Uncovering the role of 5-hydroxymethylcytosine in the epigenome

Miguel R. Branco, Gabriella Ficz and Wolf Reik

Abstract | Just over 2 years ago, TET1 was found to catalyse the oxidation of 5-methylcytosine, a well-known epigenetic mark, into 5-hydroxymethylcytosine in mammalian DNA. The exciting prospect of a novel epigenetic modification that may dynamically regulate DNA methylation has led to the rapid accumulation of publications from a wide array of fields, from biochemistry to stem cell biology. Although we have only started to scratch the surface, interesting clues on the role of 5-hydroxymethylcytosine are quickly emerging.

Epigenetic marking of the genome is central to the establishment of tissue-specific gene expression programmes and, therefore, to cell differentiation and development. Until recently, the only known epigenetic mark of DNA itself was 5-methylcytosine (5mC), which is established by DNA methyltransferases (DNMTs) preferentially at CpG dinucleotides. 5mC is generally associated with gene repression and has long been regarded as a stable, highly heritable mark. Nonetheless, DNA methylation patterns undergo genome-wide reprogramming during the establishment of primordial germ cells (PGCs) and after fertilization[1–4]. Importantly, the molecular mechanisms that are involved in this reprogramming are only beginning to be addressed.

In 2009, the enzyme ten-eleven translocation 1 (TET1; one of the three enzymes of the TET family) was found to oxidize 5mC to 5-hydroxymethylcytosine (5hmC)[5], a base that was first detected in mammalian DNA in 1972 (REF. 6). The more recent and experimentally robust detection of 5hmC in the mouse cerebellum[7] and in embryonic stem cells (ESCs)[5] has led to the hypothesis that 5hmC may be an intermediate in the removal of 5mC. Indeed, the moderately low levels of 5hmC in the genome (~10% of 5mC and ~0.4% of all cytosines) are consistent with the hypothesis that it may be a short-lived entity. Alternatively, or complementarily, 5hmC may be an epigenetic modification in its own right, attracting a unique panel of chromatin or transcriptional modifiers and thus adding another layer of complexity to the intricate network of epigenetic regulators.

> “5hmC and TETs have important roles in epigenetic reprogramming and regulation of tissue-specific gene expression”

Here we review the astonishingly numerous findings relating to the TET enzymes and 5hmC that have built up in the short time since 2009. TET enzymes and 5hmC appear to be involved in demethylation of DNA in the totipotent zygote and in pluripotent ESCs. Remarkably, this pathway to demethylation may involve further oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), which have also been detected in the genome, as well as interaction with other demethylation pathways. Genome-wide-profiling methods for these marks are being developed, showing thus far that 5hmC has a distinct distribution in the genome from 5mC and can be associated with promoters and gene expression, as well as with Polycomb-mediated silencing.

TETs as new epigenetic modifiers

The TET proteins were recently identified as new oxoglutarate- and iron-dependent dioxygenases by similarity to other members of this enzyme superfamily[8,9]. Genes from the TET subfamily are present throughout the Metazoa and arose from a common ancestor that underwent triplication (forming *Tet1*, *Tet2* and *Tet3*) in jawed vertebrates[9]. All three mouse TETs have been shown to oxidize 5mC to 5hmC *in vitro* and *in vivo*[10,11], and the presence of 5hmC depends on pre-existing 5mC *in vivo*[12,13], suggesting that this is the only route for the synthesis of genomic 5hmC.

Protein structure. All TET proteins contain predicted domains for the binding of iron and oxoglutarate, but they also contain unique domains, such as a cysteine-rich region, which is thought to be involved in DNA binding[9]. Additionally, TET1 and TET3 contain a CXXC zinc finger domain at their amino-terminus[9], which is a known DNA-binding domain that is common to other 5mC-related proteins, such as DNA (cytosine-5)-methyltransferase 1 (DNMT1; this protein maintains CpG methylation during replication). However, the CXXC domain of TET1 differs from that of several other CXXC-containing proteins, and it has been proposed that this difference prevents DNA binding[14]. By contrast, others have suggested that not only can this domain still bind DNA, but its distinct nature allows it to bind to unmethylated, methylated and hydroxymethylated DNA (unlike the DNMT1 CXXC domain, which binds almost exclusively to unmethylated DNA)[15,16]. Importantly, whereas wild-type TET1 binds to CpG islands (CGIs) in the genome (similarly to other CXXC domain proteins), a version containing a mutation in the CXXC domain does not, showing that the CXXC domain is important for the genome-wide pattern of TET1 binding[15], as discussed below.

TET expression and 5hmC levels. In trying to understand the role of 5hmC, several techniques have been developed to allow its quantification in different tissues (BOX 1). The existence of three mammalian TET

enzymes raises the possibility that each has a distinct panel of genomic targets, such that their tissue-specific expression may lead to specific physiological effects. Indeed, in ESCs, TET1 is the primary TET enzyme, and its repression following differentiation into embryoid bodies correlates with a reduction in 5hmC levels[5]. Moreover, high levels of *Tet1* expression are also detected in PGCs[17], suggesting that TET1 is associated with the pluripotent state. Expression of *Tet3* is by far highest in oocytes and zygotes, where 5hmC levels rise dramatically after fertilization in the male pronucleus[18–20]. Across adult tissues, whereas 5mC levels are fairly constant (4–5% of all cytosines), 5hmC levels vary between <0.1% and ~0.7% and are highest in tissues from the central nervous system[21] (ESCs have ~0.4%[5,22]). Although the amount of 5hmC in the genome is dependent on levels of TET activity, steady-state levels of 5hmC will also depend on activities that further modify 5hmC[23–25] and possibly on rates of DNA replication (if 5hmC is not maintained throughout replication).

