Chapter 1

Not all Is SET for Methylation: Evolution of Eukaryotic Protein Methyltransferases

Allyson A. Erlendson and Michael Freitag

Abstract

Dynamic posttranslational modifications to canonical histones that constitute the nucleosome (H2A, H2B, H3, and H4) control all aspects of enzymatic transactions with DNA. Histone methylation has been studied heavily for the past 20 years, and our mechanistic understanding of the control and function of individual methylation events on specific histone arginine and lysine residues has been greatly improved over the past decade, driven by excellent new tools and methods. Here, we will summarize what is known about the distribution and some of the functions of protein methyltransferases from all major eukaryotic supergroups. The main conclusion is that protein, and specifically histone, methylation is an ancient process. Many taxa in all supergroups have lost some subfamilies of both protein arginine methyltransferases (PRMT) and the heavily studied SET domain lysine methyltransferases (KMT). Over time, novel subfamilies, especially of SET domain proteins, arose. We use the interactions between H3K27 and H3K36 methylation as one example for the complex circuitry of histone modifications that make up the "histone code," and we discuss one recent example (*Paramecium* Ezl1) for how extant enzymes that may resemble more ancient SET domain KMTs are able to modify two lysine residues that have divergent functions in plants, fungi, and animals. Complexity of SET domain KMT function in the well-studied plant and animal lineages arose not only by gene duplication but also acquisition of novel DNA- and histone-binding domains in certain subfamilies.

Key words Euchromatin, Heterochromatin, Histone, PRMT, SET, H3K36, H3K27, Protists, Fungi, Plant, Animal

1 Introduction

Chromatin is the key architectural feature organizing most eukaryotic genomes into structurally distinct domains, resulting in varying accessibility to transcriptional machinery [1]. Chromatin is an assembly of proteins and RNA that wrap DNA into repeating units of ~150 bp, called nucleosomes, each of which contains a histone octamer of dimers of H2A, H2B, H3, and H4. In the past, much emphasis has been placed on the idea that chromatin compacts the cell's genetic material and organizes the nucleus into

Raphaël Margueron and Daniel Holoch (eds.), *Histone Methyltransferases: Methods and Protocols*,
Methods in Molecular Biology, vol. 2529, https://doi.org/10.1007/978-1-0716-2481-4_1,

complex hierarchical structures. Yet, to function properly, chromatin must be dynamic and is thus subject to regulation in space and time as development of organisms, differentiation of tissues, and responses to the environment may demand. The local dynamics of chromatin are dictated by interactions of DNA with core histone octamers, with the linker histone H1 and numerous other DNA-binding proteins providing additional structural organization. Histones and other proteins can be chemically altered by posttranslational modifications (PTMs), including methylation, acetylation, phosphorylation, ubiquitination, and more. While these modifications are found along the entire histone sequence, modification in the basic N-terminal tails are most widely studied [1]. Incisive studies by mass spectrometry have uncovered hundreds of PTMs on all histones, though it remains unclear whether all of them carry biological significance [2–6].

Changes to the PTM chromatin landscape are accomplished by histone "writers" and "erasers," while histone "reader" proteins modulate the recruitment of downstream effectors [1]. This modification landscape is complex, with mechanistic understanding still rudimentary; it involves multiple histones, and sometimes neighboring nucleosomes. Histone methylation is just one PTM that is involved in numerous fundamental processes, affecting DNA replication, DNA repair, genome maintenance, and access to the transcriptional start site by transcription factors, and all these processes require alterations in local chromatin, which are achieved by changes in histone methylation status which are coupled to concomitant changes in the status of other PTMs, e.g., acetylation, phosphorylation, and ubiquitination.

Long-range interactions, or the modulation of larger chromatin domains, are defined by characteristic protein and DNA modifications—including nucleosome occupancy, cytosine or adenine DNA methylation, and histone PTMs—resulting in changes to gene expression [1]. Often described as "loosely packed," euchromatin is found in gene-rich, transcriptionally active regions of the genome. These regions have lower nucleosome occupancy, are characterized by histone lysine acetylation, histone H3 lysine 4 di- and trimethylation (H3K4me2/3), H3K36me3, and H3K79me3 (Fig. 1a). In contrast, domains of heterochromatin are transcriptionally silent, existing in two major forms—always condensed "constitutive heterochromatin" and reversibly transcriptionally silent "facultative heterochromatin" (Fig. 1b). Constitutive heterochromatin, characterized by H3K9me3 and cytosine DNA methylation, is found primarily in regions with repetitive DNA, including centromeric and subtelomeric regions, and transposable elements. Facultative heterochromatin is generally transcriptionally silent, but is expressed under appropriate conditions, in response to external or internal stimuli, aiding in proper spatio-temporal gene expression. Chromatin thus exists along a dynamic continuum from

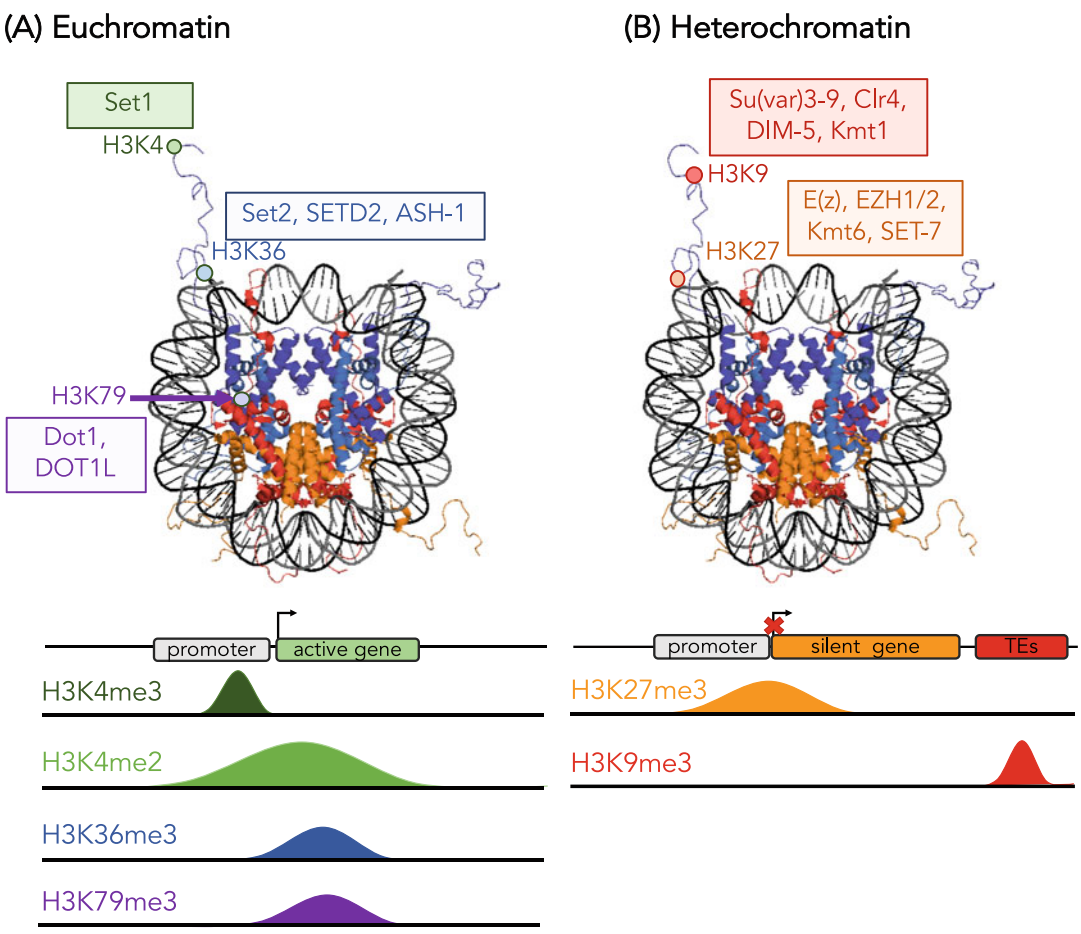

Fig. 1 Selected histone modifications that are correlated with (**a**) euchromatin or (**b**) heterochromatin and their idealized distribution on protein-coding genes. H3K4me2/3 are catalyzed by KMT2/Set1 proteins and are usually found in promoter or 5′ regions of genes. H3K36me2/3 are catalyzed by both Set2 and Ash1-like proteins and in many organisms cover all protein-coding genes, or expressed genes. While early studies showed effects of Ash1 on H3K4me2/3, recent results obtained with filamentous fungi suggest that Ash1-mediated H3K36me2/3 is correlated with subtelomeric facultative heterochromatin and affects H3K27me3. The non-SET KMT, Dot1, methylates a surface-exposed H3K79 residue in the H3 globular domain and is mostly correlated with active transcription. In many eukaryotes, facultative heterochromatin is marked by H3K27 methylation by the KMT6/E(z) subfamily of SET domain proteins, while constitutive heterochromatin is marked by H3K9me2/3, catalyzed by the KMT1/Su(var)3–9 subfamily, first discovered in *Drosophila* and called Clr4 in *S. pombe* and DIM-5 in *N. crassa*. Not shown is H4K20 methylation, which affects cell cycle regulation and DNA repair, and is correlated with gene repression. Also not shown here are the activities of the various protein arginine methyltransferases (PRMTs), as they have not been universally confirmed in most eukaryotes

inaccessible, transcriptionally silent heterochromatin to accessible, transcriptionally active euchromatin [7].

Here we will discuss the occurrence of histone methyltransferases in eukaryotes, and their shared and sometimes divergent functions in arginine and lysine methylation, with some emphasis on the interactions between two conserved SET domain histone methyltransferase complexes: PRC2, which methylates histone H3K27, and ASH1, which methylates histone H3K36. As is true for other families of chromatin-modifying proteins [8], expansion of gene families occurred in all eukaryotic clades, and novelty often arose by addition of expansion of DNA- or histone-binding motifs.

2 Meet the Organisms: Deep Phylogenetic Sequencing Makes New Models

Most scientists working on histone methyltransferases settle early on their "favorite model" organism. Thus, plant biologists are experts on the many histone modification enzymes controlling plant development or host–pathogen interactions, and the same can be said of human geneticists who may be particularly interested in pathologies caused by PTM dysregulation during development or cancer. Some organisms have been essential general models, e.g. among the ciliates *Tetrahymena* and among the fungi budding (*Saccharomyces cerevisiae*) and fission (*Schizosaccharomyces pombe*) yeast, but by now we know that they lack important PTMs present in plants or many animals, and thus filamentous ascomycete fungi (like *Neurospora crassa*, *Fusarium* spp., or *Zymoseptoria tritici*) and basidiomycete yeasts (like the human pathogen *Cryptococcus neoformans*) have become additional models to decipher the general histone methylation landscape and the interactions and dependencies between different methylation states.

Of course, eukaryotic biology is much more diverse than the choice of model organisms reflects, and thus one aim of this chapter is to explore the complement of protein methyltransferases that can affect histones, and thus chromatin structure, from all eukaryotic supergroups [9, 10]. Taxonomy has come a long way since the days of the "Five Kingdoms" hypothesis; instead, we now recognize at least seven supergroups of eukaryotes, still leaving six large clades unassigned, and we are not even quite sure yet how the supergroups form monophyletic clades [9]. Advances in genome sequencing give us access to at least the predicted proteomes of many new taxa, including important human pathogens, organisms important for carbon sequestration and other ecological issues, plants and their pathogens, and a large number of animal taxa. For our analyses, we selected representatives from most supergroups, at least one species each when high-quality genome drafts were available (Fig. 2). This still cannot do justice to the diversity of taxa in supergroups; it is clear from analyses of plants, fungi, and animals, the best-studied eukaryotes, that numerous lineages maintain or expand most known protein methyltransferases but that many lineages within kingdoms lose specific activities.

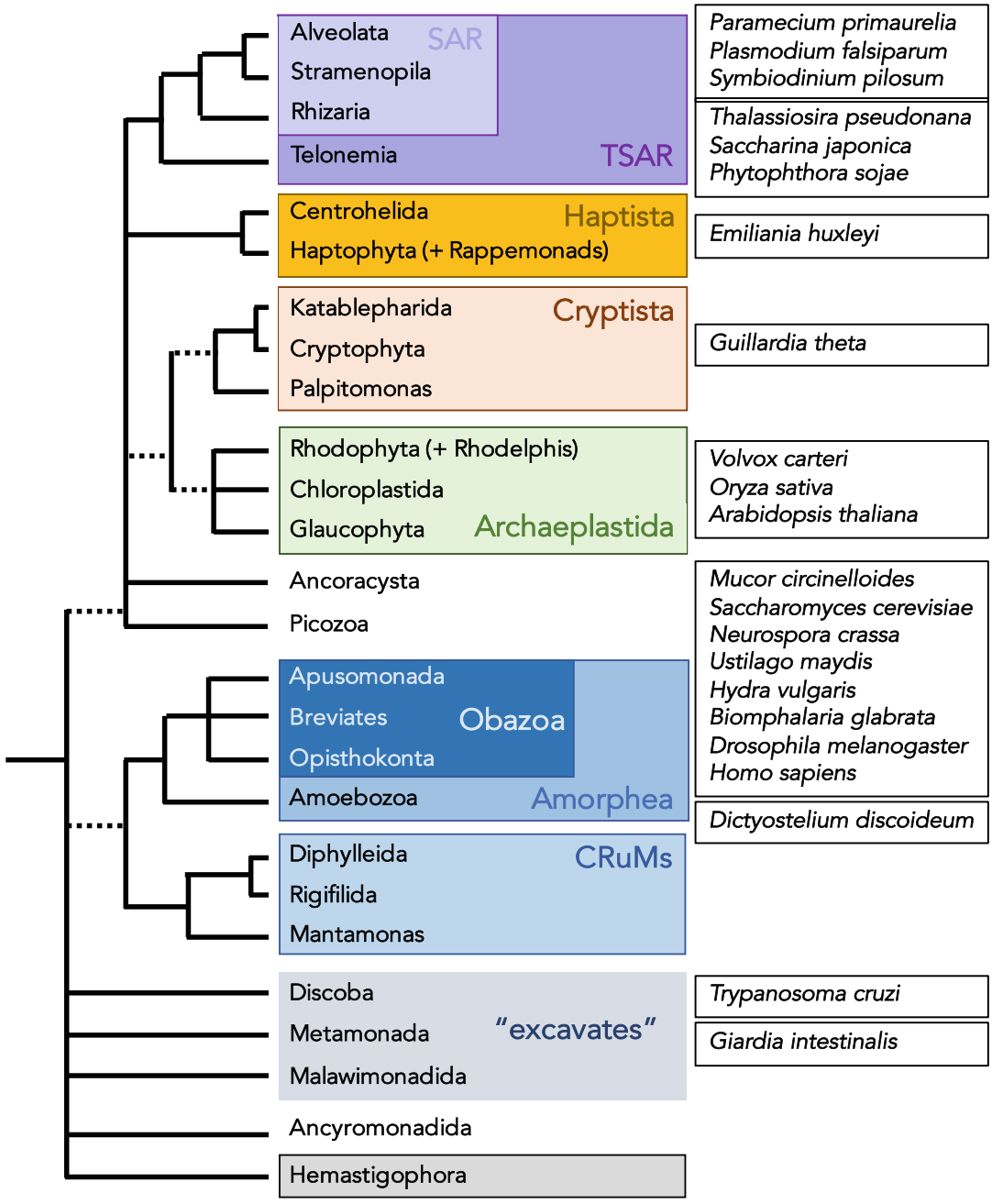

Fig. 2 Proposed supergroups within the eukaryotes (according to [9]) and species selected to investigate distribution and relationships of eukaryotic protein methyltransferases. See text for details on the organisms

The Discoba and Metamonada belong to the former supergroup of "excavates," an early and deeply diverging lineage of eukaryotes (Fig. 2). Many protists that evolved from this group are parasites and thus often have reduced genomes shaped by loss of traits caused by the evolving host–pathogen interactions

[11, 12]. The Metamonada contain many anaerobic species that are symbionts (e.g., in termite guts) or intestinal parasites of mammals. In humans, *Giardia* infection by various species causes serious diarrheal disease, and thus, *G. intestinalis* (synonyms *G. duodenalis* and *G. lamblia*) has been heavily studied and several genomes are available [12–14]. Kinetoplastids belong to the Discoba and include the genera *Leishmania*, causal agents of Leishmaniasis, and *Trypanosoma*, such as *T. brucei*, the causal agent of African sleeping sickness, and *T. cruzi*, the causal agent of Chagas disease, all classified as "neglected tropical diseases" by the NIH. Numerous genome sequences from a variety of pathovars are now available [15, 16].

