

# TET enzymes, TDG and the dynamics of DNA demethylation

Rahul M. Kohli<sup>1,2</sup> & Yi Zhang<sup>3,4,5,6,7</sup>

DNA methylation has a profound impact on genome stability, transcription and development. Although enzymes that catalyse DNA methylation have been well characterized, those that are involved in methyl group removal have remained elusive, until recently. The transformative discovery that ten-eleven translocation (TET) family enzymes can oxidize 5-methylcytosine has greatly advanced our understanding of DNA demethylation. 5-Hydroxymethylcytosine is a key nexus in demethylation that can either be passively depleted through DNA replication or actively reverted to cytosine through iterative oxidation and thymine DNA glycosylase (TDG)-mediated base excision repair. Methylation, oxidation and repair now offer a model for a complete cycle of dynamic cytosine modification, with mounting evidence for its significance in the biological processes known to involve active demethylation.

Two competing demands on the genome are the need for stability and for flexibility. Because the functional programs of cells are encoded by the genome, this information must be faithfully propagated both through development and across generations. At the same time, the genome of a multicellular organism must encode for diverse cell types, each of which must be capable of responding to a changing environment. These latter functions require the capacity for adaptive regulation of gene expression, which can be achieved by the transcription factor complexes that bind DNA, by the packaging of DNA into chromatin and by dynamic covalent modifications to either histones or DNA itself.

Covalent modification of DNA, in particular, helps to provide a means for functional variability while maintaining the information content of the base. One of the best-studied covalent modifications on DNA is 5-methylcytosine (5mC), a mark deposited by DNA methyltransferase (DNMT) enzymes<sup>1</sup>. In mammalian genomes, 5mC exists mostly in the CpG dinucleotide context and about 70–80% of CpGs are methylated. DNMTs can both introduce methylation marks (*de novo* methylation) and maintain them after the genome is replicated (maintenance methylation), making DNA methylation a long-term and potentially heritable mark<sup>1</sup>. Conventionally, 5mC is associated with a transcriptionally repressed chromatin state, and DNA methylation at specific genomic loci, including lineage-specific genes, can help to shape a cellular program during development<sup>2</sup>. 5mC-mediated long-term gene silencing also contributes to genomic imprinting, X-chromosome inactivation and suppression of mobile genetic elements<sup>3,4</sup>.

DNA methylation is relatively stable compared with most histone modifications. Nevertheless, loss of DNA methylation, or DNA demethylation, has been observed in different biological contexts and this alteration can take place actively or passively. Active DNA demethylation refers to an enzymatic process that removes or modifies the methyl group from 5mC. By contrast, passive DNA demethylation refers to loss of 5mC during successive rounds of replication in the absence of functional DNA methylation maintenance machinery. Although passive DNA demethylation is generally understood and accepted, the evidence for active DNA demethylation and how it occurs has been controversial<sup>5,6</sup>. In part, this controversy has been due to the cacophony of enzymes and pathways implicated in demethylation.

However, a series of recent discoveries has begun to harmonize, and thereby greatly advance, our understanding of active DNA demethylation. Here, we review these significant discoveries, their biological implications and the promising areas for further exploration.

## DNA demethylation and historical mechanisms

Several reviews have described the biological context in which active DNA demethylation may take place<sup>5–7</sup>. Establishing and editing genomic methylation patterns seems to be particularly relevant in several stages of mammalian embryogenesis. Initially, after the sperm penetrates the egg and before the merging of paternal and maternal genomes, the paternal genome goes through a complex remodelling process that includes deposition of histone H3.3 and remodelling of DNA methylation patterns<sup>8</sup>. Here, a rapid loss of 5mC staining is observed in the paternal, but not the maternal, genome, suggesting an active 5mC editing process<sup>9,10</sup>. After implantation, and early in development, a subset of posterior epiblast cells is instructed to become primordial germ cells (PGCs). PGCs have to go through a complex epigenetic reprogramming process, including erasure of genome-wide DNA methylation patterns<sup>11</sup>, to prepare them for germ-cell-specific processes, such as meiosis. Besides global loss of DNA methylation in zygotes and PGCs, DNA demethylation has also been observed at specific loci in rapid response to environmental stimuli or in post-mitotic cells, supporting the relevance of active demethylation in various biological settings in the absence of cellular replication<sup>12–14</sup>.

Many candidates from the known repertoire of DNA modifying enzymes have historically been proposed to function in DNA demethylation (see refs 5, 6, 15 and 16 for reviews). As we will discuss in the context of more recent discoveries, DNA cytosine deaminases that can introduce genomic mismatches, DNA glycosylases that can excise bases, other DNA repair factors and even DNMTs themselves have been suggested to be involved in DNA demethylation. Although there is some evidence to support a role for many of these DNA modifying pathways, these roles have often seemed specific to the individual biological system being examined. The lack of a unifying mechanistic process has led to ongoing dispute over the relative importance of these various pathways in DNA demethylation. Although these

<sup>1</sup>Department of Medicine, Raymond and Ruth Perelman School of Medicine, University of Pennsylvania, Philadelphia 19104, USA. <sup>2</sup>Department of Biochemistry and Biophysics, Raymond and Ruth Perelman School of Medicine, University of Pennsylvania, Philadelphia 19104, USA. <sup>3</sup>Howard Hughes Medical Institute, WAB-1496, 200 Longwood Ave., Boston, Massachusetts 02115, USA. <sup>4</sup>Program in Cellular and Molecular Medicine, Boston Children's Hospital, Boston, Massachusetts 02115, USA. <sup>5</sup>Division of Hematology/Oncology, Department of Pediatrics, Boston Children's Hospital, Boston, Massachusetts 02115, USA. <sup>6</sup>Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. <sup>7</sup>Harvard Stem Cell Institute, 200 Longwood Ave., Boston, Massachusetts 02115, USA.

multiple candidate pathways remain areas of active exploration<sup>5,6,15,16</sup>, in this Review we focus on recent developments that have brought new clarity to the field of DNA demethylation by elucidating pathways of oxidation-mediated demethylation.

