NIH Public Access
Author Manuscript
Nat Rev Mol Cell Biol. Author manuscript; available in PMC 2014 June 01.

Published in final edited form as:
Nat Rev Mol Cell Biol. 2013 June ; 14(6): 341–356. doi:10.1038/nrm3589.

TETonic shift: biological roles of TET proteins in DNA demethylation and transcription

William A. Pastor[1,3], **L. Aravind**[2], and **Anjana Rao**[1]

[1]Sanford Consortium for Regenerative Medicine and La Jolla Institute for Allergy and Immunology, 9420 Athena Circle, La Jolla, California 92037, USA.

[2]National Center for Biotechnology Information, National Library of Medicine, US National Institutes of Health, Bethesda, Maryland, USA.

Abstract

In many organisms, the methylation of cytosine in DNA has a key role in silencing 'parasitic' DNA elements, regulating transcription and establishing cellular identity. The recent discovery that ten-eleven translocation (TET) proteins are 5-methylcytosine oxidases has provided several chemically plausible pathways for the reversal of DNA methylation, thus triggering a paradigm shift in our understanding of how changes in DNA methylation are coupled to cell differentiation, embryonic development and cancer.

Since the initial description of their enzymatic activity in 2009 (REFS 1,2), proteins of the TET family have become a focus of substantial interest. TET proteins are named after the ten-eleven translocation (t(10;11)(q22;q23)) that is found in rare cases of acute myeloid and lymphocytic leukaemia. This translocation fuses the mixed-lineage leukaemia 1 (*MLL1*) gene located on human chromosome 10 with the *TET1* gene on human chromosome 11 (REFS 3,4). The three mammalian TET proteins, namely TET1, TET2 and TET3, are Fe^{2+}- and 2-oxoglutarate-dependent `dioxygenases` that successively oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) in DNA[1,2,5,6] (FIG. 1a). All three forms of oxidized methylcytosine are now known to be present in numerous mammalian tissues[1,5,7–9]. TET proteins have roles in diverse biological processes, including epigenetic regulation of gene transcription, embryonic development, stem cell function and cancer, but the mechanisms underlying these roles are still poorly defined.

Here, we review our current understanding of TET enzymes and their biological functions, focusing on recent studies not covered in previous reviews[10–14]. We discuss the established and controversial roles of TET proteins and methylcytosine oxidation products in `DNA demethylation`, gene regulation and embryonic development. The role of TET proteins in

© 2013 Macmillan Publishers Limited. All rights reserved

Correspondence to: W.A.P and A.R., wpastor@ucla.edu; arao@liai.org.
[3]Present address: Department of Molecular, Cell and Developmental Biology, University of California at Los Angeles, Los Angeles, California, USA.

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Anjana Rao's homepage: http://www.liai.org/pages/faculty-rao
Protein Data Bank: http://www.rcsb.org/pdb/home/home.do.2FD8

SUPPLEMENTARY INFORMATION
See online article: S1 (figure)

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

haematopoietic differentiation and oncogenesis has been reviewed elsewhere[10,15–17]; their role in neurons is only briefly discussed here owing to space limitations.

TET proteins are 5mC oxidases

The prediction that TET proteins might have DNA-modifying activity was based on the analysis of J-binding protein 1 (JBP1) and JBP2 from *Trypanosoma brucei*, the causative parasite of African sleeping sickness in humans[18]. *Leishmania* species, trypanosomes such as *T.brucei* and other kinetoplastids contain a modified thymine known as 'base J' (β-D-glucosyl-hydroxymethyluracil)[19]. JBP1 and JBP2 generate base J by oxidizing the methyl group of thymine to yield 5-hydroxymethyluracil, and the resulting hydroxyl group is then glucosylated by an unknown glucosyltransferase[20] (FIG. 1b). JBP enzymes are members of a large family of Fe^{2+}- and 2-oxoglutarate-dependent dioxygenases[21,22]; AlkB enzymes, which remove aberrant methylation from damaged DNA bases by an oxidative mechanism, are distantly related members of the same superfamily[21–23]. In *T.brucei* , base J is present at subtelomeric repeats, at inactive copies of the variant surface glycoprotein that is used by the parasite to evade host immune defence and at other silenced regions of the genome[24]. In *Leishmania* species, base J was recently shown to restrain elongation of the unique polycistronic transcripts of kineto plastids beyond transcription stop sites[25].

Computational screens to identify additional homologues of JBP enzymes revealed a large family of predicted nucleic acid-modifying dioxygenases from diverse eukaryotes and bacteriophages, which included the metazoan TET enzymes[2,18]. A gene encoding an enzyme of the TET–JBP family entered the common ancestor of the metazoan lineage and fused with a second gene containing a `CXXC domain` (described below), forming the TET subfamily[2]. TET enzymes are present in all metazoans that have retained cytosine methylation but are absent in organisms such as *Caenorhabditis elegans* in which methylation has been unambiguously lost[2]. The coexistence of TET proteins with DNA methylation, their association with CXXC domains which frequently bind unmethylated `CpG sequences` (see below) and the chemical similarity of thymine and 5mC oxidation all led to the proposal that TET proteins might function as 5mC oxidases and potentially as DNA demethylases[2,18].

Indeed, ectopic expression of TET proteins in cell lines reduces 5mC levels and causes the appearance of 5hmC, and this activity is abrogated by mutation of the signature His-Xaa-Asp motif (where Xaa represents any amino acid) of these dioxygenases[1,26,27] (see below). Recombinant TET catalytic domains and full-length TET proteins efficiently convert 5mC to 5hmC *in vitro* in the presence of the essential cofactors 2-oxoglutarate and Fe^{2+} (REFS 1,27). TET proteins also produce the further oxidation products 5fC and 5caC[5,6,28] (FIG. 1a). Thus, the successive actions of DNA methyltransferases (DNMTs) and TET proteins produce four distinct cytosine modifications, bringing the total number of cytosine species to five (FIG. 1a).

5hmC is found at different levels in mammalian cells: it is present at 1% of the total level of 5mC in some immune cell populations[26], ~5–10% of the level of 5mC in embryonic stem (ES) cells[1] and as high as 40% of 5mC in Purkinje neurons[9]. Consistently, the highest reported levels of 5hmC are in the brain[5,7,8]. An early report of 5hmC in mammalian DNA[29] is questionable, as in this study 5mC was not detected in mouse brain and liver DNA, whereas the level of 5hmC was unrealistically high (15–17% of all cytosines), suggesting massive oxidation of 5mC during the unconventional DNA extraction procedure devised by this group[29]. 5fC and 5caC are present in mammalian cells at much lower levels than 5hmC (0.03% and 0.01%, respectively, of the level of 5mC in mouse ES cells)[5,6,28], at least partly because there are enzymatic mechanisms for their removal. These include base excision by

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

thymine DNA glycosylase (TDG)[6,30] and possibly decarboxylation of 5caC by unknown enzymes present in ES cell lysates[31] (FIG. 1a).

Domain structure of metazoan TETs

Most metazoan TET proteins contain an amino-terminal CXXC domain (~60 amino acids) and a carboxy-termina l catalytic domain (FIG. 2a), the core of which adopts a double-stranded β-helix (DSBH) fold[1,2]. Metazoan TET proteins can be distinguished from all other members of the TET–JBP family by a Cys-rich domain in the N-terminal region of the DSBH domain. In jawed vertebrates, the *TET* gene underwent triplication, giving rise to *TET1*, *TET2* and *TET3*. *TET2* then underwent a chromosomal inversion event in which the exon containing the CXXC domain was detached and became a separate gene that encodes IDAX (inhibition of the Dvl and axin complex; also known as CXXC4)[2,32] (FIG. 2a, b).

CXXC domains occur in many chromatin-associated proteins. Three distinct subfamilies of CXXC domains can be identified by their sequence[33] (Supplementary information S1 (figure)). Subfamily 1 includes the CXXC domains of CXXC1 (also known as CFP1), DNMT1, MLL and Lys-specific demethylase (KDMA) family proteins, as well as the third CXXC domain of methyl CpG-binding protein 1 (MBD1); these CXXC domains specifically recognize unmodified CG sequences and target subfamily 1 proteins to CpG-rich sequences in DNA[34–38]. Subfamily 2 consists of the first two CXXC domains of MBD1; these CXXC domains have not yet been demonstrated to bind DNA and have no documented function[36]. Subfamily 3 includes the CXXC domains of TET1, TET3, IDAX and RINF (retinoid-inducible nuclear factor; also known as CXXC5)[2,39]. The IDAX CXXC domain preferentially binds unmethylated CpG sequences *in vitro*, and is found at CpG islands and CpG-rich promoters (which are predominantly unmethylated) in cells (Ref. 41). Curiously, however, the TET1 CXXC domain is reported not to bind DNA[33] or alternatively to bind CpG sequences regardless of whether the cytosine is modified[39,40], and the TET3 CXXC domain is reported to bind unmodified cytosine irrespective of whether it is followed by a guanine, a finding supported by affinity measurements and X-ray crystallography[42]. As subfamily 3 CXXC domains have a strong positive charge that increases their tendency to bind nonspecifically to DNA, an analysis of binding sites *in vivo* is needed to reconcile these *in vitro* findings.

Surprisingly, IDAX (a reported inhibitor of WNT signalling[43]) targets TET2, the protein from which it was separated during evolution, for destruction via a caspase-dependent mechanism[41]. Depletion of IDAX from ES cells prevented the downregulation of TET2 that normally occurs upon differentiation, whereas its depletion from a human myeloid cell line increased TET2 and 5hmC levels. IDAX cannot negatively regulate TET2 if key DNA-binding residues in the IDAX CXXC domain are mutated, suggesting that IDAX recruits TET2 to DNA before the caspase-dependent degradation mechanism is activated. The implication is that TET2, which like other TET proteins is likely to also bind DNA through the cysteine-rich region of its catalytic domain (see below), binds different sets of genomic regions depending on whether IDAX is present. The CXXC domain of TET3 negatively regulates TET3 catalytic activity perhaps through an autoinhibitory mechanism that involves a physical interaction between the CXXC and catalytic domains or, alternatively, by tethering TET3 to particular DNA elements and thus limiting its genome-wide activity.

The catalytic domains of TET proteins are characteristic of Fe^{2+}- and 2-oxoglutarate-dependent dioxygenases (reviewed in REF. 22). The DSBH fold, which comprises the catalytic domain of all these dioxygenases, contains an His-Xaa-Asp/Glu signature motif, a C-terminal conserved His residue that is involved in coordinating Fe^{2+} and a conserved Arg residue that binds 2-oxoglutarate via a salt bridge[21,22]. Because a crystal structure of the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

TET catalytic domain is not yet available, the structure of the AlkB DSBH domain and a detailed view of the active site44 are shown in FIG. 2c. As is the case for AlkB, the substrate 5mC is likely to be flipped out of the DNA double helix into the catalytic cavity of the TET protein, where it is brought in close proximity to the catalytic Fe^{2+} ion (FIG. 2c).

The inserted Cys-rich domain in the catalytic region of TET proteins is likely to chelate two or more Zn^{2+} ions via nine conserved Cys residues and one His residue, and has been postulated to be part of a DNA-binding surface that might help in target recognition[2]. Indeed, the same region contains comparable insertions in the distantly related Jumonji-like family[45] (for example, the AT-rich interaction domain (ARID; also known as BRIGHT) in Jumonji and ARID domain-containing 1 (JARID1) and the histone-binding PHD finger in SMCX (also known as KDM5C or JARID1C)). Metazoan TET proteins also contain a large low-complexity insert predicted to be on the exterior surface of the DSBH fold[2,32] (FIG. 2c). The sequence and size of this unstructured insert varies greatly between TET family members, but its continued presence indicates a function, perhaps as a site of protein–protein interaction[2,32].

TET-mediated DNA demethylation: mechanisms

Interest in TET proteins has primarily centred around the possibility that oxidized methylcytosines could serve as intermediates in one or more ‘DNA demethylation’ pathways. In such a pathway, TET would oxidize 5mC to 5hmC, 5fC or 5caC, which would then be replaced with cytosine, the net result thus being ‘demethylation’. There are at least four mechanisms by which TET proteins could mediate DNA demethylation (FIG. 1a).

Facilitation of passive DNA demethylation

Because CG sites are palindromes that in general are symmetrically methylated[46], DNA replication yields two strands with hemimethylated CG sites. Maintenance of DNA methylation patterns requires the DNA methyltransferase DNMT1 and its obligate partner UHRF1 (REF. 47) (FIG. 1c). UHRF1 binds to hemimethylated CpG sites via its SAD/ SRA (SET associated Deinococcus domain (SAD)/SET and RING associated (SRA) domain) domain and recruits DNMT1. DNMT1 then methylates the CG sites on the nascent DNA strand, thus maintaining methylation patterns through cell division[47,48]. *In vitro*, UHRF1–hemi-5hmC binding is tenfold less efficient than UHRF1–hemi-5mC binding[49]. In addition, the activity of recombinant DNMT1 is reduced 12-fold[50] or 50-fold[49] at sites of hemi-5hmC *in vitro*. Together, these results imply that the TET-mediated hydroxymethylation of a methylated CG site *in vivo* can block maintenance methylation during cell division and eliminate 5mC in a ‘passive’, replication-dependent manner. However, hydroxymethylated plasmids stably transfected into transformed human cells retain maintenance methylation through cycles of plasmid division as efficiently as a methylated plasmid[51], suggesting that 5hmC does not completely block maintenance methylation in cells.

