# DNA and Histone Methylation in Epigenetics

**Hengbin Wang, Jinrong Min, and Trygve Tollefsbol**

**Abstract** Epigenetics, the mechanism that defines gene expression patterns without changing the DNA sequence, has far-reaching consequences on normal and pathological development including but not limited to cell fate determination, maintenance of tissue differentiation, and cancer occurrence. Methylation of DNA and histones, the two components of chromatin, constitutes important epigenetic mechanisms that govern chromatin-based nuclear processes. In this chapter, we briefly summarize the key enzymes involved, mechanisms, and function of these two modifications. We envision that DNA and histone methylation will increasingly become important targets for cancer treatment.

**Keywords** Epigenetics • DNA methylation • Histone methylation

## 1 Introduction

Epigenetics refers to mechanisms that control gene expression patterns across cell generations but do not involve changes in the DNA sequence [1]. Since virtually all somatic cells contain the same DNA sequence, epigenetic mechanisms must lie outside of the DNA sequence. Specific to eukaryotes, the genomic DNA is associated with a set of highly conserved histone proteins that form the structure of chromatin. Compared to the linearly organized genetic information in the DNA sequence, which is identical in virtually all somatic cells, chromatin adapts distinct local

H. Wang (✉)
Department of Biochemistry and Molecular Genetics, University of Alabama at Birmingham, Kaul Human Genetics Building 402A, 720 20th Street South, Birmingham, AL 35294, USA
e-mail: hbwang@uab.edu

J. Min
Structural Genomics Consortium and Department of Physiology, University of Toronto, 101 College Street, MaRS South Tower, Toronto, ON M5G 1L7, Canada

T. Tollefsbol
Department of Biology, University of Alabama at Birmingham, Campbell Hall 175, 1300 University Boulevard, Birmingham, AL 35294, USA



A. Kaneda, Y.-i. Tsukada (eds.), *DNA and Histone Methylation as Cancer Targets*, Cancer Drug Discovery and Development, DOI 10.1007/978-3-319-59786-7_1

conformation, permissive or repressive to protein factor binding, allowing gene expression at programs specific to each cell type. Therefore, the organized information in chromatin constitutes an additional layer of mechanisms that control the use of genetic information (Fig. 1). This organized information is unique to each cell type and also specific to any given cellular functional status. As the physical carrier of epigenetic regulation, chromatin integrates signals from endogenous development and the exogenous environment to define the functional output of the underlying DNA sequences. It is due to the epigenetic mechanism that eukaryotes can have many different cell types, which corresponds to distinct gene expression programs (the so-called epigenomes), from a single genome. This fundamental function endows essential roles of epigenetics in normal development as well as disease progression.

Studies in the past decades have revealed multiple mechanisms that modulate chromatin structure and affect the functional output of chromatin. These mechanisms include, but are not limited to, DNA methylation, ATP-dependent chromatin remodeling, posttranslational histone modifications, and non-coding RNAs [1]. Of these mechanisms, both DNA and histone methylation involve the addition of a methyl group, a single carbon metabolite, to the essential components of chromatin. Both DNA and histone methylation may profoundly affect the chromatin conformation and functional output of the underlying DNA sequence. While the mechanism that mediates the transmission of DNA methylation from mother cells to daughter cells has been revealed, the mechanism that mediates the transmission of histone methylation remains largely obscure. Although the formation of C-C bond in DNA methylation and N-C bound in histone methylation render these modifications relatively stable, recent studies reveal the reversibility of both modifications. The stability as well as the plasticity of both modifications entitles important regulatory functions of both modifications in chromatin conformation during cell proliferation and differentiation. Both DNA and histone methylation are tightly regulated to ensure proper cellular function during normal development. Deregulation of either modification system, including mutations in the enzymes catalyzing DNA and histone methylations, dysregulation of proteins binding to and mediating the function of DNA and histone methylation, and even histone protein itself are frequently

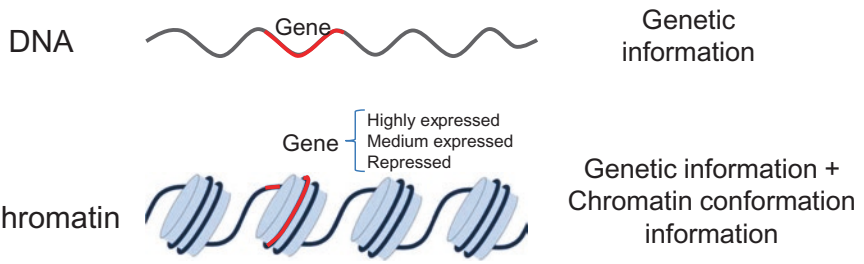


**Fig. 1** Genetics vs epigenetics. Genetic information is stored in the DNA sequence while chromatin contains organized structural information in addition to the genetic information. This epigenetic mechanism allows genes to be expressed at different levels according to requirement

observed in many human diseases including cancers. Therefore, both DNA and histone methylation are appealing targets for cancer treatment.

