Target specificity of mammalian DNA methylation and demethylation machinery

M. Ravichandran[a], R. Z. Jurkowska[b,†] and T. P. Jurkowski[c,†]

DNA methylation is an essential epigenetic modification for mammalian embryonic development and biology. DNA methylation pattern across the genome, together with other epigenetic signals, is responsible for the transcriptional profile of the cell and thus preservation of the cell's identity. Equally, the family of TET enzymes which triggers the initiation of the DNA demethylation cycle plays a vital role in the early embryonic development and lack of these enzymes at later stages leads to diseased state and dysregulation of the epigenome. DNA methylation has long been considered a very stable modification, however, it has become increasingly clear that for the establishment and maintenance of the methylation pattern both, generation of DNA methylation and its removal are important, and that a delicate balance of ongoing DNA methylation and demethylation shapes the final epigenetic methylation pattern of the cell. Although this epigenetic mark has been investigated in great detail, it still remains to be fully understood how specific DNA methylation imprints are precisely generated, maintained, read or erased in the genome. Here, we provide a biochemist's view on how both DNA methyltransferases and TET enzymes are recruited to specific genomic loci, and how their chromatin interactions, as well as their intrinsic sequence specificities and molecular mechanisms contribute to the methylation pattern of the cell.

A Introduction

Modification of DNA in CpG dinucleotides plays an important role in mammalian development and has been studied for decades. Yet, despite breakthroughs in high-resolution mapping of the distribution of DNA methylation across mammalian genomes and progress in understanding the targeting and regulation of DNA methyltransferases in cells, it remains not fully known how specific methylation patterns are precisely generated, maintained, read and erased. Similarly, although the contribution of DNA methylation to human diseases, especially cancer, has been clearly demonstrated in numerous studies, the exact molecular mechanisms leading to the aberrant methylation patterns generation are not yet fully elucidated. The recent discovery of TET enzymes showed that DNA demethylation can occur through stepwise oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and finally to 5-carboxylcytosine (5caC,) followed by the removal of the higher oxidized bases by Thymine DNA glycosylase (TDG) and base excision repair mechanism. Genetic studies revealed that TET proteins are involved in numerous biological processes, such as transcriptional regulation, hematopoietic stem cell differentiation, embryonic, and primordial germ cells (PGCs) development, and that these enzymes are commonly deregulated in cancer. While the biological functions of TET enzymes have been studied extensively, very little is known about their biochemical properties. Here, much more work is needed to understand the specificity and catalytic mechanism of TET proteins, as well as the contribution of different domains to enzymes targeting and regulation. Importantly, novel regulatory mechanisms, including allosteric regulation by protein domains and other protein partners or influence of cofactor molecules (like vitamin C) have been described recently for DNMTs and TET enzymes, respectively, providing another layer of complexity in the regulation of these fascinating proteins. In this review, we summarise the most important properties of both DNA methyltransferases and TET enzymes and dissect molecular pathways leading to their recruitment to the target sites.

A Setting DNA methylation

Methylation of mammalian DNA at CpG sites was identified in the beginning of the 1980s[1, 2] followed by the discovery of the first mammalian DNA methyltransferase (MTase), today called DNMT1[3, 4]. The other two enzymes, DNMT3A and DNMT3B were identified in 1998[5]. Surprisingly, a rodent methyltransferase family has recently been extended by another member, DNMT3C[6]. Since their discovery, mammalian DNMT enzymes have been intensively investigated and pivotal genetic, biochemical, structural and functional studies have contributed to the elucidation of enzymatic properties of these interesting enzymes, as well as their targeting mechanisms in the cells.

