Tet family proteins and 5-hydroxymethylcytosine in development and disease

Li Tan¹ and Yujiang Geno Shi¹².*

Summary
Over the past few decades, DNA methylation at the 5-position of cytosine (5-methylcytosine, 5mC) has emerged as an important epigenetic modification that plays essential roles in development, aging and disease. However, the mechanisms controlling 5mC dynamics remain elusive. Recent studies have shown that ten-eleven translocation (Tet) proteins can catalyze 5mC oxidation and generate 5mC derivatives, including 5-hydroxymethylcytosine (5hmC). The exciting discovery of these novel 5mC derivatives has begun to shed light on the dynamic nature of 5mC, and emerging evidence has shown that Tet family proteins and 5hmC are involved in normal development as well as in many diseases. In this Primer we provide an overview of the role of Tet family proteins and 5hmC in development and cancer.

Key words: Tet, 5hmC, DNA methylation, Cancer

Introduction
DNA methylation at the 5-position of cytosine (5-methylcytosine; 5mC) is one of the key epigenetic marks that play a crucial role in development and genome regulation (Bestor and Coxon, 1993; Bird, 2002; Cedar and Bergman, 2009; Suzuki and Bird, 2008). It has long been known that distinct genomic regions are differentially methylated depending on cell or tissue type and developmental stage (Suzuki and Bird, 2008). In addition to the establishment of lineage-specific DNA ‘methylomes’ during mammalian development, two waves of global DNA demethylation have also been reported: one in the fertilized zygote and the other during primordial germ cell development (Suzuki and Bird, 2008; Wu and Zhang, 2010). The enzymes that catalyze DNA methylation, namely the DNA methyltransferases (DNMTs) DNMT1, DNMT3A, DNMT3B and the regulatory subunit DNMT3L, have been identified and well characterized (Bestor et al., 1988; Bourc’his et al., 2001; Okano et al., 1998). However, there remain fundamental questions as to how genome-wide DNA methylation is differentially regulated and dynamically processed during development and to what extent the regulation of DNA methylation might be linked to downstream gene expression programs that are important for developmental processes.

The 5-hydroxymethylcytosine (5hmC) epigenetic mark was first identified in the T-even bacteriophage almost six decades ago (Wyatt and Cohen, 1953), and it was later found in the vertebrate brain and in several other tissues (Globisch et al., 2010; Kriaucionis and Heintz, 2009; Nestor et al., 2011; Penn et al., 1972; Song, C. X. et al., 2011). These studies suggested a significant biological role for 5hmC in vertebrates. Interestingly, although 5hmC exists in mouse embryonic stem (ES) cells at high levels, it decreases significantly after ES cell differentiation (Szwagierczak et al., 2010; Tahiliani et al., 2009) but rises again in terminally differentiated cells, such as Purkinje neurons (Kriaucionis and Heintz, 2009). Very recently, it was shown that 5hmC exists in mouse, bovine and rabbit zygotes, and that 5hmC accumulates specifically in the paternal pronucleus coinciding with a reduction in 5mC (Iqbal et al., 2011; Wossidlo et al., 2011), which suggests a potential biological function of 5hmC and a role in the regulation of DNA methylation dynamics in early development.

Recently, human ten-eleven translocation 1 (TET1), a member of the Tet family of proteins, was identified as a 5mC dioxygenase responsible for catalyzing the conversion of 5mC to 5hmC (Tahiliani et al., 2009). The mammalian Tet family contains three members, Tet1, Tet2 and Tet3, all of which share a high degree of homology within their C-terminal catalytic domain (Iyer et al., 2009; Loenarz and Schofield, 2009). The discovery of this family of enzymes suggested a potentially novel mechanism for the regulation of DNA methylation, with 5hmC acting as an intermediate during DNA demethylation, although the biology and regulation of 5hmC and Tet family enzymes during development remain elusive. Furthermore, it was shown that Tet proteins can oxidize 5mC or 5hmC further and convert them to 5-formylcytosine (5fC) and/or 5-carboxylcytosine (5caC) (He et al., 2011; Ito et al., 2011), adding another layer of complexity to uncovering the potential function of Tet family enzymes and 5hmC in epigenetic regulation during development. In this Primer, we provide an overview of Tet family proteins, focusing on the developmental roles of Tet proteins and 5hmC in mice. We also summarize the emerging roles for Tet proteins in human cancers.