## 2 DNA Methylation

### *2.1 DNA Methylation in Mammals*

Of the four nucleotides that make up the DNA, adenosine, thymine, cytosine, and guanine, both adenine and cytosine can be methylated; however, methylated cytosine is mainly detected in mammals [15]. The methyl group is added to the 5 position of the pyridmidine ring to form the 5-methylcytosine (5mC). The methyl group is positioned in the major groove of the DNA and this methylation does not interfere with the Watson-Crick pairing of cytosine and guanine. Therefore, cytosine methylation by itself does not affect the stability of DNA double helix. In mammals, DNA methylation occurs predominantly in a CpG dinucleotide context, where cytosine is positioned 5′ to a guanine nucleotide. 60–90% of CpGs are methylated and methylation is generally evenly distributed in mammals. In contrast, CpG methylation is often grouped into clusters in invertebrates.

In mammals, unmethylated CpGs are often clustered together and form the so-called CpG islands [5]. CpG islands are often located upstream of transcription initiation sites of protein coding genes and modulate the expression of downstream genes. These DNA sequences destabilize nucleosomes and recruit proteins that help establish a transcriptionally permissive chromatin status. When these CpGs are methylated, methylated DNA binding proteins (MBDs) are often recruited. These MBDs may complex with histone deacetylases and chromatin remodeling factors and create a chromatin status that is repressive to gene transcription. In many human diseases including cancer, CpGs are aberrantly methylated at many gene promoters. CpG methylation may result in the repression of many important genes that regulate cell cycle, apoptosis, DNA repair, etc. The silencing of these genes causes the damaged DNA unrepaired in daughter cells and may preposition these cells for cancer progression. Alterations in DNA methylation patterns have been recognized as an important feature of cancer development. On the other hand, a decrease of DNA methylation (so called hypomethylation) is linked to chromosome instability, due to essential roles of DNA methylation in repressing repetitive DNA sequences at pericentromeric regions. Genome instability also constitutes an important component of cancer development. Therefore, both DNA hyper- and hypo-methylation may contribute to the development of cancers.

DNA methylation undergoes dynamic changes during development [6]. Genomic DNA methylation is greatly reduced as the zygote develops into the gastrulation stages of embryos. During this process, DNA methylation in the female genome is passively removed by dilution as cells undergo division, whereas it is erased in the male genome by a faster, active DNA demethylation process. As a result, embryos at gastrulation stages have low levels of DNA meth-

ylation. DNA methylation is re-established via *de novo* DNA methylation during successive cell divisions as the gastrulation stage of embryos initiates differentiation and further develops into the next stages. DNA methylation is usually heritable through mitotic cell division and some methylation is also heritable through the specialized meiotic cell division that creates egg and sperm cells, resulting in genomic imprinting.

DNA methylation can be offset by the activity of ten-eleven translocation (TET) methylcytosine dioxygenase family of proteins [11]. Instead of direct removal of methyl groups, TET family of proteins hydroxylates the methyl group to form the 5-hydromethylcytisine (5hmC). The 5hmC can be further oxidized and finally removed from DNA sequence by the demethylation DNA repair system. This allows for the removal of the DNA methylation. Both DNA methylation and demethylation are tightly regulated. Misregulation of either process can result in dysregulation of gene expression that leads to increased susceptibility to human diseases such as cancer.

## *2.2 DNA Methyltransferase*

The enzyme that catalyzes DNA methylation is DNA methyltransferase (DNMT) [17]. DNMT uses S-adenosyl methionine (SAM) as the methyl group donor to methylate cytosine at the 5 position of the pyrimidine ring. In mammals, there are four members of DNMTs that share the same catalytic domains: DNMT1, DNMT2, DNMT3a, and DNMT3b (Fig. 2). One additional protein, DNMT3L, although containing the ATRX domain that is also present in DNMT3A and DNMT3B, does not contain an active catalytic domain. Based on the function of these DNMTs in DNA methylation, DNMTs can be divided into two subgroups: maintenance DNMT and *de novo* DNMT.

DNMT1 is the maintenance DNA methyltransferase (Fig. 2A). Through multiple regions, DNMT1 interacts with the replication fork, and uses the methylated cytosine from the parental strand as hemimethylated templates to methylate cytosine in the newly synthesized daughter strand during DNA replication. Therefore, this enzyme is required for preserving the DNA methylation pattern during successive cell division. Without the function of this enzyme, the replication machinery would produce daughter strands that are unmethylated, leading to passive DNA demethylation as cells undergo division. Deletion of DNMT1 results in genome-wide loss of DNA methylation. However, DNMT1 can only methylate DNA when one strand is methylated and cannot initiate DNA methylation when there is no methylated cytosine in either strand. DNMT3A and DNMT3B are the enzymes that initiate DNA methylation without the dependence on preexisting methylated cytosine (Fig. 2B). These two proteins are responsible for setting up the DNA methylation patterns during early embryonic development. After fertilization, the male genome is actively demethylated and the female genome is passively demethyated as the embryos develop to the blastocyst stages. At this stage, the majority of genome DNA methylation is removed except for the imprinting

