



Epigenetics of Modified DNA Bases: 5-Methylcytosine and Beyond

Suresh Kumar^{1*}, Viswanathan Chinnusamy² and Trilochan Mohapatra³

¹ Division of Biochemistry, Indian Agricultural Research Institute (ICAR), New Delhi, India, ² Division of Plant Physiology, Indian Agricultural Research Institute (ICAR), New Delhi, India, ³ Indian Council of Agricultural Research, New Delhi, India

OPEN ACCESS

Edited by:

Mojgan Rastegar,
University of Manitoba, Canada

Reviewed by:

James Ronald Davie,
University of Manitoba, Canada
Tetsu Kinoshita,
Yokohama City University, Japan
Ian C. G. Weaver,
Dalhousie University, Canada

*Correspondence:

Suresh Kumar
sureshkumar3_in@yahoo.co.uk

Specialty section:

This article was submitted to
Epigenomics and Epigenetics,
a section of the journal
Frontiers in Genetics

Received: 09 October 2018

Accepted: 27 November 2018

Published: 18 December 2018

Citation:

Kumar S, Chinnusamy V and
Mohapatra T (2018) Epigenetics
of Modified DNA Bases:
5-Methylcytosine and Beyond.
Front. Genet. 9:640.
doi: 10.3389/fgene.2018.00640

Modification of DNA bases plays vital roles in the epigenetic control of gene expression in both animals and plants. Though much attention is given to the conventional epigenetic signature 5-methylcytosine (5-mC), the field of epigenetics is attracting increased scientific interest through the discovery of additional modifications of DNA bases and their roles in controlling gene expression. Theoretically, each of the DNA bases can be modified; however, modifications of cytosine and adenine only are known so far. This review focuses on the recent findings of the well-studied cytosine modifications and yet poorly characterized adenine modification which serve as an additional layer of epigenetic regulation in animals and discuss their potential roles in plants. Cytosine modification at symmetric (CG, CHG) and asymmetric (CHH) contexts is a key epigenetic feature. In addition to the ROS1 family mediated demethylation, Ten-Eleven Translocation family proteins-mediated hydroxylation of 5-mC to 5-hydroxymethylcytosine as additional active demethylation pathway are also discussed. The epigenetic marks are known to be associated with the regulation of several cellular and developmental processes, pluripotency of stem cells, neuron cell development, and tumor development in animals. Therefore, the most recently discovered N⁶-methyladenine, an additional epigenetic mark with regulatory potential, is also described. Interestingly, these newly discovered modifications are also found in the genomes which lack canonical 5-mC, signifying their independent epigenetic functions. These modified DNA bases are considered to be important players in epigenomics. The potential for combinatorial interaction among the known modified DNA bases suggests that epigenetic codon is likely to be substantially more complicated than it is thought today.

Keywords: cytosine methylation, DNA modification, epigenetic marks, modified DNA base, 5-hydroxymethylcytosine, 5-methylcytosine, N⁶-methyladenine

INTRODUCTION

Modification of DNA base is one of the main epigenetic mechanisms which regulate gene expression in both plants and animals. DNA methylation indicates attachment of a methyl (CH₃) group at 5'-carbon of pyrimidine ring of cytosine nucleotide. In addition to methylation of cytosine residue, other modifications such as oxidation of methylated cytosine (5-mC) to

5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fC), 5-carboxylcytosine (5-caC), and methylation of adenine (A) to N⁶-methyladenine (6-mA), are being identified as important epigenetic regulators (Klungland and Robertson, 2017). Supplementary epigenetic diversity in regulating gene expression is accomplished by post-translational modifications of histone proteins of the nucleosome. These modifications regulate cellular machinery to modulate specific chromatin regions and mark the regions for various cellular functions. The sum total of the chemical modifications in the nuclear DNA and N-terminal tail of histone proteins constitute epigenome. Epigenetics is a branch of functional genomics which deals with regulation of gene expression through DNA base and histone modifications (Kumar, 2017). For the last two decades, the scope of epigenetic studies is expanding continuously with the identification of more and more epigenetic marks (Shi et al., 2017). Identification of additional DNA base modifications (5-hmC and 6-mA), known to have epigenetic regulatory functions in animals, resulted in the increased significance of epigenomic studies. 5-mC is a well-studied epigenetic mark among the other modified nucleosides in DNA.

Epigenomic changes play an important role in the regulation of gene expression in both animals and plants during development, suppression of transposable elements (TEs) and response to environmental stresses (Chinnusamy and Zhu, 2009; Kumar and Singh, 2016; Kumar et al., 2017a). While the genome is largely consistent within an individual throughout its life, the epigenome is dynamically influenced by environmental factors and developmental processes; hence, the epigenome may vary in different cells under different environmental conditions (Shi et al., 2017). The epigenetic changes may cause variation in the structure of chromatin and expression of the genome. The epigenetic changes may result in chromatin remodeling, which may cause an alteration in transcription level even after withdrawal of the stress (Avramova, 2015). The epigenetic marks are maintained by DNA methylases (writers) and demethylases (erasers), and the reader proteins (see the section “**Glossary**”) read and interpret the encoded information (Torres and Fujimori, 2015; Feinberg et al., 2016). Thus, a complex interactions among the different molecular factors including DNA methylation/demethylation, enzymes responsible for post-translational modifications of histone proteins, non-coding RNAs and chromatin remodelers are accountable for the epigenetic regulation of gene expression (Lauria and Rossi, 2011; Pikaard and Mittelsten-Scheid, 2014; Gallusci et al., 2016).

DNA (de)methylation regulates activation/silencing of TEs, genomic imprinting, developmental processes and stress responses in both plants and animals (Chinnusamy and Zhu, 2009; Allis and Jenuwein, 2016; Springer and Schmitz, 2017). This review focuses on the recent advances in understanding the functional consequences of the well-studied cytosine modifications and the recent studies on epigenetic modification of adenine which appears to serve as an additional layer of epigenetic regulation.

EPIGENETIC DNA MODIFICATIONS

Chemical modifications of nitrogenous bases of DNA play a vital role in the regulation of gene expression. Methylated cytosine (5-mC), also known as the fifth base of DNA, was recognized long before the DNA was identified as the genetic material. While more attention is given on the conventional modified DNA base 5-methylcytosine (5-mC), the recent discovery of additional base-modifications has resulted in an increased interest in epigenomics studies. DNA base modifications have been detected in all the kingdoms of living organisms, including eukaryotes, prokaryotes, and viruses. More importantly, the dynamics of epigenetic gene regulation requires removal of the epigenetic mark. The newly discovered diversity in epigenetic modifications and the potential for their combinatorial interaction indicate that the epigenetic codons are considerably more complicated than it is considered presently (Breiling and Lyko, 2015).

THE FIFTH BASE OF NUCLEIC ACID: 5-METHYLCYTOSINE

DNA contains four nitrogenous bases namely cytosine (C), guanine (G), adenine (A), and thymine (T). T is replaced by uracil (U) in the case of RNA. DNA may also contain modified bases like 5-mC, 5-hmC, 5-fC, 5-carboxycytosine (5-caC), and 6-mA in a small amount (Kumar, 2018b), while RNA contains more than 140 different types of modified bases (Sloan et al., 2017). Methylcytosine (5-mC) is most common among these modified bases in the genome, and hence it is considered as the fifth base of DNA. More than 4% of the cytosines present in the human genome have been reported to be methylated (Breiling and Lyko, 2015). However, the level of 5-mC content may vary significantly among animals and plants. The significance of the 5-mC cannot be demarcated merely by its abundance, it also depends on its positioning [in symmetric (CG and CHG) and asymmetric (CHH) contexts] as well as on its location in different parts of a gene. In the animal, 5-mC occurs predominantly in CG context, but in plants, it occurs in all the three contexts (CG, CHG, and CHH) (Wang et al., 2016). Symmetric methylation indicates that 5-mC is present in CG or CHG contexts in the antiparallel strands of DNA, and the methylation pattern can be truthfully reproduced during DNA replication. In the human genome, >80% of the cytosine present in CG context is methylated, which indicates ubiquitous methylation landscape of the genome; however, site-specific gaps are often present in the regulatory elements (e.g., enhancers and promoters) of the actively transcribed genes. On the contrary, an extremely low level of methylation has been reported in certain invertebrates like *Drosophila* and silkworm (Xiang et al., 2010).

Cytosine methylation affects the accessibility of the genomic regions to regulatory proteins/protein complexes, which influences chromatin structure and/or affects the rate of transcription of the gene. Cytosine methylation is the only well-studied DNA modification with established maintenance mechanisms. It has been associated with repression of the gene when present in the promoter and enhancer regions (Charlet

et al., 2016), but gene body methylation (gbM), i.e., 5-mC in the coding/transcribed region excluding the transcription start site (TSS) and transcription termination site (TTS), might repress or enhance transcriptional activity (Buck-Koehn and Defossez, 2013; Spruijt and Vermeulen, 2014).