Box 1 | **Quantifying and mapping 5-hydroxymethylcytosine**

Several techniques have been adapted or newly developed to quantify and profile 5-hydroxymethylcytosine (5hmC) in the genome and are described below. A key challenge for the future is to develop a technique that provides single-nucleotide resolution mapping of 5hmC. Bisulphite sequencing has classically been used for 5-methylcytosine (5mC) single-nucleotide mapping, as bisulphite treatment converts unmodified cytosines, but not 5mC, to uracil. However, bisulphite sequencing does not distinguish between 5mC and 5hmC[52], nor does it distinguish between unmodified cytosines and 5-carboxylcytosine (5caC)[24]. The next generation of genome-wide sequencers promises to deliver high-throughput, single-nucleotide resolution methods for 5hmC mapping. These methods involve direct sequencing of unamplified DNA by exploring differences in physical properties that allow the different bases to be distinguished, including 5mC and 5hmC[53,54].

Thin layer chromatography

This can be used for global quantification. Cytosines at the ends of restriction sites (for example, *Msp*I restriction sites) are isotopically labelled by transfer of a ^{32}P-labelled phosphate group, and they are then hydrolysed and resolved by thin layer chromatography (TLC), allowing the detection of 5hmC within particular dinucleotide contexts (for example, CpG)[5]. Nearest-neighbour analysis is a similar TLC-based method that also allows global quantification of 5hmC within specific sequence contexts[7]. These techniques are useful owing to their sequence specificity, but accurate quantification is difficult.

Liquid chromatography and mass spectrometry

This can be used for global quantification. Hydrolysed nucleotides are resolved by liquid chromatography (LC) and detected by mass spectrometry (MS), and exact quantification is achieved by the inclusion of isotopically labelled standards[21]. It constitutes a gold standard in global quantification of 5hmC but requires extensive experience in LC–MS methods.

Glucosylation

This can be used for global and local quantification. 5hmC is glucosylated *in vitro* by T4 β-glucosyltransferase (BGT). Quantification of genomic 5hmC levels can be performed by using isotopically labelled uridine 5′-diphosphate (UDP)-glucose as a substrate and comparing against known standards[12]. It is seemingly accurate and perhaps more widely accessible than LC–MS. Glucosylation also alters the activity of certain restriction enzymes (for example, *Msp*I cuts $C^{m}CGG$ and $C^{hm}CGG$ sites, but not $C^{gluc}CGG$), such that local 5hmC levels can be measured by methylation-sensitive restriction followed by quantitative PCR[13,55]. Although local absolute quantification is useful, it is restricted to sites of enzymatic cleavage.

Antibody detection

This can be used for global quantification and mapping. Several antibodies have been raised against 5hmC, allowing global quantification by dot-blot or enzyme-linked immunosorbent assay (ELISA)[35,56]. Mapping can also be done by DNA pull-down followed by genome-wide sequencing (hMeDIP–seq)[13,15,38]. An antibody against the product of bisulphite treatment of 5hmC (cytosine 5-methylenesulphonate) has also been used, which prevents an apparent bias that anti-5hmC antibodies display towards CpG-dense regions[10,39]. Global quantification using these approaches is less accurate than that achieved by LC–MS, but it is easy to do. Mapping is relative in nature, and the use of antibodies carries the risk of unspecific binding.

Chemical labelling

This can be used for global quantification and mapping. BGT also is used to transfer 6-N_3-glucose onto 5hmC, which allows the use of click chemistry specifically to tag 5hmC with, for example, biotin. 5hmC can then be quantified by the use of an avidin conjugated to horseradish peroxidase or mapped by pulling down and sequencing biotin-containing DNA fragments[50]. An alternative chemical strategy for tagging 5hmC with biotin involves oxidation of glucosylated 5hmC followed by reaction with an aldehyde-reactive probe[39]. Global quantification using these chemical-labelling methods is less accurate than LC–MS. Pull-down is efficient and highly specific, but mapping is relative.

Is 5hmC a demethylation intermediate?