### TET-mediated oxidation of 5mC

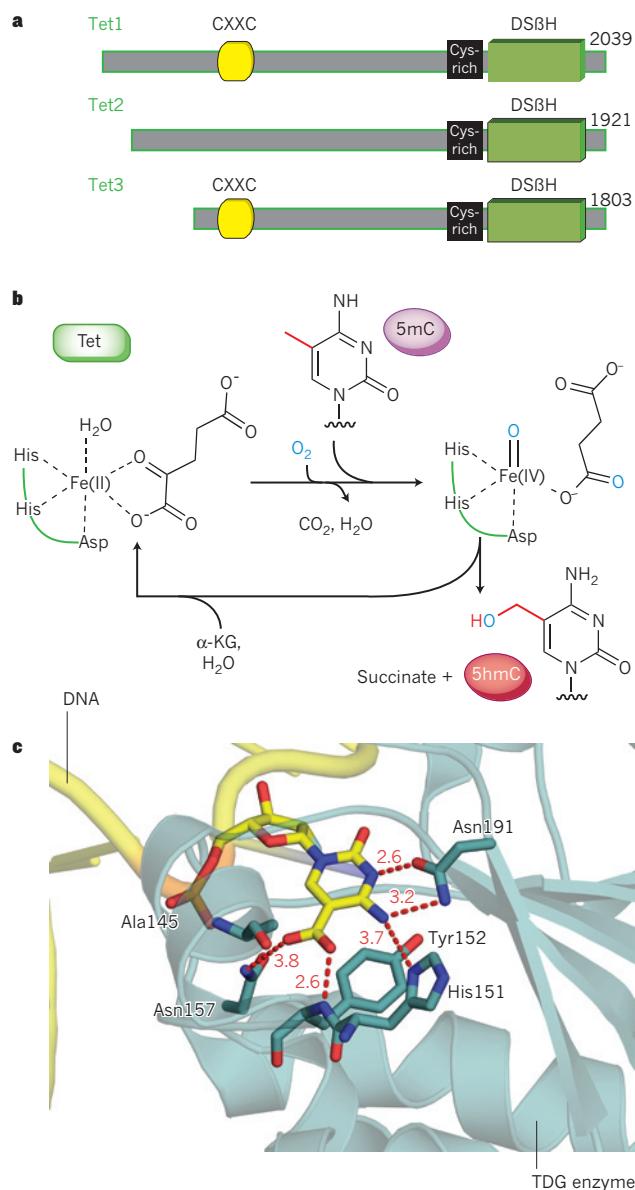
The discovery of a family of enzymes that can modify 5mC through oxidation was a watershed moment in advancing our understanding of DNA demethylation mechanisms, introducing 5-hydroxymethylcytosine (5hmC) as a key intermediate in active demethylation pathways<sup>17–19</sup>. This discovery was motivated by the study of two pathways involving oxidative modifications of T bases: one involving oxidative modifications to DNA, the other, demethylation of a nucleobase. In the parasite responsible for African sleeping sickness, *Trypanosoma brucei*, glucosylated 5-hydroxymethyluracil (Base J) functions in transcriptional regulation to modulate surface glycoprotein expression and thereby promote immune escape<sup>20</sup>. Base J biosynthesis involves oxidation of T within DNA to 5-hydroxymethyluracil (5hmU) by JBP1 and JBP2, members of the Fe(II)/ $\alpha$ -ketoglutarate ( $\alpha$ -KG)-dependent oxygenase family of enzymes. A second member of this oxygenase family, thymidine hydroxylase, acts instead on free T bases in a pyrimidine salvage pathway. Interestingly, the initial oxidation product, 5hmU, is subsequently further oxidized to 5-formyluracil and 5-carboxyluracil<sup>21</sup>. Decarboxylation by isoorotate decarboxylase<sup>22</sup> completes a cycle from T to U with potential mechanistic parallels to 5mC demethylation<sup>5</sup>.

Bioinformatic analyses by Rao and colleagues<sup>17,23</sup> revealed several mammalian paralogues of JBP1 and JBP2 (Fig. 1a). These enzymes belong to the TET family, which has previously been implicated in haematopoietic malignancies<sup>24</sup>. Surprisingly, overexpression of TET1 was associated with a reduction in genomic 5mC, suggesting that, unlike its paralogues, TET1 recognized modified C, rather than T, bases in DNA<sup>17</sup>. Indeed, purified TET enzymes modified oligonucleotide substrates containing 5mC through oxidation, and the product was authenticated as 5hmC<sup>17,18</sup> (Fig. 1b).

Although 5hmC had previously been observed in mammalian genomes<sup>25</sup>, these earlier observations did not receive attention until the discovery of TET enzymes, because these catalysts are capable of purposefully generating this oxidized base. Moreover, 5hmC was shown to be readily detectable in mouse embryonic stem (ES) cells in a manner dependent on expression of TET<sup>17,18</sup>. Even more strikingly, 5hmC was shown to be nearly 40% as abundant as 5mC in post-mitotic neuronal Purkinje cells<sup>19</sup>. Even though neuronal cells seem particularly enriched, accurate quantification methods have since demonstrated that 5hmC accumulates in most cell types, raising the possibility that this 'sixth base' in the genome may have a distinctive epigenetic role<sup>26,27</sup>.

Reflecting on the discovery of TET, two groups noted that the initial assay conditions used for detecting 5hmC were not permissive for detecting further oxidative modifications<sup>28,29</sup>. With alternative chromatographic conditions, additional products emerged, demonstrating that, like thymidine hydroxylase, TET was capable of iterative oxidation, yielding 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)<sup>28,29</sup>. Highly sensitive mass spectrometry established that these base modifications are detectable either with TET overexpression or, of more physiological relevance, within ES cells, although their levels are at least an order of magnitude less than those of 5hmC<sup>28,29</sup>. Notably, 5hmC, 5fC and 5caC are chemically distinct modifications of C that could be specifically recognized by different DNA-binding proteins. The oxidized 5-substituents also have different steric and electronic properties, which can promote alternative nucleobase tautomers or, in the case of 5fC and 5caC, destabilize the *N*-glycosidic bond<sup>30,31</sup>.

The TET family members (TET1, 2 and 3) each harbour a core catalytic domain (Fig. 1a), with a double-stranded  $\beta$ -helix fold that contains the crucial metal-binding residues found in the family of Fe(II)/ $\alpha$ -KG-dependent oxygenases<sup>32</sup>. In the putative mechanism based on the precedent of other family members (Fig. 1b), TET uses molecular oxygen as a substrate to catalyse oxidative decarboxylation of  $\alpha$ -KG, thereby



**Figure 1 | TET and TDG function in oxidation and excision of modified C bases.** **a**, Schematic of mouse Tet enzymes, showing the double-stranded  $\beta$ -helix fold core oxygenase domain, a preceding cysteine (Cys)-rich domain and a CXXC domain in Tet1 and Tet3. **b**, Catalytic pathway for generation of 5hmC by Tet enzymes. An active site Fe(II) (left) is bound by conserved His–His–Asp residues in Tet and coordinates water and  $\alpha$ -ketoglutarate ( $\alpha$ -KG). A two-electron oxidation of  $\alpha$ -KG by molecular oxygen yields  $\text{CO}_2$  and enzyme-bound succinate, and results in a high-valent Fe(IV)-oxo intermediate (right). The intermediate reacts with 5mC to yield 5hmC, with a net oxidative transfer of the single oxygen atom to the substrate, resulting in regeneration of the Fe(II) species. **c**, TDG specifically accommodates oxidized C bases. Shown is the active site of TDG, bound to DNA, containing a substrate analogue of 5caC (PDB 3UOB). Critical residues of the enzyme are labelled. The 5caC analogue is highlighted in yellow. Heteroatoms are shown with nitrogen (blue), oxygen (red) and phosphorus (orange). The distance of hydrogen bonds (dashed red lines) are measured in Å. In addition to several interactions with the Watson–Crick face of the base from Asn 191 and His 151, the carboxylate substituent in the 5-position is well-accommodated by the active site with a binding pocket defined by Ala 145, and hydrogen bonds from Asn 157 and the backbone amide of Tyr 152.

generating a reactive high-valent enzyme-bound Fe(IV)-oxo intermediate that converts 5mC to 5hmC. The core catalytic domain constitutes only a fraction of the large TET enzymes, suggesting the possibility that the non-catalytic domains may have regulatory functions (Fig. 1a). In all

TET isoforms, a cysteine-rich domain precedes the core and seems to be required for activity<sup>17,23</sup>. TET1 and TET3 also contain a chromatin-associated CXXC domain that is known to bind CpG sequences, whereas TET2 partners with IDAX, an independent CXXC-containing protein<sup>33–35</sup>.