Active DNA demethylation through DNA repair

Two replication-independent (‘active’) demethylation mechanisms have been reported to couple the methylcytosine oxidase activity of TET proteins with `base excision repair` (BER). The first, which involves 5fC, 5caC and TDG, has been confirmed by multiple laboratories[6,30,53]. The second mechanism, which involves AID (activation-induced cytidine deaminase) and APOBEC (apolipo-protein B mRNA editing enzyme, catalytic polypeptide), is still controversial[52,54]. In the first mechanism, TET proteins further oxidize 5hmC to generate 5fC and 5caC[5,6,28]. 5fC and 5caC can be excised by TDG[6,30]; their replacement with cytosine results in demethylation (FIG. 1a). Electrophoretic mobility shift assays (EMSAs) and the structure of TDG in complex with 5caC[53] indicate that TDG binds 5caC:G

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

mismatches with higher affinity than T:G mismatches (its conventional substrate). Depletion of TDG causes a 2–10-fold increase in the levels of 5fC and 5caC in ES cells[6,55,56], consistent with a model in which these bases are short-lived demethylation intermediates that are removed by TDG. However, even in a TDG-deficient cell, 5fC is rare compared with 5mC (0.2–0.3% of the total methylcytosines)[5,55,56], and 5caC is even less abundant[6,56], indicating that this demethylation pathway has limited throughput in ES cells.

The second DNA repair-based mechanism has been reported by one laboratory in transfected human cells and in mouse brain[52]. They proposed that 5hmC is deaminated to 5-hydroxyuracil (5hmU) by AID and APOBEC family enzymes, removed by SMUG1 (single-strand-selective monofunctional uracil DNA glyco-sylase 1) or TDG glycosylases and ultimately replaced by cytosine (FIG. 1a). This sequential deamination and removal of 5hmC is similar to the deamination of 5mC and the removal of the resulting T:G mismatches by TDG, a mechanism previously proposed to occur in zebrafish[57,58], human cancers[58], primordial germ cells[59] and fused cells undergoing reprogramming[60]. However, there is considerable controversy about the significance of deamination-based demethylation mechanisms in general. In support of such mechanisms, TDG can excise T:G mismatches and 5hmU:G mismatches *in vitro*[54,61], and both PGCs derived from *Aid*-deficient mice and mouse embryonic fibroblasts (MEFs) derived from *Tdg* -deficient mice have modestly increased levels of methylation at some CpG island promoters[61,62]. Arguing against a deamination mechanism, the AID enzyme primarily acts on single-stranded DNA[63], and recombinant AID and APOBEC enzymes have their strongest activity against unmodified cytosine with reduced activity against 5mC and no detectable activity against 5hmC[54,64]. It therefore seems unlikely that AID and APOBEC enzymes have a role in 5hmC-dependent demethylation pathways, although their involvement in specific situations cannot be ruled out.

Enzymatic decarboxylation of 5caC

There is some evidence that 5caC may be decarboxylated by protein(s) present in ES cell lysates. An oligonucleotide containing 5caC that was labelled with ^{15}N at both positions of the pyrimidine ring was synthesized and incubated with ES cell lysate. Following the recovery and digestion of the oligonucleotide, a small but detectable quantity of [$^{15}N_2$]-deoxycytosine was detected, indicating direct conversion of 5caC to cytosine without BER[31] (FIG. 1a). The factors that catalyse this decarboxylation reaction remain to be identified.

Dehydroxymethylation by DNMT enzymes

Remarkably, DNMT enzymes can remove the hydroxymethyl group of 5hmC *in vitro*, directly converting 5hmC to cytosine[65,66] (FIG. 1a). Reducing conditions favour the methyl-transferase activity of DNMT3A, whereas oxidizing conditions favour dehydroxymethylation[66]. Whether this reaction occurs in cells is still unknown. Two papers reported cyclical changes in the DNA methylation status of several oestrogen-responsive genes upon the addition of oestrogen[67,68], and one of these also showed cyclical recruitment of all three DNMTs[67]. It would be interesting to revisit this system to determine whether reciprocal actions of TET proteins and DNMTs were involved.

When a biological role is observed for a TET protein, there has been a strong tendency in the literature to invoke DNA demethylation as a major mechanism. However, an unequivocal demonstration of demethylation can be difficult, especially as the changes observed in most studies have been subtle. At a given locus, TET may simply be converting 5mC to 5hmC, or it may be converting 5mC to cytosine via an oxidized 5mC intermediate by one of the pathways described in FIG. 1a. Actual DNA demethylation only occurs when 5mC is converted to cytosine via an oxidized 5mC intermediate. TET depletion will

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

necessarily result in decreased 5hmC, 5fC and 5caC, and a corresponding increase in 5mC from which these oxidized methylcytosines are derived. Thus, a simple increase in 5mC upon TET depletion is not enough to prove that a TET protein is causing demethylation. Rather, the increase in 5mC has to be greater than the attendant decrease in 5hmC, 5fC and 5caC. If TET is demethylating a locus, TET depletion will result in a decrease in unmodified cytosine as judged by bisulphite sequencing. Even here, caveats apply: 5fC and 5caC will appear as unmodified cytosine. Another point to bear in mind is that binding of transcription factors to a TET-regulated locus may antagonize DNA methylation (REFS 69–72), thus if TET depletion results in altered transcription factor binding, the resulting changes in cytosine species could lead to incorrect interpretations of whether TET is causing demethylation at that locus or not. Methylated DNA immunoprecipitation (MeDIP) and other methods of measuring 5mC abundance are less useful than bisulphite sequencing, as simple conversion of 5mC to 5hmC will produce a drop in measured methylation even if TET does not mediate demethylation at the locus.

Genomic distribution of oxidized 5mC

Large-scale mapping of 5hmC, and more recently 5fC and 5caC, has been performed for mammalian genomes. The available methods for mapping these modified bases are outlined in BOX 1 and BOX 2. Commercial kits are available in many cases.

Oxidized methylcytosine enrichment at promoters

There is contention as to whether 5hmC is enriched at promoters, owing largely to how ‘promoter’ and ‘enrichment’ are defined. First, the definition of promoter varies between laboratories. Many promoters that are CpG-rich contain peaks of hydroxymethylation 500–2000 bases before and after the transcription start site (TSS) but are depleted for 5hmC at the TSS itself, where there is typically almost no modified cytosine[73,74]. In this Review, we classify these promoters as 5hmC-enriched. Second, the definition of enrichment varies depending on whether precipitation-based or single-base resolution methods are used. Precipitation-based methodologies compare how many times a given region of DNA is sequenced in a 5hmC pulldown compared with a negative control sample. As a result, regions that contain a higher than average number of 5hmC residues in a span of several hundred bases will be classified as enriched. In contrast, single-base resolution studies typically measure the fraction of total reads in which an individual CpG site is hydroxymethylated. Thus, promoters with a high CpG density, but in which each CpG site has a relatively low frequency of 5hmC, may be classified as 5hmC-enriched by a p recipitation-based methodology but as 5hmC-depleted by a single-base methodology. We will predominantly discuss results from precipitation-based studies, although it is unclear which definition of enrichment is more biologically relevant. Furthermore, any given cytosine modification is either present or absent at any given CpG site at the level of a single allele, but can seem infrequent when a large, potentially hetero-genous population of cells is examined.

In human[73,75] and mouse[11,74,76] ES cells, mouse neural progenitor cells[77], mouse neurons[77] and the cerebellum of 7-day-old mice[78] (which contains a significant population of neural progenitor cells), 5hmC is enriched at promoters in absolute terms and especially when compared with 5mC, which is generally depleted at promoters[79]. In these cell types, genes with hydroxymethylated promoters are expressed at lower levels than other genes, although this does not prove a silencing role for 5hmC; 5hmC can only be formed at promoters that have some underlying level of methylation, and promoter methyl-ation correlates with silencing. In other cell types, including PGCs[80], adult nervous tissue[78], liver cells[81] and benign nevi[82], 5hmC is depleted from TSSs. The discrepancies in 5hmC distribution among

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

cell types could be caused by differences in the underlying distribution of 5mC and TET proteins.

In ES cells, 5fC is found at largely the same set of promoters as 5hmC[83], and levels of 5fC[55] and 5caC[56] increase at these promoters upon TDG depletion, indicating that at least some TET-mediated demethylation occurs at these loci.

Oxidized methylcytosine enrichment at gene bodies

In virtually all mammalian cell types studied, including mouse and human ES cells, mouse liver and brain and human melanomas, 5hmC is enriched at gene bodies[11,73,77,78,81,82]. In mouse brain, liver and ES cells (but not in human ES cells), gene expression positively correlates with gene body hydroxymethylation[78,84,85], although the reason for this correlation is unclear. While promoter methylation strongly correlates with gene silencing, how gene body methylation correlates with gene expression varies with cell type and whether the methylated cyto-sine exists in a CpG or non-CpG context[86–88], making it unclear how 5mC oxidation and/or DNA demethylation by TET proteins in gene bodies affects gene expression.

Oxidized methylcytosine enrichment at enhancers

In all mammalian cells in which both 5hmC and enhancers have been mapped, 5hmC is strongly enriched at enhanc-ers[72,73,75,81,89], which are often regions of low CpG density with reduced levels of DNA methylation compared with neighbouring regions[72]. Single-base resolution mapping using TET-assisted bisulphite sequencing (TAB-seq) (BOX 2) indicates that 5hmC is strongly enriched immediately adjacent to transcription factor-binding sites in enhancers, but is depleted at the precise site of binding[89]. Due to the absence of data on the binding of transcription factors to enhancers in TET-deficient cells, it is difficult to know whether methylcytosine oxidation ‘opens’ chromatin structure, allowing transcription factors to bind, whether transcription factors recruit TET proteins or whether both mechanisms operate and reinforce one another. Moreover, the reduction in 5mC and 5hmC at transcription factor-binding sites may reflect demethyl-ation mediated by TET proteins, but could also be because transcription factors physically block access of DNMT[69,90] and TET proteins. Probably, both mechanisms are at work.

Notably, in differentiating ES cells, 5hmC levels increase sharply at activated enhancers[91]. 5hmC levels rise almost immediately with the onset of differentiation and either precede or accompany acetylation of Lys27 of histone H3 (H3K27), which is a mark of active enhancers. Furthermore, 5fC and 5caC are enriched at enhancers in TDG-depleted ES cells, consistent with TET-mediated demethylation of enhancers[55,56]. However, the correlation between the 5fC levels at enhancers and hypomethylation is weak[55]. Further studies will be necessary to determine the extent to which TET proteins contribute to enhancer hypomethylation and function.

Mapping TETs and TET-interacting proteins

The distributions of TET proteins and oxidized 5mC species do not overlap neatly. All three TET proteins and IDAX are strongly enriched at promoters, especially promoters that are CpG-rich[39,41,92–95]. In the case of TET1 (REF. 39) and TET3 (REF. 42), this preference is partially determined by the presence of CXXC domains, which most likely bind CpG sequences, as discussed above.

The data support a repressive role for TET1 in ES cells. In three studies in mouse ES cells, there is a strong statistical correlation between the physical presence of TET1 at a promoter and increased gene expression from this promoter upon TET1 knockdown[39,92,93]. The

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

correlation between the presence of TET1 at a promoter and decreased expression from this promoter upon TET1 knockdown is much weaker. TET2 shows precisely the opposite trend: there is a correlation between the physical presence of TET2 at a promoter and decreased expression from this promoter upon TET2 depletion, which indicates that TET2 is generally a positive regulator of gene expression[95]. The likely explanation for the different activities of TET1 and TET2 is that, in ES cells, TET1 primarily modulates transcription by recruiting the repressive histone deacetylase-containing MBD3–NURD (nucleosome remodelling and deacetylase)[96] and SIN3A (switch-independent 3A) complexes[13,94,97] to promoters. TET2 shows no association with these complexes in ES cells[93] and, when exogenously expressed in 293T cells, TET1 and TET3 associate with SIN3A, whereas TET2 does not[94]. Furthermore, in contrast to somatic cells in which methyl-ation has an impact on the expression of thousands of genes[98], methylation only regulates a handful of genes in ES cells and is dispensable for ES cell growth and survival[99]. Hence, the demethylating properties ascribed to TET proteins may be of limited importance in regulating transcription in ES cells.

Recently, three groups have reported a physical association between TET proteins and the enzyme OGT (*O*-linked β-D-*N*-acetylglucosamine (*O*-GlcNAc) transferase)[94,95,97]. Despite some contradictions in these studies, it is probable that all three TET proteins associate with OGT[100]. OGT adds a GlcNAc sugar to Ser and Thr residues of numerous proteins, including histone H2B and other chromatin modifiers. The impact of this sugar modification on the target protein is highly contextual. For instance, *O*-GlcNAcylation can serve to antagonize phosphorylation, direct proteolysis or allow protein– protein interaction[101]. OGT, like TET proteins, is strongly enriched at promoters. In ES cells, TET proteins recruit OGT to promoters, but not vice versa. Depletion of TET proteins diminishes the amount of OGT at promoters, but depletion of OGT does not influence the association of TET1 or TET2 with chromatin, although it does alter the distribution of 5hmC at certain loci[95, 97]. There are various mechanisms by which OGT could influence transcription. For example, levels of *O*-GlcNAcylation at Ser112 of H2B, a mark associated with transcriptional activation[102], drop sharply upon TET2 depletion providing a mechanism whereby TET2 increases transcription through OGT. Likewise, OGT glycosylates HCF1 (host cell factor 1), a component of the H3K4 methyltransferase SET1 complex (also known as COMPASS) associated with active RNA polymerase II (Pol II), providing a link between H3K4 trimethylation and TET proteins[94]. Given the many protein targets of OGT, it is likely that there will be additional consequences of TET-mediated OGT recruitment.

TET proteins in embryonic development

TET proteins are implicated in several stages of mouse development, particularly those in which mass demethyl-ation and *de novo* methylation take place, namely the `zygote`, the inner cell mass of the blastula and PGCs (FIG. 3). The phenotypes of TET-deficient mice are described in BOX 3.