Most of the diversity of eukaryotes resides in the supergroup now called "TSAR" [9]. We selected three Alveolata, the ciliate *Paramecium primaurelia*, the apicomplexan *Plasmodium falsiparum*, and the dinoflagellate *Symbiodinium pilosum* for our analyses (Fig. 2). Alveolates followed distinct evolutionary trajectories to yield different nuclear genome organization [17, 18]. Dinoflagellates in particular are quite distinct from other eukaryotes in that they do not use histones to organize chromatin; instead, they have condensed liquid-crystalline chromosomes [19–21]. Many *Symbiodinium* species are photosynthetic, and all coral symbionts are from this large genus. Understanding symbiosis, especially in this age of climate change, is of outstanding importance to ensure the survival of essential marine ecosystems; thus, the study of corals and their symbionts has enjoyed much attention [22]. Ciliates, like the genera *Tetrahymena* and *Paramecium*, have been important model organisms, including for studies on gene silencing [23–25]. They contain a somatic macro- and reproductive micronucleus but lack plastids. Genome structures of representatives from the large genus *Paramecium*, e.g., *P. primaurelia* and *P. tetraurelia*, have intensified after whole-genome duplications were detected in the clade [26]. Apicomplexans, such as the malarial parasite, *P. falciparum* have quite reduced genomes, with degenerate "apicoplasts" [17]. They are of general interest because malaria is one of the most important human diseases, on the rise partly because of climate change [27].

From the Stramenopila, we selected a diatom, *Thalassiosira pseudonana*, a kelp, *Saccharina japonica*, and an oomycete, *Phytophthora sojae*. Marine diatoms like *T. pseudonana* are widely distributed throughout all oceans and are models for light absorption and carbon metabolism, including how diatoms may affect global carbon cycling [28]. "Kelps" or "seaweed" belong to a large group of marine brown algae, and *S. japonica* is one of the commercially important species for food production [29–31]. The oomycete genus *Phytophthora* includes some of the most devastating plant pathogens [32, 33]. Some are relative specialists and infect only specific plants, like *P. sojae* on soybeans and *P. infestans* mainly

on potatoes, and others are generalists and are able to infect a large group of plants, e.g., *P. ramorum* or *P. cinnamomi* on many diverse woody plants. These species have the potential to be extremely invasive and change the whole ecosystems in a relatively short amount of time.

Within the Haptista, there are many species of marine and freshwater protists. *Emiliania huxleyi* belongs to photosynthetic plankton found in oceans from the equator to subpolar regions that form the basis of marine food webs [34]. It can form extensive blooms in nutrient-depleted waters that impact ocean temperatures and carbon balance but contributions of this species or plankton as a whole are not yet well understood [35–37]. The Cryptista contains species of flagellate algae that have a secondary plastid within a cytoplasm that also contains a vestigial nucleomorph, evidence of eukaryotic endosymbiosis [38]. *Guillardia theta* is the only characterized member of the genus and the first cryptophyte with a sequenced genome [39].

The Archaeplastida includes all land plants and green algae, the photoautotrophic red algae (Rhodophyta), and their non-photosynthetic sister group (Rhodelphis), as well as a distinct group of freshwater algae (Glaucophyta). From this group, we selected a multicellular green alga, *Volvox carteri*, which has become a model organism to study evolution of multicellularity, and two land plants, the monocot and most cultivated grain species, rice (*Oryza sativa*), and the dicot and best understood plant, *Arabidopsis thaliana*.

The Amorphea includes amoebae and the Opisthokonta, which includes the fungi and animals. All of these groups have been very heavily studied, and thus most analyses center on comparisons of animals to each other, or to the fungi and plants. We selected some of the obvious candidates, such as the amoeba ("slime mold") *Dictyostelium discoideum* [40], the fungi *Mucor circinelloides* (an emerging human pathogen belonging to the former "zygomycetes" [41]), the budding yeast *S. cerevisiae* [42], the ascomycete *N. crassa* [43], and the hemibasidiomycete *Ustilago maydis* (a global pathogen on maize [44]). From the animals, we chose *Hydra vulgaris* [45], *Biomphalaria glabrata* (a snail host of schistosome parasites) [46], the fruit fly *Drosophila melanogaster* [47], and human, *Homo sapiens* [48].

The take-home message from our representative sampling is that histone methylation capabilities are of ancient origin and that most arginine and lysine methyltransferase proteins are found in at least some taxa from all extant supergroups. This has been borne out by an in-depth phylogenomics analysis of many taxa that found "punctate retention" of histone methylation genes across eukaryotes [49].

3 Classes of Histone Methyltransferases

Protein methyltransferases (MTases) evolved to specifically methylate arginine or lysine residues of target substrates, and many of them play instrumental roles in regulating the structure and function of chromatin. Universally, histone methyltransferases (HMTs) use *S*-adenosyl-L-methionine (SAM or AdoMet), an intermediate metabolite of methionine metabolism, as a methyl radical donor, yielding methylated arginines or lysines and releasing the cofactor product *S*-adenosyl-L-homocysteine (SAH). HMTs fall into three groups: protein arginine methyltransferases (PRMTs), non-SET domain KMTs, and SET domain lysine methyltransferases (KMTs).

3.1 Protein Arginine Methyltransferases (PRMTs)

Arginine residues can be monomethylated, carried out by Type I, II, and III enzymes, or dimethylated, either asymmetrically (Type I) or symmetrically (Type II; reviewed in [50, 51]). While three types of PRMTs are distinguished by the methylarginine they produce, the active sites of these enzymes are similar though they differ in the substrate binding pocket, thus restricting their activity toward specific target proteins [52–54]. Based on the primary sequence of the AdoMet MTase domain and the presence of additional protein domains or motifs, at least ten distinct eukaryotic PRMTs can be distinguished (Fig. 3a). The current nomenclature for these proteins is challenging to follow; as we will also see for the much better studied SET domain proteins, there is no universally accepted numbering system. Most supergroups have genes encoding representatives of PRMT1, PRMT5, PRMT6, PRMT7, and PRMT10, while PRMT2 and PRMT8 are restricted to animals.

The N-terminus of the conserved methyltransferase region of PRMT3 is composed of a Rossmann-like fold, with five beta strands alternating with four alpha helixes to form an extended beta sheet structure. These folds create a SAM binding pocket [52]. Toward the C-terminal region of the conserved methyltransferase domain a barrel-like structure is observed, comprising the active site of the PRMT, and there has been much progress made on discovering small molecule inhibitors [50, 55].

3.1.1 Distribution of PRMTs in the Eukaryotes

Of the 22 species selected for extensive BLAST searches, only one, the diplomonad *Giardia intestinalis* in the Metamonada supergroup, did not return any reliable "hits" against plant, animal, or fungal PRMTs (Table 1). The fact that *Giardia* does not seem to have any arginine methylation has recently come to light; apparently there are functional equivalents to fulfill this important role, not just for histone but protein methylation in general [56]. Absence of PRMTs had been previously postulated [57], and as diplomonads include important human pathogens, neofunctionalization to substitute for PRMT function immediately suggested novel approaches for pharmacological intervention.

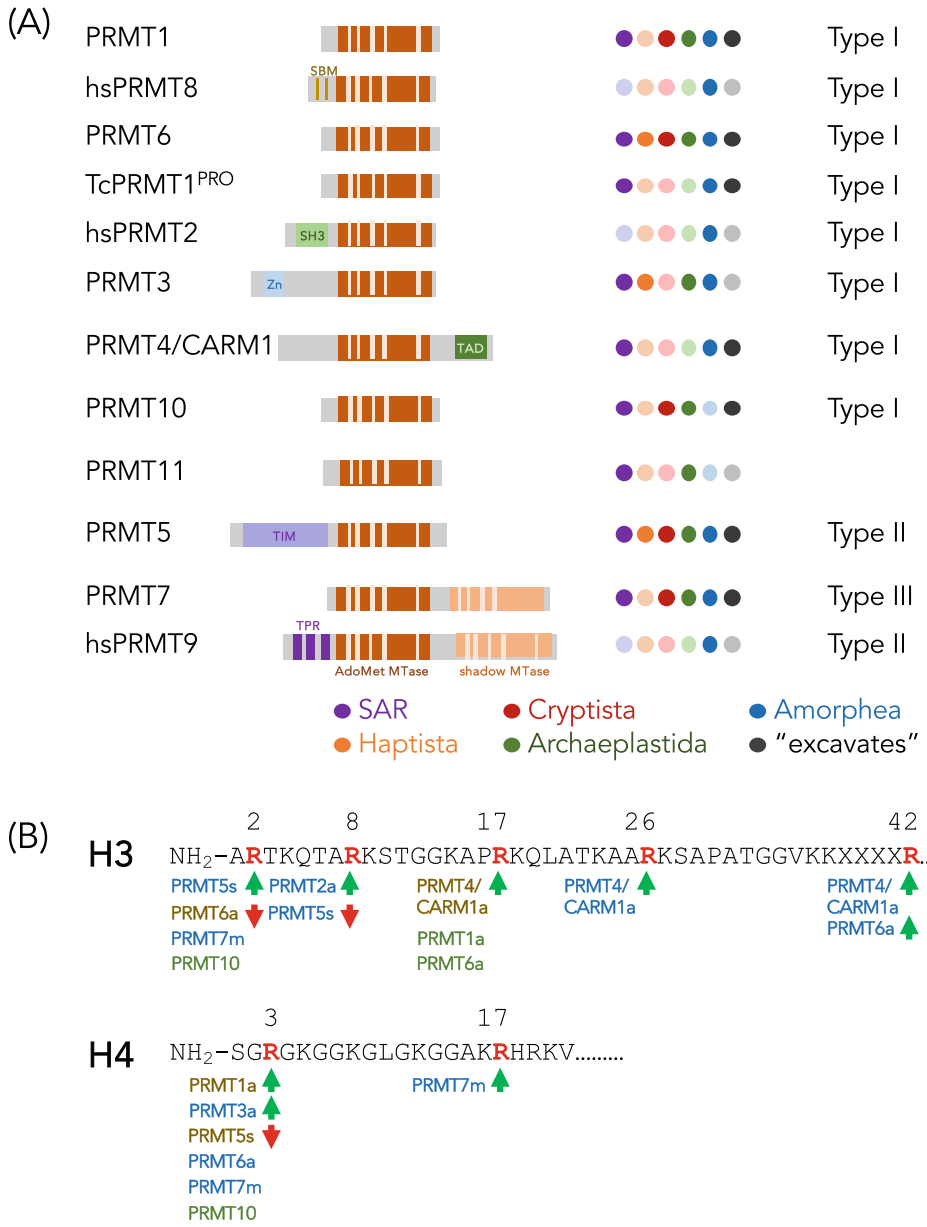

Fig. 3 Eukaryotic protein arginine methyltransferases (PRMTs) and their known histone substrate specificity. (**a**) Classification of PRMTs mostly following the subfamilies described in mammals. The PRMT catalytic domains are shown in brown, with the conserved motifs shaded in light tan. PRMT7 and PRMT9 proteins both contain discernible duplications of the AdoMet MTase domain, called "shadow MTase" here. Within the subfamilies, some novelty is generated by addition of the various functional domains shown (i.e., SBM, SH3, Zn-binding domains in the PRMT1, PRMT6, and PRMT3 groups). Presence of PRMT families in the 22 taxa examined is indicated by the shaded colored circles; solid colors denote presence and pastel colors denote absence in the taxa studied, but the latter does not indicate that other species in the respective supergroups may not encode these PRMTs. (**b**) Documented activities of eukaryotic PRMTs on histone H3 and H4 tails. Blue type denotes activity demonstrated in the Amorphea (mostly animals and fungi), green type denotes activity demonstrated in Archaeplastida, and gold type denotes activity found in both Archaeplastida and Amorphea. Green arrows denote correlation with gene expression, while red arrows denote correlation with transcriptional gene silencing

Table 1
GenBank accession numbers for PRMT homologs discussed and their most likely placement in PRMT subfamilies

Name	PRMT1	PRMT2	PRMT3	PRMT4	PRMT5	PRMT6	PRMT7	PRMT8	PRMT9	PRMT10	PRMT11
Thalassiosira pseudonana	XP_002288454.1		XP_002295201.1		XP_002293139.1	XP_002290528.1				XP_002297359.1	
			XP_002297358.1								
Symbiodinium pilosum				CAE7435993.1	CAE7718648.1	CAE7494238.1	CAE7761882.2			CAE7236492.1	CAE7513706.1
							CAE7331634.1				
Saccharina japonica				5116170.1	5102158.1						
Paramecium primaurelia	CAD8085966.1				CAD8046498.1				CAD8078653.1		
	CAD8100299.1								CAD8088540.1		
	CAD8085964.1								CAD8082635.1		
									CAD8107569.1		
Plasmodium falsiparum	PF3D7_1426200				PF3D7_1361000				PF3D7_0811500		
Phytophthora sojae	XP_009526476.1			XP_009523191.1	XP_009532297.1		XP_009536099.1			XP_009514661.1	
	XP_009526537.1										
Emiliania huxleyi			XP_005757501.1	XP_005766316.1	XP_005764086.1	XP_005765892.1				XP_005792814.1	
			XP_005768836.1		XP_005782814.1					XP_005776194.1	
										XP_005793688.1	
										XP_005764430.1	
Guillardia theta	XP_005825583.1				XP_005837555.1	XP_005836817.1	XP_005835264.1		XP_005822462.1		
	XP_005824608.1					XP_005842057.1	XP_005833971.1				
	XP_005818099.1										
	XP_005834566.1										

Volvox carteri	XP_002946963.1		XP_002953088.1	XP_002958435.1	XP_002946852.1				XP_002946350.1	XP_002958150.1
			XP_002950898.1							
Oryza sativa	XP_015612441.1		XP_015647334.1	XP_015647687.1	XP_015627032.1	XP_015614476.1	XP_015641005.1		XP_015642494.1	
Arabidopsis thaliana	NP_194680.1		NP_187835.2	NP_199713.2	OAO97588.1	NP_188637.2	OAO96767.1		NP_563720.1	
Trypanosoma cruzi	XP_815715.1				XP_808137.1	XP_811231.1	XP_815328.1	XP_822106.1		
	XP_821921.1					XP_806643.1	XP_819617.1	XP_817659.1		
Giardia Intestinalis										
Dictyostelium discoideum	XP_635288.1				XP_637240.1	XP_643270.2				
						XP_636224.1				
Mucor circinelloides	EPB85895.1		EPB81838.1	EPB90571.1	EB87598.1			EPB82839.1		
Saccharomyces cerevisiae	NP_009590.1		NP_010753.1		KZV13214.1					
Neurospora crassa	XP_963910.1		XP_956875.2		XP_011393468.1					
Ustilago maydis	XP_011392245.1			XP_011387335.1	XP_011387515.1	XP_011371078.1				
Hydra vulgaris	XP_002157035.1		XP_002165428.3	XP_002169585.2			XP_012566808.1	XP_002155504.2		
Biomphalaria glabrata	XP_013087638.1		XP_013064251.1	XP_013071413.1	XP_013084844.1	XP_013077111.1	XP_013029096.1	XP_013084436.1		
Drosophila melanogaster	NP_650017.1	NP_001285600.1	NP_731984.1	NP_001262445.1	ADU79249.1	NP_650322.1	NP_611753.4	NP_609478.1	NP_650321.1	
Homo sapiens	NP_001527.3	NP_001526.2	NP_005779.1	NP_001357018.1	NP_006100.2	NP_060607.2	NP_001338072.1	NP_062828.3	NP_612373.2	

Another "deep-branching" group in the former excavates supergroup, the Discoba, includes *Trypanosoma cruzi*, which encodes five different PRMTs, the same as found previously in *T. brucei* [57]. There has been sustained interest in PRMT function in kinetoplastids, resulting in studies on the function and interactions of all PRMTs [58] but especially the novel "PRMT3-like" protein that was found to be necessary for PRMT1 function [59, 60]. This protein has thus been renamed PRMT1^{PRO} (for "prozyme"), and PRMT1 has become PRMT^{ENZ}; a recent review summarizes these and other functional studies [61].