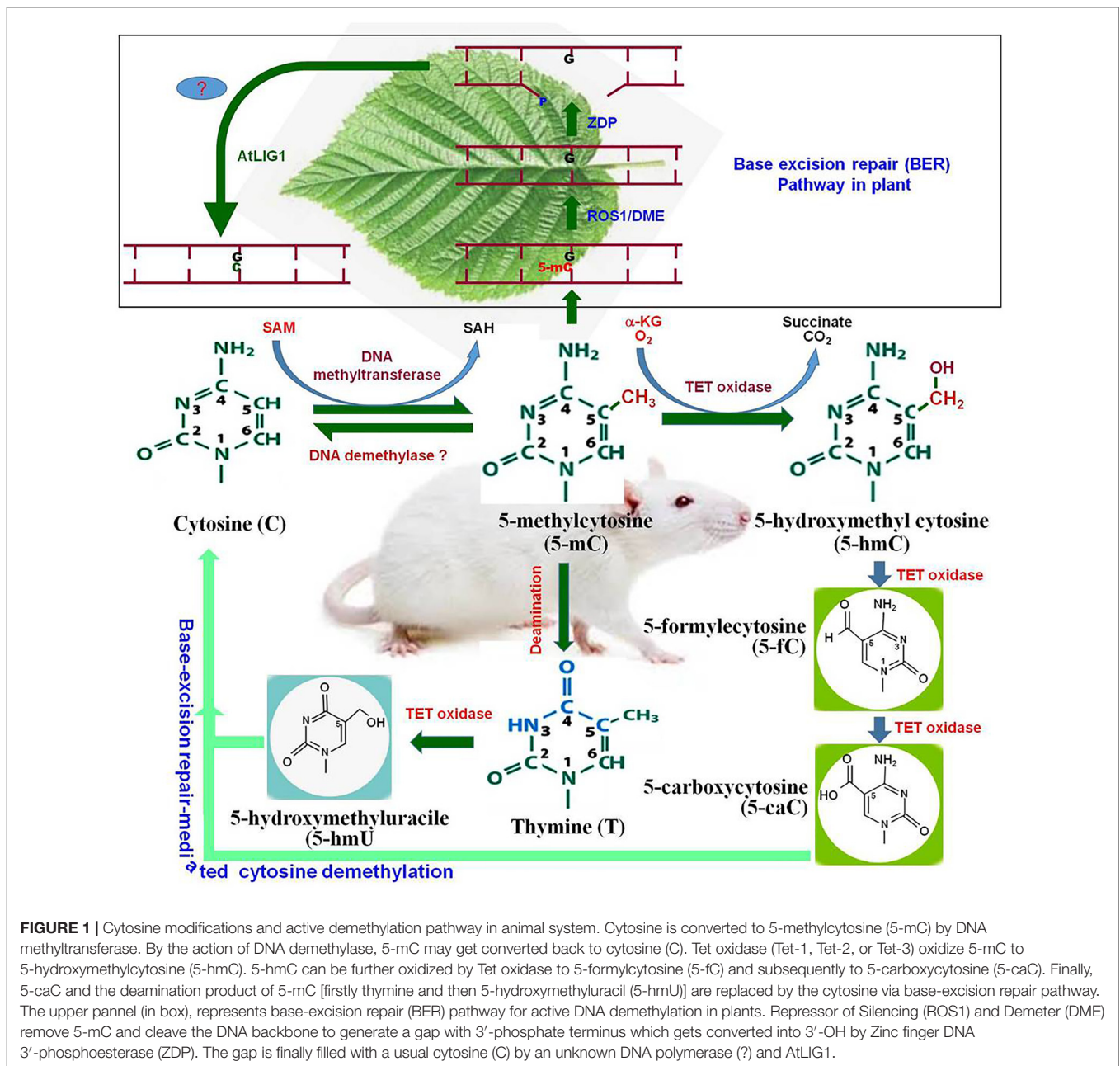
In plants, cytosine methylation at symmetric (CG and CHG) context is maintained by methyltransferase 1 (MET1) and chromomethylase 3 (CMT3), while methylation at asymmetric (CHH) context is maintained by RNA-dependent DNA methylation pathway (RdDM) pathway or by the chromatin remodeler DDM1-dependent chromomethylase 2 (Zemach et al., 2013). Whole-genome bisulfite sequencing analysis of *Arabidopsis thaliana* revealed that gbM is limited to CG context, and methylation at CHG and CHH context is found in TE- and repeats-enriched heterochromatin regions. Methylation level varies significantly in CG and non-CG contexts. In *A. thaliana*, methylation at CG context was reported to be 24%, while methylation at CHG and CHH context was found to be 7% and 2%, respectively (Cokus et al., 2008). It was found to be 86%, 74%, and 5% at CG, CHG and CHH contexts, respectively, in maize at the reproductive stage (Gent et al., 2013). Methylation at the non-CG context in plants plays a key role in silencing exogenous DNA via the RdDM pathway (Law and Jacobsen, 2010). Therefore, it is reasonable to assume that the methylation of gene/genome is the default state, and specific mechanisms are needed to maintain specific regions free from methylation. Methylation level is dynamically controlled by DNA (de)methylation processes. DNA demethylation takes place by active and/or passive methods. During active DNA demethylation, 5-mC is oxidized to 5-hmC, 5-fC, and further to 5-caC followed by base-excision repair (BER) mechanism which is independent of replication process (Figure 1). Thus, active DNA demethylation engages enzymatic removal of 5-mC. In plants, a family of DNA glycosylases (erasers) namely DME (Demeter) ROS1 (Repressor of Silencing 1), DML2 (Demeter-like 2), and DML3 initiate active DNA demethylation, and it is completed by the BER-dependent mechanism (Zhu, 2009) (Figure 2). DME directly removes 5-mC from the DNA backbone, creates an abasic site which is filled up by the BER pathway (Li et al., 2018). Mutations in the gene(s) of DNA demethylation pathway affect DNA methylation, which leads to genome-wide DNA hypermethylation, and cause silencing/activation of the TEs/genes expression of which is no longer needed and must be repressed (Lang et al., 2017; Frost et al., 2018). Transcriptional repression of MET1 was reported to be associated with genome-wide passive DNA demethylation (Ji et al., 2014). However, it is still not known how global demethylation occurs during gametogenesis and embryogenesis in plants. BER pathway cannot be the primary process for global demethylation because this would result in so many abasic sites/broken DNA strands, which will destabilize the genome (Zhu, 2009). Thus, we can speculate that another pathway might be involved in active DNA demethylation in plants, and this unknown active demethylation pathway might be similar to the known Ten-Eleven Translocation (TET) pathway in mammals; however, such pathway has not yet been identified. Activation of genes during the developmental process, environmental stresses, and genome-wide epigenetic reprogramming are mediated by

active DNA demethylation (Hsieh et al., 2009; Kumar et al., 2017b).

OXIDATION OF 5-mC CREATES ANOTHER EPIGENETIC MARK: 5-HYDROXYMETHYLCYTOSINE

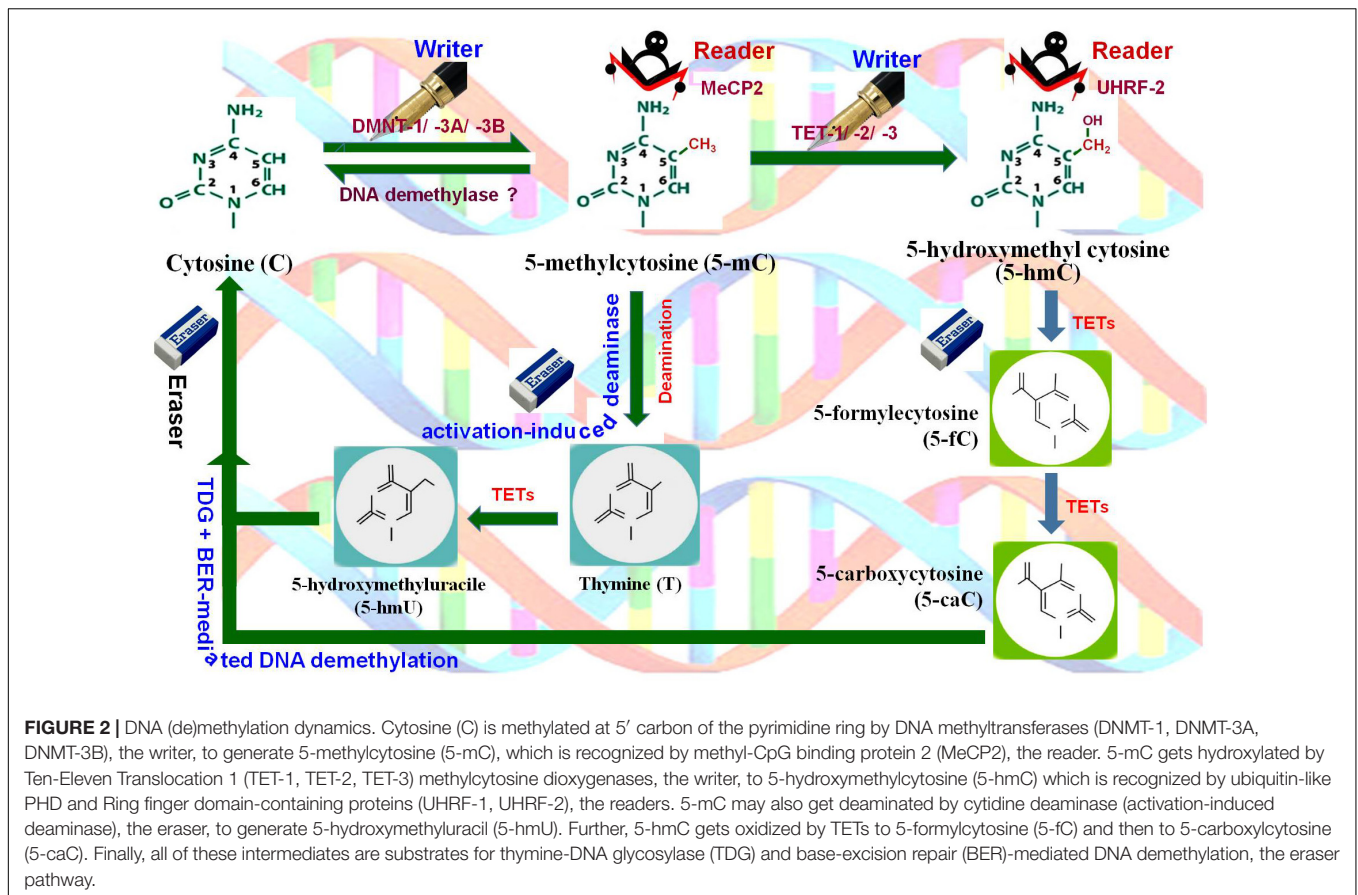
Shortly after the discovery of 5-mC in 1948, 5-hmC was identified in a bacteriophage and vertebrates, including a frog, mouse, and rat. Hydroxylation of 5-mC in viral DNA was reported earlier to be a mechanism to tag it to differentiate from the cytosine of host DNA and to avoid BER pathway of the host defense mechanism. Though the biological significance of 5-hmC has been investigated lately (Kriaucionis and Heintz, 2009), its function as an epigenetic mark is poorly understood (Shi et al., 2017). 5-hmC is an intermediate of active DNA demethylation process, and it is a stable epigenetic mark in animals. With the discovery of enzymatic functions of TET family proteins, 5-hmC and the TET-dependent oxidation products [5-fC, 5-caC, 5-hydroxymethyluracil (5-hmU)] are considered to be the demethylation intermediates of 5-mC in animals. Interestingly, these modified-cytosine bases might act as important epigenetic marks. These epigenetic marks (5-fC, 5-caC, and 5-hmU) have been reported to play an important role in the regulation of transcription process, chromatin remodeling, and recruitment of DNA repair-associated complexes in animal (Fong et al., 2013; Yue et al., 2016). Novel epigenetic modifications of DNA are now being identified with the discovery of catalytic dioxygenase activity of the TET enzymes (Tahiliani et al., 2009; Ito et al., 2010). 5-hmC content in mammalian tissues is about 0.1% but can vary greatly (Globisch et al., 2010) with the highest level in the brain where it can go up to 1% (Kriaucionis and Heintz, 2009). Three mammalian TET homologs (TET-1, TET-2, and TET-3) are involved in the conversion of 5-mC to 5-hmC, 5-fC, and to 5-caC. While *Tet1* was found to express in the embryonic stem cells (ESCs), *Tet2* and *Tet3* exhibited a similar expression pattern in various tissues. Thus, *Tet* genes show cell/organ-specific expression (Ito et al., 2011; Tollervy and Lunyak, 2012). In mouse ESCs (mESCs), about 30000 5-mC, 1300 5-hmC, 20 5-fC, and only 3 5-caC per million cytosine residues were observed (He et al., 2011; Ito et al., 2011), which indicate the sporadic presence of 5-fC and 5-caC. These unusual modified bases (5-fC and 5-caC, and deamination products of 5-mC: thymine or 5-hmU), are removed by BER mechanism (He et al., 2011; Ito et al., 2011). Thymine-DNA glycosylase (TDG) removes T from G:T mismatches, and initiates BER of the deaminated 5-mC. TDG was found to be active on 5-fC and 5-caC, but not active on 5-mC or 5-hmC (He et al., 2011; Maiti and Drohat, 2011). Thus, cytosine methylation dynamics is managed by the activities of the DNA methyltransferases, TET enzymes and TDG pathway.

