

# Reversing DNA Methylation: Mechanisms, Genomics, and Biological Functions

Hao Wu<sup>1,2,3,4</sup> and Yi Zhang<sup>1,2,3,4,\*</sup>

<sup>1</sup>Howard Hughes Medical Institute

<sup>2</sup>Program in Cellular and Molecular Medicine, Boston Children's Hospital

<sup>3</sup>Department of Genetics

<sup>4</sup>Harvard Stem Cell Institute

Harvard Medical School, WAB-149G, 200 Longwood Avenue, Boston, MA 02115, USA

\*Correspondence: [yzhang@genetics.med.harvard.edu](mailto:yzhang@genetics.med.harvard.edu)

<http://dx.doi.org/10.1016/j.cell.2013.12.019>

Methylation of cytosines in the mammalian genome represents a key epigenetic modification and is dynamically regulated during development. Compelling evidence now suggests that dynamic regulation of DNA methylation is mainly achieved through a cyclic enzymatic cascade comprised of cytosine methylation, iterative oxidation of methyl group by TET dioxygenases, and restoration of unmodified cytosines by either replication-dependent dilution or DNA glycosylase-initiated base excision repair. In this review, we discuss the mechanism and function of DNA demethylation in mammalian genomes, focusing particularly on how developmental modulation of the cytosine-modifying pathway is coupled to active reversal of DNA methylation in diverse biological processes.

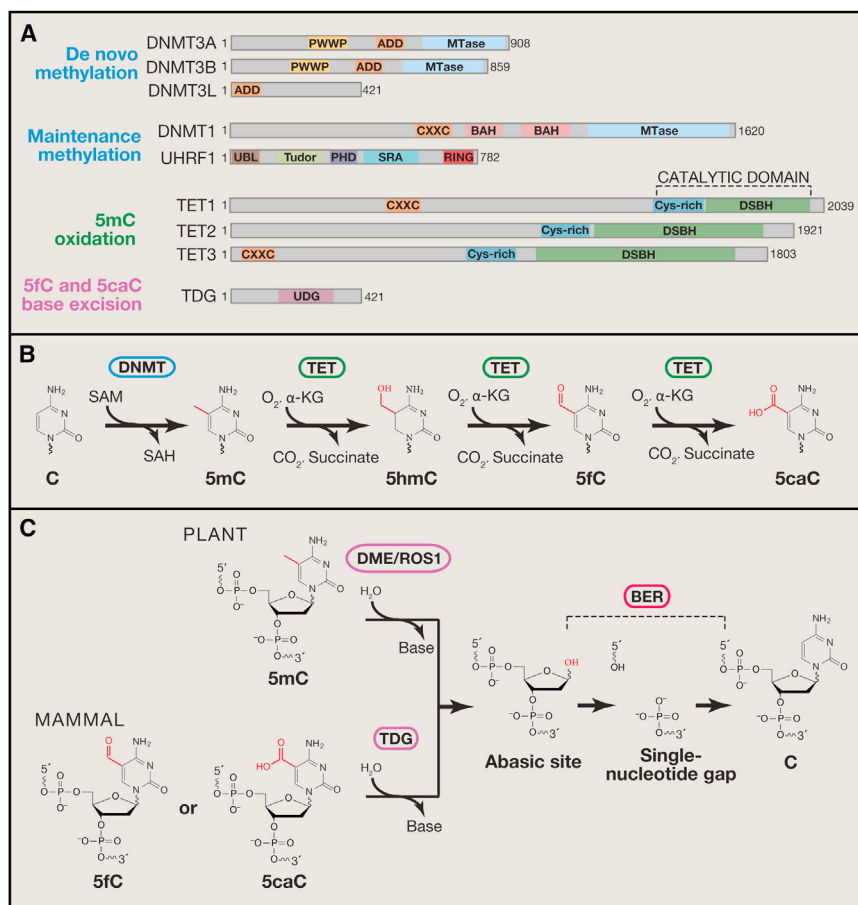
## Introduction

It has long been recognized that cytosines in the genome, as part of the genetic code, also carry epigenetic information through chemical modification of its pyrimidine ring (Holliday and Pugh, 1975; Riggs, 1975). The dual functions associated with cytosines provide a means by which developmental stage- and cell-type-specific epigenetic memory can be directly deposited onto DNA itself (Bird, 2002). Methylation of the fifth position of cytosine (5-methylcytosine, 5mC) is a highly conserved epigenetic modification of DNA found in most plant, animal and fungal models (Law and Jacobsen, 2010) and has a profound impact on genome stability, gene expression, and development (Jae-nisch and Bird, 2003; Smith and Meissner, 2013).

In mammals, new DNA methylation pattern is established by de novo DNA methyltransferases, DNMT3A and DNMT3B (Okano et al., 1999; Okano et al., 1998) (Figures 1A and 1B). Their activity can be modulated by a catalytically inactive family member, DNMT3L (Goll and Bestor, 2005). In somatic cells, 5mC is primarily restricted to palindromic CpG dinucleotides, which are typically methylated in a symmetric manner (de novo methylation in Figure 2A). Methylation of cytosine in non-CpG context (CpH, H = A, T, C) is prevalent in plants (Law and Jacobsen, 2010) but is rare in most mammalian cell-types. Recent work suggests that non-CpG methylation is relatively abundant in oocytes, pluripotent embryonic stem cells (ESCs), and mature neurons (Lister et al., 2013; Lister et al., 2009; Shirane et al., 2013; Xie et al., 2012), but the function of mammalian non-CpG methylation remains unclear. Of the roughly 28 million CpGs in the human genome, 60%–80% are methylated in somatic cells (Smith and Meissner, 2013). During mitosis, the global CpG

methylation pattern is faithfully maintained in daughter cells through the action of maintenance DNA methyltransferase DNMT1 and its obligate partner, the ubiquitin-like plant homeo-domain and RING finger domain 1 (UHRF1), which preferentially recognizes hemimethylated CpGs (Bostick et al., 2007; Hermann et al., 2004; Sharif et al., 2007) (maintenance methylation in Figures 1A and 2A). Such inheritability of CpG methylation suggests a role for 5mC in long-term epigenetic regulation required for diverse biological processes, such as stable silencing of gene expression, maintenance of genome stability, and establishment of genomic imprinting (Bird, 2002).

Although DNA methylation pattern in somatic cells is stably maintained, genome-wide loss of 5mC, or DNA demethylation, has been observed in specific developmental stages such as preimplantation embryos and developing primordial germ cells (PGCs) (Hajkova et al., 2002; Mayer et al., 2000; Oswald et al., 2000; Sasaki and Matsui, 2008). Global DNA demethylation is important for setting up pluripotent states in early embryos and for erasing parental-origin-specific imprints in developing PGCs (Feng et al., 2010b). Mounting evidence indicates that the rapid erasure of 5mC during these two major waves of epigenetic reprogramming could not be fully explained by replication-dependent passive loss of 5mC, suggesting the existence of enzymatic activities capable of actively removing or modifying methyl groups on cytosines (Wu and Zhang, 2010). However, a unifying mechanistic understanding of active DNA demethylation processes in mammalian cells did not emerge until recently. As we will discuss below, the transformative discovery of ten-eleven translocation (TET) proteins as 5mC oxidase has provided major insights into mechanisms of active DNA demethylation.



**Figure 1. Domain Architecture and Enzymatic Activities of Cytosine Methylation and Demethylation Machinery**

(A) Schematic diagrams of predicted functional domains in proteins involved in de novo methylation, maintenance methylation, 5mC oxidation, and 5fC/5caC excision. All TET proteins contain a C-terminal catalytic domain that includes a Cys-rich insert and a large double-stranded  $\beta$ -helix (DSBH) domain. TET1 and TET3 proteins also contain a N-terminal CXXC domain. TDG contains a uracil DNA glycosylase (UDG) domain. The number of amino acids of the full-length isoform is indicated, and the numbering corresponds to mouse proteins. ADD, ATRX-DNMT3-DNMT3L; PWWP, proline-tryptophan-tryptophan-proline motif; MTase, DNA methyltransferase domain; BAH, bromo-adjacent homology; CXXC, zinc-finger Cys-X-X-Cys; SRA, SET and RING associated; RING, Really interesting new gene; PHD, Plant homeodomain; UBL, ubiquitin-like domain.

(B) The step-wise cytosine modification pathway that includes cytosine methylation by DNMTs and iterative oxidation of 5mC by TET proteins. DNMTs use S-adenosylmethionine (SAM) as a methyl donor to catalyze methylation at the 5-position of cytosine, yielding S-adenosylhomocysteine (SAH). As Fe(II)/ $\alpha$ -ketoglutarate ( $\alpha$ -KG)-dependent dioxygenases, TET proteins use a base-flipping mechanism to flip the target base out of the duplex DNA into the catalytic site (Hashimoto et al., 2013; Hu et al., 2013). The active site Fe(II) is bound by conserved His-His-Asp residues in TETs and coordinates water and  $\alpha$ -KG. TET enzymes use molecular oxygen as a substrate to catalyze oxidative decarboxylation of  $\alpha$ -KG, yielding CO<sub>2</sub>, enzyme-bound succinate, and a reactive high-valent Fe(IV)-oxo intermediate. The enzyme-bound Fe(IV)-oxo intermediate reacts with 5mC/5hmC/5fC to generate 5hmC/5fC/5caC, with a

net oxidative transfer of the single oxygen atom to the substrate, resulting in regeneration of the Fe(II) species.

(C) Base excision repair of 5mC (in plants) or 5fC/5caC (in mammals) to complete DNA demethylation. In plants, the DME/ROS1 family of bifunctional DNA glycosylases is capable of catalyzing the release of 5mC base from DNA by cleavage of the N-glycosidic bond, generating an abasic (apurinic and apyrimidinic [AP]) site. The DNA backbone is then nicked by AP lyase activity of DME/ROS1 enzymes. The 3' sugar group is then cleaved by an AP endonuclease and the resulting single-nucleotide gap is filled in with an unmodified C by DNA polymerase and ligase activities. In mammals, oxidized 5mC bases (5fC/5caC), could be excised by TDG. TDG is a monofunctional DNA glycosylase, so other proteins are required to provide the AP lyase activity to remove the sugar ring to generate a single-nucleotide gap.

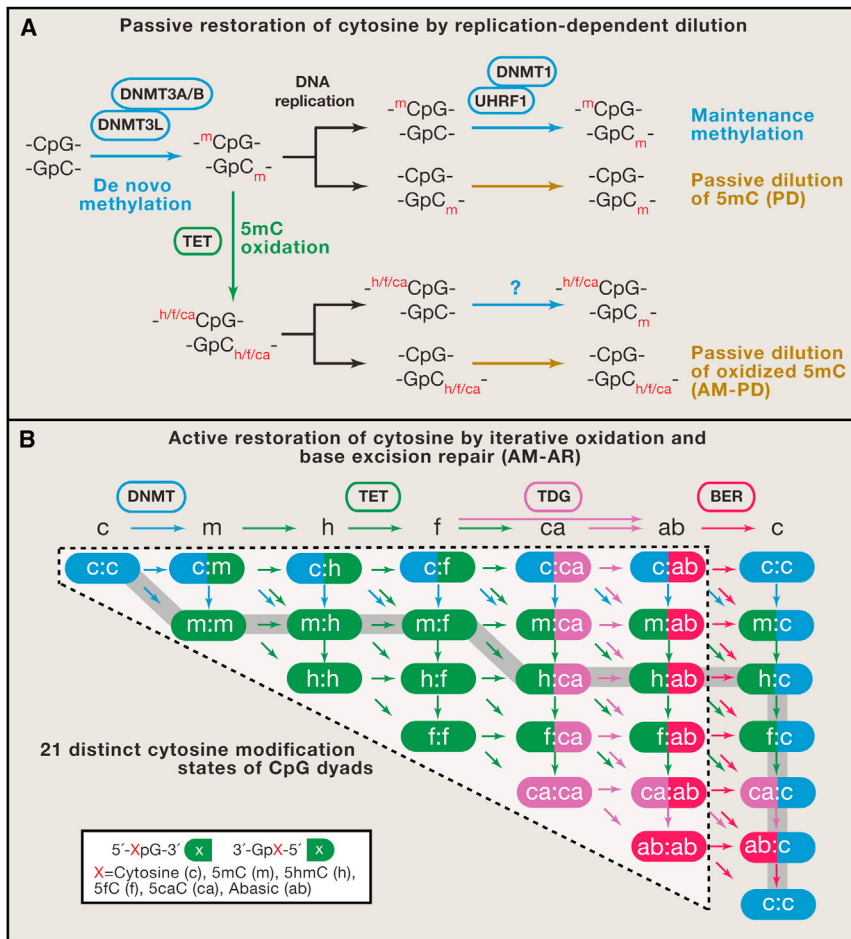
The biochemical basis of TET enzymes in oxidative modification of 5mC has recently been reviewed elsewhere (Kohli and Zhang, 2013; Pastor et al., 2013). In this review, we focus on an integrated understanding of mechanisms, genomics, and biological functions of mammalian DNA demethylation process. First, we summarize the current mechanistic understanding of passive and active DNA demethylation pathways. Second, we examine the recent advances in development of genomic mapping technologies for 5mC oxidation derivatives as well as the understanding of potential regulatory functions of oxidized cytosine bases in the mammalian genome. Finally, we discuss how distinct modes of DNA demethylation dynamics can be achieved through modulation of the cytosine-modifying enzymatic pathway to satisfy the needs of diverse biological processes.

#### DNA Methylation and Demethylation in Mammals: A Cyclic Enzymatic Cascade

DNA demethylation can take place either passively or actively. Upon DNA replication, the maintenance machinery DNMT1/

UHRF1 restores the symmetrical CpG methylation pattern by methylating the unmodified cytosine in the nascent DNA strand (Bostick et al., 2007; Sharif et al., 2007). Passive loss of 5mC can therefore be achieved through successive cycles of DNA replication in the absence of functional DNMT1/UHRF1 (passive dilution of 5mC [PD] in Figure 2A). One example of replication-dependent passive loss of 5mC is the global erasure of 5mC in the maternal genome during mouse preimplantation development (see below).

Compelling evidence indicates that 5mC can also be rapidly erased independently of DNA replication in specific biological settings in both plants and mammals (Law and Jacobsen, 2010; Wu and Zhang, 2010). Such change involves active DNA demethylation mechanisms, which require one or more enzymatic reactions to modify or remove 5mC. In flowing plants (e.g., *Arabidopsis thaliana*), active erasure of 5mC is mediated by the Demeter (DME)/repressor of silencing 1 (ROS1) family of DNA glycosylases and base excision repair (BER) machinery (Zhu, 2009) (Figure 1C). However, mammalian orthologs of



**Figure 2. Mechanisms of Passive and Active Reversal of CpG DNA Methylation**

(A) Schematic diagrams of replication-dependent passive loss of 5mC or oxidized 5mC within CpG dyads. Replication-dependent passive dilution (PD) of 5mC occurs in the absence of the DNA methylation maintenance machinery (DNMT1/UHRF1). By contrast, 5mC oxidation derivatives, 5hmC (h) (potentially 5fC [f] and 5caC [ca]) may facilitate passive demethylation as hemihydroxymethylated CpGs is an inefficient substrate for DNMT1. This form of active DNA demethylation is termed as active modification (AM) followed by passive dilution (AM-PD). Whether DNMT1 or DNMT3 can efficiently methylate CpG dyads hemimodified with 5fC/5caC is unclear.

(B) Schematic diagrams of replication-independent DNA demethylation within CpG dyads. TET and TDG mediate sequential 5mC oxidation and 5fC/5caC excision. The resulting abasic site is repaired by BER to regenerate unmodified cytosines. All possible intermediate products (a total of 21 distinct cytosine modification states for a single CpG dyad) of the DNMT/TET/TDG/BER enzymatic cascade are depicted. For each state, the left half represents the top strand of CpG dyad, whereas the right half indicates the bottom strand. The arrows represent enzymatic reactions and are color-coded according to that of the corresponding enzyme. One of many potential combinations of intermediate steps for the TET/TDG-mediated active DNA demethylation process is highlighted in gray. Of note, 5fC can be directly excised by TDG to generate abasic site (e.g., f:ca → ab:ab). For clarity, all f → ab reactions are not depicted.

### Iterative Oxidation of 5mC by TET Proteins

The discovery of enzymatic activity modifying 5mC through oxidation was motivated by two known pathways involved in oxidative modification of thymine (T). The first involves the biosynthesis of “base J” (β-D-glucosyl-hydroxymethyluracil), a modified base present in the genome of the parasite *Trypanosoma brucei* that causes African sleeping sickness (Borst and Sabatini, 2008). Base J biosynthesis is achieved by a two-step process. J-binding protein (JBP) 1 and 2 first oxidize T in the DNA to 5-hydroxymethyluracil (5hmU); then a glycosyltransferase completes the synthesis of base J by adding glucose groups to 5hmU (Borst and Sabatini, 2008). JBP1/2 proteins are members of the Fe(II)/α-ketoglutarate (α-KG)-dependent dioxygenase family (Loenarz and Schofield, 2011). Systematic bioinformatic analysis identified the human gene, TET1, as a homolog of JBP1/2 (Iyer et al., 2009). TET1, initially discovered as a fusion partner of the histone H3 Lys4 methyltransferase mixed-lineage leukemia 1 (MLL1) (t(10;11)(q22;q23)) in certain acute myeloid leukemia (AML) (Lorsbach et al., 2003; Ono et al., 2002), contains a C-terminal catalytic domain consisting of a Cys-rich region and a double-stranded β-helix (DSBH) fold (Figure 1A). Overexpression of TET1 in cultured cells results in reduction of genomic 5mC level, and recombinant TET1 proteins can oxidize 5mC in vitro and the resulting product was determined as 5hmC (Tahiliani et al., 2009) (Figure 1B). Similar enzymatic activity

DME/ROS1 enzymes have not been identified. Mammalian cells may therefore achieve active DNA demethylation using different mechanisms. Early efforts focused on testing various known DNA modifying proteins, such as DNA repair enzymes, cytosine deaminases, and DNA glycosylases (Box 1A). However, experimental evidence for these proposed candidates is often restricted to very specific biological settings, and in many cases the activity cannot be substantiated by follow-up studies (Ooi and Bestor, 2008; Wu and Zhang, 2010). Thus, these studies did not reveal a consensus mechanistic framework for active DNA demethylation in mammals.

The breakthrough came in 2009 when the presence of 5-hydroxymethylcytosine (5hmC), an oxidized derivative of 5mC, was unambiguously demonstrated in the mammalian genome and TET family of DNA dioxygenases were identified as the enzyme that convert 5mC to 5hmC (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). Subsequent studies have revealed that TET proteins can further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxycytosine (5caC) (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011). This series of discoveries have established the biochemical basis for understanding the molecular mechanism of active DNA demethylation in mammals.