### Replication, repair and demethylation

Detection of oxidized 5mC bases within ES cells has suggested potential functional relevance for these bases in the dynamic regulation of the genome and led to the next question: how might these oxidized bases be altered to regenerate unmodified C? Three potential pathways for demethylation following 5mC oxidation have been entertained: passive dilution of the oxidized base, direct removal of the oxidized 5-position substituent and DNA repair-mediated excision of modified nucleotides.

As had been previously entertained with 5mC, passive dilution of 5hmC or the highly oxidized bases may contribute to demethylation. Indeed, significant evidence (discussed later) has pointed to the importance of this DNA replication-dependent pathway. Although confusion exists in the literature as to whether this pathway should be designated as active or passive (see Perspectives), we find it most useful to consider this as an active demethylation pathway that results from active modification of 5mC followed by passive dilution of the oxidized base to regenerate unmodified C in a replication-dependent manner.

What about pathways that might promote active restoration of unmodified C? Whereas direct removal of a methyl group has a high energetic barrier, the removal of the oxidized methyl group is more feasible. For example, similar to the precedent of isoorotate decarboxylase, decarboxylation of 5caC could revert the base to unmodified C. One study with isotopic labelling of 5caC has suggested this possibility<sup>36</sup>; however, a 5caC decarboxylase has yet to be identified. Interestingly, in the absence of the methyl-donor S-adenosylmethionine (SAM), DNMTs can potentially promote the addition or removal of oxidized 5-position substituents,

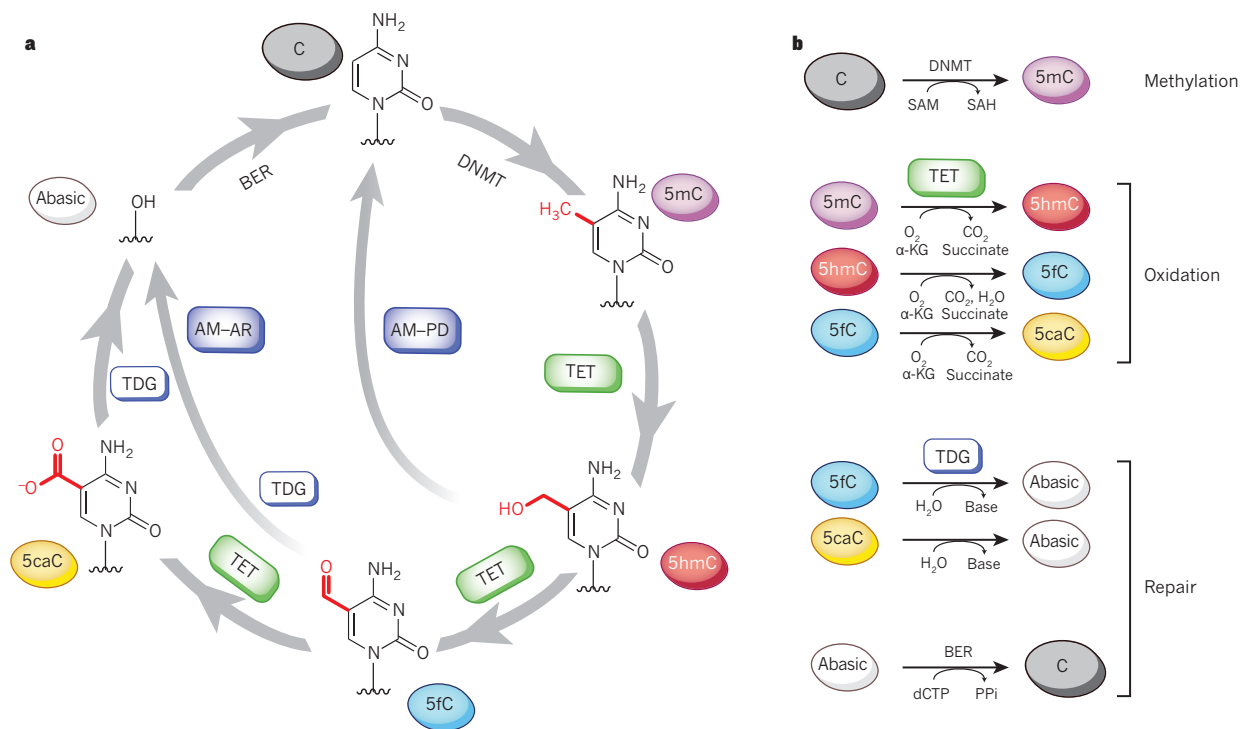
including reacting with 5hmC *in vitro*<sup>37</sup>. Thus, DNMTs could theoretically function in demethylation, raising interesting regulatory implications. The biological relevance of this 'reverse' DNMT reaction remains unknown because SAM is a general methyl-group donor that is present in all cell types.

On firmer ground, an alternative pathway for active restoration of C could involve DNA repair enzymes. Although pathways involving nucleotide excision repair have been considered in demethylation<sup>38,39</sup>, the bulk of the focus has been on base excision repair (BER), a pathway involving the removal of an entire modified base and its subsequent repair to replace the residue with unmodified C (see ref. 40 for review of BER). Several key components of the BER pathway are present at crucial transitions of DNA methylation patterns<sup>41–43</sup>, and this line of inquiry, as detailed in the next section, has proven fruitful.

### TDG-mediated repair completes the cycle

The suggested involvement of BER in demethylation prompted an active search for glycosylase enzymes that might excise modified C bases. Plants use such a mechanism to excise 5mC directly, and some early reports suggested that either methyl-binding domain protein 4 (MBD4) or thymine DNA glycosylase (TDG) may have similar activity in mammals<sup>44</sup>. Although these possibilities have since been discounted given these enzymes' marginal 5mC glycosylase activity, there is mounting biochemical evidence for a role of TDG in DNA demethylation<sup>45</sup>. In particular, TDG has been shown to interact with numerous transcription factors, chromatin modifying enzymes and DNMTs, raising the possibility of a functional role for TDG in modulating gene transcription, either through its glycosylase activity or as a transcriptional coactivator<sup>45</sup>.

After the discovery of TET, the next significant milestone in DNA demethylation came when two groups demonstrated that, unlike other



**Figure 2 | A complete pathway for dynamic modifications of C.** **a**, A biochemically validated pathway for modification of C within DNA is shown. 5mC bases, introduced by DNA methyltransferase (DNMT) enzymes, can be oxidized iteratively to 5hmC, 5fC and 5caC. In the pathway of active modification (AM) followed by passive dilution (PD), 5hmC is diluted in a replication-dependent manner to regenerate unmodified C. For clarity, PD of highly oxidized 5fC and 5caC is not depicted. In the pathway of AM followed by active restoration (AR), 5fC

or 5caC is excised by TDG generating an abasic site as part of the base excision repair (BER) process that regenerates unmodified C. **b**, The individual reactions in the pathway are shown with all reactants depicted. The BER pathway involves excision of the abasic site, replacement of the nucleotide using unmodified deoxycytidine triphosphate (dCTP) by a DNA polymerase (generating pyrophosphate, PPi) and ligation to repair the nick.  $\alpha$ -KG,  $\alpha$ -ketoglutarate; SAM, S-adenosylmethionine; SAH, S-adenosylhomocysteine.



DNA glycosylases, TDG is required for embryonic development<sup>46,47</sup>. Molecular studies on TDG-null embryos, or a catalytically inactive mutant, have pointed to an epigenetic abnormality. Among other alterations, the mutants showed marked decreases in the expression of developmental transcription factors, such as *hox* gene family members, with perturbed methylation at their regulatory sequences<sup>46,47</sup>. Although these genetic studies raised the possibility of TDG as an important player in DNA demethylation, the nature of its genomic target remained unclear.