The role of 5-hmC as an epigenetic mark and the specific functions of TET enzymes in animal system are beginning to emerge. *TET1* and *TET2* genes are highly expressed in mESCs, but a depleted expression of either of these genes did not affect pluripotency or development (Dawlaty et al., 2011; Ko et al., 2011; Koh et al., 2011; Li et al., 2011). Homozygous



TET3 mutant mice showed proper embryonic development, but they died at birth (Gu et al., 2011), which suggests that *TET3* is not necessary for embryonic development. Mice ESCs with defective *TET1* and *TET2* showed no significant change in 5-hmC level and retained pluripotency, but developmental defects associated with the ectopic hypermethylation were observed (Dawlaty et al., 2013). However, triple-mutant (*TET-1*, *TET-2*, and *TET-3*) mESCs were found to be viable, pluripotent and showed depleted 5-hmC content (Dawlaty et al., 2014; Lu et al., 2014). They showed hypermethylation of promoters and impaired differentiation potential (Dawlaty et al., 2014), which indicate a major role of active DNA demethylation via *TET*-mediated oxidation in maintaining the regulatory regions,

particularly enhancers and promoters, free from 5-mC (Lu et al., 2014). The potential role of 5-hmC as active epigenetic mark was further supported by mass spectrometric analysis of radiolabeled DNA from mammalian and mice cells having stable 5-hmC without any transient intermediate (Bachman et al., 2014). The 5-hmC level was found to increase during neuron differentiation, and stable intragenic 5-hmC content was observed in many active neuron-specific genes (Hahn et al., 2013). The abundance of 5-hmC was found in constitutive exons compared to that in the alternatively spliced exons (Khare et al., 2012). Therefore, it is plausible to assume that 5-hmC plays a role in alternate splicing, wherein it contributes to the binding of regulatory proteins to DNA. Thus, 5-hmC may



be considered as a stable modified-base found in mammalian promoters, enhancers, near the TSS, coding region of the actively transcribed genes, 3' UTR and intragenic regions (Wu and Zhang, 2011; Pastor et al., 2013). It is also believed that 5-hmC acts as *cis*-element to repress or promote expression of the gene. Besides, 5-hmC can be associated with histone-modification which may collectively alter the configuration of chromatin as well as with switching “on” or “off” the genes in heterochromatic and/or euchromatic regions (Shi et al., 2017).

While there is evidence for epigenetic functions of 5-hmC, at least in some tissues of animals, the similar function of its oxidation-derivatives (5-fC and 5-caC) appears to be less likely. Although the presence of 5-fC at the 5-hmC – marked regions suggest its potential role as an independent epigenetic mark, yet it needs to be confirmed. The level of 5-fC and 5-caC was found to increase in thymine-DNA-glycosylase-deficient mESCs, which suggests that 5-caC represents the site of active DNA demethylation (Raiber et al., 2012; Song et al., 2013). It has also been observed that TET proteins can oxidize thymine (particularly those produced by deamination of 5-mC) to 5-hmU which enters into BER pathway (Figure 1) (Pffaffeneder et al., 2014). Though some of the reports suggest the existence of 5-hmC in plants, they have conflicting inferences. Erdmann et al. (2015) investigated the existence of 5-hmC in *A. thaliana* and other plants using various techniques like

thin-layer chromatography, ELISA, ChIP-chip, enzymatic-radio-labeling, and mass spectrometry. While the antibody-based techniques suggested a lower level of 5-hmC in the plant genome, the most sensitive techniques like enzymatic-radio-labeling and mass spectrometry failed to detect it in genomic DNA from different tissues. The findings suggest that 5-hmC may not be present in biologically relevant quantity in the plant genome (Erdmann et al., 2015). Studies on three rice cultivars indicated that 5-hmC is present at very low quantity (1.39 to 2.17 per million nucleosides) especially in TEs and heterochromatin regions (Wang et al., 2015). Moreover, the occurrence of 5-hmC was detected in the inactive TEs.

Neither TET proteins nor UHRF2 [Ubiquitin-like, PHD and Ring Finger Domains 2, which specifically recognize 5-hmC, the reader (Figure 2)], well characterized in the animal system (Zhou et al., 2014), have been identified yet in plants. Therefore, it is speculated that DNA demethylation is not processed through TET pathway in plants. It is not yet clear that 5-hmC is implicated in what type of cellular processes in the animal system, and its epigenetic relevance is yet to be understood (Klungland and Robertson, 2017). Hence, existence and epigenetic role of 5-hmC, if any, is yet to be discovered in plants (Terragni et al., 2012; Moricová et al., 2013; Wang et al., 2015). These findings rejuvenate interest in understanding the epigenetics of 5-hmC, and hence 5-hmC is now considered as the sixth base of the epigenome (Shi et al., 2017).

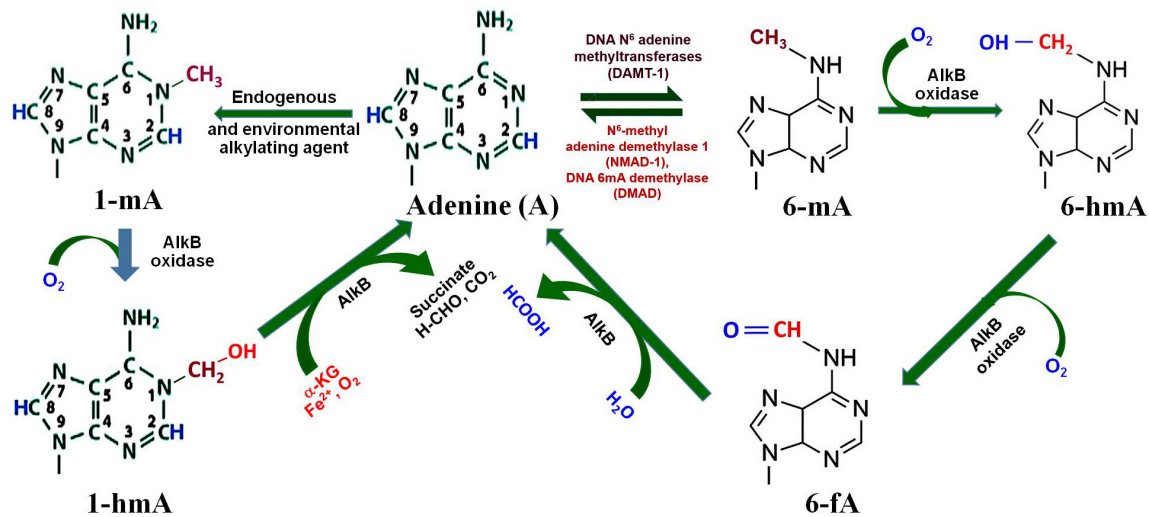


FIGURE 3 | Adenine modification dynamics. Adenine gets methylated by DNA N⁶-adenine methyltransferases (DAMT-1) to produce N⁶-methyladenine (6-mA), as observed in *C. elegans*. 6-mA can be demethylated by the action of N⁶-methyl adenine demethylase-1 (NMAD-1) or DNA methyladenine demethylase (DMAD). Oxidation of methyl group of 6-mA by AlkB oxidase results in the formation of N⁶-hydroxymethyl adenine (6-hmA) and N⁶-formyl adenine (6-fA), and finally back to adenine. Adenine may also get methyl adduct by endogenous or environmental alkylating agents to N¹-methyladenine (1-mA), which may further get demethylated by AlkB oxidase to adenine via N¹-methylhydroxy adenine (1-hmA).