With the exception of *Plasmodium* [61–63], there are no published data on activities or function on PRMTs in the selected species in the Haptista (*E. huxleyi*), Cryptista (*G. theta*), and the diverse TSAR supergroup; these taxa show the most diverse patterns of presence or absence of PRMTs (Table 1). For example, the diatom *T. pseudonana* has two potential PRMT3 homologs, but neither includes a Zn-finger motif, and they are either predicted to be much longer or shorter than the homologs found in fungi and animals. It encodes another four PRMTs, similar to trypanosomes but has PRMT10 rather than PRMT7 in addition to PRMT1, PRMT5, and PRMT6. In contrast, the oomycete *P. sojae* has two PRMT1s, PRMT4, PRMT5, PRMT7, and PRMT10, but lacks PRMT3 and PRMT6. The ciliate *P. primaurelia* and the apicomplexan *P. falsiparum* encode just three PRMTs, clear homologs of PRMT1 and PRMT5, and another PRMT that is similar to *T. brucei* PRMT^{PRO}. Whether this protein has the same function in activating PRMT^{ENZ} is not resolved. Only two predicted proteins with PRMT signatures were detected in the genome of the kelp, *S. japonica*. While dinoflagellates have long been known to lack histones, there are at least six PRMTs predicted from recently published genome sequences, including "PRMT11," which was also detected in *Volvox* in our representative sampling of eukaryotic genomes. All *Symbiodinium* PRMTs are predicted to modify other proteins though it is conceivable that some modify the histones of the hosts, corals.

Compared to the taxa previously mentioned, most Archaeplastida (plants and green, as well as some red algae) have expanded this gene family, encoding eight or nine PRMTs [64], some of which (e.g., PRMT1 and PRMT4) are present in pairs that are at least partially redundant [51]. The green alga *V. carteri* seems to lack clear homologs for PRMT6 and PRMT7 but has two proteins, PRMT11 and PRMT12, that may be the best homologs for these PRMTs (Table 1).

Within the Amorphea, amoebae like *D. discoideum* encode three PRMTs (PRMT1, PRMT5, and two isoforms of PRMT6). Basal lineages of fungi, like the former "zygomycetes," represented here by *M. circinelloides*, encode five PRMTs, including a homolog of the PRMT1^{PRO} protein found in *T. brucei* (Table 1). This lineage

and the basidiomycetes also have PRMT4/CARM1 homologs that the ascomycetes (e.g., *S. cerevisiae*, *S. pombe*, and *N. crassa*) lack. Instead, these well-studied model organisms all have a minimal complement of PRMTs, namely PRMT1, PRMT3, and PRMT5 [65, 66]. Animals like *Hydra* and the snail *B. glabrata* have five and seven PRMTs, respectively; there are no functional studies available. *Drosophila* and mammals encode nine different PRMTs [67], with human PRMT2 and PRMT9 arising from the PRMT6 and PRMT7 family, respectively. Similarly, PRMT8 seems to be most closely related to the PRMT1 family. It is important to remember that the numbering system of nonmammalian animal PRMTs does not necessarily match protein similarity based on the human numbering system [67]; some of the animal PRMTs may also have been missed in previous analyses.

Overall, while comparisons of the whole protein sequences or just the MTase domain across the selected taxa resolve some protein phylogenies well, e.g., PRMT5, the PRMT1^{PRO} group, PRMT7, PRMT10, PRMT4, and PRMT1, placement of PRMT2, PRMT3, and PRMT6 is more difficult to resolve, largely because in some taxa the specific motifs found in animal PRMTs are lacking. Compared to some of the other histone modification gene families, PRMTs fall into the group that have been most widely retained since the "Last Common Eukaryotic Ancestor," LECA [49]. Intriguingly, at least one report has also suggested the presence of methylated arginine in bacteria [53].

3.1.2 Histone Methylation Catalyzed by PRMTs

PRMTs are well known to posttranslationally modify many proteins, including transcription factors, co-activators and co-repressors, and signaling factors involved in the cell cycle and oncogenesis; an in-depth review is beyond the purpose of this chapter but is available elsewhere [50]. PRMT activities on core histone tails constitute a rather minor part of their substrate repertoire, but they have been reviewed [50, 51, 67, 68] and are best-studied in mammalian cells (Fig. 3b). Methylated arginine residues of H3 and H4 are correlated with both active and silenced transcription, for example, H4R3me2a, catalyzed by PRMT1 or PRMT3, is a mark for active transcription, but H4R3me2s, catalyzed by PRMT5, is correlated with gene silencing. Similarly, on the H3 tail, H3R2me2s is a mark for active transcriptionm but H3R8me2s is a repressive mark, and both the reactions are catalyzed by PRMT5. PRMT6 catalyzes H2AR29me2a, which results in repression of transcription. Because methylarginines do not just correlate with active or silent transcription but are also involved in crosstalk between other histone modifications it is clear that there are several layers and potentially redundant circuits for specific outcomes of the "histone code." For example, there is evidence for crosstalk between PRMT7 and PRMT5, as H4R17me by PRMT7 may activate PRMT5 to yield H4R3me2s, a mark repressive for transcription [69].

Less is known from plants but early work in both *Arabidopsis* and rice revealed multiple histone methylation sites (Fig. 3b). In rice, PRMT1 generates H3R17me2 and H4R3me2, PRMT4 generates H3R17me2, PRMT5 generates H4R3me2, PRMT6b generates H3R2me2 and H3R17me2, and PRMT10 generates H3R2me2 and H4R3me2 [64]. Much work remains to be done, especially in the protists and filamentous fungi, to uncover the full scope of gene regulation by histone arginine methylation.

3.2 Non-SET Domain Histone Lysine Methyltransferases (KMTs)

Unlike HMTs that target histone tails for methylation, the non-SET-domain-containing methyltransferase Dot1 ("Disruptor of telomeric silencing 1") methylates H3K79, a surface-exposed residue in the H3 globular core. Dot1 is still the sole non-SET domain lysine methyltransferase, first discovered in budding yeast in a genetic screen for proteins involved in position effect variegation [70] and shown to modify gene silencing [71]. The mammalian homolog, DOT1L ("DOT1-like"), is important for transcriptional regulation, cell cycle regulation, and the DNA damage response [72]. In one classification, Dot1 homologs are labeled KMT4 [73].

Most eukaryotes have a single gene encoding Dot1 homologs, though some, like *T. brucei*, have two enzymes, one for H3K79 mono- and dimethylation (Dot1A) and one for H3K79 trimethylation (Dot1B) [72]. In many animals, either several genes (e.g., in *Caenorhabditis elegans*) or splice variants (in mammals) have been detected. Some fungi and most plants do not have genes encoding Dot1 homologs [74].

The overall size and structure of DOT1 homologs vary greatly, with the highest levels of sequence similarity in the N-terminus. Both yeast and human Dot1 have active sites capable of mono-, di-, and trimethylation. Crystal structures of DOT1L in complex with the methyl donor, SAM, showed that the N-terminal HMT domain is comprised of a series of open α/β structures surprisingly similar to that of PRMTs [75]. Differences surrounding the Dot1 active sites confer target specificity. Through incisive studies over the past two decades, Dot1L has emerged as one of the paradigms for histone modification crosstalk. Dot1 is activated by ubiquitination of H2B lysine 120 (H2BK120ub) [76] and structural work on how this is accomplished has recently been reviewed [77]. Additional structural studies showed how not only H2BK120ub but also H4K16 acetylation (H4K16ac) results in allosteric stimulation of Dot1 activity, both in vivo and in vitro [78].

3.3 SET Domain KMTs

SET domain-containing proteins, named after three proteins that were first discovered in *D. melanogaster*, namely suppressor of variegation 3–9 [Su(var)3–9], enhancer of zeste [E(z)], and trithorax (Trx), can be found in the genomes of all eukaryotes and in some bacteria [8, 51, 56, 74, 79–83]. The recent advances in whole-genome and metagenome sequencing have uncovered that histone methylation by SET domain group proteins is an

ancient process; phylogenies of selected SET domain proteins show that early diverging eukaryotes carry genes for many of the well-studied subfamilies. Many heavily studied model organisms, like budding or fission yeast, however, have lost genes for specific KMTs, and while the proteomes of budding and fission yeast harbor 12 and 13 SET domain proteins, respectively, many filamentous fungi (like *N. crassa*, *M. circinelloides*, and *Ustilago maydis*) encode as many as 20 SET domain proteins [82]. This protein family is even more expanded in mammalian and plant proteomes, where between 40 and 60 SET domain proteins are found [84]. Especially in the non-model eukaryotes, protein KMT activity has been experimentally attributed to only a subset of these proteins by either in vivo or in vitro methods, and fewer still are histone methyltransferases with known activity on specific lysine residues.

3.3.1 Distribution of SET Domain Proteins in Eukaryotes

Plant and mammalian SET domain proteins are well studied, and most previous work characterized seven or eight only partially overlapping subfamilies (Table 2); phylogenetic analyses of ciliate SET domain proteins uncovered 13 monophyletic eukaryotic clades [80], but based on the uncertain relationship in several subfamilies, there may be as many as 15 SET domain subfamilies. Many of the plant and animal proteins in these subfamilies have homologs and orthologs in the fungi, amoebae, and the SAR clade, though some of the truly well-studied plant, fungal, and mammalian KMTs have no obvious homologs in the SAR group or the early and deeply branching clades Discoba and Metamonada (Fig. 4a). In many of these clades, especially the plant-type SMYD and SETD subfamilies of KMTs, carrying both a SET domain and a zinc-finger MYND domain or a Rubisco LSMT substrate-binding domain, respectively, appear to be expanded to include many more family members than in plants, fungi, and animals. A complete accounting and curation for all SET domain proteins in the SAR clade is beyond this review, but it is curious that *Symbiodinium*, i.e., a genus without histones, seems to encode dozens of proteins with SET domains of the SMYD and SETD type.

Subfamilies in mammalian genomes have been renamed according to a system proposed after numerous model genome sequences had been nearly completed, relying on ordering KMT subfamilies by date of discovery [73]. Subfamily numbering in plants does not adhere to this classification, and even in single species, there are often multiple names for the same gene or isoform, as is common in mammals as well. For our purposes, we grouped subfamilies by known or predicted substrates and followed the mammalian nomenclature (Fig. 4a).

Homologs of *Drosophila* Su(var)3-9 belong to the KMT1 or plant Suv subfamily and methylate H3K9; most studied enzymes are capable of catalyzing mono-, di-, and trimethylation and are essential for gene silencing in constitutive heterochromatin

Table 2
GenBank accession numbers for SET domain KMTs found in selected eukaryotes and their most likely placement in KMT subfamilies

	Supergroup	SAR				Haptista	Cryptista	Archaeaplastida	Amorphea					
	Group	Diatom	Ciliate	Apicomplexa	Oomycete			Chloroplastida	Amoebozoa	Fungi			Animals	
Subfamily	Names	*Tp*	*Pt*	*Pf*	*Ps*	*Eh*	*Gt*	*At*	*Dd*	*Mc*	*Sc*	*Nc*	*Dm*	*Hs*
KMT1	Su(var)3-9, KMT1			PF3D7_0827800_SET3	XP_009518856.1	XP_005778404.1	XP_005825525.1	At5G04940	XP_646062.1	EPB89792.1	none	XP_957479.2	NP_524357.2	
SUV39	SUV39H1, KMT1A							At1G73100						NP_001269095.1
H3K9me	SUV39H2, KMT1B							At1G17770						NP_003164.1
	G9a, KMT1C			PF3D7_0508100_SET9				At2G24740					NP_569834.1	AAD21812.1
	GLP/ EuHMT1, KMT1D							At2G05900						BAB56104.1
	ESET/ SETDB1, KMT1E							At5G13960					NP_611966.3	NP_001380889.1
	SETDB2, KMT1F							At4G13460						NP_114121.2
								At3G03750 At2G35160 At2G22740 At1G04050 At5G43990						
								At3G04380						
								At2G23740						
KMT2	SET1, KMT2	XP_002291362.1	GSPATG00013040001	PF3D7_0629700_SET1	XP_009536801.1	XP_005764528.1	XP_005837587.1	At2G31650	XP_636258.1	EPB84754.1	NP_011987.1	XP_961572.3		
SET1	Trx, MLL1, KMT2A	XP_002296329.1					XP_005839674.1	At1G05830					NP_476769.1	NP_005924.2
H3K4me	Trx, MLL2, KMT2B		GSPATG00014017001					At3G61740						AAD56420.1
	Trr, MLL3, KMT2C		GSPATG00025368001	PF3D7_0910000_SET4				At4G27910						AAK00583.1
	Trr, MLL4, KMT2D		GSPATG00031547001	PF3D7_1355300_SET6				At5G53430						AAH09337.2
	hSET1A, KMT2F		GSPATG00018768001					At4G15180					NP_001015221.1	NP_055527.1

	hSET1B, KMT2G		GSPATG00035094001						At5G42400						NP_001340274.1
	ATXR1-4	XP_002290717.1													
KMT3	SET2, KMT3	XP_002294263.1	GSPATG00013040001	PF3D7_1322100	XP_009537692.1	XP_005794164.1	XP_005832326.1	At2G44150	XP_647576.1	EPB87143.1	NP_012367.2	XP_957740.1	NP_001263029.1		
SET2	SET2, KMT3A	XP_002296152.1	GSPATG00003275001				XP_005818219.1	At3G59960		EPB88931.1			NP_572888.2	NP_054878.5	
H3K36me	NSD1, KMT3B	XP_002296152.1	GSPATG00004957001		XP_009539893.1			At4G30860						AAK92049.1	
	SMYD2, KMT3C	XP_002294263.1						At1G76710						AAF68983.1	
	ASH1, set-3, KMT2H				XP_009518980.1			At1G77300				XP_964116.3	AAB01100.1	AAH11635.1	
KMT5	Set9, KMT5	XP_002291638.1			XP_009522567.1	XP_005776989.1		At2G33290		EPB87004.1	none	XP_963033.2			
SUV4-20	Pr-SET7/8, SET8, KMT5A			PF3D7_0403900_SET8						EPB81915.1			NP_001247100.1	Q9NQR1.3	
H4K20me	SUV4-20H1, KMT5B													NP_001356355.1	
	SUV4-20H2, KMT5C												NP_001245453.1		
KMT6	E(z), KMT6	XP_002290191.1	XP_001436830.1_EZL1		XP_009524339.1		XP_005836861.1	At1G02580		EPB84754.1	none	XP_965043.2	AAC46462.1		
E9z)	EZH1, KMT6A		GSPATG00032888001		XP_009519967.1		XP_005834965.1	At2G23380						XP_011522819.1	
H3K27me	EZH2, E(z), KMT6B		GSPATG00012695001		XP_009516860.1			At4G02020						NP_004447.2	
			GSPATG00013305001		XP_009521676.1										
			PTETG1700020001		XP_009526799.1										
ATXR	ATXR5, ATXR6	XP_002293206.1	GSPATG00025951001					At5G09790							
H3K27me1			GSPATG00033097001					At5G24330							
			GSPATG00036078001												
			GSPATG00035182001												
			GSPATG00021145001												
KMT7	SET7/9						XP_005842301.1		XP_636856.1						
														NP_085151.1	
KMT8	RIZ, PRDM			PF3D7_1214200_SET5											

