



New Insights into 5hmC DNA Modification: Generation, Distribution and Function

Dong-Qiao Shi*, Iftikhar Ali, Jun Tang and Wei-Cai Yang*

State Key Laboratory of Molecular Developmental Biology, National Center for Plant Gene Research (Beijing), Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, University of Chinese Academy of Sciences, Beijing, China

OPEN ACCESS

Edited by:

Trygve Tollefsbol,
University of Alabama at Birmingham,
United States

Reviewed by:

Alexey Ruzov,
University of Nottingham,
United Kingdom
Xiaojing Yue,
La Jolla Institute for Allergy
and Immunology, United States

*Correspondence:

Dong-Qiao Shi
dqshi@genetics.ac.cn
Wei-Cai Yang
wcyang@genetics.ac.cn

Specialty section:

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

Received: 01 April 2017

Accepted: 05 July 2017

Published: 19 July 2017

Citation:

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

Dynamic DNA modifications, such as methylation/demethylation on cytosine, are major epigenetic mechanisms to modulate gene expression in both eukaryotes and prokaryotes. In addition to the common methylation on the 5th position of the pyrimidine ring of cytosine (5mC), other types of modifications at the same position, such as 5-hydroxymethyl (5hmC), 5-formyl (5fC), and 5-carboxyl (5caC), are also important. Recently, 5hmC, a product of 5mC demethylation by the Ten-Eleven Translocation family proteins, was shown to regulate many cellular and developmental processes, including the pluripotency of embryonic stem cells, neuron development, and tumorigenesis in mammals. Here, we review recent advances on the generation, distribution, and function of 5hmC modification in mammals and discuss its potential roles in plants.

Keywords: 5-hydroxymethylcytosine, DNA hydroxylation, DNA demethylation, TET proteins, epigenetics

INTRODUCTION

From two decades of studies, it is clear that the primary sequence information of DNA can be enhanced by epigenetic modifications. There can be various epigenomes in different cell types, although there is only one genome of an organism. Since the discovery of the first modified base, 5-methylcytosine (5mC, methylation of cytosine at the 5-carbon position) in calf thymus DNA by Hotchkiss (1948), many studies showed that DNA methylation regulates gene expression. As the most common DNA modification in eukaryotes, 5mC is sometimes regarded as the fifth genetic code, in addition to the canonical “A, G, C, T” in DNA. Of the approximately 28 million CpG sites present in the human genome, 60–80% of the cytosines are methylated as 5mC (Smith and Meissner, 2013). In angiosperms, the genome-wide methylation level (GML) is as high as 43% in *Beta vulgaris* (with mean of 16%) (Alonso et al., 2015; Niederhuth et al., 2016; Vidalis et al., 2016). Increasing data suggest that the epigenetic information of the epigenome is maintained and translated by the dynamic activity of DNA methylases (writers) and demethylases (erasers) and reader proteins who recognize and interpret the information in both mammals and plants (He X.J. et al., 2011; Torres and Fujimori, 2015; Feinberg et al., 2016). However, in some organisms, such as yeast, *Caenorhabditis elegans*, *Drosophila melanogaster* and many other invertebrates, either no or only trace amounts of methylated cytosine are found (Wenzel et al., 2011; Capuano et al., 2014).

Other DNA modifications, such as 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC), 5-carboxylcytosine (5caC), and N^6 -methyladenine (6mA) are found to play critical roles in many biological processes (Wion and Casadesus, 2006; Kriaucionis and Heintz, 2009; Tahiliani et al., 2009; He Y.F. et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011; Fu and He, 2012; Liu et al., 2016) (Figure 1).

The methyl moiety of 5mC can be lost or eliminated in a cell type-specific manner. This loss can be achieved passively during DNA replication, or actively through enzymatic DNA demethylation. During active DNA demethylation, 5mC is oxidized to produce 5hmC, 5fC, and further 5caC in a replication-independent manner; subsequently, the DNA base-excision repair (BER) pathway can also remove the methylated cytosine by filling in an unmodified cytosine. In addition, passive elimination of 5mC is likely enhanced by active DNA demethylation, as DNMT1 (DNA methyltransferase 1) activity can be depleted by DNA with 5hmC *in vitro* (Valinluck and Sowers, 2007; Ji et al., 2014). More and more studies indicate that 5hmC acts not only as an intermediate during 5mC demethylation, but also plays important roles during maintenance of pluripotency in embryo stem cells (ESCs), neural system development and tumorigenesis (Thomson and Meehan, 2017). However, many questions remain to be answered. How DNA is actively demethylated? Where does 5hmC exist, and what roles does 5hmC have during development? These questions are the focus of this review.

5hmC, A RETROSPECTIVE VIEW

Shortly after the discovery of 5mC, the identification of 5hmC was published in *Nature* in 1952 (Wyatt and Cohen, 1952). Wyatt and Cohen (1953) showed that 5hmC exists as “a major constituent of the nucleic acid of a virus”, instead of cytosine or 5mC, in bacteriophages *T2*, *T4*, and *T6* of *Escherichia coli*. The 5hmC modification of the viral DNA makes it different from host DNA and thereby avoids degradation by the host (Wyatt and Cohen, 1953; Moréra et al., 1999). β -glucosyltransferase of bacteriophages can catalyze the transfer of glucose (Glc) from uridine diphosphoglucose (UDP-Glc) to 5hmC in double-stranded DNA (dsDNA), subsequently the glucosylation protects the infecting viral DNA from host restriction enzymes (Moréra et al., 1999). In 1972, 5hmC was also found in vertebrates, including rat, mouse, and frog (Penn et al., 1972), but 5hmC was only poorly understood as an epigenetic mark on DNA.

Further work on 5hmC had not attracted much attention until two important publications in *Science* in 2009 (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Kriaucionis and Heintz (2009) reported that 5hmC comprises 0.6 and 0.2% nucleotides in Purkinje cells and granule cells, respectively; soon after 5hmC was detected in all tissues and cell types in mouse, with the highest levels (0.3–0.7%) found in the central nervous system (CNS) (Kriaucionis and Heintz, 2009; Globisch et al., 2010). Another breakthrough on 5hmC was the identification of TET1 (Ten-Eleven Translocation 1), an enzyme that oxidizes 5mC to

5hmC (Tahiliani et al., 2009). These new findings undoubtedly revitalized interest on 5hmC function, so much so that now 5hmC is accepted as the sixth DNA base in mammalian genomic DNA.

HOW IS 5hmC PRODUCED AND ELIMINATED?

TET Proteins and 5hmC

Inspired by the biosynthesis of an unusual modified thymine J (β -D-glucosyl-hydroxymethyluracil) in Trypanosomes, Tahiliani et al. (2009) identified homologs of J-binding protein 1 (JBP1) and JBP2, two thymine (T) hydroxylases of the 2-oxoglutarate (2OG)- and Fe(II)-dependent oxygenase superfamily in mammals (Yu et al., 2007; Cliffe et al., 2009; Tahiliani et al., 2009). Human TET1, TET2, and TET3, as well as other orthologs from metazoa, fungi, and algae, were soon identified when databases were queried with the predicted oxygenase domains of JBP1 and JBP2 (Tahiliani et al., 2009). *TET1* was reported to fuse with the *MLL* gene to produce a chimeric transcript in acute myeloid leukemia (AML) (Ono et al., 2002). TET1 is directly responsible for the production and abundance of 5hmC. The oxidation of 5mC to 5hmC is executed by the predicted catalytic domain of TET1 (TET1-CD) in a Fe(II)- and 2OG-dependent manner (Tahiliani et al., 2009). Three TET proteins, TET1, TET2, and TET3, all exhibit oxidation activity to convert 5mC to 5hmC (Ito et al., 2010). In mouse embryonic stem cells (mESCs), the 5hmC level is decreased at the promoter/TSS (transcription start site) when TET1 is depleted, whereas TET2 is predominantly associated with 5hmC in gene bodies and at boundaries of actively expressed exons (Huang et al., 2014).