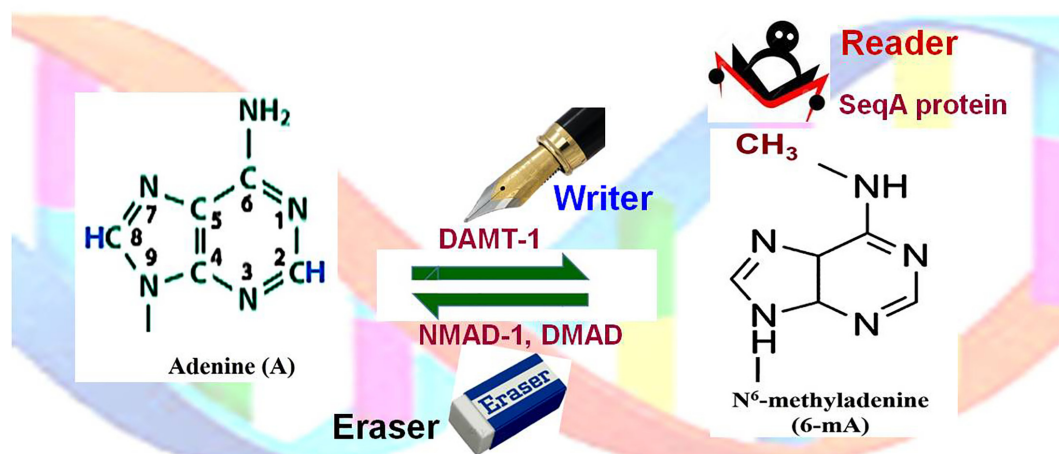


FIGURE 4 | Adenine (de)methylation dynamics. Adenine (A) is methylated at N6 of the purine ring by DNA N⁶-adenine methyltransferases 1 (DAMT-1), the writer, to generate N⁶-methyladenine (6-mA). SeqA protein, the reader, preferentially binds to hemimethylated (6-mA) DNA. 6-mA may get demethylated by the action of N⁶-methyl adenine demethylase-1 (NMAD-1) or DNA methyladenine demethylase (DMAD), the eraser.

NON-CYTOSINE METHYLATION OF DNA: N⁶-METHYLADENINE

Methylation of adenine, i.e., 6-mA, in GATC context has been reported for the survival of several bacteria, as DNA adenine methyl transferase (DAMT, the writer) creates specific methylation marks (Figure 3) that are important for DNA replication, mismatch repair, segregation, and regulation of gene expression (Ratel et al., 2006). 6-mA is known to play an important regulatory role in RNAs, and recent studies suggest the presence of 6-mA in eukaryotic genomes (Kigar et al.,

2017; Yao et al., 2017, 2018). Moreover, oxidation of the methyl group of 6-mA by AlkB family of dioxygenases (e.g., 6-mA demethylases) results in the formation of 6-hmA and 6-fA, which can restore the original base (Figure 3) by releasing formaldehyde (Fu et al., 2013). A recent study on *Caenorhabditis elegans* demonstrated that NMAD-1 (an AlkB family enzyme) demethylates 6-mA in DNA (Fu et al., 2015). The study suggests that the enzymes of the AlkB family are critical players in demethylation of 6-mA in DNA (Iyer et al., 2016). Adenine may also get methylated to N¹-methyladenine (1-mA) by the endogenous or environmental alkylating agents (Sedgwick et al.,

2007), which may get demethylated indirectly by AlkB oxidase to adenine via N¹-methylhydroxy adenine (1-hmA).

In *Escherichia coli*, 1-mA is generated due to the environmental alkylation agents which creates a local lesion, prohibits the formation of regular Watson–Crick base pairing, and consequently blocks DNA replication (Sedgwick et al., 2007). *AlkB* is an inducible gene in *E. coli* for adaptive response to the alkylating agents. Human AlkB homologs perform similar base modifications and exhibit overwhelming functional roles (Westbye et al., 2008). Similarly, N⁷-methylguanine (7-mG) is also produced due to alkylation by the endogenous and/or environmental alkylating agents. A comprehensive understanding of these base modifications and their functional roles would be necessary to discover their contribution, if any, in epigenetic regulation of gene expression in animals and plants.

In several unicellular eukaryotes, including *Chlamydomonas reinhardtii*, comparatively higher level of 6-mA has been reported (Ratel et al., 2006). Several powerful techniques have been developed for detection of 6-mA, and this has resulted in the identification of some key features of 6-mA in *Chlamydomonas* (Fu et al., 2015). The algal adenine-methylome consists of about 85000 6-mA in AT sequence context, mainly in the promoter and the linker regions. Therefore, it was proposed that 6-mA marks the position of nucleosomes near TSS (Fu et al., 2015). Moreover, *Chlamydomonas* genome has been characterized to have a low level of cytosine methylation in CG, CHG, and CHH contexts in the genes, which corroborates with the methylation pattern in plants (Wang et al., 2016). *C. elegans* and *Drosophila melanogaster* possess the negligible amount of 5-mC or 5-hmC, but the significantly high level of 6-mA. In these organisms, such DNA base modification is considered as a new epigenetic mark. Experimental data from *C. elegans* suggest a functional correlation between 6-mA and a histone modification H3K4me2 (Greer et al., 2015). In *D. melanogaster*, mutations in DNA 6-mA – demethylase (DMAD, a TET-homolog) caused increased TE activity (Zhang et al., 2015). In these organisms, mutations in DMAD (eraser) caused significant phenotypic aberrations including developmental defects and infertility, suggesting epigenetic functions of 6-mA. SeqA protein serves as a reader of 6-mA (Figure 4), preferentially binds to hemimethylated DNA, and affects gene expression/chromatin activity. Therefore, future research needs to discover the functions of the newly identified epigenetic marks (5-hmC, 5-fC, 5-caC, 5-hmU, and 6-mA), enzymes that are involved in setting up (writing) and removing (erasing) these modifications, and the reader that recognizes/binds the epigenetic mark.

Report on the occurrence of 6-mA in higher eukaryotes is sparse, and the finding has been often inconclusive (Ratel et al., 2006). Yao et al. (2017) demonstrated the dynamics of 6-mA in mouse brain under environmental stress. They reported that the levels of 6-mA get significantly elevated upon the stress. They also observed an inverse relationship between 6-mA dynamics and expression of certain upregulated neuronal genes or downregulated transposons. However, the presence and function of 6-mA in the mammal remain unclear. Highly sensitive mass spectrometry technique detected less than one

6-mA per million nucleotides in mouse genomic DNA (Fong et al., 2013). This suggests that 6-mA is either not a common (like 5-mC/5-hmC) modified base or its turnover by demethylation process is very rapid. However, accumulation of 6-mA by deactivating DMAD has been demonstrated in *D. melanogaster* (Zhang et al., 2015). It has also been found that demethylation of 6-mA is mediated by a TET-like enzyme in *D. melanogaster* (Zhang et al., 2015). Their finding indicates that cytosine and adenine (de)methylation occur in a coordinated, dynamic and sequence context-specific manner. Recently, Yao et al. (2018) reported epigenetic regulation of a group of genes involved in neurodevelopment and neuronal functions in *D. melanogaster*. Accumulation of 6-mA due to deactivated DMAD coordinates with Polycomb proteins and contributes to transcriptional repression of the genes. Hence, it would be interesting to investigate the interplay between different epigenetic marks to explore the complexity of the epigenetic code, which might answer several biological enigmas in the near future (Kumar, 2017).

Study on *C. elegans* revealed that about 0.3% adenine are methylated in GAGG and AGAA sequence contexts. Interestingly, 6-mA was observed to accumulate in *C. elegans* deficient for *spr-5* (codes for H3K4me2 demethylase) (Greer et al., 2015). While 5-mC increases DNA helix stability, 6-mA destabilizes the helical structure of DNA. Therefore, 5-mC is considered to be a repressor of transcription process when it occurs in the promoter region, and 6-mA functions as an activator of transcription.

ADVANCES IN DETECTION OF MODIFIED DNA BASES

Epigenetic changes are reported to be important players in determining cellular growth, development, differentiation processes, and tolerance to abiotic and biotic stresses in the organisms. There is evidence suggesting an association between variations in DNA methylation and environmental perturbations in animals and plants. Therefore, detection of epigenetic changes has become critically important for scientific/diagnostic purposes. Information about alterations in epigenetic marks might be useful in identifying the stress level and the associated problems, such as the occurrence of cancer in an animal might be detected right at the early stage of development. One of the exciting technological advancements in the field of epigenomics has been the invention of different efficient methods having certain advantages and disadvantages to detect/analyze DNA base modifications (Cokus et al., 2008; McIntyre et al., 2017). Considerable progress is being made in the identification/quantification of the modified DNA bases, which has significantly broadened the area of epigenomics research.