Domain structure of Tet family proteins
The mouse Tet gene family has three members: Tet1, Tet2 and Tet3 (Ito et al., 2010; Tahiliani et al., 2009). All Tet proteins contain a catalytic C-terminal CD domain (Cys-rich and DSBH regions) that belongs to the Cupin-like dioxygenase superfamly and exhibits 2-oxoglutarate (2-OG)- and iron (II)-dependent dioxygenase activity (Fig. 1). Tet proteins oxidize 5mC into 5hmC via these CD domains and require α-ketoglutarate as a co-substrate for enzymatic activity (Tahiliani et al., 2009). Recently, using radiolabeled thin-layer chromatography and high-sensitivity HPLC/mass spectrometry assays, Zhang and colleagues demonstrated that the CD domain of Tet proteins can also oxidize 5mC and 5hmC into 5fC and/or 5caC, although at very low levels (Ito et al., 2011). In addition, Xu and colleagues reported that Tet proteins can substantially oxidize 5mC and 5hmC into 5caC via their CD domains in the presence of ATP in vitro (He et al., 2011).

Another distinct feature of Tet family proteins is the CXXC zinc-finger domain, which was first identified and defined in DNMT1 (Bestor, 1992). A CXXC domain can be found in the N-terminus...
Development 139 (11) 1896

Mechanism of action of Tet proteins: a role for 5hmC in the regulation of DNA methylation and gene expression

Recent studies suggest that there might be multiple pathways or mechanisms by which 5mC and Tet family proteins regulate DNA methylation dynamics and gene transcription (Fig. 3). Song and colleagues presented a model in which 5hmC is more sensitive than 5mC to deamination by activation-induced deaminase (AID), and the deamination product, 5-hydroxymethyluracil (5hmU), activates base-excision repair (BER) pathway-mediated demethylation (Guo et al., 2011), suggesting one route to achieve Tet protein-mediated active DNA demethylation (Fig. 3). Furthermore, recent findings have shown that Tet proteins can oxidize 5mC not only to 5hmC but also 5fC and 5caC, which can subsequently be recognized and excised by thymine DNA glycosylase (TDG) in vitro and in vivo (He et al., 2011; Ito et al., 2011; Pfaffeneder et al., 2011; Zhang et al., 2012). In addition to these mechanisms of active DNA demethylation, the recent finding that 5hmC, 5fC and 5caC associated with the paternal genome in the zygote are gradually lost during preimplantation development suggested a dominant mechanism of passive DNA demethylation at this developmental stage (Inoue and Zhang, 2011; Inoue et al., 2011). Although these studies propose alternative routes for Tet-mediated active or passive DNA demethylation, it is suggested that the initial oxidation of 5mC to 5hmC by Tet family proteins is a prerequisite for the subsequent demethylation processes, regardless of how these final steps are mediated (e.g. by deamination, BER or TDG action), to complete the cycle of DNA demethylation (He et al., 2011; Ito et al., 2011; Maiti and Drohat, 2011). These studies thus suggest that 5hmC acts as an intermediate during DNA demethylation and that Tet proteins might function as DNA rate-limiting demethylation regulators. However, these studies also raise a number of questions (see Box 1). For example, why has 5hmC been detected as a relatively stable epigenetic mark in ES cells as well as in many normal tissues if the primary role of 5hmC is as an intermediate of 5mC demethylation? Why are 5fC and 5caC much less abundant than 5hmC in the investigated cells and tissues if they are also intermediates of 5mC demethylation? It is of interest to investigate how the extent of Tet-catalyzed 5mC oxidation is regulated in vivo and whether there exist DNA decarboxylases for 5caC or DNA deformylases for 5fC.

Alternatively, 5hmC might also function as a new landmark in the epigenetic landscape and may recruit specific readers that subsequently direct the dynamic remodeling and organization of the chromatin structures that are important for proper gene transcription. Although many methyl-CpG-binding domain (MBD) proteins [such as MBD1, MBD2, methyl CpG-binding protein 2 (MECP2) and

Fig. 1. Structure and function of mouse Tet family proteins. The mouse Tet protein family contains three members: Tet1, Tet2 and Tet3. All have a C-terminal CD domain (containing the Cys-rich and DSBH regions) that exhibits 2-oxoglutarate (2-OG)- and iron (II)-dependent dioxygenase activity. The CD domain also includes a spacer region, the length of which varies between Tet family members. The N-termini of Tet1 and Tet3, but not Tet2, contain a CXXC domain, which mediates their direct DNA-binding ability (Zhang et al., 2010; Xu, Y. et al., 2011). There is also a unique PRK12323 (DNA polymerase III subunits gamma and tau, provisional) domain in Tet3. This PRK12323 domain is not assigned to any domain superfamily and its function remains unknown. a.a., amino acids.

domain. This unique spacer region is common to all Tet family members, although its length varies (Fig. 1). The functional importance of this domain is currently unknown. Interestingly, Upadhyay et al. revealed that the spacer region of Tet1 has significant sequence similarity to the C-terminal domain (CTD) of RNA polymerase II of *Saccharomyces cerevisiae* (Upadhyay et al., 2011). Although not identical, many key residues important for post-translational protein modification, such as phosphorylation and methylation, are invariant and conserved between the two sequences. Furthermore, *TET2* is one of the most frequently mutated genes in myelodysplastic syndrome (MDS) (Kosmider et al., 2009), and about one-quarter of mutations were found in the region corresponding to the spacer domain of *TET2*, highlighting the functional importance of this region (Ko et al., 2010). Although future structural, deletion/mutational and/or biochemical analyses are required to uncover its function, it is possible that the spacer region of Tet family proteins, similar to the CTD of RNA polymerase II, is post-translationally modified, which might in turn play a ‘switch’ role in orchestrating the complex action of Tet proteins on 5mC or during targeted gene regulation.