### Box 1. Known and Putative Mechanisms of Active Erasure of 5-Methylcytosines

#### A. 5mC Oxidation-Independent Active DNA Demethylation Mechanisms

##### 1. Enzymatic Removal of the Methyl Group from 5mC

The most straightforward way to achieve DNA demethylation is direct removal of the methyl-group. However, this is a thermodynamically unfavorable reaction. One study proposed that methyl-CpG-binding domain protein 2 (MBD2) can catalyze such a reaction leading to the release of methanol, but other laboratories could not reproduce this finding and active erasure of paternal 5mC occurs normally in fertilized oocytes lacking Mbd2 (Wu and Zhang, 2010). To date, no compelling evidence supports that this biochemical reaction could take place under physiological conditions.

##### 2. Nucleotide Excision Repair to Erase 5mC

Excision repair of short genomic regions that contain methylated cytosine nucleotides can indirectly result in erasure of 5mC. There are two major DNA excision repair mechanisms to repair single-stranded DNA damage: base excision repair (BER) and nucleotide excision repair (NER). BER is used to correct damaged bases or mismatched base pairs (e.g., G:T mismatch), whereas NER can repair bulky DNA lesions formed by exposure to radiation or chemicals. Once the bulky DNA lesion is recognized, specific enzymatic activities introduce dual incisions flanking the damage region. The resulting single-stranded gap (usually 24–32 nucleotides) is then filled in by DNA repair polymerases and ligases. The Gadd45 (growth arrest and DNA-damage-inducible protein 45) family of proteins was reported to stimulate active DNA demethylation via NER (Barreto et al., 2007). While some evidence suggests that Gadd45 may induce active DNA demethylation in *Xenopus laevis* embryos and mammalian cells (Ma et al., 2009; Rai et al., 2008; Schmitz et al., 2009), conflicting results were also reported (Jin et al., 2008). Furthermore, *Gadd45a*- and *Gadd45b* null mice are grossly normal and have no global alteration in DNA methylation levels (Engel et al., 2009; Ma et al., 2009). Thus, the role of Gadd45 in DNA demethylation remains uncertain.

##### 3. Direct Base Excision Repair of 5mC by DNA Glycosylases

Active DNA demethylation can be achieved by a DNA glycosylase that efficiently excises 5mC base from DNA. Strong genetic and biochemical evidence supports the use of this mechanism in plants (Zhu, 2009). Mammalian DNA glycosylases, including TDG and methyl-CpG binding domain-containing protein 4 (MBD4), were also proposed to have excision activity against 5mC (Zhu et al., 2000). However, 5mC excision activity of these enzymes is significantly (30- to 40-fold) lower than their activity for T or U in T·G or U·G mismatches. In addition, *Mbd4* null mice are viable and have normal DNA methylation pattern during development. Interestingly, *Tdg* null mice die around E12.5 and exhibit modest increase in level of 5mC at some CpG-rich gene promoters (Cortázar et al., 2011; Cortellino et al., 2011). Because TDG, but not MBD4, has robust activity toward 5fC and 5caC, it is likely that 5fC/5caC excision activity, instead of the marginal 5mC excision activity of TDG, is responsible for methylation defects observed in *Tdg*-deficient cells (see main text).

##### 4. Deamination and Repair of 5mC Base

Deamination of 5mC generates thymine and the resulting G·T mismatch can be repaired by multiple DNA glycosylases (e.g., TDG) and BER to replace the mismatched T with C. Activation-induced deaminase (AID)/APOBEC-family of cytidine deaminases are known to function in generating mutations in DNA and RNA,

### Box 1. Continued

which are biologically important for antibody diversification in B cells and for RNA editing (Conticello, 2008). In vitro biochemical analysis suggests that AID/APOBEC could also deaminate 5mC to T (Morgan et al., 2004). However, AID/APOBEC only efficiently acts on single-strand but not double-strand DNA (Bransteitter et al., 2003), and exhibit substantially lower activity on 5mC (10- to 20-fold) relative to cytosine, their canonical substrate (Nabel et al., 2012). In addition, AID- or APOBEC1-knockout mice exhibit the expected B cell and immunological defects, but are viable and fertile (Wu and Zhang, 2010). Nevertheless, studies in zebrafish embryos (Rai et al., 2008), mouse PGCs (Popp et al., 2010), mouse ESC/human fibroblast fused heterokaryons (Bhutani et al., 2010), and iPSCs (Kumar et al., 2013), support a potential role of AID/APOBEC-mediated 5mC deamination in DNA demethylation. Further studies are needed to clarify the exact biochemical mechanism and function of AID/APOBEC proteins in DNA demethylation. Moreover, de novo DNMTs have been reported to possess in vitro 5mC deaminase activity in the absence of the methyl-donor S-adenosylmethionine (SAM) (Métivier et al., 2008). Given that SAM is relatively abundant in all living cells and is a general methyl-group donor for many essential enzymes, the in vivo relevance of DNMT-mediated deamination of 5mC remains unknown (Ooi and Bestor, 2008).

##### 5. Radical SAM Mechanisms

Elongator complex protein 3 (ELP3) was reported to promote paternal genome demethylation in mouse zygotes (Okada et al., 2010). While it is proposed that the Fe-S radical SAM domain of ELP3 may be involved in the demethylation process (Wu and Zhang, 2010), direct biochemical evidence demonstrating the demethylation activity of ELP3 is currently lacking.

#### B. 5mC Oxidation-Dependent Active DNA Demethylation Mechanisms

##### 1. Replication-Dependent Passive Dilution of Oxidized 5mC

Iterative oxidation of 5mC by TET proteins generates oxidized cytosine bases (5hmC/5fC/5caC). These cytosine analogs may facilitate DNA demethylation by impairing the binding and/or activity of maintenance methylation machinery (DNMT1/UHRF1) during DNA replication. Compelling evidence suggests that this mechanism is used in the erasure of paternal genome methylation during preimplantation development as well as in the complete demethylation of imprinting control regions in developing PGCs (see main text).

##### 2. DNA Repair-Mediated Excision of Modified 5mC

*TDG-mediated excision of 5fC/5caC followed by BER.* TET proteins further oxidize 5hmC to 5fC and 5caC, which can be efficiently excised by TDG and repaired by BER pathway. Multiple studies have biochemically validated this pathway (see main text). Compelling genetic and genomic evidence suggests that TET/TDG-mediated active DNA demethylation takes place in mouse ESCs, but how extensive this mechanism is utilized in vivo and in other cell-types is currently unknown.

*AID/APOBEC-mediated deamination of 5hmC followed by BER.*

Similar to 5mC deamination, AID/APOBEC proteins may theoretically deaminate 5hmC to generate 5-hydroxyuracil (5hmU). One study has shown that 5hmC deamination by AID/APOBEC may take place in cultured cells and in mouse brain (Guo et al., 2011). The resulting 5hmU·G mismatch can be repaired by DNA glycosylases such as TDG and SMUG1 (Cortellino et al., 2011). However, subsequent studies show that AID/APOBEC proteins exhibit no detectable deaminase activity against 5hmC in in vitro

(Continued on next page)



**Box 1. Continued**

biochemical assays or in cells overexpressing these enzymes (Nabel et al., 2012), challenging the biochemical plausibility of this proposed mechanism. One potential explanation for the negligible activity of AID/APOBEC on 5hmC is that the relatively bulky steric size of 5-position substituent (cytosine [ $\text{H}^-$ ] < 5mC [ $\text{CH}_3$ ] < 5hmC [ $\text{OHCH}_2$ ]) may have a significant negative impact on deaminase activity (Nabel et al., 2012).

3. Direct Removal of the Oxidized 5-Position Substituent

**Dehydroxymethylation by DNMTs.** Although direct removal of a methyl group from 5mC is thermodynamically challenging, the removal of the oxidized methyl group is more feasible. In vitro biochemical studies suggest that DNMT3A/3B can remove the hydroxymethyl group of 5hmC to generate unmodified cytosine in the absence of SAM and under oxidizing conditions (Chen et al., 2012; Liutkeviciute et al., 2009). Given that SAM is present at relatively high levels in all cell types, the physiological relevance of this reaction remains unclear.

**Decarboxylation of 5caC.** In the thymine salvage pathway, T-to-U conversion is completed by iso-oxotetrahydrofolate-mediated decarboxylation of 5-carboxyuracil (Smiley et al., 2005). Given the chemical similarity between THase and TET enzymes, it has been proposed that decarboxylation activity may exist in mammalian cells to directly convert 5caC to unmodified cytosine (Wu and Zhang, 2010). One study suggests that 5caC decarboxylase activity may exist in mouse ESC lysates (Schiesser et al., 2012), but the responsible enzyme awaits to be identified.

was also demonstrated in all three TET proteins in mouse (TET1-3) (Ito et al., 2010). Supporting the physiological relevance of TET-mediated 5mC oxidation, 5hmC can be readily detected in most cell-types (1%–5% of 5mC) and is particularly enriched in the genome of adult neurons (15%–40% of 5mC) (Globisch et al., 2010; Kriaucionis and Heintz, 2009).

The second pathway that provides clues on enzymatic activity of TET proteins involves T-to-uracil (U) conversion in the pyrimidine salvage pathway. In this process, T is first converted by thymine hydroxylase (THase), also a Fe(II)/ $\alpha$ -KG-dependent dioxygenase, to 5hmU, 5-formyluracil (5fU), and 5-carboxyluracil (5caU) through three successive oxidation reactions; then isoorotate decarboxylase completes the T-to-U conversion through decarboxylation of 5caU (Smiley et al., 2005). The similarity in chemistry between T-to-U conversion and oxidative modification of 5mC suggests the possibility that TET enzymes may function similarly to THase and can mediate iterative oxidation of 5mC (Wu and Zhang, 2010). With improved chromatographic separation conditions, TET proteins have been shown to be capable of further oxidizing 5hmC to 5fC and 5caC (He et al., 2011; Ito et al., 2011) (Figure 1B). Mass spectrometry analysis has shown that 5fC is present in a wide-range of cell types and 5caC is detectable in mouse ESCs, although their levels are at least an order of magnitude less than that of 5hmC (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011). The low abundance of 5fC and 5caC is partially due to the presence of specific enzymatic activity capable of efficiently excising these highly oxidized bases from DNA (see below). Thus, TET-mediated 5mC oxidation reactions extend the previously recognized binary state of cytosine modification (C/5mC) to five distinct states (C/5mC/5hmC/

5fC/5caC) in the mammalian genome, raising the possibility that these new cytosine analogs may play a role in dynamic regulation of 5mC.

**Passive Restoration of Unmodified Cytosines by Replication-Dependent Dilution of Oxidized 5mC**

The presence of oxidized 5mC bases at CpG sites may facilitate replication-dependent passive loss of 5mC. While there are conflicting results regarding whether UHRF1 can efficiently recognize hemihydroxymethylated CpG sites ( $^h\text{CG:GC}$ ) (Frauer et al., 2011; Hashimoto et al., 2012), independent studies have established that DNMT1 is significantly less efficient (10~60-fold) in methylating  $^h\text{CG:GC}$  than hemimethylated ( $^m\text{CG:GC}$ ) sites in vitro (Hashimoto et al., 2012; Valinluck et al., 2004). Whether hemimodified CpGs carrying 5fC or 5caC ( $\text{XG:GC}$ ,  $\text{X} = ^f\text{C}$  or  $^{\text{ca}}\text{C}$ ) also impair DNMT1 functions is unclear. In summary, TET proteins may initiate a two-step demethylation process in proliferating cells that involves initial active modification (AM) of 5mC through oxidation and subsequent replication-dependent passive dilution (PD) of 5hmC and potentially 5fC/5caC (AM-PD in Figure 2A). This mode of active DNA demethylation is distinct from simple passive dilution of 5mC (PD in Figure 2A) (Kohli and Zhang, 2013), as AM-PD may be effective even in the presence of functional methylation maintenance machinery. Replication-dependent DNA demethylation in the presence of DNMT1/UHRF1 has been reported during erythropoiesis (Shearstone et al., 2011), it will be interesting to examine whether 5mC oxidation may contribute to DNA demethylation in this system.

Interestingly, DNMT3A/B has comparable in vitro methylase activity toward  $^m\text{CG:GC}$  and  $^h\text{CG:GC}$  sites (Hashimoto et al., 2012; Otani et al., 2013b), suggesting that DNMT3 may be involved in maintaining 5mC levels in dividing cells with high TET activity. In support of this notion, mouse ESCs deficient in both DNMT3A and DNMT3B become globally hypomethylated after prolonged culturing (Chen et al., 2003). Because 5hmC/5fC/5caC in the paternal genome can be passively diluted in the presence of maternal DNMT3A proteins during early preimplantation development (Inoue et al., 2011; Inoue and Zhang, 2011), DNMT3A alone may not be sufficient to effectively methylate hemimodified sites ( $\text{XG:GC}$ ,  $\text{X} = ^h\text{C}$ ,  $^f\text{C}$ , or  $^{\text{ca}}\text{C}$ ). Further studies are needed to elucidate the interplay between 5mC oxidation and replication-dependent demethylation.

**Active Restoration of Unmodified Cytosines by Enzymatic Removal of Oxidized 5mC**

In addition to passive dilution of modified cytosines, several TET initiated but replication-independent active demethylation mechanisms have been proposed (Box 1B). These active DNA demethylation processes also involve initial oxidative modification of 5mC but complete the demethylation through replication-independent active restoration (AR) of unmodified C. First, removal of oxidized 5-position substituent, through either dehydroxymethylation of 5hmC by DNMTs (Chen et al., 2012; Liutkeviciute et al., 2009) or decarboxylation of 5caC by a putative decarboxylase (Schiesser et al., 2012; Wu and Zhang, 2010), may directly revert the oxidized 5mC base to unmodified cytosine. Second, DNA glycosylase-mediated excision of oxidized 5mC bases, including removal of 5hmU (deamination product

of 5hmC) (Guo et al., 2011) or 5fC/5caC (He et al., 2011; Maiti and Drohat, 2011), may result in abasic sites that are further repaired by BER to restore unmethylated C.

Among these proposed TET-dependent AM-AR demethylation pathways, the mechanism that involves TDG-mediated excision of 5fC/5caC has received the most experimental support. As a member of uracil DNA glycosylase (UDG) superfamily (Figure 1A), TDG is known to use a base-flipping mechanism to remove the pyrimidine base from T·G or U·G mismatches to initiate BER (Stivers and Jiang, 2003). Given the precedent of DME/ROS1-mediated excision repair of 5mC in plants (Figure 1C), DNA glycosylases have long been suspected in playing a role in active DNA demethylation in mammals. TDG is a particularly attractive candidate as it can act as a scaffold protein to bridge transcriptional coactivator CBP/p300 to numerous transcription factors, supporting a gene regulatory role of TDG (Dalton and Bellacosa, 2012). Early studies suggested that TDG might promote DNA demethylation through either direct excision of 5mC or T (the deamination product of 5mC). Although biochemically feasible, these mechanisms probably play a limited role in DNA demethylation given the marginal 5mC excision activity of TDG and AID/APOBEC deaminases' selectivity for single-stranded DNA (Box 1A). The role of TDG in DNA demethylation was revisited after additional oxidized cytosine analogs were discovered in the genome. While TDG cannot efficiently remove 5mC or 5hmC, it exhibits robust *in vitro* excision activity toward 5fC or 5caC (5fC·G or 5caC·G pairs) in duplex DNA (He et al., 2011; Maiti and Drohat, 2011). In fact, TDG has a higher activity for 5fC·G compared to T·G mismatch, its cognate substrate (Maiti and Drohat, 2011). The molecular basis for excision of 5fC/5caC by TDG stems from the effect of 5fC and 5caC in destabilizing the base-sugar linkage (*N*-glycosidic bond) as C bases with a weakened *N*-glycosidic bond can be efficiently excised by DNA glycosylases (Bennett et al., 2006). Further biophysical analysis indicates that the active site of TDG may possess specific structural features to mediate the recognition of oxidized 5mC bases (Zhang et al., 2012). Supporting the physiological relevance of this mechanism, overexpression of TET and TDG in the HEK293 cells rapidly depleted 5fC and 5caC (Nabel et al., 2012), whereas *Tdg*-deficiency in mouse ESCs caused a 5- to 10-fold increase in levels of 5fC and 5caC (He et al., 2011; Shen et al., 2013; Song et al., 2013a). Thus, methylation, iterative oxidation and excision repair offer a complete cytosine-modifying cascade that permits replication-independent erasure of DNA methylation (Figures 1C and 2B).

While other members of the UDG family of DNA glycosylases (UNG, MBD4, and SMUG1) are dispensable for embryonic development (Kemmerich et al., 2012; Millar et al., 2002), *Tdg* null mouse embryos exhibit a wide range of developmental defects and die around embryonic day (E) 12.5 (Cortázar et al., 2011; Cortellino et al., 2011). Interestingly, among all UDG DNA glycosylases, only TDG can efficiently excise 5fC/5caC (He et al., 2011; Maiti and Drohat, 2011), suggesting a unique role of TDG's 5fC/5caC excision activity in embryonic development. Indeed, mutant embryos carrying the glycosylase activity defective mutation (N151A in mouse TDG) phenocopy developmental defects of *Tdg* null embryos (Cortellino et al., 2011). Given its

robust excision repair activity for mismatch mutations, TDG may potentially ensure proper development by reducing genetic mutations; however, *Tdg*-deficiency induced embryonic lethality phenotype was fully penetrant, which is incompatible with an antimutagenic DNA repair defect that would exhibit highly variable phenotypes due to stochastic mutations in different genes. Thus, inactivation of TDG may primarily result in epigenetic rather than genetic defects in the genome. Dys-regulation of C/5mC/5hmC levels and/or accumulation of ectopic 5fC/5caC at specific gene regulatory sequences in *Tdg*-deficient mice may potentially affect transcription and cause developmental defects. Further *in vivo* analysis is needed to clarify how extensive TET/TDG-mediated DNA demethylation pathway is utilized *in vivo* and to elucidate the exact mechanism by which TDG activity contributes to transcription and development.

### Intermediate Steps of TET/TDG-Mediated Replication-Independent DNA Demethylation

Considering the cytosine methylation and demethylation process in the context of TET/TDG-mediated step-wise enzymatic cascade prompts key mechanistic questions: does TET enzyme act on DNA substrates processively or distributively for the three successive oxidation reactions? What are the preferred intermediate steps through which TET and TDG enzymatically revert 5mC to unmodified cytosine? Does TDG-mediated excision of 5fC/5caC preferentially generate single-strand break (SSB) or double-strand break (DSB)? Base-resolution mapping studies showed that steady-state 5hmC is almost exclusively found in the CpG context (>99% in CpGs), even in mouse ESCs and neurons where non-CpG methylation is prevalent (Lister et al., 2013; Yu et al., 2012). Thus, oxidative modification of 5mC by TET proteins may occur mostly in the CpG context. In line with this, recent structural and biochemical analyses of human TET2 catalytic domain or *Naegleria* Tet-like dioxygenase (NgTet1) indicate that these 5mC oxidases have a stronger preference to oxidize 5mC in CpG sites than in non-CpG context (Hashimoto et al., 2013; Hu et al., 2013). Alternatively, iterative oxidation of 5mC at non-CpG sites could occur at a much faster rate such that 5hmC at these sites is only transiently generated and rapidly converted to 5fC/5caC, which are removed by TDG/BER. Base-resolution mapping of 5fC/5caC in *Tdg* mutant cells can potentially capture these short-lived oxidized 5mC bases within non-CpG sites.