Detection of 5-mC

To detect methylcytosine (5-mC) at the genome level, methylation-sensitive amplified polymorphism (MSAP) technique has been being used, which uses a modified protocol

of amplified fragment length polymorphism (AFLP). In MSAP, isoschizomers *HpaII* and *MspI* are used at the place of *MseI* (a frequent cutter), while the hexa-cutter *EcoRI* remains the same (Xiong et al., 1999). Though *HpaII* and *MspI* have the same restriction site (5'-CCGG-3'), they show differential sensitivity to 5-mC. This is why the genomic profile generated using MSAP presents the variations in methylation at the enzyme's recognition site only. Another technique being used for quantification of 5-mC at whole-genome level utilizes monoclonal antibody specific to 5-mC. This requires a smaller amount of genomic DNA, provides reliable results at the lower cost and higher throughput compared to those obtained by HPLC based detection of 5-mC. Several efficient methods have been devised to determine 5-mC content at single-base resolution, and most of them take advantage of sodium-bisulfite modification reaction which deaminates cytosine (but not the 5-mC) to uracil (U). This results in the conversion of C to thymine (T) in DNA during the synthesis of the complementary strand.

This allows to distinguishing unmethylated cytosine (C) from 5-mC. Combination of bisulfite conversion and high-throughput DNA sequencing for determining DNA methylation status is referred to as whole-genome bisulfite sequencing (WGBS). Presence of C in the sequencing read typically indicates that the cytosine was 5-mC, as the 5-mC is protected from conversion by the sodium bisulfite treatment. This allows determination of the content, location, and context (CG, CHG, CHH) of 5-mC for better interpretation of the results. This also generates quantitative data for DNA (de)methylation at the individual site because of the multiple independent reads that are aligned with the reference sequence. Although WGBS cannot distinguish 5-mC from 5-hmC, yet there are chemistries to resolve this problem. Moreover, certain variations in bisulfate sequencing often affect the detection/measurement of 5-mC in tissue samples, which may differ even within the same epigenome-wide association studies (EWAS). Variants like bisulfite-conversion efficiency, DNA polymerase efficiency/fidelity, number of cycles of PCR, amount of the template DNA used, bias due to Gibbs law (more CG sites on probe = greater hybridization, regardless of modification), the DNA strand (leading, lagging or both) used to read, the number of reads performed during sequence, etc. affect detection efficiency of the technique. For example, in case of Illumina 450K only 500 ng DNA is required, >100 cycles of amplification are performed using an enzyme with much lower fidelity (compared to that of the high-fidelity enzyme like Phusion DNA Polymerase having 50× higher fidelity than Taq DNA polymerase and 6× lower Pfu DNA Polymerase). All of these cause variations lead to misinterpretation of the results, unable us to compare data from different studies, and reduce reproducibility of data even when the same tissue and technology is used. Therefore, in addition to focusing on epigenomics of single cell type and increasing data sets, methylation data should always be confirmed by more than one assays (e.g., bisulfite-pyrosequencing, MeDIP-qPCR, MS-qPCR, RRBS (reduced representation bisulfate sequencing), etc.), which would provide a more

balanced, comprehensive and critical view of the researchable issue.

Recently, nanopore sequencing has improved the read-length (10–100 kb), sequencing throughput, and more importantly enabled direct detection of DNA base modifications. Nanopore sequencing method has been utilized to detect DNA base modifications (such as 5-mC) based on the sensitivity of electrolytic current signals to base modifications. Simpson et al. (2017) reported quantification of 5-mC using this effect in the Oxford Nanopore Technologies MinION sequencer to sequence the human methylome. Further advances in the nanopore sequencing are expected to revolutionize epigenomic studies through high throughput next-generation sequencing.

Detection of 5-hmC

Classical methods for quantification of 5-hydroxymethylcytosine (5-hmC) have been based on sophisticated instruments such as liquid chromatography, mass spectrometric techniques (LC-MS, HPLC-MS). More recently, quantification of 5-hmC based on immunoassay principles (utilizing antibodies specific to 5-hmC) has been commercialized. Although, these methods are quantitative and reproducible; they are complicated, expensive, lack sensitivity, and they are less suitable for high throughput analysis. Thus, several methods for detection and quantification of 5-hmC do exist, yet none of these techniques meet the requirements of advanced epigenetic studies. Besides, a larger amount of DNA is required when the sample contains a lower level of 5-hmC. Moreover, the detection limit of these techniques is approximately 0.03% 5-hmC per dNTP which further limits their utility. Interestingly, these techniques provide only a relative value for 5-hmC content; hence, for absolute measurement of 5-hmC a calibration curve needs to be generated first.

Another method utilizes phage T₄ β-glucosyltransferase (β-GT) enzyme which catalyzes attachment of β-D-glucosyl residues of uridine diphosphoglucose (UDP-Glu) with the hydroxyl group of 5-hmC. In this method, a reactive azide group is used to label 5-hmC, and the azide group is subsequently tagged with a fluorescent-labeled strained alkyne like dibenzocyclooctyne-Cy5 reporter molecule. Although this is an accurate and high-throughput quantification method for quantification of 5-hmC, a larger amount of DNA (~6 μg) is required for the estimation. Hence, to minimize the amount of DNA sample required for detection of 5-hmC, the ultra-sensitive single-molecule approach was utilized by Gilat et al. (2017).

Detection of 5-fC and 5-caC

Genome-wide analysis of 5-fC in mESCs revealed preferential occurrence of 5-fC at enhancers among the other gene regulatory elements. 5-fC exhibits a preference to the poised enhancers, suggesting a role for 5-fC in epigenetic priming of enhancers (Song et al., 2013). Accumulation of 5-caC at promoter regions of several genes in hepatic cells at differentiation corresponds with the beginning of their expression. Since 5-fC and 5-caC show similar behavior to cytosine in bisulfate sequencing and occur at very low abundance in mammalian genomic DNA

(Ito et al., 2011), their detection becomes a challenging task using antibody-based immunoprecipitation technique. Song et al. (2013) used 5-fC-selective chemical labeling approach and a 5-fC chemically assisted bisulfite sequencing method for genome-wide profiling of 5-fC. Wheldon et al. (2014) and Lewis et al. (2017) reported the use of 5-caC-DNA immunoprecipitation followed by bisulfite sequencing for genome-wide profiling of 5-caC in mESCs.

Detection of 6-mA

Initially, the presence of N⁶-methyladenine (6-mA) in bacteria was confirmed by using a combination of the ultraviolet absorption spectrum, electrophoretic mobility, and paper chromatography. These methods are relatively insensitive, hence the presence of 6-mA in eukaryotes is not detectable. Later on, researchers realized that restriction enzymes can be utilized to identify 6-mA residues. However, limitations of this approach are the dependence of the restriction enzyme on occurrence of 6-mA in the appropriate restriction enzyme target motif. Therefore, this method cannot detect the presence of 6-mA in different sequence contexts. High-performance liquid chromatography (HPLC) was then used to detect 6-mA in *E. coli*.

Subsequently, liquid chromatography and recently ultra-high performance liquid chromatography combined with mass spectrometry (UHPLC-MS/MS) are used to detect 6-mA. While the above-mentioned techniques have been proven useful for detection of the presence of 6-mA, they do not provide any information about its genomic location. When methylated DNA immunoprecipitation (MeDIP) technique is coupled with UHPLC-MS/MS, which can distinguish 1-mA from 6-mA, it can give unambiguous confirmation about the genomic location of 6-mA (Greer et al., 2015). Therefore, to convincingly identify 6-mA a combination of complementary techniques would be essential, as each technique has its own set of limitations. Recently, McIntyre et al. (2017) used nanopore sequencing to detect 6-mA in bacteria deploying a new generation base-caller method (nucleic acid modification caller, mCaller) to detect the base modification.

However, the pattern of DNA modifications are dynamic in nature during cellular differentiation and development stages with introduction and/or erasing of different sets of base modification(s) in the genomic regions specific for the developmental stage. Cellular differentiation appears to be governed by methylation homeostasis in the mammalian genome. However in epigenetic research, the challenge has been in linking specific epigenetic and environmental factors to cellular function/behavior. In fact, most of the research findings discussed in this review is merely based on correlation analysis using mixed cell populations often without any attempt to functionally confirm the linkage between the epigenetic mark and the change in expression of a gene in the cell/tissue. Merely because an epigenetic variation has been found to appear with a change in phenotype, it does not necessarily mean that the phenotypic change is because of the epigenetic action. Generally, the normal cell growth, development and

pathology in plants and animals are based on the multi-dimensional information. Therefore, the need of the day is to integrate epigenetic information with phenotypic change using modern techniques like optogenetics (Guru et al., 2015), sensor-based live imaging (Ingouff et al., 2017) epigenome editing (Liao et al., 2017), etc. While optogenetics uses light to control cells in living tissue to monitor the expression of light-sensitive ion channels in genetically modified cells, the sensor-based live imaging provides insight into global DNA modification dynamics at single-cell level with a high temporal resolution during development stages and in response to environmental stress. On the other hand, epigenome editing aims at targeted *in vivo* editing of the genomic region(s) without altering DNA sequence using the tools like CRISPR–dCas9.

Recently, 6-mA was identified in *Oryza sativa* and *Zea mays* utilizing more sensitive detection methods like HPLC-MS/MS (Huang et al., 2015). DRM2 has been shown to catalyze both cytosine- and adenine-methylation in Arabidopsis, but in wheat N⁶-adenine DNA-methyltransferase was reported to perform this job. The data on 6-mA in plants must be viewed with caution unless it is validated that the detected 6-mA was not from mitochondrial DNA of the plant or from contaminating symbionts. It has been observed that the organisms having a higher content of 6-mA (e.g., bacteria and unicellular eukaryotes) generally have a lower content of 5-mC, while the organisms with higher content of 5-mC (e.g., plants and mammals) possess a lower content of 6-mA (O’Brown and Greer, 2016). If 6-mA is found to be present in a significantly higher quantity in eukaryotes also, it might turn out to be the seventh base and carrier of the epigenetic information in living organisms.