Several mechanisms for the removal of 5mC have been proposed[26,27] but, in mammals, the only widely accepted pathway is passive demethylation during replication. Nonetheless, 5mC is largely lost from the paternal genome after fertilization before replication is initiated[2,3], and demethylation has been observed in non-replicating cells (for example, in neurons[23]), suggesting active demethylation pathways. In plants, DNA glycosylases can remove 5mC, leaving an abasic site that is repaired by the base excision repair (BER) machinery. By contrast, known mammalian glycosylases (for example, thymine DNA glycosylase (TDG) and methyl-CpG-binding domain protein 4 (MBD4)) only show weak activity on 5mC *in vitro*[27]. However, these enzymes have strong activity towards T:G mismatches that can be created through deamination of 5mC by cytidine deaminases of the activation-induced cytidine deaminase (AID)/apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC) family[27], suggesting that, in mammals, active DNA demethylation may occur in a two-step process that involves deamination followed by mismatch repair (FIG. 1). Indeed, AID contributes to DNA demethylation in PGCs[28]; however, other activities might also be involved and, interestingly, *Tet1* is highly expressed in PGCs[17]. The discovery of 5mC oxidation by TET enzymes has opened new avenues for exploring potential active demethylation mechanisms (FIG. 1). We first discuss the evidence that 5hmC is part of the demethylation process *in vivo* and then look at the biochemical pathways involved in more detail.

***5hmC and demethylation* in vivo.** Initial support for a role of TETs and 5hmC in the removal of 5mC came from the observation that TET1 overexpression in cultured cells leads to a decrease in 5mC levels and a small increase in the levels of unmodified cytosine[5]. Conversely, TET1 depletion in ESCs leads to an accumulation of 5mC both globally[15] and at specific genomic regions, such as LINE1 retrotransposons and transcription factor binding sites[13]. In the mouse brain, TET1 appears to be important for the

demethylation of the brain-derived neurotrophic factor (*Bdnf*) and fibroblast growth factor 1 (*Fgf1*) gene promoters as part of a physiological response to electroconvulsive stimulation[23]. One seemingly inconsistent example is that of myeloid tumours carrying *TET2* mutations that decrease its activity, where decreased global 5hmC levels are actually associated with CpG hypomethylation, but it is possible that the methylation machinery is also impaired in these tumours[10]. Finally, loss of 5mC from the paternal genome in the fertilized egg correlates with an increase in 5hmC in the male pronucleus at a time when the female pronucleus remains methylated and contains low levels of 5hmC[18–20]. This striking asymmetry is highly suggestive of an involvement of 5hmC in demethylation of the paternal genome, and indeed TET3-depleted zygotes fail to demethylate the male pronucleus[18,20] and, in particular, certain promoter regions (for example, *Oct4* (also known as *Pou5f1*) and *Nanog*)[18]. This is compelling evidence that TET3 is involved in DNA demethylation, potentially through oxidation of 5mC. Nonetheless, a direct link to the catalytic activity of the enzyme is still missing, and it is possible that there is a non-catalytic function of TET3 in demethylation. More recently, it was shown by immunostaining that paternal 5hmC is lost passively through replication during pre-implantation development[29], suggesting that 5mC is converted to 5hmC, which is then removed through replication. Because immunostaining is not quantitative, it remains possible that only a portion of 5mC is converted to 5hmC, and the remaining 5mC is removed by a distinct pathway. Indeed, demethylation of transposable elements is apparent by bisulphite sequencing before replication is initiated[30].