Within each CpG dyad, the cytosine-modifying pathway (DNMT-TET-TDG/BER) can theoretically generate 21 distinct cytosine modification states (XG:GX, X = C, <sup>m</sup>C, <sup>hm</sup>C, <sup>f</sup>C, <sup>ca</sup>C or abasic), six of which are in symmetric form (Figure 2B). Base-resolution mapping of 5hmC in mouse ESCs indicates that less than 20% of 5hmC is potentially present in symmetrically modified form (<sup>hm</sup>CG:GC<sup>hm</sup>) (Yu et al., 2012). Furthermore, crystal structures of TET2 or NgTet1 in complex with DNA containing <sup>m</sup>CG:GC<sup>m</sup> suggest that TET enzymes initiate the oxidation reaction by DNA bending and base flipping, and act on a single 5mC from a symmetrically methylated CpG site (Hashimoto et al., 2013; Hu et al., 2013). Thus, hemimodified CpG dyads may be the predominant pool of intermediates generated from 5mC oxidation reactions. Strand asymmetry associated with 5hmC (or 5fC and 5caC) implies that TET/TDG-mediated active demethylation may frequently generate single abasic site at

methylated CpGs, which typically leads to formation of SSB. Reducing the frequency of symmetrically oxidized CpGs (XG:GX, X = <sup>5</sup>C or <sup>5</sup>caC) may be physiologically relevant as two abasic sites in a single CpG dyad (XG:GX, X = abasic) could frequently generate DSB, increasing the chance of introducing deleterious mutations. Curiously, in vitro biochemical assay showed that TDG has higher excision activity for hemicarboxyl-methylated CpG site (<sup>5</sup>caCG:GC) than for fully carboxylmethylated one (<sup>5</sup>caCG:GC<sup>ca</sup>) (He et al., 2011), which may also help reduce frequency of generating DSB. Sequencing technology that allows simultaneous mapping of 5fC and 5caC at base-resolution may help determine the relative frequency of SSBs versus DSBs introduced during TET/TDG-mediated active demethylation process.

### Regulatory Mechanisms of TET-Dependent DNA Demethylation

Maintaining proper gene expression and cellular identity requires precise control of the DNA methylation pattern (Ji et al., 2010; Ziller et al., 2013). 5mC is likely to be the default epigenetic state of CpGs (except for those within CpG-rich promoters) in the mammalian genome, thus direct regulation of DNA methylation machineries generally leads to global changes in 5mC levels; however, regulatory mechanisms that modulate substrate accessibility, enzymatic activity, expression levels, and genomic targeting of TET enzymes may provide more precise control of the pattern of 5mC (or oxidized 5mC) at specific loci in the genome.

#### Regulation by Substrate Preference and Accessibility of TET Enzymes

Quantitative mass spectrometry analysis of mouse ESCs suggests that 5hmC (~5% of 5mC) is significantly more prevalent than 5fC (0.06%–0.6% of 5mC) and 5caC (0.01% of 5mC) (Ito et al., 2011; Pfaffeneder et al., 2011), suggesting that TET enzymes tend to stall at 5hmC. The processivity of 5mC oxidation may be regulated by substrate preference and/or the accessibility of TET enzymes. In vitro biochemical analysis suggests that the initial reaction rate of TET enzymes for oxidizing 5mC to 5hmC is significantly higher than that of oxidizing 5hmC or 5fC (Hashimoto et al., 2013; Ito et al., 2011), indicating that TET proteins may be more efficient in utilizing 5mC (as compared to 5hmC or 5fC) as a substrate. Structural analysis of TET2 catalytic domain shows that the methyl group of 5mC is not involved in the TET2-DNA interaction, thus allowing the active site to accommodate oxidized 5mC derivatives for further oxidation (Hu et al., 2013). It remains to be determined whether the active site of TET enzymes has distinct affinity for 5mC, 5hmC, and 5fC. Interestingly, a SRA domain-containing protein, UHRF2, has been recently identified as a 5hmC-binding protein and reported to be able to enhance the processivity of TET enzymes (Spruijt et al., 2013), potentially via its ability to flip modified cytosine from DNA duplex and increase the substrate accessibility of TET proteins.

#### Regulation of Enzymatic Activity by Metabolites and Cofactors

Emerging evidence indicates that changes of metabolic states in cells may affect the relative concentration of enzymatic cofac-

tors and in turn influence the activity of DNA and histone-modifying enzymes (Kaelin and McKnight, 2013). TET enzymes require oxygen, Fe (II) and  $\alpha$ -KG to catalyze 5mC oxidation reactions.  $\alpha$ -KG is primarily produced from isocitrate in the tricarboxylic acid (TCA) cycle by isocitrate dehydrogenases 2 and 3 (IDH2 and IDH3) in the mitochondria. Cytosolic IDH1 can provide an alternative source of  $\alpha$ -KG. Interestingly, inactivating mutations in the TCA cycle enzymes, fumarate hydratase (FH) and succinate dehydrogenase (SDH), or tumor-associated missense mutations in IDH1/2 result in accumulation of high concentrations of fumarate, succinate, and the oncometabolite 2-hydroxyglutarate (2-HG), all of which share structural similarity with  $\alpha$ -KG and competitively inhibit Fe(II)/ $\alpha$ -KG-dependent dioxygenases (Kaelin and McKnight, 2013), including TET enzymes and JmjC-domain-containing histone demethylases (Klose et al., 2006).

Vitamin C (also known as ascorbic acid), a micronutrient known for its antiscorvy activity in humans, is also implicated in modulation of TET activity. Recent studies have supported a role of vitamin C in stimulating the catalytic activity of TET proteins (Blaschke et al., 2013; Minor et al., 2013; Yin et al., 2013). Vitamin C may directly interact with the catalytic domain of TET proteins and act as a positive modulator of TET activity by facilitating conformational changes and/or recycling of the cofactor Fe (II) (Yin et al., 2013). In mouse ESCs, vitamin C significantly increases levels of all 5mC oxidation products, particularly 5fC and 5caC (>10-fold). Moreover, ground-state mouse ESCs cultured in the presence of vitamin C adopt a methylome more resembling that of blastocysts in vivo (Blaschke et al., 2013).

#### Regulation of Genomic Targeting of TET Enzymes

In addition to the C-terminal catalytic domain, *TET1* and *TET3* encode a zinc finger Cysteine-X-X-Cysteine (CXXC) domain at their N-termini (Figure 1A). The CXXC domain is present in many chromatin-associated proteins and has a strong preference for unmethylated CpGs (Long et al., 2013). Although *TET2* lacks a CXXC domain, a neighboring gene *IDAX* (inhibition of the Dvl and axin complex; also known as CXXC4) encodes a CXXC domain highly similar to those in *TET1* (also known as CXXC6) and *TET3* (CXXC10). It is possible that *IDAX* was originally part of an ancestral *TET2* gene that underwent a chromosomal inversion, which separated the CXXC domain from the catalytic domain (Iyer et al., 2009; Ko et al., 2013). Genome-wide occupancy analysis suggests that *TET1* is preferentially localized to unmethylated CpG-rich promoters and genic regions in mouse ESCs (Williams et al., 2011; Wu et al., 2011b), and *IDAX* is also enriched at unmethylated CpG sequences and directly interacts with *TET2* (Ko et al., 2013). Further biochemical and structural analyses indicate that *TET1* CXXC domain may also recognize methylated cytosines (Xu et al., 2011b; Zhang et al., 2010), and the *TET3* CXXC domain can target unmethylated cytosines within both CpG and non-CpG contexts (Xu et al., 2012). Thus, unlike other CXXC domains that specifically recognize unmethylated CpG sites, CXXC domains of *TET1/3* and *IDAX* may have increased flexibility in sequence selectivity and play a role in facilitating the recruitment of TET enzymes to their specific genomic targets.

### Transcriptional and Posttranscriptional Regulation of TET Levels

Three TET genes exhibit different expression patterns in a developmental-stage- and cell-type-specific manner (Ito et al., 2010; Szwagierczak et al., 2010). A high-level of TET1 is specifically detected in mouse ESCs, the inner cell mass of blastocysts, and developing PGCs (E9.5–12.5), whereas TET2 and TET3 are broadly expressed in various mouse adult tissues (Ito et al., 2010; Yamaguchi et al., 2012). Interestingly, TET3 is the only TET enzyme present in mouse oocytes and one-cell zygotes (Iqbal et al., 2011; Wossidlo et al., 2011). Transcriptional control of TET genes is likely mediated by cell-type-specific transcription factors. For instance, the upstream promoter region of the *TET1* gene contains a large cluster of binding sites for pluripotency-related transcription factors (TFs) (Ficz et al., 2011), suggesting a role of pluripotency-related TFs in maintaining TET1 expression. Indeed, TET1 is highly expressed in mouse ESCs, but its level is rapidly downregulated upon differentiation (Ito et al., 2010; Koh et al., 2011).

Given the relatively large size of mRNA and 3' untranslated regions (UTR) of *TET1–3* genes, TET mRNAs may also be regulated posttranscriptionally by microRNAs. Indeed, two recent studies have reported a role of oncogenic microRNA miR-22 in negatively regulating TET protein levels in the context of breast cancer development and hematopoietic stem cell transformation (Song et al., 2013b; Song et al., 2013c). In addition, IDAX may target TET2 proteins for degradation via a caspase-dependent mechanism; similarly, the CXXC domain of TET3 may negatively regulate its own protein stability (Ko et al., 2013), suggesting an auto-inhibitory mechanism of regulating TET protein levels.

### Genomic Distribution of Oxidized Methylcytosines and Their Potential Roles in Transcription Regulation

Genome-wide mapping of oxidized 5mC bases in the mammalian genome represents an unbiased approach to investigate the potential roles of TET-mediated demethylation pathway and these oxidized cytosines in epigenome function and transcriptional regulation. In principle, DNMT and TET enzymes can modify each of the approximately 28 million CpGs in the diploid mammalian genome to adopt one of the five epigenetic states (C/5mC/5hmC/5fC/5caC).

#### Genome-wide Mapping Technologies for Oxidized 5mC Bases

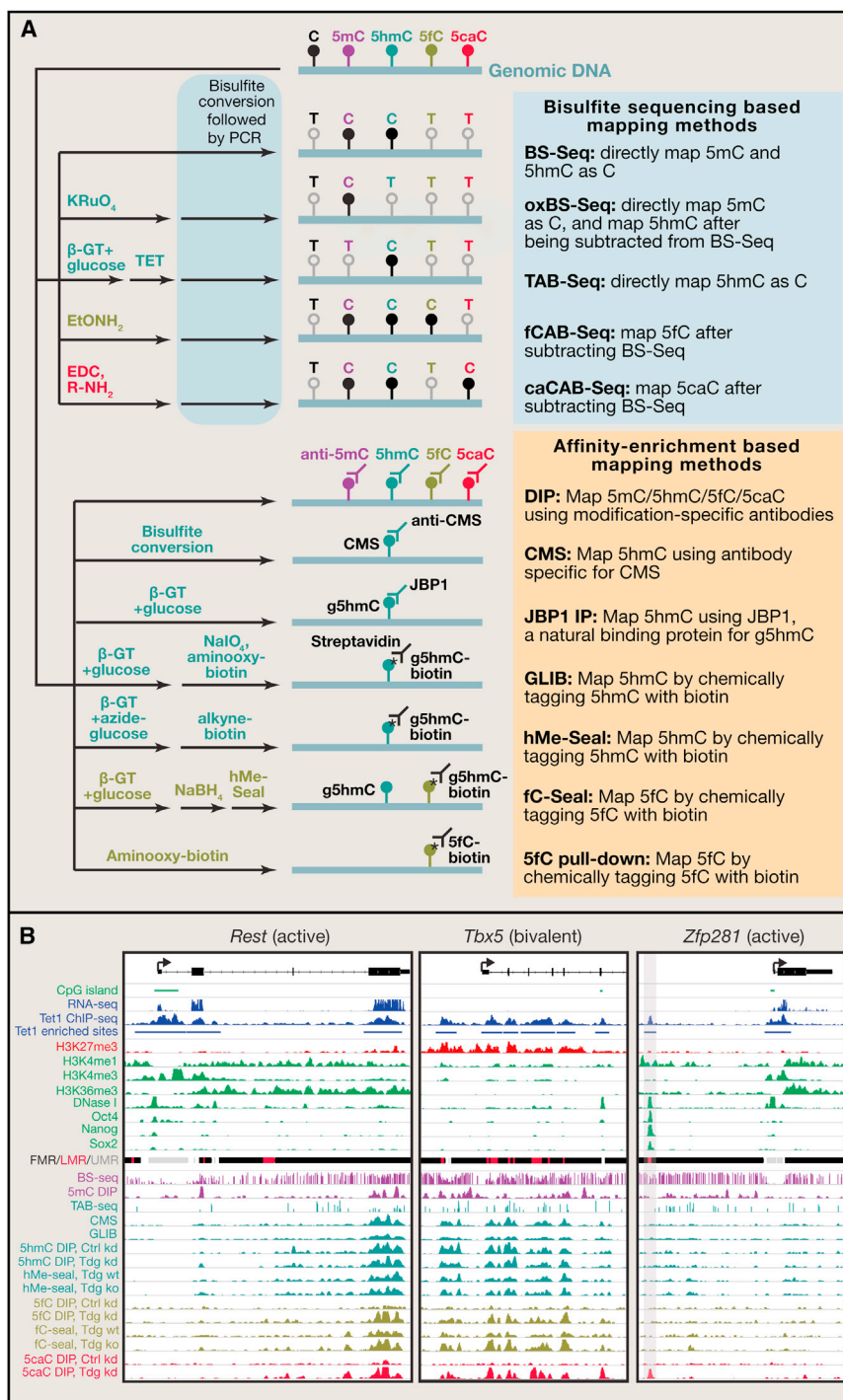
Multiple methods exploiting specific biochemical or biophysical properties of modified cytosine bases have been developed to map their distribution across the genome (Figure 3A). These methods can be categorized into two groups: first, affinity-enrichment-based approaches, which utilize modification-specific antibodies or selective chemical labeling of modified cytosines (Box 2A); second, base-resolution mapping methods such as bisulfite sequencing (BS-Seq) (Box 2B). The affinity-based methods are generally straightforward to perform and more cost-effective. However, the resolution of these methods is limited by the size of DNA fragments (several hundred base pairs), and affinity-enrichment methodology identifies regions enriched for modified bases by comparing signals (normalized

sequencing read counts) in pull-down and those in negative control (IgG or input), thereby providing a measurement of relative fold-enrichment (pull-down/control). By contrast, base-resolution mapping of modified bases can determine which cytosine (including exact position and strand) is modified and typically measures the percentage of modification level ((number of modified cytosines/number of all cytosines) × 100%) at a given cytosine. The probability that a cytosine can be confidently identified as modified is governed by the sequencing depth and abundance of the modification at that cytosine in a specific cell type. Given the low abundance of oxidized methylcytosines, base-resolution methods require a relatively high sequencing depth to detect them. It has been estimated that at a normal sequencing depth (20–30× for BS-Seq), base-resolution mapping of 5hmC (TAB-seq in Figure 3A and Box 2B) can confidently detect cytosines with an average 20% or higher level of hydroxymethylation regardless of the local density of cytosine substrates (Yu et al., 2012). As a result, base-resolution methods have enhanced sensitivity for detecting individual cytosine that is modified at medium-to-high level but sparsely distributed (that is, not clustered with other 5hmC). Indeed, while 80%–90% of 5hmC peaks identified by affinity-based methods (e.g., 5hmC antibody IP and GLIB in Figure 3A) contain at least one of the approximately 2.06 million high-confidence 5hmC-marked cytosines detected by TAB-seq in mouse ESCs, 5hmC peaks from affinity-based methods only cover 30%–40% of high-confidence 5hmC identified by TAB-seq (Yu et al., 2012). Because affinity-based methods integrate signals from multiple cytosines within a given genomic region, they are capable of detecting closely clustered cytosines with low level of modification, which may otherwise be missed by base-resolution methods. Further increasing the sequencing depth of base-resolution mapping experiments may help identify cytosines carrying low-level modification, although it is unclear whether these infrequently modified cytosines are biologically relevant. In summary, both affinity-based and single-base mapping methods can consistently identify regions enriched for cytosines that are frequently modified (medium-to-high level), but their results may differ at sparsely distributed or infrequently modified cytosines. Discrepancies between published genome-wide mapping studies may be due to the use of different methodologies and data analysis strategies as well as varying definition of a given genomic feature. Here, we focus on recent advances on mapping of oxidized 5mC bases including 5fC/5caC in stem cells and their potential gene regulatory roles at various genomic features.

#### Oxidized Methylcytosines at Gene Promoters

Approximately 70% of mammalian promoters are associated with short CpG-dense sequences (termed CpG islands [CGIs], on average 1 kilobase [kb] long) (Deaton and Bird, 2011). CpG-rich promoters (also known as high-CpG promoters [HCPs]) are generally unmethylated and transcriptionally active as H3K4 methylation and histone variant H2A.Z enriched at these HCPs antagonize DNMT binding (Ooi et al., 2007; Zilberman et al., 2008). The lack of 5mC at HCPs predicts that TET-mediated oxidative modification of 5mC is minimal at these promoters. Indeed, the current consensus view from both affinity-enrichment and base-resolution mapping studies suggests that





**Figure 3. Genome-wide Mapping Methods for Cytosine Modifications**

(A) Summary of BS-seq or affinity-enrichment based genome-wide mapping methods of DNA cytosine modifications. In conjunction with various chemical and enzymatic pretreatment, BS-seq or affinity-enrichment (antibody or chemical tagging) methods have been developed to map 5mC (BS-seq and oxBS-seq; 5mC DNA immunoprecipitation (DIP)), 5hmC (oxBS-seq and TAB-seq; 5hmC DIP, CMS IP, JBP1 IP, GLIB, and hMe-Seal), 5fC (fCAB-seq; 5fC DIP, fC-Seal, and 5fC pull-down), and 5caC (caCAB-seq; 5caC DIP) (see Box 2). Notably, oxBS-seq, fCAB-seq, and caCAB-seq require subtracting signals of conventional BS-seq from those of modified BS-seq to indirectly determine the position and abundance of oxidized 5mC bases. 5fC and 5caC have not been systematically mapped at base-resolution, as they are relatively rare in the wild-type cells and require unusually high sequencing coverage to confidently determine their position and abundance.