In summary, these studies have shown that Tet family proteins, via their CD domains, possess an intrinsic enzymatic capability to convert 5mC to 5hmC, 5fC or 5caC depending on the presence of distinct co-factors such as ATP (Fig. 2). Tet proteins also contain a number of additional domains with as yet undefined functions. However, it appears that the Tet family proteins alone are not able to complete the demethylation of 5mC to unmodified cytosine.

of Tet1 and Tet3 but not in Tet2 (Fig. 1). Unlike the CXXC domains in other proteins [for example, DNMT1, myeloid/lymphoid or mixed-lineage leukemia (MLL) and CXXC finger protein 1 (CFP1, or CXXC1)], which are known to bind unmethylated CpG dinucleotides (Cierpicki et al., 2010; Song, J. et al., 2011; Xu, C. et al., 2011), the function of this domain in Tet1 and Tet3 is largely unknown. It is known, however, that the CXXC domain of TET1 recognizes not only unmodified cytosine but also 5mC and 5hmC, and it prefers to bind to regions in the genome of high CpG content (Zhang et al., 2010; Xu, Y. et al., 2011). Consistent with this feature, genome-wide mapping of Tet1 binding by ChIP-seq approaches revealed its enrichment around transcription start sites (TSSs) in mouse ES cells (Xu, Y. et al., 2011; Williams et al., 2011; Wu et al., 2011b). Although future investigation is required to elucidate the importance of this domain in Tet1 or Tet3 function, we are tempted to speculate that the CXXC domain serves as the functional unit that targets these enzymes to specific genomic regions for their action.

In addition to the above-described functional domains responsible for executing Tet protein action, there is a spacer region that bridges the two parts of the disconnected DSBH enzymatic domain. This unique spacer region is common to all Tet family members, although its length varies (Fig. 1). The functional importance of this spacer region is currently unknown. Interestingly, Upadhyay et al. revealed that the spacer region of Tet1 has significant sequence similarity to the C-terminal domain (CTD) of RNA polymerase II of *Saccharomyces cerevisiae* (Upadhyay et al., 2011). Although not identical, many key residues important for post-translational protein modification, such as phosphorylation and methylation, are invariant and conserved between the two sequences. Furthermore, *TET2* is one of the most frequently mutated genes in myelodysplastic syndrome (MDS) (Kosmider et al., 2009), and about one-quarter of mutations were found in the region corresponding to the spacer domain of *TET2*,
MBD4] do not bind 5hmC in vitro (Jin et al., 2010), a recent study proposed that Tet1 is associated with the Mbd3/NuRD complex in mouse ES cells and that Mbd3 can bind 5hmC and unmodified cytosine, but not 5mC, in vitro (Yildirim et al., 2011). Given that most, if not all, 5hmC comes from 5mC (Ficz et al., 2011), establishing DNA methylation at the hydroxymethylated sites during DNA replication appears to be a prerequisite for 5hmC maintenance. UHRF1 (ubiquitin-like, containing PHD and RING finger domains 1) is required for maintenance of methylation by interacting with DNMT1 and hemi-methylated CpG (Bostic et al., 2007; Sharif et al., 2007). The finding that UHRF1 can bind both 5mC and 5hmC (Frauer et al., 2011) provides a possible mechanism by which UHRF1 may also target DNMT1 to the hemi-hydroxymethylated replication forks. More recently, Cheng and colleagues showed that DNMT1 has a high intrinsic activity toward the hemi-methylated CpG substrate but a greatly reduced activity for the hemi-hydroxymethylated CpG substrate in methyl transfer assays in vitro (Hashimoto et al., 2012). Interestingly, they found that DNMT3A and DNMT3B exhibited equal activities on unmodified, hemi-methylated and hemi-hydroxymethylated CpG substrates (Hashimoto et al., 2012). Collectively, these studies suggest that, instead of being merely an intermediate of DNA demethylation, 5hmC acts as a stable epigenetic mark that directly influences genome structure and function. Future experiments that aim to identify additional specific 5hmC ‘readers’ and the cellular complexes associated with 5hmC should distinguish between 5mC and 5hmC in recruiting their specific readers (see Box 1). In addition, such studies will provide new insight into our understanding of the multiple mechanisms of action of 5hmC in the dynamics of DNA methylation and chromatin structure.