TET3 in the zygote

Immediately after fertilization, before the genetic material from the sperm and egg have fused to form one nucleus, the male pronucleus loses almost all methylation as assessed by staining with antibodies specific for 5mC[103–105] (FIG. 3). This process occurs before replication and is not inhibited by DNA polymerase inhibitors. In contrast, bisulphite sequencing, which does not distinguish between 5mC and 5hmC[106–108] (BOX 2), indicates that this loss of 5mC is much slower and less complete than the abrupt loss measured by antibody staining[104,109–111]. These results are reconciled by the observation that the zygotic demethylation event is in fact a mass 5mC oxidation event; the apparent loss of 5mC stems from the fact that 5mC-specific antibodies do not recognize 5hmC or other oxidized

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

methylcytosines. Instead, the male pronucleus shows a selective increase in 5hmC[111–113] and other oxidized methylcytosines[114], as observed by staining with appropriate antibodies. The increase in 5hmC is abrogated by siRNA-mediated knockdown of *Tet3* (REF. 112), and is not observed in fertilizations that involve TET3-deficient oocytes[113]. This indicates that TET3 oxidizes 5mC in the male pronucleus.

The protein DPPA3 (development pluripotency-associated 3; also known as Stella or PGC7) protects maternal DNA and a few `imprinted loci` in the paternal genome from TET3-mediated oxidation[115]. DPPA3 is reported to be recruited to the protected regions by dimethylation of H3 at Lys9 (H3K9me2), a mark that is found selectively at these sites[115]. In the absence of DPPA3, both paternal and maternal genomes become hydroxymethylated[112,115] and most show abnormalities by the 2-cell or 4-cell stage[116–118]. In contrast, *Tet 3* -deficient oocytes fertilized with wild-type sperm have reduced expression of OCT4 during the morula stage but form blastocysts normally. Moreover, about half of these animals show serious abnormalities by embryonic day E11.5 and fail to develop[113]. Thus, methylcytosine oxidation in the male pronucleus enhances, but is not required for, survival of the embryo. It is not clear why some zygotes survive, whereas others do not, and why it takes 11 days for the failure of zygotic oxidation to result in lethality.

The bulk of the evidence suggests that cytosine modification is lost through a replication-dependent process that may be enhanced by, but is not completely dependent on, TET enzymes or BER. In fact, maintenance methyl-ation is very inefficient in the early embryo because of the cytoplasmic localization of DNM1O, which is a special splice variant of DNMT1 present in the early mouse embryo[80,103,105]. As a result, 5mC on the maternal DNA and 5hmC on the paternal DNA are diluted out through a passive replication-dependent process (FIG. 3). Substantial levels of 5hmC are retained at least through the 8-cell stage[111,119], arguing against the possibility that it is excised or converted directly to cytosine. However, some loss of 5mC is observed by bisulphite sequencing before any cell replication event[113]. A likely explanation for this is that TET3 is oxidizing 5mC to 5fC and 5caC[114], creating the appearance of demethylation by bisulphite sequencing (BOX 2). It is also plausible that some proportion of 5fC and 5caC is removed by TDG and replaced by cytosine[30]. Evidence showing BER in the male pronucleus supports this possibility[110,120]. Arguing against direct removal of 5fC and 5caC is evidence demonstrating that the modified bases remain in the embryo at least until the 4-cell stage[114] and the fact that large-scale BER would be expected to cause increased rates of mutations in the zygote, which would be deleterious to the organism and its progeny.

As methylation is lost passively anyway, why are methylcytosines oxidized in the mammalian zygote? The most likely possibility is that methylcytosine oxidation in the male pronucleus accelerates global DNA demethylation, either through 5fC and 5caC formation and the subsequent removal of these modified bases or by ensuring that maintenance methylation is especially inefficient in the earliest cell divisions. No proof of either mechanism has been demonstrated. It is also possible that cytosine oxidation promotes broad transcriptional activation previously observed in the male pronucleus[121]. However, depletion of TET3 in the early zygote is reported to have no effect on global transcriptional levels[122]. In short, it is clear that TET3 oxidizes the male pronucleus and that this oxidation event enhances survival of the early embryo, but why this oxidation occurs remains unknown.

TET1 and TET2 in ES cells

When pluripotent ES cells are injected into immunocompromised mice, they form tumours called teratomas that contain tissues from all three germ layers (ectoderm, mesoderm and endoderm). *Tet1*$^{-/-}$ ES cells[123], *Tet1*$^{-/-}$ *Tet2r*$^{-/-}$ ES cells[124] and ES cells stably expressing

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Tet1 short hairpin RNA (shRNA)[125] form large haemorrhagic teratomas that are enriched for trophoblast cells, suggesting that TET1 regulates the first lineage commitment step in the embryo by suppressing differentiation towards the extra-embryonic lineage. Notably, however, *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ ES cells contribute efficiently to chimaeras when injected into blastocysts[124], indicating that the skewing towards trophoblast differentiation observed in *Tet1*$^{-/-}$ and *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ teratomas can be overcome by other regulatory influences during embryonic development. TET1 depletion influences gene expression and differentiation in mouse ES cells by negatively regulating the expression of key `trophectoderm` regulators, such as caudal-type homeo box 2 (CDX2)[27,125,126], eomesodermin (EOMES)[93,123,125] and ELF5 (REF. 125), and positively regulating expression of the neuroectoderm factors paired box 6 (PAX6)[39,93,123,125,126] and neurogenic differentiation factor 2 (NeuroD2)[39,93,123]. However, neural progenitor cells can be derived from *Tet1*$^{-/-}$ `embryoid bodies`[123], and *Tet1*$^{-/-}$ and *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ mice develop all three germ layers and survive, suggesting that the role of TET1 and TET2 in differentiation is minor. Overall, the data indicate that, despite the abundance of 5hmC in ES cells, TET1 and TET2 have modulatory but not essential roles in mouse development, ES cell survival or pluripotency. Whether TET3 can compensate for the loss of TET1 and TET2 has not yet been established.

Nevertheless, the catalytic activity of TET proteins may be important in ES cells and the inner cell mass from which they are derived, because of the underlying biology of these cells. ES cells approximately recapitulate the stage of development in which large-scale *de novo* methylation of the genome occurs (FIG. 3). Because transcriptional initiation antagonizes methylation[71,87], genes with lower levels of transcription are presumably more susceptible to aberrant increases in DNA methylation. As methylation is generally maintained through cell divisions, aberrant methylation may be inconsequential at early embryonic stages but deleterious during later development. Thus, the function of TET proteins in the inner cell mass may be to repress lineage-specific genes, while simultaneously antagonizing methylation to permit activation of these genes later in development. Accordingly, in ES cells, both TET1 and 5hmC are enriched at promoters that contain dual H3K4me3 and H3K27me3 marks, which are so-called ‘poised’ promoters that can be activated later in development[74,76]. Loci with high levels of 5fC in ES cells show increased cyto-sine methylation in TDG-deficient MEFs[83], even though TET1 and TET2 are almost absent from MEFs. Thus, a demethylation defect in early development seems to result in excess methylation at a later stage.

Generally, the data support a real but limited role for TET1 and TET2 in demethylation in ES cells. Bisulphite sequencing of selected loci in TET1-deficient ES cells does indicate a very modest increase in methylation levels[39]. Mass spectrometry indicates that cells deficient in both TET1 and TET2 show an increase in 5mC levels (from 5.3% to 5.8% of all cytosines) and a smaller absolute drop in 5hmC levels (from 0.13% of all cytosine to undetectable)[124]. Thus, there is a ~1.05-fold increase in the total levels of cytosine modification (5.43% to 5.8%). If the increase in cytosine modification is concentrated at promoters at which CpG levels are high and TET proteins are most active, this could have a strong impact on transcription and development. However, caution must be used in extrapolating changes in methylation observed in ES cells, which undergo many generations of division in culture, to the inner cell mass that differentiate s quickly.

Most tissues in *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ mice show modest increases in total cytosine modification levels as measured by mass spectrometry, although it is unclear whether this is due to inadequate demethylation in the blastula, during later embryonic development or in adult tissue. The incompletely penetrant lethal phenotype of *Tet*$^{-/-}$*Tet2*$^{-/-}$ mice (BOX 3) could reflect a model in which, by stochastic variation, some mice acquire excess methylation at

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

key promoters in early development, leading to subsequent lethality, whereas other mice do not. However, the perinatal lethality of *Tet1*$^{-/-}$*Tet2*$^{-/-}$ mice cannot be assumed to be caused by alterations in DNA methylation. Furthermore, virtually all of the *Tet1*$^{-/-}$*Tet2*$^{-/-}$ mice develop to birth with fully developed organs, so any dysregulation of methylation is necessarily fairly mild.

TET proteins in PGCs

PGCs undergo a rapid drop in 5mC levels between E9.5 and E10.5 (FIG. 3). As in zygotes, this apparent demethylation event corresponds to mass conversion of 5mC to 5hmC, in this case mediated by TET1 and TET2 (REFS 80,127). There is no detectable formation of 5fC or 5caC as measured by immunocytochemistry[80]. Measurement of individual loci during germ cell development indicates that 5mC is replaced by 5hmC between E9.5 and E11.5, and that both marks are subsequently lost at a rate consistent with dilution by replication, as opposed to active removal[80]. Thus, if TET proteins are mediating demethylation in PGCs, it is by facilitated passive demethylation. However, even this mechanism may not be critical, as UHRF1 and DNMT3B levels are quite low in PGCs[128] and thus passive demethylation could occur without TET activity.

Germ cells derived from mouse *Tet1* `gene-trap` mutants (BOX 3) have almost normal methylation levels[129]. At E13.5, when demethylation is complete, an increase in methylation corresponding to less than 1% of the total cytosines at CpG sites is observed in TET1-deficient cells[129]. A much greater effect is seen at the level of gene expression: loss of TET1 causes dys regulation of ~1000 genes, 90% of which are positively regulated by TET1, indicating that TET1 primarily functions to activate transcription in PGCs. A number of critical meiosis-related genes are targets of TET1 activation, including malate dehydrogenase 1 (*Mae1*), synaptonemal complex protein 1 (*Sycp1*) and *Sycp3*. Half of the developing TET1-deficient female gametes subsequently display a defect in `meiotic synapsis` and suffer a developmental arrest between E16.5 and E18.5 (REF. 129), and accordingly *Tet1* gene-trap and *Tet1*$^{-/-}$ mice have reduced litter size. Although it is possible that TET1 facilitates critical demethylation events at key genes, it seems unlikely that such a modest change in methylation levels has such a drastic effect on gene expression. It is more likely that TET1 modulates gene expression by recruiting other proteins or that oxidized cytosines are influencing transcription in a way that does not involve their removal.

The simultaneous shRNA-mediated depletion of TET1 and TET2 (REF. 80), or the depletion of TET1 from *Tet2*$^{-/-}$ cells[127], seems to partly antagonize demethyl-ation at a few loci in PGC-like cells derived *in vitro* from precursors. Furthermore, there is evidence that oocytes deficient in TET1 and TET2 may not completely erase all parental imprints during PGC demethyl-ation. Although *Tet1*$^{+/-}$*Tet2*$^{+/-}$ mice had no intrinsic survival defect, half of the progeny of *Tet1*$^{-/-}$*Tet2*$^{-/-}$ females crossed with wild-type males display perinatal lethality[124]. This indicates that some fraction of the *Tet1*$^{-}$ *Tet2*$^{-}$ oocytes completed meiosis successfully but were nevertheless defective, as judged by the perinatal lethal phenotype observed in mice developing from the *Tet1*$^{-}$*Tet2*$^{-}$ oocytes. Accordingly, some imprinted loci in the progeny of *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ mice show aberrant methylation. The occurrence and degree of increased methylation varies across loci and between individual mice. This is consistent with mass genomic demethyl-ation proceeding normally in TET-deficient animals, but the subsequent erasure of imprints being variably impaired to cause fatal defects in some, but not all, progeny.

Thus, in the two instances of mass demethylation in mammals, essentially the same pattern occurs. 5mC is oxidized enmasse by TET3 in zygotes and TET1 and TET2 in PGCs, predominantly producing 5hmC, and there is little evidence that the oxidized methylcytosines are excised. Rather, all modified cytosines are diluted out during

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

subsequent cell divisions, due to the expression of DNMT1O in zygotes and the downregulation of UHRF1 in PGCs. Insofar as oxidized methylcytosines are promoting demethylation in these systems, they probably achieve this by antagonizing maintenance methylation under conditions in which this process is already in efficient. Potentially, the oxidized methylcytosines have important functions that are not related to their removal as, for example, epigenetic marks.

TET proteins in reprogramming

The transfection of developmentally-committed cells with certain combinations of transcription factors can reprogramme them, with low efficiency, into induced pluripotent stem (iPS) cells that are phenotypically similar to ES cells[130]. The classic combination of transcription factors used in these experiments is OCT4 (also known as POU5F1), SOX2, Krüppel-like factor 4 (KLF4) and MYC (collectively referred to as OSKM), although other combinations have also been used successfully[131]. For efficient reprogramming, endogenous pluripotency factors must be activated and their promoters and enhancers demethylated[132]. Although TET proteins are apparently dispensable for pluripotency in ES cells and *in vivo*, several recent publications indicate roles for TET1 and/or TET2 in the generation of iPS cells, although they do not report a common mechanism. Specifically, one study showed that the depletion of TET2 by shRNA completely ablated reprogramming of fibroblasts by OSKM, and that the *Nanog* promoter was hydroxymethylated and demethylated during the first 4 days after reprogramming[133]. These findings, together with the fact that TET1 and TET3 expression was not observed during this time frame, suggest that TET2 has a role in generating iPS cells. In another study, TET1 and TET2 were shown to physically interact with Nanog in ES cells, and co-transfection of Nanog with either TET1 or TET2 greatly enhanced OCT4, KLF4 and MYC-mediated reprogramming of neural stem cells[134]. The authors proposed that Nanog and TET1 or TET2 could co-activate the endogenous *Oct4* and *Esrrb* (oestrogen-related receptor beta) genes via a mechanism that involves at least some demethylation. A third study showed that co-transfecting TET1 with OSKM factors greatly enhances the reprogramming of fibroblasts[135]. These data suggest that TET1 mediates the oxidation and demethylation of the *Oct4* promoter and proximal enhancer, and that TET1 can substitute for OCT4 in OSKM-dependent reprogramming transfection. In all of the above cases, demethylation of the target loci is slow enough to be caused by a passive mechanism. The requirement for TET proteins in reprogramming clearly varies depending on the precise conditions and status of the starting cell lines. However, the ability of TET protein s to generate oxidized methylcytosines and antagonize methylation of select loci may help to establis h the pluripotent state.