In addition, *TET* genes show cell/organ-specific expression. *TET1* is preferentially expressed in ESCs, while *TET2* and *TET3* exhibit similar expression profiles in various tissues (Ito et al., 2011; Nestor et al., 2012; Tollervy and Lunyak, 2012). *TET3* is abundantly expressed in the cerebellum, cortex, and hippocampus. TET1 and TET2 can catalyze the oxidation of 5mC and 5hmC in DNA (Gu et al., 2011; He Y.F. et al., 2011). In HEK293 cells transfected with *TET2*, 90% of the 5mC or 5hmC can be converted to 5caC. Using a two dimensional TLC (2D-TLC) assay, Ito et al. (2011) demonstrated that TET proteins have enzymatic activity able to oxidize 5mC to 5hmC, 5fC and 5caC *in vitro*. TET3 contributes to 5hmC production during gametogenesis and embryogenesis (Gu et al., 2011). Sperm and eggs possess distinct epigenomes that are subjected to reprogramming after fertilization. *TET3* is specifically expressed at the male pronuclear stage, and consistent with this, oxidation of 5mC occurs in the paternal genome and results in the accumulation of 5hmC (Gu et al., 2011; Amouroux et al., 2016). Furthermore, 5hmC generation in zygotes is also dependent on Dnmt3a and Dnmt1, two enzymes that produce 5mC. The fresh 5hmC comes from *de novo* oxidation of 5mC in zygotes (Amouroux et al., 2016).

Protein Lin28A regulates production of 5hmC by recruiting TET proteins during the conversion of 5mC to

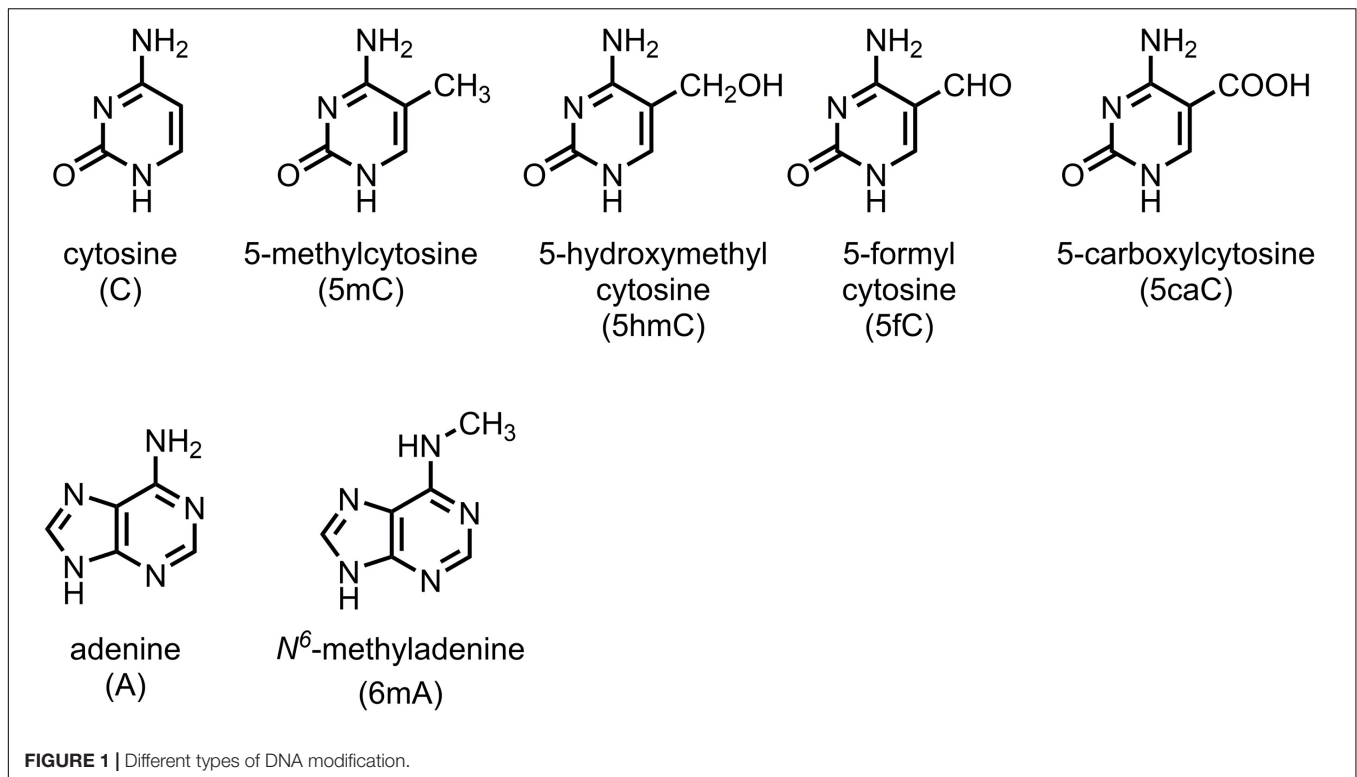


FIGURE 1 | Different types of DNA modification.

5hmC. Lin28 was first discovered in *C. elegans* as an RNA binding protein. However, Zeng et al. (2016) reported that Lin28A, a paralog of Lin28 that is preferentially expressed in mouse embryos and ESCs before differentiation, can recruit TET1 to DNA and facilitate the conversion of 5mC to 5hmC and demethylation of gene bodies (Tan and Yeo, 2016).

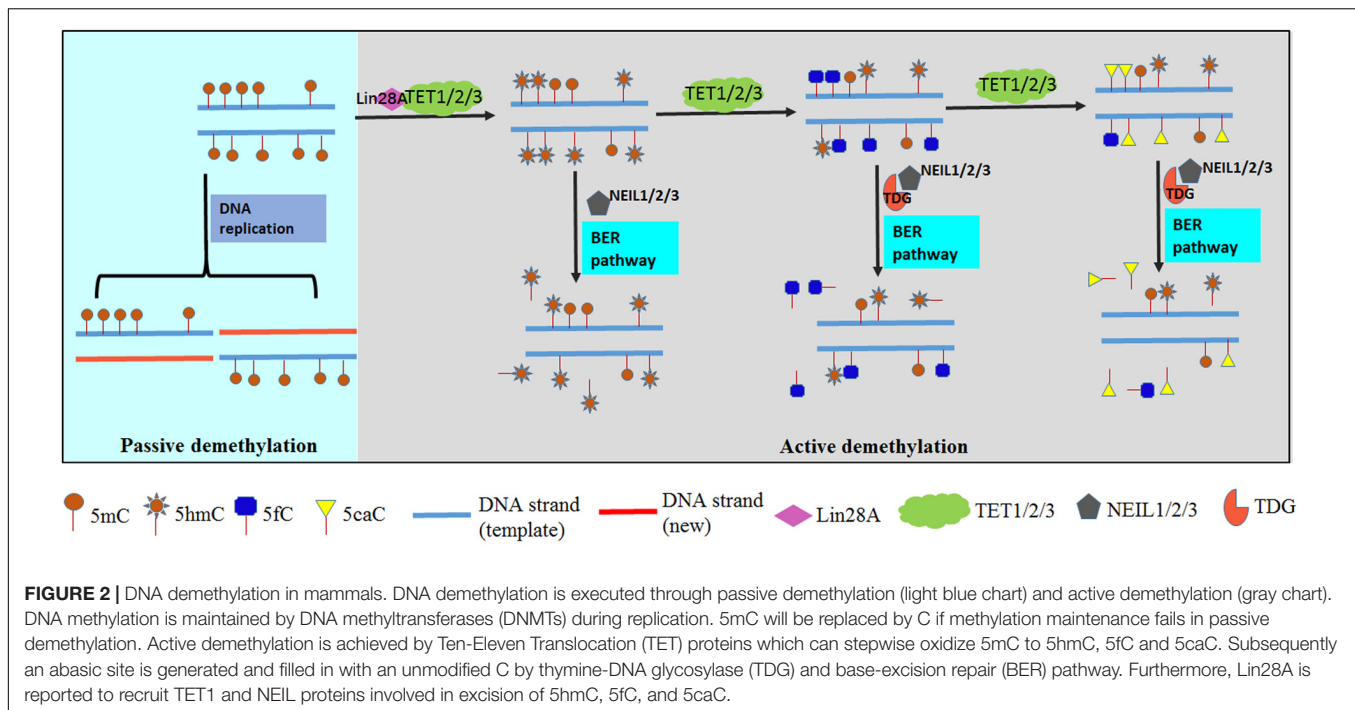
Excision of 5fC and 5caC

5hmC, 5fC, and 5caC show different abundances and tissue specificities. 5hmC is 10- to 100-fold more prevalent than 5fC/5caC and it is relatively enriched in neurons, stem cells, and much decreased in cancer cells (Tahiliani et al., 2009; Globisch et al., 2010; Ito et al., 2011; Song et al., 2011; Mellén et al., 2012). These differences are due to different preference and activity of TET proteins for these bases. Both human TET1 and TET2 show higher activity for 5mC than for 5hmC/5fC. Once 5hmC is established in genomic DNA, it is not easy to oxidize it to 5fC and 5caC (Hu et al., 2015). However, TET3 acts as 5caC reader during 5caC excision in BER, since its CXXC domain shows high affinity for 5caC (Jin et al., 2016).