(B) Shown are genomic distributions of oxidized 5mC bases at representative loci in mouse ESCs. Genomic distribution of 5hmC/5fC/5caC in control (WT: wild-type; Ctrl kd: control knockdown) and *Tdg*-depleted (ko: knockout; Tdg kd: *Tdg* knockdown) mouse ESCs are shown for three representative loci. 5hmC distribution was measured by both affinity-based method (5hmC DIP, CMS, hMe-Seal, and GLIB) (Pastor et al., 2011; Shen et al., 2013; Song et al., 2013a) and base-resolution method (TAB-seq, scale: 0%–50%) (Yu et al., 2012). Distributions of 5fC and 5caC were determined by antibody (5fC/5caC DIP) (Shen et al., 2013) or chemical tagging (fC-Seal) (Song et al., 2013a) methods. DNA methylation levels were estimated by 5mC DIP (Shen et al., 2013) or conventional BS-seq (5mC+5hmC, scale: 0%–100%) in wild-type mouse ESCs (Stadler et al., 2011). Based on local methylation levels, the genome can be categorized into FMRs (fully methylated regions, 50%–100% methylation), LMRs (low methylated regions, 10%–50%), and UMRs (unmethylated regions, < 10%) (Stadler et al., 2011). Also shown are other genomic data sets in wild-type mouse ESCs, including Tet1 ChIP-seq (Wu et al., 2011b), RNA-seq (Ficz et al., 2011), major histone modifications (Mikkelsen et al., 2007), DNase I hypersensitivity site (modENCODE project), and binding sites of pluripotency TFs (Whyte et al., 2012). A region located upstream of *Zfp281* gene is highlighted in gray, where *Tdg*-deficiency induced ectopic 5caC is overlapped with chromatin features of active enhancer (DNase I<sup>+</sup>/H3K4me1<sup>+</sup>/H3K4me3<sup>+</sup>) and binding sites of pluripotency factors (Oct4/Nanog/Sox2).

proximal regions surrounding the transcriptional start site (TSS) of active promoters (mostly HCPs) are generally devoid of 5hmC/5fC/5caC in both wild-type and *Tdg*-depleted mouse ESCs (exemplified by promoters of *Rest* and *Zfp281* in Figure 3B, overlapping with unmethylated regions [UMR]) (Shen et al., 2013; Song et al., 2013a; Szulwach et al., 2011a; Wu et al., 2011a; Yu et al., 2012). TET proteins enriched at

CpG-rich active promoters may therefore facilitate steady-state transcription through catalytic activity-independent functions such as stabilizing or recruiting other chromatin-modifying complexes (Box 3).

Affinity-based methods showed that 5hmC is preferentially enriched at promoters of genes expressed at medium-to-low levels in mouse and human ESCs (Pastor et al., 2011; Szulwach et al., 2011a; Wu et al., 2011a; Xu et al., 2011b), and these 5hmC-enriched promoters are generally associated with intermediate

## Box 2. Genome-wide Mapping Technologies for 5mC Oxidation Products

### A. Affinity-Enrichment Methods

Affinity-enrichment is achieved by using either modification-specific antibodies or chemical-tagging-specific 5mC oxidation derivatives (Figure 3A).

1. **Antibody-Based Immunoprecipitation (IP) of 5hmC/5fC/5caC**  
Modification-specific antibodies are available for all three oxidized 5mC bases and the IP procedure is simple (Song et al., 2012b; Wu and Zhang, 2011a). However, there is evidence that antibody-based method may pull-down certain type of simple repeats (e.g., poly-CA) nonspecifically (Matarese et al., 2011). Subtracting IgG control signals from IP signals can partially alleviate this problem (Shen et al., 2013).
2. **CMS Antibody-Based IP of 5hmC**  
This method is based on sodium bisulfite-mediated conversion of 5hmC to cytosine methylene sulphonate (CMS). Highly specific antibody against CMS can be generated and used to enrich 5hmC-containing DNA fragment (Huang et al., 2012). Anti-CMS serum may have lower background compared to anti-5hmC serum (Pastor et al., 2011). However, the presence of bulky CMS moiety in DNA can introduce PCR biases.
3. **JBP1 IP of 5hmC**  
The method involves initial conversion of 5hmC to glycosylated 5hmC (g5hmC) using  $\beta$ -glucosyltransferase ( $\beta$ -GT) from T4 bacteriophage. JBP1, a natural g5hmC-binding protein, is then used to enrich g5hmC-containing genomic DNA (Robertson et al., 2011). However, a recent study suggests that JBP1 IP generally produces weaker signals compared to 5hmC antibody IP or hMe-Seal (Thomson et al., 2013).
4. **Selective Chemical Labeling 5hmC with Biotin (hMe-Seal or GLIB)**  
By exploiting the naturally occurring glucosylation process of 5hmC by  $\beta$ -GT, several methods were developed to chemically and enzymatically label 5hmC with the biotin tag (Pastor et al., 2011; Song et al., 2011; Szwagierczak et al., 2010). The highly selective and strong interaction between biotin and streptavidin allows efficient pull-down of 5hmC-containing DNA. In the first method, termed hMe-Seal (Song et al., 2011), purified  $\beta$ -GT is used to transfer an azide-modified glucose analog (UDP-6-N<sub>3</sub>-glucose) to 5hmC. Then, a biotin group can be attached to g5hmC with alkyne-biotin probe via click chemistry (Song et al., 2011). The second approach, named GLIB (glucosylation, periodate oxidation, biotinylation), first uses  $\beta$ -GT to transfer an unmodified glucose to 5hmC. Next, sodium periodate (NaIO<sub>4</sub>) is used to oxidize the vicinal hydroxyl group in the glucose to aldehyde group, which is further modified with aldehyde-reactive probe (aminooxy-biotin or NH<sub>2</sub>O-biotin) to add biotin tag to g5hmC (Pastor et al., 2011). These chemical labeling methods have enhanced sensitivity and lower background as compared to the antibody-based method (Song et al., 2012b). However, similar to 5hmC antibody IP, these methods still exhibit some degree of density-dependency and tend to more efficiently pull-down DNA fragment containing multiple modified bases (Shen et al., 2013).
5. **Selective Chemical Labeling 5fC with Biotin (fC-Seal or 5fC Pull-Down)**  
A strategy for direct labeling 5fC with biotin tag using aminooxy-biotin probes has been suggested (Pfaffeneder et al., 2011) and was used to generate the genome-wide distribution map of 5fC (Raiber et al., 2012). However, this method may have side reactions with other cytosines or abasic sites, resulting in relatively high background. Based on hMe-Seal, an indirect 5fC labeling

## Box 2. Continued

method (termed fC-Seal) has been developed (Song et al., 2013a). In fC-Seal,  $\beta$ -GT is first used to block endogenous 5hmC in the genome with unmodified glucose. Next, sodium borohydride (NaBH<sub>4</sub>) is used to reduce 5fC to 5hmC, which is further converted to g5hmC-biotin (derived from endogenous 5fC) as in hMe-Seal.

### B. Single-Nucleotide Resolution Methods

Base-resolution mapping of 5mC oxidation derivatives involves the use of bisulfite sequencing (BS-seq) or third-generation sequencing platforms.

1. **BS-Seq (5mC+5hmC)**  
In standard BS-seq, 5mC and 5hmC are resistant to deamination and consequently be read out as C after PCR amplification (Clark et al., 1994). In contrast, unmodified C, 5fC and 5caC are deaminated in bisulfite treatment and read out as T in sequencing. Thus, methylation signals (C) in standard BS-seq represent the sum of 5mC and 5hmC, complicating the interpretation of BS-seq results (Huang et al., 2010). By exploiting distinct chemical properties of 5hmC/5fC/5caC in bisulfite treatment (NaHSO<sub>3</sub>), traditional BS-seq can be adapted to map oxidized 5mC (Figure 3A).
2. **oxBS-Seq (5hmC)**  
In oxidative bisulfite sequencing (oxBS-seq) (Booth et al., 2012), potassium perruthenate (K<sub>2</sub>ReO<sub>4</sub>) was used to specifically oxidize 5hmC to 5fC, which can be deaminated by bisulfite treatment. After the DNA sample is treated with K<sub>2</sub>ReO<sub>4</sub> and NaHSO<sub>3</sub>, 5hmC/5fC/5caC is detected as T in sequencing, whereas 5mC remains intact and is read as C. Therefore, subtracting signals of oxBS-seq maps from standard BS-seq signals would reveal the location and abundance of 5hmC (Booth et al., 2012).
3. **TAB-Seq (5hmC)**  
In TET-assisted bisulfite sequencing (TAB-seq) (Yu et al., 2012), 5hmC is first protected from TET-mediated oxidation by glucosylation using  $\beta$ -GT. Next, 5mC and 5fC are oxidized to 5caC in the presence of highly active recombinant TET proteins.
4. **fCAB-Seq (5fC)**  
In this chemical modification assisted BS-seq (CAB) method, O-ethylhydroxylamine (EtONH<sub>2</sub>) is used to prevent 5fC from bisulfite-mediated deamination (Song et al., 2013a). Subtracting standard BS-seq signals from signals of fCAB-seq will reveal the location and abundance of 5fC.
5. **caCAB-Seq (5caC)**  
Similar to fCAB-seq, 5caC can be modified by 1-ethyl-3-[3-dimethylaminopropyl]-carbodiimide hydrochloride (EDC) to resist deamination (Lu et al., 2013). However, this method may have side reactions with the genomic DNA.
6. **Third-Generation Sequencing Technology**  
These single-molecule sequencing technologies directly sequence DNA molecules without the need of PCR amplification. In single-molecule real-time (SMRT) sequencing, real-time recording of nucleotide incorporations generates not only the sequence identity but also DNA polymerase kinetics. The polymerase replication rate at and around modified cytosines is slower compared to unmodified bases. When combined with chemical labeling or enzymatic modification (adding a glucosyl group to 5hmC; oxidizing 5mC to 5caC) to increase kinetic differences, SMRT can be used to detect strand-specific 5mC and oxidized bases at base-resolution with high confidence (Clark et al., 2013; Song et al., 2012a). Nanopore sequencing monitors the current signal when individual DNA molecular pass through nanoscale protein pores and has also shown the potential to detect oxidized 5mC bases (Li et al., 2013; Wanunu et al., 2011). While third

(Continued on next page)

**Box 2. Continued**

generation sequencing technologies have the potential to directly map various cytosine modifications, significant technological advances are required to further improve the detection accuracy and to increase the sequencing throughput for analyzing complex mammalian genomes.

or low CpG-density (intermediate-CpG promoter [ICP] or low-CpG promoter [LCP]). Base-resolution mapping of 5hmC also indicated that 5hmC is more abundant at ICPs/LCPs than at HCPs (Yu et al., 2012). Recent studies showed that 5fC/5caC tend to accumulate at transcriptionally inactive or poised promoters in the absence of TDG in mouse ESCs (Shen et al., 2013; Song et al., 2013a). In fact, some of the highest levels of 5hmC/5fC/5caC in *Tdg*-depleted mouse ESCs are found at promoters and flanking regions of lineage-specific transcription factor (mostly ICPs) (*Tbx5* in Figure 3B) (Shen et al., 2013), indicating that TET/TDG-mediated active demethylation occurs at these loci. 5hmC/5fC/5caC peaks at or surrounding these ICPs of lineage-specific genes are frequently overlapped with low-methylation regions (LMRs) that are previously identified by standard BS-seq as partially methylated regions (Stadler et al., 2011), as well as bivalent domains, which contain H3K4me3 (active histone mark) and H3K27me3 (polycomb repression complex 2 [PRC2]-deposited repressive mark) (Bernstein et al., 2006). TET proteins and oxidized methylcytosines may contribute to the establishment and/or maintenance of transcriptionally poised/inactive state at lineage-specific promoters through several mechanisms, including maintenance of partially methylation state by mediating active demethylation, recruitment of TET-interacting corepressor complexes such as SIN3A (Box 3), and attracting/repelling specific reader proteins for oxidized 5mC bases (see below). Further studies are needed to clarify the relative importance of these distinct mechanisms in gene regulation.

**Oxidized Methylcytosines within Gene Bodies**

Enrichment of 5hmC within intragenic regions, particularly at 3' portion of gene bodies, is found to be a hallmark of active transcription in virtually all cell-types investigated, including mouse/human ESCs (Ficz et al., 2011; Pastor et al., 2011; Szulwach et al., 2011a; Williams et al., 2011; Wu et al., 2011a; Xu et al., 2011b) (Rest in Figure 3B), mouse cerebellum (Mellén et al., 2012; Song et al., 2011), and mouse/human frontal cortex (Lister et al., 2013). Despite their scarcity in the bulk genome, 5fC and 5caC also exhibit a tendency to accumulate within actively transcribed gene bodies in the absence of TDG (Shen et al., 2013; Song et al., 2013a) (Rest in Figure 3B). The exact role of intragenic 5hmC/5fC/5caC in transcriptional regulation is currently unclear. One possible function of intragenic 5fC/5caC is to regulate the rate of RNA polymerase II transcription (Kellinger et al., 2012). In mouse ESCs, 5hmC/5fC/5caC is also more enriched at exons relative to introns, raising the possibility that 5mC oxidation may regulate mRNA processing events such as splicing.

**Oxidized Methylcytosines at Distal cis-Regulatory Elements**

Both affinity enrichment-based and base-resolution mapping analyses indicate that 5hmC is highly enriched at active distal-

**Box 3. Catalytic Activity-Independent Role of TET Proteins in Transcriptional Regulation**

Genomic mapping of TET proteins in ESCs, HEK293T cells and bone marrow tissues suggest that all three TET proteins are preferentially localized to unmethylated CpG-rich promoters (Chen et al., 2013b; Deplus et al., 2013; Vella et al., 2013; Williams et al., 2011; Wu et al., 2011b). In mouse ESCs, TET1 is highly enriched at actively transcribed CpG-rich promoters as well as bivalent gene promoters, which are repressed by Polycomb repression complex 2 (PRC2). The preference for CGIs is likely due to its N-terminal CXXC domains. Depletion of TET1 causes transcriptional activation and repression of many direct TET1 targets (Williams et al., 2011; Wu et al., 2011b; Xu et al., 2011b). These results suggest that TET proteins may have dual functions in transcriptional regulation via distinct mechanisms. The transcriptional activation role of TET1 may be dependent on its 5mC oxidation activity to maintain unmethylated states at active promoters or distal enhancers (Ficz et al., 2011; Shen et al., 2013; Wu et al., 2011b); in contrast, a role of TET1 in transcriptional repression may be related to its physical interaction with the SIN3A repressor complex (Williams et al., 2011), or its ability to facilitate the recruitment of the MBD3/NURD (Yildirim et al., 2011) and PRC2 (Wu et al., 2011b) to chromatin. TET2 and TET3 (to a lesser extent for TET1), can directly interact with O-linked  $\beta$ -N-acetylglucosamine (O-GlcNAc) transferase (OGT), and recruit OGT to CpG-rich gene promoters in mouse ESCs and differentiated cells (Chen et al., 2013b; Deplus et al., 2013; Vella et al., 2013). Although this physical interaction appears not to regulate enzymatic activity of TET proteins, depletion of TET proteins reduces OGT binding to chromatin and its activity to glycosylates histone H2B (Chen et al., 2013b) and other chromatin regulators such as HCF1 (host cell factor 1) (Deplus et al., 2013), a component of the H3K4 methyltransferase SET1/COMPASS complexes. Thus, TET proteins may promote transcription of target genes by enhancing histone modifications such as O-GlcNAcylation at Ser 112 of H2B and trimethylation of Lys 4 on H3 (H3K4me3), both of which are associated with gene activation. Together, these results suggest that TET proteins possess catalytic activity-independent gene regulatory functions.

regulatory elements (Stroud et al., 2011; Yu et al., 2012), which are marked by DNase I hypersensitivity, enhancer-associating histone modifications (H3K4me1 and H3K27ac) (Shen et al., 2012) and LMRs. Interestingly, base-resolution mapping indicates that 5hmC is actually depleted at the binding motif of transcription factors or CTCF insulators but enriched at immediately adjacent regions (Lister et al., 2013; Yu et al., 2012). Furthermore, base-resolution mapping of 5mC and 5hmC during human and mouse frontal cortex development showed that for adult-brain-specific enhancers (active in adult brains but inactive in fetal brains), both 5mC and 5hmC are enriched at the center regions of fetal brains but are lost in adult brains (Lister et al., 2013), supporting a role of TET-mediated 5mC oxidation in enhancer activation in postnatal brains.

In mouse ESCs, 5hmC and 5fC tend to enrich at poised tissue-specific enhancers (H3K4me1-positive, but H3K27ac-negative) (Shen et al., 2013; Song et al., 2013a), indicating that these poised enhancers are premarked by 5hmC/5fC for subsequent demethylation during later development. Furthermore, 5caC (to a lesser extent for 5fC) is preferentially accumulated at distally located active enhancers in *Tdg*-depleted mouse ESCs (Shen et al., 2013; Song et al., 2013a). For instance, 5caC is enriched



at a cohort of distal pluripotency TF (Oct4/Nanog/Sox2)-binding sites in *Tdg*-depleted mouse ESCs (*Zfp281* in Figure 3B). These results suggest that TET/TDG-mediated active demethylation is preferentially targeted to these active distal-regulatory elements. DNA binding proteins are capable of inducing localized DNA demethylation (Stadler et al., 2011), but it is currently unclear how TET/TDG activity is recruited to these *cis*-regulatory elements. TET and TDG may be recruited directly by transcription factors or coactivators (CBP/p300). Alternatively, transcription factor binding may first facilitate promoter-enhancer interaction via looping, which allows promoter-bound TET/TDG proteins to access distal-regulatory elements.

#### **Oxidized 5mC Bases as Potential Epigenetic Marks**

It is known that specific protein domains exist to recognize unmodified, hemi- or fully methylated CpG sites. For example, unmethylated CpG sites (CG:GC) can be specifically recognized by CXXC-domain-containing proteins (Long et al., 2013), whereas the SRA domain in UHRF1 and methyl-CpG-binding domains (MBD) can specifically recognize hemimethylated (<sup>m</sup>CG:GC) and symmetrically methylated (<sup>m</sup>CG:GC<sup>m</sup>) CpG dyads, respectively (Hashimoto et al., 2010). Besides acting as intermediates of active DNA demethylation, growing evidence suggests that 5hmC is relatively stable in the genome and may potentially influence transcription by attracting or repelling specific DNA-binding proteins. There is evidence that several MBD proteins, including MBD3 (Yildirim et al., 2011), MeCP2 (Mellén et al., 2012), and MBD4 (Otani et al., 2013a), may also bind to hemi- or fully hydroxymethylated CpG sites, suggesting that these proteins have dual binding capacity for 5mC and 5hmC. Recently, mass spectrometry-based unbiased analyses of nuclear proteins in mouse ESCs, neural progenitor cells (NPC), and adult brains have identified a large number of candidate proteins bound selectively to or repelled by 5hmC-containing oligonucleotides (Spruijt et al., 2013). While some 5hmC interacting protein candidates (e.g., WDR76, THY28) are ubiquitously expressed, many of them are expressed in a cell-type-specific manner. For instance, 5hmC-specific reader protein UHRF2 and THAP11 are preferentially enriched in NPCs and adult brains, respectively, suggesting that 5hmC may mediate cell-type-specific functions by recruiting distinct reader proteins.