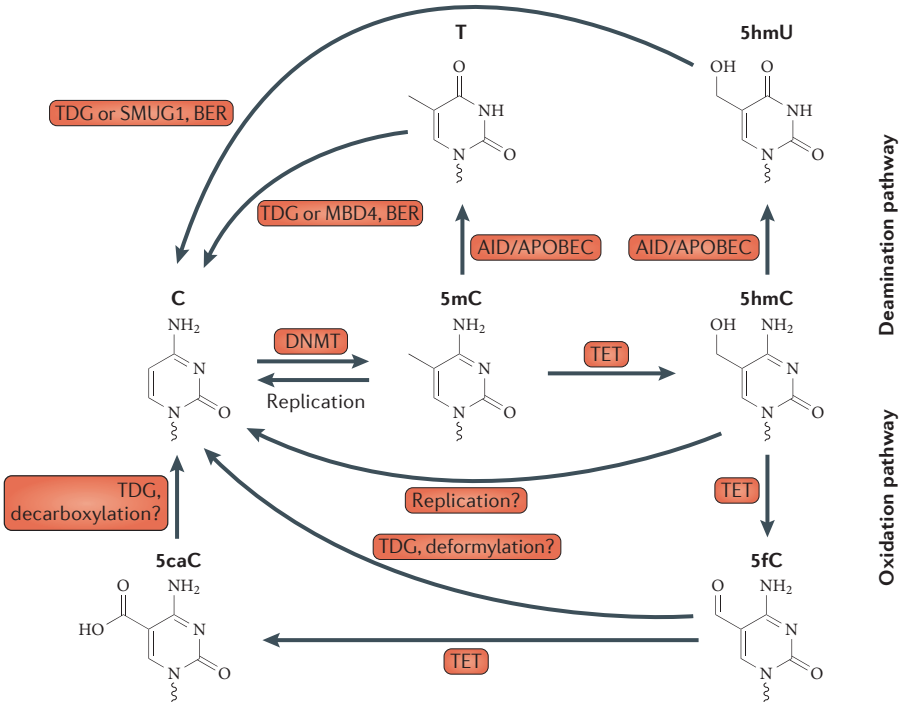

Figure 1 | **Potential pathways for DNA demethylation.** Genomic 5-methylcytosine (5mC) can be removed passively during replication, but several pathways for active demethylation have also been proposed, including those in which 5-hydroxymethylcytosine (5hmC) is an intermediate. Deamination of either 5mC or 5hmC by the activation-induced cytidine deaminase (AID)/apolipoprotein B mRNA editing enzyme, catalytic polypeptide (APOBEC) family of cytidine deaminases produces mismatches that are recognized by DNA glycosylases, producing an abasic site that is then repaired by the base excision repair (BER) machinery[27,33]. Alternatively, 5hmC may be further oxidized to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) by TET enzymes[24,25]. Although it remains possible that deformylation of 5fC and decarboxylation of 5caC convert these intermediates directly back to cytosine, no enzymatic activities have been described to date. Instead, thymine DNA glycosylase (TDG) has been shown to cleave 5fC and 5caC[24,34], again implicating the BER pathway in DNA demethylation. 5hmU, 5-hydroxymethyluracil; DNMT, DNA methyltransferase; MBD4, methyl-CpG-binding domain protein 4; SMUG1, strand-selective monofunctional uracil-DNA glycosylase 1.

Potential mechanisms. Conversion of 5mC to 5hmC could promote passive demethylation at the time of replication by preventing DNMT1 activity at hydroxymethylated CpGs. But, whereas DNMT1 is not active at hemi-hydroxymethylated CpGs[31] *in vitro*, the chaperone UHRF1 (which is also known as NP95 and is crucial for DNMT1 function) binds both 5mC and 5hmC with similar affinities[32], potentially promoting DNMT1 activity at hydroxymethylated CpGs. Current *in vivo* evidence from pre-implantation embryos suggests that 5hmC is passively removed through replication[29], but further studies are required to clarify this point.

In terms of potential active demethylation pathways, 5hmC may be removed through a deamination pathway that is similar to that discussed above, in which AID/APOBEC deaminases act on 5mC (FIG. 1). Indeed, overexpression of AID/APOBEC deaminases in neural cells promotes the removal of 5hmC, and both TET1 and AID overexpression lead to a global accumulation of 5-hydroxymethyluracil (5hmU), the deamination product of 5hmC[23]. Furthermore, both single-strand-selective monofunctional uracil-DNA glycosylase 1 (SMUG1) and TDG glycosylases have strong activity towards 5hmU:G mismatches and appear to be involved in this pathway[23,33]. The potential redundancies between the multiple deaminase and glycosylase activities that operate *in vivo* are likely to complicate future studies of such mechanisms.

Another active pathway that has been proposed involves further oxidation steps that would modify 5hmC first to 5fC and then to 5caC (FIG. 1). Indeed, all three TETs were recently shown to catalyse these additional oxidation steps, and both 5fC and 5caC were detected in ESCs, albeit at very low levels[22,24,25]. Reduced preference of TETs towards 5hmC as a substrate (in comparison to 5mC)[25] and/or fast removal of 5fC and 5caC could explain their low abundance. Interestingly, TDG can remove both 5fC and 5caC[24,34], raising the suggestion that the BER machinery might operate in this pathway as well. However, it is possible that activities that deformylate 5fC or decarboxylate 5caC (generating C) also exist.

It is likely that there are multiple pathways for the removal of 5mC from the genome and that TET enzymes are involved in several of them. One interesting possibility is that different tissues might use different demethylation pathways, depending on whether demethylation is genome-wide or local and on its targets in the genome, for example.