TDG has long been the focus of compelling biochemical and structural studies because of its interesting role as a DNA repair enzyme that can remove a normal base, T, from a genomic T·G mismatch<sup>48</sup>. Initial speculation therefore focused on the possibility that TDG activity would be coupled to deamination of 5mC or 5hmC because T·G or hmU·G mismatches were known substrates of TDG for which repair could regenerate unmodified C<sup>47,49</sup>. In this model, AID/APOBEC enzymes, the adaptive or innate immune system enzymes that normally target unmodified C, were considered to be the likely candidates for catalysing deamination. Indeed, some studies have suggested a role for deaminases in the reprogramming of stem cells or in embryogenesis<sup>50–53</sup>. These deamination-mediated pathways for demethylation could also involve the DNA damage response protein GADD45 or even MBD4 as an alternative glycosylase for excision of T·G mismatches<sup>38,53,54</sup>, although evidence to the contrary exists<sup>55,56</sup>. Notably, however, deamination of 5hmC by AID/APOBEC enzymes is not detectable *in vitro* or in cells<sup>57,58</sup>, challenging the plausibility of proposed pathways that have invoked 5hmC deamination in DNA demethylation<sup>49</sup>. By contrast, deamination of 5mC by AID/APOBEC family members does occur at a detectable rate *in vitro*, about 10-fold slower than with unmodified C<sup>57</sup>. Although feasible, we anticipate that the role of deaminases in demethylation is probably limited given that 5mC is much less abundant than the unmodified C in mammalian genomes, as well as the enzyme's selectivity for single-stranded DNA and its preference for particular sequence contexts. This view is supported by the observation that there are no significant developmental defects associated with AID/APOBEC deficiency<sup>16</sup>.

A role for TDG in processing T·G mismatches potentially generated by AID/APOBEC deamination of 5mC remained one possible explanation for its requirement in embryonic development. However, the scope of TDG's role in demethylation was reconsidered on revisiting a previous observation that some correctly base paired, modified C bases can also be targeted by TDG<sup>59</sup>. Specifically, C bases with 5-position substituents that destabilize the N-glycosidic bond by electronic effects, such as 5-fluorocytosine, have been shown to be efficiently excised by the glycosylase. These observations opened up the possibility that TDG could directly excise TET oxidation products. Indeed, although no significant *in vitro* base excision activity has been observed with C, 5mC and 5hmC, TDG has robust *in vitro* base excision activity on 5fC and 5caC properly base paired to G in duplex DNA<sup>29,60</sup> (Fig. 1c). This *in vitro* activity is relevant in cells, as knockdown of the gene encoding TDG leads to elevated 5caC levels in ES cells<sup>29</sup>. Furthermore, simultaneous TET and TDG overexpression in the HEK293 cell line leads to a depletion of TET-associated 5caC<sup>29,57</sup> or 5fC<sup>57</sup>. Thus, in a striking example of synergy, studies demonstrating a requirement for TDG in development could be reconciled with insights into TET-mediated oxidation. TDG, acting on TET-generated 5fC and 5caC, mediates the first biologically and biochemically validated, complete pathway for active DNA demethylation (Fig. 2).

Biochemical and biophysical studies have started to shed light on the molecular basis for excision of 5fC and 5caC by TDG. In line with earlier studies on TDG's requirements for excision<sup>59</sup>, computational studies have suggested that 5fC and 5caC have destabilized N-glycosidic bonds relative to C, 5mC and 5hmC<sup>30,59</sup>. TDG also seems to have structural features that mediate recognition of these oxidized C bases, including a binding pocket that can specifically accommodate the 5-carboxyl substituent<sup>61</sup> (Fig. 1c). Interestingly, the determinants for

5fC excision seem to be separable from 5caC recognition<sup>31</sup>, an insight that will probably prove useful in assessing the relative importance of 5fC compared with 5caC to demethylation.

## Revisiting biological roles for demethylation

As our biochemical knowledge of TET, TDG and other DNA modifying pathways has evolved, the many biological processes in which DNA demethylation seems to be relevant have been re-examined with a fresh perspective. 5hmC resides at a potentially crucial branch point in the DNA demethylation pathway (Box 1). Here we summarize recent advances, focusing on studies that help to establish the role of 5hmC in various biological and pathological settings in which dynamic DNA methylation takes place both globally and locally.

## Pre-implantation global methylation dynamics

The specific and rapid loss of 5mC from the paternal genome of zygotes has been re-examined in light of the discovery of 5hmC (Fig. 3a). Immunostaining using a 5hmC-specific antibody revealed that loss of 5mC coincides with the appearance of 5hmC, suggesting that TET is involved in the rapid disappearance of 5mC<sup>62–65</sup>. Interestingly, 5fC and 5caC can also be detected in the paternal chromosome, although the significance of this observation is still unknown<sup>66</sup> (Fig. 3a). Knockdown or targeted deletion of the gene encoding Tet3 — the only highly expressed TET protein in the zygote — abolished the loss of 5mC and the generation of 5hmC, indicating that Tet3 is responsible for the oxidation of 5mC in this context<sup>62,65</sup>. Immunostaining and sequencing studies have shown that, after the two pronuclei fuse, both the maternal genome containing 5mC and the paternal genome with Tet3-generated 5hmC are diluted in a replication-dependent manner<sup>63,66,67</sup>. Thus, Tet3 seems to mediate active demethylation of the paternal genome through active oxidation of 5mC followed by passive dilution, resulting in restoration of unmodified C. The reason why the male genome undergoes an additional oxidation step is currently unknown. However, the process is likely to be biologically important, because female mice depleted of Tet3 in the germ line show reduced fecundity and their heterozygous mutant offspring suffer an increased incidence of developmental failure<sup>62</sup>.

What is the mechanism underlying this asymmetric DNA demethylation? Although factors in the paternal genome that attract Tet3 cannot be ruled out, available data suggest that Tet3 may be actively excluded from the maternal genome. A recent study has shown that the dimethylation of histone H3 lysine 9 (H3K9me2) present predominantly on maternal chromatin provides a binding site for the recruitment of PGC7 (also known as Dppa3 and Stella), which in turn excludes Tet3 from binding to the maternal pronucleus<sup>68</sup>. Interestingly, some imprinted loci on the paternal genome that do not undergo demethylation are also not targeted by TET3 (ref. 62). These imprinted sites show similar hallmarks of H3K9me2 and PGC7, suggesting a potential common mechanism for Tet3 exclusion<sup>68</sup>.