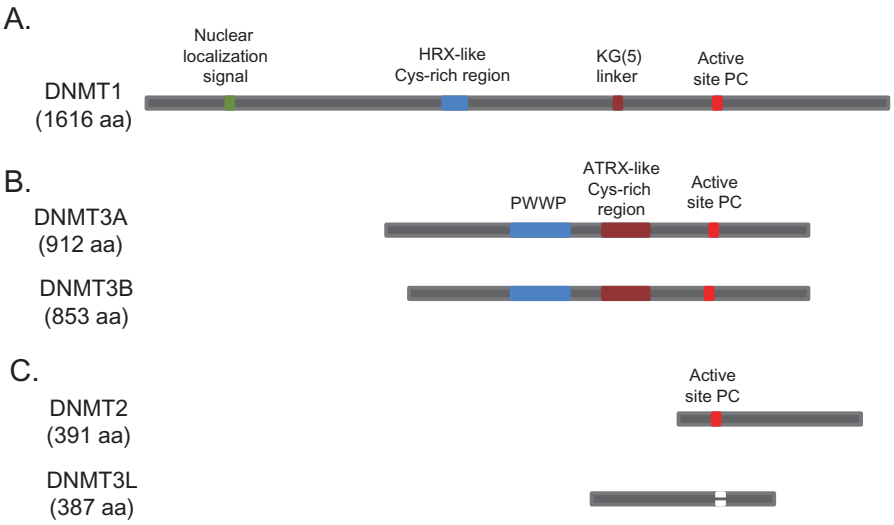


**Fig. 2** Mammalian DNA methyltransferases. (**a**) Maintenance DNA methyltransferase DNMT1. (**b**) De novo DNA methyltransferases DNMT3A and DNMT3B. (**c**) Proteins sharing sequence homologues with DNA methyltransferase. Functional domains are denoted

regions. As embryos further develop and progress into the next stage, the DNA methylation pattern is re-established and DNMT3A and DNMT3B are responsible for the establishment of the DNA methylation pattern during this development stage. These two proteins may be partially redundant, as only deletion of both DNMT3A and DNMT3B in mouse leads to a failure to initiate *de novo* methylation after implantation.

DNMT2 also contains the catalytic domain for DNA methylation (Fig. 2C); however, the substrate for DNMT2 is not DNA but is the cytosine 38 in the anticodon loop of aspartic acid transfer RNA (tRNA-Asp) [8]. Methylation of tRNA-Asp increases its affinity for aspartyl-tRNA synthetase and therefore controls the synthesis of a group of proteins containing poly-Asp sequences. Although DNM3L itself does not have DNA methylation activity (Fig. 2C), this protein can interact with the *de novo* DNMT 3A and 3B and contribute to the establishment of DNA methylation patterns. Interaction with DNMT3L increases the binding affinity of these two DNMTs to DNA and stimulates the methyltransferase activity.

## *2.3 Methylated DNA Binding Protein*

DNA methylation may affect the conformation and function of chromatin through several ways [9]. First, the methyl group on cytosine itself may physically impede the binding of transcription factors or other proteins to the underlying DNA template. Second, a group of proteins, named methylated DNA binding proteins (MBDs), can

specifically recognize the methylated DNA. The binding of these MBDs can block the binding of transcription factors or other protein factors to the underlying DNA template. Third and more generally, these MBDs recruit histone deacetylases (HDACs) and other chromatin remodeling proteins to methylated DNA regions. These enzymatic activities modify local chromatin conformation through histone deactylation and chromatin remodeling, leading to the formation of a repressive chromatin environment. This repressive chromatin environment inhibits transcription and other processes that require access to the underlying DNA template.

There are five members of proteins, MeCP2 and MBD1-4 that can specifically recognize methylated CpGs in mammals (Fig. 3) [9]. These proteins bind to the methylated CpGs through a methyl-CpG-binding domain (MBD). The MBD domain contains 75 amino acids and adapt to a structural fold of four beta strands, three loops and one alpha helix, forming the interfacial surface with methylated DNA. The interaction occurs in the major groove of the double DNA helix. The interaction is through hydrophilic and hydrophobic interactions. Of these MBDs, MeCP2 appears to be specifically important for the normal function of nerve cells. Mutations of the *MeCP2* gene are the cause of most cases of Rett syndrome, a progressive neurologic developmental disorder and one of the most common causes of mental retardation in females. In addition to the MBD domain, the MBD1 protein contains two to three cysteine-rich (CXXC) type zinc finger domains, and an additional transcriptional repression domain (TRD). MBD1 not only represses transcription from methylated promoters but also repress unmethylated promoter activity through the variation of the CXXC domain. MBD2 has been reported to interact with HDAC1/2, MBD3, and Sin3, and mediates the repressive function of DNA methylation, by histone deacetylation. MBD3 is a subunit of the NuRD, a multisubunit complex containing nucleosome remodeling and histone deacetylase activities. MBD3 has been reported to interact with MTA2, HDAC1/2, and MBD2, and mediates the repressive function of DNA methylation. In addition to the MBD domain, MBD4 also contains a DNA glycosylase domain at its C-terminus. MBD4 can also bind to the deamination derivatives of CpG G:U and G:T base pairs, and function in the initial step of base excision repair. MBD4 can specifically remove T and U paired with guanine (G) within the CpG sites, and contributes to the stability

