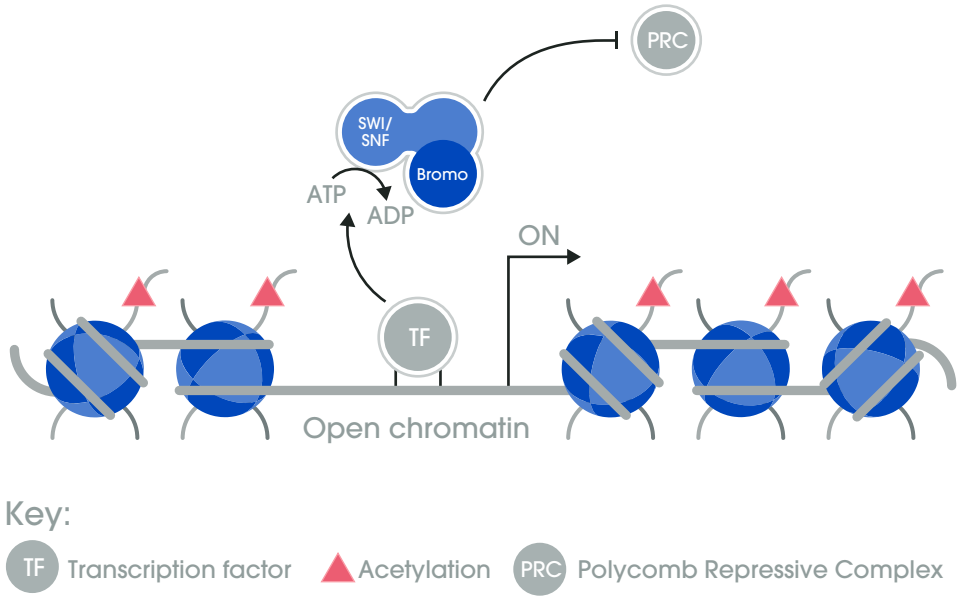


Figure 4. SWI/SNF protein complexes maintain an open chromatin environment through nucleosome remodeling. SWI/SNF (BAF/PBAF) protein complexes are recruited to CGIs at the promoters of transcribed genes, where they contribute to maintaining an open, permissive chromatin environment through ATPase-dependent chromatin remodeling and exclusion of PRCs.

Recently, a third type of mammalian SWI/SNF complex, non-canonical BAF or GBAF (named from its discovery in glioma) was identified. Interestingly, it lacks the SMARCB1 component, which provides chromatin targeting and stimulation of remodeling activity to the canonical BAF/PBAF complexes. Thus, like many other chromatin complexes, the BAF complexes are formed by modular assembly with shared and distinct subunits providing distinct molecular functions[8].

Chromatin binding and transcriptional regulation by TrxG proteins

In *Drosophila*, both PcG and TrxG proteins are recruited to Polycomb or Trithorax response elements (PREs/TREs). In mammalian cells, unmethylated CpG islands provide docking sites for both types of complexes in response to transcriptional cues. While PcG proteins are recruited to non-transcribed genes, TrxG complexes are recruited to actively transcribed genes, where they contribute to maintaining an open, permissive chromatin environment through nucleosome remodeling and H3K4 methylation (Figure 4). With their opposing functions on chromatin states and mutual inhibition of their respective binding or catalytic function, it appears that the outcome on transcriptional activity depends on the achieved balance of repressive vs activating signals[7,8].

Compared to PRCs, much less is known about the molecular mechanisms underlying the recruitment of TrxG proteins. Non-coding RNAs have been implicated, but similar to the PcG proteins, the generality of this model seems dubious. Rather, recruitment seems to rely on direct chromatin binding by several subunits. For COMPASS complexes, these include the CXXC domain of MLL1/2 and CXXC1 of the SET-COMPASS complex. Association with transcription factors may also contribute to recruitment, as seen by the role of the pioneering factor FOXA1 in recruiting MLL3/4 to enhancers.

Furthermore, a direct affinity for the H3 tail and certain histone modifications may contribute to directing chromatin binding[1]. Similarly, BAF/PBAF complexes contain subunits with several chromatin-binding modules, including zinc finger motifs, AT-hooks, and bromodomains, which contribute to direct DNA binding and/or recruitment to acetylated lysine residues[12].

TrxG proteins in development and disease

In knockout mouse studies, developmental phenotypes of TrxG loss-of-function range from early embryonic lethality to later or milder defects, depending on the targeted component. However, both SWI/SNF and COMPASS complexes are required for normal development[7].