DNA repair proteins for DNA damage response play key roles in active DNA demethylation (Wossidlo et al., 2010; Spruijt et al., 2013). The distribution of 5fC can be regulated through TET-mediated oxidation and excision by thymine-DNA glycosylase (TDG). TDG is a DNA repair enzyme that excises T from G:T mismatch and BER of deaminated 5mC or 5hmC. TDG is active on both 5fC and 5caC, but not 5mC or 5hmC (He Y.F. et al.,

2011; Maiti and Drohat, 2011). The dynamic methylation status of 5mC is maintained by the antagonistic action of TET and TDG. On one hand, 5mC is stepwise oxidized by TET proteins to produce 5hmC, 5fC, and 5caC. On the other hand, TDG acts on 5fC and 5caC as a glycosylase to regenerate unmodified cytosine (C) through the BER pathway (Figure 2). This iterative oxidation is the main way active demethylation of 5mC is accomplished in mammals (Figure 2). Besides TDG, there are three orthologs of the prototypical endonuclease VIII (Nei). The Nei-like NEIL1-3 enzymes are alternative DNA glycosylases to TDG (Prakash et al., 2014). NEIL1, 2, and 3 can partially rescue the loss of TDG (Müller et al., 2014). Schomacher et al. (2016) reported that NEIL1 and NEIL2 promote TDG-mediated excision of 5fC and 5caC. It is likely that TDG first hydrolyses 5fC or 5caC, then TDG is displaced by NEIL1 and NEIL2, and subsequently the DNA backbone is cut in the BER pathway (Spruijt et al., 2013).

It is also speculated that 5hmC is alternatively deaminated by AID (Activation-induced deaminase) (Hendrich et al., 1999), to generate 5hmU and then to BER to complete 5mC demethylation. However, as AID cannot work on dsDNA, and the rate of deamination of 5hmC in single-stranded or dsDNA is also very low, the spontaneous deamination of 5hmC to 5hmU may not be able to compete with the oxidation step from 5hmC to 5fC. Furthermore, no detectable deamination of 5hmC has been found, even for cells overexpressing AID/APOBEC (Nabel et al., 2012). It is still not clear whether 5hmC can proceed with the deamination reaction. In this case, the development of sensitive methods for 5hmU detection and demonstration of 5hmC deamination in both *in vivo* and *in vitro* will be



needed to answer the question (Grand et al., 2014; Lu et al., 2015).

WHERE ARE 5hmC BASES?

Genome Distribution of 5hmC and Gene Activity

5hmC is generated from 5mC by TETs, however, the TET protein level does not always reflect the existence of 5hmC. There may be post-transcriptional regulation or interaction partners of TETs, since 5hmC abundance is remarkably higher during neuronal development in mouse, even though there is no significant increase in *TET* family gene expression (Szulwach et al., 2011b).

First, to what extent can the TET proteins transfer 5mC to 5hmC? 5mC often occurs in pairs at CpG in reversely complemented DNA strands. Can both 5mCs be converted to two 5hmCs? Song et al. (2016) developed a technology, single-molecule fluorescence resonance energy transfer (smFRET), to image and quantify 5mC and 5hmC simultaneously (Branco et al., 2012). In this method, DNA is end-labeled with biotin and immobilized on a slide, then 5hmC and 5mC are labeled with Cy5 and Cy3 and act as donor (Cy5-5hmC) and receptor (Cy3-5mC), respectively. The hemihydroxymethylated/hemimethylated CpG sites (5hmC/5mCpGs) correspond to a high-FRET (~ 0.78) state. By calculating the FRET, it was estimated that 5hmC/5mCpGs account for about 60% of 5hmCs (Song et al., 2016). This suggests that 5hmC, apart from CpG sites, also occurs in other parts of the genome.

5hmC is associated with gene transcription or translation. More 5hmC is found in gene bodies of active genes, and TET1

is often observed at TSS of genes with high CpG promoters that are marked by the bivalent histone signature of H3K27me3 and H3K4me3. It is therefore assumed that 5hmC and TET proteins may regulate gene expression through modulating chromatin accessibility of the transcriptional machinery, or by inhibiting repressor binding. This is consistent with 5hmC enrichment within gene bodies, promoters, and transcription factor (TF)-binding regions (Nestor et al., 2012). Immunostaining also showed that 5hmC often accumulated in gene-rich regions marked by H3K4me2/3 (Wion and Casadesus, 2006; Ficiz et al., 2011; Szulwach et al., 2011a; Branco et al., 2012; Kubiura et al., 2012). The relationship between 5hmC peaks and gene expression level is complicated. For example, genes with active transcription show depleted 5hmC in their TSS regions, while genes of low expression show abundant 5hmC at promoters in both ESCs and NPCs (neural progenitor cells). However, in gene bodies, 5hmC peaks are positively correlated with gene expression levels in ESCs, but with lower expression levels in NPCs. Consistently, 5hmC distribution is vastly different in ESCs and NPCs. For example, the majority of 5hmC peaks of ESCs are lost in NPCs, yet some *de novo* DNA hydroxymethylation occurs at gene loci associated with mature neuronal functions (Tan et al., 2013). Apart from the different distribution of 5hmC on gene models, the global 5hmC level of NPCs is much lower than that of ESCs, which suggests that differentiation of ESCs to NPCs requires a genomic reduction of 5hmC (Tan et al., 2013). Consistently, *TET1* and *TET2* are highly expressed in mESCs, and 5hmC is enriched in mouse and human ESCs. Enrichment of 5hmC at the promoter/TSS is most likely associated with TET1, while the 5hmC level in gene bodies and boundaries of exons in active genes is correlated with TET2 activity (Huang et al., 2014).

Different studies about 5hmC abundance and gene expression are inconsistent. It seems that this is rather a cell type-dependent regulatory network than simple activation or repression of gene activity. Differentially hydroxymethylated regions (DHMRs) were also uncovered in mouse ESCs and NPCs (Tan et al., 2013). The tissue type is likely a main modifier for 5hmC, although 5hmC is accumulated in actively transcribed genes, for a gene in similar transcribed level, there may be 20-fold change of 5hmC on the gene body in different tissues (Nestor et al., 2012).

5hmC may be involved in regulating gene expression by association with various regulatory elements and processes (Szulwach et al., 2011a,b). Distribution of 5hmC is affected by histone modification, binding proteins of epigenetic marks, and chromatin configuration during cell differentiation and specification. 5hmC accumulated at the TSSs of genes whose promoters are decorated with dual histone marks, H3K27me3 for transcription repression, and H3K4me3 for transcription activation (Pastor et al., 2011), and at “poised” and active enhancers labeled with H3K4me1, H3K18ac, and H3K27ac (Szulwach et al., 2011a,b). Genes involved in development have “bivalent domains” (Bernstein et al., 2006) in promoters. In pluripotent ESCs, the “bivalent domains” can poise genes with both activating (H3K4me3) and repressive (H3K27me3) marks. The TET-mediated generation and distribution of 5hmC, global 5hmC/5mC change, and reprogramming of *de novo* “bivalent histone code” in CGI (CpG island) promoters contribute to bivalent domains directly (Hammoud et al., 2014; Grosser et al., 2015; Kong et al., 2016). TET1 forms a complex with Polycomb repressive complex 2 (PRC2) at H3K27me3 positive sections in the “bivalent” regions (Neri et al., 2013). 5hmC is abundant in both repressed (bivalent, TET1/PRC2-cobound) and activated (TET1-only) genes (Wu et al., 2011). This suggests that 5hmC plays a role in the machinery responsible for pluripotency switch.