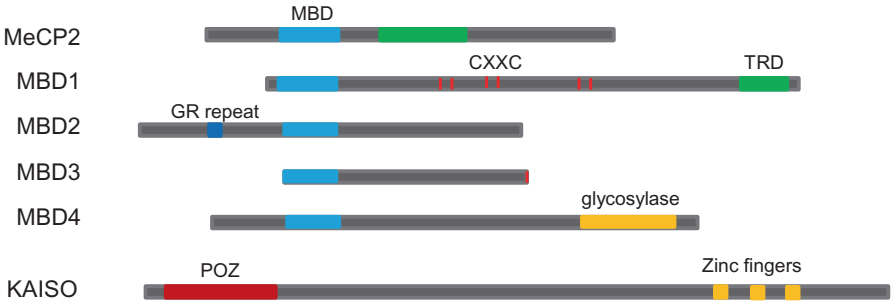


**Fig. 3** Methyl DNA binding proteins. Functional domains are denoted

of CpG at promoters. In addition to these MBD proteins, mammals also contain a KAISO protein. This protein exhibits bimodal DNA-binding specificity. KAISO binds to methylated DNA and also to the non-methylated DNA within the TCCTGCNA sequence. KAISO can recruit the N-CoR repressor complex, which contains the histone deacetylation activity and helps form the repressive chromatin structures in target gene promoters.

## *2.4 Function of DNA Methylation*

Although methylated cytosine can be further hydroxylated to form 5hmC, DNA methylation is often stable and therefore the effects of DNA methylation on gene expression are normally permanent and unidirectional. The stable nature of DNA methylation helps the maintenance of the gene expression program specific to each cell type and prevents cells from reverting to stem cells or trans-converting into different cell types. The function of DNA methylation in mammals includes transcription silencing, X chromosome inactivation, genome stability, as well as in many other biological processes [6, 13].

DNA methylation has been generally associated with transcription repression. Through the recruitment of HDACs and chromatin remodeling complexes, methylation of DNA often results in a compact chromatin conformation and transcription silencing. One special case of DNA methylation-mediated gene silencing is gene imprinting. Gene imprinting refers to a phenomenon where the expression of a given gene depends on which allele the gene is located on, i.e., only one allele from either the paternal (sperm derived) or maternal (egg-derived) genome is expressed. DNA methylation on imprinted genes escapes the global DNA demethylation after fertilization. One of the best characterized examples of gene imprinting is the insulin-like growth factor *Igf2/H19* genes. These two genes are located 70 kb away in human chromosome 11. The *Ifg2* gene is preferentially expressed from the paternal allele while the *H19* gene is preferentially expressed from the maternal allele. This allele-specific expression is controlled by a region located upstream of the *H19* gene. This region is methylated in the paternal allele, which recruits MBDs to form a closed chromatin conformation, and represses the expression of *H19* gene. In this case, the tissue-specific enhancer interacts with the promoter of the *Igf2* gene and allows the expression of this gene from the paternal allele. In contrast, this region is not methylated in maternal allele, which recruits CTCF to form an open chromatin conformation, allowing the expression of *H19* gene only from maternal allele. The engagement of tissue-specific enhancer with the *H19* gene prevents the expression of *Igf2* gene from the maternal allele. Gene imprinting prevents the reproduction of organisms from unfertilized eggs and contributes to the stability and variability of species. The gene imprinting system is often disrupted in congenital malformation syndromes, tumors, or cloned animals.

Another important function of DNA methylation is X inactivation. In female mammals, one of the two X chromosomes is silenced to achieve equal gene doses

between male and female individuals. The silencing is achieved by packaging the entire X chromosome into a highly compact inactive structure called heterochromatin. This chromosome-wide phenomenon has long been considered a paradigm for the study of the effect of heterochromatin formation and DNA methylation on gene expression in mammals. The inactivated X chromosome contains high levels of DNA methylation, which may contribute to the packaging of the X chromosome into the highly compacted inactive status. The inactivation process consists of three components—initiation, spreading and maintenance. DNA methylation may play a role in all three of these processes; however, the exact mechanism remains unclear.

DNA methylation is also required for the establishment of heterochromatic regions such as the centromere and telomere regions. These heterochromatic regions play important roles in genome stability. The centromere is required for proper segregation of mitotic chromosomes to daughter cells. Malfunction of the centromeric region is often accompanied by inappropriate partition of genetic materials into two daughter cells. DNA methylation may also be responsible for maintaining telomere integrity through indirect regulation. DNMT knockout mice exhibit increased telomeric recombination and variations in telomere length. One particular example of DNA methylation in affecting genome stability is the ICF (immunodeficiency, centromere instability, facial anomalies) syndrome. The majority of ICF patients carry mutations in the *de novo* DNA methyltransferase DNMT3B. Of these patients, hypomethylation is observed in centromeric heterochromatic regions, which may account for the genomic instability in ICF syndrome patients. Genome instability arising from aberrant DNA methylation particularly contributes to cancer initiation and development. DNA methylation has appeared as an appealing target for cancer treatment. Currently, small inhibitors for DNA methylation have undergone clinical trials for certain cancer treatment and have been approved by the FDA.