Although 5hmC has been implicated as either a stable epigenetic modification (the ‘sixth base’) or an intermediate of DNA demethylation, correlation between genome-wide 5hmC enrichment and gene transcription activity in mouse ES cells and brain tissue remains elusive and the findings remain somewhat controversial (Ficz et al., 2011; Jin et al., 2011a; Pastor et al., 2011; Robertson et al., 2011; Song, C. X. et al., 2011; Stroud et al., 2011; Williams et al., 2011; Wu et al., 2011a; Xu, Y. et al., 2011). Zhang and colleagues showed that the knockdown of Tet1 by shRNA increases the 5mC level on Nanog promoters and repress the transcription of Nanog mRNA, which subsequently skews the mouse ES cells toward a differentiated state (Ito et al., 2010). A recent study also showed that acute depletion of Tet1-dependent 5hmC levels by shRNA impairs leukemia inhibitory factor (LIF)/Stat3 signaling and results in a loss of ES cell identity. This supports the idea that Tet1 is essential for the regulation of key genes involved in ES cell self-renewal and lineage determination (Freudenberg et al., 2011). However, knockout of Tet1 by homologous recombination or knockdown of Tet1 by shRNA performed by other laboratories did not affect Nanog expression or the self-renewal of mouse ES cells (Dawlaty et al., 2011; Koh et al., 2011; Williams et al., 2011; Xu, Y. et al., 2011). Thus, the involvement of Tet family proteins and 5hmC in ES cell self-renewal or pluri potency can be determined only if we can clearly illustrate the mechanism by which 5hmC and Tet are involved in the regulation of those critical genes.

**Box 1. Tet proteins and 5hmC in development and cancer: open questions**

**Tet-mediated DNA demethylation: active or passive?**
Active DNA demethylation occurs during at least two stages of mouse development: in fertilized zygotes and in primordial germ cells (PGCs). In fertilized zygotes, Tet3-mediated 5mC oxidation can lead to active demethylation (with the help of some DNA repair enzymes) or may result in passive demethylation due to rapid DNA replication in the early embryo. Do these mechanisms also apply to the global demethylation that occurs in PGCs? Are there decarboxylases that directly convert 5caC to unmodified cytosine to complete the demethylation of 5mC?

**What are the 5hmC readers?**
Many 5mC-binding proteins and other proteins function as readers of 5mC, whereas little is known about how the 5hmC mark is recognized. Does a true 5hmC-specific reader exist? The identification of 5hmC-specific binding proteins will help to decipher the functions of 5hmC and 5mC.

**How do the catalytic activity-independent functions of Tet proteins contribute to their role?**
Recent work has revealed that Tet1 can exert catalytic activity-independent functions. How Tet proteins act to regulate gene transcription in a catalytic activity-independent fashion needs to be investigated, as does the distinction between the catalytic activity-dependent and -independent functions of Tet proteins during mouse development and in disease.

**Are there evolutionary roles for 5hmC?**
As an epigenetic modification based on 5mC, 5hmC might have a less fundamental role than its precursor and might rather serve to increase the complexity of DNA methylation regulation in mammals. Does 5hmC share the same history as 5mC during evolution? When did Tet proteins and 5hmC appear in evolutionary history? Given that brain tissues exhibit the highest levels of 5hmC in adult mice, have 5hmC and Tet proteins contributed to the increasingly complex function (such as in neural circuitry and behaviors) of mammals during evolution?

**How does 5hmC function in cancer?**
Evidence suggests that 5hmC is greatly reduced or suppressed in multiple cancers. However, what is the role of 5hmC in cancer prevention and/or tumorigenesis? How is 5hmC downregulated in cancers? Mechanistically, answering these questions might offer new strategies for anti-cancer drug development and cancer therapy.

**Fig. 2. Oxidative reactions carried out by Tet proteins.** Tet family proteins oxidize 5mC into 5hmC (and to 5fC and 5caC at low levels) in the presence of 2-OG and iron. 5mC and 5hmC can be further oxidized mainly to 5caC by Tet proteins if the reaction system is supplemented with ATP. 5mC, 5-methylcytosine; 5hmC, 5-hydroxymethylcytosine; 5fC, 5-formylcytosine; 5caC, 5-carboxylcytosine.
the CXXC domain, or the dioxygenase activities of each with which the Tet family members associate, the DNA-binding expression via different mechanisms. The multi-protein complexes involved in the regulation of DNA methylation and/or gene activity of the Tet family proteins is important for their function, a catalytic activity-independent manner. These results suggest that Tet1 might repress gene transcription in the upregulation of Tet1 target genes upon Tet1 knockdown can mouse ES cells (Williams et al., 2011). Importantly, they found that and colocalizes with the Sin3a co-repressor complex in 293T and that study, Helin and colleagues demonstrated that Tet1 associates independently of their catalytic activity (Williams et al., 2011). In Finally, although many studies have indicated that the enzymatic activity of the Tet family proteins is important for their function, a recent study suggested that Tet proteins might also exert functions independently of their catalytic activity (Williams et al., 2011). In that study, Helin and colleagues demonstrated that Tet1 associates and colocalizes with the Sm3a co-repressor complex in 293T and mouse ES cells (Williams et al., 2011). Importantly, they found that the upregulation of Tet1 target genes upon Tet1 knockdown can also be observed in DNMT triple knockout ES cells in which both 5mC and 5hmC modifications are absent (Williams et al., 2011). These results suggest that Tet1 might repress gene transcription in a catalytic activity-independent manner. Taken together, it is likely that different Tet family members are involved in the regulation of DNA methylation and/or gene expression via different mechanisms. The multi-protein complexes with which the Tet family members associate, the DNA-binding activity of the CXXC domain, or the dioxygenase activities of each Tet member serve as potential molecular mechanisms that differentially contribute to the regulation of DNA methylation status and gene expression. Future studies are needed to clarify how Tet proteins are regulated by their post-translational modifications or their associated co-factors. Furthermore, the spacer region of Tet family proteins is currently understudied. Its potential regulatory function, e.g., its role in orchestrating the formation of Tet complexes and modulating Tet enzymatic activities, warrants future investigation.