Despite their scarcity in the genome, 5fC and 5caC possess unique chemical groups for protein recognition, raising the possibility that they may act as dynamic/transient epigenetic marks (Song and He, 2013). For example, TDG and a number of DNA repair proteins were found to specifically bind to 5fC- and 5caC-containing DNA oligonucleotides (Spruijt et al., 2013). Another study also identified many candidate proteins from mouse ESC nuclear extract that specifically bind to 5fC, including transcription factors, DNA repair proteins, and chromatin regulators (Iurlaro et al., 2013). However, these studies cannot distinguish between proteins that bind directly and indirectly to 5fC/5caC, and some candidates may simply bind to formyl or carboxyl groups nonspecifically. In addition to the possibility of acting as epigenetic marks, 5fC/5caC may contribute to transcriptional regulation through eliciting “scheduled” DNA repair response at specific gene regulatory elements (Fong et al., 2013).

#### **Biological Functions of DNA Demethylation in Diverse Biological Processes**

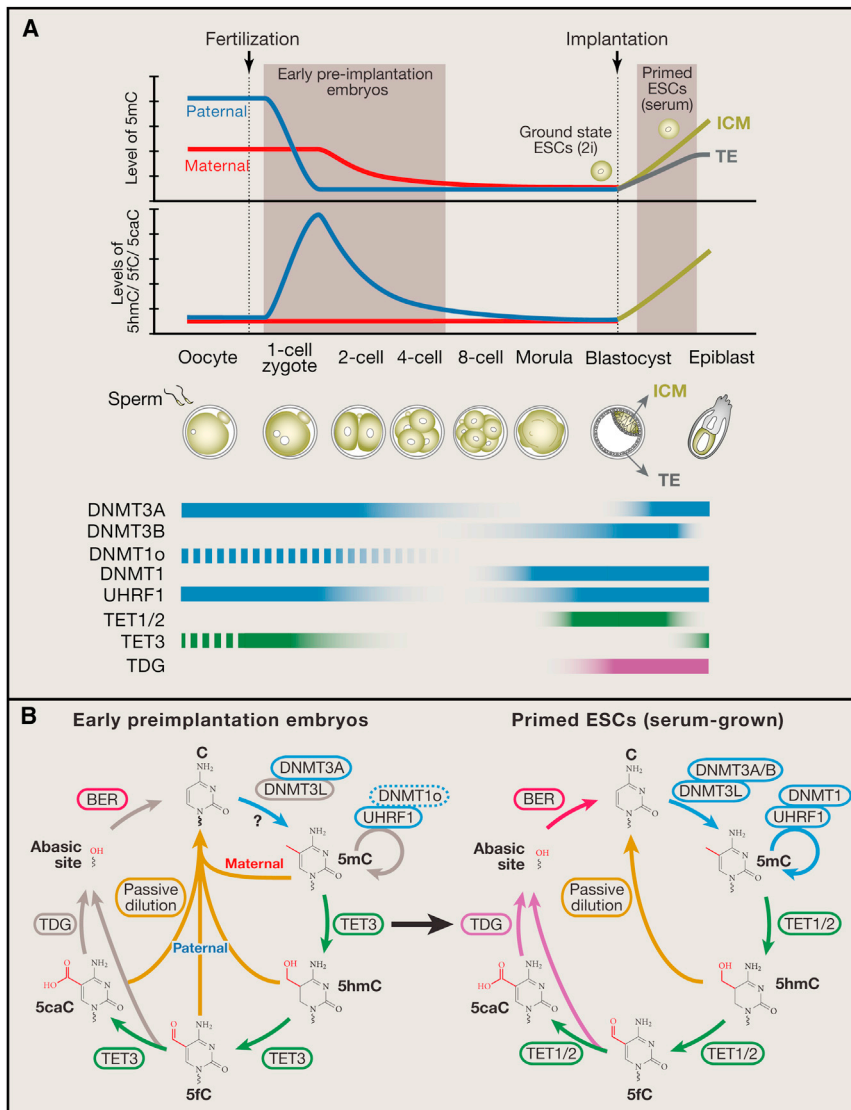
Global erasure of 5mC is known to occur in specific stages of mammalian development, including early preimplantation embryos and developing germ cells. A better understanding of the biochemical basis of DNA demethylation pathway and emergence of new technologies for mapping various cytosine modifications have opened new avenues to study the role of DNA demethylation in these biological processes. Intriguingly, the broad expression patterns of TET enzymes and other components of active DNA demethylation pathway in somatic tissues such as brains imply that dynamic DNA demethylation may be more prevalent than previously thought. Below, we review recent advances in our understanding on biological functions of DNA demethylation in various biological and pathological settings.

#### **DNA Demethylation Dynamics in Preimplantation Development**

After the sperm fertilizes the oocyte and before two pronuclei merge, the paternal genome goes through a complex epigenetic remodeling process that includes global erasure of DNA methylation. The sperm undergoes a rapid loss of 5mC before the onset of first DNA replication as measured by immunostaining and locus-specific bisulfite sequencing (Mayer et al., 2000; Oswald et al., 2000); by contrast, the maternal genome is not affected in one-cell zygotes, but become gradually demethylated during subsequent cleavage divisions. Recent studies indicate that the loss of paternal DNA methylation coincides with a rapid increase in 5hmC (Inoue and Zhang, 2011; Iqbal et al., 2011; Wossidlo et al., 2011) and 5fC/5caC (Inoue et al., 2011), suggesting that TET-mediated 5mC oxidation is involved in the erasure of 5mC. Knockdown or deletion of TET3, the only 5mC oxidase present at this stage, abolished the loss of 5mC and 5hmC generation in the male pronucleus (Gu et al., 2011; Wossidlo et al., 2011). *Tet3*-deficient oocytes fertilized with wild-type sperms can develop normally to blastocysts, but many of these heterozygous embryos exhibit developmental failure (Gu et al., 2011), indicating that TET3-mediated rapid and specific oxidation of 5mC from paternal genome is biologically relevant. It will be interesting to identify which regions in the paternal genome are preferentially targeted by TET3 and to determine why 5mC oxidation of paternal genome is functionally important for early embryogenesis.

The striking asymmetry of epigenetic reprogramming between paternal and maternal genomes in one-cell zygotes can potentially be achieved through either specific targeting of TET3 to paternal genome or active protection of maternal genome from TET3-mediated 5mC oxidation. A recent study indicates that Stella (also known as PGC7 or Dppa3), a maternal factor essential for early development (Payer et al., 2003), is recruited to oocyte-derived genome by dimethylation of Lys9 on H3 (H3K9me2), a histone mark that is preferentially enriched on maternal chromatin (Nakamura et al., 2012). In the absence of Stella, TET3 is localized to both paternal and maternal genomes (Nakamura et al., 2012; Wossidlo et al., 2011), and the maternal genome also undergoes 5mC oxidation (Nakamura et al., 2007; Nakamura et al., 2012). Locus-specific bisulfite sequencing analysis further demonstrated that Stella is required for protecting a





**Figure 4. DNA Methylation Dynamics during Mouse Preimplantation Development**

(A) Dynamic changes in cytosine modifications and relevant enzymes during preimplantation development. Immediately after fertilization, paternal 5mC is rapidly oxidized to 5hmC/5fC/5caC by TET3 proteins. In early preimplantation embryos, oocyte-derived DNMT1o is largely excluded from nucleus (dash line) and consequently maintenance methylation is inefficient. Oxidized 5mC bases in the paternal genome and 5mC in the maternal genome are therefore passively diluted. The global 5mC level reaches the lowest point around blastocyst stage (E3.5). After implantation, DNA methylation pattern is re-established by DNMT3A/3B in inner cell mass (ICM) cells, but not in trophectoderm (TE) cells. The genome of ground-state ESC (2i-condition) is hypomethylated (5mC: 1% of all C) and is more similar to the methylome of preimplantation ICM cells, whereas primed ESCs (serum) possess a methylome (5mC: 4% of all C) that recapitulates overall methylation pattern in epiblast cells.

(B) Developmental stage-specific usage of the cyclic cytosine-modifying enzymatic cascade during early embryonic development. Upon fertilization, oocyte-derived TET3 proteins specifically localize to paternal nucleus and convert sperm-derived 5mC to 5hmC/5fC/5caC. DNA replication drives passive loss of 5hmC/5fC/5caC in the paternal genome and 5mC in the maternal genome (left). In primed ESCs and cells during ICM-to-epiblast transition *in vivo*, all enzymes of the cytosine-modifying cascade are expressed at relatively high levels. Dnmt3A/3B/3L and DNMT1/UHRF1 establish and maintain 5mC patterns. At some genomic regions (e.g., active enhancers and bivalent promoters) in these cells, TET1/2 oxidize 5mC to 5hmC/5fC/5caC, and TDG/BER excises 5fC/5caC to regenerate unmodified cytosine.

2003; Smallwood et al., 2011; Smith et al., 2012). Recent evidence suggests that maintenance of 5mC at ICRs is perhaps conferred through targeting DNMT1 and/or DNMT3A to these regions by zinc finger transcription factor ZFP57 and its partner, KAP1 (also known

as TRIM28) (Li et al., 2008; Messerschmidt et al., 2012; Quenneville et al., 2011). In line with the role of this complex in maintenance of 5mC at imprinting loci, ZFP57 prefers to bind to its target sequence (TGC<sup>m</sup>CGC, present in all known murine ICRs) when it is methylated, but has reduced binding affinity when its target sequence is unmodified or hydroxymethylated DNA (Liu et al., 2012).

If inefficient maintenance of DNA methylation is a major driver for the genome-wide erasure of 5mC in early embryos (Figure 4B), why the paternal genome needs to be oxidized first? One possible reason might be to rapidly equalize 5mC levels of the two parental genomes as the CpG methylation level differs significantly between sperms (~90%) and oocytes (~40%) before fertilization (Kobayashi et al., 2012). Thus, oxidized 5mC bases in paternal genome may accelerate the demethylation process of paternal genome. 5fC and 5caC generated from further oxidation of 5hmC provide a substrate for TDG-mediated

excision repair pathway (Inoue et al., 2011), and there is some evidence supporting a role of BER in paternal genome demethylation (Hajkova et al., 2010). However, bulk 5fC and 5caC signals in the paternal genome remain detectable by immunostaining in 4-cell stage and also exhibit replication-dependent dilution (Inoue et al., 2011), suggesting that 5fC/5caC are not rapidly excised. In support of this view, single-cell RNA-seq analysis suggests that the *Tdg* gene is expressed at extremely low levels in oocytes and during preimplantation development (Tang et al., 2011) (Figure 4A), indicating that TDG may only play a limited role in paternal genome demethylation (e.g., locus-specific demethylation). Alternatively, after DNA replication, CpG sites with hemimodified state (XG:GC, X = <sup>h</sup>C, <sup>1</sup>C, <sup>ca</sup>C) in the paternal genome may be more resistant to enzymatic activity of maternal DNMT3A or residual DNMT1o in the nucleus than hemimethylated CpG sites (<sup>m</sup>CG:GC), thereby ensuring more efficient paternal demethylation.

Overall, DNA demethylation in early embryos is a highly orchestrated process, requiring both active and passive demethylation. In addition, specific factors, such as Stella and ZFP57/KAP1, may protect 5mC patterns at selective loci (e.g., ICRs) from either active (TET3-mediated 5mC oxidation) or passive (inefficient methylation maintenance) demethylation. Interestingly, in contrast to the situation in mice, recent methylome analyses in zebrafish suggest that paternal methylation pattern is maintained during early embryogenesis; instead, the relatively hypomethylated maternal genome is progressively reprogrammed to a pattern similar to that of the sperm methylome (Jiang et al., 2013; Potok et al., 2013). Furthermore, 5mC oxidation is not detected during zebrafish early embryo development (Jiang et al., 2013). Thus, 5mC oxidation-mediated paternal genome demethylation seems to be a unique regulatory mechanism for mammalian embryogenesis. The evolutionary significance of this mechanism awaits further investigation.

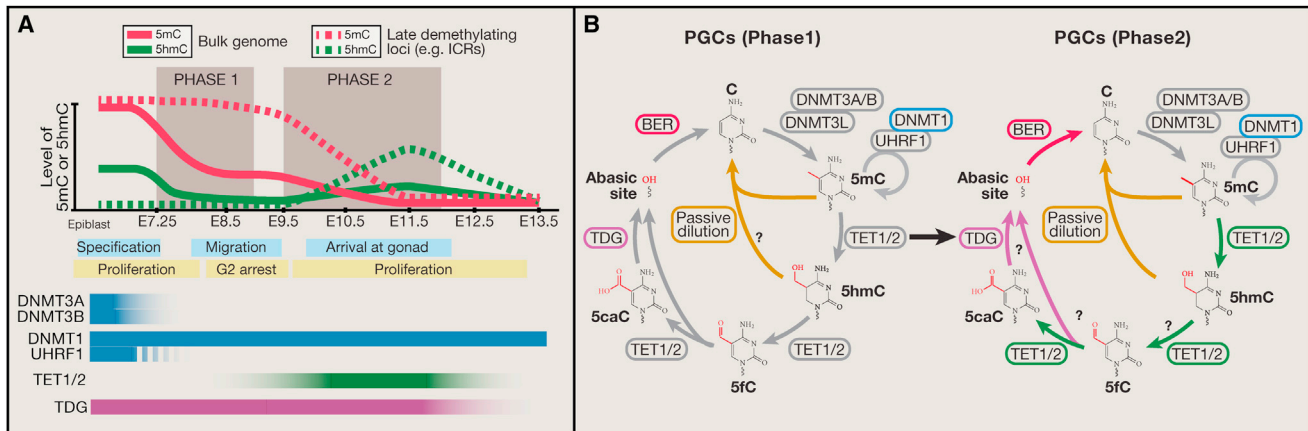
#### **DNA Demethylation in Pluripotent Stem Cells and Reprogramming**

As the zygote develops into the blastocyst (E3.5), the genome of inner cell mass (ICM) cells becomes hypomethylated (~21% CpG methylation) as compared to sperm (90%) or oocyte (40%) methylome after replication-dependent dilution of maternal 5mC and paternal 5hmC/5fC/5caC (Kobayashi et al., 2012). Thereafter, DNMT3A/3B mediated de novo methylation takes places during the developmental transition from blastocyst to epiblast stage (70%) by E6.5 (Seisenberger et al., 2012; Smith et al., 2012) (Figure 4A). ESCs are initially derived from the ICM of blastocysts, and these pluripotent stem cells possess unique properties of self-renewing while maintaining potential to differentiate into any somatic cell lineages (Hanna et al., 2010). Traditionally, mouse ESCs are cultured in serum supplemented with leukemia inhibitory factor (LIF) and their methylomes resemble those of somatic cells (70% CpG methylation). Mouse ESCs grown in serum exhibit greater heterogeneity and consist of at least two distinguishable populations, one with a transcriptome similar to the naive ICM-like state and the other similar to the primed epiblast-like state.

Inhibition of the mitogen-activated protein (MAPK) pathway and the glycogen synthase kinase (GSK3) pathway, via two kinase inhibitors (PD0325901 for MAPK and CHIR99021 for

GSK3, known as 2i condition), is sufficient to maintain mouse ESCs in the naive ICM-like state (Ying et al., 2008). Mouse ESCs grown in the 2i condition appear to be a homogeneous population that has a transcriptome and histone modification profile distinct from that of serum-grown mouse ESCs (Marks et al., 2012). Using mass spectrometry or whole-genome BS-seq (WGBS), recent studies showed that 2i-grown mouse ESCs adopt a globally hypomethylated genome recapitulating that of ICM cells in vivo (Ficz et al., 2013; Habibi et al., 2013; Leitch et al., 2013). Strikingly, switching from serum to 2i condition induces genome-wide DNA demethylation in mouse ESCs, which is on a scale and pattern similar to the early embryos. While the maintenance methylation system (DNMT1/UHRF1) and TET1/2 are expressed at comparable levels between ground state (2i) and primed state (serum) mouse ESCs, PR-domain-containing 14 (PRDM14) is upregulated in 2i-grown mouse ESCs and mediates rapid transcriptional downregulation of the de novo DNA methylation machinery (DNMT3A/3B/3L) (Yamaji et al., 2013). Available evidence suggests that global demethylation during the serum-to-2i transition is primarily induced by the downregulation of DNMT3A/3B/3L and TET1/2-mediated transient 5mC oxidation (Blaschke et al., 2013; Ficz et al., 2013; Habibi et al., 2013). Overall, globally hypomethylated genomes observed in ground state mouse ESCs (2i-condition) closely resemble the methylome of ICM cells in preimplantation blastocysts, whereas serum-grown mouse ESCs have a hypermethylated genome similar to that of postimplantation embryos and somatic cells.