## 3 Histone Methylation

### *3.1 Methylation Sites*

As the building block of chromatin, nucleosomes are subject to a variety of modifications that modulate the dynamics and metabolism of chromatin. Once thought merely as a static structural component for DNA packaging, histones are now recognized as important regulators for nucleosome function. Histones, particularly the N-terminal tails, are modified by a variety of posttranslational modifications. These modifications include, but are not limited to, methylation, acetylation, phosphorylation, and ubiquitination [16]. These modifications regulate the interaction between DNA and histones within the nucleosome, the interaction between adjacent nucleosomes, and even the interaction between nucleosomes from different chromosomes. Among these modifications, histone methylation is a prevalent modification and can modulate gene expression in either a positive or negative manner.

Of the four histones, histones H3 and H4 are predominately methylated (Fig. 4). Both histone lysine (K) and arginine (R) residues, which contain the free amino groups on their side chains, can be modified by methylation. The prominent methylation sites on histone H3 include R2, K4, K9, R17, R26, K27, K36, and K79 (Fig. 4). The prominent sites on histone H4 include R3 and K20 (Fig. 4). Furthermore, lysine residues can be methylated to mono-, di-, and tri-methylation status. Meanwhile, the arginine residue can be mono- and di- methylated, and based on the location of the methyl groups, arginine methylation can occur in symmetrical or asymmetrical manners. While methylation of arginine is generally linked to gene activation, lysine methylation can lead to both gene activation and repression, depending on the specific amino acids being modified and the methylation status. It should be pointed out that the advance of the mass spectrometry technology and the development of the specific software allow more methylation sites being identified.

The formation of N-C bond in histone methylation renders the stable nature of this modification. Once thought to be a unidirectional modification, histone methylation has now been recognized as a reversible modification [10]. Two families of proteins targeting lysine methylation have been identified, and both employ an oxidation mechanism. The lysine-specific demethylase 1 (LSD1) is a flavin-dependent monoamino oxidase, which coverts mono- and di- methylated H3K4 or H3K9 to unmethylated lysine residues. The JmjC domain-containing protein family catalyzes lysine demethylation with Fe(II) and α-ketoglutarate as cofactors. In contrast to LSD1, JmjC domain-containing demethylase can remove all three lysine methylation states and constitutes the larger family of histone demethylases. As compared to lysine demethylation, arginine methylation is reversed in a different way. While the enzymes for demethylation of arginine residues remains to be identified, a dimi-

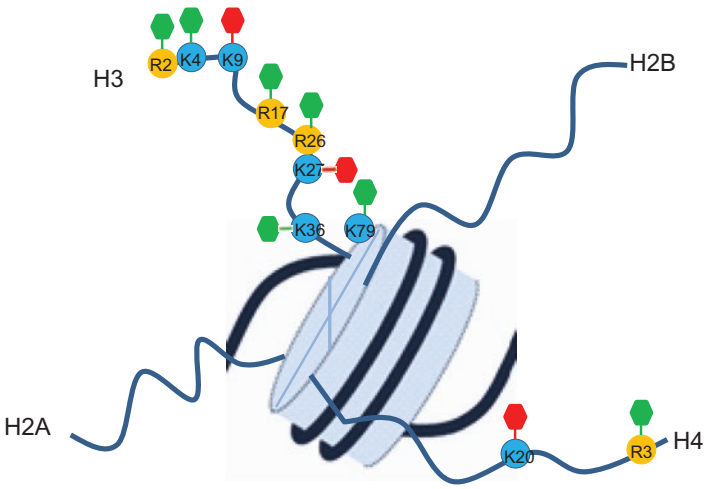


**Fig. 4** Histone methylation sites. Repressive histone methylations are marked with *red* while active histone methylation sites are marked with *green*. Both lysine and arginine methylation are indicated

nase PADI4 can convert the methylated arginine to citrulline residue and thus antagonize the effects of arginine methylation. However, PADI4 can also convert non-methylated arginine residues to citrulline residue, rendering the specificity of this enzyme in antagonizing the function of arginine methylation in question.

## *3.2 Histone Methyltransferases*

The enzymes that catalyze histone methylation are histone methyltransferases (HMTs). Based on specific residues targeted by these enzymes, HMTs can be further divided into different subgroups [14]. HMTs that specifically target the arginine residues are protein arginine methyltransferases (PRMTs) [7]. All PRMTs contain the conserved catalytic core, with signature motif I, post-I, II, III, and the THW loop (Fig. 5). The catalytic core adapts a doughnut-like head-to-tail homodimer. Each monomer contains an active site that binds the methyl group donor SAM and the targeted peptide or protein substrate. A pair of highly conserved glutamate residues in the active core uses its negative charge to coordinate the positively charged guanidino group of the arginine residue into the correct orientation for methylation.