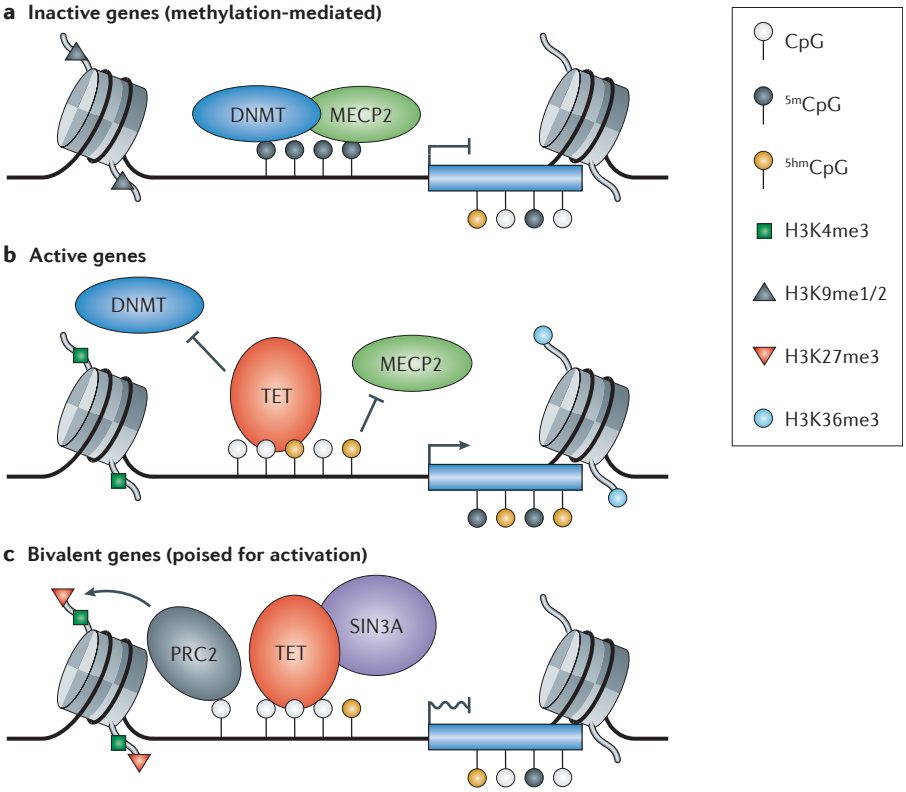

Figure 2 | **Mechanisms of gene expression control by 5-hydroxymethylcytosine and TETs.** **a** | DNA methylation at promoter regions is normally associated with silenced genes, where binding of methyl-CpG binding domain (MBD)-containing proteins, such as MECP2, to 5-methylcytosine (5mC) mediates chromatin remodelling and gene repression. **b** | The oxidation of 5mC to 5-hydroxymethylcytosine (5hmC) by ten-eleven translocation (TET) proteins may therefore relieve the repressive effects of 5mC and, indeed, MECP2 does not bind to 5hmC[43]. Additionally, TET binding may prevent access to DNA methyltransferases (DNMTs), further contributing to the maintenance of an unmethylated promoter. **c** | However, 5hmC and TET enzymes are also associated with repressed genes, namely through an interaction with the SIN3A co-repressor complex[38]. Removal of 5mC also promotes recruitment of Polycomb repressive complex 2 (PRC2)[36] and, therefore, methylation of histone H3 at lysine 27 (H3K27me) at bivalent genes that are silent but transcriptionally poised for activation.

Biological functions of 5hmC and TETs

Roles in pluripotent cells. Oocytes have the capacity to reprogram somatic nuclei during cloning (somatic cell nuclear transfer) and the demethylation activity of the oocyte, in part through TET3[18,20], is important for this process. Similarly, pluripotent cells, such as ESCs or embryonic germ cells (EGCs), have the ability to reprogram somatic cell nuclei when fused with them, including demethylation of specific genes, such as *Oct4* or *Nanog*. It was therefore of interest that, in ESCs, levels of 5hmC and *Tet1* transcripts are relatively high and are substantially reduced following differentiation into embryoid bodies[5] (*Tet2* is also downregulated[35]). A number of recent studies have thus looked at the roles of 5hmC and TET1 in ESCs, adding to the vast amount of existing epigenomic data for this cell type[13,15,36–41]. Immunofluorescent labelling of mouse and human ESC chromosomes showed that, although 5mC is enriched in pericentromeric heterochromatin (albeit not exclusively), 5hmC appears to be mainly euchromatic, suggesting that it is associated with gene activity[13,40]. Higher resolution mapping has been achieved by the development of 5hmC profiling strategies that make use of genome-wide sequencing technologies (BOX 1). In mouse ESCs, both TET1 and 5hmC are found to be enriched within gene bodies (specifically at exons) and also at transcription start sites (TSSs) and promoters[13,15,37–39]. Surprisingly, whereas TET1 binds CGIs[15,36,38] in ESCs, 5hmC is preferentially located at regions of intermediate CpG density and is not enriched at promoter CGIs (but it is at intragenic CGIs)[15]. Nonetheless, TET1-targeted CGIs gain 5mC following TET1 depletion[37], suggesting that TET1 maintains CGI hypomethylation, either by blocking access to the DNA methylation machinery or by rapidly demethylating 5mC (FIG. 2). Additional evidence for a regulatory role of 5hmC in ESCs is suggested by its enrichment at binding sites for pluripotency-associated transcription factors and CCCTC binding factor (CTCF)[13,37]. As binding of these proteins to DNA is impaired by DNA methylation, 5hmC may relieve the repressive effect of 5mC. Many of the findings in human ESCs are consistent with those in mouse ESCs, such as enrichment of 5hmC at TSSs and gene bodies; additionally, enrichment at enhancers has also been described in human ESCs[40,41].

What effect does the distribution of 5hmC have on gene expression in ESCs? In gene bodies, the presence of 5hmC appears to be consistently correlated with gene expression[13,15,37,39] (FIG. 2). However, the evidence of whether there is a correlation between promoter 5hmC and transcription is inconsistent between studies — an issue that will be important to address in future work. Analyses are further confounded by the fact that 5hmC and 5mC often co-exist in the genome[13,38], and promoter 5mC has a repressive effect on transcription. Other than its potential involvement in demethylation, the presence of 5hmC may relieve the silencing effect of 5mC by preventing binding of methyl-binding proteins[43] (FIG. 2). In addition, it is possible that proteins involved in gene regulation exist that specifically bind to 5hmC. In line with a potential activating role of 5hmC, several genes are downregulated upon TET1 knockdown in ESCs, but a surprisingly large number of genes are derepressed[15,36,38]. Interestingly, many of the derepressed genes are Polycomb repressive complex 2 (PRC2) targets[36]. In ESCs, such genes are often bivalently marked with both active marks (trimethylation of histone H3 and lysine 4 (H3K4me3)) and repressive marks (H3K27me3) and are enriched for developmental genes that are transcriptionally 'poised', so that they can be rapidly activated or silenced, depending on the specific differentiation pathway that is taken. Both TET1 and 5hmC are enriched at bivalent genes, and TET1 depletion impairs PRC2 binding to these targets[13,36–39]. It has been suggested that maintenance of hypomethylation by TET1 allows PRC2 binding[36] (FIG. 2) and/or that 5hmC is perhaps involved in promoter pausing of RNA polymerase II[40]. Increased methylation dynamics at PRC2 target genes is in line with the concept of bivalency. A non-catalytic role of TET1 in silencing has also been suggested through an interaction with the SIN3A co-repressor

complex, which is involved in histone deacetylation[38] (FIG. 2). Altogether, these findings indicate that TET1 and 5hmC have multiple regulatory roles in ESCs that depend on factors such as CpG density and the chromatin environment.