TET proteins in the nervous system

The brain consistently has the highest levels of 5hmC of any mammalian tissue. This may in part reflect the abundance of long-lived post-mitotic cells in the nervou s system, such that 5hmC can accumulate in neurons without being depleted through replication. It is still largely unknown how TET deficiency influences mouse brain development. Increased 5hmC and decreased DNA methylation clearly correlate with higher gene expression levels and a more open chromatin state in neural cells[85] but causality has not been established. Depletion of TET2 and TET3 in the developing mouse cortex, by *in utero* electroporation of shRNA, causes a block in the differentiation of neural progenitor cells into neurons[77].

Xenopus laevis embryos injected with morpholinos targeting *tet3* develop a gross defect in early development, that features loss of expression of critical eye and neural markers and results in small heads and no eyes[42]. The phenotype was partly rescued by injection of mRNA coding for a catalytically inactive TET3, whereas injection with a TET3 CXXC

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

domain mutant provided no rescue at all. Thus, TET3 has a role in normal head development in *X. laevis*. It is unknown whether the perinatal lethality observed in *Tet3*$^{-/-}$ mice[113] is linked to a neurological development defect.

Interpreters of 5mC oxidation

As the extent to which 5mC oxidation products are demethylation intermediates is unclear, other mechanisms by which they could influence transcription should be considered.

Some transcriptional regulators and chromatin-associate d proteins specifically recognize 5hmC or other oxidized methylcytosines. Mass spectrometry has been used to identify proteins bound selectively to 5hmC-containing DNA oligonucleotides in ES cells, neural progenitor cells and mouse adult brain[136]. Proteins that were found to bind to 5hmC included the neural progenitor cell-specific protein UHRF2, which presumably recognizes 5hmC-containing DNA via its SAD/SRA domain, transcription factors such as zinc-fingers and homeoboxes protein 1 (ZHX1), ZHX2 and THAP domain-containing protein 11 (THAP11), as well as several uncharacterized proteins. An important limitation of this study is that it does not distinguish between proteins that bind directly and indirectly to 5hmC. Despite the inability of glycosylases to cleave 5hmC *in vitro*[52], several glycosylases and DNA repair proteins were found to associate with 5hmC in this study, suggesting that 5hmC may be a target of DNA repair[136]. In addition, two methyl-binding proteins, methyl CpG binding-protein 2 (MeCP2) and methyl-CpG binding domain protein 4 (MBD4), were shown in this and other studies to bind to 5hmC[85,136,137]. However, an initial finding that MBD3 ;REF 96) preferentially binds to 5hmC over 5mC has not been reproducible[49,136].

TDG specifically binds 5fC and 5caC *in vitro*[53] and might therefore be a reader of 5fC and 5caC. Indeed, 5fC and 5caC were found to bind TDG in mouse ES cell, as well as a substantial population of DNA repair proteins, several transcription factors and a number of miscellaneous proteins, some of which may simply bind formyl or carboxyl groups nonspecifically[136]. A suggestion that TDG recruits the histone acetyltransferase p300 ;REF 138) is countered by the finding that the number of p300-binding sites in the genome actually increases in TDG-deficient ES cells[55]. However, TDG could nevertheless be a transcriptional regulator, as it has been reported to interact with multiple DNA-binding transcription factors[139].

Oxidized methylcytosines can affect Pol II processivity directly. 5mC and 5hmC have a relatively small effect on polymerase processivity *in vitro*, but a single 5fC or 5caC can drastically stall the progression of Pol II on a recombinant substrate[140]. This phenomenon is functionally and perhaps also mechanistically analogous to that observed with base J, which has a crucial role in transcription termination in kinetoplastids[25].

Finally, methylcytosine oxidation could antagonize 5mC-mediated silencing. Six different mechanisms for methylation-based silencing have been demonstrated; BOX 4). The relative importance of each mechanism is unclear, but their relevance clearly varies across cell types and loci. The impact of methylcytosine oxidation on these silencing mechanisms is also likely to be variable. Thus, in order to understand how 5hmC influences transcription at a locus, it may first be necessary to determine how methylation mediates silencing at that locus.

Conclusions and future directions

The discovery of TET proteins has provided a new perspective on how DNA modification influences gene expression. There are now several potential mechanisms by which TET proteins and oxidized methylcytosines might mediate DNA demethylation, but whether and

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

how these mechanisms contribute to gene regulation in different cell types remains to be elucidated. It will be important to determine how proteins that bind oxidized methylcytosines contribute to gene regulation and how the enzymatic activities and biological roles of TET proteins are influenced by interactions with other protein complexes. The ongoing development of improved methods for mapping oxidized methylcytosines promises to advance the field substantially, especially if it can be applied at single-base resolution to small numbers of differentiating cells. A careful analysis of changes in the distribution of oxidized methylcytosines, transcription levels, transcription factor binding and chromatin modifications in cells from animals lacking one or multiple TET proteins will also be essential to further our understanding of how TET proteins modulate gene expression, cell differentiation and function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank S. Evans and A. Clark for comments on the manuscript. This work was supported by US National Institutes of health (NIH) R01 grants AI44432, HD065812 and CA151535, grant RM-01729 from the California Institute of Regenerative Medicine, and Translational Research grant TRP 6187–12 from the Leukemia and Lymphoma Society (to A.R). Work in the laboratory of L.A. is supported by intramural funds of the National Library of Medicine, NIH. W.A.P. was supported by a predoctoral graduate research fellowship from the National Science Foundation and is currently supported by a postdoctoral fellow ship from the Jane Coffin Childs Memorial Fund for Medical Research.

Glossary

Dioxygenases	Enzymes that catalyse the addition of both oxygen atoms from molecular oxygen to one or two organic substrates
DNA demethylation	Here, defined as replacement of 5-methylcytosine, the major methylated base in mammalian DNA, with unmodified cytosine, either directly or through intermediates
CXXC domain	A Zn^{2+}-chelating domain typified by the signature amino acid sequence $CGXCXXC(X)_NC$, in which X represents any amino acid. CXXC domains in metazoans always contain two such sequences
CpG sequences	Any instance of a cytosine followed immediately by a guanine on the same strand of DNA. Most 3DNA methylation in mammals occurs at CpG sites
Base excision repair	(BER). A DNA repair pathway in which a DNA base is removed by a glycosylase enzyme and ultimately replaced by a new basel
Primordial germ cells	(PGCs). Precursors of mature germ cells (egg in female and sperm in male)
Click chemistry	Chemistry involving high-yield, highly specific reactions that are compatible with physiological conditions and maintain the integrity of biological molecules
Sequencing coverage	Average number of times that a genome or a DNA region is sequenced using a next-generation sequencing instrument
Zygote	Cell formed by fertilization of the oocyte (egg) with a sperm cell

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Imprinted locus	In epigenetics this describes a genomic region with a methylation mark that is present only on the maternally or paternally derived copy of an allele
Trophectoderm	Cells that give rise to the placenta and other extra-embryonic tissue
Embryoid bodies	Aggregates of cells formed by allowing embryonic stem cells to differentiate without contact with a solid surface
Inbred mouse strain	Experiments are typically conducted using inbred mouse strains, in which all mice are genetically extremely similar. C57BL/6 is one of the most frequently used strains. 129P2/OlaHsd is another inbred mouse strain. The same mutation can have different effects in different backgrounds
Gene-trap	A mutant in which a gene is disrupted by the random insertion of transgenic DNA that contains a splice acceptor site followed by stop codons
Meiotic synapsis	The event in meiosis prophase I in which homologous chromosomes align to allow recombination of genetic material (known as ‘crossing over’)

References

1. Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET 1. Science. 2009; 324:930–935. [PubMed: 19372391] Discovery that TET proteins oxidize 5mC to 5hmC
2. Iyer LM, 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. 2009; 8:1698–1710. [PubMed: 19411852] Describes the evolution of TET proteins and the presence of TET homologues in non-metazoan species
3. 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. 2002; 62:4075–4080. [PubMed: 12124344]
4. Lorsbach RB, et al. TET1, pa member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10; 11)(q22;q23). Leukemia. 2003; 17:637–641. [PubMed: 12646957]
5. Ito S, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011; 333:1300–1303. [PubMed: 21778364] Shows that TET proteins can produce 5fC and 5caC and quantifies the level of modified cytosines in a range of cell types
6. He YF, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Sdence. 2011; 333:1303–1307.Together with references 30 and 53, shows that TDG excises 5fC and 5caC. Demonstrates the presence of 5caC in mammalian DNA and shows that 5caC levels increase upon TDG depletion
7. Szwagierczak A, Bultmann S, Schmidt CS, Spada F, Leonhardt H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Adas Res. 2010; 38:el81.
8. Globisch D, et al. Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. PloS one. 2010; 5:e15367. [PubMed: 21203455]
9. Kriaucionis S, Heintz N. The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science. 2009; 324:929–930. [PubMed: 19372393] Together with reference 1, convincingly demonstrates the presence of 5hmC in mammals
10. Cimmino L, Abdel-Wahab O, Levine RL, Aifantis I. TET family proteins and their role in stem cell differentiation and transformation. Cell Stem Cell. 2011; 9:193–204. [PubMed: 21885017]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

11. Wu H, Zhang Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. Genes Dev. 2011; 25:2436–2452. [PubMed: 22156206]
12. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. Nature Rev. Genet. 2012; 13:7–13. [PubMed: 22083101]
13. Williams K, Christensen J, Helin K. DNA methylation: TET proteins-guardians of CpG islands? EMBO Rep. 2012; 13:28–35. [PubMed: 22157888]
14. Tan L, Shi YG. Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development. 2012; 139:1895–1902. [PubMed: 22569552]
15. Bejar R, Levine R, Ebert BL. Unraveling the molecular pathophysiology of myelodysplastic syndromes. J. Clin. Oncol. 2011; 29:504–515. [PubMed: 21220588]
16. Pronier E, Delhommeau F. Role of TET2 mutations in myeloproliferative neoplasms. Curr. Hematol. Malignancy Rep. 2012; 7:57–64.
17. Mercher T, et al. TET2, a tumor suppressor in hematological disorders. Biochim. Biophys. Acta. 2012; 1825:173–177. [PubMed: 22240200]
18. Iyer LM, Anantharaman V, Wolf MY, Aravind L. Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. Int. J. Parasitol. 2008; 38:1–31. [PubMed: 17949725] First paper to predict a 5mC oxidase and demethylase activity for TET1 (which was then termed CXXC6)
19. Gommers-Ampt JH, et al. β-D-glucosyl-hydroxymethyluracil: a novel modified base present in the DNA of the parasitic protozoan. T. brucei. Cell. 1993; 75:1129–1136.
20. Yu Z, et al. The protein that binds to DNA base J in trypanosomatids has features of a thymidine hydroxylase. Nucleic Acids Res. 2007; 35:2107–2115. [PubMed: 17389644]
21. Loenarz C, Schofield CJ. Expanding chemical biology of 2-oxoglutarate oxygenases. Nature Chern Biol. 2008; 4:152–156.
22. Loenarz C, Schofield CJ. Physiological and biochemical aspects of hydroxylations and demethylations catalyzed by human 2-oxoglutarate oxygenases. Trends Biochem Sci. 2011; 36:7–18. [PubMed: 20728359]
23. Aravind L, Koonin EV. The DNA-repair protein AlkB, EGL-9, and leprecan define new families of 2-oxoglutarate- and iron-dependent dioxygenases. Genome Biol. 2001; 2 RESEARCH0007.
24. Borst P, Sabatini R. Base J: discovery, biosynthesis, and possible functions. Annu. Rev. Microbiol. 2008; 62:235–251. [PubMed: 18729733]
25. van Luenen HG, et al. Glucosylated hydroxymethyluracil, DNA base j, prevents transcriptional readthrough in leishmania. Cell. 2012; 150:909–921. [PubMed: 22939620]
26. Ko M, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature. 2010; 468:839–843. [PubMed: 21057493]
27. Ito S, et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature. 2010; 466:1129–1133. [PubMed: 20639862]
28. Pfaffeneder T, et al. The discovery of 5-formylcytosine in embryonic stem cell DNA. Angewandte Chemie. 2011; 50:7008–7012. [PubMed: 21721093]
29. Penn NW, Suwalski R, O'Riley C, Bojanowski K, Yura R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic aciddeoxyribonucleic acid. Biochem. J. 1972; 126:781–790. [PubMed: 4538516]
30. Maiti A, Drohat AC. Thymine DNAglycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J. Biol. Chern. 2011; 286:35334–35338.
31. Schiesser S, et al. Mechanism and stem-cell activity of 5-carboxycytosine decarboxylation determined by isotope tracing. Angewandte Chemie. 2012; 51:6516–6520. [PubMed: 22644704]
32. Iyer LM, Abhiman S, Aravind L. Natural history of eukaryotic DNA methylation systems. Prog. Mol. Biol. Trans!. Sci. 2011; 101:25–104.
33. Frauer C, et al. Different binding properties and function of CXXC zinc finger domains in Dnmtl and Tet 1. PloS one. 2011; 6:e16627. [PubMed: 21311766]
34. Lee JH, Voo KS, Skalnik DG. Identification and characterization of the DNA binding domain of CpG-binding protein. J. Biol. Chem. 2001; 276:44669–44676. [PubMed: 11572867]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