HMTs that specifically catalyze the methylation of lysine resides can be further divided into two subgroups [14]. One subgroup contains a conserved SET (stands for **S**u(var)3-9, **E**nhancer of Zeste, **T**rithorax) domain, a pre-SET domain, and a post-SET domain (Fig. 6). The structures involved in the methyltransferase activity are the SET domain, which is composed of approximately 130 amino acids, the pre-SET, and the post-SET domains. The pre-SET domain often contains cysteine residues and forms zinc clusters to bind zinc ions. The SET domain is enriched in β-strands, which together with the β-sheets in pre-SET domain, forms the catalytic core. For methylation, the catalytic core binds to both the SAM and the substrate

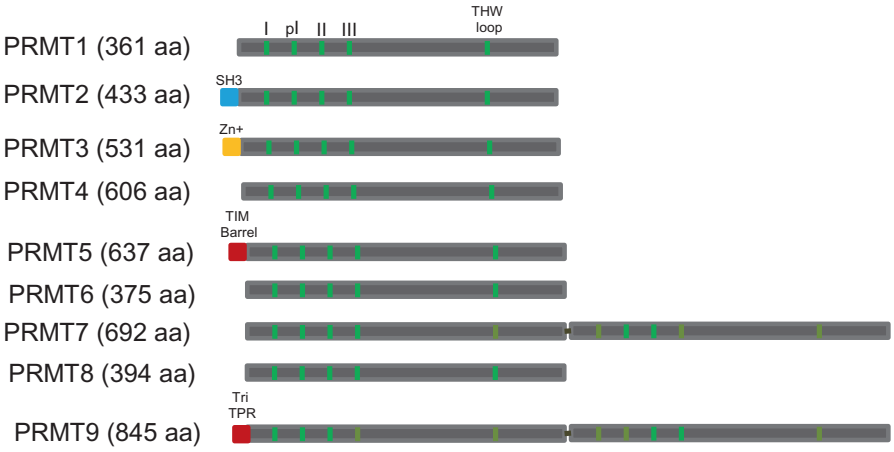


**Fig. 5** Protein arginine methyltransferases. Functional domains are denoted

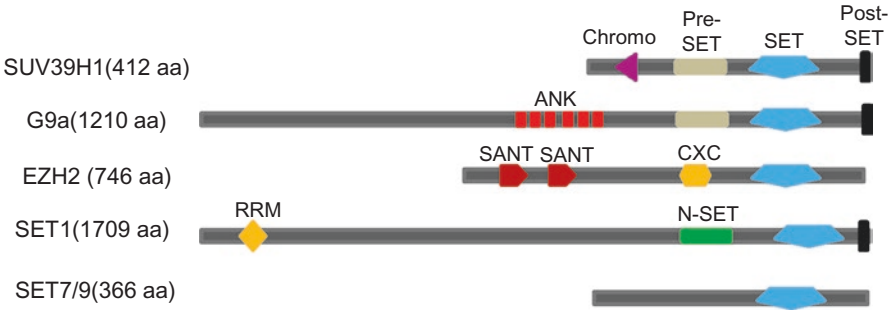


**Fig. 6** Selected Histone lysine methyltransferases. Functional domains are denoted

histone tail and orients them in a proper position. A tyrosine residue in the catalytic core then deprotonates the ε-amino group of the lysine residue allowing the lysine chain to make a nucleophilic attack on the methyl group on the sulfur atom of the SAM molecule and thereby transferring the methyl group to the lysine side chain. This family of HMTs constitutes the largest family of histone methyltransferases. The variations of the SET domain structure allow these SET domain methyltransferases to target many different residues and perform different degrees of methylation. Interestingly, some of the HMTs targeting lysine methylation do not contain the SET domain. The only enzyme in this family of HMTs is Dot1 or Dot1L in humans. Dot1 or Dot1L methylates histone H3 at K79 that is at its globular region. The active site of Dot1 is at its N-terminal. The methyl group donor SAM binds to a loop linking the N-terminal catalytic domain and the C-terminal domain. The C-terminal is important for the substrate specificity and Dot1 binding to DNA. The structural constraints define that Dot1 can only methylate histone H3.

## *3.3 Methylated Histone Binding Protein*

Although methylation does not change the overall charge of the residue, each replacement of the proton from the ε-amino decreases the possibility of hydrogen bound formation and increases hydrophobicity. Therefore, methylation of histones by itself could affect the interactions of protein factors with the chromatin template. More generally, methylation of histones provides a binding platform for other downstream proteins [12]. The diverse function of histone methylation in chromatin regulation is largely attributed to the various binding proteins for lysine residues modified at different sites and different states (Fig. 7).

Domains that specifically recognize methylated lysine residues include the plant homeodomain (PHD), the Royal family domains, CW domain and some WD40 domains, such as those of WDR5 and EED. The Royal family includes the Tudor, plant agent, chromo, PWWP, and malignant brain tumor (MBT) domains. The PHD finger constitutes a large family of proteins that specifically bind to methylated