The 5hmC level was changed in LINE (long tandem repeat), SINE (short interspersed nuclear element), and other repetitive elements in umbilical cord DNA from gestational diabetes mellitus (GDM) or preeclampsia patients (Sun et al., 2016). It is unclear whether 5hmC plays a role in regulating transposon activity.

As a reader for both 5mC and 5hmC (Mellén et al., 2012; Spruijt et al., 2013; Diotel et al., 2017), MeCP2 (methyl-CpG-binding protein 2) has similar affinity to 5hmC and 5mC (Mellén et al., 2012). However, the 5hmC level is negatively correlated with MeCP2 abundance. The binding of MeCP2 with 5mC may possibly hinder the production of 5hmC (Szulwach et al., 2011a). 5hmC accumulates in MeCP2- and H3K4me2-enriched nuclear foci. Thus 5hmC and MeCP2 might constitute a cell-specific epigenetic mechanism for the regulation of chromatin structure and gene expression.

Cell Type-Specific and Developmentally Related Abundance of 5hmC

Global detection of 5hmC has been carried out in human, mouse, zebrafish, and *Xenopus* (Globisch et al., 2010; Munzel et al., 2010; Chen and Riggs, 2011; Jin et al., 2011; Song et al., 2011; Szulwach et al., 2011a,b; Almeida et al., 2012; Han et al., 2016; Diotel et al.,

2017). These studies showed that the dynamics and abundance of 5hmC are cell type-dependent and developmentally regulated. It is interesting that in amphibians, high levels of 5hmC were detected in *Xenopus* spinal cord and axolotl neural tube cells, or distributed in amphibian skin and connective tissue in a mosaic manner (Almeida et al., 2012). Significant increases of Sertoli cell-specific global 5hmC were detected during rat puberty. From juvenile to adult rat, the functions of Sertoli cells are different, and genes which lose or gain 5hmC belong to different functional pathways (Landfors et al., 2017).

Recently, Yu et al. (2012) adapted traditional bisulfite sequencing to map 5hmC globally. TET-Assisted Bisulfite Sequencing (TAB-Seq) showed that 5hmC was 10-fold more abundant in CNS and ESCs than in peripheral tissues (Branco et al., 2012). The distribution of 5hmC is somewhat negatively correlated with 5mC along genomic DNA in ESCs. Its frequency is biased with DNA sequence and strand asymmetry in a tissue-dependent manner. In mouse, for example, 5hmC is developmentally enriched and age-dependent in brain. Furthermore, genes that have acquired 5hmC during aging are associated with age-related neurodegenerative disorders (Song et al., 2011; Szulwach et al., 2011a). 5hmC acquisition occurs in developmentally programmed neuronal cells, since very low or no 5hmC was detected in the immature neurons. However, a significant increase of 5hmC was found from postnatal day 7 (P7) to adult in both cerebellum and hippocampus (Szulwach et al., 2011a). This leads to the establishment of DHMRs and the transcriptional programming of tissue-dependent genes in cerebellum and hippocampus (Yao et al., 2012).

Recently, a nano-hmC Seal method (Song et al., 2012; Han et al., 2016) enabled genome profiling of 5hmC from DNA isolated from 1000 cells. The 5hmC pattern was associated with hematopoiesis, as the enhancer sites were dynamically hydroxymethylated during hematopoietic stem cell (HSC) differentiation. Chromatin Immunoprecipitation sequencing (ChIP-seq) and Assay for Transposase-Accessible Chromatin sequencing (ATAC-seq) were used to show that the 5hmC density in gene bodies correlates positively with histone modification marks of gene activation, such as H3K4me1 and H3K4me2, in all cell types during hematopoiesis. However, 5hmC and H3K4me1 were not concentrated in the center of ATAC-seq peaks with high chromatin accessibility, but instead were enriched in ATAC-seq peaks with lower signal intensity or the less active chromatin elements (Chen and Riggs, 2011).

Furthermore, 5hmC modification is associated with tumorigenesis and stress response. Compared to healthy tissues, the abundance of 5hmC is reduced up to eightfold in cancer cells (Jin et al., 2011; Lian et al., 2012). TET2 depletion led not only to the loss, but also the redistribution of 5hmC in *tet2*-mutant AML murine stem cells (Han et al., 2016). Potential stress-related target genes were proposed to be regulated by 5hmC in early-life stressed mouse (Papale et al., 2017). The glucocorticoid receptor Nr3c1 in the hippocampus is highly mediated in stress response. 5hmC is increased about 1.8-fold at the seven CpGs in the 3' UTR of *Nr3c1* gene after 30 min restraint in a conical vial and 1 h recovery (Li et al., 2015).

In total, 458 hyper- and 174 hypo-DhMRs were identified in hippocampal DNA, including 470 and 166 genes that have hyper- and hypo-DhMRs, respectively, and 2 genes that showed both hyper- and hypo-DhMRs. Among them, more than 1/3 of the genes (240 in 638) are involved in stress responses. All the functional DhMRs contain at least one TF binding motif. These results suggest that 5hmC possibly acts through TFs, in that the affinity of TFs to DNA is changed in accordance with 5hmC modification (Li et al., 2016).

Many studies showed that 5hmC is linked with neurological and psychiatric disorders. Rett syndrome (Szulwach et al., 2011a), aging and Alzheimer's disease (Chouliaras et al., 2013; Shu et al., 2016; López et al., 2017), melanoma (Papale et al., 2017), Huntington's disease (Villar-Menéndez et al., 2013; Wang et al., 2013), Fragile X-associated tremor/ataxia syndrome (FXTAS) (Yao et al., 2014), Ataxia-telangiectasia (Jiang et al., 2015), Schizophrenia, bipolar disorder, major depressive disorder (Dong et al., 2012; Matrisciano et al., 2013; Tseng et al., 2014), Autism (Penagarikano et al., 2011), and intracerebral hemorrhage (Tang et al., 2016), all feature obvious changes of 5hmC. Aberrant life will be caused when the dynamics and distribution of 5hmC, or even the ratio of 5hmC to 5mC, are disturbed. It is widely accepted that when the abundance and distribution of 5hmC change, disease occurs.

WHAT ARE THE MECHANISM AND FUNCTION OF 5hmC?

Rediscovery of 5hmC in DNA inspired studies on the distribution and function of 5hmC. 5hmC modification of the phage DNA prevents restriction and degradation during infection into *E. coli* host (Hattman and Fukasawa, 1963; Shedlovsky and Brenner, 1963; Erikson and Szybalski, 1964). Some new enzymes are produced in the bacterium infected by phage with 5hmC modified DNA (Flaks and Cohen, 1959; Kornberg et al., 1959; Koerner et al., 1960). Substitution of C for 5hmC in vegetative *T4* DNA depresses the synthesis of late proteins (Hosoda and Levinthal, 1968; Kutter and Wiberg, 1969), moreover, transcription of some late genes can only occur from 5hmC-containing, but not from C-containing DNA (Kutter and Wiberg, 1969).

The function and mechanism of 5hmC are still elusive. Distribution of 5hmC is diversified at enhancer, promoter, TSS, gene body, 3' UTR or intragenic region. It is conceivable that 5hmC takes roles as *cis* element, for example, to promote or repress gene expression by binding to TFs, such as activators or repressors in regulatory regions of a gene.

The second way for 5hmC to work as *cis* player is to associate with histone modifications for the alteration of chromatin configuration, to switch "on" or "off" genes in heterochromatic or euchromatic chromatin, respectively. As mentioned above, TET proteins, 5hmC, 5mC and histone modifications such as H3K27me3 and H3K4me3, combine together to form the complicated bivalent domains of poised genes to maintain pluripotency of ESCs.