In line with their importance for maintaining gene expression patterns and cell identity, many TrxG proteins show deregulation in cancer. The MLL proteins of the COMPASS-like complexes were named for their discovery as a region frequently translocated to produce fusion proteins in mixed-lineage leukemia (MLL), and they have since been found mutated in many other cancers. SWI/SNF components, including SMARCB1 and ARID1, are mutated in many ovarian cancers and rhabdoid tumors. Mutations in the catalytic domain of the ATPase BRG1 are frequent in a range of tumors and have been shown to promote aberrant recruitment and gene repression by PRCs, thus providing the basis for some of the observed co-dependencies and the rationale for treating some SWI/SNF-mutated tumors with PRC2 inhibitors[1,8].

Therapeutic targeting of TrxG proteins

Due to their frequent misregulation or mutation in human cancers, several compounds targeting TrxG proteins are in preclinical or clinical development. These include small-molecule inhibitors disrupting the WDR5-MLL interaction, thus inhibiting the function of COMPASS-like complexes, and several molecules targeting various SWI/SNF components. Given the described dependencies between SWI/SNF-mutated tumors and PRC2 inhibition, the development of novel drugs targeting SWI/SNF components may provide exciting opportunities for combination treatment in cancers with loss-of-function PRC2 mutations[7,8].

Conclusions and future perspectives

Since their discovery in *Drosophila*, the PcG and TrxG families of chromatin-associated proteins have been extensively studied in organisms ranging from yeast to humans. In recent years, *in vitro* and *in vivo* structure-function studies have yielded important information about molecular mechanisms of the PcG and TrxG complexes. This, along with the extensive phenotypic characterization of model organisms and large-scale genomic data from human cancers, has greatly advanced our understanding of how these complexes exert their biochemical and biological roles, both in normal biological contexts and disease.

Despite extensive studies, we still do not fully understand how PcG and TrxG complexes impact and are impacted by transcription and how the complexes affect each other to process transcriptional signals. Recent technical advances, including single-molecule tracking experiments, low-input techniques, cellular systems of rapid degradation, are helping to detangle cause from consequence between chromatin binding, modification, and transcriptional regulation. Research also remains focused on increasing our molecular understanding of chromatin regulation and linking it with biological phenotype. In the setting of human disease, it is a crucial task to further characterize co-dependencies and robust biomarkers for the optimized therapeutic use of anti-cancer therapies targeting PcG and TrxG proteins.

To summarize, despite the progress in the PcG and TrxG research field, many questions remain open, including molecular details of recruitment of PcG and TrxG complexes, their mechanistic impact on transcription, their biological roles in cancer, and their use as biomarkers and therapeutic targets in anti-cancer therapies.