CONCLUSION AND FUTURE PERSPECTIVES

Epigenetic modifications influence the accessibility of the genomic regions for regulatory proteins, which affect chromatin structure, enhancer/promoter activity, and the transcriptional process. Cytosine modifications, particularly methylation, are the known epigenetic changes with established maintenance mechanisms and some of the functions which enable their use as an epigenetic mark (Kumar, 2018a). Cytosine methylation has generally been associated with repression of gene expression when present in promoter regions, but it might also enhance the gene expression either by recruiting methylation-specific transcription factors (Buck-Koehntop and Defossez, 2013) or by yet to be discovered mechanisms (Baubec et al., 2015; Kumar et al., 2017c). The dynamic epigenetic regulation also requires removal of the modified base, which involves DNA demethylases and TET proteins. Though 5-hmC has been identified as an intermediate product of the active demethylation process, and a potential epigenetic mark in metazoan animals, its existence in plant is still controversial. Until recently, 6-mA has been thought to be present in protozoans (bacteria, archaea, and protists); however, its recently discovered existence in eukaryotes, including animals and plants, indicates that 6-mA might serve as

an additional epigenetic mark. In *C. elegans*, this novel epigenetic mark shows a functional interaction with the established histone mark (H3K4me2), and in *D. melanogaster* it supports increasing transposon activity.

Nevertheless, reports on the existence of 6-mA in eukaryotes have been sparse, and often inconclusive (Ratel et al., 2006). Future research needs to focus on investigating the conserved epigenetic marks, the enzymes involved in setting, recognizing and removing these modifications, and their heritable components if any. Interestingly, 6-mA demethylation was observed to be mediated by a TET-like enzyme in *D. melanogaster* (Takayama et al., 2014), which indicates interplay between cytosine and adenine methylation/demethylation. Therefore, further studies should also focus on unraveling the correlation between the known epigenetic marks, which may render clues on their biological significance and evolutionary roles (Kumar, 2018b).

Genome editing technologies, such as CRISPR-Cas9, are continuously being improved for targeted edit purposes. Deactivated endonuclease Cas9 (dCas9) has transformed synthetic biology platforms to achieve desired gene regulation, genome editing and fluorescent labeling (Xu et al., 2016). Discovery of flexibility in sgRNA has created the possibility of inserting additional RNA elements (Koneremann et al., 2015) which can be recognized by RNA-specific binding protein effectors leading to improved efficacy of targeted dCas9-mediated functional moieties (Zalatan et al., 2015). By modulating DNA (de)methylation level at specific sites,

we may explore how (de)methylation at CG sites impact cellular biochemistry, physiology, and pathology. Thus, RNA-guided dCas9 (de)methylation would have versatile and broader applications in basic research as well as therapeutics. This technology might soon be improved to the point that plant epigenomes could also be manipulated, and in the coming years, we may witness epigenome engineered crop plants for certain desirable traits. However, it will be essential to identify the epigenetic mark(s) linked with the trait of interest and understand the epigenetic machinery so that we can deploy epigenetic engineering to solve clinical problems as well as to tailor make the crop plants to cope up with the deleterious effects of global climate changes.

AUTHOR CONTRIBUTIONS

SK and TM conceived the review. SK and VC prepared the manuscript. SK, VC, and TM revised the manuscript and approved the final draft.

FUNDING

The epigenomics research work is being carried out with the financial support from National Agricultural Science Fund (NASF/ABP-70161/2018-1), Indian Council of Agricultural Research, New Delhi, India.

REFERENCES

- Allis, C. D., and Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nat. Rev. Genet.* 17, 487–500. doi: 10.1038/nrg.2016.59
- Avramova, Z. (2015). Transcriptional “memory” of a stress: transient chromatin and memory (epigenetic) marks at stress-response genes. *Plant J.* 83, 149–159. doi: 10.1111/tpj.12832
- Bachman, M., Uribe-Lewis, S., Yang, X., Williams, M., Murrell, A., and Balasubramanian, S. (2014). 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat. Chem.* 6, 1049–1055. doi: 10.1038/nchem.2064
- Baubec, T., Colombo, D. F., Wirbelauer, C., Schmidt, J., Burger, L., Krebs, A. R., et al. (2015). Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520, 243–247. doi: 10.1038/nature14176
- Breiling, A., and Lyko, F. (2015). Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. *Epigenetics Chromatin* 8:24. doi: 10.1186/s13072-015-0016-6
- Buck-Koehntop, B. A., and Defossez, P. A. (2013). On how mammalian transcription factors recognize methylated DNA. *Epigenetics* 8, 131–137. doi: 10.4161/epi.23632
- Charlet, J., Duymich, C. E., Lay, F. D., Mundbjerg, K., Dalsgaard Sørensen, K., Liang, G., et al. (2016). Bivalent regions of cytosine methylation and H3K27 acetylation suggest an active role for DNA methylation at enhancers. *Mol. Cell* 62, 422–431. doi: 10.1016/j.molcel.2016.03.033
- Chinnusamy, V., and Zhu, J.-K. (2009). Epigenetic regulation of stress responses in plants. *Curr. Opin. Plant Biol.* 12, 133–139. doi: 10.1016/j.pbi.2008.12.006
- Cokus, S. J., Feng, S., Zhang, X., Chen, Z., Merriman, B., Haudenschild, C. D., et al. (2008). Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452, 215–219. doi: 10.1038/nature06745
- Dawlaty, M. M., Breiling, A., Le, T., Barrasa, M. I., Raddatz, G., Gao, Q., et al. (2014). Loss of Tet enzymes compromises proper differentiation of embryonic stem cells. *Dev. Cell* 29, 102–111. doi: 10.1016/j.devcel.2014.03.003
- Dawlaty, M. M., Breiling, A., Le, T., Raddatz, G., Barrasa, M. I., Cheng, A. W., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* 24, 310–323. doi: 10.1016/j.devcel.2012.12.015
- Dawlaty, M. M., Ganz, K., Powell, B. E., Hu, Y. C., Markoulaki, S., Cheng, A. W., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* 9, 166–175. doi: 10.1016/j.stem.2011.07.010
- Erdmann, R. M., Souza, A. L., Clish, C. B., and Gehring, M. (2015). 5-Hydroxymethylcytosine is not present in appreciable quantities in Arabidopsis DNA. *Genes Genome Genet.* 5, 1–8. doi: 10.1534/g3.114.014670
- Feinberg, A. P., Koldobskiy, M. A., and Gündör, A. (2016). Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat. Rev. Genet.* 17, 284–299. doi: 10.1038/nrg.2016.13
- Fong, Y. W., Cattoglio, C., and Tjian, R. (2013). The intertwined roles of transcription and repair proteins. *Mol. Cell* 52, 291–302. doi: 10.1016/j.molcel.2013.10.018
- Frost, J. M., Kim, M. Y., Park, G. T., Hsieh, P.-H., Nakamura, M., Lin, S. J. H., et al. (2018). FACT complex is required for DNA demethylation at heterochromatin during reproduction in Arabidopsis. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4720–E4729. doi: 10.1073/pnas.1713333115
- Fu, Y., Jia, G., Pang, X., Wang, R. N., Wang, X., Li, C. J., et al. (2013). FTO-mediated formation of N6-hydroxymethyladenosine and N6-formyladenosine in mammalian RNA. *Nat. Commun.* 4:1798. doi: 10.1038/ncomms2822
- Fu, Y., Luo, G. Z., Chen, K., Deng, X., Yu, M., Han, D., et al. (2015). N(6)-methyldeoxyadenosine marks active transcription start sites in chlamydomonas. *Cell* 161, 879–892. doi: 10.1016/j.cell.2015.04.010