## TET proteins in PGC reprogramming

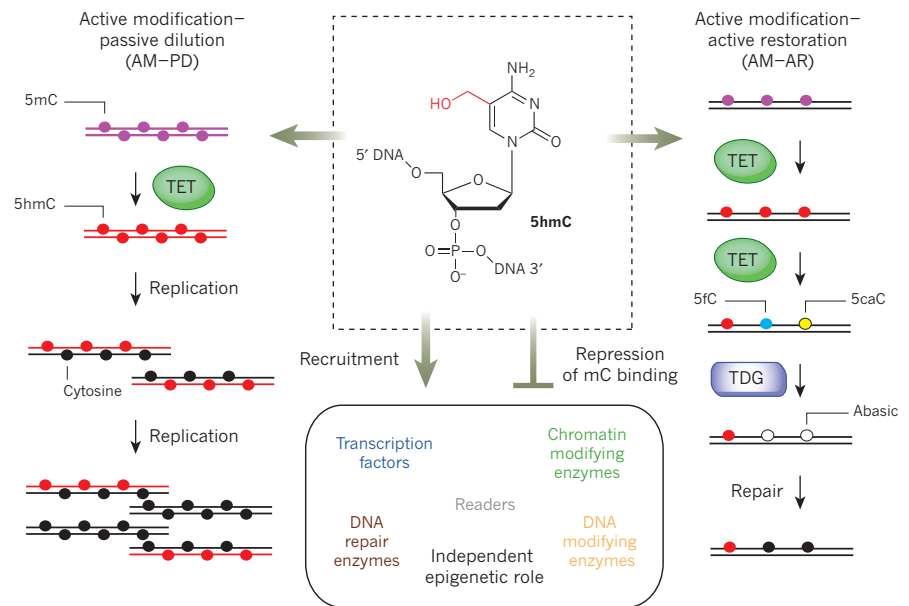
After methylation patterns are established in the embryo, the specialized group of PGCs undergo a further, complex epigenetic reprogramming process that includes erasure of genome-wide DNA methylation patterns<sup>11</sup> (Fig. 3b). Although this process is largely believed to be an active process, careful studies using complementary immunostaining and sequencing techniques have revealed that both passive and active processes contribute to the global loss of 5mC<sup>69–72</sup>. After initial passive dilution of 5mC, 5hmC subsequently accumulates actively and is then lost in an apparent replication-dependent manner<sup>69,71</sup> (Fig. 3b). Just as in early embryonic development, specific loci can deviate from these global patterns, and these differentially methylated loci can persist even into mature oocytes<sup>69–71</sup>.

Although both Tet1 and Tet2 are expressed during PGC reprogramming, only Tet1 is upregulated in reprogramming germ cells<sup>69,73</sup>. However, targeted deletion in mice<sup>73,74</sup> or knockdown in ES cells followed by *in vitro* PGC differentiation<sup>75</sup> revealed that Tet1 does not affect global

## BOX 1

## Roles of 5hmC in DNA demethylation

With the discovery of TET, 5hmC has taken on a new central role in epigenetics. The base sits at an important branch point, with at least three potential outcomes. 5hmC could have an independent epigenetic role related to the base's interaction with chromatin-associated proteins, either through direct recruitment by 5hmC or through disruption of 5mC-specific interactions. 5hmC also can feed into two different pathways that we define as active demethylation (see Perspectives). The pathway involving active modification and passive dilution (AM-PD) (Box Fig., left) is marked by 5mC conversion to 5hmC (or higher oxidized species), replication-dependent depletion of the modified base and reversion to unmodified C. In the active restoration pathway (AM-AR) (Box Fig., right), iterative oxidation by TET can yield 5fC or 5caC, which can be excised by TDG to generate abasic sites as part of the DNA repair pathway that ultimately regenerates unmodified C.



DNA demethylation. Nevertheless, loss of function of Tet1 does impact locus-specific DNA demethylation, particularly at meiotic genes<sup>73</sup>. In addition, although *Tet2* knockout alone does not lead to any PGC phenotype<sup>76–78</sup>, demethylation of some imprinted loci is affected in *Tet1* and *Tet2* double knockout mice<sup>79</sup>. Thus, although there is some consensus for a function for TETs in PGCs, further studies are needed to clarify the exact contribution of Tet1 and Tet2 and their possible redundancy in shaping the PGC methylome.

### TET proteins in stem cells

ES cells are a model for understanding demethylation dynamics, because maintenance of ES cells is associated with a distinct methylation pattern that supports expression of pluripotency factors while silencing lineage-specification factors. Both Tet1 and Tet2 are expressed in mouse ES cells<sup>18,80</sup>. TET proteins are probably part of the pluripotency regulatory circuit and may act by directly regulating expression of key ES cell transcription factors<sup>80,81</sup>. In a series of studies consistent with this hypothesis, short-hairpin-mediated knockdown of *Tet1* alone<sup>18,82</sup> or in combination with *Tet2* (ref. 80) resulted in a defect in ES cell maintenance, as well as skewed differentiation toward trophectoderm and primitive endoderm. However, some ambiguity remains about the role of TET in ES cell maintenance, given that other knockdown studies do not result in similar phenotypes<sup>83</sup> and mice deficient in Tet1 can be derived from *Tet1*-knockout ES cells<sup>74</sup>. The use of different cell lines and culture conditions may contribute to these different results, although potential off-target activity of short hairpin RNAs cannot be ruled out.

Genome-wide and single-base resolution methods have been adapted to discriminate between modified C bases in ES cells. A consensus of these studies (see ref. 84 for a review) has demonstrated a probable regulatory role for 5hmC with particular enrichment at transcribed gene bodies, bivalent and silent promoters, and distal *cis*-regulatory elements. More strikingly, recent genome-wide mapping in ES cells has pointed to the functional relevance of TDG, 5fC and 5caC. Using an immunoprecipitation approach in *Tdg*-deficient ES cells, a significant enrichment of 5fC and 5caC was observed in non-repetitive

regions, particularly at distal regulatory elements<sup>85</sup>. A second study that used a chemical-labelling pull-down approach for detection of 5fC demonstrated 5fC enrichment in enhancer regions<sup>86</sup>. These studies strongly suggest that dynamic C modification involving TDG-mediated 5fC and 5caC removal takes place widely in mouse ES cells.

Although the exact function of the TET proteins in ES cells needs further study, several recent publications are supportive of a role for TET in reprogramming of somatic cells to generate induced pluripotent stem cells (iPSCs) (see Review by Apostolou and Hochedlinger<sup>87</sup> in this Insight). For example, at the early stage of transduction with the transcription factors Oct4, Klf4, Sox2 and c-Myc (collectively referred to as OKSM), Tet2 is recruited to the *Nanog* and *Esrrb* loci to activate their transcription<sup>88</sup>. In addition, both Tet1 and Tet2 can associate with Nanog and facilitate iPSC generation in an enzymatic activity-dependent manner<sup>89</sup>. Remarkably, Tet1 overexpression can not only enhance reprogramming efficiency by promoting demethylation and reactivation of Oct4, but can also replace Oct4 in the iPSC reprogramming cocktail<sup>90</sup>. Furthermore, beyond reprogramming mediated by OKSM, Tet1 and Tet2 seem to have distinct roles in reprogramming mediated by fusion of somatic cells to pluripotent cells<sup>91</sup>.

### Locus-specific active demethylation in somatic cells

The key players in the C modifying pathway have also been implicated in locus-specific demethylation, independent of replication. For example, TDG has been observed at loci at which rapid cycling of C and 5mC is associated with hormonal<sup>12,13</sup> or cytokine-mediated<sup>14</sup> regulation, and TET has been associated with demethylation in the post-mitotic adult brain<sup>49</sup>. These studies imply that active demethylation with TET and TDG may be operational when transcriptional control must be modulated in the absence of DNA replication. However, we still have much more to learn at the level of individual promoters.