(continued)

Table 2
(continued)

Subfamily	Supergroup Group Names	SAR Diatom *Tp*	Ciliate *Pt*	Apicomplexa *Pf*	Oomycete *Ps*	Haptista *Eh*	Cryptista *Gt*	Archaeaplastida Chloroplastida *At*	Amorphea Amoebozoa *Dd*	Fungi *Mc*	*Sc*	*Nc*	Animals *Dm*	*Hs*
H3K9me													NP_001261444.1	NP_001380915.1
SMYD				PF3D7_1115200_SET7			XP_005818844.1	At2G17900	XP_629856.1					Q8NB12.1
		XP_002292103.1			XP_009533210.1		XP_005827049.1	At2G19640	XP_644412.1_SMYD3L			XP_959981.1		NP_064582.2
		XP_002290198.1			XP_009533210.1	XP_005777382.1	XP_005826792.1	At1G26760	XP_002649106.1	EPB91989.1		XP_957968.2	NP_650955.1	NP_001161212.1
								At3G21820	XP_629629.1			XP_965349.2		Q8IYR2
								At5G06620						NP_006053.2
SETD		XP_002287161.1	GSPATG00003961001		XP_009533919.1	XP_005783180.1	XP_001713614.1	At1G01920				XP_964786.2		
		XP_002286075.1	GSPATG00027281001				XP_005828987.1	At1G14030				XP_963144.2		
		XP_002295030.1	GSPATG00020579001					At1G24610			NP_015160.1	XP_963594.3		NP_001153777.1
		XP_002296912.1	GSPATG00031631001					At2G18850						
		XP_002294212.1	GSPATG00024000001					At3G07670						
			GSPATG00015535001					At3G55080						
			GSPATG00012103001					At3G56570						
			GSPATG00016863001					At4G20130						
			GSPATG00037631001					At5G14260						
			GSPATG00010387001					At5G17240						
			GSPATG00038239001											
			GSPATG00007235001											
			GSPATG00012599001											
			GSPATG00013397001											
			GSPATG00012308001											
SETMAR														NP_006506
H3K4, H3K36														

Tt, Thalassiosira pseudonana, *Pt, Paramecium tetraurelia*, *Pf, Plasmodium falsiparum*, *Ps, Phytophthora sojae*, *Eh, Emiliania huxleyi*, *Gt, Guillardia theta*, *At, Arabidopsis thaliana*, *Dd, Dictyostelium discoideum, Mc, Mucor circinelloides, Sc, S. cerevisiae, Nc, Neurospora crassa, Dm, Drosophila melanogaster, Hs, Homo sapiens*

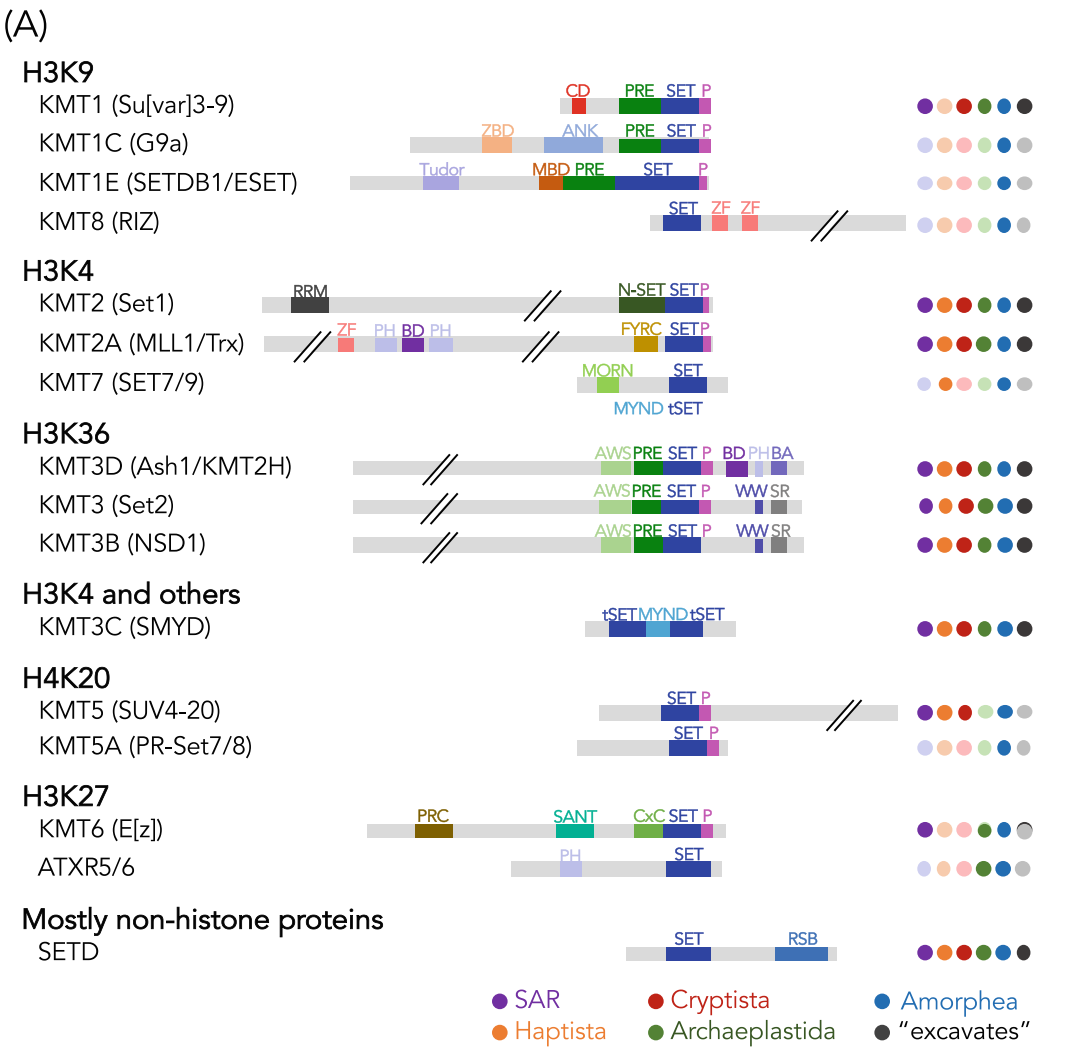

Fig. 4 Eukaryotic SET domain lysine methyltransferases (KMTs) and their known histone substrate specificity. (**a**) Classification of KMTs grouped by known substrate specificity and mostly following subfamilies described in mammals. See text for details. Presence of PRMT families in the 22 taxa examined is indicated by the shaded colored circles; solid colors denote presence and pastel colors denotes absence in the taxa studied, but the latter does not indicate that other species in the respective supergroups may not encode these KMTs. (**b**) Documented activities of eukaryotic KMTs on histone H3 and H4 tails. Blue type denotes activity

(Fig. 4b). Within the SAR clade, there are potential homologs in *Plasmodium* (SET3) and *P. sojae*, and some of the other taxa studied here, like *Giardia*, also have putative KMT1 homologs [56], none of which have been studied in great detail. Ciliates like *Paramecium* and especially *Tetrahymena* have been models for gene silencing for decades; here H3K9 methylation is carried out by a protein more similar to KMT6 (E[z]) [85], while true Su(var)3–9 homologs are absent [80]. Within the Amorphea, *Dictyostelium* has one potential homolog that may be mis-annotated as the protein is quite long and is predicted to include a NimA kinase motif (Table 2). Fungi have single KMT1 homologs, though whole families, like budding yeast and its relatives, have lost the ability to methylate H3K9. Only the animals have additional members of H3K9-specific KMTs that catalyze H3K9me in euchromatin or under certain conditions, such as KMT1C (G9a), KMT1E (SETDB1/ESET), and a KMT with a quite different primary structure, KMT8 (RIZ) (Fig. 4a). With between eight and ten proteins, the KMT1 family is expanded in most plant species that have been studied [74, 83, 86]. Shared features among all the KMT1s include pre- and post-SET domains, but the N-terminal chromo domain (CD) is lacking in many family members; the animal-specific KMT1s have additional motifs that are often involved in binding chromatin proteins or histones, such as the Tudor, zinc-binding (ZBD), ankyrin, or methyl-binding domains (MBD).

Homologs for *Saccharomyces* Set1 and *Drosophila* Trx belong to the KMT2 or plant Trx subfamily and catalyze H3K4 methylation, a histone modification associated with active transcription (Fig. 1a); KMT2 homologs were found in most taxa examined (Table 2). Again, while many taxa have single KMT2 homologs, this family is greatly expanded in plants and animals with at least seven different proteins. Pre- and post-SET domains are present, as are PWWP, PHD, ring finger, and FYRC motifs (Fig. 4a). Two additional subfamilies of KMTs are known to act on H3K4, namely KMT7 (SET7/9), found in animals, and KMT3C (SMYD) that in animals includes ~400–450 aa proteins with a large SET domain that is interrupted by a MYND domain. These proteins may act not only on H3K4 but also on other histone residues and indeed non-histone substrates, as many of the protist- or plant-type SMYD proteins may do.

Fig. 4 (continued) demonstrated in the Amorphea (mostly animals and fungi; SET32 is a novel H3K23 MTase from *C. elegans*), green type denotes activity demonstrated in Archaeplastida, gold type denotes activity found in both Archaeplastida and Amorphea, and black type denotes activity found in Archaeplastida, Amorphea, and "excavates." Green arrows denote correlation with gene expression, while red arrows denote correlation with transcriptional gene silencing

Saccharomyces Set2, *Drosophila* Ash1, and mammalian NSD KMTs belong to the KMT3 or plant Ash subfamily, which is well conserved across eukaryotes and is recognized to methylate H3K36. Traditionally, this histone mark has been associated with active transcription, because its appearance is correlated with transcript elongation in *S. cerevisiae*; however, subsequent studies showed that its intrinsic function is to interfere with transcription efficiency [87]. Thus, it is perhaps not surprising that H3K36me2/3 are also involved in the generation and maintenance of facultative heterochromatin, as will be discussed below. Again, plants and animals have more family members than the other taxa examined. One group of proteins originally grouped with KMT2 or Set1 homologs are ASH1 proteins [73]; however, they, as all other KMT3 proteins, contain an AWS domain, and they have by now been shown to catalyze H3K36 methylation, even though earlier studies showed that ASH1 affected H3K4me. In KMT3 proteins, the C-terminal motifs vary widely between proteins among the various taxa examined (Fig. 4a).

Schizosaccharomyces Set9 and animal SUV4-20 proteins constitute the KMT5 subfamily, known to methylate H4K20 and also correlated with the maintenance or generation of heterochromatin; there are overall fewer homologs in the eukaryotes studied here. In fungi, several taxa lack this protein, and in *Arabidopsis*, a Suv subfamily protein, SUVH2, is capable of methylating H4K20. Animals have additional, shorter SET domain proteins (PR-Set7/8) that carry out H4K20 mono-methylation.

Homologs of *Drosophila* E(z) belong to the KMT6 or plant E(z) subfamily and carry out H3K27 methylation, the canonical histone mark for facultative heterochromatin. This family is expanded in plants and animals but not in fungi; *S. cerevisiae* and *S. pombe*, and the industrially or medically important taxa *Aspergillus* and *Penicillium,* lack this protein. Many protists have potential KMT6 homologs (Table 2). Belonging to a different KMT subfamily, the *Arabidopsis* ATXR5 and ATXR6 proteins (plant subfamily IV) carry out H3K27me1 and are involved in regulation of re-replication of heterochromatin [88, 89]; no obvious homologs for these proteins exist in fungi or animals.

As mentioned above, the SMYD (plant subfamily VI) and SETD (plant subfamily VII) are still poorly defined groups in terms of sequence and function and appear greatly expanded in the SAR and deeply branching clades (Table 2, Fig. 4a); many of these KMTs have nonhistone substrates but activity on specific histone residues has been observed both in vivo and in vitro. In conclusion, the distribution and relationships between the extant SET domain subfamilies allows the assertion that histone methylation by these KMTs is an ancient process that was lost in many lineages over evolutionary time. There is strong support for an

ancient origin of the KMT2 (Set1), KMT3 (Set2/Ash1), KMT6 (E[z]), ATXR, SMYD, and SETD subfamilies [80]. In contrast, the KMT1 (Su[var]3-9), KMT5 (Su[var]4-20), KMT7, and KMT8 subfamilies appear to be more recent additions to the ensemble of SET domain KMTs.

3.3.2 Function of Selected SET Domain-Containing Protein Complexes

While discussing the distributions of SET domain KMTs, we already mentioned their preferred histone substrates; there is insufficient space to discuss functional studies that contributed to this general understanding for all KMT function in plants, fungi, and animals. Instead, we will focus on selected aspects of the relationships between KMT2 (Set1), KMT3 (Ash1), and KMT6 (E[z]) complexes. This quickly expanding subject of chromatin biology aims to decipher regulation of opposing chromatin features, for example, how bivalent chromatin promoters influence gene expression, and how the balance of PcG (Polycomb Group)-mediated gene silencing and TrxG (Trithorax Group)-dependent expression affects development and disease [90, 91].

The H3K4 methyltransferase complex, COMPASS (Complex Proteins Associated with Set1), is a highly conserved family of proteins functioning—in combination with other complexes—to maintain developmentally appropriate patterns of gene expression. The subunits of the COMPASS of yeast are comprised of the KMT2 subfamily member Set1, as well as Bre2 (Cps60), Swd1 (Cps50), Spp1 (Cps40), Swd2 (Cps35), Swd3 (Cps30), Sdc1 (Cps25), and Shg1 (Cps15) [92]. While a KMT2 homolog is always present, the subunits can vary greatly among the eukaryotes. Fungi have one COMPASS complex, containing a single homolog of *S. cerevisiae* Set1, but the number of complexes is greatly expanded in *Drosophila* and humans, containing at least three or six COMPASS families, and each capable of H3K4 methylation with non-redundant functions in the cell [93, 94]. Responsible for the "bulk" H3K4me2/3 at promoters and gene bodies of actively transcribed genes, Set1A (KMT2F) COMPASS is important for the regulation of stem cell differentiation [95, 96]. MLL (KMT2A) COMPASS is primarily responsible for the deposition of H3K4me3 marks specifically regulating Hox genes clusters [97], while MLL2 (KMT2B) COMPASS has a role in maintaining bivalent chromatin [98]. Monomethylase activity has been primarily attributed to the MLL3/MLL4 (KMT2C/D) COMPASS—deletion of MLL3 and 4 resulted in substantial losses of monomethylation, particularly at enhancer regions [98]. H3K4 methylation has long been associated with actively transcribed regions as early studies found a correlation between levels of H3K4 methylation and transcriptional activation in *Tetrahymena* macronuclei [99]. Later work established a connection between MLL (KMT2A) activity and Hox gene expression [100]. Chromatin patterns of H3K4 methylation are dependent not only on COMPASS but also on

RNA polymerase II (RNAPII)-mediated transcription. An association between COMPASS and Paf1 (polymerase associated factor 1) has been shown integral for the recruitment of COMPASS to RNAPII and therefore actively transcribed chromatin [101].