- Gallusci, P., Hodgman, C., Teyssier, E., and Seymour, G. B. (2016). DNA methylation and chromatin regulation during fleshy fruit development and ripening. *Front. Plant Sci.* 7:807. doi: 10.3389/fpls.2016.00807
- Gent, J. I., Ellis, N. A., Guo, L., Harkess, A. E., Yao, Y., Zhang, X., et al. (2013). CHH islands: de novo DNA methylation in near-gene chromatin regulation in maize. *Genome Res.* 23, 628–637. doi: 10.1101/gr.146985.112
- Gilat, N., Tabachnik, T., Shwartz, A., Shahal, T., Torchinsky, D., Michaeli, Y., et al. (2017). Single-molecule quantification of 5-hydroxymethylcytosine for diagnosis of blood and colon cancers. *Clin. Epigenetics* 9:70. doi: 10.1186/s13148-017-0368-9
- Globisch, D., Munzel, M., Muller, M., Michalakos, S., Wagner, M., Koch, S., et al. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS One* 5:e15367. doi: 10.1371/journal.pone.0015367
- Greer, E. L., Blanco, M. A., Gu, L., Sendinc, E., Liu, J., Aristizabal-Corralles, D., et al. (2015). DNA methylation on N-adenine in *C. elegans*. *Cell* 161, 868–878. doi: 10.1016/j.cell.2015.04.005
- Gu, T. P., Guo, F., Yang, H., Wu, H. P., Xu, G. F., Liu, W., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* 477, 606–610. doi: 10.1038/nature10443
- Guru, A., Post, R. J., Ho, Y. Y., and Warden, M. R. (2015). Making sense of optogenetics. *Int. J. Neuropsychopharmacol.* 18:pyv079. doi: 10.1093/ijnp/pyv079
- Hahn, M. A., Qiu, R., Wu, X., Li, A. X., Zhang, H., Wang, J., et al. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. *Cell Rep.* 3, 291–300. doi: 10.1016/j.celrep.2013.01.011
- He, Y. F., Li, B. Z., Li, Z., Liu, P., Wang, Y., Tang, Q., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307. doi: 10.1126/science.1210944
- Hsieh, T. F., Ibarra, C. A., Silva, P., Zemach, A., Eshed-Williams, L., Fischer, R. L., et al. (2009). Genome-wide demethylation of Arabidopsis endosperm. *Science* 324, 1451–1454. doi: 10.1126/science.1172417
- Huang, W., Xiong, J., Yang, Y., Liu, S. M., Yuan, B. F., and Feng, Y. Q. (2015). Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. *R. Soc. Chem. Adv.* 5, 64046–64054. doi: 10.1039/C5RA05307B
- Ingouff, M., Selles, B., Michaud, C., Vu, T. M., Berger, F., Schorn, A. J., et al. (2017). Live-cell analysis of DNA methylation during sexual reproduction in Arabidopsis reveals context and sex-specific dynamics controlled by noncanonical RdDM. *Genes Dev.* 31, 72–83. doi: 10.1101/gad.289397.116
- Ito, S., D'Alessio, A. C., Taranova, O. V., Hong, K., Sowers, L. C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466, 1129–1133. doi: 10.1038/nature09303
- Ito, S., Shen, L., Dai, Q., Wu, S. C., Collins, L. B., Swenberg, J. A., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303. doi: 10.1126/science.1210597
- Iyer, L. M., Zhang, D., and Aravind, D. (2016). Adenine methylation in eukaryotes: apprehending the complex evolutionary history and functional potential of an epigenetic modification. *Bioessays* 38, 27–40. doi: 10.1002/bies.201500104
- Ji, D., Lin, K., Song, J., and Wang, Y. (2014). Effects of Tet-induced oxidation products of 5-methylcytosine on Dnmt1- and DNMT3a-mediated cytosine methylation. *Mol. Biosyst.* 10, 1749–1752. doi: 10.1039/c4mb00150h
- Khare, T., Pai, S., Koncevicius, K., Pal, M., Kriukiene, E., Liutkeviciute, Z., et al. (2012). 5-hmC in the brain is abundant in synaptic genes and shows differences at the exon-intron boundary. *Nat. Struct. Mol. Biol.* 19, 1037–1043. doi: 10.1038/nsmb.2372
- Kigar, S. L., Chang, L., Guerrero, C. R., Sehring, J. R., Cuarenta, A., Parker, L. L., et al. (2017). N6-methyladenine is an epigenetic marker of mammalian early life stress. *Sci. Rep.* 7:e18078. doi: 10.1038/s41598-017-18414-7
- Klungland, A., and Robertson, A. B. (2017). Oxidized C5-methyl cytosine bases in DNA: 5-hydroxymethylcytosine; 5-formylcytosine; and 5-carboxylcytosine. *Free Radic. Biol. Med.* 107, 62–68. doi: 10.1016/j.freeradbiomed.2016.11.038
- Ko, M., Bandukwala, H. S., An, J., Lamperti, E. D., Thompson, E. C., Hastie, R., et al. (2011). Ten-eleven-translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. U.S.A.* 108, 14566–14571. doi: 10.1073/pnas.1112317108
- Koh, K. P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 8, 200–213. doi: 10.1016/j.stem.2011.01.008
- Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517, 583–588. doi: 10.1038/nature14136
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930. doi: 10.1126/science.1169786
- Kumar, S. (2017). Epigenetic control of apomixis: a new perspective of an old enigma. *Adv. Plants Agric. Res.* 7:00243. doi: 10.15406/apar.2017.07.00243
- Kumar, S. (2018a). Epigenetic memory of stress responses in plants. *J. Phytochem. Biochem.* 2:e102.
- Kumar, S. (2018b). Epigenomics of plant responses to environmental stress. *Epigenomes* 2:e6. doi: 10.3390/epigenomes2010006
- Kumar, S., Beena, A. S., Awana, M., and Singh, A. (2017a). Physiological, biochemical, epigenetic and molecular analyses of wheat (*Triticum aestivum*) genotypes with contrasting salt tolerance. *Front. Plant Sci.* 8:120. doi: 10.3389/fpls.2017.01151
- Kumar, S., Beena, A. S., Awana, M., and Singh, A. (2017b). Salt-Induced tissue-specific cytosine methylation downregulates expression of HKT genes in contrasting wheat (*Triticum aestivum* L.) genotypes. *DNA Cell Biol.* 36, 283–294. doi: 10.3389/fpls.2017.01151
- Kumar, S., Singh, A. K., and Mohapatra, T. (2017c). Epigenetics: history, present status and future perspective. *Indian J. Genet. Plant Breed.* 77, 445–463. doi: 10.5958/0975-6906.2017.00061.X
- Kumar, S., and Singh, A. (2016). Epigenetic regulation of abiotic stress tolerance in plants. *Adv. Plants Agric. Res.* 5:00179. doi: 10.15406/apar.2016.05.00179
- Lang, Z., Wang, Y., Tang, K., Tang, D., Datsenka, T., Cheng, J., et al. (2017). Critical roles of DNA demethylation in the activation of ripening-induced genes and inhibition of ripening-repressed genes in tomato fruit. *Proc. Natl. Acad. Sci. U.S.A.* 114, E4511–E4519. doi: 10.1073/pnas.1705233114
- Lauria, M., and Rossi, V. (2011). Epigenetic control of gene regulation in plants. *Biochim. Biophys. Acta* 1809, 369–378. doi: 10.1016/j.bbagr.2011.03.002
- Law, J. A., and Jacobsen, S. E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220. doi: 10.1038/nrg2719
- Lewis, L. C., Lo, P. C., Foster, J. M., Dai, N., Corrèa, I. R., Durczak, P. M., et al. (2017). Dynamics of 5-carboxylcytosine during hepatic differentiation: potential general role for active demethylation by DNA repair in lineage specification. *Epigenetics* 12, 277–286. doi: 10.1080/15592294.2017.1292189
- Li, Y., Kumar, S., and Qian, W. (2018). Active DNA demethylation: mechanism and role in plant development. *Plant Cell Rep.* 37, 77–85. doi: 10.1007/s00299-017-2215-z
- Li, Z., Cai, X., Cai, C. L., Wang, J., Zhang, W., Petersen, B. E., et al. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* 118, 4509–4518. doi: 10.1182/blood-2010-12-325241
- Liao, H. K., Hatanaka, F., Araoka, T., Reddy, P., Wu, M. Z., Sui, Y., et al. (2017). In vivo target gene activation via CRISPR/Cas9-mediated trans-epigenetic modulation. *Cell* 171, 1495–1507. doi: 10.1016/j.cell.2017.10.025
- Lu, F., Liu, Y., Jiang, L., Yamaguchi, S., and Zhang, Y. (2014). Role of Tet proteins in enhancer activity and telomere elongation. *Genes Dev.* 28, 2103–2119. doi: 10.1101/gad.248005.114
- Maiti, A., and Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286, 35334–35338. doi: 10.1074/jbc.M111.284620
- McIntyre, A. B. R., Alexander, N., Burton, A. S., Castro-Wallace, S., Chiu, C. Y., John, K. K., et al. (2017). Nanopore detection of bacterial DNA base modifications. *bioRxiv* [Preprint]. doi: 10.1101/127100
- Moricová, P., Ondřej, V., Navrátilová, B., and Luhová, L. (2013). Changes of DNA methylation and hydroxymethylation in plant protoplast cultures. *Acta Biochim. Pol.* 60, 33–36.