### DNA demethylation in cancer

Aberrant DNA methylation is a prominent feature of cancer cells<sup>92</sup>, raising the possibility that demethylation pathways may contribute to cancer development<sup>93</sup>. *TET1* was initially identified owing

to its fusion to *MLL* (also known as *KMT2A*) in patients with acute myeloid leukaemia<sup>24</sup>, and inactivating *TET2* mutations have since been demonstrated to be frequent lesions in myeloid lineage malignancies<sup>94,95</sup>. Interestingly, these same myeloid-lineage conditions are susceptible to therapy aimed at inhibiting DNA methylation<sup>96</sup>. Further supporting a role for Tet2 in normal haematopoiesis, mouse models have shown that the enzyme is a crucial regulator of self-renewal and differentiation in haematopoietic stem cells<sup>76–78,97</sup>. Although most studies have focused on haematological malignancies, downregulation of TET expression has been observed in human breast, liver, lung, pancreatic and prostate cancers<sup>98</sup>. Despite discrepancies in the levels of 5mC in these various settings, TET mutations are consistently associated with a decrease in 5hmC, which has been suggested as a potential diagnostic biomarker<sup>98,99</sup>. Regarding the other players in demethylation, although the relevance of 5fC and 5caC in cancer has not yet been explored, TDG has also been implicated in various cancers<sup>45</sup>. It remains to be established if this association is due to TDG's role in mismatch repair or in active DNA demethylation.

Interestingly, in acute myeloid leukaemia, TET2 mutations were found to be mutually exclusive with a neomorphic mutation in the isocitrate dehydrogenase genes *IDH1* and *IDH2* (ref. 100). Wild-type *IDH1* and *IDH2* catalyse the conversion of isocitrate to  $\alpha$ -KG, the cofactor for the TET and histone demethylase family of oxygenase enzymes<sup>32</sup>. The neomorphic mutation in *IDH1* and *IDH2* leads to the production of  $\alpha$ -hydroxyglutarate, an oncometabolite that can competitively inhibit these  $\alpha$ -KG-dependent enzymes<sup>101</sup>. These studies suggest that neomorphic *IDH1* and *IDH2* mutations may alter DNA methylation patterns by recapitulating TET2 mutations, although alternative mechanisms have also been postulated<sup>102</sup>.

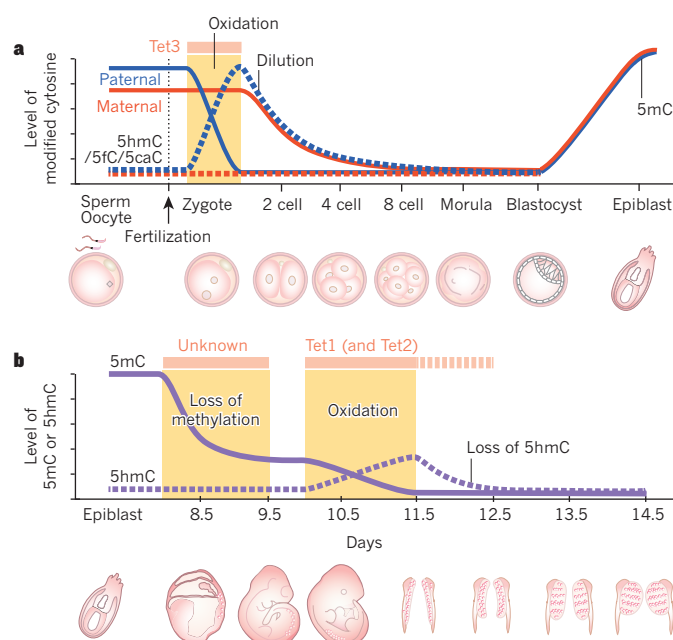
### Perspective and open questions

Across various physiological developmental niches, non-physiological settings such as iPSCs, and even pathological settings such as cancer, loss of TET proteins and 5hmC is associated with dysregulated DNA methylation. These biological studies, on the heels of a series of transformative biochemical discoveries on TET and TDG, have established 5hmC as a key intermediate in active DNA demethylation.

### Revisiting the definition of active demethylation

Recent advances require the classical definitions of passive and active DNA demethylation to be revisited. As we have noted, passive demethylation seems to be best suited for describing the replication-dependent dilution of 5mC only, as this pathway does not involve any active enzymatic processes that alter the base itself. Given our current understanding, active demethylation involving TET is best viewed as two pathways, both of which initially involve active modification (AM) of 5mC to generate 5hmC. This base can be further processed through either passive dilution (PD) to regenerate unmodified C through DNA replication, or active restoration (AR) through further enzymatic modification (Box 1). This framework should also be fitting for other potential pathways for demethylation, such as a 5mC deamination–BER pathway, which would be described as an AM–AR active demethylation pathway.

AM–AR has the advantage of achieving rapid conversion of 5mC to unmodified C, yet the pathway also poses the potential risk of genomic damage given the involvement of DNA breaks in BER. By contrast, the dependence of AM–PD on replication means that functions that might be associated with 5mC modification are quickly achieved, whereas reversion to unmodified C awaits DNA replication. AM–AR therefore seems particularly well suited to locus-specific demethylation processes that require a rapid response to environmental stimuli, whereas AM–PD might be better suited to developmental processes in which cellular replication is tied to lineage specification, such as preimplantation development and PGC reprogramming. With this framework, future studies can evaluate



**Figure 3 | DNA methylation dynamics in pre-implantation embryos and primordial germ cells.**

**a**, Dynamics of 5mC and its oxidation products in pre-implantation embryos. Although the maternal DNA goes through passive demethylation, the paternal genome is demethylated in two steps. Tet3 first oxidizes the 5mC in the paternal genome, and the oxidation products are then diluted through a replication-dependent process. For clarity, although the absolute levels of 5hmC, 5fC and 5caC differ, the bases are schematically shown together (dotted line) given that their increase and subsequent depletion follow similar patterns. DNA methylation patterns are re-established by *de novo* DNMTs at the blastocyst stage. **b**, Illustration of the 5mC and 5hmC dynamics in primordial germ cells (PGCs) during their reprogramming. DNA demethylation in PGCs goes through three stages: loss of bulk DNA methylation in a Tet-independent manner; oxidation of remaining 5mC to 5hmC by Tet1 and potentially Tet2 proteins; and loss of 5hmC through replication-dependent passive dilution. 5fC and 5caC are not shown in this panel because no dynamic change in their levels was observed by immunostaining<sup>69–72</sup>. Figure scale is shown in embryonic days post-fertilization.

the theoretical risks and benefits of the AM–PD and AM–AR pathways and better delineate the cellular context in which either or both pathways are active.

### Regulation of the demethylation pathway

Viewing C modification as a series of step-wise modifications (Fig. 2) prompts the key question: what regulates stalling at various intermediates in the pathway or progression through the cycle?

5hmC is significantly more prevalent than 5fC and 5caC<sup>84</sup>. Given that TET enzymes can iteratively oxidize, it remains unclear what factors dictate that the modification pathway halts at 5hmC. Stalling of the pathway at 5hmC could be regulated through modulating TET's accessibility to 5hmC, through either post-translational modifications or interaction with protein partners. Alternatively, stalling at 5hmC could be regulated at a biochemical level through altered enzyme kinetics. In this regard, a crucial question that remains unresolved refers to TET's relative ability to oxidize 5mC, 5hmC and 5fC. The basal reactivity of TET with each of these substrates and regulation of its substrate preferences will need to be addressed. Structural insights into the TET catalytic domain could prove key to deciphering regulatory mechanisms that govern iterative oxidation.

At the next stage of the pathway, do 5fC and 5caC have distinctive roles and what, if any, significance do they have beyond serving as intermediates in demethylation? 5fC and 5caC are similar marks in that they both result from iterative oxidation and both can be excised by TDG, yet they



could have different roles. Indeed, TDG shows a higher affinity for 5fC<sup>60</sup> and different mechanisms may be used in excision of 5fC and 5caC<sup>31</sup>. Given the relative scarcity of 5fC and 5caC, it is not clear if these bases are simply intermediates in an active demethylation pathway, or if they have functionally significant interactions with genomic 'readers'<sup>103,104</sup>. Further efforts to perturb TDG function and to localize 5fC and 5caC in the genome in high resolution and at specific genomic loci should help to resolve some of these issues.