In contrast to the TrxG COMPASS complex, PcG proteins form complexes that promote and maintain the formation of repressive facultative heterochromatin. These proteins thus act in direct opposition to COMPASS and H3K4 methylation. Their proper regulation is essential for multicellular development and differentiation, X-chromosome inactivation, and the repression of cancer development. In humans, there are at least two such complexes, Polycomb Repressive Complex 1 and 2 (PRC1 and PRC2). PRC2 catalyzes the deposition of H3K27me3, and PRC1 is believed to either maintain this heterochromatic mark and directly interfere with transcription [102] or act as a guide to bring PRC2 to the appropriate regions [103]; these two options are not mutually exclusive. Subunits of PRC1 vary greatly, but in animals include Polycomb (Pc) a chromatin "reader" protein that binds H3K27me3, suggesting a mutualistic function between PRC1 and PRC2 [104], a catalytic RING protein, which is known to ubiquitinate lysine 119 of histone H2A, and homologs of *Drosophila* polyhomeotic protein (Phc). Even though PRC1 is required in animals, plants and fungi lack clear homologs for most PRC1 subunits [82, 105], although in plants Pc is replaced by a version of HP1, a protein that binds H3K9me3 in other organisms [106]. In fungi, functional homologs for Pc remain to be discovered; most likely there is a completely different group of protein complexes involved.

PRC2 is conserved and likely an ancient protein complex. Three subunits are essential: E(z)/EZH1/2, EED (Early Ectoderm Development), and SUZ12 (Suppressor of Zeste). E(z) homologs are KMT6 lysine methyltransferases, catalyzing H3K27 mono-, di-, and trimethylation. The WD40 beta propeller domain of EED recognizes H3K27me3 and is believed to aid in propagation of the repressive mark. The function of SUZ12 is still not completely understood, though its presence is required for the establishment and maintenance of H3K27me3 [107, 108], and it is likely serving as a "recruitment platform" for additional PRC2 subunits, such as p55/RbAp46/48 and others. In plants, with three E(z) homologs, and in mammals, with two EZH proteins, multiple PRC2 complexes are formed. Targeting of PRC2 in mammals seems to involve CG-rich DNA [108], targeting in *Drosophila* is accomplished by binding to "Polycomb response elements" (PREs) [109, 110], and specific binding motifs may also be important for PRC2 targeting in plants [111]. There are also examples in mammals (ES cells), where non-coding RNAs have been shown as another means to target PRC2 to specific genes [112]. No such elements have been conclusively identified in fungi, suggesting

other mechanisms for PRC2 targeting. Because of their relative simplicity, PRC2 complexes of the basidiomycete yeast, *Cryptococcus*, and the ascomycetes *Neurospora*, *Fusarium*, and *Zymoseptoria* have become models to aid in the general understanding of how PRC2 interacts with other histone marks [113–115].

3.3.3 The Relationship Between H3K36 and H3K27 Methylation

Studies in animals and fungi have suggested antagonism not just between TrxG (H3K4 methylation) and PcG (H3K27 methylation) proteins but also between H3K36 methylation by KMT3/Ash1 and H3K27 methylation by KMT6/E(z). Members of the original KMT3 (ScSet2) subfamily bind to the elongating RNAPII and mono-, di-, or trimethylate H3K36. Distribution of H3K36me3, catalyzed by the single Set2 enzyme in budding yeast, is correlated with active transcription; in other fungi, H3K36me3 covers most annotated genes, though it is more pronounced near the 3′ end of genes. The true function for Set2-catalyzed H3K36me3 is repression of transcription [87], and ScSet2 interacts with the two largest RNAPII subunits by binding to phosphorylated serine 2 of the C-terminal domain [116]. In other fungi, plants, and animals, a second group of KMT3s, the Ash1 homologs, are also capable of H3K36 methylation. All KMT3 enzymes have AWS (associated with SET), WW, and SRI (Set2-Rpb1-interacting) domains but plant and fungal Ash1-like KMT3s lack the C-terminal PHD, BAH, or Bromo domains found in animal Ash1 (Fig. 4a; [117]).

In vitro studies have shown that H3K4me and H3K36me peptides or nucleosomes inhibit reconstituted PRC2 [118, 119] and mutations that change or eliminate H3K36me3 result in mislocalized PcG proteins in animals [120, 121]. Deletion of KMT3/Ash1/SET-3 in *Fusarium fujikuroi* [122] resulted in region-specific increases of H3K27me3 in subtelomeric regions but H3K36 methylation catalyzed by KMT3/Set2 on active genes appeared largely undisturbed; this was similar to findings in *Drosophila* where H3K36me2 inhibited H3K27me3 [123]. Deletion of *kmt6/set-7* in *N. crassa* had no effect on KMT3/Ash1-mediated H3K36me2 in a *kmt3/set-2* strain, while genome-wide loss of Ash1-mediated H3K36me2 resulted in loss (180 regions) or gain (128 regions) of H3K27 methylation and upregulation or transcriptional silencing of genes [124]. Deleting the SRI domain of *Neurospora* KMT3/SET-2 removed most H3K36me3 yet was slightly additive in combination with a catalytically inactive $ash\text{-}1^{Y888F}$ mutation, suggesting that *Neurospora* KMT3/Ash1 catalyzes not just H3K36me2 but also at least some H3K36me3; Ash1 deletion was shown to be lethal in this fungus [124]. Overall, H3K36me2-methylated regions depending on KMT3/Ash1 are associated with poorly transcribed and usually transcriptionally silent genes mostly in subtelomeric regions. Thus, H3K36 methylation on subtelomeric transcriptionally silenced genes is necessary

for the proper accumulation and maintenance of H3K27 methylation in the same regions in both *Neurospora* and *Fusarium,* but sometimes with opposite consequences for transcription. In *F. fujikuroi*, the absence of KMT3/Ash1-catalyzed H3K36me3 resulted in enhanced chromosome instability, measured by the frequency of loss of an "accessory," i.e., conditionally dispensable, chromosome [122]. An increase of H3K27me3 in inappropriate regions, e.g., after loss of H3K9me3, also resulted in increased genome instability in *Z. tritici* [125].

Earlier studies showed that *Drosophila* Ash1 inhibits H3K27me3 accumulation [123] and that the presence of H3K36me histone tails inhibits PRC2 activity [119]. Recent structural studies by cryoelectron microscopy revealed how H3K36me-modified and -unmodified histone tails affect KMT6 directly [126]. PRC2 contacts two nucleosomes: the substrate nucleosome is bound by the EZH2 CXC domain, and the allosteric nucleosome is contacted by EED and the EZH2 SBD and SANT1 domains. In this configuration, H3K36 lies directly opposite to the EZH2-CXC-DNA interaction surface. Positioning of the H3K27 residue in the catalytic center is sensitive to the chemistry of the H3K36 side chain; mutations of H3K36A or H3K36R do not provide a correct fit. Methylation of H3K36 appears to directly interfere with PRC2 catalysis.

Previous studies showed that H3K36 methylation can repress PRC2 activity by PRC2-associated Polycomb-like proteins via their H3K36me3-binding Tudor domains [127–129]. Comparing activity of a full-length PHF1-PRC2 on unmodified and H3Kc36me3 (an H3K36 analog) on mononucleosomes showed that H3K27 methylation was inhibited on H3Kc36me3 mononucleosomes [126], even though one function of PHF1 is to increase PRC2 residence time on nucleosomes [130]. Genetic studies with H3K36R and H3K36A mutant larvae confirmed reduction of H3K27me3 levels, including on HOX genes [126]. Although all of these results suggest a direct influence of KMT3/Ash1-catalyzed H3K36 methylation on PRC2 activity, genetic experiments support the idea that Ash1 catalytic activity may also contribute indirectly, for example, by methylation of non-histone substrates like KMT2/Trx [131, 132].

3.3.4 A Single Enzyme that Methylates both H3K9 and H3K27

The ciliate *Paramecium tetraurelia* has been shown to methylate both H3K9 and H3K27, both in vivo and in vitro, by use of a single enzyme, Ezl1 [85]. The main overlapping function of both histone modifications is transcriptional silencing of transposable elements (TEs), as loss of Ezl1 results in transcription from TEs but not from core genes. This contrasts with the well-studied function of H3K27 methylation in plants, fungi, and animals, where presence or absence of H3K27 methylation controls development and differentiation.

This observation suggested that E(z) proteins are the more ancient subfamily as they are conserved in the SAR clade as well as the Archaeplastida and Amorphea [80, 85]. This "double-marking" is found in several fungi naturally, e.g., in the subtelomeric regions of the ascomycete fungi *N. crassa* [133, 134], *F. graminearum* [135], and *Zymoseptoria* [125] and the basidiomycete yeast *Cryptococcus neoformans* [136]. Marking with both H3K9me3 and H3K27me2/3 is enhanced to cover usually H3K9-methylated regions when HP1 is lacking [134], suggesting that the *Neurospora* E(z) homolog, SET-7, or PRC2 as a whole have intrinsic abilities to be directed toward constitutive heterochromatin. This has also been found in bryophytes like *Marchantia* [137], mammals, and *C. elegans*, at least in certain regions [138–140]. Similarly, when H3K9 methylation is abolished by mutation of the KMT1 homologs of *N. crassa*, *C. neoformans*, or *Z. tritici*, H3K27me3 moves from its usual locations into formerly H3K9-methylated regions again, revealing intrinsic abilities to be directed toward constitutive heterochromatin; this phenomenon has also been observed in mammalian H3K9me3 mutants [141].

The results obtained with *Paramecium* also suggest that H3K9me3 modification predates the evolution of a dedicated enzyme, like the members of the KMT1 subfamily, and this is borne out by results in other ciliates, diatoms, and *Chlamydomonas* where KMT1 homologs are lacking [80, 85, 105, 142, 143]. How then is recognition of the correct target sequence accomplished? Three-dimensional structures of the C-terminal SET domains of HMTs reveal much of what we know about the domain's function. In protein databases one will find many structures of SET domains and SET domain-containing proteins, both in *apo* form or complexed with ligands and cofactors, such as the methyl donor SAM, the cofactor product SAH, substrate peptides, and zinc. While structural variations exist between SET domains of different KMT subfamilies, key structural features apply to all SET domains. There are two distinct architectural features, a conserved β-barrel and a pseudoknot structure comprising the enzyme's active site. These anti-parallel β-sheets position the catalytic residues of the SET domain (e.g., N688, H689, and Y726 in human EZH2), separated by approximately 36 amino acids in the primary sequence, into the active site fold. While there is debate in the field as to the exact mechanism of the methylation of a substrate peptide of a SET domain-containing protein, catalysis by KMT requires these conserved residues and a protonated amino group on the substrate lysine. As the substrate enters the active site, a hydrogen bond between the catalytic tyrosine and the amide proton is enough to change the electronic chemistry of the nitrogen at the N-terminus of the lysine, promoting its nucleophilic attack on the sulphonium methyl group of the SAM cofactor. These catalytic intermediates

are further stabilized by other active site residues (N688 and H689), and they, in combination with the binding pocket, promote the release of the cofactor product SAH [144].

What dictates the precise chemical specificity against substrates of SET domain KMTs is heavily studied, especially with an eye to pharmacological interventions [50, 55]. It results from a combination of amino acids comprising the enzyme's substrate binding site as well as the consensus amino acid motifs in biological targets. Histone tails are quite basic; the interactions that drive substrate specificity are largely polar in nature [145, 146]. For the two well-studied targets of methylation by KMT1 and KMT6, H3K9 and H3K27, the sequences directly flanking the target lysines in the histone H3 tail are identical: A-R-K-S, and yet the two enzymes that catalyze these reactions, KMT1/Su(var)3-9 and KMT6/EZH2, specifically recognize which lysine to methylate. Adjacent to the conserved ARKS region, the consensus is different, T-A-R-K-S-T for Su(var)3–9, compared to A-A-R-K-S-A for EZH2, and thus the two enzymes may achieve specificity by recognizing a different binding pocket or tunnel, and by their different catalytic sites. The intriguing study on *Paramecium* Ezl1 compared the SET domain features of *Neurospora* KMT1 (DIM-5), human EZH2, and Ezl1, and modeled potential catalytic site interactions that would allow dual specificity [85]. Based on this, and additional studies of protist and fungal enzymes, one goal for the future is to take advantage of the ancient origin of KMT2, KMT3, and KMT6 enzymes to study the catalytic characteristics of extant or "reverse evolved" enzymes, as has been done successfully with families of extant and ancient transcription factors [147, 148].

4 Summary

Reports from the literature and our representative sampling of eukaryotic supergroups allow the conclusion that protein methyltransferases are ancient. There is evidence that PRMTs arose with the last common eukaryotic ancestor but that SET domain KMTs are even more ancient. Phylogenies of PRMTs for PRMT5, the PRMT1^{PRO} group, PRMT7, PRMT10, PRMT4, and PRMT1 suggest monophyletic placement but the relationships between PRMT2, PRMT3, and PRMT6 are more difficult to resolve. Dot1-like non-SET KMTs occur most often as single proteins in eukaryotes, although genome or gene duplications can result in species with two specialized homologs (e.g., in *T. brucei*), which allows for functional specialization (e.g., for mono-, di, or trimethylation). In animals, several Dot1L genes or splice variants are found. However, Dot1 is not universally conserved in eukaryotes, as some fungi and most plants do not have homologs. Phylogenetic analyses of SET domain proteins suggest the

existence of more than a dozen different subfamilies, based on the uncertain relationship in several subfamilies, and much work remains to establish the substrates and functions for many of these proteins, especially in the less well understood taxa.

Overall, the distribution between extant PRMT and SET domain subfamilies allows the assertion that histone methylation by these proteins is an ancient process that was lost in many lineages over evolutionary time. For example, there is much evidence to support the ancient origin of the KMT2 (Set1), KMT3 (Set2/Ash1), KMT6 (E[z]), ATXR, SMYD, and SETD subfamilies, while the KMT1 (Su[var]3-9), KMT5 (Su[var]4-20), KMT7 (SET7/9), and KMT8 (RIZ) subfamilies are of more recent origin. Studies of *Paramecium* Ezl1, a dual-specificity KMT6-type protein that catalyzes both H3K9me and H3K27me, open the door to interesting new mechanistic studies that may allow the "reverse evolution" of some extant KMTs to uncover their evolutionary origin and capabilities.

Acknowledgments

We thank the members of the Freitag Lab for discussions. We apologize to our colleagues whose original research is not cited because of space constraints. Accession numbers and sequences of proteins discussed are available upon request. Work in the Freitag lab is supported by grants from the NSF (MCB1818006), the NIH (R01GM132644), and the US-Israel BSF (#2019034).