Increasing evidence from both ESCs and induced pluripotent stem cells (iPSCs) suggests that TET proteins and 5hmC are involved in regulating pluripotency and differentiation potential. In a reversal of what is seen during ESC differentiation, reprogramming of differentiated cells into iPSCs is associated with the activation of *Tet1* and *Tet2* and accumulation of 5hmC[35]. A link between 5hmC and pluripotency is further highlighted by the existence of a cluster of binding sites for pluripotency-related transcription factors upstream of *Tet1* (REF. 13) and, among those, OCT4 and SOX2 appear to control the levels of both *Tet1* and *Tet2* directly (REF. 35). Importantly, TET1 depletion in ESCs leads to the 5mC-mediated downregulation of several pluripotency-associated genes[13,36,38], and genes that are silenced during embryoid body formation lose 5hmC and gain 5mC at their promoters[13]. Furthermore, several extra-embryonic lineage markers (for example, caudal type homeobox 2 (*Cdx2*), eomesodermin (*Eomes*), GATA binding protein 4 (*Gata4*) and *Gata6*) are derepressed in TET1-depleted cells[13,35,36]. The expression of such markers is detected during the formation of wild-type embryoid bodies, and the specific depletion of TET1 increases the transdifferentiation potential of ESCs to extra-embryonic tissues[35,44]. Accordingly, TET1-depleted cells have a skewed differentiation potential in teratomas towards trophoblast-like cells[35,44] and, unlike wild-type cells, they are able to contribute to the trophectoderm lineage in embryos[11,35].

Interestingly, unlike the findings for TET1, depletion of TET2 does not increase the transdifferentiation potential of ESCs, and simultaneous knockdown of TET1 and TET2 appears to be less efficient in increasing the transdifferentiation potential when compared to TET1 knockdown alone[35]. These results suggest that the balance of expression of the different TET proteins and hydroxymethylation of their specific targets may contribute to the establishment of the inner cell mass and trophectoderm compartments during the morula-to-blastocyst transition. However, both *Tet1*-[44] and *Tet2*-knockout[45–48] mice are viable, although TET1 mice display various degrees of growth restriction[44]. This suggests that, although TET1 activity is not crucial for embryonic development, misregulation of its targets during pre-implantation development may have a long-term effect on development. Notably, there is only a 35% reduction in 5hmC levels in *Tet1*-knockout ESCs and, therefore, generation of triple TET-knockout mice would be necessary to elucidate whether 5hmC itself is required for embryogenesis.

TET2 in haematopoiesis. The role of TET enzymes and 5hmC in differentiation has also been studied in the context of haematopoiesis. *TET2* mutations are frequently involved in human myeloid malignancies and often affect the catalytic activity of TET2. Accordingly, 5hmC levels in bone marrow from patients carrying *TET2* mutations are, in general, lower than their wild-type counterparts, which is independent of zygosity of the mutation[10]. The wide range of malignancies in which *TET2* mutations are found further suggests that 5hmC deficiency is involved in an early stage of tumorigenesis, possibly in the stem cell compartment. Indeed, four studies using mouse *Tet2*-knockout models support this hypothesis[45–48]. Although *Tet2*-null mice are viable and appear to develop normally, as they age, they start dying from haematopoietic malignancies that involve differentiation defects[46–48]. Importantly, *Tet2*-null mice display an increase in haematopoietic stem cell numbers, as well as of other myeloid progenitor cells[45–48]. In fact, $Tet2^{-/-}$ and $Tet2^{+/-}$ stem cells have a higher self-renewal ability in culture than wild-type cells do, suggesting that *Tet2* expression promotes haematopoietic differentiation. This finding is consistent with the higher level of *Tet2* expression that is seen in differentiated haematopoietic lineages when compared with progenitor cells[10,48]. Therefore, 5hmC

Glossary

Blastocyst
A pre-implantation embryo that contains a fluid-filled cavity (the blastocoel), a focal cluster of cells from which the embryo develops (the inner cell mass (ICM)), an epithelial layer of cells on the surface of the ICM (the primitive endoderm) and peripheral trophectoderm cells, which form the extra-embryonic part of the placenta.

CCCTC binding factor
(CTCF). A zinc finger transcriptional repressor that recognizes and binds many sites in the genome and has insulator activity, as well as being important for higher-order chromatin organization.

Click chemistry
A general term for chemical reactions with very high chemical yields and a large thermodynamic driving force. Reactions of this type are wide in scope, are stereospecific, are simple to perform and create products that are stable at physiological conditions and by-products that are easily removable.