35. Birke M, et al. The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation. Nucleic Acids Res. 2002; 30:958–965. [PubMed: 11842107]
36. Jorgensen HF, Ben-Porath I, Bird AP. Mbdl is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. Mol. Cell. Biol. 2004; 24:3387–3395. [PubMed: 15060159]
37. Blackledge NP, et al. CpG islands recruit a histone H3 lysine 36 demethylase. Mol. Cell. 2010; 38:179–190. [PubMed: 20417597]
38. Allen MD, et al. Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyl transferase. EMBO J. 2006; 25:4503–4512. [PubMed: 16990798]
39. Xu Y, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tetl hydroxylase in mouse embryonic stem cells. Mol. Cell. 2011; 42:451–464. [PubMed: 21514197]
40. Zhang H, et al. TET 1 is a DNA-binding protein that modulates DNA methylation and gene transcription via hydroxylation of 5-methylcytosine. Cell Res. 2010; 20:1390–1393. [PubMed: 21079648]
41. Ko M, et al. Modulation of TET2 expression and 5-methylcytosine oxidation by the CXXC domain protein IDAX. Nature. 2013; 497:122–126. [PubMed: 23563267] Describes the role of IDAX in regulating TET2
42. Xu Y, et al. Tet3 CXXC domain and dioxygenase activity cooperatively regulate key genes for *Xenopus* eye and neural development. Cell. 2012; 151:1200–1213. [PubMed: 23217707]
43. Hino S, et al. Inhibition of the Wnt signaling pathway by Idax, a novel Dvl-binding protein. Mol. Ceil. Biol. 2001; 21:330–342.
44. Yu B, et al. Crystal structures of catalytic complexes of the oxidative DNA/RNA repair enzyme AlkB. Nature. 2006; 439:879–884. [PubMed: 16482161]
45. Iyer LM, Abhiman S, de Souza RF, Aravind L. Origin and evolution of peptide-modifying dioxygenases and identification of the wybutosine hydroxylase/hydroperoxidase. Nucleic Acids Res. 2010; 38:5261–5279. [PubMed: 20423905]
46. Bird A. The dinucleotide CG as a genomic signalling module. J. Mol. Biol. 2011; 409:47–53. [PubMed: 21295585]
47. Bostick M, et al. UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science. 2007; 317:1760–1764. [PubMed: 17673620]
48. Sharif J, et al. The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmtl to methylated DNA. Nature. 2007; 450:908–912. [PubMed: 17994007]
49. Hashimoto H, et al. Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. Nucleic Acids Res. 2012; 40:4841–4849. [PubMed: 22362737]
50. Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res. 2007; 67:946–950. [PubMed: 17283125]
51. Kubosaki A, et al. CpG site-specific alteration of hydroxymethylcytosine to methylcytosine beyond DNA replication. Biochem. Biophys. Res. Commn. 2012; 426:141–147.
52. Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET 1 promotes active DNA demethylation in the adult brain. Cell. 2011; 145:423–434. [PubMed: 21496894]
53. Zhang L, et al. Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nature Chem. Biol. 2012; 8:328–330. [PubMed: 22327402] Together with references 6 and 30, shows that TDG excises 5fC and 5caC. This paper also shows a crystal structure of TDG with 5caC
54. Nabel CS, et al. AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. Nature Chem. Biol. 2012; 8:751–758. [PubMed: 22772155]
55. Song CX, et al. Genome-wide profiling of 5-formylcytosine reveals its roles in epigenetic priming. Cell. 2013; 153:678–691. [PubMed: 23602153]
56. Shen L, et al. Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. Cell. 2013; 153:692–706. [PubMed: 23602152]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

57. Rai K, et al. DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and gadd45. Cell. 2008; 135:1201–1212. [PubMed: 19109892]
58. Rai K, et al. DNA demethylase activity maintains intestinal cells in an undifferentiated state following loss of APC. Cell. 2010; 142:930–942. [PubMed: 20850014]
59. Popp C, et al. Genome-wide erasure of DNA methylation in mouse primordial germ cells is affected by AID deficiency. Nature. 2010; 463:1101–1105. [PubMed: 20098412]
60. Bhutani N, et al. Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature. 2010; 463:1042–1047. [PubMed: 20027182]
61. Cortellino S, et al. Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell. 2011; 146:67–79. [PubMed: 21722948]
62. Cortazar D, et al. Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature. 2011; 470:419–423. [PubMed: 21278727]
63. Bransteitter R, Pham P, Scharff MD, Goodman MF. Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. Proc. Natl Acad Sci. USA. 2003; 100:4102–4107. [PubMed: 12651944]
64. Rangam G, Schmitz KM, Cobb AJ, Petersen-Mahrt SK. AID enzymatic activity is inversely proportional to the size of cytosine C5 orbital cloud. PloS one. 2012; 7:e43279. [PubMed: 22916236]
65. Liutkeviciute Z, Lukinavicius G, Masevicius V, Daujotyte D, Klimasauskas S. Cytosine-5-methyltransferases add aldehydes to DNA. Nature Chem. Biol. 2009; 5:400–402. [PubMed: 19430486]
66. Chen CC, Wang KY, Shen CK. The mammalian de novo DNA. methyltransferases, DNMT3A and DNMT3B are also DNA 5-hydroxymethylcytosine dehydroxy me myiases. J. Biol. Chem. 2012; 287:33116–33121. [PubMed: 22898819]
67. Metivier R, et al. Cyclical DNA methylation of a transcriptionally active promoter. Nature. 2008; 452:45–50. [PubMed: 18322525]
68. Kangaspeska S, et al. Transient cyclical methylation of promoter DNA. Nature. 2008; 452:112–115. [PubMed: 18322535]
69. Hsieh CL. Evidence that protein binding specifies sites of DNA demethylation. Mol. Cell. Biol. 1999; 19:46–56. [PubMed: 9858530]
70. Lin IG, Tomzynski TJ, Ou Q, Hsieh CL. Modulation of DNA binding protein affinity directly affects target site demethylation. Mol. Cell. Biol. 2000; 20:2343–2349. [PubMed: 10713158]
71. Brandeis M, et al. Sp 1 elements protect a CpG island from *de novo* methylation. Nature. 1994; 371:435–438. [PubMed: 8090226]
72. Stadler MB, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011; 480:490–495. [PubMed: 22170606]
73. Szulwach KE, et al. Integrating 5-Hydroxymethylcytosine into the epigenomic landscape of human embryonic stem cells. PLoS Genet. 2011; 7:el002154.
74. Wu H, Zhang Y. Tetl and 5-hydroxymethylation: a genome-wide view in mouse embryonic stem cells. Cell Cycle. 2011; 10:2428–2436. [PubMed: 21750410]
75. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 2011; 12:R54. [PubMed: 21689397]
76. Pastor WA, et al. Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells. Mature. 2011; 473:394–397.
77. Hahn MA, et al. Dynamics of 5-hydroxymethylcytosine and chromatin marks in mammalian neurogenesis. Cell Rep. 2013; 3:291–300. [PubMed: 23403289]
78. Szulwach KE, et al. 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nature Neurosci. 2011; 14:1607–1616. [PubMed: 22037496]
79. Weber M, et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. Nature Genet. 2007; 39:457–466. [PubMed: 17334365]
80. Hackett JA, et al. Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine. Science. 2013; 339:448–452. [PubMed: 23223451]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

81. Thomson JP, et al. Non-genotoxic carcinogen exposure induces defined changes in the 5-hydroxymethylome. Genome Biol. 2012; 13:R93. [PubMed: 23034186]
82. Lian CG, et al. Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell. 2012; 150:1135–1146. [PubMed: 22980977]
83. Raiber EA, et al. Genome-wide distribution of 5-formylcytosine in ES cells is associated with transcription and depends on thymine DNA glycosylase. Genome Biol. 2012; 13:R69. [PubMed: 22902005]
84. Wu H, et al. Genome-wide analysis of 5-hydroxymethylcytosine distribution reveals its dual function in transcriptional regulation in mouse embryonic stem cells. Genes Dev. 2011; 25:679–684. [PubMed: 21460036]
85. Mellen M, Ayata P, Dewell S, Kriaucionis S, Heintz N. MeCP2 binds to 5hmC enriched within active genes and accessible chromatin in the nervous system. Cell. 2012; 151:1417–1430. [PubMed: 23260135]
86. Lister R, et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature. 2009; 462:315–322. [PubMed: 19829295]
87. Straussman, Rv, et al. Developmental programming of CpG island methylation profiles in the human genome. Nature Struct. Mol. Biol. 2009; 16:564–571. [PubMed: 19377480]
88. Weber M, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature Genet. 2005; 37:853–862. [PubMed: 16007088]
89. Yu M, et al. Base-re solution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell. 2012; 149:1368–1380. [PubMed: 22608086] First genome-wide map of 5hmC at single-molecule resolution
90. Hsieh CL. Dependence of transcriptional repression on CpG methylation density. Mol. Cell. Biol. 1994; 14:5487–5494. [PubMed: 7518564]
91. Serandour AA, et al. Dynamic hydroxymethylation of deoxyribonucleic acid marks differentiation-associated enhancers. Nucleic Acids Res. 2012; 40:8255–8265. [PubMed: 22730288]
92. Wu H, et al. Dual functions of Tet 1 in transcriptional regulation in mouse embryonic stem cells. Nature. 2011; 473:389–393. [PubMed: 21451524]
93. Williams K, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011; 473:343–348. [PubMed: 21490601] Describes the distribution of TET1 and 5hmC in ES cells and demonstrates, using several methods, that TET1 recruits SIN3A to target genes
94. Deplus R, et al. TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET 1 /COMPASS. EMBO J. 2013; 32:645–655. [PubMed: 23353889]
95. Chen Q, Chen Y, Bian C, Fujiki R, Yu X. TET2 promotes histone O-GlcNAcylation during gene transcription. Nature. 2013; 493:561–564. [PubMed: 23222540]
96. Yildirim O, et al. Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell. 2011; 147:1498–1510. [PubMed: 22196727]
97. Vella P, et al. Tet proteins connect the *O*-linked *N*-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. Mol. Cell. 2013; 49:645–656. [PubMed: 23352454]
98. Jackson-Grusby L, et al. Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nature Genet. 2001; 27:31–39. [PubMed: 11137995]
99. Fouse SD, et al. Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation. Cell Stem Cell. 2008; 2:160–169. [PubMed: 18371437]
100. Balasubramani A, Rao A. *O*-GlcNAcylation and 5-methylcytosine oxidation: an unexpected association between OGT and TETs. Mol. Cell. 2013; 49:618–619. [PubMed: 23438858]
101. Hart GW, Slawson C, Ramirez-Correa G, Lagerlof O. Cross talk between *O*-GlcNAcylation and phosphorylation: roles in signaling transcription, and chronic disease. Annu. Rev. Biochem. 2011; 80:825–858. [PubMed: 21391816]
102. Fujiki R, et al. GlcNAcylation of histone H2B facilitates its monoubiquitination. Nature. 2011; 480:557–560. [PubMed: 22121020]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

103. Mayer W, Niveleau A, Walter J, Fundele R, Haaf T. Demethylation of the zygotic paternal genome. Nature. 2000; 403:501–502. [PubMed: 10676950]

104. Oswald J, et al. Active demethylation of the paternal genome in the mouse zygote. Curr. Biol. 2000; 10:475–478. [PubMed: 10801417]

105. Santos F, Hendrich B, Reik W, Dean W. Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev. Biol. 2002; 241:172–182. [PubMed: 11784103]

106. Huang Y, et al. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. PLoS ONES. 2010:e8888.

107. Nestor C, Ruzov A, Meehan R, Dunican D. Enzymatic approaches and bisulfite sequencing cannot distinguish between 5-methylcytosine and 5-hydroxymethylcytosine in DNA. Biotechniques. 2010; 48:317–319. [PubMed: 20569209]

108. Jin SG, Kadam S, Pfeifer GP. Examination of the specificity of DNA methylation profiling techniques towards 5-methylcytosine and 5-hydroxymethylcytosine. Nucleic Acids Res. 2010; 38:el25.

109. Okada Y, Yamagata K, Hong K, Wakayama T, Zhang Y. A role for the elongator complex in zygotic paternal genome demethylation. Nature. 2010; 463:554–558. [PubMed: 20054296]

110. Wossidlo M, et al. Dynamic link of DNA demethylation, DNA strand breaks and repair in mouse zygotes. EMBO J. 2010; 29:1877–1888. [PubMed: 20442707]

111. Iqbal K, Jin SG, Pfeifer GP, Szabo PE. Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine. Proc. Natl Acad. Sci. USA. 2011; 108:3642–3647. [PubMed: 21321204]

112. Wossidlo M, et al. 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. Nature Commun. 2011; 2:241. [PubMed: 21407207]

113. Gu TP, et al. The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature. 2011; 477:606–610. [PubMed: 21892189] Together with reference 112, proves that TET3 is responsible for mass methylcytosine oxidation in the male pronucleus. Reference 113 also establishes the phenotype of TET3-deficient oocytes

114. 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. 2011; 21:1670–1676. [PubMed: 22124233]

115. Nakamura T, et al. PGC7 binds histone H3K9me2 to protect against conversion of 5mC to 5hmC in early embryos. Nature. 2012; 486:415–419. [PubMed: 22722204] Shows that the maternal pronucleus of the zygote and select loci on the paternal pronucleus are protected from TET3-mediated oxidation by DPPA3

116. Nakamura T, et al. PGC7/Stella protects against DNA demethylation in early embryogenesis. Nature Cell Biol. 2007; 9:64–71. [PubMed: 17143267]

117. Bortvin A, Goodheart M, Liao M, Page DC. Dppa3/ Pgc7 / stella is a maternal factor and is not required for germ cell specification in mice. BMC Dev. Biol. 2004; 4:2. [PubMed: 15018652]

118. Payer B, et al. Stella is a maternal effect gene required for normal early development in mice. Curr. Biol. 2003; 13:2110–2117. [PubMed: 14654002]

119. Inoue A, Zhang Y. Replication-dependent loss of 5-hydoxymethylcytosine in mouse preimplantation embryos. Science. 2011; 334:194. [PubMed: 21940858]