In serum-grown mouse ESCs, all major components of the cytosine-modifying cascade (DNMT, TET, TDG, and BER proteins) are expressed (Figure 4B). Thus, ESCs in primed state provide a model system for understanding DNA demethylation dynamics and its regulatory mechanisms. Single or combined depletion of *Tet1* and *Tet2* in mouse ESCs, by short-hairpin RNA (shRNA)-mediated knockdown or genetic inactivation, produced variable in vitro phenotypes (Dawlaty et al., 2013; Dawlaty et al., 2011; Ficz et al., 2011; Freudenberger et al., 2012; Ito et al., 2010; Koh et al., 2011; Williams et al., 2011). Multiple studies confirmed that depletion of TET1 alone or in combination with TET2 skews in vitro differentiation toward trophoblast and primitive endoderm, and teratomas derived from *Tet1*<sup>-/-</sup> or *Tet1/2*<sup>-/-</sup> ESCs are enriched for trophoblast cells. In support of a role of TET1 in regulating early lineage commitment, TET1 negatively regulates the expression of key regulators of trophoblast lineage, such as *Cdx2* and *Eomes*, possibly by facilitating binding of repressor complexes to target gene promoters (Williams et al., 2011; Wu et al., 2011b). However, there are conflicting results concerning the role of TET1 in the maintenance of mouse ESCs (Dawlaty et al., 2011; Freudenberger et al., 2012; Ito et al., 2010; Williams et al., 2011; Wu and Zhang, 2011b). The use of different cell lines and culturing conditions may contribute to these discrepancies, although potential off-target effects of specific shRNAs cannot be ruled out. Because all these studies focused on serum-grown ESCs, fluctuating epigenetic state associated with this heterogeneous ESC population may also contribute to the variable phenotype in *Tet1*-depleted mouse ESCs. Overall, current data indicate that, despite the presence of relatively high level of TET1/2 proteins and oxidized



**Figure 5. Global Reprogramming of DNA Methylation Patterns during Early PGC Development**

(A) Dynamic changes in cytosine modifications and expression levels of relevant enzymes during early stage of mouse PGC development. Global erasure of DNA methylation in developing PGCs goes through two phases. In the first phase (E7.25 to E9.0), both de novo and maintenance methylation machineries are functionally impaired due to the repression of DNMT3A/3B/3L and UHRF1. There is also evidence that remaining UHRF1 is excluded from nucleus (dash line). Thus, loss of bulk genomic DNA methylation (solid line, inherited from epiblast cell stage) takes place in a replication-dependent manner in phase 1. However, a portion of the genome (dash line) including ICRs and germline-specific genes are partially or fully protected from the first wave of passive demethylation. In the second phase (E9.5 to E12.0), TET1 (possibly TET2) are upregulated and oxidize remaining 5mC to 5hmC. In phase 2, TET1/2 proteins oxidize remaining 5mC at late demethylating loci. Replication-dependent loss of 5hmC completes the global demethylation process. 5fC/5caC is not specifically enriched in developing PGCs as compared to surrounding somatic cells, so they may be actively excised by TDG/BER pathway that is present in PGCs.

methylcytosines in ESCs, TET enzymes are largely dispensable for ESC maintenance but may play a role in guiding ESCs to properly differentiate into specific lineages and/or regulating later development of the embryos. Indeed, while *Tet1*<sup>-/-</sup> and *Tet2*<sup>-/-</sup> single-mutant mice are viable, roughly half of *Tet1*<sup>-/-</sup>*Tet2*<sup>-/-</sup> double knockout (DKO) mice from intercrossing double-heterozygous animals die perinatally with severe developmental abnormalities, indicating a role of TET1/2 proteins in regulating embryonic development (Dawlaty et al., 2013). The partially penetrant perinatal lethality of TET1/2 DKO mutants suggests TET3 may compensate for the loss of TET1/2 during development. In line with this notion, all *Tet3*<sup>-/-</sup> single mutants die perinatally (Gu et al., 2011).

While additional studies are needed to clarify the exact function of TET proteins in mouse ESCs and embryonic development, recent studies support a role for TET enzymes in transcription factor (Oct4/Klf4/Sox2/c-Myc, referred to as OKSM)-mediated reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) (Hanna et al., 2010). One study showed that TET2 is recruited to the *Nanog* and *Esrrb* loci (both are pluripotency-related TFs) to facilitate their transcription in the early stage of reprogramming (Doege et al., 2012). In addition, TET1 and TET2 may interact with Nanog protein and promote iPSC generation in a TET catalytic activity-dependent manner (Costa et al., 2013). More strikingly, TET1 overexpression not only promotes transcriptional reactivation of *Oct4* but can also replace Oct4 in the reprogramming OKSM cocktail (Gao et al., 2013). A recent study also implicates a role of vitamin C in modulating TET functions during iPSC generation (Chen et al., 2013a).

In addition to TET proteins, TDG may also play a role in stem cell functions and embryonic development. TDG is broadly

expressed during early development and in adult tissues. Depletion of *Tdg* by shRNA or genetic inactivation in mouse ESCs does not cause any discernible phenotype in ESC proliferation or early differentiation (Cortázar et al., 2011; Shen et al., 2013; Song et al., 2013a), suggesting that TDG may regulate later steps of embryonic development. Indeed, *Tdg*-deficiency in mice causes widespread development deficits and embryonic lethality (Cortázar et al., 2011; Cortellino et al., 2011). Further studies are needed to clarify the exact mechanism by which TDG contributes to developmental gene regulation.

#### DNA Demethylation Dynamics during Germ Cell Development

Around E6.5, a subset of posterior epiblast cells (~40 cells) are instructed by specific factors such as BLIMP1 (also known as PRDM1) and PRDM14 to become PGCs (Sasaki and Matsui, 2008). These committed PGCs suppress somatic transcriptional program and reactivate expression of pluripotency-associated factors. Around E7.5, PGCs start to migrate through the hindgut to the developing gonad, where they colonize by E10.5 to initiate epigenetic reprogramming (Figure 5A) (Saitou et al., 2012). Early PGCs (E7.25) are thought to inherit a methylation pattern similar to that of epiblast cells (~70% CpG methylation), but they undergo global erasure of DNA methylation and become nearly completely demethylated in gonadal PGCs around E13.5 (7%–14%) (Seisenberger et al., 2012).

Recent genome-wide DNA methylation studies suggest that global erasure of DNA methylation in PGCs takes place in two distinct phases (Seisenberger et al., 2012; Vincent et al., 2013). The first phase of demethylation (phase 1 in Figure 5A) occurs before or during PGC migration to genital ridge and appears to primarily involve replication-dependent passive dilution of 5mC. Upon PGC specification (E7.25), all components of de



novo methylation systems are transcriptionally downregulated (Kagiwada et al., 2013), possibly due to PRDM14-mediated repression. UHRF1 is also acutely downregulated in PGCs (Kagiwada et al., 2013; Seisenberger et al., 2012), which prevents localization of DNMT1 to replication foci. In the absence of efficient de novo and maintenance methylation, global erasure of DNA methylation occurs indiscriminately at nearly all genomic features (promoters, exons, introns, repeats and intergenic regions) in early replicating PGCs. By E9.5, the overall CpG methylation level reduces to ~30% in PGCs (Seisenberger et al., 2012). In line with this model, recent evidence suggests that TET proteins are largely dispensable for the initial global DNA demethylation (Figure 5B). Mass spectrometry and WGBS analysis of PGC-like cells differentiated in vitro from *Tet1/2*-depleted mouse ESCs (shRNA-mediated depletion of *Tet1* in *Tet2*<sup>-/-</sup> cells) indicate that global loss of 5mC in early PGCs still occurs in the absence of TET1/2 (Vincent et al., 2013). Furthermore, WGBS analysis of *Tet1* mutant PGCs derived from gene-trap female mice indicates that *Tet1*-deficiency does not affect global demethylation (Yamaguchi et al., 2012).

Methylation levels at some genomic regions, including ICRs, CpG-rich promoters of germ-cell-specific genes, CGIs on X chromosome genes, and specific type of repeats (e.g., IAPs), are largely maintained during the first phase of global demethylation in PGCs. Demethylation of these sequences is only completed once PGCs settle in gonad and is most likely initiated by TET1/2-mediated 5mC oxidation (phase 2 in Figure 5A). Expression of TET1/2 is upregulated in gonadal PGCs (relative to early PGCs and surrounding somatic cells) (Hackett et al., 2013; Yamaguchi et al., 2012), and antibody-based genome-wide mapping of 5mC and 5hmC in PGCs indicates that 5mC is oxidized to 5hmC at many late demethylating loci between E9.5 to E11.5 (Hackett et al., 2013). Time-course analysis suggests that both 5mC and 5hmC are lost at a rate consistent with replication-dependent dilution during this period (Hackett et al., 2013; Kagiwada et al., 2013). Given that TDG and BER components are expressed throughout the PGC development (Kagiwada et al., 2013) (Figure 5B), it is likely that TDG/BER may play a role in removing 5fC/5caC at specific loci during phase 2. Integrated genomic analysis showed that TET1 is specifically required for promoter demethylation and transcriptional activation of germline-specific genes such as meiosis-related genes (e.g., *Mae1*, *Sycp1*, and *Sycp3*) in developing female PGCs (Yamaguchi et al., 2012). The locus-specific demethylation at meiotic gene promoters by TET1 in female PGCs is biologically important, as half of the *Tet1*-deficient female gametes exhibits a defect in meiotic synapsis and become developmentally arrested later (Yamaguchi et al., 2012).

Conversion of 5mC to 5hmC at ICRs in late PGCs suggests that TET proteins are also required for the complete 5mC erasure at parental imprints (Hackett et al., 2013). Analysis of methylated ICRs of somatic cells fused with embryonic germ cells (EGCs) suggests that TET1 can induce 5mC oxidation specifically at ICRs and mediate eventual erasure of imprints in this in vitro system (Piccolo et al., 2013). Locus-specific methylation analysis in progenies of *Tet1*<sup>-/-</sup>*Tet2*<sup>-/-</sup> mice revealed aberrant methylation pattern at some ICRs, further supporting a mechanistic link between TET proteins and erasure of parental imprints

in PGCs (Dawlaty et al., 2013). More definitive evidence comes from analysis of progenies derived from crossing *Tet1*<sup>-/-</sup> male with wild-type female (Yamaguchi et al., 2013). Despite their identical genotype, these heterozygous progenies exhibit a number of variable phenotypes including placental, fetal and post-natal growth defects, and early embryonic lethality. Several lines of evidence suggest that these developmental defects can be largely explained by the *Tet1*-deficiency induced epigenetic dysregulation of imprinted genes in PGCs. First, many dysregulated imprinting genes, such as *Peg10* and *Peg3*, exhibit aberrant hypermethylation in the paternal allele of differential methylated regions (DMRs) in progenies of *Tet1*<sup>-/-</sup> male. Second, RNA-seq analysis reveals extensive dysregulation of imprinted genes in these heterozygous progenies. Lastly, reduced representation bisulfite sequencing (RRBS) analysis of E13.5 PGCs and sperms of *Tet1*<sup>-/-</sup> male mice reveals that aberrantly hypermethylated ICRs observed in *Tet1*<sup>-/-</sup> sperms can be traced back to PGCs. Overall, these data suggest that TET1 plays a critical role in resetting parental origin-specific genomic imprinting in PGCs by facilitating the complete erasure of methylation at ICRs (phase 2 in Figure 5). Curiously, it is currently unclear why and how late demethylating loci such as promoters of meiotic genes and ICRs of imprinting genes are protected from the initial phase of global demethylation in early PGCs (phase 1 in Figure 5).

In summary, both passive and active demethylation mechanisms are critical to complete genome-wide erasure of DNA methylation in developing PGCs. This is reminiscent of global demethylation in early embryos. In both cases, passive dilution is facilitated by inefficient maintenance methylation systems (exclusion of Dnmt1o in early embryos and transcriptional repression of Uhrf1 in developing PGCs). Meanwhile, 5mC oxidation by TET3 in zygotes and TET1 (potentially TET2) in PGCs is indispensable for achieving efficient and complete erasure of 5mC in paternal genome of early embryos and late demethylating loci (e.g., meiotic genes and ICRs) in PGCs, respectively.

### Role of DNA Demethylation in Neural Development

The development of mammalian brain is a spatiotemporally orchestrated process that requires precisely controlled gene regulatory program to generate diverse, functionally distinct populations of neurons and glia. Inactivation of DNMTs in the nervous system results in neurodevelopmental deficits and impaired neuronal functions (Fan et al., 2001; Fan et al., 2005; Feng et al., 2010a; LaPlant et al., 2010; Nguyen et al., 2007; Wu et al., 2010), suggesting that precise control of DNA methylation pattern is required for proper brain development and maturation. Interestingly, distinct parts of mammalian brains, including frontal cortex (Lister et al., 2013), hippocampus (Münzel et al., 2010), and cerebellum (Song et al., 2011; Szulwach et al., 2011b), all exhibit age-dependent acquisition of 5hmC (e.g., 0.2% of cytosines in fetal cortex versus 0.9% of cytosines in adult cortex) (Lister et al., 2013). In fact, the genome of mature neurons in adult central nervous system contains the highest level of 5hmC of any mammalian cell-type (~40% as abundant as 5mC in Purkinje neurons in cerebellum) (Kriaucionis and Heintz, 2009). These observations indicate that 5mC oxidation and potentially DNA demethylation



may be functionally important for neuronal differentiation and maturation processes.

All three *Tet* genes are expressed at detectable levels in developing mouse brains (Hahn et al., 2013; Szwagierczak et al., 2010). In embryonic cortex, an increase in expression levels of *Tet2* and *Tet3* is observed when the Nestin-positive neural progenitor cells (NPCs) differentiate into Tubb3-expressing neurons, coinciding with an elevation of 5hmC level during neuronal differentiation (Hahn et al., 2013). Knockdown of *Tet2/3* in developing embryonic cortex via in utero electroporation results in neuronal differentiation deficits; in contrast, overexpression of *Tet2/3* seems to promote embryonic cortical neurogenesis (Hahn et al., 2013). In addition, depletion of *tet3* in *Xenopus* (*X. laevis*) embryos by antisense morpholinos causes defects in early eye and neural development. Interestingly, intact CXXC domain and catalytic activity of TET3 protein are both essential for its role in target gene regulation and normal head development (Xu et al., 2012). Thus, these results support a conserved role of specific TET proteins in regulating early stages of brain development.

Growing evidence suggests that TET proteins may also regulate adult brain functions. While *Tet1*<sup>-/-</sup> mice have no detectable defects in embryonic development, a recent study demonstrates that TET1 may promote adult hippocampal neurogenesis by regulating neural progenitor proliferation and *Tet1*-deficiency may cause defects in spatial learning as well as short-term memory (Zhang et al., 2013). Analysis of DNA methylation by RRBS and gene expression in *Tet1*-deficient NPCs isolated from adult dentate gyrus (DG) of hippocampus identified a cohort of genes that become promoter hypermethylated and downregulated in the absence of TET1 (Zhang et al., 2013). Another study of adult *Tet1*<sup>-/-</sup> mice revealed that multiple neuronal activity-regulated genes, including *Npas4*, *c-Fos*, and *Arc*, are transcriptionally downregulated, and promoter of *Npas4* becomes hypermethylated in absence of TET1 (Rudenko et al., 2013). This study also demonstrated that *Tet1*<sup>-/-</sup> mice are associated with abnormal hippocampal long-term depression (LTD) and impaired memory extinction. Interestingly, viral-mediated overexpression of TET1 or a catalytically inactive form of TET1 in adult hippocampus resulted in transcriptional up-regulation of several memory-associated genes and impaired contextual fear memory (Kaas et al., 2013), suggesting a 5mC oxidase-independent role of TET1 in adult brains. Given that TET2 and TET3 are also expressed in adult mouse brains, it will be interesting to examine the overlapping and distinct function of different TET proteins in adult brains.

The high abundance of 5hmC in adult brains raises the possibility that 5hmC may act as an epigenetic mark. Indeed, current evidence suggests that 5hmC may recruit MeCP2 (Mellén et al., 2012) and other putative binding protein candidates (Spruijt et al., 2013) to facilitate transcription in postnatal brains. Interestingly, a Rett syndrome mutation, R133C, preferentially disrupts the ability of MeCP2 to bind to 5hmC, but not 5mC (Mellén et al., 2012), implicating impairment of 5hmC binding of MeCP2 in this debilitating autism spectrum disorder. Comparative analysis of 5mC, 5hmC and gene expression in specific cell types in cerebellum or olfactory sensory neurons suggest that enrichment of 5hmC and depletion of 5mC within intragenic regions positively correlate with transcription and chromatin

accessibility (Colquitt et al., 2013; Mellén et al., 2012). However, the functional significance of the interplay between 5hmC and its potential binding proteins remains unclear.

While most studies focused on 5mC in CpG context, non-CpG methylation is relatively prevalent in adult mouse brains (Xie et al., 2012). Recently, a comprehensive base-resolution analyses of 5mC and 5hmC in mammalian frontal cortex in both fetal and adult stages indicate that non-CpG methylation (<sup>m</sup>CH) and CpG hydroxymethylation (<sup>h</sup>CG) drastically build up in cortical neurons after birth (Lister et al., 2013), coinciding with the peak of synaptogenesis and synaptic pruning in the cortex. Intriguingly, this study demonstrated that <sup>m</sup>CH could become a dominant form of cytosine modifications in adult brains, accounting for 53% in adult human cortical neuronal genome. Furthermore, the genomic positions of <sup>m</sup>CH in adult brains are highly conserved between individuals, indicating a precisely controlled process catalyzing non-CpG methylation in adult brain. Current evidence suggests that the developmental acquisition of neuronal <sup>m</sup>CH is mediated by DNMT3A, whose expression is transiently spiked during the same period. In mature neurons, intragenic <sup>m</sup>CH is preferentially enriched at inactive nonneuronal lineage-specific genes, indicating a role in negative regulation of the associated transcripts. By contrast, genic <sup>h</sup>CG is positively correlated with gene expression levels. In addition to the 5mC/5hmC dynamics detected in genic regions, distal *cis*-regulatory regions are also associated with developmental change in cytosine modification states. In fetal brains, <sup>h</sup>CG and <sup>m</sup>CG are preferentially enriched at a large cohort of poised distal-regulatory regions that are subsequently demethylated (converting to unmodified CG) and activated in the adult brain. *Tet2*-deficiency partially blocks the demethylation process at these <sup>h</sup>CG-marked poised regulatory elements. However, it is currently unclear whether this TET-mediated demethylation process at poised neuronal enhancers primarily occurs in proliferating neuronal progenitors or postmitotic neurons. Publically accessible RNA-seq data sets indicate that *Tdg* gene is expressed in both fetal and adult mouse cortex, suggesting that TET/TDG-active demethylation may take place in both progenitors and neurons. Further studies are needed to reveal the mechanism and biological significance of widespread methylome reconfiguration in the neuronal genome during the maturation process of brain.

#### DNA Demethylation and Cancer

Aberrant DNA methylation is one of the hallmarks associated with cancer cells (Baylin and Jones, 2011; Esteller, 2007). Growing evidence suggests that impairment in TET-mediated DNA demethylation may contribute to tumorigenesis (Cimmino et al., 2011; Wu and Zhang, 2011a). The initial evidence implicating TET proteins in cancer development came from the identification of human *TET1* as a rare fusion partner of MLL in patients with acute myeloid leukemia (AML) (Ono et al., 2002). Subsequent studies reveal a link between recurrent mutations in the *TET2* gene and myelodysplastic disorders (Delhommeau et al., 2009; Langemeijer et al., 2009). Mouse genetic studies have demonstrated that *Tet2*-deficiency causes enhanced self-renew of hematopoietic stem cells (Ko et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011), supporting a role of TET2 as tumor suppressor for leukemia. *TET2* mutations associated with myeloid malignancies impair

enzymatic activity and cause consistent loss of 5hmC in patient bone marrows (Ko et al., 2010).