### The role of Tet family proteins and 5hmC during mouse development

Dramatic epigenetic reprogramming, including DNA methylation and demethylation, is essential for mammalian development. Recently, the developmental roles of Tet family proteins and 5hmC have been investigated using gene knockout strategies (see Table 1). Below, we summarize the function of Tet family proteins at different stages throughout development.

#### Tet proteins in primordial germ cells and gametes

Although embryonic development starts with oocyte fertilization, the oocyte and sperm develop from primordial germ cells (PGCs). Mouse PGCs are specified from epiblast cells and migrate into the gonad region at embryonic day (E) 7.25. This is followed by widespread epigenetic reprogramming at E11.5, which includes genome-wide DNA demethylation, erasure of genomic imprints, and large-scale chromatin remodeling (Hajkova et al., 2008; Kota and Feil, 2011). Both Tet1 and Tet2 are expressed in E11.5 and E12.5 PGCs but less so in oocytes and zygotes, whereas Tet3 is highly expressed in oocytes but not in PGCs (Gu et al., 2011; Hajkova et al., 2010). The development of PGCs, and subsequently of spermatocytes or oocytes, is normal in Tet1 or Tet2 knockout mice, and these knockout adults display similar fecundity to wild-type mice (Dawlaty et al., 2011; Li et al., 2011; Ko et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011). Xu and colleagues established PGC-conditional Tet3 knockout mice and found that oogenesis and spermatogenesis are normal when Tet3 is absent in PGCs (Gu et al., 2011). These findings support the conclusion that, despite the dramatic reprogramming events that occur in PGCs and gametes, the knockout of Tet1, Tet2 or Tet3 alone has no significant effect on the development of PGCs and gametes.

#### Tet proteins in the zygote

DNA demethylation also occurs in the zygote, when the paternal genome loses its DNA methylation marks rapidly after fertilization (Oswald et al., 2000). It was shown recently that 5mC in the paternal but not the maternal pronucleus is oxidized to 5hmC (Gu et al., 2011). The development of PGCs but less so in oocytes and zygotes, whereas Tet3 is highly expressed in oocytes but not in PGCs (Gu et al., 2011; Hajkova et al., 2010). The development of PGCs, and subsequently of spermatocytes or oocytes, is normal in Tet1 or Tet2 knockout mice, and these knockout adults display similar fecundity to wild-type mice (Dawlaty et al., 2011; Li et al., 2011; Ko et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011). Xu and colleagues established PGC-conditional Tet3 knockout mice and found that oogenesis and spermatogenesis are normal when Tet3 is absent in PGCs (Gu et al., 2011). These findings support the conclusion that, despite the dramatic reprogramming events that occur in PGCs and gametes, the knockout of Tet1, Tet2 or Tet3 alone has no significant effect on the development of PGCs and gametes.

#### Tet proteins in the zygote

DNA demethylation also occurs in the zygote, when the paternal genome loses its DNA methylation marks rapidly after fertilization (Oswald et al., 2000). It was shown recently that 5mC in the paternal but not the maternal pronucleus is oxidized to 5hmC (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). The expression levels of Tet1 and Tet2 are very low in oocytes or

### Table 1. Phenotypes of Tet knockout mice

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Preimplantation development</th>
<th>Postimplantation development</th>
<th>Postnatal development</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tet1&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>Normal</td>
<td>Small body size</td>
<td>Small body size</td>
<td>(Dawlaty et al., 2011)</td>
</tr>
<tr>
<td>Tet2&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>Normal</td>
<td>Normal</td>
<td>Spontaneous myeloid leukemia</td>
<td>(Ko et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011)</td>
</tr>
<tr>
<td>Tet3&lt;sup&gt;−/−&lt;/sup&gt;</td>
<td>Normal</td>
<td>Normal</td>
<td>Neonatal lethality</td>
<td>(Gu et al., 2011)</td>
</tr>
<tr>
<td>Tet3&lt;sup&gt;mat−/pat+&lt;/sup&gt;</td>
<td>Normal, but blockage in paternal genome reprogramming</td>
<td>Normal</td>
<td>Neonatal lethality and small body size</td>
<td>(Gu et al., 2011)</td>
</tr>
</tbody>
</table>