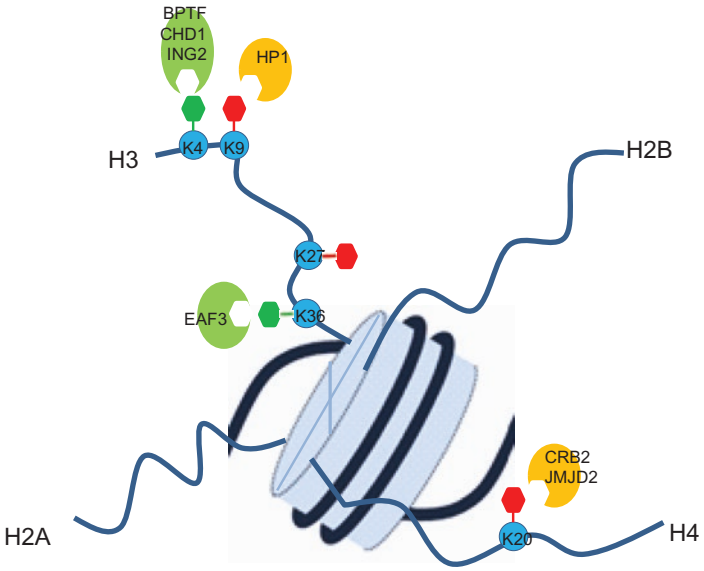


**Fig. 7** Proteins bound to methylated lysine residues in histones

H3K4. The PHD fingers of ING family proteins bind to H3K4me3 through aromatic side chains of Y215 and W238 residues. The binding of H3K4me3 stabilizes the association of mSin3a-HDAC1 complex, where ING2 is a core subunit, to chromatin. The association results in histone deacetylation and formation of repressive chromatin conformation in response to DNA damage. On the other hand, the PHD finger in the BPTF protein recognizes H3K4me3 through the anti-parallel β-sheet and the H3K4me3 is inserted into the deep packets on the BPTF surface. Since BPTF is a component of the ATP-dependent chromatin remodeling complex, the binding may contribute to chromatin remodeling during transcription activation. The PHD finger in RAG2 proteins can influence V(D)J recombination through recognition of H3K4me3. Another big family of proteins that recognize methylated lysine residues is the "Royal family" domain. The chromodomains in HP1 and PC recognize methylated H3K9 and H3K27, respectively. The binding may help trigger the formation of heterochromatin and/or recruitment of downstream of repressive complexes. The Tudor domain has high degrees of structural and functional diversity. Similar to the chromodomain, the Tudor domain recognizes methylated lysine through an hydrophobic packet formed by 2–4 aromatic amino acids. The tandem Tudor domain in 53BP1 and JMJD2A has been implicated in the binding of methylated histones. In this case, the tandem Tudor domain of 53BP1 has been implicated in binding of H4K20me2. The two Tudor domains in JMJD2A are interdigitated and form the binding surface of H3K4me3 and H4K20mes. Many MBT domain proteins belong to the Polycomb protein family. This domain shows structural similarity to the "Royal family" of histone binding domains. However, MBT domain may bind to methylated lysine with low methylation level (e.g. mono- and di-).

A subgroup of Tudor domains, such as those in the SMN, SPF30 and TDRDs, can specifically recognize arginine methylation [2, 3]. Yet, it is difficult to predict which Tudor domain recognizes methylated lysine or arginine just based on the protein sequences. Tudor domain proteins that bind to methylarginine are often linked to RNA metabolism. For example, both SMN and SPF30 are involved in regulation of pre-mRNA splicing, and many members of the TDRD family of proteins regulate small interference RNA (siRNA) silencing pathway through recognition of methylated arginine residues. Interestingly, the expression of most TDRD proteins is particularly high in the germ cells, which might implicate their involvement in the piRNA pathway during gametogenesis. TDRD3 is the only known Tudor protein that was shown to recognize arginine-methylated histones. TDRD3 preferentially recognizes H3R17me2a and H4R3me2a mark and promotes transcription.

## *3.4 Function of Histone Methylation*

Although histone methylation does not alter chromatin function directly, methylation can affect the binding of protein factors, in turn repressing or activating transcription (Fig. 4, repressive marks are colored red and permissive marks are colored green) [14]. Furthermore, different degrees of histone methylation can have different functional outputs. For example, methylation of H3K4 at di- or tri-methylated status usually locates on the promoter regions of active genes. Methylation of H3K4 at mono- methylated status often locates at enhancer regions. These methylation events are believed to activate transcription, directly or indirectly. Indeed, the enzymes that catalyze H3K4 methylation are associated with actively transcribed RNA polymerases and add these histone marks as transcription occurs. These methylation events may help establish chromatin status for the next round of transcription. In contrast, methylated H3K36 is localized in gene coding regions. This mark is deposited on chromatin by association of the enzymes with elongating form of RNA polymerases. This mark may help contribute to the stability of nucleosomes after transcription and may also modulate the pre-mRNA splicing processes. Methylation of H3K79 is also located in gene coding regions. Both H3K4 and H3K79 methylation are controlled by H2B ubiquitination and ongoing transcription. In addition to these lysine modifications, H3R2, H3R16, and R26 are also correlated with gene activation. Methylation of these three arginine residues is carried out by the co-activator CARM1. Arginine methylation may enhance the histone acetylation and contribute to gene activation. Similarly, H4R3 can also be methylated, and this methylation facilitates downstream histone acetylation and positively contributes to gene activation.