The third way in which 5hmC functions is possibly to modulate alternative splicing facilitated by 5hmC distribution in gene bodies, and exon/intron boundaries. Alternative splicing is a conserved way to diversify the transcriptome and proteome of a given genome. Transcripts from about 90% genes undergo alternative splicing (Wang et al., 2008). About 30 years ago it was shown that, splicing can occur co-transcriptionally in *Drosophila* (Beyer and Osheim, 1988). RNA polymerase II (Pol II) can recruit the 5' cap-binding complex (CAP), splicing and pre-spliceosome factors, and the polyadenylation complex in the context of nucleosome-containing chromatin (Luco et al., 2011). DNA modification and alternative splicing are linked by protein bridge of a variety of components such as CCCTC-binding factor (CTCF), MeCP2, and HP1 (Heterochromatin Protein 1) (Iannone and Valcarcel, 2013; Maunakea et al., 2013; Lev Maor et al., 2015; Yearim et al., 2015). 5hmC is more abundant in constitutive exons than in alternatively spliced exons (Khare et al., 2012). It is likely that 5hmC, similar to 5mC, plays a role in alternative splicing through these regulatory proteins. DNA modification contributes to binding of these proteins to DNA. Pol II elongation, recognition of alternative exons by spliceosome, and the subsequent inclusion or exclusion of exons are promoted or paused by DNA methylation (Lev Maor et al., 2015). Rett syndrome is caused by the inhibition of binding of MeCP2 to 5hmC (Mellén et al., 2012), additionally, an increase of 5hmC and decrease of CTCF occupancy at the *FXN* 5' UTR were also reported in Friedreich ataxia (Al-Mahdawi et al., 2013).

In addition, 5hmC is possibly linked with small RNA pathways. MiR29b, a key member of miR-29 family, is upregulated during differentiation of mESCs. MiR29b targets the 3' UTR of *TET1/TET2* mRNAs, however, 5hmC level is decreased by repressing of *TET1*, but not *TET2* (Tu et al., 2015). The miR29b-TET1 axis promotes formation of mesendoderm lineage by inducing the Nodal signaling pathway and expression of involved genes. MiR29b may associate with active DNA demethylation by repressing TDG. Taken together, 5hmC are closely linked with miRNA regulation in formation of mesendoderm lineage (Tu et al., 2015).

5hmC MODIFICATION AND ACTIVE DNA DEMETHYLATION IN PLANTS

In plants, 5mC is often found in contexts of CG, CHG, and CHH (H represents either A, C or T). The *de novo* DNA methylation and maintenance are associated with different pathways. In *Arabidopsis*, Domains Rearranged Methyl-transferase 2 (DRM2), DNA Methyl-Transferase 1 (MET1), Chromomethylase 3 (CMT3), CMT2 and DRM2 are responsible to maintain DNA methylation in different pathways or sequence contexts. DNA demethylation also occurs passively or positively in plants. However, TET proteins or UHRF2 (Ubiquitin-like with PHD and Ring Finger Domains 2) (Spruijt et al., 2013; Zhou et al., 2014), the writers or readers of 5hmC found in mammals, have not been identified in plants, thus DNA demethylation might not be processed through TET pathway. Active DNA demethylation by BER pathway is initiated by ROS1 (Repressor of Silencing

1) and other DME family members, DME, DML2, and DML3, the bifunctional DNA glycosylase and apyrimidinic (AP) lyase (Gehring et al., 2006; Morales-Ruiz et al., 2006). The DME glycosylases can remove 5mC from the DNA backbone directly and create an abasic site subject to the BER pathway (Gong and Zhu, 2011). However, mutations in the known demethylases in *Arabidopsis* do not cause global demethylation, but only affect the methylation status of some specific loci (Gehring et al., 2006; Lister et al., 2008). It is still elusive that how global demethylation occurs during gametogenesis and embryogenesis in plants. The BER pathway may not be the main pathway for global demethylation, because it will result in too many abasic sites and broken strands simultaneously, which could destabilize the whole genome (Zhu, 2009). Does this mean that any other pathway is involved in active DNA demethylation in plants? Is the unknown active demethylation pathway in plant similar to the TET pathway in mammals?

Repressor of Silencing 1 and DME purified from *E. coli* can excise 35-mer oligonucleotides containing cytosine, 5mC, 5hmC, 5fC, and 5caC, indicating that they are able to cleave both 5mC and 5hmC *in vitro*. However, 5hmC was not detected in the *Arabidopsis* genome (Jang et al., 2014). Several other groups also tried to detect if there is 5hmC in plants. Up to now, the existence of 5hmC in plants is still under debate (Terragni et al., 2012; Yao et al., 2012; Liu et al., 2013; Moricová et al., 2013; Wang et al., 2015). Wang et al. studied three rice cultivars and concluded that 5hmC modification is present in rice DNA, although in low abundance as 1.39 ± 0.16 and 2.17 ± 0.03 per million nucleosides in leaf and panicle, respectively. 5hmC distribution tends to localize in TE and heterochromatin regions in rice (Wang et al., 2015). Based on the extremely low abundance of 5hmC, the lack of genes encoding proteins for the generation, recognition, and functioning of 5hmC, and activity of ROS1 and DMEs to eliminate 5mC efficiently and directly, it is assumed that 5hmC, if any, may be an intermediate of passive demethylation in local regions, or not an enzymatic product in plants.

PERSPECTIVE: MORE QUESTIONS ARE WAITING

Benefiting from the development and progress of sequencing technology, biochemistry techniques, and cellular and molecular strategies, epigenetic research has been blooming in the past 20 years. It is no doubt that the content of epigenetic research has been enriched from studies on 5hmC and other recently determined epigenetic marks, such as methylated adenine in DNA, or even in RNA. It is suggested that 5hmC most likely plays a role in gene expression, pluripotency of stem cells, stress response, disease and aging. However, functional studies on 5hmC are still ahead. There are still some fundamental questions to be addressed:

(1) There is a long journey from global detection of 5hmC modification to precise elucidation of functional genes, pathways and regulatory mechanisms.

(2) More proteins that produce, recognize, and eliminate 5hmC must be identified. It is badly needed to uncover the mechanism underlying 5hmC function. The functional studies will be aided, in turn, by research on 5hmC writers, readers, and erasers.

(3) Does 5hmC exist in mitochondrial DNA (mtDNA)? Given that 5hmC is present in mtDNA, does it share the similar pattern with 5hmC in nuclear DNA? Is 5hmC involved in retrograde signal transduction from organelle to nucleus? And is there any evolutionary relation between 5hmC in mtDNA with that in bacteriophage DNA?

(4) RNA hydroxymethylcytosine (hmrC or hm5C, for the difference of 5hmC in DNA), but not 5mC, or 5hmC in DNA was reported in *Drosophila* (Delatte et al., 2016). Does it mean any preference of DNA or RNA modification in some organisms? Do the proteins which are involved in DNA epigenetic modification act in similar way in RNA?

(5) Since the distribution and abundance of 5hmC in DNA correlate with some mental or neuronal disorders, is the change of 5hmC a cause or a result of the diseases? Will the diseases be cured or prevented when 5hmC distribution alteration is prohibited by some gene therapy, if it is a cause of the disorders?

(6) Is it possible that 5hmC is enriched in some special cells or at appropriate developmental stages, but cannot be detected globally in plants? Can we obtain 5hmC genomic DNA if we introduce and express *TET* genes *in planta*? What will happen if 5hmC modification occurs in plant DNA?

AUTHOR CONTRIBUTIONS

D-QS and W-CY conceived the review. D-QS prepared manuscript. IA, JT, and W-CY discussed the proposal and revised manuscript.

FUNDING

This work was supported by projects from the Ministry of Science and Technology of China and National Natural Science Foundation of China (2013CB126904, 31370345 to D-QS, and 31330053 to W-CY).

ACKNOWLEDGMENT

We thank Prof. Sheila McCormick at UC Berkeley for English edition of the manuscript.