---

**Fig. 3. Involvement of Tet family proteins in the regulation of DNA methylation and demethylation: potential mechanisms.** Tet family proteins (TET1/2/3) catalyze 5mC oxidation to 5hmC. Tet family proteins can also convert 5mC or 5hmC further into 5fC and 5caC, and the latter may be recognized and excised by thymine DNA glycosylase (TDG) to generate thymine and thereby complete demethylation. Alternatively, because 5hmC is more sensitive than 5mC to deamination (via activation-induced deaminase, AID), it can be converted to 5-hydroxymethyluracil (5hmU), which in turn can be converted to cytosine following base-excision repair (BER) pathway-mediated demethylation (Guo et al., 2011). It remains unknown whether there are decarboxylases or deformylases that can remove the modification directly. Importantly, 5mC and its oxidative derivatives may undergo passive demethylation by dilution during DNA replication (Inoue and Zhang, 2011; Inoue et al., 2011).
fertilized zygotes (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). By contrast, Tet3 is highly expressed in oocytes and fertilized zygotes but disappears rapidly during cleavage (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). In line with this, Xu and colleagues showed recently that maternal Tet3 is the key enzyme catalyzing 5mC into 5hmC in the paternal genome (paternal pronuclei) of developing zygotes (Gu et al., 2011). DNA demethylation and 5hmC generation can also occur in the maternal pronuclei of Stella (Dppa3) null zygotes, suggesting that the Stella protein protects the maternal pronucleus from DNA demethylation and oxidation (Nakamura et al., 2007; Wossidlo et al., 2011). Similarly, Tet3-mediated 5mC oxidation in oocytes is also required for epigenetic reprogramming of the donor nuclear DNA during somatic cell nuclear transfer (Gu et al., 2011). Furthermore, Zhang and colleagues observed that 5hmC and other 5mC derivatives can be simply eliminated by passive dilution through DNA replication during the division of zygotes (Inoue and Zhang, 2011; Inoue et al., 2011). It is conceivable that converting 5mC to 5hmC via Tet3-mediated oxidation is a key step in initiating the zygotic ‘resetting’ of DNA methylation patterns, while subsequent passive dilution of 5hmC through DNA replication during zygote cleavage serves as one of the mechanisms for ultimate erasure of paternal 5mC (Fig. 4). Nonetheless, although these new findings shed light on the molecular mechanisms that underlie the first wave of DNA demethylation in the earliest developmental stage of life, it remains unclear to what extent 5mC oxidation, followed by passive DNA demethylation, accounts for the demethylation of parental pronuclear DNA in the zygote.

Tet proteins in pre- and postimplantation embryos and in ES cells

Both Tet1 and 5hmC levels are abundant in mouse ES cells, in the inner cell mass of the mouse blastocyst, and in cells of the early mouse epiblast (Ito et al., 2010; Koh et al., 2011; Ruzov et al., 2011). Recent work from Ito et al. found that the shRNA-mediated knockdown of Tet1, but not Tet2 or Tet3, led to the spontaneous differentiation of mouse ES cells even in the presence of LIF (Ito et al., 2010). However, others believe that Tet1 is not required for maintaining the self-renewal of mouse ES cells (Koh et al., 2011; Williams et al., 2011; Xu, Y. et al., 2011); recently, Jaenisch and colleagues found that Tet1 null ES cells maintain their self-renewal ability under mouse ES cell culture conditions (LIF plus fetal bovine serum) in vitro and develop normally in vivo (Dawlaty et al., 2011). Furthermore, whereas both knockdown and knockout of Tet1 force the transdifferentiation of mouse ES cells into trophoblasts in vitro, such effects were not detected in vivo (Dawlaty et al., 2011; Ito et al., 2010; Koh et al., 2011), and Tet1+− and Tet2−/− mice appear to develop normally, appear healthy through adulthood and are fertile (Dawlaty et al., 2011; Ko et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011) (Table 1). It is still unclear why Tet1 expression and 5hmC are maintained at high levels in mouse ES cells. As such, the biological role of Tet1/Tet2 or 5hmC in postimplantation embryonic development remains unclear and controversial.

One explanation for the discrepancies between different research groups regarding the functional roles of Tet1 and 5hmC in postimplantation embryonic development might be that other Tet proteins play compensatory roles, as knockout or knockdown of Tet1 or Tet2 results in only a partial decrease in 5hmC levels (Dawlaty et al., 2011; Ito et al., 2010; Koh et al., 2011; Xu, Y. et al., 2011). Interestingly, homozygous mutation of Tet3 led to neonatal lethality, highlighting its importance in the development of multiple organs during embryogenesis (Gu et al., 2011) (Table 1). Thus, Tet family proteins may have important yet redundant roles in the precise regulation of postimplantation embryonic development.