CpG islands
(CGIs). DNA sequences that have a considerably higher CG content than expected from the genome average. Most CGIs are depleted of DNA methylation and more than half of the CGIs in the genome are found in promoter regions of genes in mammals.

Embryoid bodies
Spherical structures formed by differentiating embryonic stem cells in culture, which resemble the early embryo.

Embryonic stem cells
(ESCs). Pluripotent cells that are established in culture from the inner cells mass of the blastocyst. These cells can self-renew indefinitely in culture and can differentiate into derivatives of all three embryonic cell lineages *in vitro* and *in vivo*.

Inner cell mass
(ICM). A cluster of undifferentiated cells in the blastocyst, which give rise to the entire fetus and to some of its extra-embryonic (placental) tissues.

Morula
A stage of pre-implantation development in which the embryo is a ball containing 8–32 cells inside the zona pellucida and precedes the formation of the blastocyst. The outer cells have a tendency to become the trophectoderm, and the inner cells have a tendency to become the inner cell mass, from which the epiblast and subsequently the fetal tissues form.

Primordial germ cells
(PGCs). These cells arise in the early embryo and give rise to the mature gametes (oocyte and sperm). They are set aside early in development and migrate to the gonads, where they eventually undergo meiosis and develop into mature gametes.

Pronucleus
Separate pronuclei form around the maternal and paternal genomes, respectively, shortly after fertilization in the mammalian embryo. Following DNA replication, the pronuclei dissolve and the zygote undergoes the first mitotic division.

Teratomas
Tumours that contain differentiated cells from all three germ layers (ectoderm, mesoderm and endoderm) and that arise from germ cells or pluripotent cells.

Trophectoderm
The outer layer of the blastocyst-stage embryo that gives rise to the trophoblast after implantation and will provide the bulk of the extra-embryonic lineages of the placenta.

Trophoblast
The cell lineage that first emerges as the outer trophectoderm layer of the blastocyst and after implantation becomes the predominant cell type in extra-embryonic tissues of the placenta.

appears to promote differentiation of haematopoietic cells, which is in contrast to its role in ESCs and demonstrates that 5hmC has distinct functions in different cell types. It is likely that these different functions of 5hmC depend on the expression of specific TET enzymes, as well as on the cross-talk with other levels of gene regulation.

TETs and 5hmC in the brain. The highest levels of 5hmC of any tissue are found in the adult brain, particularly in the hypothalamus and in the cerebral cortex, but they are also high in other compartments[12,49]. This is of particular interest, as most neuronal cells in the adult brain have ceased to divide mitotically. Hence, any erasure of DNA methylation would appear to require active demethylation, for which there is some preliminary evidence[23]. The possibility has been raised that, in neuronal tissues, one pathway for demethylation may involve oxidation of the methyl group by TET1, followed by deamination to 5hmU by AID or APOBEC1, and the 5hmU might then be excised by the glycosylases SMUG1 or TDG[23,33]. 5hmC profiling has also been performed on the mouse cerebellum and revealed that 5hmC is enriched within gene bodies and upstream of the TSS but is depleted at the TSS[50], and a similar pattern has been described for the human cortex[42]. This profile resembles that of 5mC; 5hmC that is located in gene bodies was found to be associated with higher levels of transcription, as in ESCs. It is possible that 5hmC in gene bodies is a more general epigenetic feature, whereas its presence in promoters may particularly be a feature of pluripotent cell types.

Conclusion and perspective

It is now clear that 5hmC and TETs have important roles in epigenetic reprogramming and regulation of tissue-specific gene expression. Although a role for 5hmC as a bona fide epigenetic mark remains possible, no specific 5hmC-binding proteins have been described to date. By contrast, its potential as a demethylation intermediate is supported by several studies, suggesting that perhaps this constitutes its main cellular role. It is particularly interesting that connections are emerging between oxidation and deamination pathways of 5mC, and the resolution of the reaction products of DNA repair and replication (FIG. 1). On the other hand, TET enzymes appear to have epigenetic functions that go beyond their catalytic role, and it will be interesting to investigate these mechanisms further and determine how they co-evolved with the catalytic role. An attractive aspect of the oxidation of 5mC by TETs is its potential sensitivity to environmental conditions. Namely, its dependence on oxoglutarate and oxygen raises the possibility that TETs may sense metabolic and oxidative cell states and may have a role in the cellular adaptation to altered conditions[51]. Future studies will undoubtedly involve additional TET-knockout models and evaluate an even wider span of the potential roles for 5hmC. These may include roles in PGC reprogramming, iPSC generation and reprogramming by fusion with ESCs or EGCs, as well as in imprinting and retrotransposon regulation, given the importance of DNA methylation in these contexts.