B Mammalian DNA methyltransferases

DNA methylation is introduced during early stages of mammalian development and during maturation of germ cells by two related DNA methyltransferases, DNMT3A and DNMT3B, with the help of the stimulatory factor DNMT3L[7]. The methyl group is set at the C5 position of the cytosine residues, primarily in the CpG dinucleotides. However, only certain CpG sites are methylated, thereby generating a tissue and cell type-specific patterns of methylation. After their establishment, DNA methylation patterns are preserved, with small tissue-specific changes. During DNA replication, new unmethylated DNA strands are synthesized, leading to the conversion of fully methylated CpG sites into hemimethylated sites that are then re-methylated by a maintenance methyltransferase DNMT1, which works essentially as a molecular copy machine[8, 9]. This elegant inheritance mechanism enables DNA methylation function as a main epigenetic mark mediating long-term transcriptional silencing. In this respect, DNA methylation is involved in silencing of repetitive elements, genomic imprinting, X-chromosome inactivation and regulation of gene expression during development and cellular specialization (reviewed in[10, 11]). Considering its important biological roles, it is not surprising that aberrant DNA methylation changes are associated with human diseases[12-15] and that DNA methyltransferase among other epigenetic factors are attractive therapeutic targets[16, 17].

B Structural organization of DNA methyltransferases

In the general architecture of mammalian DNA methyltransferases, two functional parts can be distinguished, a large N-terminal

regulatory part and a smaller catalytic domain residing at the protein C-terminus (Figure 1)[18-20]. The N-terminal parts of DNMTs, which differ between DNMT1 and DNMT3 enzymes, contain several domains with regulatory and targeting functions. They guide the nuclear localization of the methyltransferases and mediate their interaction with chromatin and other proteins. The C-terminal domains harboring the catalytic centers are required for binding of the cofactor *S*-adenosyl-*L*-methionine (AdoMet) and DNA substrate, and for catalysis. C-terminal domains contain several amino acid motifs conserved among prokaryotic and eukaryotic C5 DNA methyltransferases and fold into a conserved structure called AdoMet-dependent MTase fold, characteristic for all DNA-(cytosine-C5) methyltransferase families. Interestingly, recent structural and biochemical studies revealed that the arrangement of the particular domains in DNMTs that can be influenced by posttranslational modifications and protein partners, plays a critical role in the regulation of enzymes' activity and specificity (reviewed in[21]).

C Domain composition of DNMT1

DNMT1 is a large protein containing several functional domains located in its N-terminal part that is linked to the catalytic part by a flexible linker composed of lysine-glycine (KG) repeats (Figure 1). The DMAP1 (DNA methyltransferase-associated protein 1) interaction domain that is located at the very N-terminus of DNMT1, is involved in the targeting of Dnmt1 to replication foci[22]. It is followed by the PCNA (proliferating cell nuclear antigen) binding domain (PBD) that contributes to the recruitment of DNMT1 to the replication fork during S phase. The interaction with PCNA supports efficient DNA methylation in the cell[23, 24]. The third domain involved in the targeting of DNMT1 to replication foci[25] and to centromeric chromatin[26] is the replication foci-targeting domain (RFTD). This domain is followed by the CXXC domain, which binds unmethylated DNA and might contribute to the specificity of DNMT1[27-29], but its exact role in enzyme's function is still controversial. Finally, the BAH1 and BAH2 (bromo-adjacent homology 1 and 2) domains located at the end of the N-terminal part of DNMT1 are necessary for the folding of the enzyme, but their molecular function awaits elucidation. Despite presence of all conserved catalytic motifs required for catalysis, the isolated catalytic domain of DNMT1 is inactive[30, 31], suggesting that it is controlled by the N-terminal domain of the enzyme. Indeed, structural and biochemical studies demonstrated that various domains in the N-terminal part of DNMT1 surround and contact the catalytic domain[27, 32-34], providing first understanding of the allosteric regulation of DNMT1 (reviewed in[21]).

C Domain composition of DNMT3s family

The DNMT3 family comprises four members: DNMT3A, DNMT3B and DNMT3C (only in rodents), which are enzymatically active, and DNMT3L, which does not possess methyltransferase activity, but works as a stimulatory factor of DNMT3A and DNMT3B. DNMT3C, which has been identified only recently, is a male germline-specific variant that arose from duplication of the DNMT3B gene and is required for retrotransposon methylation during mouse spermatogenesis[6]. Its orthologue has however not been identified in humans. All DNMT3 proteins share considerable sequence similarity (Figure 1). In their N-terminal part, two functional domains are present, the ADD (ATRX-DNMT3-DNMT3L) domain and the PWWP domain, which is absent in DNMT3L and DNMT3C.