Histone methylation sites that are linked to gene repression include H3K9, H3K27, and H4K20 methylation (Fig. 4). These modifications may contribute to the closed chromatin conformation and transcription repression, directly or indirectly. Methylations of H3K9 di- and tri- status are involved in heterochromatin formation and thus gene repression. These modifications are mediated by SUV39H1 and are

particularly enriched at percentermatic regions. These modifications recruit the HP1 proteins and serve as seeding sites for heterochromatin spreading. This modification is important in the formation of constitutive heterochromatic structure and thus genomic stability. In contrast, H3K9 di- and tri- methylation in eurochromatic regions is mediated by G9a. This modification facilitates the formation of facultative heterochromatin and transcription repression. Methylation of H3K27 is catalyzed by Polycomb repressive complex 2 and therefore this modification is believed to create a binding site for downstream PRC1. Therefore, H3K27 methylation is specifically linked to polycomb group protein-mediated gene silencing. Methylation of H4K20 is also associated with a closed chromatin status and may function in DNA damage repair. In this case, H4K20 methylation is recognized by 53BP1 protein and helps localize 53BP1 to damaged DNA foci, in conjunction with H2AK15 ubiquitination. Although histone demethylases have been identified, histone modifications are generally stable and can be passed down to progeny with the mechanism remains to be elucidated.

## 4 Perspective

Methylation of DNA and histones are important components of the epigenetic machinery. These two modifications, together with other epigenetic regulatory mechanisms, contribute to the stable expression of gene expression program specific to a given cell type. The epigenetic mechanism is particularly important as cells undergo normal programming during development and undertake reprograming to acquire a cancerous phenotype. Compared to gene mutation and deletion, epigenetic mechanisms are reversible and can be targeted by small molecule inhibitors, and thus epigenetics has appeared as an appealing target for cancer treatment.

As important components of epigenetics, both DNA and histone methylation depends on the availability of methyl group donor SAM. In humans, there are three major resources of methyl group donor: methionine (~ 10 mmol of methyl/day), one carbon metabolism via methylfolate (~5–10 mmol of methyl/day), and choline (~30 mmoles methyl/day). The three methyl group donors can compensate each other; however, when severely depleted from diet, methylation of DNA and histones can be affected. Since proper methylation of DNA and histones may be important for normal cell functions, scientists recommend adequate methyl donors in the diet such as green tea, red wine, spinach, walnuts, and pomegranate [4]. This diet may be of benefit in repressing the expression of harmful genes.

Since the epigenetic system plays a key role in the cell gene expression program, this system has to undergo profound reprograming as cells revert to acquire uncontrolled proliferative features. As important components of epigenetics, there has been a wealth of literature that both DNA and histone methylation undergo significant changes during the carcinogenesis processes. Consequently, DNA methylation inhibitors have been used in certain cancer treatments, in conjunction with other treatments. The specific effects on cancer cells are enhanced by the relatively rapid

division of these cells *vs* normal somatic cells. Along this line, a number of novel inhibitors for HMTs are under way for clinical trial. One side effect for these inhibitors and possibly all inhibitors targeting epigenetic revenue is that epigenetics are also essential for normal cell function. How to specifically target these inhibitors to cancer cells may be a long way for future studies. However, strong basic science and mechanism-based clinical trials will overcome the issue.

## References

1. Allis CD, Caparros M-L, Jenuwein T, Reinberg D, Lachlan M (2015) Epigenetics, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
2. Beaver JE, Waters ML (2016) Molecular recognition of Lys and Arg methylation. ACS Chem Biol 11:643–653
3. Chen C, Nott TJ, Jin J, Pawson T (2011) Deciphering arginine methylation: Tudor tells the tale. Nat Rev Mol Cell Biol 12:629–642
4. Choi SW, Friso S (2010) Epigenetics: a new bridge between nutrition and health. Adv Nutr (Bethesda, Md) 1:8–16
5. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25:1010–1022
6. Hackett JA, Surani MA (2013) DNA methylation dynamics during the mammalian life cycle. Philos Trans R Soc Lond Ser B Biol Sci 368:20110328
7. Jahan S, Davie JR (2015) Protein arginine methyltransferases (PRMTs): role in chromatin organization. Adv Biol Regul 57:173–184
8. Jeltsch A, Ehrenhofer-Murray A, Jurkowski TP, Lyko F, Reuter G, Ankri S, Nellen W, Schaefer M, Helm M (2016) Mechanism and biological role of Dnmt2 in nucleic acid methylation. RNA Biol:1–16
9. Klose RJ, Bird AP (2006) Genomic DNA methylation: the mark and its mediators. Trends Biochem Sci 31:89–97
10. Klose RJ, Zhang Y (2007) Regulation of histone methylation by demethylimination and demethylation. Nat Rev Mol Cell Biol 8:307–318
11. Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. Nature 502:472–479
12. Patel DJ, Wang Z (2013) Readout of epigenetic modifications. Annu Rev Biochem 82:81–118
13. Rottach A, Leonhardt H, Spada F (2009) DNA methylation-mediated epigenetic control. J Cell Biochem 108:43–51
14. Shilatifard A (2006) Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. Annu Rev Biochem 75:243–269
15. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14:204–220
16. Tessarz P, Kouzarides T (2014) Histone core modifications regulating nucleosome structure and dynamics. Nat Rev Mol Cell Biol 15:703–708
17. Turek-Plewa J, Jagodzinski PP (2005) The role of mammalian DNA methyltransferases in the regulation of gene expression. Cell Mol Biol Lett 10:631–647