- O'Brown, Z. K., and Greer, E. L. (2016). N6-methyladenine: a conserved and dynamic DNA mark. *Adv. Exp. Med. Biol.* 945, 213–246. doi: 10.1007/978-3-319-43624-1_10
- Pastor, W. A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* 14, 341–356. doi: 10.1038/nrm3589
- Pfaffeneder, T., Spada, F., Wagner, M., Brandmayr, C., Laube, S. K., Eisen, D., et al. (2014). Tet oxidizes thymine to 5-hydroxymethyluracil in mouse embryonic stem cell DNA. *Nat. Chem. Biol.* 10, 574–581. doi: 10.1038/nchembio.1532
- Pikaard, C. S., and Mittelsten-Scheid, O. (2014). Epigenetic regulation in plants. *Cold Spring Harb. Perspect. Biol.* 6:a019315. doi: 10.1101/cshperspect.a019315
- Raiber, E. A., Beraldi, D., Ficiz, G., Burgess, H. E., Branco, M. R., Murat, P., et al. (2012). Genome-wide distribution of 5-formylcytosine in embryonic stem cells is associated with transcription and depends on thymine DNA glycosylase. *Genome Biol.* 13:R69. doi: 10.1186/gb-2012-13-8-r69
- Ratel, D., Ravanat, J. L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *Bioessays* 28, 309–315. doi: 10.1002/bies.20342
- Sedgwick, B., Bates, P. A., Paik, J., Jacobs, S. C., and Lindahl, T. (2007). Repair of alkylated DNA: recent advances. *DNA Repair* 6, 429–442. doi: 10.1016/j.dnarep.2006.10.005
- Shi, D.-Q., Ali, I., Tang, J., and Yang, W.-C. (2017). New insights into 5hmC DNA modification: generation, distribution and function. *Front. Genet.* 8:100. doi: 10.3389/fgene.2017.00100
- Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nat. Methods* 14, 407–410. doi: 10.1038/nmeth.4184
- Sloan, K. E., Warda, A. S., Sharma, S., Entian, K.-D., Lafontaine, D. L. J., and Bohnsack, M. T. (2017). Tuning the ribosome: the influence of rRNA modification on eukaryotic ribosome biogenesis and function. *RNA Biol.* 14, 1138–1152. doi: 10.1080/15476286.2016.1259781
- Song, C. X., Szulwach, K. E., Dai, Q., Fu, Y., Mao, S. Q., Lin, L., et al. (2013). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* 153, 678–691. doi: 10.1016/j.cell.2013.04.001
- Springer, N. M., and Schmitz, R. J. (2017). Exploiting induced and natural epigenetic variation for crop improvement. *Nat. Rev. Genet.* 18, 563–575. doi: 10.1038/nrg.2017.45
- Spruijt, C. G., and Vermeulen, M. (2014). DNA methylation: old dog, new tricks? *Nat. Struct. Mol. Biol.* 21, 949–954. doi: 10.1038/nsmb.2910
- Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935. doi: 10.1126/science.1170116
- Takayama, S., Dhahbi, J., Roberts, A., Mao, G., Heo, S. J., Pachter, L., et al. (2014). Genome methylation in *D. melanogaster* is found at specific short motifs and is independent of DNMT2 activity. *Genome Res.* 24, 821–830. doi: 10.1101/gr.162412.113
- Terragni, J., Bitinaite, J., Zheng, Y., and Pradhan, S. (2012). Biochemical characterization of recombinant b-glucosyltransferase and analysis of global 5-hydroxymethylcytosine in unique genomes. *Biochemistry* 51, 1009–1019. doi: 10.1021/bi2014739
- Tollervy, J. R., and Lunyak, V. V. (2012). Epigenetics: judge, jury and executioner of stem cell fate. *Epigenetics* 7, 823–840. doi: 10.4161/epi.21141
- Torres, I. O., and Fujimori, D. G. (2015). Functional coupling between writers, erasers and readers of histone and DNA methylation. *Curr. Opin. Struct. Biol.* 35, 68–75. doi: 10.1016/j.sbi.2015.09.007
- Wang, X., Li, Q., Yuan, W., Cao, Z., Qi, B., Kumar, S., et al. (2016). The cytosolic Fe-S cluster assembly component MET18 is required for the full enzymatic activity of ROS1 in active DNA demethylation. *Sci. Rep.* 6:26443. doi: 10.1038/srep26443
- Wang, X. L., Song, S. H., Wu, Y. S., Li, Y. L., Chen, T. T., Huang, Z. Y., et al. (2015). Genome-wide mapping of 5-hydroxymethylcytosine in three rice cultivars reveals its preferential localization in transcriptionally silent transposable element genes. *J. Exp. Bot.* 66, 6651–6663. doi: 10.1093/jxb/erv372
- Westbye, M. P., Feyzl, E., Aas, P. A., Vagbø, C. B., Talstad, V. A., Kavll, B., et al. (2008). Human AlkB homolog I is a mitochondrial protein that demethylates 3-methylcytosine in DNA and RNA. *J. Biol. Chem.* 283, 25046–25056. doi: 10.1074/jbc.M803776200
- Wheldon, L. M., Abakir, A., Ferjentsik, Z., Dudnakova, T., Strohbuecker, S., Christie, D., et al. (2014). Transient accumulation of 5-carboxylcytosine indicates involvement of active demethylation in lineage specification of neural stem cells. *Cell Rep.* 7, 1353–1361. doi: 10.1016/j.celrep.2014.05.003
- Wu, H., and Zhang, Y. (2011). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* 25, 2436–2452. doi: 10.1101/gad.179184.111
- Xiang, H., Zhu, J., Chen, Q., Dai, F., Li, X., Li, M., et al. (2010). Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat. Biotechnol.* 28, 516–520. doi: 10.1038/nbt.1626
- Xiong, L. Z., Xu, C. G., Saghai-Marouf, M. A., and Zhang, Q. (1999). Patterns of cytosine methylation in an elite rice hybrid and its parental lines, detected by a methylation-sensitive amplification polymorphism technique. *Mol. Gen. Genet.* 261, 439–446. doi: 10.1007/s004380050986
- Xu, X., Tao, Y., Gao, X., Zhang, L., Li, X., Zou, W., et al. (2016). A CRISPR-based approach for targeted DNA demethylation. *Cell Discov.* 2:1600. doi: 10.1038/celldisc.2016.9
- Yao, B., Cheng, Y., Wang, Z., Li, Y., Chen, L., Huang, L., et al. (2017). DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nat. Commun.* 8:1122. doi: 10.1038/s41467-017-01195-y
- Yao, B., Li, Y., Wang, Z., Chen, L., Poidevin, M., Zhang, C., et al. (2018). Active N6-methyladenine demethylation by DMAD regulates gene expression by coordinating with polycomb protein in neurons. *Mol. Cell* 6, 848–857. doi: 10.1016/j.molcel.2018.07.005
- Yue, X., Trifari, S., Åijö, T., Tsagaratou, A., Pastor, W. A., Zepeda-Martinez, J. A., et al. (2016). Control of Foxp3 stability through modulation of TET activity. *J. Exp. Med.* 213, 377–397. doi: 10.1084/jem.20151438
- Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., LaRussa, M., et al. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* 160, 339–350. doi: 10.1016/j.cell.2014.11.052
- Zemach, A., Kim, M. Y., Hsieh, P. H., Coleman-Derr, D., Eshed-Williams, L., Thao, K., et al. (2013). The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* 153, 193–205. doi: 10.1016/j.cell.2013.02.033
- Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., et al. (2015). N(6)-methyladenine DNA modification in *Drosophila*. *Cell* 161, 893–906. doi: 10.1016/j.cell.2015.04.018
- Zhou, T., Xiong, J., Wang, M., Yang, N., Wong, J., Zhu, B., et al. (2014). Structural basis for hydroxymethylcytosine recognition by the SRA domain of UHRF2. *Mol. Cell* 54, 879–886. doi: 10.1016/j.molcel.2014.04.003
- Zhu, J.-K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* 43, 143–166. doi: 10.1146/annurev-genet-102108-134205

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kumar, Chinnusamy and Mohapatra. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

APPENDIX

Glossary

Active DNA Demethylation

It involves removal of methylcytosine (5-mC) and replacement with cytosine (C) using enzymatic machinery.

Asymmetric Context

When a cytosine (C) occurs in a sequence context (e.g., CHH, where H = A, C or T) which is not the same in the complementary strand, then it is named as asymmetric context.

Base Excision Repair

This is the mechanism which repairs DNA damages caused due to alkylation, deamination, and oxidation. DNA glycosylase initiates the process by recognizing and removing the damaged base, leaving an abasic site, which is subsequently repaired by another enzyme complex.

Chromatin

It is constituted of packaged nucleosomes (DNA wrapped around a bundle of histone proteins) and plays a vital role in the regulation of gene expression.

CRISPR-dCas9

The bacterial defense system based clustered regularly interspaced short palindromic repeats (CRISPR) along with the deactivated CRISPR-associated protein 9 (dCas9) constitute one of the advanced (epi)genome editing tools.

DNA Demethylation

This is the process of removal of methyl group from a nucleotide base, which may occur either through the active or passive mechanism.

DNA Methyltransferase

It is a family of enzymes involved in catalyzing the transfer of methyl (CH₃) group to DNA base using S-adenosyl methionine (SAM) as a methyl group donor.

Epigenetics

This is the study of functionally relevant modifications (affecting gene expression) in the genome without any change in the nucleotide sequence.

Epigenetic Code

It is hypothesized to be the combinations of epigenetic modification(s), including DNA and histone modifications, which affect the expression of the gene.

Epigenetic Eraser

A group of enzymes that catalyze the removal of an epigenetic mark.

Epigenetic Mark

A feature not directly related to the change in genetic information (nucleotide sequence), but depends on the epigenetic modifications (DNA base and histone modifications) affecting gene expression.

Epigenetic Reader

A protein or protein complex that recognizes and binds to the epigenetic modification to affect the expression of the gene.

Epigenetic Writer

The enzyme or enzyme complex that catalyzes and establishes a particular epigenetic modification.

Epigenome

The sum total of chemical modifications in DNA bases (without affecting the nucleotide sequence), histone proteins and biogenesis of non-coding RNAs in a cell constitutes the epigenome.

Euchromatin

The loosely packaged, gene enriched part of the chromatin that is actively transcribed as per the needs is called euchromatin.

Gene Body Methylation

It refers to methylation of DNA base(s) present in the transcribed/coding region of the gene.

Genomic Imprinting

Generally the genes inherited from both the parents are equally active/expressed, but due to genomic imprinting (an epigenetic phenomenon) set of genes inherited from one of the parents are exclusively expressed.

H3K4me2

Indicates that the fourth lysine (K) residue of histone protein H3 is dimethylated (me₂).

Heterochromatin

Heterochromatin is a tightly packaged form of chromatin, contains repetitive DNA sequences, devoid of a functional gene, and a rarely transcribed portion of the genome.

Histone Proteins

Highly alkaline proteins found in the nucleus of a eukaryotic cell that helps to package DNA into structural units nucleosomes.

Histone Modification

It includes post-translational modifications such as acetylation, methylation, phosphorylation, sumoylation, and ubiquitylation of particular amino acid of specific histone proteins.

Methylome

The sum total of methylated DNA bases in the genome of a particular cell/tissue under specified conditions or the stage of development.

Nucleosome

It is the preliminary structure in the packaging of DNA in eukaryotes, wherein 146 base pairs of DNA is wound around a bundle of eight histone proteins.

Passive DNA Demethylation

It refers to the removal of 5-methylcytosine (5-mC) from DNA during the process of replication, particularly in absence of the maintenance methyltransferases.

Post-translational Modification

Modification of protein or its amino acid which occurs after translation of the protein.

Symmetric Context

When cytosine (C) occurs in a context (e.g., CG and CHG, where H = A, C or T) having the same reading sequence on the antiparallel strands of DNA, it is called symmetric context.

Whole-Genome Bisulfite Sequencing

It is a next-generation sequencing technique used to determine DNA methylation (5-mC) status at single-base resolution after treating the genomic DNA with sodium-bisulfate [that converts unmethylated cytosine (C) into uracil (U)] followed by whole genome sequencing.