Interestingly, TET2 mutations are found to be mutually exclusive with neomorphic mutations in IDH1/2 (R132 of IDH1 and R140/172 of IDH2) in AML (Dang et al., 2009; Ward et al., 2010). Wild-type IDH1/2 converts isocitrate to  $\alpha$ -KG, whereas mutant IDH enzymes produce 2-hydroxyglutarate (2-HG), an oncometabolite that competitively inhibit many  $\alpha$ -KG-dependent dioxygenases, including TET proteins (Figuerola et al., 2010; Xu et al., 2011a). Similarly, inactivation mutations in SDH and FH, which have been identified in a subset of cancers, lead to marked accumulation of succinate and fumarate, both of which can inhibit  $\alpha$ -KG-dependent dioxygenases (Xiao et al., 2012). Thus, mutations in these TCA cycle enzymes may lead to inactivation of TET proteins and alteration of methylome in cancer cells.

Recent studies suggest that transcriptional downregulation of all three *TET* genes and *IDH1/2* is observed in many solid cancers that are not known to be associated with mutations in these enzymes, and global loss of 5hmC may be broadly used as a diagnostic biomarker for human melanoma, breast, liver, lung pancreatic and prostate cancers (Hsu et al., 2012; Lian et al., 2012; Yang et al., 2013). Overexpression of oncogenic miRNAs that target *TET* mRNAs may be one possible mechanism to downregulate TET proteins and 5hmC levels (Song et al., 2013c). In summary, multiple mechanisms may cause inactivation of TET-mediated 5mC oxidation, but relevant molecular targets of TET proteins that are causally linked to cancer development are not systematically investigated. It is also unclear whether TDG and 5fC/5caC may play a role in cancer.

### Conclusion and Perspective

The discovery of oxidative modification of 5mC and related repair mechanism has provided a new perspective on how heritable DNA methylation pattern may be dynamically regulated in the mammalian genome. The realization that epigenetic information can be reversibly “written” and “erased” on mammalian DNA through the cyclic cytosine-modifying cascade raises new mechanistic and biological questions: Which 5-methylcytosines in the genome are selectively targeted by TET and TDG-mediated demethylation activity? Can oxidized 5mC bases (5hmC/5fC/5caC) act as epigenetic marks by recruiting or repelling specific reader proteins? How does targeted excision repair of 5fC/5caC at distal *cis*-regulatory elements contribute to gene regulation? How is TET-mediated active erasure of 5mC regulated in response to developmental signaling, environmental stimuli and changes in metabolic states? How does this dynamic epigenetic regulatory process functionally contribute to development and diseases in vivo? New technologies such as TAL effectors and CRISPR/Cas9 that allow manipulation of the epigenetic state of cytosine modification at specific loci (Maeder et al., 2013) as well as base-resolution mapping of all five cytosine modification states promise to advance this exciting field substantially in the coming years.

### ACKNOWLEDGMENTS

We thank members of the Zhang Lab for their comments on the manuscript. We apologize to colleagues whose work cannot be cited owing to space con-

straints. This work was supported by NIH grants GM68804 and U01DK089565 (to Y.Z.). H.W. is supported by a postdoctoral fellowship from the Jane Coffin Childs Memorial Fund for Medical Research. Y.Z. is an Investigator of the Howard Hughes Medical Institute.

### REFERENCES

- Barreto, G., Schäfer, A., Marhold, J., Stach, D., Swaminathan, S.K., Handa, V., Döderlein, G., Maltry, N., Wu, W., Lyko, F., and Niehrs, C. (2007). Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* 445, 671–675.
- Baylin, S.B., and Jones, P.A. (2011). A decade of exploring the cancer epigenome - biological and translational implications. *Nat. Rev. Cancer* 11, 726–734.
- Bennett, M.T., Rodgers, M.T., Hebert, A.S., Ruslander, L.E., Eisele, L., and Drohat, A.C. (2006). Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* 128, 12510–12519.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* 125, 315–326.
- Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* 463, 1042–1047.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.
- Blaschke, K., Ebata, K.T., Karimi, M.M., Zepeda-Martínez, J.A., Goyal, P., Mahapatra, S., Tam, A., Laird, D.J., Hirst, M., Rao, A., et al. (2013). Vitamin C induces Tet-dependent DNA demethylation and a blastocyst-like state in ES cells. *Nature* 500, 222–226.
- Booth, M.J., Branco, M.R., Ficiz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* 336, 934–937.
- Borst, P., and Sabatini, R. (2008). Base J: discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol.* 62, 235–251.
- Bostick, M., Kim, J.K., Estève, P.O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760–1764.
- Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. USA* 100, 4102–4107.
- Chen, T., Ueda, Y., Dodge, J.E., Wang, Z., and Li, E. (2003). Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* 23, 5594–5605.
- Chen, C.C., Wang, K.Y., and Shen, C.K. (2012). The mammalian de novo DNA methyltransferases DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxymethylases. *J. Biol. Chem.* 287, 33116–33121.
- Chen, J., Guo, L., Zhang, L., Wu, H., Yang, J., Liu, H., Wang, X., Hu, X., Gu, T., Zhou, Z., et al. (2013a). Vitamin C modulates TET1 function during somatic cell reprogramming. *Nat. Genet.* 45, 1504–1509.
- Chen, Q., Chen, Y., Bian, C., Fujiki, R., and Yu, X. (2013b). TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* 493, 561–564.
- Cimmino, L., Abdel-Wahab, O., Levine, R.L., and Aifantis, I. (2011). TET family proteins and their role in stem cell differentiation and transformation. *Cell Stem Cell* 9, 193–204.
- Clark, S.J., Harrison, J., Paul, C.L., and Frommer, M. (1994). High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* 22, 2990–2997.
- Clark, T.A., Lu, X., Luong, K., Dai, Q., Boitano, M., Turner, S.W., He, C., and Koriach, J. (2013). Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tet1 oxidation. *BMC Biol.* 11, 4.

- Colquhitt, B.M., Allen, W.E., Barnea, G., and Lomvardas, S. (2013). Alteration of genic 5-hydroxymethylcytosine patterning in olfactory neurons correlates with changes in gene expression and cell identity. *Proc. Natl. Acad. Sci. USA* **110**, 14682–14687.
- Coticello, S.G. (2008). The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **9**, 229.
- Cortázar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423.
- Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79.
- Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013). NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370–374.
- Dalton, S.R., and Bellacosa, A. (2012). DNA demethylation by TDG. *Epigenomics* **4**, 459–467.
- Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* **462**, 739–744.
- Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C., and Jaenisch, R. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**, 166–175.
- Dawlaty, M.M., Breiling, A., Le, T., Raddatz, G., Barrasa, M.I., Cheng, A.W., Gao, Q., Powell, B.E., Li, Z., Xu, M., et al. (2013). Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24**, 310–323.
- Deaton, A.M., and Bird, A. (2011). CpG islands and the regulation of transcription. *Genes Dev.* **25**, 1010–1022.
- Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Massé, A., Kosmider, O., Le Couedic, J.P., Robert, F., Alberdi, A., et al. (2009). Mutation in TET2 in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301.
- Deplus, R., Delatte, B., Schwinn, M.K., Defrance, M., Méndez, J., Murphy, N., Dawson, M.A., Volkmar, M., Putmans, P., Calonne, E., et al. (2013). TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J.* **32**, 645–655.
- Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A., et al. (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652–655.
- Engel, N., Tront, J.S., Erinle, T., Nguyen, N., Latham, K.E., Sapienza, C., Hoffman, B., and Liebermann, D.A. (2009). Conserved DNA methylation in Gadd45a(-/-) mice. *Epigenetics* **4**, 98–99.
- Esteller, M. (2007). Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat. Rev. Genet.* **8**, 286–298.
- Fan, G., Beard, C., Chen, R.Z., Csankovszki, G., Sun, Y., Siniaia, M., Biniszkie-wicz, D., Bates, B., Lee, P.P., Kuhn, R., et al. (2001). DNA hypomethylation perturbs the function and survival of CNS neurons in postnatal animals. *J. Neurosci.* **21**, 788–797.
- Fan, G., Martinowich, K., Chin, M.H., He, F., Fouse, S.D., Hutnick, L., Hattori, D., Ge, W., Shen, Y., Wu, H., et al. (2005). DNA methylation controls the timing of astroglialogenesis through regulation of JAK-STAT signaling. *Development* **132**, 3345–3356.
- Feng, J., Zhou, Y., Campbell, S.L., Le, T., Li, E., Sweatt, J.D., Silva, A.J., and Fan, G. (2010a). Dnmt1 and Dnmt3a maintain DNA methylation and regulate synaptic function in adult forebrain neurons. *Nat. Neurosci.* **13**, 423–430.
- Feng, S., Jacobsen, S.E., and Reik, W. (2010b). Epigenetic reprogramming in plant and animal development. *Science* **330**, 622–627.
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402.
- Ficz, G., Hore, T.A., Santos, F., Lee, H.J., Dean, W., Arand, J., Krueger, F., Oxley, D., Paul, Y.L., Walter, J., et al. (2013). FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency. *Cell Stem Cell* **13**, 351–359.
- Figueroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567.
- Fong, Y.W., Cattoglio, C., and Tjian, R. (2013). The intertwined roles of transcription and repair proteins. *Mol. Cell* **52**, 291–302.
- Frauer, C., Hoffmann, T., Bultmann, S., Casa, V., Cardoso, M.C., Antes, I., and Leonhardt, H. (2011). Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306.
- Freudenberg, J.M., Ghosh, S., Lackford, B.L., Yellaboina, S., Zheng, X., Li, R., Cuddapah, S., Wade, P.A., Hu, G., and Jothi, R. (2012). Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res.* **40**, 3364–3377.
- Gao, Y., Chen, J., Li, K., Wu, T., Huang, B., Liu, W., Kou, X., Zhang, Y., Huang, H., Jiang, Y., et al. (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469.
- Globisch, D., Münzel, M., Müller, M., Michalak, S., Wagner, M., Koch, S., Brückl, T., Biel, M., and Carell, T. (2010). Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367.
- Goll, M.G., and Bestor, T.H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514.
- Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610.
- Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434.
- Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13**, 360–369.
- Hackett, J.A., Sengupta, R., Zylisc, J.J., Murakami, K., Lee, C., Down, T.A., and Surani, M.A. (2013). Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448–452.
- Hahn, M.A., Qiu, R., Wu, X., Li, A.X., Zhang, H., Wang, J., Jui, J., Jin, S.G., Jiang, Y., Pfeifer, G.P., and Lu, Q. (2013). Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. *Cell Rep.* **3**, 291–300.
- Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El-Maari, O., Reik, W., Walter, J., and Surani, M.A. (2002). Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**, 15–23.
- Hajkova, P., Jeffries, S.J., Lee, C., Miller, N., Jackson, S.P., and Surani, M.A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78–82.
- Hanna, J.H., Saha, K., and Jaenisch, R. (2010). Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143**, 508–525.
- Hashimoto, H., Vertino, P.M., and Cheng, X. (2010). Molecular coupling of DNA methylation and histone methylation. *Epigenomics* **2**, 657–669.
- Hashimoto, H., Liu, Y., Upadhyay, A.K., Chang, Y., Howerton, S.B., Vertino, P.M., Zhang, X., and Cheng, X. (2012). Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Res.* **40**, 4841–4849.

- Hashimoto, H., Pais, J.E., Zhang, X., Saleh, L., Fu, Z.-Q., Dai, N., Corrêa, I.R., Zheng, Y., and Cheng, X. (2013). Structure of a Naegleria Tet-like dioxygenase in complex with 5-methylcytosine DNA. *Nature*. Published online December 25, 2013. <http://dx.doi.org/10.1038/nature12905>.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.
- Hermann, A., Goyal, R., and Jeltsch, A. (2004). The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. *J. Biol. Chem.* 279, 48350–48359.
- Holliday, R., and Pugh, J.E. (1975). DNA modification mechanisms and gene activity during development. *Science* 187, 226–232.
- Hsu, C.H., Peng, K.L., Kang, M.L., Chen, Y.R., Yang, Y.C., Tsai, C.H., Chu, C.S., Jeng, Y.M., Chen, Y.T., Lin, F.M., et al. (2012). TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases. *Cell Rep.* 2, 568–579.
- Hu, L., Li, Z., Cheng, J., Rao, Q., Gong, W., Liu, M., Shi, Y.G., Zhu, J., Wang, P., and Xu, Y. (2013). Crystal Structure of TET2-DNA Complex: Insight into TET-Mediated 5mC Oxidation. *Cell* 155, 1545–1555.
- Huang, Y., Pastor, W.A., Shen, Y., Tahiliani, M., Liu, D.R., and Rao, A. (2010). The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* 5, e8888.
- Huang, Y., Pastor, W.A., Zepeda-Martínez, J.A., and Rao, A. (2012). The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. *Nat. Protoc.* 7, 1897–1908.
- Inoue, A., and Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 334, 194.
- Inoue, A., Shen, L., Dai, Q., He, C., and Zhang, Y. (2011). Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* 21, 1670–1676.
- Iqbal, K., Jin, S.G., Pfeifer, G.P., and Szabó, P.E. (2011). Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl. Acad. Sci. USA* 108, 3642–3647.
- 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.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303.
- Iurlaro, M., Ficiz, G., Oxley, D., Raiber, E.A., Bachman, M., Booth, M.J., Andrews, S., Balasubramanian, S., and Reik, W. (2013). A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* 14, R119.
- Iyer, L.M., Tahiliani, M., Rao, A., and Aravind, L. (2009). Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* 8, 1698–1710.
- Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet. Suppl.* 33, 245–254.
- Ji, H., Ehrlich, L.I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M.J., Irizarry, R.A., Kim, K., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* 467, 338–342.
- Jiang, L., Zhang, J., Wang, J.J., Wang, L., Zhang, L., Li, G., Yang, X., Ma, X., Sun, X., Cai, J., et al. (2013). Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell* 153, 773–784.
- Jin, S.G., Guo, C., and Pfeifer, G.P. (2008). GADD45A does not promote DNA demethylation. *PLoS Genet.* 4, e1000013.
- Kaas, G.A., Zhong, C., Eason, D.E., Ross, D.L., Vachhani, R.V., Ming, G.L., King, J.R., Song, H., and Sweatt, J.D. (2013). TET1 controls CNS 5-methylcytosine hydroxylation, active DNA demethylation, gene transcription, and memory formation. *Neuron* 79, 1086–1093.
- Kaelin, W.G., Jr., and McKnight, S.L. (2013). Influence of metabolism on epigenetics and disease. *Cell* 153, 56–69.
- Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M., and Saitou, M. (2013). Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J.* 32, 340–353.
- Kellinger, M.W., Song, C.X., Chong, J., Lu, X.Y., He, C., and Wang, D. (2012). 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 19, 831–833.
- Kemmerich, K., Dingler, F.A., Rada, C., and Neuberger, M.S. (2012). Germline ablation of SMUG1 DNA glycosylase causes loss of 5-hydroxymethyluracil and UNG-backup uracil-excision activities and increases cancer predisposition of Ung<sup>-/-</sup>Msh2<sup>-/-</sup> mice. *Nucleic Acids Res.* 40, 6016–6025.
- Klose, R.J., Kallin, E.M., and Zhang, Y. (2006). JmJc-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* 7, 715–727.
- Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* 468, 839–843.
- Ko, M., Bandukwala, H.S., An, J., Lamperti, E.D., Thompson, E.C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S.B., and Rao, A. (2011). Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl. Acad. Sci. USA* 108, 14566–14571.
- Ko, M., An, J., Bandukwala, H.S., Chavez, L., Aijō, T., Pastor, W.A., Segal, M.F., Li, H., Koh, K.P., Lähdesmäki, H., et al. (2013). Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* 497, 122–126.
- Kobayashi, H., Sakurai, T., Imai, M., Takahashi, N., Fukuda, A., Yayoi, O., Sato, S., Nakabayashi, K., Hata, K., Sotomaru, Y., et al. (2012). Contribution of intra-genic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* 8, e1002440.
- Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cuniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., 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.
- Kohli, R.M., and Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502, 472–479.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Kumar, R., DiMenna, L., Schrodde, N., Liu, T.C., Franck, P., Muñoz-Descalzo, S., Hadjantonakis, A.K., Zarrin, A.A., Chaudhuri, J., Elemento, O., and Evans, T. (2013). AID stabilizes stem-cell phenotype by removing epigenetic memory of pluripotency genes. *Nature* 500, 89–92.
- Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003). Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* 35, 88–93.
- Langemeijer, S.M., Kuiper, R.P., Berends, M., Knops, R., Aslanyan, M.G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A.G., Raymakers, R.A., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat. Genet.* 41, 838–842.
- LaPlant, Q., Vialou, V., Covington, H.E., 3rd, Dumitriu, D., Feng, J., Warren, B.L., Maze, I., Dietz, D.M., Watts, E.L., Iñiguez, S.D., et al. (2010). Dnmt3a regulates emotional behavior and spine plasticity in the nucleus accumbens. *Nat. Neurosci.* 13, 1137–1143.
- 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.
- Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. *Nat. Struct. Mol. Biol.* 20, 311–316.
- Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. *Dev. Cell* 15, 547–557.