REFERENCES

- Al-Mahdawi, S., Sandi, C., Mouro Pinto, R., and Pook, M. A. (2013). Friedreich ataxia patient tissues exhibit increased 5-hydroxymethylcytosine modification and decreased CTCF binding at the FXN locus. *PLoS ONE* 8:e74956. doi: 10.1371/journal.pone.0074956
- Almeida, R. D., Sottile, V., Loose, M., De Sousa, P. A., Johnson, A. D., and Ruzov, A. (2012). Semi-quantitative immunohistochemical detection of 5-hydroxymethyl-cytosine reveals conservation of its tissue distribution between amphibians and mammals. *Epigenetics* 7, 137–140. doi: 10.4161/epi.7.2.18949
- Alonso, C., Pérez, R., Bazaga, P., and Herrera, C. M. (2015). Global DNA cytosine methylation as an evolving trait: phylogenetic signal and correlated evolution with genome size in angiosperms. *Front. Genet.* 6:4. doi: 10.3389/fgene.2015.00004
- Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P. W., D'Souza, Z., et al. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* 18, 225–233. doi: 10.1038/ncb3296
- Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326. doi: 10.1016/j.cell.2006.02.041
- Beyer, A. L., and Osheim, Y. N. (1988). Splice site selection, rate of splicing, and alternative splicing on nascent transcripts. *Genes Dev.* 2, 754–765. doi: 10.1101/gad.2.6.754
- Branco, M. R., Ficz, G., and Reik, W. (2012). Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat. Rev. Genet.* 13, 7–13. doi: 10.1038/nrg3080
- Capuano, F., Müllleder, M., Kok, R., Blom, H. J., and Ralser, M. (2014). Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Anal. Chem.* 86, 3697–3702. doi: 10.1021/ac500447w
- Chen, Z. X., and Riggs, A. D. (2011). DNA methylation and demethylation in mammals. *J. Biol. Chem.* 286, 18347–18353. doi: 10.1074/jbc.R110.205286
- Chouliaras, L., Mastroeni, D., Delvaux, E., Grove, A., Kenis, G., Hof, P. R., et al. (2013). Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. *Neurobiol. Aging* 34, 2091–2099. doi: 10.1016/j.neurobiolaging.2013.02.021
- Cliffe, L. J., Kieft, R., Southern, T., Birkeland, S. R., Marshall, M., Sweeney, K., et al. (2009). JBP1 and JBP2 are two distinct thymidine hydroxylases involved in J biosynthesis in genomic DNA of African trypanosomes. *Nucleic Acids Res.* 37, 1452–1462. doi: 10.1093/nar/gkn1067
- Delatte, B., Wang, F., Ngoc, L. V., Collignon, E., Bonvin, E., Deplus, R., et al. (2016). RNA biochemistry. Transcriptome-wide distribution and function of RNA hydroxymethylcytosine. *Science* 351, 282–285. doi: 10.1126/science.aac5253
- Diotel, N., Mérot, Y., Coumailleau, P., Gueguen, M. M., Sérandour, A. A., Salbert, G., et al. (2017). 5-Hydroxymethylcytosine marks postmitotic neural cells in the adult and developing vertebrate central nervous system. *J. Comp. Neurol.* 525, 478–497. doi: 10.1002/cne.24077
- Dong, E., Gavin, D. P., Chen, Y., and Davis, J. (2012). Upregulation of TET1 and downregulation of APOBEC3A and APOBEC3C in the parietal cortex of psychotic patients. *Transl. Psychiatry* 2, e159. doi: 10.1038/tp.2012.86
- Erikson, R. L., and Szybalski, W. (1964). The Cs₂SO₄ equilibrium density gradient and its application for the study of T-even phage DNA: glucosylation and replication. *Virology* 22, 111–124. doi: 10.1016/0042-6822(64)90053-4
- Feinberg, A. P., Koldobskiy, M. A., and Göndör, A. (2016). Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nat. Rev. Genet.* 17, 284–299. doi: 10.1038/nrg.2016.13
- Ficz, G., Branco, M. R., Seisenberger, S., Santos, F., Krueger, F., Hore, T. A., et al. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* 473, 398–402. doi: 10.1038/nature10008
- Flaks, J. G., and Cohen, S. S. (1959). Virus-induced acquisition of metabolic function. 1. Enzymatic formation of 5-hydroxymethyldeoxycytidylate. *J. Biol. Chem.* 234, 1501–1506.
- Fu, Y., and He, C. (2012). Nucleic acid modifications with epigenetic significance. *Curr. Opin. Chem. Biol.* 16, 516–524. doi: 10.1016/j.cbpa.2012.10.002
- Gehring, M., Huh, J. H., Hsieh, T. F., Penterman, J., Choi, Y., Harada, J. J., et al. (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* 124, 495–506. doi: 10.1016/j.cell.2005.12.034
- Globisch, D., Münzel, M., Müller, M., Michalakakis, 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
- Gong, Z., and Zhu, J. K. (2011). Active DNA demethylation by oxidation and repair. *Cell Res.* 21, 1649–1651. doi: 10.1038/cr.2011.140
- Grand, A., Jorge, N., Morell, C., Cadet, J., and Eriksson, L. A. (2014). Deamination features of 5-hydroxymethylcytosine, a radical and enzymatic DNA oxidation product. *J. Mol. Model.* 20, 2290–2299. doi: 10.1007/s00894-014-2290-9
- Grosser, C., Wagner, N., Grothaus, K., and Horsthemke, B. (2015). Altering TET dioxygenase levels within physiological range affects DNA methylation dynamics of HEK293 cells. *Epigenetics* 10, 819–833. doi: 10.1080/15592294.2015.1073879
- 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
- Hammoud, S. S., Low, D. H., Yi, C., Carrell, D. T., Guccione, E., and Cairns, B. R. (2014). Chromatin and transcription transitions of mammalian adult germline stem cells and spermatogenesis. *Cell Stem Cell* 15, 239–253. doi: 10.1016/j.stem.2014.04.006
- Han, D., Lu, X., Shih, A. H., Nie, J., You, Q., Xu, M. M., et al. (2016). A highly sensitive and robust method for genome-wide 5hmC profiling of rare cell populations. *Mol. Cell* 63, 711–719. doi: 10.1016/j.molcel.2016.06.028
- Hattman, S., and Fukasawa, T. (1963). Host-induced modification of T-even phages due to defective glucosylation of their DNA. *Proc. Nat. Acad. Sci. U.S.A.* 50, 297–300. doi: 10.1073/pnas.50.2.297
- He, X. J., Chen, T., and Zhu, J. K. (2011). Regulation and function of DNA methylation in plants and animals. *Cell Res.* 21, 442–465. doi: 10.1038/cr.2011.23
- 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
- Hendrich, B., Hardeland, U., Ng, H. H., Jiricny, J., and Bird, A. (1999). The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. *Nature* 401, 301–304. doi: 10.1038/4584
- Hosoda, J., and Levinthal, C. (1968). Protein synthesis by *E. coli* infected with bacteriophage T4D. *Virology* 34, 709–727. doi: 10.1016/0042-6822(68)90092-5
- Hotchkiss, R. D. (1948). The quantitative separation of purines, pyrimidines and nucleosides by paper chromatography. *J. Biol. Chem.* 175, 315–332.
- Hu, L., Lu, J., Cheng, J., Rao, Q., Li, Z., Hou, H., et al. (2015). Structural insight into substrate preference for TET-mediated oxidation. *Nature* 527, 118–122. doi: 10.1038/nature15713
- Huang, Y., Chavez, L., Chang, X., Wang, X., Pastor, W. A., Kang, J., et al. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc. Nat. Acad. Sci. U.S.A.* 111, 1361–1366. doi: 10.1073/pnas.1322921111
- Iannone, C., and Valcarcel, J. (2013). Chromatin's thread to alternative splicing regulation. *Chromosoma* 122, 465–474. doi: 10.1007/s00412-013-0425-x
- 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
- Jang, H., Shin, H., Eichman, B. F., and Huh, J. H. (2014). Excision of 5-hydroxymethylcytosine by DEMETER family DNA glycosylases. *Biochem. Biophys. Res. Commun.* 446, 1067–1072. doi: 10.1016/j.bbrc.2014.03.060
- 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