References

1. Allis CD (2015) Epigenetics, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
2. Johnson L, Mollah S, Garcia BA, Muratore TL, Shabanowitz J, Hunt DF, Jacobsen SE (2004) Mass spectrometry analysis of *Arabidopsis* histone H3 reveals distinct combinations of post-translational modifications. Nucleic Acids Res 32:6511–6518
3. Wu T, Yuan T, Tsai SN, Wang C, Sun SM, Lam HM, Ngai SM (2009) Mass spectrometry analysis of the variants of histone H3 and H4 of soybean and their post-translational modifications. BMC Plant Biol 9:98
4. Lu CC, Coradin M, Porter EG, Garcia BA (2021) Accelerating the field of epigenetic histone modification through mass spectrometry-based approaches. Mol Cell Proteomics 20:100006
5. Sidoli S, Kori Y, Lopes M, Yuan ZF, Kim HJ, Kulej K, Janssen KA, Agosto LM, da Cunha JPC, Andrews AJ, Garcia BA (2019) One minute analysis of 200 histone posttranslational modifications by direct injection mass spectrometry. Genome Res 29:978–987
6. Garcia B (2016) Quantitative proteomics for understanding the histone code. J Biomol Tech 24(supplement):S10
7. Allshire RC, Madhani HD (2018) Ten principles of heterochromatin formation and function. Nat Rev Mol Cell Bio 19:229–244
8. Iyer LM, Anantharaman V, Wolf MY, Aravind L (2008) Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int J Parasitol 38:1–31
9. Burki F, Roger AJ, Brown MW, Simpson AGB (2020) The new tree of eukaryotes. Trends Ecol Evol 35:43–55
10. Burki F (2014) The eukaryotic tree of life from a global Phylogenomic perspective. CSH Perspect Biol 6:a016147

11. Leger MM, Kolisko M, Kamikawa R, Stairs CW, Kume K, Cepicka I, Silberman JD, Andersson JO, Xu FF, Yabuki A, Eme L, Zhang QQ, Takishita K, Inagaki Y, Simpson AGB, Hashimoto T, Roger AJ (2017) Organelles that illuminate the origins of *trichomonas* hydrogenosomes and *giardia* mitosomes. Nat Ecol Evol 1:0092
12. Morrison HG, McArthur AG, Gillin FD, Aley SB, Adam RD, Olsen GJ, Best AA, Cande WZ, Chen F, Cipriano MJ, Davids BJ, Dawson SC, Elmendorf HG, Hehl AB, Holder ME, Huse SM, Kim UU, Lasek-Nesselquist E, Manning G, Nigam A, Nixon JEJ, Palm D, Passamaneck NE, Prabhu A, Reich CI, Reiner DS, Samuelson J, Svard SG, Sogin ML (2007) Genomic minimalism in the early diverging intestinal parasite *Giardia lamblia*. Science 317:1921–1926
13. Xu FF, Jex A, Svard SG (2020) A chromosome-scale reference genome for *Giardia intestinalis* WB. Sci Data 7:38
14. Adam RD (2000) The *Giardia lamblia* genome. Int J Parasitol 30:475–484
15. Sodre CL, Chapeaurouge AD, Kalume DE, Lima LD, Perales J, Fernandes O (2009) Proteomic map of *Trypanosoma cruzi* CL Brener: the reference strain of the genome project. Arch Microbiol 191:177–184
16. El-Sayed NM, Myler PJ, Bartholomeu DC, Nilsson D, Aggarwal G, Tran AN, Ghedin E, Worthey EA, Delcher AL, Blandin G, Westenberger SJ, Caler E, Cerqueira GC, Branche C, Haas B, Anupama A, Arner E, Aslund L, Attipoe P, Bontempi E, Bringaud F, Burton P, Cadag E, Campbell DA, Carrington M, Crabtree J, Darban H, da Silveira JF, de Jong P, Edwards K, Englund PT, Fazelina G, Feldblyum T, Ferella M, Frasch AC, Gull K, Horn D, Hou LH, Huang YT, Kindlund E, Ktingbeil M, Kluge S, Koo H, Lacerda D, Levin MJ, Lorenzi H, Louie T, Machado CR, McCulloch R, McKenna A, Mizuno Y, Mottram JC, Nelson S, Ochaya S, Osoegawa K, Pai G, Parsons M, Pentony M, Pettersson U, Pop M, Ramirez JL, Rinta J, Robertson L, Salzberg SL, Sanchez DO, Seyler A, Sharma R, Shetty J, Simpson AJ, Sisk E, Tammi MT, Tarteton R, Teixeira S, Van Aken S, Vogt C, Ward PN, Wickstead B, Wortman J, White O, Fraser CM, Stuart KD, Andersson B (2005) The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science 309:409–415
17. Gardner MJ, Hall N, Fung E, White O, Berriman M, Hyman RW, Carlton JM, Pain A, Nelson KE, Bowman S, Paulsen IT, James K, Eisen JA, Rutherford K, Salzberg SL, Craig A, Kyes S, Chan MS, Nene V, Shallom SJ, Suh B, Peterson J, Angiuoli S, Pertea M, Allen J, Selengut J, Haft D, Mather MW, Vaidya AB, Martin DMA, Fairlamb AH, Fraunholz MJ, Roos DS, Ralph SA, McFadden GI, Cummings LM, Subramanian GM, Mungall C, Venter JC, Carucci DJ, Hoffman SL, Newbold C, Davis RW, Fraser CM, Barrell B (2002) Genome sequence of the human malaria parasite *plasmodium falciparum*. Nature 419:498–511
18. Aranda M, Li Y, Liew YJ, Baumgarten S, Simakov O, Wilson MC, Piel J, Ashoor H, Bougouffa S, Bajic VB, Ryu T, Ravasi T, Bayer T, Micklem G, Kim H, Bhak J, LaJeunesse TC, Voolstra CR (2016) Genomes of coral dinoflagellate symbionts highlight evolutionary adaptations conducive to a symbiotic lifestyle. Sci Rep 6:39734
19. Bouligand Y, Norris V (2001) Chromosome separation and segregation in dinoflagellates and bacteria may depend on liquid crystalline states. Biochimie 83:187–192
20. Rizzo PJ (1991) The enigma of the dinoflagellate chromosome. J Protozool 38:246–252
21. Rizzo PJ (2003) Those amazing dinoflagellate chromosomes. Cell Res 13:215–217
22. Gonzalez-Pech RA, Stephens TG, Chen YB, Mohamed AR, Cheng YY, Shah S, Dougan KE, Fortuin MDA, Lagorce R, Burt DW, Bhattacharya D, Ragan MA, Chan CX (2021) Comparison of 15 dinoflagellate genomes reveals extensive sequence and structural divergence in family Symbiodiniaceae and genus Symbiodinium. BMC Biol 19:73
23. Preer JR Jr (2000) Epigenetic mechanisms affecting macronuclear development in *paramecium* and *Tetrahymena*. J Eukaryot Microbiol 47:515–524
24. Ruiz F, Vayssie L, Klotz C, Sperling L, Madeddu L (1998) Homology-dependent gene silencing in *paramecium*. Mol Biol Cell 9:931–943
25. Cummings DJ, Tait A, Goddard JM (1974) Methylated bases in DNA from *Paramecium aurelia*. Biochim Biophys Acta 374:1–11
26. Aury JM, Jaillon O, Duret L, Noel B, Jubin C, Porcel BM, Segurens B, Daubin V, Anthouard V, Aiach N, Arnaiz O, Billaut A, Beisson J, Blanc I, Bouhouche K, Camara F, Duharcourt S, Guigo R, Gogendeau D, Katinka M, Keller AM, Kissmehl R, Klotz C, Koll F, Le Mouel A, Lepere G, Malinsky S, Nowacki M, Nowak JK, Plattner H, Poulain J, Ruiz F, Serrano V, Zagulski M, Dessen P, Betermier M, Weissenbach J, Scarpelli C, Schachter V, Sperling L, Meyer E, Cohen J,

Wincker P (2006) Global trends of whole-genome duplications revealed by the ciliate *paramecium tetraurelia*. Nature 444:171–178

27. Agyekum TP, Botwe PK, Arko-Mensah J, Issah I, Acquah AA, Hogarh JN, Dwomoh D, Robins TG, Fobil JN (2021) A systematic review of the effects of temperature on *anopheles* mosquito development and survival: implications for malaria control in a future warmer climate. Int J Environ Res Public Health 18:7255
28. Clement R, Dimnet L, Maberly SC, Gontero B (2016) The nature of the CO2-concentrating mechanisms in a marine diatom, *Thalassiosira pseudonana*. New Phytol 209:1417–1427
29. Duan DL, Wang XL, Yao JT, Liu DY, Hong WG, Hu ZM, Xia YH, Shao ZR, Li QY, Zhang J (2013) Draft genome sequences of *Saccharina japonica*. Phycologia 52:27–27
30. Liu T, Wang XM, Wang GL, Jia SG, Liu GM, Shan GL, Chi S, Zhang J, Yu YH, Xue T, Yu J (2019) Evolution of complex thallus alga: genome sequencing of *Saccharina japonica*. Front Genet 10:378
31. Ye NH, Zhang XW, Miao M, Fan X, Zheng Y, Xu D, Wang JF, Zhou L, Wang DS, Gao Y, Wang YT, Shi WY, Ji PF, Li DM, Guan Z, Shao CW, Zhuang ZM, Gao ZW, Qi J, Zhao FQ (2015) *Saccharina* genomes provide novel insight into kelp biology. Nat Commun 6:6986
32. Mao YX, Tyler BM (1996) The *Phytophthora sojae* genome contains tandem repeat sequences which vary from strain to strain. Fungal Genet Biol 20:43–51
33. Zhang X, Liu B, Zou F, Shen DY, Yin ZY, Wang RB, He F, Wang YC, Tyler BM, Fan W, Qian WQ, Dou DL (2019) Whole genome re-sequencing reveals natural variation and adaptive evolution of Phytophthora sojae. Front Microbiol 10:2792
34. Read BA, Kegel J, Klute MJ, Kuo A, Lefebvre SC, Maumus F, Mayer C, Miller J, Monier A, Salamov A, Young J, Aguilar M, Claverie JM, Frickenhaus S, Gonzalez K, Herman EK, Lin YC, Napier J, Ogata H, Sarno AF, Shmutz J, Schroeder D, de Vargas C, Verret F, von Dassow P, Valentin K, Van de Peer Y, Wheeler G, Dacks JB, Delwiche CF, Dyhrman ST, Glockner G, John U, Richards T, Worden AZ, Zhang XY, Grigoriev IV, Allen AE, Bidle K, Borodovsky M, Bowler C, Brownlee C, Cock JM, Elias M, Gladyshev VN, Groth M, Guda C, Hadaegh A, Iglesias-Rodriguez MD, Jenkins J, Jones BM, Lawson T, Leese F, Lindquist E, Lobanov A, Lomsadze A, Malik SB, Marsh ME, Mackinder L, Mock T, Mueller-Roeber B, Pagarete A, Parker M, Probert I, Quesneville H, Raines C, Rensing SA, Riano-Pachon DM, Richier S, Rokitta S, Shiraiwa Y, Soanes DM, van der Giezen M, Wahlund TM, Williams B, Wilson W, Wolfe G, Wurch LL, Annotation EH (2013) Pan genome of the phytoplankton *Emiliania* underpins its global distribution. Nature 499:209–213
35. Westbroek P, Brown CW, Vanbleijswijk J, Brownlee C, Brummer GJ, Conte M, Egge J, Fernandez E, Jordan R, Knappertsbusch M, Stefels J, Veldhuis M, Vanderwal P, Young J (1993) A model system approach to biological climate forcing - the example of *Emiliania huxleyi*. Glob Planet Chang 8:27–46
36. Seitz R (2011) Bright water: hydrosols, water conservation and climate change. Clim Chang 105:365–381
37. Lefebvre SC, Benner I, Stillman JH, Parker AE, Drake MK, Rossignol PE, Okimura KM, Komada T, Carpenter EJ (2012) Nitrogen source and pCO(2) synergistically affect carbon allocation, growth and morphology of the coccolithophore *Emiliania huxleyi*: potential implications of ocean acidification for the carbon cycle. Glob Chang Biol 18:493–503
38. Archibald JM, Lane CE (2009) Going, going, not quite gone: Nucleomorphs as a case study in nuclear genome reduction. J Hered 100:582–590
39. Curtis BA, Tanifuji G, Burki F, Gruber A, Irimia M, Maruyama S, Arias MC, Ball SG, Gile GH, Hirakawa Y, Hopkins JF, Kuo A, Rensing SA, Schmutz J, Symeonidi A, Elias M, Eveleigh RJM, Herman EK, Klute MJ, Nakayama T, Obornik M, Reyes-Prieto A, Armbrust EV, Aves SJ, Beiko RG, Coutinho P, Dacks JB, Durnford DG, Fast NM, Green BR, Grisdale CJ, Hempel F, Henrissat B, Hoppner MP, Ishida KI, Kim E, Koreny LK, Kroth PG, Liu Y, Malik SB, Maier UG, McRose D, Mock T, Neilson JAD, Onodera NT, Poole AM, Pritham EJ, Richards TA, Rocap G, Roy SW, Sarai C, Schaack S, Shirato S, Slamovits CH, Spencer DF, Suzuki S, Worden AZ, Zauner S, Barry K, Bell C, Bharti AK, Crow JA, Grimwood J, Kramer R, Lindquist E, Lucas S, Salamov A, McFadden GI, Lane CE, Keeling PJ, Gray MW, Grigoriev IV, Archibald JM (2012)

Algal genomes reveal evolutionary mosaicism and the fate of nucleomorphs. Nature 492: 59–65

40. Eichinger L, Pachebat JA, Glockner G, Rajandream MA, Sucgang R, Berriman M, Song J, Olsen R, Szafranski K, Xu Q, Tunggal B, Kummerfeld S, Madera M, Konfortov BA, Rivero F, Bankier AT, Lehmann R, Hamlin N, Davies R, Gaudet P, Fey P, Pilcher K, Chen G, Saunders D, Sodergren E, Davis P, Kerhornou A, Nie X, Hall N, Anjard C, Hemphill L, Bason N, Farbrother P, Desany B, Just E, Morio T, Rost R, Churcher C, Cooper J, Haydock S, van Driessche N, Cronin A, Goodhead I, Muzny D, Mourier T, Pain A, Lu M, Harper D, Lindsay R, Hauser H, James K, Quiles M, Babu MM, Saito T, Buchrieser C, Wardroper A, Felder M, Thangavelu M, Johnson D, Knights A, Loulseged H, Mungall K, Oliver K, Price C, Quail MA, Urushihara H, Hernandez J, Rabbinowitsch E, Steffen D, Sanders M, Ma J, Kohara Y, Sharp S, Simmonds M, Spiegler S, Tivey A, Sugano S, White B, Walker D, Woodward J, Winckler T, Tanaka Y, Shaulsky G, Schleicher M, Weinstock G, Rosenthal A, Cox EC, Chisholm RL, Gibbs R, Loomis WF, Platzer M, Kay RR, Williams J, Dear PH, Noegel AA, Barrell B, Kuspa A (2005) The genome of the social amoeba *Dictyostelium discoideum*. Nature 435:43–57
41. Tang X, Zhao L, Chen H, Chen YQ, Chen W, Song Y, Ratledge C (2015) Complete genome sequence of a high lipid-producing strain of *Mucor circinelloides* WJ11 and comparative genome analysis with a low lipid-producing strain CBS 277.49. PLoS One 10: e0137543
42. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, Feldmann H, Galibert F, Hoheisel JD, Jacq C, Johnston M, Louis EJ, Mewes HW, Murakami Y, Philippsen P, Tettelin H, Oliver SG (1996) Life with 6000 genes. Science 274:546–567
43. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne-M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre S, Kamal M, Kamvysselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catcheside D, Li W, Pratt RJ, Osmani SA, DeSouza CP, Glass L, Orbach MJ, Berglund JA, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA, Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, Birren B (2003) The genome sequence of the filamentous fungus *Neurospora crassa*. Nature 422:859–868
44. Kämper J, Kahmann R, Bolker M, Ma LJ, Brefort T, Saville BJ, Banuett F, Kronstad JW, Gold SE, Muller O, Perlin MH, Wosten HA, de Vries R, Ruiz-Herrera J, Reynaga-Pena CG, Snetselaar K, McCann M, Perez-Martin J, Feldbrügge M, Basse CW, Steinberg G, Ibeas JI, Holloman W, Guzman P, Farman M, Stajich JE, Sentandreu R, Gonzalez-Prieto JM, Kennell JC, Molina L, Schirawski J, Mendoza-Mendoza A, Greilinger D, Munch K, Rossel N, Scherer M, Vranes M, Ladendorf O, Vincon V, Fuchs U, Sandrock B, Meng S, Ho EC, Cahill MJ, Boyce KJ, Klose J, Klosterman SJ, Deelstra HJ, Ortiz-Castellanos L, Li W, Sanchez-Alonso P, Schreier PH, Hauser-Hahn I, Vaupel M, Koopmann E, Friedrich G, Voss H, Schluter T, Margolis J, Platt D, Swimmer C, Gnirke A, Chen F, Vysotskaia V, Mannhaupt G, Guldener U, Munsterkotter M, Haase D, Oesterheld M, Mewes HW, Mauceli EW, DeCaprio D, Wade CM, Butler J, Young S, Jaffe DB, Calvo S, Nusbaum C, Galagan J, Birren BW (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. Nature 444:97–101
45. Chapman JA, Kirkness EF, Simakov O, Hampson SE, Mitros T, Weinmaier T, Rattei T, Balasubramanian PG, Borman J, Busam D, Disbennett K, Pfannkoch C, Sumin N, Sutton GG, Viswanathan LD, Walenz B, Goodstein DM, Hellsten U, Kawashima T, Prochnik SE, Putnam NH, Shu SQ, Blumberg B, Dana CE, Gee L, Kibler DF, Law L, Lindgens D, Martinez DE, Peng JS, Wigge PA, Bertulat B, Guder C, Nakamura Y, Ozbek S, Watanabe H, Khalturin K, Hemmrich G, Franke A, Augustin R, Fraune S, Hayakawa E, Hayakawa S, Hirose M, Hwang JS, Ikeo K, Nishimiya-Fujisawa C, Ogura A, Takahashi T, Steinmetz PRH, Zhang XM, Aufschnaiter R, Eder MK, Gorny AK, Salvenmoser W, Heimberg AM, Wheeler BM, Peterson KJ, Boettger A, Tischler P, Wolf A, Gojobori T, Remington KA, Strausberg RL, Venter JC, Technau U,