Tet proteins in postnatal development

Although the genomic DNA sequence is the same, each organ or tissue in the mouse has a tissue-specific DNA methylome (Liang et al., 2011; Previti et al., 2009). As a derivative of 5mC, 5hmC can be detected in almost all organs and tissues, but unlike the relatively constant levels of 5mC, 5hmC levels vary among tissues (Ito et al., 2010; Globisch et al., 2010; Li et al., 2011). For example, in the brain, genome-wide analysis of 5hmC distribution and localization revealed that 5hmC levels in the cerebellum and hippocampus of adult mice (6 weeks and 1 year of age) are higher than those observed in postnatal day (P) 7 mice (Szulwach et al., 2011). MECP2 is a methyl-CpG-binding protein that plays an important role in neural development and brain function (Lewis et al., 1992). Accordingly, mutations in human MECP2 cause the neurodevelopmental disorder Rett syndrome (Amir et al., 1999). Using MeCP2 knockout and knockin mice, Szulwach et al. found that the overall abundance of 5hmC in the cerebellum was negatively correlated with MeCP2 dosage (Szulwach et al., 2011). These studies suggested that 5hmC dynamics might be involved in postnatal neurodevelopment and brain function and, when abnormal, may lead to neural disorders. Tet family genes are expressed in multiple regions of the brain and therefore it will be worthwhile to carefully analyze the brain structures, neurophysiology and behaviors of Tet knockout mice to uncover the potential function of Tet proteins in brain development, function and pathology (see Box 1).

Recent evidence suggests that Tet2 and 5hmC are also associated with postnatal development of the hematopoietic lineage, and alteration of Tet2 expression or the levels of 5hmC may disrupt the homeostasis of cells and ultimately lead to abnormalities in

Fig. 4. Tet gene expression and 5mC and 5hmC levels during mouse preimplantation development. The expression levels of Tet1 and Tet3 are very low in oocytes and fertilized zygotes. By contrast, Tet3 is highly expressed in oocytes and fertilized zygotes but disappears rapidly during cleavage (Gu et al., 2011; Iqbal et al., 2011; Wossidlo et al., 2011). Tet3-mediated 5mC oxidation and subsequent dilution by DNA replication during embryo cleavage might contribute to the previous ‘active’ DNA demethylation observed in the developing preimplantation embryo (Gu et al., 2011; Inoue and Zhang, 2011). The expression levels of Tet1 and Tet2 increase during preimplantation development and both are highly expressed in the inner cell mass of the blastocyst, which also exhibits high 5hmC levels (Ruzov et al., 2011).
hematopoietic cell proliferation and maturation (Abdel-Wahab et al., 2009; Albano et al., 2011; Chou et al., 2011; Euba et al., 2011; Figueroa et al., 2010; Pronier et al., 2011; Weissmann et al., 2011). Although Tet2−/− mice appeared to develop normally overall, they displayed an abnormality in myeloid lineage differentiation and developed leukemia spontaneously (Ko et al., 2011; Li et al., 2011; Moran-Crusio et al., 2011; Quivoron et al., 2011). Since the Tet3−/− phenotype was neonatal lethal, the exact functions of Tet3 protein in postnatal development are largely unknown and need further investigation (Gu et al., 2011). We believe that conditional knockout of Tet3 in various tissues or organs postnatally will provide a good strategy for probing the biological role of Tet3 during mouse postnatal development.

**Tet family genes and 5hmC in human cancers**

*TET1* was originally discovered as a gene that, as result of a chromosome translocation, was fused to the *MLL* gene in certain leukemia patients (Lorsbach et al., 2003). However, the role of *TET1* in hematopoietic development and leukemogenesis remains ambiguous. By contrast, *TET2*, although only recently identified, is the most studied member of the Tet family genes in MDS and other types of leukemia. It has been reported as one of the most frequently mutated genes in myeloid malignancies (Abdel-Wahab et al., 2009; Albano et al., 2011; Chou et al., 2011; Euba et al., 2011; Figueroa et al., 2010; Weissmann et al., 2011). Whereas some of the mutation sites in *TET2* correspond to residues within the catalytic domain and could thus impair catalytic activity, many mutations appear unrelated to enzymatic activity (Ko et al., 2010). Therefore, the molecular mechanism underlying *TET2* loss-of-function in leukemogenesis is still unknown and requires further investigation.