- Jiang, D., Zhang, Y., Hart, R. P., Chen, J., Herrup, K., and Li, J. (2015). Alteration in 5-hydroxymethylcytosine-mediated epigenetic regulation leads to Purkinje cell vulnerability in ATM deficiency. *Brain* 138, 3520–3536. doi: 10.1093/brain/awv284
- Jin, S. G., Jiang, Y., Qiu, R., Rauch, T. A., Wang, Y., Schackert, G., et al. (2011). 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res.* 71, 7360–7365. doi: 10.1158/0008-5472.CAN-11-2023
- Jin, S. G., Zhang, Z. M., Dunwell, T. L., Harter, M. R., Wu, X., Johnson, J., et al. (2016). Tet3 reads 5-carboxylcytosine through Its CXXC domain and is a potential guardian against neurodegeneration. *Cell Rep.* 14, 493–505. doi: 10.1016/j.celrep.2015.12.044
- Khare, T., Pai, S., Koncivicius, 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
- Koerner, J. F., Smith, M. S., and Buchanan, J. M. (1960). Deoxycytidine triphosphatase, an enzyme induced by bacteriophage infection. *J. Biol. Chem.* 235, 2691–2697.
- Kong, L., Tan, L., Lv, R., Shi, Z., Xiong, L., Wu, F., et al. (2016). A primary role of TET proteins in establishment and maintenance of De Novo bivalency at CpG Islands. *Nucleic Acids Res.* 44, 8682–8692. doi: 10.1093/nar/gkw529
- Kornberg, A., Zimmerman, S. B., Kornberg, S. R., and Josse, J. (1959). Enzymatic synthesis of deoxyribonucleic acid. VI. Influence of bacteriophage T2 on the synthetic pathway in host cells. *Proc. Nat. Acad. Sci. U.S.A.* 45, 772–785. doi: 10.1073/pnas.45.6.772
- 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
- Kubiura, M., Okano, M., Kimura, H., Kawamura, F., and Tada, M. (2012). Chromosome-wide regulation of euchromatin-specific 5mC to 5hmC conversion in mouse ES cells and female human somatic cells. *Chromosome Res.* 20, 837–848. doi: 10.1007/s10577-012-9317-9
- Kutter, E. M., and Wiberg, J. S. (1969). Biological effects of substituting cytosine for 5-hydroxymethylcytosine in the deoxyribonucleic acid of bacteriophage T4. *J. Virol.* 4, 439–453.
- Landfors, M., Johansen, J., Aronsen, J. M., Vågbo, C. B., Doré, L. C., He, C., et al. (2017). Genome-wide profiling of DNA 5-hydroxymethylcytosine during rat Sertoli cell maturation. *Cell Discov.* 3:17013. doi: 10.1038/celldisc.2017.13
- Lev Maor, G., Yearim, A., and Ast, G. (2015). The alternative role of DNA methylation in splicing regulation. *Trends Genet.* 31, 274–280. doi: 10.1016/j.tig.2015.03.002
- Li, S., Papale, L. A., Zhang, Q., Madrid, A., Chen, L., Chopra, P., et al. (2015). Hippocampal increase of 5hmC in the glucocorticoid receptor gene following acute stress. *Behav. Brain Res.* 286, 236–240. doi: 10.1016/j.bbr.2015.03.002
- Li, S., Papale, L. A., Zhang, Q., Madrid, A., Chen, L., Chopra, P., et al. (2016). Genome-wide alterations in hippocampal 5-hydroxymethylcytosine links plasticity genes to acute stress. *Neurobiol. Dis.* 86, 99–108. doi: 10.1016/j.nbd.2015.11.010
- Lian, C. G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., et al. (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* 150, 1135–1146. doi: 10.1016/j.cell.2012.07.033
- Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., et al. (2008). Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536. doi: 10.1016/j.cell.2008.03.029
- Liu, J., Zhu, Y., Luo, G. Z., Wang, X., Yue, Y., Wang, X., et al. (2016). Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nat. Commun.* 7:13052. doi: 10.1038/ncomms13052
- Liu, S., Dunwell, T. L., Pfeifer, G. P., Dunwell, J. M., Ullah, I., and Wang, Y. (2013). Detection of oxidation products of 5-methyl-2'-deoxycytidine in *Arabidopsis* DNA. *PLoS ONE* 8:e84620. doi: 10.1371/journal.pone.0084620
- López, V., Fernández, A. F., and Fraga, M. F. (2017). The role of 5-hydroxymethylcytosine in development, aging and age-related diseases. *Ageing Res. Rev.* 37, 28–38. doi: 10.1016/j.arr.2017.05.002
- Lu, X., Zhao, B. S., and He, C. (2015). TET family proteins: oxidation activity, interacting molecules, and functions in diseases. *Chem. Rev.* 115, 2225–2239. doi: 10.1021/cr500470n
- Luco, R. F., Allo, M., Schor, I. E., Kornblihtt, A. R., and Misteli, T. (2011). Epigenetics in alternative pre-mRNA splicing. *Cell* 144, 16–26. doi: 10.1016/j.cell.2010.11.056
- 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.C111.284620
- Matriciano, F., Tueting, P., Dalal, I., Kadriu, B., Grayson, D. R., Davis, J. M., et al. (2013). Epigenetic modifications of GABAergic interneurons are associated with the schizophrenia-like phenotype induced by prenatal stress in mice. *Neuropharmacology* 68, 184–194. doi: 10.1016/j.neuropharm.2012.04.013
- Maunakea, A. K., Chepelev, I., Cui, K., and Zhao, K. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* 23, 1256–1269. doi: 10.1038/cr.2013.110
- Mellén, M., Ayata, P., Dewell, S., Kriaucionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430. doi: 10.1016/j.cell.2012.11.022
- Morales-Ruiz, T., Ortega-Galisteo, A. P., Ponferrada-Marin, M. I., Martínez-Macias, M. I., Ariza, R. R., and Roldán-Arjona, T. (2006). DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Nat. Acad. Sci. U.S.A.* 103, 6853–6858. doi: 10.1073/pnas.0601109103
- Moréra, S., Imberty, A., Aschke-Sonnenborn, U., Rüger, W., and Freemont, P. S. (1999). T4 phage beta-glucosyltransferase: substrate binding and proposed catalytic mechanism. *J. Mol. Biol.* 292, 717–730. doi: 10.1006/jmbi.1999.3094
- 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.
- Müller, U., Bauer, C., Siegl, M., Rottach, A., and Leonhardt, H. (2014). TET-mediated oxidation of methylcytosine causes TDG or NEIL glycosylase dependent gene reactivation. *Nucleic Acids Res.* 42, 8592–8604. doi: 10.1093/nar/gku552
- Munzel, M., Globisch, D., Bruckl, T., Wagner, M., Welzmilller, V., Michalakis, S., et al. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* 49, 5375–5377. doi: 10.1002/anie.201002033
- Nabel, C. S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H. L., Stivers, J. T., et al. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* 8, 751–758. doi: 10.1038/nchembio.1042
- Neri, F., Incarnato, D., Krepelova, A., Rapelli, S., Pagnani, A., Zecchina, R., et al. (2013). Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biol.* 14:R91. doi: 10.1186/gb-2013-14-8-r91
- Nestor, C. E., Ottaviano, R., Reddington, J., Sproul, D., Reinhardt, D., Dunican, D., et al. (2012). Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* 22, 467–477. doi: 10.1101/gr.126417.111
- Niederhuth, C. E., Bewick, A. J., Ji, L., Alabady, M. S., Kim, K. D., Li, Q., et al. (2016). Widespread natural variation of DNA methylation within angiosperms. *Genome Biol.* 17, 194–212. doi: 10.1186/s13059-016-1059-0
- Ono, R., Taki, T., Taketani, T., Taniwaki, M., Kobayashi, H., and Hayashi, Y. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res.* 62, 4075–4080.
- Papale, L. A., Madrid, A., Li, S., and Alisch, R. S. (2017). Early-life stress links 5-hydroxymethylcytosine to anxiety-related behaviors. *Epigenetics* 12, 264–276. doi: 10.1080/15592294.2017.1285986
- Pastor, W. A., Pape, U. J., Huang, Y., Henderson, H. R., Lister, R., Ko, M., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* 473, 394–397. doi: 10.1038/nature10102
- Penagarikano, O., Abrahams, B. S., Herman, E. I., Winden, K. D., Gdalyahu, A., Dong, H., et al. (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell* 147, 235–246. doi: 10.1016/j.cell.2011.08.040
- Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K., and Yura, R. (1972). The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* 126, 781–790. doi: 10.1042/bj1260781

- Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C. A., et al. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* 50, 7008–7012. doi: 10.1002/anie.201103899
- Prakash, A., Carroll, B. L., Sweasy, J. B., Wallace, S. S., and Doublé, S. (2014). Genome and cancer single nucleotide polymorphisms of the human NEIL1 DNA glycosylase: activity, structure, and the effect of editing. *DNA Repair.* 14, 17–26. doi: 10.1016/j.dnarep.2013.12.003
- Schomacher, L., Han, D., Musheev, M. U., Arab, K., Kienhöfer, S., von Seggern, A., et al. (2016). Neil DNA glycosylases promote substrate turnover by Tdg during DNA demethylation. *Nat. Struct. Mol. Biol.* 23, 116–124. doi: 10.1038/nsmb.3151
- Shedlovsky, A., and Brenner, S. (1963). A chemical basis for the host-induced modification of T-even bacteriophages. *Proc. Nat. Acad. Sci. U.S.A.* 50, 300–305. doi: 10.1073/pnas.50.2.300
- Shu, L., Sun, W., Li, L., Xu, Z., Lin, L., Xie, P., et al. (2016). Genome-wide alteration of 5-hydroxymethylcytosine in a mouse model of Alzheimer's disease. *BMC Genomics* 17:381. doi: 10.1186/s12864-016-2731-1
- Smith, Z. D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220. doi: 10.1038/nrg3354
- Song, C. X., Diao, J., Brunger, A. T., and Quake, S. R. (2016). Simultaneous single-molecule epigenetic imaging of DNA methylation and hydroxymethylation. *Proc. Nat. Acad. Sci. U.S.A.* 113, 4338–4343. doi: 10.1073/pnas.1600223113
- Song, C. X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* 29, 68–72. doi: 10.1038/nbt.1732
- Song, C. X., Yi, C., and He, C. (2012). Mapping recently identified nucleotide variants in the genome and transcriptome. *Nat. Biotechnol.* 30, 1107–1116. doi: 10.1038/nbt.2398
- Spruijt, C. G., Gnerlich, F., Smits, A. H., Pfaffeneder, T., Jansen, P. W., Bauer, C., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159. doi: 10.1016/j.cell.2013.02.004
- Sun, M., Song, M. M., Wei, B., Gao, Q., Li, L., Yao, B., et al. (2016). 5-Hydroxymethylcytosine-mediated alteration of transposon activity associated with the exposure to adverse in utero environments in human. *Hum. Mol. Genet.* 25, 2208–2219. doi: 10.1093/hmg/ddw089
- Szulwach, K. E., Li, X., Li, Y., Song, C. X., Han, J. W., Kim, S., et al. (2011a). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* 7:e1002154. doi: 10.1371/journal.pgen.1002154
- Szulwach, K. E., Li, X., Li, Y., Song, C. X., Wu, H., Dai, Q., et al. (2011b). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* 14, 1607–1616. doi: 10.1038/nn.2959
- 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
- Tan, F. E., and Ye, G. W. (2016). Blurred boundaries: the RNA binding protein Lin28A is also an epigenetic regulator. *Mol. Cell* 61, 1–2. doi: 10.1016/j.molcel.2015.12.018
- Tan, L., Xiong, L., Xu, W., Wu, F., Huang, N., Xu, Y., et al. (2013). Genome-wide comparison of DNA hydroxymethylation in mouse embryonic stem cells and neural progenitor cells by a new comparative hMeDIP-seq method. *Nucleic Acids Res.* 41:e84. doi: 10.1093/nar/gkt091
- Tang, Y., Han, S., Asakawa, T., Luo, Y., Han, X., Xiao, B., et al. (2016). Effects of intracerebral hemorrhage on 5-hydroxymethylcytosine modification in mouse brains. *Neuropsychiatr. Dis. Treat.* 12, 617–624. doi: 10.2147/NDT.S97456
- Terragni, J., Bitinaite, J., Zheng, Y., and Pradhan, S. (2012). Biochemical characterization of recombinant β -glucosyltransferase and analysis of global 5-hydroxymethylcytosine in unique genomes. *Biochemistry* 51, 1009–1019. doi: 10.1021/bi2014739
- Thomson, J. P., and Meehan, R. R. (2017). The application of genome-wide 5-hydroxymethylcytosine studies in cancer research. *Epigenetics* 9, 77–91. doi: 10.2217/epi-2016-0122
- Tollervey, 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
- Tseng, P. T., Lin, P. Y., Lee, Y., Hung, C. F., Lung, F. W., Chen, C. S., et al. (2014). Age-associated decrease in global DNA methylation in patients with major depression. *Neuropsychiatr. Dis. Treat.* 10, 2105–2114. doi: 10.2147/NDT.S71997
- Tu, J., Ng, S. H., Luk, A. C., Liao, J., Jiang, X., Feng, B., et al. (2015). MicroRNA-29b/Tet1 regulatory axis epigenetically modulates mesoderm differentiation in mouse embryonic stem cells. *Nucleic Acids Res.* 43, 7805–7822. doi: 10.1093/nar/gkv653
- Valinluck, V., and Sowers, L. C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA Maintenance methyltransferase DNMT1. *Cancer Res.* 67, 946–950. doi: 10.1158/0008-5472.CAN-06-3123
- Vidalis, A., Živković, D., Wardenaar, R., Roquis, D., Tellier, A., and Johannes, F. (2016). Methylome evolution in plants. *Genome Biol.* 17, 264–277. doi: 10.1186/s13059-016-1127-5
- Villar-Menéndez, I., Blanch, M., Tyebji, S., Pereira-Veiga, T., Albasanz, J. L., Martín, M., et al. (2013). Increased 5-methylcytosine and decreased 5-hydroxymethylcytosine levels are associated with reduced striatal A2AR levels in Huntington's disease. *Neuromolecular. Med.* 15, 295–309. doi: 10.1007/s12017-013-8219-0
- Wang, E. T., Sandberg, R., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., et al. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509
- Wang, F., Yang, Y., Lin, X., Wang, J. Q., Wu, Y. S., Xie, W., et al. (2013). Genome-wide loss of 5hmC is a novel epigenetic feature of Huntington's disease. *Hum. Mol. Genet.* 22, 3641–3653. doi: 10.1093/hmg/ddt214
- Wang, X. L., Song, S. H., Luo, S., Khrebtkova, I., Zhang, L., Mayr, C., 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
- Wenzel, D., Palladino, F., and Jedrusik-Bode, M. (2011). Epigenetics in *C. elegans*: facts and challenges. *Genesis* 49, 647–661. doi: 10.1002/dvg.20762
- Wion, D., and Casadesus, J. (2006). N⁶-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* 4, 183–192. doi: 10.1038/nrmicro1350
- Wossidlo, M., Arand, J., Sebastiano, V., Lepikhov, K., Boiani, M., Reinhardt, R., et al. (2010). Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* 29, 1877–1888. doi: 10.1038/emboj.2010.80
- Wu, H., D'Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., et al. (2011). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* 25, 679–684. doi: 10.1101/gad.2036011
- Wyatt, G. R., and Cohen, S. S. (1952). A new pyrimidine base from bacteriophage nucleic acids. *Nature* 170, 1072–1073. doi: 10.1038/1701072a0
- Wyatt, G. R., and Cohen, S. S. (1953). The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochem. J.* 55, 774–782. doi: 10.1042/bj0505774
- Yao, B., Lin, L., Street, R. C., Zalewski, Z. A., Galloway, J. N., Wu, H., et al. (2014). Genome-wide alteration of 5-hydroxymethylcytosine in a mouse model of fragile X-associated tremor/ataxia syndrome. *Hum. Mol. Genet.* 23, 1095–1107. doi: 10.1093/hmg/ddt504
- Yao, Q., Song, C. X., He, C., Kumaran, D., and Dunn, J. J. (2012). Heterologous expression and purification of *Arabidopsis thaliana* VIM1 protein: in vitro evidence for its inability to recognize hydroxymethylcytosine, a rare base in Arabidopsis DNA. *Protein Expr. Purif.* 83, 104–111. doi: 10.1016/j.pep.2012.03.003
- Yearim, A., Gelfman, S., Shayevitch, R., Melcer, S., Glaich, O., Mallm, J. P., et al. (2015). HP1 is involved in regulating the global impact of DNA methylation on alternative splicing. *Cell Rep.* 10, 1122–1134. doi: 10.1016/j.celrep.2015.01.038
- Yu, M., Hon, G. C., Szulwach, K. E., Song, C. X., Jin, P., Ren, B., et al. (2012). Tet-assisted bisulfite sequencing of 5-hydroxymethylcytosine. *Nat. Protoc.* 7, 2159–2170. doi: 10.1038/nprot.2012.137
- Yu, Z., Genest, P. A., ter Riet, B., Sweeney, K., DiPaolo, C., Kieft, R., et al. (2007). The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. *Nucleic Acids Res.* 35, 2107–2115. doi: 10.1093/nar/gkm049

- Zeng, Y., Yao, B., Shin, J., Lin, L., Kim, N., Song, Q., et al. (2016). Lin28A binds active promoters and recruits Tet1 to regulate gene expression. *Mol. Cell.* 61, 153–160. doi: 10.1016/j.molcel.2015.11.020
- 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 © 2017 Shi, Ali, Tang and Yang. 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) or licensor 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.