Hobmayer B, Bosch TCG, Holstein TW, Fujisawa T, Bode HR, David CN, Rokhsar DS, Steele RE (2010) The dynamic genome of *hydra*. Nature 464:592–596

46. Adema CM, Hillier LW, Jones CS, Loker ES, Knight M, Minx P, Oliveira G, Raghavan N, Shedlock A, do Amaral LR, Arican-Goktas HD, Assis JG, Baba EH, Baron OL, Bayne CJ, Bickham-Wright U, Biggar KK, Blouin M, Bonning BC, Botka C, Bridger JM, Buckley KM, Buddenborg SK, Caldeira RL, Carleton J, Carvalho OS, Castillo MG, Chalmers IW, Christensens M, Clifton S, Cosseau C, Coustau C, Cripps RM, Cuesta-Astroz Y, Cummins SF, Di Stephano L, Dinguirard N, Duval D, Emrich S, Feschotte C, Feyereisen R, FitzGerald P, Fronick C, Fulton L, Galinier R, Gava SG, Geusz M, Geyer KK, Giraldo-Calderon GI, Gomes MD, Gordy MA, Gourbal B, Grunau C, Hanington PC, Hoffmann KF, Hughes D, Humphries J, Jackson DJ, Jannotti-Passos LK, Jeremias WD, Jobling S, Kamel B, Kapusta A, Kaur S, Koene JM, Kohn AB, Lawson D, Lawton SP, Liang D, Limpanont Y, Liu SJ, Lockyer AE, Lovato TL, Ludolf F, Magrini V, McManus DP, Medina M, Misra M, Mitta G, Mkoji GM, Montague MJ, Montelongo C, Moroz LL, Munoz-Torres MC, Niazi U, Noble LR, Oliveira FS, Pais FS, Papenfuss AT, Peace R, Pena JJ, Pila EA, Quelais T, Raney BJ, Rast JP, Rollinson D, Rosse IC, Rotgans B, Routledge EJ, Ryan KM, Scholte LLS, Storey KB, Swain M, Tennessen JA, Tomlinson C, Trujillo DL, Volpi EV, Walker AJ, Wang T, Wannaporn I, Warren WC, Wu XJ, Yoshino TP, Yusuf M, Zhang SM, Zhao M, Wilson RK (2017) Whole genome analysis of a schistosomiasis-transmitting freshwater snail. Nat Commun 8:15451
47. Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Henderson SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YHC, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Miklos GLG, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies A, de Pablos B, Delcher A, Deng ZM, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gabrielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong FC, Gorrell JH, Gu ZP, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Houston DA, Howland TJ, Wei MH, Ibegwam C, Jalali M, Kalush F, Karpen GH, Ke ZX, Kennison JA, Ketchum KA, Kimmel BE, Kodira CD, Kraft C, Kravitz S, Kulp D, Lai ZW, Lasko P, Lei YD, Levitsky AA, Li JY, Li ZY, Liang Y, Lin XY, Liu XJ, Mattei B, McIntosh TC, McLeod MP, McPherson D, Merkulov G, Milshina NV, Mobarry C, Morris J, Moshrefi A, Mount SM, Moy M, Murphy B, Murphy L, Muzny DM, Nelson DL, Nelson DR, Nelson KA, Nixon K, Nusskern DR, Pacleb JM, Palazzolo M, Pittman GS, Pan S, Pollard J, Puri V, Reese MG, Reinert K, Remington K, Saunders RDC, Scheeler F, Shen H, Shue BC, Siden-Kiamos I, Simpson M, Skupski MP, Smith T, Spier E, Spradling AC, Stapleton M, Strong R, Sun E, Svirskas R, Tector C, Turner R, Venter E, Wang AHH, Wang X, Wang ZY, Wassarman DA, Weinstock GM, Weissenbach J, Williams SM, Woodage T, Worley KC, Wu D, Yang S, Yao QA, Ye J, Yeh RF, Zaveri JS, Zhan M, Zhang GG, Zhao Q, Zheng LS, Zheng XQH, Zhong FN, Zhong WY, Zhou XJ, Zhu SP, Zhu XH, Smith HO, Gibbs RA, Myers EW, Rubin GM, Venter JC (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195
48. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XQH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang JH, Miklos GLG, Nelson C, Broder S, Clark AG, Nadeau C, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng ZM, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge WM, Gong FC, Gu ZP, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke ZX, Ketchum KA, Lai ZW, Lei YD, Li ZY, Li

JY, Liang Y, Lin XY, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue BX, Sun JT, Wang ZY, Wang AH, Wang X, Wang J, Wei MH, Wides R, Xiao CL, Yan CH, Yao A, Ye J, Zhan M, Zhang WQ, Zhang HY, Zhao Q, Zheng LS, Zhong F, Zhong WY, Zhu SPC, Zhao SY, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An HJ, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigo R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi HY, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu XJ, Lopez J, Ma D, Majoros W, McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen MY, Wu D, Wu M, Xia A, Zandieh A, Zhu XH (2001) The sequence of the human genome. Science 291:1304–1351

49. Weiner AKM, Ceron-Romero MA, Yan Y, Katz LA (2020) Phylogenomics of the epigenetic toolkit reveals punctate retention of genes across eukaryotes. Genome Biol Evol 12:2196–2210
50. Hwang JW, Cho Y, Bae GU, Kim SN, Kim YK (2021) Protein arginine methyltransferases: promising targets for cancer therapy. Exp Mol Med 53:788–808
51. Liu CY, Lu FL, Cui X, Cao XF (2010) Histone methylation in higher plants. Annu Rev Plant Biol 61:395–420
52. Zhang X, Zhou L, Cheng XD (2000) Crystal structure of the conserved core of protein arginine methyltransferase PRMT3. EMBO J 19:3509–3519
53. Zhang M, Xu JY, Hu H, Ye BC, Tan MJ (2018) Systematic proteomic analysis of protein methylation in prokaryotes and eukaryotes revealed distinct substrate specificity. Proteomics 18:1700300
54. Tewary SK, Zheng YG, Ho MC (2019) Protein arginine methyltransferases: insights into the enzyme structure and mechanism at the atomic level. Cell Mol Life Sci 76:2917–2932
55. Kaniskan HU, Martini ML, Jin J (2018) Inhibitors of protein methyltransferases and demethylases. Chem Rev 118:989–1068
56. Emery-Corbin SJ, Hamey JJ, Ansell BRE, Balan B, Tichkule S, Stroehlein AJ, Cooper C, McInerney BV, Hediyeh-Zadeh S, Vuong D, Crombie A, Lacey E, Davis MJ, Wilkins MR, Bahlo M, Svard SG, Gasser RB, Jex AR (2020) EukaryoteConservedMethylarginine is absent in diplomonads and functionally compensated in *giardia*. Mol Biol Evol 37:3525–3549
57. Fisk JC, Read LK (2011) Protein arginine methylation in parasitic protozoa. Eukaryot Cell 10:1013–1022
58. Lott K, Zhu L, Fisk JC, Tomasello DL, Read LK (2014) Functional interplay between protein arginine methyltransferases in *Trypanosoma brucei*. Microbiol Open 3:595–609
59. Hashimoto H, Kafkova L, Raczkowski A, Jordan KD, Read LK, Debler EW (2020) Structural basis of protein arginine methyltransferase activation by a catalytically dead homolog (Prozyme). J Mol Biol 432: 410–426
60. Kafkova L, Debler EW, Fisk JC, Jain K, Clarke SG, Read LK (2017) The major protein arginine methyltransferase in *Trypanosoma brucei* functions as an enzyme-Prozyme complex. J Biol Chem 292:2089–2100
61. Campagnaro GD, Nay E, Plevin MJ, Cruz AK, Walrad PB (2021) Arginine methyltransferases as regulators of RNA-binding protein activities in pathogenic Kinetoplastids. Front Mol Biosci 8:692668
62. Fan Q, Miao J, Cui L, Cui LW (2009) Characterization of PRMT1 from *plasmodium falciparum*. Biochem J 421:107–118

63. Zeeshan M, Kaur I, Joy J, Saini E, Paul G, Kaushik A, Dabral S, Mohmmed A, Gupta D, Malhotra P (2017) Proteomic identification and analysis of arginine-methylated proteins of *plasmodium falciparum* at asexual blood stages. J Proteome Res 16:368–383
64. Ahmad A, Dong YZ, Cao XF (2011) Characterization of the PRMT gene family in Rice reveals conservation of arginine methylation. PLoS One 6:e22664
65. Borkovich KA, Alex LA, Yarden O, Freitag M, Turner GE, Read ND, Seiler S, Bell-Pedersen-D, Paietta J, Plesofsky N, Plamann M, Goodrich-Tanrikulu M, Schulte U, Mannhaupt G, Nargang FE, Radford A, Selitrennikoff C, Galagan JE, Dunlap JC, Loros JJ, Catcheside D, Inoue H, Aramayo R, Polymenis M, Selker EU, Sachs MS, Marzluf GA, Paulsen I, Davis R, Ebbole DJ, Zelter A, Kalkman ER, O'Rourke R, Bowring F, Yeadon J, Ishii C, Suzuki K, Sakai W, Pratt R (2004) Lessons from the genome sequence of *Neurospora crassa*: tracing the path from genomic blueprint to multicellular organism. Microbiol Mol Biol Rev 68:1–108
66. Schmoll M, Dattenbock C, Carreras-Villasenor N, Mendoza-Mendoza A, Tisch D, Aleman MI, Baker SE, Brown C, Cervantes-Badillo MG, Cetz-Chel J, Cristobal-Mondragon GR, Delaye L, Esquivel-Naranjo EU, Frischmann A, Gallardo-Negrete Jde J, Garcia-Esquivel M, Gomez-Rodriguez EY, Greenwood DR, Hernandez-Onate M, Kruszewska JS, Lawry R, Mora-Montes HM, Munoz-Centeno T, Nieto-Jacobo MF, Nogueira Lopez G, Olmedo-Monfil V, Osorio-Concepcion M, Pilsyk S, Pomraning KR, Rodriguez-Iglesias A, Rosales-Saavedra MT, Sanchez-Arreguin JA, Seidl-Seiboth V, Stewart A, Uresti-Rivera EE, Wang CL, Wang TF, Zeilinger S, Casas-Flores S, Herrera-Estrella A (2016) The genomes of three uneven siblings: footprints of the lifestyles of three *Trichoderma* species. Microbiol Mol Biol Rev 80:205–327
67. Wang YC, Li C (2012) Evolutionarily conserved protein arginine methyltransferases in non-mammalian animal systems. FEBS J 279: 932–945
68. Jahan S, Xu WN, Ilic A, Davie JR (2017) Protein arginine methyltransferases (PRMTs) and transcriptionally active chromatin domains. Biochem Cell Biol 95:182–182
69. Jain K, Clarke SG (2019) PRMT7 as a unique member of the protein arginine methyltransferase family: a review. Arch Biochem Biophys 665:36–45
70. Singer MS, Kahana A, Wolf AJ, Meisinger LL, Peterson SE, Goggin C, Mahowald M, Gottschling DE (1998) Identification of high-copy disruptors of telomeric silencing in *Saccharomyces cerevisiae*. Genetics 150: 613–632
71. van Leeuwen F, Gafken PR, Gottschling DE (2002) Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell 109:745–756
72. Kim W, Choi M, Kim JE (2014) The histone methyltransferase Dot1/DOT1L as a critical regulator of the cell cycle. Cell Cycle 13:726–738
73. Allis CD, Berger SL, Cote J, Dent S, Jenuwein T, Kouzarides T, Pillus L, Reinberg D, Shi Y, Shiekhattar R, Shilatifard A, Workman J, Zhang Y (2007) New nomenclature for chromatin-modifying enzymes. Cell 131:633–636
74. Pontvianne F, Blevins T, Pikaard CS (2010) *Arabidopsis* histone lysine methyltransferases. Adv Bot Res 53:1–22
75. Min J, Feng Q, Li Z, Zhang Y, Xu RM (2003) Structure of the catalytic domain of human DOT1L, a non-SET domain Nucleosomal histone methyltransferase. Cell 112:711–723
76. Ng HH, Xu RM, Zhang Y, Struhl K (2002) Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. J Biol Chem 277:34655–34657
77. Worden EJ, Wolberger C (2019) Activation and regulation of H2B-ubiquitin-dependent histone methyltransferases. Curr Opin Struc Biol 59:98–106
78. Valencia-Sanchez MI, De Ioannes P, Wang M, Truong DM, Lee R, Armache JP, Boeke JD, Armache KJ (2021) Regulation of the Dot1 histone H3K79 methyltransferase by histone H4K16 acetylation. Science 371(363): eabc6663
79. Cui LW, Fan Q, Cui L, Miao J (2008) Histone lysine methyltransferases and demethylases in *plasmodium falciparum*. Int J Parasitol 38:1083–1097
80. Lhuillier-Akakpo M, Frapporti A, Denby Wilkes C, Matelot M, Vervoort M, Sperling L, Duharcourt S (2014) Local effect of enhancer of Zeste-like reveals cooperation of epigenetic and cis-acting determinants for zygotic genome rearrangements. PLoS Genet 10:e1004665