When the pathway is viewed as a complete cycle of C methylation, oxidation and repair, it immediately begs the question: does recurrent cycling of this pathway occur? Rapid cycling between C and 5mC has been observed at some promoters<sup>12,13</sup> and disruption of TDG leads to accumulation of intermediates in the pathway<sup>85,86</sup>. However, bona fide evidence of multiple cycling events through an AM-AR pathway at a single locus has yet to be shown.

Finally, although we have mainly emphasized the emerging role of TET and TDG, numerous other DNA modifying enzymes that can recognize modified C bases will undoubtedly influence the physiological function of the dynamic demethylation pathway. An active search for potential 5caC decarboxylase enzymes, further studies to elucidate the relevance of AID/APOBEC deamination of 5mC, the role of DNMTs in a 'reverse' reaction, and interactions of modified C bases with DNA-binding proteins are just a few areas in which advances may come in the next phase of discovery.

Oxidative modifications of 5mC and related repair mechanisms have expanded the possibilities by which the genome can retain great flexibility while maintaining the integrity of its coding information. A few years ago, it would have been hard to imagine how much our knowledge of active DNA demethylation would change, and we anticipate that the years ahead will be marked by many more exciting discoveries regarding the role of dynamic regulation of DNA methylation in development, gene regulation and genome stability. ■

Received 2 May; accepted 6 August 2013.

- Goll, M. G. & Bestor, T. H. Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481–514 (2005).
- Klose, R. J. & Bird, A. P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89–97 (2006).
- Bestor, T. H. & Bourc'his, D. Transposon silencing and imprint establishment in mammalian germ cells. *Cold Spring Harb. Symp. Quant. Biol.* **69**, 381–387 (2004).
- Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (suppl.), 245–254 (2003).
- Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nature Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
- Ooi, S. K. & Bestor, T. H. The colorful history of active DNA demethylation. *Cell* **133**, 1145–1148 (2008).
- Surani, M. A. & Hajkova, P. Epigenetic reprogramming of mouse germ cells toward totipotency. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 211–218 (2010).
- Jenkins, T. G. & Carrell, D. T. Dynamic alterations in the paternal epigenetic landscape following fertilization. *Front. Genet.* **3**, 143 (2012).
- Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
- Oswald, J. et al. Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478 (2000).
- Hackett, J. A., Zyllicz, J. J. & Surani, M. A. Parallel mechanisms of epigenetic reprogramming in the germline. *Trends Genet.* **28**, 164–174 (2012).
- Kangaspeska, S. et al. Transient cyclical methylation of promoter DNA. *Nature* **452**, 112–115 (2008).
- Métivier, R. et al. Cyclical DNA methylation of a transcriptionally active promoter. *Nature* **452**, 45–50 (2008).
- Thilainadesan, G. et al. TGF- $\beta$ -dependent active demethylation and expression of the p15ink4b tumor suppressor are impaired by the ZNF217/CoREST complex. *Mol. Cell* **46**, 636–649 (2012).
- Bhutani, N., Burns, D. M. & Blau, H. M. DNA demethylation dynamics. *Cell* **146**, 866–872 (2011).
- Nabel, C. S., Manning, S. A. & Kohli, R. M. The curious chemical biology of cytosine: deamination, methylation, and oxidation as modulators of genomic potential. *ACS Chem. Biol.* **7**, 20–30 (2012).
- Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009). **This manuscript reports on the discovery of TET enzymes and their biochemical ability to convert 5mC to 5hmC.**
- Ito, S. et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**, 1129–1133 (2010). **This report describes the discovery of a role for TET enzymes in 5mC oxidation within ES cells.**
- Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
- Borst, P. & Sabatini, R. Base J: discovery, biosynthesis, and possible functions. *Annu. Rev. Microbiol.* **62**, 235–251 (2008).
- Liu, C. K., Hsu, C. A. & Abbott, M. T. Catalysis of three sequential dioxygenase reactions by thymine 7-hydroxylase. *Arch. Biochem. Biophys.* **159**, 180–187 (1973).
- Smiley, J. A., Kundracik, M., Landfried, D. A., Barnes, V. R. S. & Axheim, A. A. Genes of the thymidine salvage pathway: thymine-7-hydroxylase from a *Rhodotorula glutinis* cDNA library and iso-orotate decarboxylase from *Neurospora crassa*. *Biochim. Biophys. Acta* **1723**, 256–264 (2005).
- Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8**, 1698–1710 (2009).
- Ono, R. et al. 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 (2002).
- Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem. J.* **126**, 781–790 (1972).
- Globisch, D. et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
- Song, C. X. et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotechnol.* **29**, 68–72 (2011).
- Ito, S. et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011). **This report describes the discovery that TET can iteratively oxidize 5mC to 5hmC, 5fC and 5caC.**
- He, Y. F. et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011). **This manuscript shows the generation of 5caC by TET and its excision by TDG.**
- Williams, R. T. & Wang, Y. A density functional theory study on the kinetics and thermodynamics of N-glycosidic bond cleavage in 5-substituted 2'-deoxycytidines. *Biochemistry* **51**, 6458–6462 (2012).
- Hashimoto, H., Zhang, X. & Cheng, X. Selective excision of 5-carboxylcytosine by a thymine DNA glycosylase mutant. *J. Mol. Biol.* **425**, 971–976 (2013).
- Loenarz, C. & Schofield, C. J. Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. *Trends Biochem. Sci.* **36**, 7–18 (2011).
- Xu, Y. et al. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. *Cell* **151**, 1200–1213 (2012).
- Zhang, H. et al. TET1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. *Cell Res.* **20**, 1390–1393 (2010).
- Ko, M. et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. *Nature* **497**, 122–126 (2013).
- Schiesser, S. et al. Mechanism and stem-cell activity of 5-carboxylcytosine decarboxylation determined by isotope tracing. *Angew. Chem. Int. Edn Engl.* **51**, 6516–6520 (2012).
- Litkeviciute, Z., Lukinavicius, G., Masevicius, V., Daujotyte, D. & Klimasauskas, S. Cytosine-5-methyltransferases add aldehydes to DNA. *Nature Chem. Biol.* **5**, 400–402 (2009).
- Niehrs, C. & Schafer, A. Active DNA demethylation by Gadd45 and DNA repair. *Trends Cell Biol.* **22**, 220–227 (2012).
- Ma, D. K., Guo, J. U., Ming, G. L. & Song, H. DNA excision repair proteins and Gadd45 as molecular players for active DNA demethylation. *Cell Cycle* **8**, 1526–1531 (2009).
- Fromme, J. C. & Verdine, G. L. Base excision repair. *Adv. Protein Chem.* **69**, 1–41 (2004).
- Hajkova, P. et al. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78–82 (2010).
- Gehring, M., Reik, W. & Henikoff, S. DNA demethylation by DNA repair. *Trends Genet.* **25**, 82–90 (2009).
- Wossidlo, M. et al. Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* **29**, 1877–1888 (2010).
- Zhu, J. K. Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43**, 143–166 (2009).
- Dalton, S. R. & Bellacosa, A. DNA demethylation by TDG. *Epigenomics* **4**, 459–467 (2012).
- Cortázar, D. et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature* **470**, 419–423 (2011).
- Cortellino, S. et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
- Cortázar, D., Kunz, C., Saito, Y., Steinacher, R. & Schar, P. The enigmatic thymine DNA glycosylase. *DNA Repair (Amst.)* **6**, 489–504 (2007). **References 47 and 48 demonstrate that deletion of TDG is associated with embryonic lethality and epigenetic abnormalities.**
- Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
- Bhutani, N. et al. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042–1047 (2010).
- Popp, C. et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. *Nature* **463**, 1101–1105 (2010).
- Kumar, R. et al. AID stabilizes stem-cell phenotype by removing epigenetic