120. Hajkova P, et al. Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science. 2010; 329:78–82. [PubMed: 20595612]

121. Aoki F, Worrad DM, Schultz RM. Regulation of transcriptional activity during the first and second cell cycles in the preimplantation mouse embryo. Dev. Biol. 1997; 181:296–307. [PubMed: 9013938]

122. Inoue A, Matoba S, Zhang Y. Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation. Cell Res. 2012; 22:1640–1649. [PubMed: 23184059]

123. Dawlaty MM, et al. Tet 1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell. 2011; 9:166–175. [PubMed: 21816367]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

124. Dawlaty MM, et al. Combined deficiency of Tet 1 and Tet2 causes epigenetic abnormalities but is compatible with postnatal development. Dev. Cell. 2013; 24:310–323. [PubMed: 23352810] Reports the phenotype of *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ mice and cells and demonstrates defective imprinting in these mice
125. Koh KP, et al. Tetl and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Ceil Stem Cell. 2011; 8:200–213.
126. Ficz G, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 2011; 473:398–402. [PubMed: 21460836]
127. Vincent JJ, et al. Stage-specific roles for Tetl and Tet2 in DNA demethylation in primordial germ cells. Cell stem cell. 2013; 12:470–478. [PubMed: 23415914]
128. Hackett JA, Zylicz JJ, Surani MA. Parallel mechanisms of epigenetic reprogramming in the germline. Trends Genet. 2012; 28:164–174. [PubMed: 22386917]
129. Yamaguchi S, et al. Tetl controlsmeiosis by regulating meiotic gene expression. Nature. 2012; 492:443–447. [PubMed: 23151479] Demonstrates that TET1 positively regulates gene expression in PGCs and that TET 1 deficiency impairs meiosis
130. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006; 126:663–676. [PubMed: 16904174]
131. Cox JL, Rizzino A. Induced pluripotent stem cells: what lies beyond the paradigm shift. Exp. Biol. Med. 2010; 235:148–158.
132. Mikkelsen TS, et al. Dissecting direct reprogramming through integrative genomic analysis. Nature. 2008; 454:49–55. [PubMed: 18509334]
133. Doege CA, et al. Early-stage epigenetic modification during somatic cell reprogramming by Parp 1 and Tet2. Nature. 2012; 488:652–655. [PubMed: 22902501]
134. Costa Y, et al. NANOG-dependent function of TET 1 and TET2 in establishment of pluripotency. Nature. 2013; 495:370–374. [PubMed: 23395962]
135. Gao Y, et al. Replacement of Oct4 by Tetl during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell. 2013; 12:453–469. [PubMed: 23499384]
136. Spruijt CG, et al. Dynamic readers for 5-(hydroxy) methylcytosine and its oxidized derivatives. Cell. 2013; 152:1146–1159. [PubMed: 23434322]
137. Otani J, et al. Structural basis of the versatile DNA recognition ability of the methyl-CpG binding domain of methyl-CpG binding domain protein 4. J. Biol. Chem. 2013; 288:6351–6362. [PubMed: 23316048]
138. Tini M, et al. Association of CBP/p300 acetylase and thymine DNA glycosylase links DNA repair and transcription. Mol. Cell. 2002; 9:265–277. [PubMed: 11864601]
139. Cortazar D, Kunz C, Saito Y, Steinacher R, Schar P. The enigmatic thymine DNA glycosylase. DNA Repair. 2007; 6:489–504. [PubMed: 17116428]
140. Kellinger MW, et al. 5-formylcytosine 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. Nature Struct. Mol. Biol. 2012; 19:831–833. [PubMed: 22820989]
141. Jin SG, Wu X, Li AX, Pfeifer GP. Genomic mapping of 5-hydroxymethylcytosine in the human brain. Nucleic Acids Res. 2011; 39:5015–5024. [PubMed: 21378125]
142. Matarese F, Carrillo-de Santa Pau E, Stunnenberg HG. 5-Hydroxymethylcytosine: a new kid on the epigenetic block? Mol. Systems Biol. 2011; 7:562.
143. Hayatsu H, Shiragami M. Reaction of bisulfite with the 5-hydroxymethyl group in pyrimidines and in phage DNAs. Biochemistry. 1979; 18:632–637. [PubMed: 420806]
144. Huang Y, Pastor WA, Zepeda-Martinez JA, Rao A. The anti-CMS technique for genome-wide mapping of 5-hydroxymethylcytosine. Nature Protoc. 2012; 7:1897–1908. [PubMed: 23018193]
145. Pastor WA, Huang Y, Henderson HR, Agarwal S, Rao A. The GLIB technique for genome-wide mapping of 5-hydroxymethylcytosine. Nature Protoc. 2012; 7:1909–1917. [PubMed: 23018194]
146. Robertson AB, et al. A novel method for the efficient and selective identification of 5-hydroxymethylcytosine in genomic DNA. Nucleic Acids Res. 2011; 39:e55. [PubMed: 21300643]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

147. Song CX, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nature Biotechnol. 2011; 29:68–72. [PubMed: 21151123]

148. Zilberman D, Henikoff S. Genome-wide analysis of DNA methylation patterns. Development. 2007; 134:3959–3965. [PubMed: 17928417]

149. Krueger F, Kreck B, Franke A, Andrews SR. DNA methylome analysis using short bisulfite sequencing *date* . Nature Methods. 2012; 9:145–151. [PubMed: 22290186]

150. Rein I, DePamphilis ML, Zorbas H. Identifying 5-methylcytosine and related modifications in DNA genomes. Nucleic Acids Res. 1998; 26:2255–2264. [PubMed: 9580672]

151. Booth MJ, et al. Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science. 2012; 336:934–937. [PubMed: 22539555]

152. Song CX, et al. Sensitive and specific single-molecule sequencing of 5-hydroxymethylcytosine. Nature Methods. 2012; 9:75–77. [PubMed: 22101853]

153. Clark TA, et al. Enhanced 5-methylcytosine detection in single-molecule, real-time sequencing via Tetl oxidation. BMC Biol. 2013; 11:4. [PubMed: 23339471]

154. Quivoron C, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell. 2011; 20:25–38. [PubMed: 21723201]

155. Moran-Crusio K, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell. 2011; 20:11–24. [PubMed: 21723200]

156. Ko M, et al. Ten-eleven-translocation 2 (TET2] negatively regulates homeostasis differentiation of hematopoietic stem cells in mice. Proc. Natl Acad. Sci. USA. 2011; 108:14566–14571. [PubMed: 21873190]

157. Li Z, et al. Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood. 2011; 118:4509–4518. [PubMed: 21803851]

158. Shide K, et al. TET2 is essential for survival and hematopoietic stem cell homeostasis. Leukemia. 2012; 26:2216–2223. [PubMed: 22469782]

159. Ng HH, et al. MBD2 is a transcriptional repressor belonging to the MeCP 1 histone deacetylase complex. Nature Genet. 1999; 23:58–61. [PubMed: 10471499]

160. Kass SU, Landsberger N, Wolffe AP. DNA methylation directs a time-dependent repression of transcription initiation. Curr. Biol. 1997; 7:157–165. [PubMed: 9395433]

161. Ng HH, Jeppesen P, Bird A. Active repression of methylated genes by the chromosomal protein MBD1. Mol. Cell. Biol. 2000; 20:1394–1406. [PubMed: 10648624]

162. Fujita N, et al. MCAF mediates MBD1-dependent transcriptional repression. Mol. Cell. Biol. 2003; 23:2834–2843. [PubMed: 12665582]

163. Sarraf SA, Stancheva I. Methyl-CpG binding protein MBD 1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. Mol. Cell. 2004; 15:595–605. [PubMed: 15327775]

164. Prokhortchouk A, et al. The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. Genes Dev. 2001; 15:1613–1618. [PubMed: 11445535]

165. 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. 2004; 32:4100–4108. [PubMed: 15302911]

166. Prokhortchouk A, et al. Kaiso-deficient mice show resistance to intestinal cancer. Mol. Cell. Biol. 2006; 26:199–208. [PubMed: 16354691]

167. Zhao X, et al. Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. Proc. Natl Acad. Sci USA. 2003; 100:6777–6782. [PubMed: 12748381]

168. Hendrich B, Guy J, Ramsahoye B, Wilson VA, Bird A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. Genes Dev. 2001; 15:710–723. [PubMed: 11274056]

169. Jones PL, et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. Nature Genet. 1998; 19:187–191. [PubMed: 9620779]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

170. Ben-Shachar S, Chahrour M, Thaller C, Shaw CA, Zoghbi HY. Mouse models of MeCP2 disorders share gene expression changes in the cerebellum and hypothalamus. Hum. Mol. Genet. 2009; 18:2431–2442. [PubMed: 19369296]
171. Skene PJ, et al. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. Mol. Cell. 2010; 37:457–468. [PubMed: 20188665]
172. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends Biochem. Sci. 2006; 31:89–97. [PubMed: 16403636]
173. Lee JH, Skalnik DG. CpG-binding protein (CXXC finger protein 1) is a component of the mammalian Set1 histone H3-Lys4 methyltransferase complex, the analogue of the yeast Set1/COMPASS complex. J. Biol. Chem. 2005; 280:41725–41731. [PubMed: 16253997]
174. Choy JS, et al. DNA methylation increases nucleosome compaction and rigidity. J. Am Chem. Soc. 2010; 132:1782–1783. [PubMed: 20095602]
175. Watt F, Molloy PL. Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. Genes Dev. 1988; 2:1136–1143. [PubMed: 3192075]
176. Zheng Y, et al. Role of conserved non-coding DNA elements in the Foxp3 gene in regulatory T-cell fate. Nature. 2010; 463:808–812. [PubMed: 20072126]
177. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature. 2000; 405:482–485. [PubMed: 10839546]
178. Miranda TB, Jones PA. DNA methylation: the nuts bolts of repression. J. Cell. Physiol. 2007; 213:384–390. [PubMed: 17708532]
179. Wu H, et al. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. Science. 2010; 329:444–448. [PubMed: 20651149]
180. Brinkman AB, et al. Sequential ChIP–bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. 2012; 22:1128–1138. [PubMed: 22466170]
181. Acosta-Silva C, Branchadell V, Bertran J, Oliva A. Mutual relationship between stacking and hydrogen bonding in DNA. Theoretical study of guanine-cytosine, guanine-5-methylcytosine, and their dimers. J. Phys. Chem. B. 2010; 114:10217–10227. [PubMed: 20684646]
182. Thalhammer A, Hansen AS, El-Sagheer AH, Brown T, Schofield CJ. Hydroxylation of methylated CpG dinucleotides reverses stabilisation of DNA duplexes by cytosine 5-methylation. Chem. Commun. (Camb.). 2011; 47:5325–5327. [PubMed: 21451870]
183. Borgel J, et al. Targets and dynamics of promoter DNA methylation during early mouse development. Nature Genet. 2010; 42:1093–1100. [PubMed: 21057502]
184. Lane N, et al. Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. Genesis. 2003; 35:88–93. [PubMed: 12533790]
185. Gkountela S, et al. The ontogeny of cKIT^{+} human primordial germ cells proves to be a resource for human germ line reprogramming, imprint erasure and *in vitro* differentiation. Nature Cell Biol. 2013; 15:113–122. [PubMed: 23242216]
186. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science. 2001; 293:1089–1093. [PubMed: 11498579]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Box 1|Detecting and mapping modified cytosines by DNA precipitation

5hmC immunoprecipitation

This is the most straightforward method for detecting 5-hydroxymethylcytosine (5hmC) and has been used by many laboratories[39,84,93,126,141]. However, the efficiency of precipitation strongly depends on the density of 5hmC in the samples[26,76]; there are notable variations in the results obtained in different laboratories[39,84,93,126], and there is evidence that the 5hmC-specific antibody precipitates poly-CA repeats, resulting in spurious peaks immediately adjacent to such repeats[142].

CMS immunoprecipitation

When 5hmC is treated with sodium bisulphite, it is converted into a compound called cytosine methylene sulphonate (CMS)[143]. CMS is highly immunogenic, and CMS immuno precipitation recovers 5hmC-containing DNA with high specificity and low background[76,144].

GLIB (glucosylation, periodate oxidation, biotinylation)

Glucose is attached to the hydroxyl group of 5hmC using the enzyme β-glucosyl transferase (BGT) from phage T4. Glucose is then oxidized with sodium periodate and treated with aldehyde reactive probe (ARP) to generate an adduct containing two biotins at the site of every 5hmC[76,145]. The hydroxymethylated DNA is then precipitated with streptavidin beads.

JBP1

5hmC is glucosylated by BGT The glucosylated base is bound by J-binding protein 1 (JBP1), which is conjugated to a bead, allowing for selective precipitation of hydroxymethylated DNA[146]. The precipitation of 5hmC is highly specific, but not very efficient. This method has not yet been used for the genome-wide mapping of 5hmC.

5-hmC selective chemical labelling (hMe-Seal)

BGT adds a glucose-azide conjugate to 5hmC, which can then be biotinylated under gentle conditions using `click chemistry`[73,147].

5fC-DNA pulldown

An ARP molecule biotinylates the formyl group of 5fC[83]. This method was used to generate the first published genome-wide map of 5fC[83], but it also biotinylates abasic sites to some extent[55].

5fC-Seal

BGT is used to protect 5hmC by linking it to an unmodified glucose. 5fC is then reduced to 5hmC using sodium borohydride ($NaBH_4$), and the product then reacts with gluc ose-azide and is biotinylated as in hMe-Seal. Thus, 5fC is selectively biotinylated[55].

5fC and 5caC immunoprecipitation

Antibodies specific for 5fC and 5caC have been used to precipitate these modifications for sequencing, although the efficiency of precipitation is fairly low[56].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Box 2|Mapping 5hmC at base pair resolution

The simplest and most widespread method for distinguishing cytosine from 5-methylcytosine (5mC) — treatment with sodium bisulphite followed by PCR amplification and bisulphite sequencing (BS-seq)[148–150] — conflates the five cytosine species into two groups. 5mC and 5-hydroxymethylcytosine (5hmC) are resistant to deamination by sodium bisulphite and thus are read as cytosine, whereas cytosine, 5-carboxylcytosine (5caC) and 5-formylcytosine (5fC) are deaminated and so are read as thymine[6,89,151] (see the table). Methods have been developed that distinguish 5hmC or 5fC from each other and from other cytosine species at base resolution, although there is no method yet that allows all five bases to be distinguished simultaneously.