The authors are part of the Epigenetics Programme, The Babraham Institute, Cambridge CB22 3AT, UK.
e-mails: miguel.branco@babraham.ac.uk; gabriella.ficz@babraham.ac.uk; wolf.reik@babraham.ac.uk.

doi:10.1038/nrg3080
Published online 15 November 2011

1. Hajkova, P. *et al.* Epigenetic reprogramming in mouse primordial germ cells. *Mech. Dev.* **117**, 15–23 (2002).
2. Lee, J. *et al.* Erasing genomic imprinting memory in mouse clone embryos produced from day 11.5 primordial germ cells. *Development* **129**, 1807–1817 (2002).
3. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. Demethylation of the zygotic paternal genome. *Nature* **403**, 501–502 (2000).
4. Oswald, J. *et al.* Active demethylation of the paternal genome in the mouse zygote. *Curr. Biol.* **10**, 475–478 (2000).
5. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
6. 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).
7. Kriaucionis, S. & Heintz, N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* **324**, 929–930 (2009).
8. Iyer, L. M., Anantharaman, V., Wolf, M. Y. & Aravind, L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* **38**, 1–31 (2008).
9. 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).
10. Ko, M. *et al.* Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant *TET2*. *Nature* **468**, 839–843 (2010).
11. 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).
12. Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res.* **38**, e181 (2010).
13. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**, 398–402 (2011).
14. Frauer, C. *et al.* Different binding properties and function of CXXC zinc finger domains in Dnmt1 and Tet1. *PLoS ONE* **6**, e16627 (2011).
15. Xu, Y. *et al.* Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol. Cell* **42**, 451–464 (2011).
16. 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).
17. Hajkova, P. *et al.* Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. *Science* **329**, 78–82 (2010).
18. Wossidlo, M. *et al.* 5-hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nature Commun.* **2**, 241 (2011).
19. Iqbal, K., Jin, S.-G., Pfeifer, G. P. & Szabó, 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).
20. Gu, T.-P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
21. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5**, e15367 (2010).
22. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angew. Chem. Int. Edn Engl.* **50**, 7008–7012 (2011).
23. 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).
24. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333**, 1303–1307 (2011).
25. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333**, 1300–1303 (2011).
26. Wu, S. C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nature Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
27. Zhu, J.-K. Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43**, 143–166 (2009).
28. 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).
29. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**, 194 (2011).
30. Wossidlo, M. *et al.* Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. *EMBO J.* **29**, 1877–1888 (2010).
31. Valinluck, V. & Sowers, L. C. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res.* **67**, 946–950 (2007).
32. Frauer, C. *et al.* Recognition of 5-hydroxymethylcytosine by the Uhrf1 SRA domain. *PLoS ONE* **6**, e21306 (2011).
33. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
34. 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).
35. 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).
36. Wu, H. *et al.* Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* **473**, 389–393 (2011).
37. Wu, H. *et al.* Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. *Gene. Dev.* **25**, 679–684 (2011).
38. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473**, 343–348 (2011).
39. Pastor, W. A. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. *Nature* **473**, 394–397 (2011).
40. Szulwach, K. E. *et al.* Integrating 5-hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. *PLoS Genet.* **7**, e1002154 (2011).
41. Stroud, H., Feng, S., Morey Kinney, S., Pradhan, S. & Jacobsen, S. E. 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* **12**, R54 (2011).

42. Jin, S.-G., Wu, X., Li, A. X. & Pfeifer, G. P. Genomic mapping of 5-hydroxymethylcytosine in the human brain. *Nucleic Acids Res.* **39**, 5015–5024 (2011).
43. Valinluck, V. *et al.* Oxidative damage to methyl-CpG sequences inhibits the binding of the methyl-CpG binding domain (MBD) of methyl-CpG binding protein 2 (MeCP2). *Nucleic Acids Res.* **32**, 4100–4108 (2004).
44. 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).
45. 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).
46. 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).
47. 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).
48. Moran-Crusio, K. *et al.* Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. *Cancer Cell* **20**, 11–24 (2011).
49. Münzel, M. *et al.* Quantification of the sixth DNA base hydroxymethylcytosine in the brain. *Angew. Chem. Int. Edn Engl.* **49**, 5375–5377 (2010).
50. Song, C.-X. *et al.* Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nature Biotech.* **29**, 68–72 (2011).
51. Wang, L., Chia, N. C., Lu, X. & Ruden, D. M. Hypothesis: environmental regulation of 5-hydroxymethylcytosine by oxidative stress. *Epigenetics* **6**, 853–856 (2011).
52. Huang, Y. *et al.* The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS ONE* **5**, e8888 (2010).
53. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* **7**, 461–465 (2010).
54. Wanunu, M. *et al.* Discrimination of methylcytosine from hydroxymethylcytosine in DNA molecules. *J. Am. Chem. Soc.* **133**, 486–492 (2011).
55. Kinney, S. M. *et al.* Tissue-specific distribution and dynamic changes of 5-hydroxymethylcytosine in mammalian genomes. *J. Biol. Chem.* **286**, 24685–24693 (2011).
56. Li, W. & Liu, M. Distribution of 5-hydroxymethylcytosine in different human tissues. *J. Nucleic Acids* **2011**, 870726 (2011).

Acknowledgements
We would like to thank all members of the Reik laboratory for stimulating discussions. Our work is supported by the Wellcome Trust, the Biotechnology and Biological Sciences Research Council (BBSRC), the Medical Research Council (MRC), the EU Network of Excellence (NoE) EpiGeneSys and EU SP BLUEPRINT.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Wolf Reik's homepage:
http://www.babraham.ac.uk/devgen/reik.html

ALL LINKS ARE ACTIVE IN THE ONLINE PDF