References, further reading, and resources

References

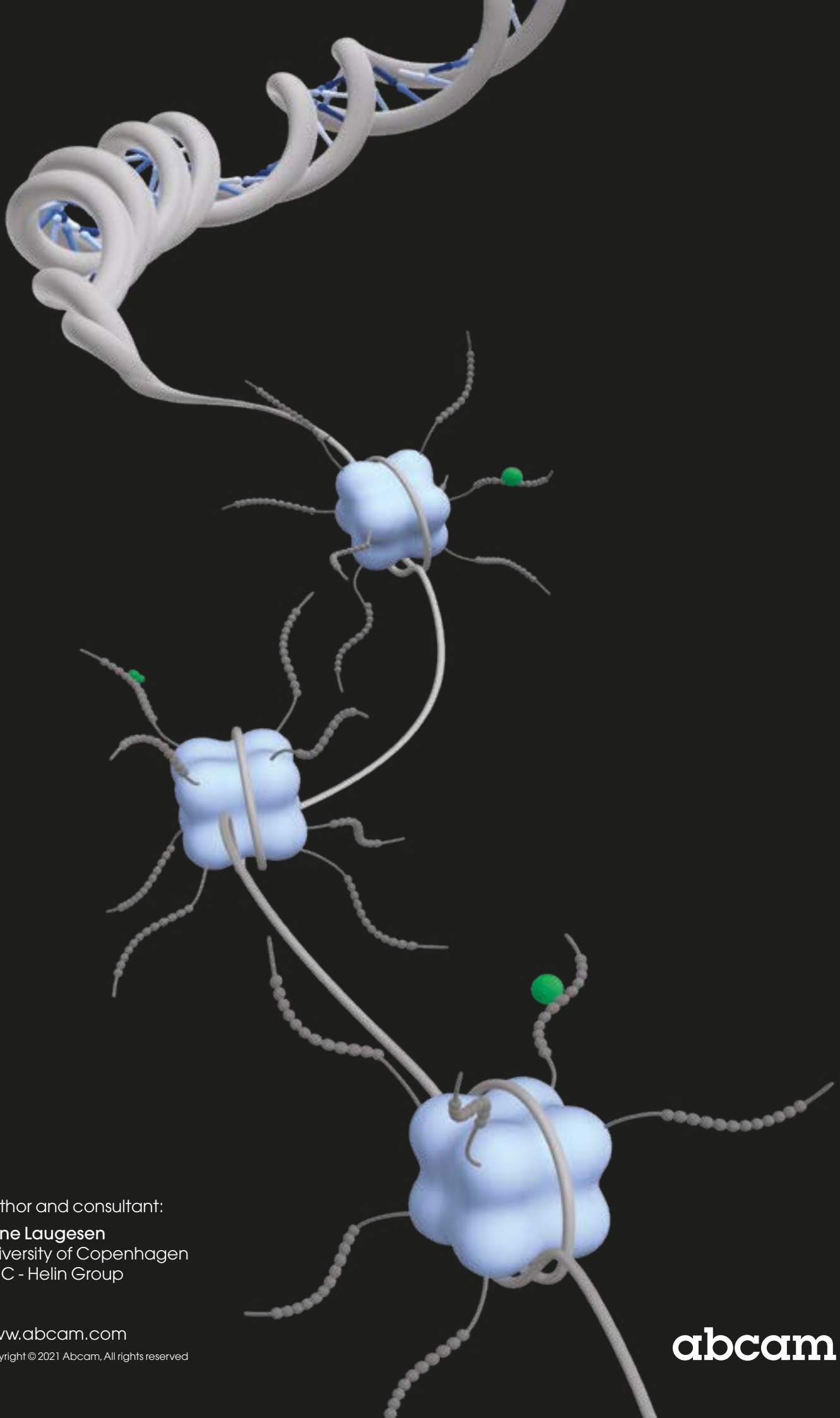
1. Schuettengruber, B., Bourbon, H.M., Di Croce, L., and Cavalli, G. Genome Regulation by Polycomb and Trithorax: 70 Years and Counting. *Cell* **171**, 34–57 (2017).
2. Chittock, E.C., Latwiel, S., Miller, T.C., and Muller, C.W. Molecular architecture of polycomb repressive complexes. *Biochem Soc Trans* **45**, 193–205 (2017).
3. Laugesen, A., Hojfeldt, J.W., and Helin, K. Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol Cell* **74**, 8–18 (2019).
4. Blackledge, N.P., Rose, N.R., and Klose, R.J. Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nat Rev Mol Cell Biol* **16**, 643–649 (2015).
5. Di Croce, L., and Helin, K. Transcriptional regulation by Polycomb group proteins. *Nat Struct Mol Biol* **20**, 1147–1155 (2013).
6. Deevy, O., and Bracken, A.P. PRC2 functions in development and congenital disorders. *Development* **146** (2019).
7. Piunti, A., and Shilatifard, A. Epigenetic balance of gene expression by Polycomb and COMPASS families. *Science* **352**, aad9780 (2016).
8. Bracken, A.P., Brien, G.L., and Verrijzer, C.P. Dangerous liaisons: interplay between SWI/SNF, NuRD, and Polycomb in chromatin regulation and cancer. *Genes Dev* **33**, 936–959 (2019).
9. Laugesen, A., Hojfeldt, J.W., and Helin, K. Role of the Polycomb Repressive Complex 2 (PRC2) in Transcriptional Regulation and Cancer. *Cold Spring Harb Perspect Med* **6** (2016).
10. Dockerill, M., Gregson, C., and DH, O.D. Targeting PRC2 for the treatment of cancer: an updated patent review (2016 – 2020). *Expert Opin Ther Pat.* (2020).
11. Kingston, R.E., and Tamkun, J.W. Transcriptional regulation by trithorax–group proteins. *Cold Spring Harb Perspect Biol* **6**, a019349 (2014).
12. Alfert, A., Moreno, N., and Kerl, K. The BAF complex in development and disease. *Epigenetics Chromatin* **12**, 19. (2019).

Further reading and resources

For more in-depth information on gene regulation by Polycomb and Trithorax proteins, we recommend the following review article: **Schuettengruber et al 2017**. For a better understanding of what chromatin is and how it functions, you can check out the detailed review by **Misteli T. 2020**.

If want to learn more about the key techniques used in epigenetics research, refer to our **Epigenetics application guide**. Also, you can check out our posters on **histone modifications** and **RNA modifications**.

If you are studying epigenetic targets in cancer, see our comprehensive **Cancer epigenetics guide**, which covers histone regulation, DNA and RNA modifications as well as polycomb and chromatin remodeling.



Author and consultant:

Anne Laugesen
University of Copenhagen
BRIC - Helin Group

www.abcam.com

abcam