So far, a direct correlation between TET3 and cancer has not been reported. A possible explanation for this is that, among the three TET proteins, TET3 is the most important regulator and critical *TET3* mutations or dysregulation might lead to lethality. Besides the hematopoietic malignancies, 5hmC levels are decreased in a broad range of solid tumors, including glioma, colon cancer, breast cancer and melanoma (Li and Liu, 2011; Jin et al., 2011b; Haffner et al., 2011; Xu, W. et al., 2011). Since global DNA hypomethylation and locus-specific DNA hypermethylation have been identified as key features of many cancers (Feinberg et al., 2006; Suzuki and Bird, 2008), investigating whether low 5hmC levels contribute to aberrant DNA methylation patterns or other aspects in cancers could prove to be very informative (see Box 1).

**Conclusions**

The role of TET proteins and 5hmC in development and disease has been widely studied in the past three years since the first biochemical connection between Tet proteins and 5hmC was established. In the future, studies using conditional knockouts of Tet genes will decipher the stage- and organ-specific functions of Tet proteins and 5hmC in development. Double or triple knockout of Tet genes may circumvent their redundant/compensatory effects on 5mC oxidation and gene regulation, offering a better understanding of the role of 5hmC and Tet family genes in development. Moreover, further investigations are needed to distinguish the catalytic activity-dependent and -independent functions of Tet proteins. Since most studies of Tet proteins and 5hmC have focused on mice, rats and humans, expanding these investigations to other model systems (such as *Xenopus*, zebrafish and even non-vertebrates) will elucidate the evolutionarily conserved functions of these proteins. Finally, although the current data from solid and hematopoietic tumors support a role for 5hmC as an epigenetic modification requisite for preventing tumorigenesis, it remains unknown whether the decrease in 5hmC level is a cause or a consequence of preventing tumorigenesis. In summary, many important issues need to be explored in the future (see Box 1). Clearly, the roles and mechanisms of action of 5hmC and Tet family members in ES cell biology and during embryonic development, as well as in cancer biology, will be the main areas of focus. In the next five years, new and exciting discoveries in these areas are expected to emerge.

**Acknowledgements**

We thank members of the Y.G.S. laboratory for helpful discussion and Abraham Khorasani for critical reading of the manuscript. We apologize to those colleagues whose publications could not be cited owing to space limitations.

**Funding**

Work in the Y.G.S. laboratory is supported by the National Institutes of Health and partly by the ‘985’ Project of the Ministry of Education of China and ‘973’ National Basic Research Program of China. Y.G.S. is a Pew scholar. Deposited in PMC for release after 12 months.

**Competing interests statement**

The authors declare no competing financial interests.

**References**


Differential expression of DNA methylation enzymes in mouse embryonic stem cells.


Lorsbach, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T. and Downing, J. R. (2003). TET1 is a member of a novel protein family, is fused to MLL in adult myeloid leukemia containing the t(10;11)(q22;q23). Leukemia 17, 637-641.


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.

Song, C. X., Szulwach, K. E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C. H.,
Zheng, W., Jin, X. et al. (2011). Selective chemical labeling reveals the
genome-wide distribution of 5-hydroxymethylcytosine. Nat. Biotechnol. 29, 68-
72.

Song, J., Rechkoblit, O., Bestor, T. H. and Patel, D. J. (2011). Structure of
DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA
methylation. Science 331, 1036-1040.

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.

Surani, M. A. and Hajkova, P. (2010). Epigenetic reprogramming of mouse germ


Szulwach, K. E., Li, X., Li, Y, Song, C. X., Wu, H., Dai, Q., Irier, H., Upadhyay,
A. K., Gearing, M., Levey, A. I., VasanthaKumar, A. et al. (2011). S-hmC-
mediated epigenetic dynamics during postnatal neurodevelopment and aging.

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.,
methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner
TET1 Science 324, 930-935.

methyl-lyase erasure: structural and functional linkage of a Lumonji demethylase

Weissmann, S., Alpermann, T., Grossmann, V., Kowarsch, A., Nadarajah, N.,

Williams, K., Christensen, J., Pedersen, M. T., Johansen, J. V., Cloos, P. A.,
Rapp, T., and Helin, K. (2011). TET1 and hydroxymethylcytosine in

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

Wu, H., D’Alessio, A. C., Ito, S., Wang, Z., Cui, K., Zhao, K., Sun, Y. E. and
distribution reveals its dual function in transcriptional regulation in mouse
embryonic stem cells. Genes Dev. 5, 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

Wu, S. C. and Zhang, Y. (2010). Active DNA demethylation: many roads lead to

bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine.
Biochem. J. 55, 774-782.

Xu, C., Bian, C., Lam, R., Dong, A. and Min, J. (2011). The structural basis for
selective binding of non-methylated CpG islands by the CFP1 CXXC domain.

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

Zhang, H., Huang, S. et al. (2011). Genome-wide regulation of 5hmC, 5mC,
and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol.
Cell 42, 451-464.

Yildirim, O., Li, R., Hung, J. H., Chen, P. B., Dong, X., Es, L. S., Weng, Z.,
expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells.
Cell 147, 1498-1510.

TET1 is a DNA-binding protein that modulates DNA methylation and gene

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-