- Li, Z., Cai, X., Cai, C.L., Wang, J., Zhang, W., Petersen, B.E., Yang, F.C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* 118, 4509–4518.
- Li, W.W., Gong, L., and Bayley, H. (2013). Single-molecule detection of 5-hydroxymethylcytosine in DNA through chemical modification and nanopore analysis. *Angew. Chem. Int. Ed. Engl.* 52, 4350–4355.
- Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., et al. (2012). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* 150, 1135–1146.
- Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322.
- Lister, R., Mukamel, E.A., Nery, J.R., Urich, M., Puddifoot, C.A., Johnson, N.D., Lucero, J., Huang, Y., Dwork, A.J., Schultz, M.D., et al. (2013). Global epigenomic reconfiguration during mammalian brain development. *Science* 341, 1237905.
- Liu, Y., Toh, H., Sasaki, H., Zhang, X., and Cheng, X. (2012). An atomic model of Zfp57 recognition of CpG methylation within a specific DNA sequence. *Genes Dev.* 26, 2374–2379.
- Liutkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D., and Klimasauskas, S. (2009). Cytosine-5-methyltransferases add aldehydes to DNA. *Nat. Chem. Biol.* 5, 400–402.
- Loenarz, C., and Schofield, C.J. (2011). Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem. Sci.* 36, 7–18.
- Long, H.K., Blackledge, N.P., and Klose, R.J. (2013). ZF-CxxC domain-containing proteins, CpG islands and the chromatin connection. *Biochem. Soc. Trans.* 41, 727–740.
- Lorsbach, R.B., Moore, J., Mathew, S., Raimondi, S.C., Mukatira, S.T., and Downing, J.R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* 17, 637–641.
- Lu, X., Song, C.X., Szulwach, K., Wang, Z., Weidenbacher, P., Jin, P., and He, C. (2013). Chemical modification-assisted bisulfite sequencing (CAB-Seq) for 5-carboxylcytosine detection in DNA. *J. Am. Chem. Soc.* 135, 9315–9317.
- Ma, D.K., Jang, M.H., Guo, J.U., Kitabatake, Y., Chang, M.L., Pow-Anpongkul, N., Flavell, R.A., Lu, B., Ming, G.L., and Song, H. (2009). Neuronal activity-induced Gadd45b promotes epigenetic DNA demethylation and adult neurogenesis. *Science* 323, 1074–1077.
- Maeder, M.L., Angstman, J.F., Richardson, M.E., Linder, S.J., Cascio, V.M., Tsai, S.Q., Ho, Q.H., Sander, J.D., Reyon, D., Bernstein, B.E., et al. (2013). Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat. Biotechnol.* 31, 1137–1142.
- 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.
- Marks, H., Kalkan, T., Menafrá, R., Denissov, S., Jones, K., Hofemeister, H., Nichols, J., Kranz, A., Stewart, A.F., Smith, A., and Stunnenberg, H.G. (2012). The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* 149, 590–604.
- Matarese, F., Carrillo-de Santa Pau, E., and Stunnenberg, H.G. (2011). 5-Hydroxymethylcytosine: a new kid on the epigenetic block? *Mol. Syst. Biol.* 7, 562.
- Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. *Nature* 403, 501–502.
- Mellén, M., Ayata, P., Dewell, S., Kriakionis, S., and Heintz, N. (2012). MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* 151, 1417–1430.
- Messerschmidt, D.M., de Vries, W., Ito, M., Solter, D., Ferguson-Smith, A., and Knowles, B.B. (2012). Trim28 is required for epigenetic stability during mouse oocyte to embryo transition. *Science* 335, 1499–1502.
- Métivier, R., Gallais, R., Tiffoche, C., Le Péron, C., Jurkowska, R.Z., Carmouche, R.P., Ibberson, D., Barath, P., Demay, F., Reid, G., et al. (2008). Cyclical DNA methylation of a transcriptionally active promoter. *Nature* 452, 45–50.
- Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560.
- Millar, C.B., Guy, J., Sansom, O.J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P.D., Bishop, S.M., Clarke, A.R., and Bird, A. (2002). Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* 297, 403–405.
- Minor, E.A., Court, B.L., Young, J.I., and Wang, G. (2013). Ascorbate induces ten-eleven translocation (Tet) methylcytosine dioxygenase-mediated generation of 5-hydroxymethylcytosine. *J. Biol. Chem.* 288, 13669–13674.
- Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* 20, 11–24.
- Morgan, H.D., Dean, W., Coker, H.A., Reik, W., and Petersen-Mahrt, S.K. (2004). Activation-induced cytidine deaminase deaminates 5-methylcytosine in DNA and is expressed in pluripotent tissues: implications for epigenetic reprogramming. *J. Biol. Chem.* 279, 52353–52360.
- Münzel, M., Globisch, D., Brückl, T., Wagner, M., Welzmler, V., Michalakakis, S., Müller, M., Biel, M., and Carell, T. (2010). Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Ed. Engl.* 49, 5375–5377.
- Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* 8, 751–758.
- Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., et al. (2007). PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nat. Cell Biol.* 9, 64–71.
- Nakamura, T., Liu, Y.J., Nakashima, H., Umehara, H., Inoue, K., Matoba, S., Tachibana, M., Ogura, A., Shinkai, Y., and Nakano, T. (2012). PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* 486, 415–419.
- Nguyen, S., Meletis, K., Fu, D., Jhaveri, S., and Jaenisch, R. (2007). Ablation of de novo DNA methyltransferase Dnmt3a in the nervous system leads to neuromuscular defects and shortened lifespan. *Dev. Dyn.* 236, 1663–1676.
- Okada, Y., Yamagata, K., Hong, K., Wakayama, T., and Zhang, Y. (2010). A role for the elongator complex in zygotic paternal genome demethylation. *Nature* 463, 554–558.
- Okano, M., Xie, S., and Li, E. (1998). Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* 19, 219–220.
- Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257.
- 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.
- Ooi, S.K., and Bestor, T.H. (2008). The colorful history of active DNA demethylation. *Cell* 133, 1145–1148.
- Ooi, S.K., Qiu, C., Bernstein, E., Li, K., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S.P., Allis, C.D., et al. (2007). DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714–717.
- Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* 10, 475–478.

- Otani, J., Arita, K., Kato, T., Kinoshita, M., Kimura, H., Suetake, I., Tajima, S., Ariyoshi, M., and Shirakawa, M. (2013a). Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. *J. Biol. Chem.* **288**, 6351–6362.
- Otani, J., Kimura, H., Sharif, J., Endo, T.A., Mishima, Y., Kawakami, T., Koseki, H., Shirakawa, M., Suetake, I., and Tajima, S. (2013b). Cell cycle-dependent turnover of 5-hydroxymethyl Cytosine in mouse embryonic stem cells. *PLoS ONE* **8**, e82961.
- Pastor, W.A., Pape, U.J., Huang, Y., Henderson, H.R., Lister, R., Ko, M., McLoughlin, E.M., Brudno, Y., Mahapatra, S., Kapranov, P., et al. (2011). Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397.
- 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.
- Payer, B., Saitou, M., Barton, S.C., Thresher, R., Dixon, J.P.C., Zahn, D., Colledge, W.H., Carlton, M.B.L., Nakano, T., and Surani, M.A. (2003). Stella is a maternal effect gene required for normal early development in mice. *Curr. Biol.* **13**, 2110–2117.
- Pfaffeneder, T., Hackner, B., Truss, M., Münzel, M., Müller, M., Deiml, C.A., Hagemeyer, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Ed. Engl.* **50**, 7008–7012.
- Piccolo, F.M., Bagci, H., Brown, K.E., Landeira, D., Soza-Ried, J., Feytout, A., Mooijman, D., Hajkova, P., Leitch, H.G., Tada, T., et al. (2013). Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion. *Mol. Cell* **49**, 1023–1033.
- Popp, C., Dean, W., Feng, S., Cokus, S.J., Andrews, S., Pellegrini, M., Jacobsen, S.E., and Reik, W. (2010). Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105.
- Potok, M.E., Nix, D.A., Parnell, T.J., and Cairns, B.R. (2013). Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* **153**, 759–772.
- Quenneville, S., Verde, G., Corsinotti, A., Kapopoulou, A., Jakobsson, J., Offner, S., Baglivo, I., Pedone, P.V., Grimaldi, G., Riccio, A., and Trono, D. (2011). In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* **44**, 361–372.
- Quivoron, C., Couronné, L., Della Valle, V., Lopez, C.K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M.H., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* **20**, 25–38.
- Rai, K., Huggins, I.J., James, S.R., Karpf, A.R., Jones, D.A., and Cairns, B.R. (2008). DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201–1212.
- Raiber, E.A., Beraldi, D., Ficiz, G., Burgess, H.E., Branco, M.R., Murat, P., Oxley, D., Booth, M.J., Reik, W., and Balasubramanian, S. (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.
- Riggs, A.D. (1975). X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* **14**, 9–25.
- Robertson, A.B., Dahl, J.A., Vågbo, C.B., Tripathi, P., Krokan, H.E., and Klungland, A. (2011). A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **39**, e55.
- Rudenko, A., Dawlaty, M.M., Seo, J., Cheng, A.W., Meng, J., Le, T., Faull, K.F., Jaenisch, R., and Tsai, L.H. (2013). Tet1 is critical for neuronal activity-regulated gene expression and memory extinction. *Neuron* **79**, 1109–1122.
- Saitou, M., Kagiwada, S., and Kurimoto, K. (2012). Epigenetic reprogramming in mouse pre-implantation development and primordial germ cells. *Development* **139**, 15–31.
- Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.* **9**, 129–140.
- Schiesser, S., Hackner, B., Pfaffeneder, T., Müller, M., Hagemeyer, C., Truss, M., and Carell, T. (2012). Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Ed. Engl.* **51**, 6516–6520.
- Schmitz, K.M., Schmitt, N., Hoffmann-Rohrer, U., Schäfer, A., Grummt, I., and Mayer, C. (2009). TAF12 recruits Gadd45a and the nucleotide excision repair complex to the promoter of rRNA genes leading to active DNA demethylation. *Mol. Cell* **33**, 344–353.
- Seisenberger, S., Andrews, S., Krueger, F., Arand, J., Walter, J., Santos, F., Popp, C., Thienpont, B., Dean, W., and Reik, W. (2012). The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862.
- Seisenberger, S., Peat, J.R., and Reik, W. (2013). Conceptual links between DNA methylation reprogramming in the early embryo and primordial germ cells. *Curr. Opin. Cell Biol.* **25**, 281–288.
- Sharif, J., Muto, M., Takebayashi, S., Suetake, I., Iwamatsu, A., Endo, T.A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., et al. (2007). The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450**, 908–912.
- Shearstone, J.R., Pop, R., Bock, C., Boyle, P., Meissner, A., and Socolovsky, M. (2011). Global DNA demethylation during mouse erythropoiesis in vivo. *Science* **334**, 799–802.
- Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V.V., and Ren, B. (2012). A map of the cis-regulatory sequences in the mouse genome. *Nature* **488**, 116–120.
- Shen, L., Wu, H., Diep, D., Yamaguchi, S., D'Alessio, A.C., Fung, H.L., Zhang, K., and Zhang, Y. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706.
- Shirane, K., Toh, H., Kobayashi, H., Miura, F., Chiba, H., Ito, T., Kono, T., and Sasaki, H. (2013). Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* **9**, e1003439.
- Smallwood, S.A., Tomizawa, S., Krueger, F., Ruf, N., Carli, N., Segonds-Pichon, A., Sato, S., Hata, K., Andrews, S.R., and Kelsey, G. (2011). Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* **43**, 811–814.
- Smiley, J.A., Kundracik, M., Landfried, D.A., Barnes, V.R., Sr., and Axheim, A.A. (2005). Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-oxoate decarboxylase from *Neurospora crassa*. *Biochim. Biophys. Acta* **1723**, 256–264.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220.
- Smith, Z.D., Chan, M.M., Mikkelsen, T.S., Gu, H., Gnirke, A., Regev, A., and Meissner, A. (2012). A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344.
- Song, C.X., and He, C. (2013). Potential functional roles of DNA demethylation intermediates. *Trends Biochem. Sci.* **38**, 480–484.
- Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X., et al. (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat. Biotechnol.* **29**, 68–72.
- Song, C.X., Clark, T.A., Lu, X.Y., Kislyuk, A., Dai, Q., Turner, S.W., He, C., and Korlach, J. (2012a). Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. *Nat. Methods* **9**, 75–77.
- Song, C.X., Yi, C., and He, C. (2012b). Mapping recently identified nucleotide variants in the genome and transcriptome. *Nat. Biotechnol.* **30**, 1107–1116.
- Song, C.-X., Szulwach, K.E., Dai, Q., Fu, Y., Mao, S.-Q., Lin, L., Street, C., Li, Y., Poidevin, M., Wu, H., et al. (2013a). Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691.
- Song, S.J., Ito, K., Ala, U., Kats, L., Webster, K., Sun, S.M., Jongen-Lavrencic, M., Manova-Todorova, K., Teruya-Feldstein, J., Avigan, D.E., et al. (2013b). The oncogenic microRNA miR-22 targets the TET2 tumor suppressor to

promote hematopoietic stem cell self-renewal and transformation. *Cell Stem Cell* 13, 87–101.

Song, S.J., Poliseno, L., Song, M.S., Ala, U., Webster, K., Ng, C., Beringer, G., Brikbak, N.J., Yuan, X., Cantley, L.C., et al. (2013c). MicroRNA-antagonism regulates breast cancer stemness and metastasis via TET-family-dependent chromatin remodeling. *Cell* 154, 311–324.

Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W.T.C., Bauer, C., Münzel, M., Wagner, M., Müller, M., Khan, F., et al. (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* 152, 1146–1159.

Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495.

Stivers, J.T., and Jiang, Y.L. (2003). A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* 103, 2729–2759.

Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S., and Jacobsen, S.E. (2011). 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* 12, R54.

Szulwach, K.E., Li, X., Li, Y., Song, C.X., Han, J.W., Kim, S., Namburi, S., Hermetz, K., Kim, J.J., Rudd, M.K., et al. (2011a). Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* 7, e1002154.

Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I., et al. (2011b). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. *Nat. Neurosci.* 14, 1607–1616.

Szwagierczak, A., Bultmann, S., Schmidt, C.S., Spada, F., and Leonhardt, H. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* 38, e181.

Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.

Tang, F., Barbacioru, C., Nordman, E., Bao, S., Lee, C., Wang, X., Tuch, B.B., Heard, E., Lao, K., and Surani, M.A. (2011). Deterministic and stochastic allele specific gene expression in single mouse blastomeres. *PLoS ONE* 6, e21208.

Thomson, J.P., Hunter, J.M., Nestor, C.E., Dunican, D.S., Terranova, R., Moggs, J.G., and Meehan, R.R. (2013). Comparative analysis of affinity-based 5-hydroxymethylation enrichment techniques. *Nucleic Acids Res.* 41, e206.

Valinluck, V., Tsai, H.H., Rogstad, D.K., Burdzy, A., Bird, A., and Sowers, L.C. (2004). Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* 32, 4100–4108.

Vella, P., Scelfo, A., Jammula, S., Chiacchiera, F., Williams, K., Cuomo, A., Roberto, A., Christensen, J., Bonaldi, T., Helin, K., and Pasini, D. (2013). Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* 49, 645–656.

Vincent, J.J., Huang, Y., Chen, P.Y., Feng, S., Calvopiña, J.H., Nee, K., Lee, S.A., Le, T., Yoon, A.J., Faull, K., et al. (2013). Stage-specific roles for tet1 and tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell* 12, 470–478.

Wanunu, M., Cohen-Karni, D., Johnson, R.R., Fields, L., Benner, J., Peterman, N., Zheng, Y., Klein, M.L., and Drndic, M. (2011). Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. *J. Am. Chem. Soc.* 133, 486–492.

Ward, P.S., Patel, J., Wise, D.R., Abdel-Wahab, O., Bennett, B.D., Collier, H.A., Cross, J.R., Fantin, V.R., Hedvat, C.V., Perl, A.E., et al. (2010). The common feature of leukemia-associated IDH1 and IDH2 mutations is a neomorphic enzyme activity converting alpha-ketoglutarate to 2-hydroxyglutarate. *Cancer Cell* 17, 225–234.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. *Nature* 482, 221–225.

Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473, 343–348.

Wossidlo, M., Nakamura, T., Lepikhov, K., Marques, C.J., Zakhartchenko, V., Boiani, M., Arand, J., Nakano, T., Reik, W., and Walter, J. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat. Commun.* 2, 241.

Wu, S.C., and Zhang, Y. (2010). Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* 11, 607–620.

Wu, H., and Zhang, Y. (2011a). Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev.* 25, 2436–2452.

Wu, H., and Zhang, Y. (2011b). Tet1 and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. *Cell Cycle* 10, 2428–2436.

Wu, H., Coskun, V., Tao, J., Xie, W., Ge, W., Yoshikawa, K., Li, E., Zhang, Y., and Sun, Y.E. (2010). Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. *Science* 329, 444–448.

Wu, H., D'Alessio, A.C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011a). Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Genes Dev.* 25, 679–684.

Wu, H., D'Alessio, A.C., Ito, S., Xia, K., Wang, Z., Cui, K., Zhao, K., Sun, Y.E., and Zhang, Y. (2011b). Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473, 389–393.

Xiao, M., Yang, H., Xu, W., Ma, S., Lin, H., Zhu, H., Liu, L., Liu, Y., Yang, C., Xu, Y., et al. (2012). Inhibition of  $\alpha$ -KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev.* 26, 1326–1338.

Xie, W., Barr, C.L., Kim, A., Yue, F., Lee, A.Y., Eubanks, J., Dempster, E.L., and Ren, B. (2012). Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* 148, 816–831.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.H., Ito, S., Yang, C., Wang, P., Xiao, M.T., et al. (2011a). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Cancer Cell* 19, 17–30.

Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S., et al. (2011b). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* 42, 451–464.

Xu, Y., Xu, C., Kato, A., Tempel, W., Abreu, J.G., Bian, C., Hu, Y., Hu, D., Zhao, B., Cerovina, T., et al. (2012). Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell* 151, 1200–1213.

Yamaguchi, S., Hong, K., Liu, R., Shen, L., Inoue, A., Diep, D., Zhang, K., and Zhang, Y. (2012). Tet1 controls meiosis by regulating meiotic gene expression. *Nature* 492, 443–447.

Yamaguchi, S., Shen, L., Liu, Y., Sandler, D., and Zhang, Y. (2013). Role of Tet1 in erasure of genomic imprinting. *Nature* 504, 460–464.

Yamaji, M., Ueda, J., Hayashi, K., Ohta, H., Yabuta, Y., Kurimoto, K., Nakato, R., Yamada, Y., Shirahige, K., and Saitou, M. (2013). PRDM14 ensures naive pluripotency through dual regulation of signaling and epigenetic pathways in mouse embryonic stem cells. *Cell Stem Cell* 12, 368–382.

Yang, H., Liu, Y., Bai, F., Zhang, J.Y., Ma, S.H., Liu, J., Xu, Z.D., Zhu, H.G., Ling, Z.Q., Ye, D., et al. (2013). Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* 32, 663–669.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. *Cell* 147, 1498–1510.

- Yin, R., Mao, S.-Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., et al. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J. Am. Chem. Soc.* *135*, 10396–10403.
- Ying, Q.L., Wray, J., Nichols, J., Battle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. *Nature* *453*, 519–523.
- Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* *149*, 1368–1380.
- Zhang, H., Zhang, X., Clark, E., Mulcahey, M., Huang, S., and Shi, Y.G. (2010). TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* *20*, 1390–1393.
- Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H., and He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nat. Chem. Biol.* *8*, 328–330.
- Zhang, R.R., Cui, Q.Y., Murai, K., Lim, Y.C., Smith, Z.D., Jin, S., Ye, P., Rosa, L., Lee, Y.K., Wu, H.P., et al. (2013). Tet1 regulates adult hippocampal neurogenesis and cognition. *Cell Stem Cell* *13*, 237–245.
- Zhu, J.K. (2009). Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* *43*, 143–166.
- Zhu, B., Zheng, Y., Hess, D., Angliker, H., Schwarz, S., Siegmund, M., Thiry, S., and Jost, J.P. (2000). 5-methylcytosine-DNA glycosylase activity is present in a cloned G/T mismatch DNA glycosylase associated with the chicken embryo DNA demethylation complex. *Proc. Natl. Acad. Sci. USA* *97*, 5135–5139.
- Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008). Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* *456*, 125–129.
- Ziller, M.J., Gu, H., Müller, F., Donaghey, J., Tsai, L.T., Kohlbacher, O., De Jager, P.L., Rosen, E.D., Bennett, D.A., Bernstein, B.E., et al. (2013). Charting a dynamic DNA methylation landscape of the human genome. *Nature* *500*, 477–481.