81. Alvarez-Venegas R (2014) Bacterial SET domain proteins and their role in eukaryotic chromatin modification. Front Genet 5:65
82. Freitag M (2017) Histone methylation by SET domain proteins in fungi. Annu Rev Microbiol 71:413–439
83. Zhou HY, Liu YH, Liang YW, Zhou D, Li SF, Lin S, Dong H, Huang L (2020) The function of histone lysine methylation related SET domain group proteins in plants. Protein Sci 29:1120–1137
84. Herz HM, Garruss A, Shilatifard A (2013) SET for life: biochemical activities and biological functions of SET domain-containing proteins. Trends Biochem Sci 38: 621–639
85. Frapporti A, Pina CM, Arnaiz O, Holoch D, Kawaguchi T, Humbert A, Eleftheriou E, Lombard B, Loew D, Sperling L, Guitot K, Margueron R, Duharcourt S (2019) The Polycomb protein Ezl1 mediates H3K9 and H3K27 methylation to repress transposable elements in paramecium. Nat Commun 10: 2710
86. Chen DH, Qiu HL, Huang Y, Zhang L, Si JP (2020) Genome-wide identification and expression profiling of SET DOMAIN GROUP family in *Dendrobium catenatum*. BMC Plant Biol 20:40
87. Strahl BD, Grant PA, Briggs SD, Sun ZW, Bone JR, Caldwell JA, Mollah S, Cook RG, Shabanowitz J, Hunt DF, Allis CD (2002) Set2 is a nucleosomal histone H3-selective methyltransferase that mediates transcriptional repression. Mol Cell Biol 22:1298–1306
88. Jacob Y, Stroud H, Leblanc C, Feng S, Zhuo L, Caro E, Hassel C, Gutierrez C, Michaels SD, Jacobsen SE (2010) Regulation of heterochromatic DNA replication by histone H3 lysine 27 methyltransferases. Nature 466:987–991
89. Jacob Y, Feng S, LeBlanc CA, Bernatavichute YV, Stroud H, Cokus S, Johnson LM, Pellegrini M, Jacobsen SE, Michaels SD (2009) ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin structure and gene silencing. Nat Struct Mol Biol 16:763–768
90. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K, Jaenisch R, Wagschal A, Feil R, Schreiber SL, Lander ES (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell 125:315–326
91. Kuroda MI, Kang H, De S, Kassis JA (2020) Dynamic competition of Polycomb and Trithorax in transcriptional programming. Annu Rev Biochem 89:235–253
92. Miller T, Krogan NJ, Dover J, Erdjument-Bromage H, Tempst P, Johnston M, Greenblatt JF, Shilatifard A (2001) COMPASS: a complex of proteins associated with a trithorax-related SET domain protein. Proc Natl Acad Sci U S A 98:12902–12907
93. Thornton JL, Westfield GH, Takahashi YH, Cook M, Gao X, Woodfin AR, Lee JS, Morgan MA, Jackson J, Smith ER, Couture JF, Skiniotis G, Shilatifard A (2014) Context dependency of Set1/COMPASS-mediated histone H3 Lys4 trimethylation. Genes Dev 28:115–120
94. Eissenberg JC, Shilatifard A (2010) Histone H3 lysine 4 (H3K4) methylation in development and differentiation. Dev Biol 339:240–249
95. Sze CC, Cao KX, Collings CK, Marshall SA, Rendleman EJ, Ozark PA, Chen FX, Morgan MA, Wang L, Shilatifard A (2017) Histone H3K4 methylation-dependent and -independent functions of Set1A/COMPASS in embryonic stem cell self-renewal and differentiation. Genes Dev 31:1732–1737
96. Wu M, Wang PF, Lee JS, Martin-Brown S, Florens L, Washburn M, Shilatifard A (2008) Molecular regulation of H3K4 Trimethylation by Wdr82, a component of human Set1/COMPASS. Mol Cell Biol 28:7337–7344
97. Wang PF, Lin CQ, Smith ER, Guo H, Sanderson BW, Wu M, Gogol M, Alexander T, Seidel C, Wiedemann LM, Ge K, Krumlauf R, Shilatifard A (2009) Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II. Mol Cell Biol 29:6074–6085
98. Hu DQ, Garruss AS, Gao X, Morgan MA, Cook M, Smith ER, Shilatifard A (2013) The Mll2 branch of the COMPASS family regulates bivalent promoters in mouse embryonic stem cells. Nat Struc Mol Biol 20:1093–1097
99. Strahl BD, Ohba R, Cook RG, Allis CD (1999) Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. Proc Natl Acad Sci U S A 96:14967–14972
100. Milne TA, Briggs SD, Brock HW, Martin ME, Gibbs D, Allis CD, Hess JL (2002) MLL targets SET domain methyltransferase activity

to Hox gene promoters. Mol Cell 10:1107–1117

101. Krogan NJ, Dover J, Wood A, Schneider J, Heidt J, Boateng MA, Dean K, Ryan OW, Golshani A, Johnston M, Greenblatt JF, Shilatifard A (2003) The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. Mol Cell 11:721–729
102. Aranda S, Mas G, Di Croce L (2015) Regulation of gene transcription by Polycomb proteins. Science Adv 1:e1500737
103. Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang XM, Lin JC, Liang GN, Jones PA, Tanay A (2008) Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proc Natl Acad Sci U S A 105: 12979–12984
104. Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469:343–349
105. Shaver S, Casas-Mollano JA, Cerny RL, Cerutti H (2010) Origin of the polycomb repressive complex 2 and gene silencing by an E (z) homolog in the unicellular alga *Chlamydomonas*. Epigenetics 5:301–312
106. Merini W, Calonje M (2015) PRC1 is taking the lead in PcG repression. Plant J 83:110–120
107. Pasini D, Bracken AP, Jensen MR, Lazzerini Denchi E, Helin K (2004) Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. EMBO J 23: 4061–4071
108. Laugesen A, Hojfeldt JW, Helin K (2019) Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. Mol Cell 74:8–18
109. Kahn TG, Dorafshan E, Schultheis D, Zare A, Stenberg P, Reim I, Pirrotta V, Schwartz YB (2016) Interdependence of PRC1 and PRC2 for recruitment to Polycomb response elements. Nucleic Acids Res 44:10132–10149
110. Horard B, Tatout C, Poux S, Pirrotta V (2000) Structure of a polycomb response element and in vitro binding of polycomb group complexes containing GAGA factor. Mol Cell Biol 20:3187–3197
111. Xiao J, Jin R, Yu X, Shen M, Wagner JD, Pai A, Song C, Zhuang M, Klasfeld S, He C, Santos AM, Helliwell C, Pruneda-Paz JL, Kay SA, Lin X, Cui S, Garcia MF, Clarenz O, Goodrich J, Zhang X, Austin RS, Bonasio R, Wagner D (2017) Cis and trans determinants of epigenetic silencing by Polycomb repressive complex 2 in *Arabidopsis*. Nat Genet 49:1546–1552
112. Rinn JL, Kertesz M, Wang JK, Squazzo SL, Xu X, Brugmann SA, Goodnough LH, Helms JA, Farnham PJ, Segal E, Chang HY (2007) Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell 129:1311–1323
113. Lewis ZA (2017) Polycomb Group Systems in Fungi: new models for understanding Polycomb repressive complex 2. Trends Genet 33: 220–231
114. Wiles ET, Selker EU (2017) H3K27 methylation: a promiscuous repressive chromatin mark. Curr Opin Genet Dev 43:31–37
115. Ridenour JB, Möller M, Freitag M (2020) Polycomb repression without bristles: facultative heterochromatin and genome stability in fungi. Genes 11:638
116. Smolle M, Workman JL, Venkatesh S (2013) reSETting chromatin during transcription elongation. Epigenetics 8:10–15
117. Veerappan CS, Avramova Z, Moriyama EN (2008) Evolution of SET domain protein families in the unicellular and multicellular Ascomycota fungi. BMC Evol Biol 8:190
118. Schmitges FW, Prusty AB, Faty M, Stützer A, Lingaraju GM, Aiwazian J, Sach R, Hess D, Li L, Bunker RD, Wirth U, Bouwmeester T, Bauer A, Ly-Hartig N, Zhao K, Chan H, Gu J, Gut H, Fischle W, Müller J, Thomä N (2011) Histone methylation by PRC2 is inhibited by active chromatin marks. Mol Cell 42:330–341
119. Yuan W, Xu M, Huang C, Liu N, Chen S, Zhu B (2011) H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. J Biol Chem 286:7983–7989
120. Lu C, Jain SU, Hoelper D, Bechet D, Molden RC, Ran LL, Murphy D, Venneti S, Hameed M, Pawel BR, Wunder JS, Dickson BC, Lundgren SM, Jani KS, De Jay N, Papillon-Cavanagh S, Andrulis IL, Sawyer SL, Grynspan D, Turcotte RE, Nadaf J, Fahiminiyah S, Muir TW, Majewski J, Thompson CB, Chi P, Garcia BA, Allis CD, Jabado N, Lewis PW (2016) Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science 352:844–849
121. Gaydos LJ, Rechtsteiner A, Egelhofer TA, Carroll CR, Strome S (2012) Antagonism between MES-4 and Polycomb repressive complex 2 promotes appropriate gene expression in C. elegans germ cells. Cell Rep 2: 1169–1177

122. Janevska S, Baumann L, Sieber CMK, Münsterkotter M, Ulrich J, Kämper J, Güldener U, Tudzynski B (2018) Elucidation of the two H3K36me3 histone methyltransferases Set2 and Ash1 in *fusarium fujikuroi* unravels their different chromosomal targets and a major impact of Ash1 on genome stability. Genetics 208:153–171

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

124. Bicocca VT, Ormsby T, Adhvaryu KK, Honda S, Selker EU (2018) ASH1-catalyzed H3K36 methylation drives gene repression and marks H3K27me2/3-competent chromatin. elife 7:e41497

125. Möller M, Schotanus K, Soyer JL, Haueisen J, Happ K, Stralucke M, Happel P, Smith KM, Connolly LR, Freitag M, Stukenbrock EH (2019) Destabilization of chromosome structure by histone H3 lysine 27 methylation. PLoS Genet 15:e1008093

126. Finogenova K, Bonnet J, Poepsel S, Schafer IB, Finkl K, Schmid K, Litz C, Strauss M, Benda C, Muller J (2020) Structural basis for PRC2 decoding of active histone methylation marks H3K36me2/3. elife 9:e61964

127. Musselman CA, Avvakumov N, Watanabe R, Abraham CG, Lalonde ME, Hong Z, Allen C, Roy S, Nunez JK, Nickoloff J, Kulesza CA, Yasui A, Cote J, Kutateladze TG (2012) Molecular basis for H3K36me3 recognition by the Tudor domain of PHF1. Nat Struct Mol Biol 19:1266–1272

128. Qin S, Guo Y, Xu C, Bian C, Fu M, Gong S, Min J (2013) Tudor domains of the PRC2 components PHF1 and PHF19 selectively bind to histone H3K36me3. Biochem Biophys Res Commun 430:547–553

129. Cai L, Rothbart SB, Lu R, Xu B, Chen WY, Tripathy A, Rockowitz S, Zheng D, Patel DJ, Allis CD, Strahl BD, Song J, Wang GG (2013) An H3K36 methylation-engaging Tudor motif of polycomb-like proteins mediates PRC2 complex targeting. Mol Cell 49: 571–582

130. Choi J, Bachmann AL, Tauscher K, Benda C, Fierz B, Muller J (2017) DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. Nat Struct Mol Biol 24:1039–1047

131. Dorafshan E, Kahn TG, Glotov A, Savitsky M, Walther M, Reuter G, Schwartz YB (2019) Ash1 counteracts Polycomb repression independent of histone H3 lysine 36 methylation. EMBO Rep 20:e46762

132. Dorafshan E, Kahn TG, Glotov A, Savitsky M, Schwartz YB (2019) Genetic dissection reveals the role of Ash1 domains in counteracting polycomb repression. G3 (Bethesda) 9:3801–3812

133. Basenko EY, Sasaki T, Ji L, Prybol CJ, Burckhardt RM, Schmitz RJ, Lewis ZA (2015) Genome-wide redistribution of H3K27me3 is linked to genotoxic stress and defective growth. Proc Natl Acad Sci U S A 112: E6339–E6348

134. Jamieson K, Wiles ET, McNaught KJ, Sidoli S, Leggett N, Shao Y, Garcia BA, Selker EU (2016) Loss of HP1 causes depletion of H3K27me3 from facultative heterochromatin and gain of H3K27me2 at constitutive heterochromatin. Genome Res 26:97–107

135. Connolly LR, Smith KM, Freitag M (2013) The *fusarium graminearum* histone H3 K27 methyltransferase KMT6 regulates development and expression of secondary metabolite gene clusters. PLoS Genet 9:e1003916

136. Dumesic PA, Homer CM, Moresco JJ, Pack LR, Shanle EK, Coyle SM, Strahl BD, Fujimori DG, Yates JR 3rd, Madhani HD (2015) Product binding enforces the genomic specificity of a yeast polycomb repressive complex. Cell 160:204–218

137. Montgomery SA, Tanizawa Y, Galik B, Wang N, Ito T, Mochizuki T, Akimcheva S, Bowman JL, Cognat V, Marechal-Drouard L, Ekker H, Hong SF, Kohchi T, Lin SS, Liu LD, Nakamura Y, Valeeva LR, Shakirov EV, Shippen DE, Wei WL, Yagura M, Yamaoka S, Yamato KT, Liu C, Berger F (2020) Chromatin Organization in Early Land Plants Reveals an ancestral association between H3K27me3, transposons, and constitutive heterochromatin. Curr Biol 30:573–588 e577

138. Ho JWK, June YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A, Parker SCJ, Gu TT, Kundaje A, Riddle NC, Bishop E, Egelhofer TA, Hu SS, Alekseyenko AA, Rechtsteiner A, Asker D, Belsky JA, Bowmanm SK, Chens QB, Chen RAJ, Day DS, Dong Y, Dose AC, Duan XK, Epstein CB, Ercan S, Feingold EA, Ferrari F, Garrigues JM, Gehlenborg N, Good PJ, Haseley P, He D, Herrmann M, Hoffman MM, Jeffers TE, Kharchenko PV, Kolasinska-Zwierz P, Kotwaliwale CV, Kumar N, Langley SA, Larschan EN, Latorre I, Libbrecht MW, Lin XQ, Park R, Pazin MJ, Pham HN, Plachetka A, Qin B, Schwartz YB, Shoresh N, Stempor P, Vielle A, Wang CY, Whittle CM, Xue HL, Kingstonm RE, Kim JH, Bernstein BE, Dernburg AF, Pirrotta V, Kuroda MI, Noble WS,

Tullius TD, Kellis M, MacAlpine DM, Strome S, Elgin SCR, Liu XS, Lieb JD, Ahringer J, Karpen GH, Park PJ (2014) Comparative analysis of metazoan chromatin organization. Nature 512:449–452

139. Hawkins RD, Hon GC, Lee LK, Ngo Q, Lister R, Pelizzola M, Edsall LE, Kuan S, Luu Y, Klugman S, Antosiewicz-Bourget J, Ye Z, Espinoza C, Agarwahl S, Shen L, Ruotti V, Wang W, Stewart R, Thomson JA, Ecker JR, Ren B (2010) Distinct Epigenomic landscapes of pluripotent and lineage-committed human cells. Cell Stem Cell 6: 479–491
140. Mozzetta C, Pontis J, Fritsch L, Robin P, Portoso M, Proux C, Margueron R, Ait-Si-Ali S (2014) The histone H3 lysine 9 methyltransferases G9a and GLP regulate polycomb repressive complex 2-mediated gene silencing. Mol Cell 53:277–289
141. Peters AH, Kubicek S, Mechtler K, O'Sullivan RJ, Derijck AA, Perez-Burgos L, Kohlmaier A, Opravil S, Tachibana M, Shinkai Y, Martens JH, Jenuwein T (2003) Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. Mol Cell 12:1577–1589
142. Taverna SD, Li H, Ruthenburg AJ, Allis CD, Patel DJ (2007) How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat Struct Mol Biol 14:1025–1040
143. Veluchamy A, Rastogi A, Lin X, Lombard B, Murik O, Thomas Y, Dingli F, Rivarola M, Ott S, Liu XY, Sun YZ, Rabinowicz PD, McCarthy J, Allen AE, Loew D, Bowler C, Tirichine L (2015) An integrative analysis of post-translational histone modifications in the marine diatom *Phaeodactylum tricornutum*. Genome Biol 16:102
144. Qian C, Zhou MM (2006) SET domain protein lysine methyltransferases: structure, specificity and catalysis. Cell Mol Life Sci 63: 2755–2763
145. Xiao B, Jing C, Kelly G, Walker PA, Muskett FW, Frenkiel TA, Martin SR, Sarma K, Reinberg D, Gamblin SJ, Wilson JR (2005) Specificity and mechanism of the histone methyltransferase Pr-Set7. Genes Dev 19: 1444–1454
146. Couture JF, Dirk LMA, Brunzelle JS, Houtz RL, Trievel RC (2008) Structural origins for the product specificity of SET domain protein methyltransferases. Proc Natl Acad Sci U S A 105:20659–20664
147. Starr TN, Thornton JW (2017) Exploring protein sequence-function landscapes. Nat Biotechnol 35:125–126
148. Anderson DW, McKeown AN, Thornton JW (2015) Intermolecular epistasis shaped the function and evolution of an ancient transcription factor and its DNA binding sites. elife 4:e07864