- memory of pluripotency genes. *Nature* **500**, 89–92 (2013).
53. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. *Cell* **135**, 1201–1212 (2008).
  54. Barreto, G. *et al.* Gadd45a promotes epigenetic gene activation by repair-mediated DNA demethylation. *Nature* **445**, 671–675 (2007).
  55. Jin, S. G., Guo, C. & Pfeifer, G. P. GADD45A does not promote DNA demethylation. *PLoS Genet.* **4**, e1000013 (2008).
  56. Wong, E. *et al.* Mbd4 inactivation increases C-T transition mutations and promotes gastrointestinal tumor formation. *Proc. Natl Acad. Sci. USA* **99**, 14937–14942 (2002).
  57. Nabel, C. S. *et al.* AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature Chem. Biol.* **8**, 751–758 (2012).
  58. Rangam, G., Schmitz, K. M., Cobb, A. J. & Petersen-Mahrt, S. K. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. *PLoS ONE* **7**, e43279 (2012).
  59. Bennett, M. T. *et al.* Specificity of human thymine DNA glycosylase depends on N-glycosidic bond stability. *J. Am. Chem. Soc.* **128**, 12510–12519 (2006).
  60. Maiti, A. & Drohat, A. C. 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 (2011).
  61. Zhang, L. *et al.* Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. *Nature Chem. Biol.* **8**, 328–330 (2012).
  62. Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
- This report demonstrates that TET3 is responsible for paternal genome 5mC oxidation in zygotes.**
63. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**, 194 (2011).
- This report demonstrates that 5hmC generated in zygotes is passively diluted during preimplantation development.**
64. Iqbal, K., Jin, S. G., Pfeifer, G. P. & Szabo, P. E. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. *Proc. Natl Acad. Sci. USA* **108**, 3642–3647 (2011).
  65. Wossidlo, M. *et al.* 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Commun.* **2**, 241 (2011).
  66. Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* **21**, 1670–1676 (2011).
  67. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 (2012).
  68. Nakamura, T. *et al.* PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. *Nature* **486**, 415–419 (2012).
  69. Hackett, J. A. *et al.* Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. *Science* **339**, 448–452 (2013).
  70. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
  71. Yamaguchi, S. *et al.* Dynamics of 5-methylcytosine and 5-hydroxymethylcytosine during germ cell reprogramming. *Cell Res.* **23**, 329–339 (2013).
  72. Kagiwada, S., Kurimoto, K., Hirota, T., Yamaji, M. & Saitou, M. Replication-coupled passive DNA demethylation for the erasure of genome imprints in mice. *EMBO J.* **32**, 340–353 (2013).
  73. Yamaguchi, S. *et al.* Tet1 controls meiosis by regulating meiotic gene expression. *Nature* **492**, 443–447 (2012).
  74. Dawlaty, M. M. *et al.* Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell* **9**, 166–175 (2011).
  75. Vincent, J. J. *et al.* Stage-specific roles for Tet1 and Tet2 in DNA demethylation in primordial germ cells. *Cell Stem Cell* **12**, 470–478 (2013).
  76. Li, Z. *et al.* Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. *Blood* **118**, 4509–4518 (2011).
  77. Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).
  78. Quivoron, C. *et al.* TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell* **20**, 25–38 (2011).
  79. Dawlaty, M. M. *et al.* Combined deficiency of Tet1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. *Dev. Cell* **24**, 310–323 (2013).
  80. Koh, K. P. *et al.* Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**, 200–213 (2011).
  81. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389–393 (2011).
  82. Freudenberger, J. M. *et al.* 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 (2012).
  83. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
  84. Song, C. X., Yi, C. & He, C. Mapping recently identified nucleotide variants in the genome and transcriptome. *Nature Biotechnol.* **30**, 1107–1116 (2012).
- This review details the novel methods that have been developed to quantify and sequence modified C bases in the genome.**
85. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706 (2013).
  86. Song, C. X. *et al.* Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. *Cell* **153**, 678–691 (2013).
  87. Apostolou, E. & Hochedlinger, K. Chromatin dynamics during somatic cell reprogramming. *Nature* **502**, 462–471 (2013).
  88. Doege, C. A. *et al.* Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. *Nature* **488**, 652–655 (2012).
  89. Costa, Y. *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* **495**, 370–374 (2013).
  90. Gao, Y. *et al.* 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 (2013).
  91. Piccolo, F. M. *et al.* Different roles for Tet1 and Tet2 proteins in reprogramming-mediated erasure of imprints induced by EGC fusion. *Mol. Cell* **49**, 1023–1033 (2013).
  92. Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome — biological and translational implications. *Nature Rev. Cancer* **11**, 726–734 (2011).
  93. Cimmino, L., Abdel-Wahab, O., Levine, R. L. & Aifantis, I. TET family proteins and their role in stem cell differentiation and transformation. *Cell Stem Cell* **9**, 193–204 (2011).
  94. Delhommeau, F. *et al.* Mutation in TET2 in myeloid cancers. *N. Engl. J. Med.* **360**, 2289–2301 (2009).
  95. Langemeijer, S. M. *et al.* Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nature Genet.* **41**, 838–842 (2009).
  96. Quintás-Cardama, A., Santos, F. P. & Garcia-Manero, G. Therapy with azanucleosides for myelodysplastic syndromes. *Nature Rev. Clin. Oncol.* **7**, 433–444 (2010).
  97. Ko, M. *et al.* Ten-eleven-translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. *Proc. Natl Acad. Sci. USA* **108**, 14566–14571 (2011).
  98. Yang, H. *et al.* Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* **32**, 663–669 (2013).
  99. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature* **468**, 839–843 (2010).
  100. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell* **18**, 553–567 (2010).
  101. Xu, W. *et al.* Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of  $\alpha$ -ketoglutarate-dependent dioxygenases. *Cancer Cell* **19**, 17–30 (2011).
  102. Losman, J. A. *et al.* (R)-2-hydroxyglutarate is sufficient to promote leukemogenesis and its effects are reversible. *Science* **339**, 1621–1625 (2013).
  103. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
  104. Mellén, M., Ayata, P., Dewell, S., Kriacounis, S. & Heintz, N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. *Cell* **151**, 1417–1430 (2012).

**Acknowledgements** We would like to thank S. Yamaguchi for preparing Figure 3; D. Crawford, A. Inoue, C. Nabel, E. Schutsky and S. Yamaguchi for their helpful comments. We apologize to the people whose work cannot be cited due to space limitations. Our DNA methylation-related work is supported by the NIH (U01DK089565 and GM068804 to Y.Z., and K08-AI089242 to R.M.K.), the Rita Allen Foundation (R.M.K.) and HHMI (Y.Z.). Y.Z. is an investigator of the HHMI.

**Author Information** Reprints and permission information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at [go.nature.com/ezvkfv](http://go.nature.com/ezvkfv). Correspondence should be addressed to R.K. (rkohli@upenn.edu) or Y. Z. (yzhang@genetics.med.harvard.edu).