Single molecule real time (SMRT) sequencing

Using chemistry similar to 5hmC selective chemical labelling (hMe-Seal) (BOX 1), a bulky conjugate is added to sites of 5hmC. The conjugate stalls a specially designed polymerase during DNA replication, and this stalling is monitored in real time to identify 5hmC[152]. This method, as well as the oxidative bisulphite sequencing (Ox-BS-seq) and TET-assisted bisulphite sequencing (TAB-seq) methods (see below), provide single-base resolution mapping of 5hmC. 5fC and 5caC produce substantial delays in replication even without the addition of a conjugate, although it is hard to distinguish them from each other[153]. SMRT sequencing can handle long DNA fragments (6–10 kb) but does not yet exhibit the throughput required for genome-wide mapping in mammalian cells.

OxBS-seq

5hmC is oxidized to 5fC using potassium perruthenate[151]. Bisulphite sequencing is performed on the same sample both before and after oxidation. As 5hmC is unaffected by sodium bisulphite treatment and 5fC is deaminated, the extent of 5hmC at a site can be approximated by subtracting the frequency of deamination after perruthenate oxidation from the frequency of deamination before oxidation. 5hmC is a rare base, so very high `sequencing coverage` is required to determine the extent of hydroxymethylation.

TAB-seq

5hmC is protected from oxidation by β-glucosyl transferase (3BGT)-mediated glucosylation, and then all 5mC bases are oxidized to 5caC using a recombinant ten-eleven translocation *(TET)* enzyme[89]. After bisulphite treatment, all cytosine species are therefore deaminated, except those that were omC, so these can be read specifically. TAB-seq has revealed that 5hmC is often asymmetric with cytosine or 5mC at CpG sites. Like Ox-BS, this method requires very high sequencing coverage.

5fC chemically assisted bisulphite sequencing (fCAB-seq)

5fC is normally deaminated by sodium bisulphite treatment, but when it reacts with *O*-ethylhydroxylamine, deamination is prevented. The extent of 5fC at a site can be approximated by subtracting the frequency of deamination after the bisulphite treatment of unreacted samples from the frequency of deamination after the bisulphite treatment of *O*-ethylhydroxylamine-treated sample. Although the technique is selective, 5fC is very rare, so extremely high sequencing coverage (>1000 fold) must be used to accurately map the base.

Base	Read by BS-seq	Read by oxBS-seq	Read by TAB-seq	Read by fCAB-seq
Cytosine	T	T	T	T

5mC	C	C	T	C
5hmC	C	T	C	C
5fC	T	T	T	C
5caC	T	T	T	T

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Box 3|Phenotypes of TET-deficient mice

Ten-eleven translocation *(Tet1)* $^{-/-}$ *mice* on a mixed C57BL/6 × 129 inbred mouse strain background are born at Mendelian ratio and appear healthy despite a low birth weight[123]. Mouse *Tet1* gene-trap mutant animals show embryonic lethality on a 129P2/OlaHsd mouse strain background, with noticeable abnormalities by embryonic day E8.5, but they are viable and fertile on a C57BL/6 background[129]. Female *Tet1* gene-trap mutants have smaller litters, similarly to *Tet1*$^{-/-}$ mice. In female *Tet1* gene-trap mutants, which have smaller ovaries, about half of the developing female gametes display a defect in meiotic synapsis and suffer a developmental arrest between E16.5 and E18.5 (REF. 129). Whether *Tet1*$^{-/-}$ mice have a similar meiotic defect is unknown.

Tet2$^{-/-}$ mice on C57BL/6 or mixed backgrounds are born at Mendelian ratios and are fertile, but display clear haematological abnormalities. They have more haematopoietic stem cells (HSCs) than normal mice, and their HSCs have enhanced self-renewal and proliferative potential in culture and in experiments in which progenitor cells are serially transferred in mice[154–157]. At 4 months old, one strain of *Tet2*$^{-/-}$ mice has been reported to develop a condition similar to human chronic myelomonocytic leukaemia (CMML), a malignancy typified by a gross abundance of monocytes[155, 157]. A *Tet2* gene-trap strain shows perinatal lethality[158], but because this phenotype is not observed in several conventional *Tet2* gene-disrupted strains, it is most likely an artefact.

Tet1$^{-/-}$ *Tet2*$^{-/-}$ mice on a mixed C57BL/6 × 129 background are born at near Mendelian ratios and have fully developed organs, but roughly half of these animals die perinatally, often with visible defects in head development[124]. Surviving mice have a smaller birth and adult weight, and females have reduced ovary size and fertility, producing litters of about one fourth the normal litter size. *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ *male* mice crossed with wild-type females produced healthy progeny, but more than half of the progeny of *Tet1*$^{-/-}$ *Tet2*$^{-/-}$ females crossed with wild-type males died perinatally, probably due to an imprinting defect in the *Tet1*$^{-}$ *Tet2*$^{-}$ oocytes (see main text).

Roughly half of the embryos that arise from *Tet3*$^{-}$ oocytes, regardless of the sperm genotype, arrest around E11.5 and do not survive. The *Tet3*$^{-/-}$ mice that survive embryonic development die perinatally for unknown reasons[113].

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Box 4|Effects of hydroxymethylation on methylation-based silencing

Methyl CpG-binding protein 1 (MBD1), MBD2 and Kaiso specifically recognize methylated DNA and recruit histone deacetylases or repressive histone H3 Lys9 methyltransferases (H3K9MTs) to 5mC[159–164]. 5-hydroxy methylcytosine (5hmC) strongly inhibits the binding of these MBD proteins to DNA[49,108,165] and therefore is thought to induce transcriptional activation (see the figure, part **a**). However the role of MBD proteins in regulating gene expression is relatively subtle. Mice deficient in Kaiso[166], MBD1 (REF. 167) and MBD2 (REF 168) develop and grow normally, although MBD1- and MBD2-deficient mice have mild neurological impairments. Methyl CpG binding-protein 2 (MeCP2), an MBD protein with a more complex regulatory role[169–171], may bind 5hmC in addition to 5-methylcytosine (5mC)[49,85]. Various other transcriptional modulators bind 5mC selectively, and whether and how 5mC oxidation perturbs their activity is unclear[136].

DNA methyltransferases (DNMTs), including DNMT1, recruit histone deacetylases (HDACs) and H3K9 methyltransferases to DNA. At promoters, this perpetuates a repressive chromatin state through cell division[172] (see the figure, part **b**). As 5hmC potentially prevents DNMT1 recruitment, hydroxymethylation could block this mechanism of methylation-mediated silencing in dividing cells (FIG. 1c).

5mC can prevent chromatin remodellers that possess CXXC domains from binding to DNA (see the figure, part **c**)[34]. These include the H3K4 methyltransferase mixed-lineage leukaemia (MLL)[38], CXXC1, a component of the SET1 H3K4 methyltransferase complex[173], and Lys-specific demethylase 2A (KDM2A), which removes H3K36 methylation marks[37]. It is unlikely that these proteins could bind oxidized 5mCs.

Methylation increases nucleosome compaction (see the figure, part **d**), and this is likely to reduce the accessibility of adjacent DNA to transcription factors *in vitro*[174]. How hydroxymethylation affects nucleosome compaction is unknown, although 5hmC-rich Purkinje neurons have markedly decondensed nuclei.

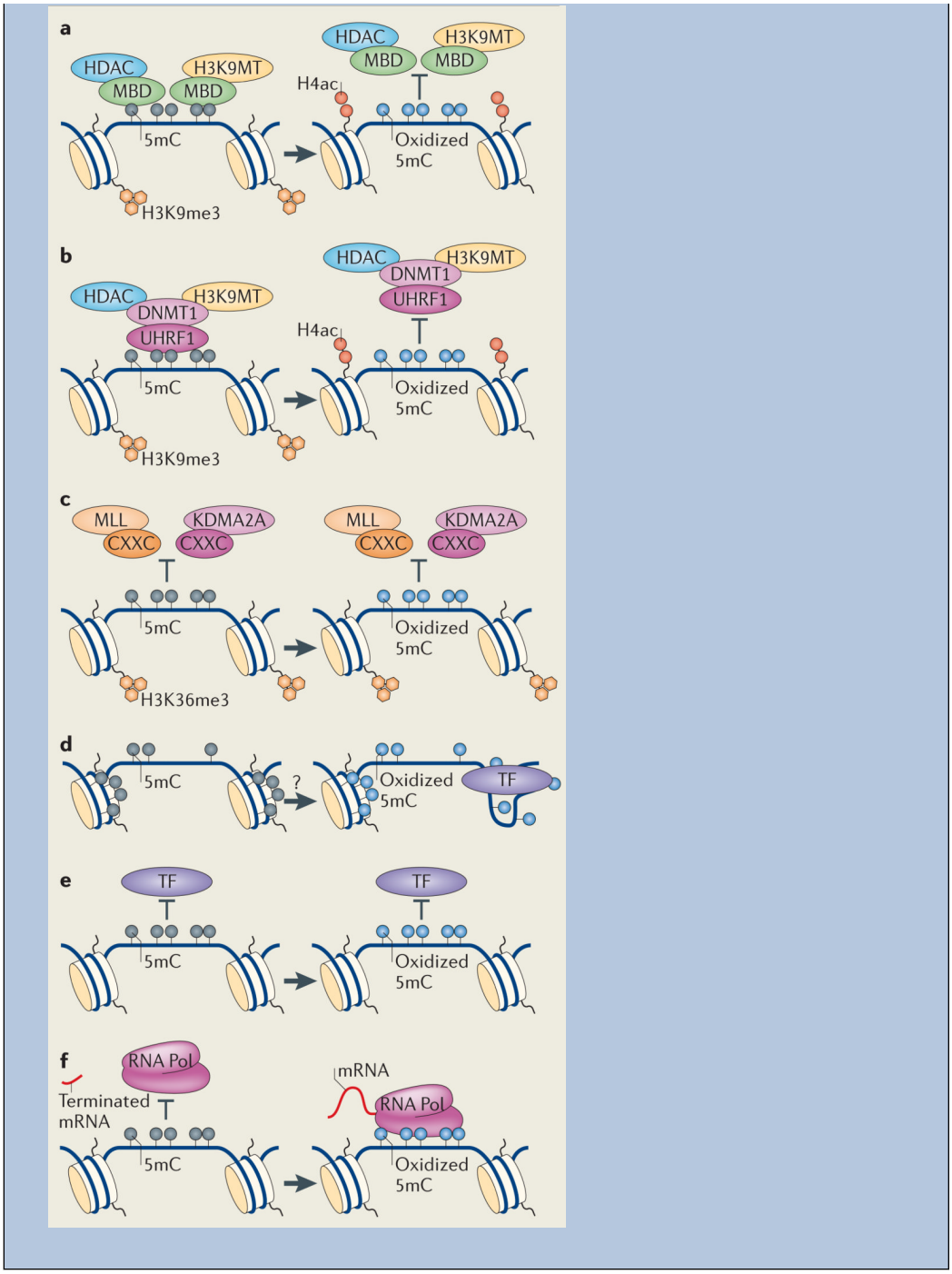
In some instances, cytosine methylation physically blocks the binding of a transcription factor (TF) to its target sequence[175, 176] (see the figure, part **e**). Alternatively, by blocking binding of repressors or insulators, methylation can actually increase the expression of a locus[177–180]. If the binding of a transcription factor is inhibited by methylation, it is unlikely that methylcytosine oxidation would restore binding. Furthermore, as 5hmC is bulkier than 5mC, it may block the binding of proteins that tolerate 5mC.

Finally, quantum mechanical calculations indicate an inherent role for 5mC in stabilizing DNA duplexes by promoting base stacking interactions[181]. Recent experiments confirm that methylated DNA has a higher melting temperature than unmodified DNA, suggesting that methylation may directly inhibit transcriptional initiation or elongation by RNA polymerase (RNA Pol) by preventing necessary melting of the template (see the figure, part **f**)[182]. By contrast, 5hmC reduces the melting temperature of DNA duplexes to promote transcriptional elongation.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript



NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

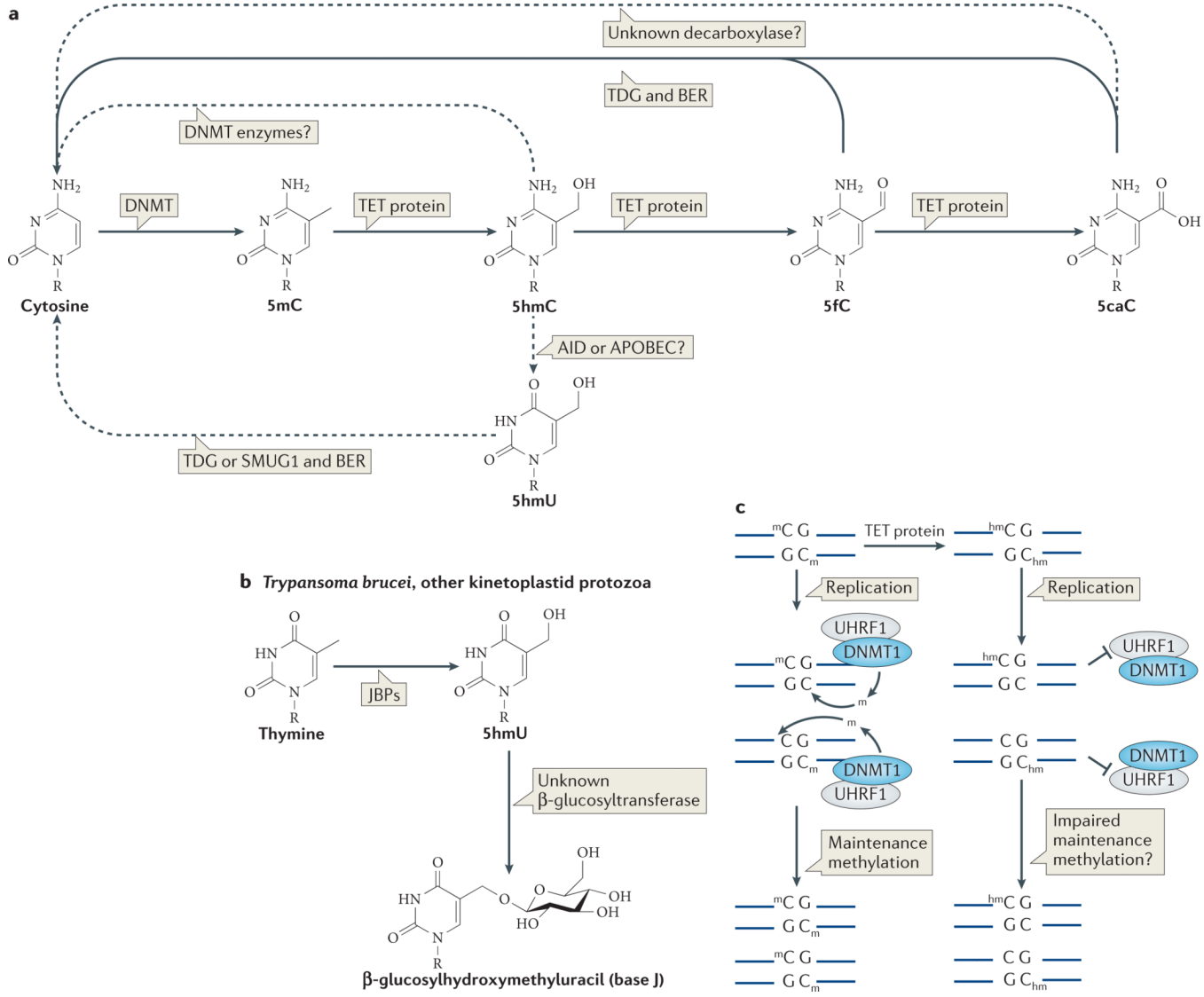


Figure 1. Mechanisms of TET-mediated demethylation

a | Known and putative pathways of DNA demethylation that involve oxidized methylcytosine intermediates. Ten-eleven translocation (TET) proteins sequentially oxidize 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC). 5fC and 5caC can be removed by thymine DNA glycosylase (TDG) and replaced by cytosine via base excision repair (BER), although the extent to which this mechanism operates in specific cell types during development is unknown. Other proposed mechanisms of demethylation are less well established, including decarboxylation of 5caC, DNA methyltransferase (DNMT)-mediated removal of the hydroxymethyl group of 5hmC and deamination of 5hmC (and 5mC) (see main text) by the cytidine deaminases AID (activation-induced cytidine deaminase) and APOBEC (apolipoprotein B mRNA editing enzyme, catalytic polypeptide). AID enzymes deaminate cytosine bases in DNA to yield uracil. AID and the larger family of APOBEC enzymes have been proposed to effect DNA demethylation by deaminating 5mC and 5hmC in DNA to yield thymine and 5hmU, respectively. As these are present in mismatched T:G and 5hmU:G basepairs, they have been proposed to be excised by SMUG1 (single-strand-selective monofunctional uracil DNA glycosylase) or TDG. This mechanism is controversial, however (see main text). **b** |

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

The mechanism of base J (β-D-glucosyl-hydroxymethyluracil) biosynthesis. The thymidine oxidation step mediated by J-binding protein 1 (JBP1) or JBP2, to produce 5-hydroxyuracil (5hmU), is analogous to the 5mC oxidation mediated by TET proteins. JBPs are the founding members of the TET–JBP superfamily: the predicted oxygenase domains of JBP1 and JBP2 were used as the starting point for the sequence profile searches that recovered the homologous domains of the three mammalian TET proteins. **c** | Mechanism by which 5hmC could facilitate replication-dependent DNA demethylation. A symmetrically-methylate d CpG sequence is converted during DNA replication into two asymmetrically methylated DNA strands (left panel). Hemimethylated CpG sites are recognized by UHRFI, the obligate partner of the maintenance DNA methyltransferase DNMT1, which restores symmetrical methylation. TET proteins act at methylated CpG sites to generate symmetrically hydroxymethylated CpG sequences. 5hmC and other oxizided methylcytosines may impair maintenance methylation by inhibiting UHRF1 binding, DNMT1 activity, or both (right panel). As a result, the CpG sequence progressively loses DNA methylation through successive DNA replication cycles.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

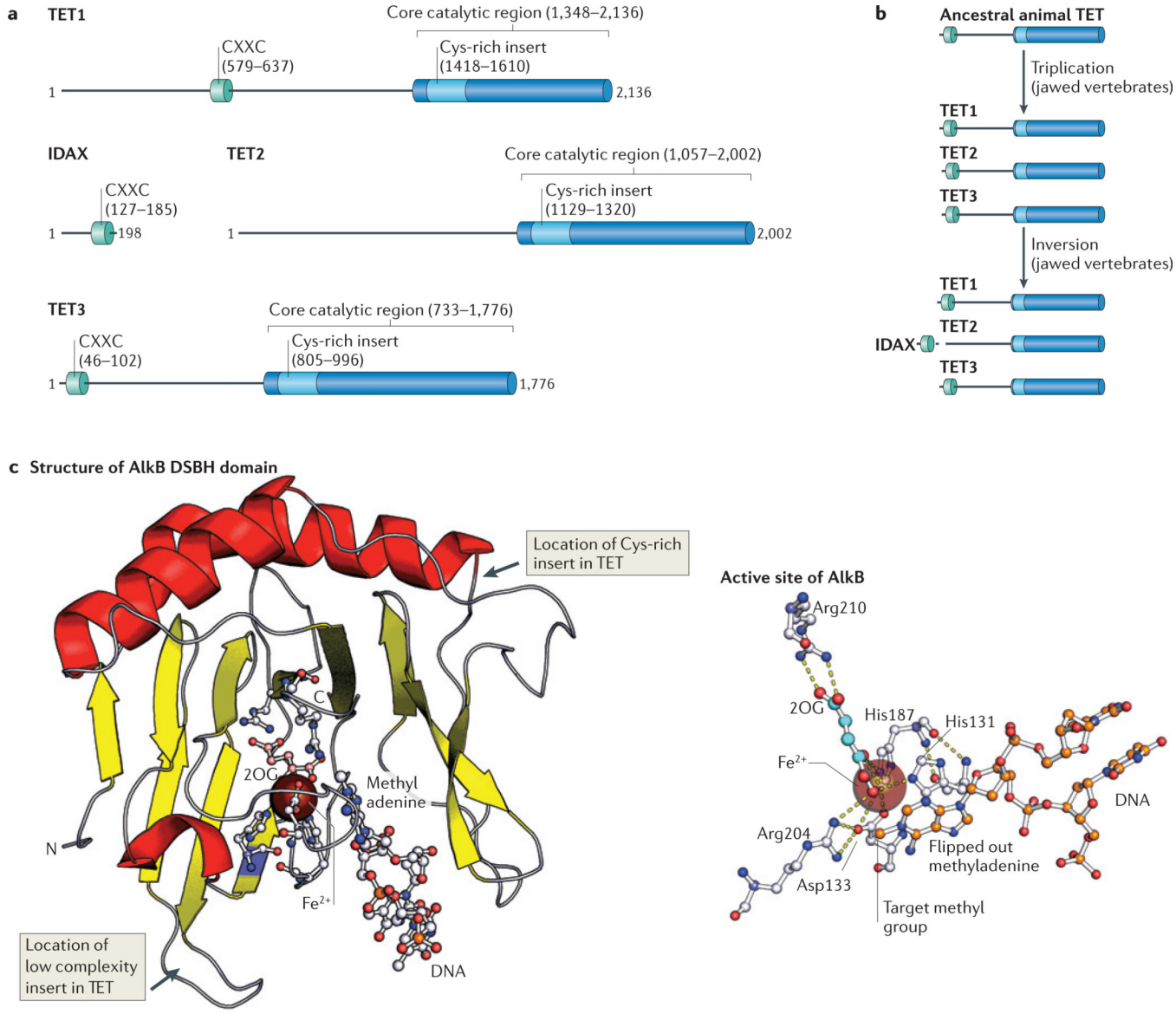


Figure 2. Known protein domains of TET family members

a | Ten-eleven translocation (TET) proteins contain a DNA-binding CXXC domain towards the ami no terminus and a carboxy-terminal catalytic core region that includes a Cys-rich insert and a larger double-stranded β-helix (DSBH) domain. The number of amino acids is indicated, and the numbering corresponds to the human proteins. **b** | Evolutionary changes in the domain structure of TET proteins. A gene triplication event that occurred in jawed vertebrates resulted in the generation of three TET family members. A chromosomal inversion then detached the catalytic domain of TET2 from its CXXC domain, which became a separate gene (which encodes IDAX (inhibition of the Dvl and axin complex)). **c** | A cartoon representation of AlkB (Protein Databank (PDB) identifier: 2FD8), a protein that belongs to the same superfamily as the TET proteins and shares a common fold with them. AlkB is shown as a complex with its substrate methyladenine and its cofactor 2-oxoglutarate (2OG). In the 2OG structure, carbon atoms are shown pink, in the rest of the structure carbons are shown in grey. Nitrogens are shown in blue, oxygens in red and phosphates in orange (left panel). A stripped-down view of the active site of AlkB in complex with its

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

substrate methlyadenine. Note the series of interactions, including pi–pi stacking interactions, between the His residues and the target base, and cation–pi interaction with the active site metal. Such interactions are likely to be retained in the TET– J-binding protein (JBP) family. In the 2OG structure, carbon atoms are shown in cyan and oxygens in red. In the protein structure, carbons are shown in grey, nitrogens in blue and oxygens in red. In methyladenine, carbons are orange and nitrogens are blue. Dashed yellow lines represent hydrogen bonds (right panel).

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

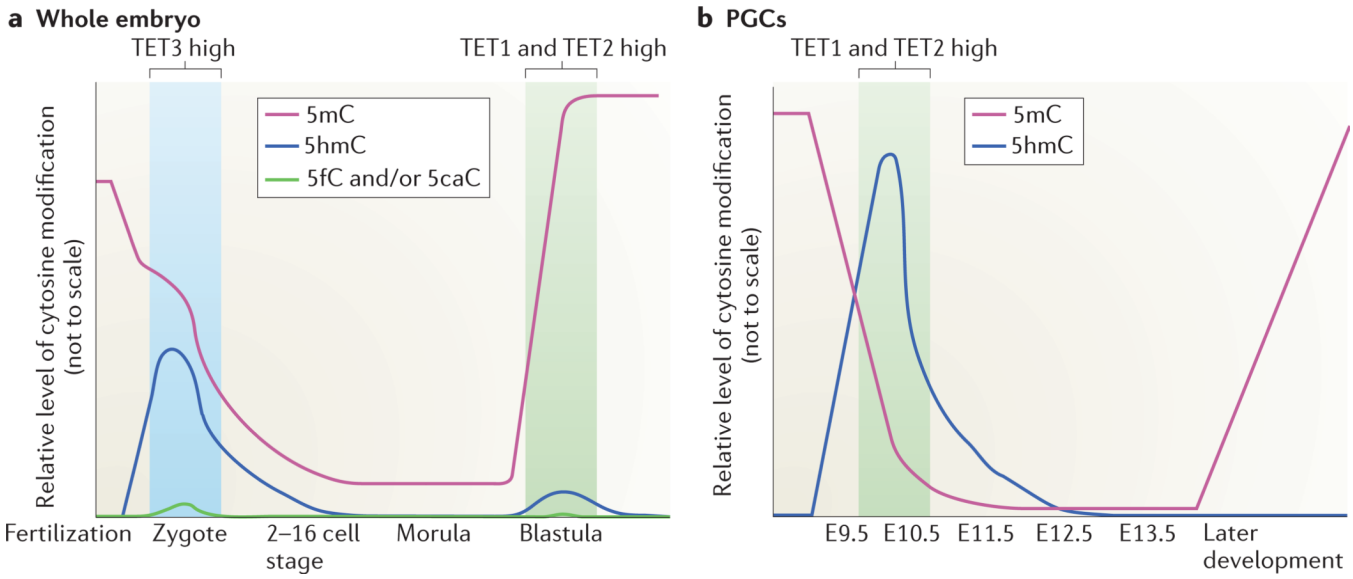


Figure 3. Methylation dynamics in mammalian development

a | Immediately after fertilization, the male pronucleus undergoes mass cytosine oxidation[111–113], mediated by ten-eleventranslocation 3 (TET3). B-methylcytosine (5mC) and oxidized cytosines are then lost from the early embryo in a 'passive' or replication-dependent manner, resulting in the loss of nearly all modified cytosines by the 16-cell stage[103,105]. Imprinted loci retain methylation[183] and some repetitive element classes[184] retain partial methylation. Approximately when the blastula implants into the uterus, the inner cell mass, which gives rise to the embryo, undergoes mass *de novo* DNA methylation[105,183].TET1 and TET2 are highly expressed at this stage, potentially fine-tuning methylation patterns, **b** | Demethylat ion also occurs in primordial germ cells (PGCs) between embryonic days E9.B and E13.B of embryonic development[80,120,128]. This event also entails both mass BmC oxidation by TETl and TET2 and loss of modified cytosine by passive demethylation, resulting in the loss of imprints. A similar process of 5mC oxidation and demethylation occurs more slowly in human germ cells[185]. This demethylat ion of imprints is critical because whereas somatic cells of an organism contain male and female imprints, the germ cells of an organism contain the imprints that correspond exclusively to the gender of the organism. Germ cells are then gradually re-methylated and imprints placed, starting at E1B in males and after birth in females[186].