The ADD domains of DNMT3 proteins specifically recognise and bind histone H3 tails unmethylated at lysine K4[35-38]. In addition, the ADD domain mediates the interaction of DNMT3A with numerous epigenetic factors and is involved in the allosteric control of DNMT3A activity[38, 39]. The PWWP domains of DNMT3A and DNMT3B are essential for the targeting of the enzymes to pericentromeric chromatin[40, 41] and gene bodies, through specific interaction with histone H3 tails trimethylated at lysine 36[42-44]. The part of DNMT3A and DNMT3B located at the very N-terminus is the most variable region between both enzymes. It binds DNA[45] and is important for anchoring of the enzymes to nucleosomes[44, 46]. However, its specific molecular or biological function still awaits elucidation.

The C-terminal domains of DNMT3A, DNMT3B and DNMT3C share ~80% sequence identity and contain the catalytic centres of the enzymes. They are active in isolated form[47]. In contrast, despite clear homology with the other family members, the C-terminal domain of DNMT3L is catalytically inactive due to several amino acid exchanges and deletions within the conserved DNA-(cytosine C5)-MTase motifs (for general reviews on DNA methyltransferases, cf.[8, 19, 48-50]).

B Catalytic properties of DNMTs: processivity and oligomerisation

All cytosine C5 methyltransferases share similar catalytic mechanism and use base flipping to rotate the target base out of the DNA duplex and insert it in the catalytic pocket (reviewed in[51, 52]). Since DNA is a long polymer, numerous target sites are available for methylation in one substrate molecule. These sites can be methylated without dissociation of the enzyme from the DNA (processive methylation) or with dissociation of the enzyme after each round of binding and methylation (distributive methylation). DNMT1 is a highly processive enzyme, able to methylate long stretches of hemimethylated DNA[53-55]. This property allows efficient methylation of the newly synthetized daughter strand during DNA replication before the chromatin is reassembled. The structure of DNMT1 with substrate DNA revealed that the enzyme enwraps the DNA, enabling sliding of the protein along the substrate and catalysis of successive methylation reactions[32]. Biochemical studies with purified DNMT3A and DNMT3B revealed interesting differences in the mechanism of both enzymes. DNMT3A was shown to methylate DNA in a distributive manner[47, 56] and to cooperatively bind DNA, forming large multimeric protein/DNA filaments[57-60]. Cooperative binding to substrate DNA allows DNMT3A to methylate multiple sites on the same DNA molecule, thereby increasing its activity and efficiency[61]. Indeed, a recent study demonstrated that multimerization of DNMT3A on the DNA also occurs in cells and allows efficient spreading of DNA methylation over a larger region[62]. Although controversial, processive methylation mechanism for DNMT3a has also been reported[63], but is incompatible with cooperative DNA binding. In contrast to DNMT3A, DNMT3B is able to methylate multiple CpG sites by a processive mechanism and in a non-cooperative manner[47, 56]. These observations illustrate that minor amino acid sequence differences in the catalytic domains of DNMT3A and DNMT3B have a profound impact on the catalytic mechanism of the enzymes.

C Intrinsic DNA sequence specificity of DNMTs

In mammals, DNA methylation is predominantly found within CpG dinucleotides. These short palindromic sequences have the advantage that both strands of DNA can be modified and therefore, after DNA replication, the methylation information on the daughter strand can be restored based on the methylation information in the parental strand. Re-methylation of the DNA in each replication cycle is enabled by the strong

preference of DNMT1 towards hemimethylated over unmethylated DNA[29, 31, 32, 55, 64]. Structural studies provided molecular explanation for this preference and revealed that the methyl group of the cytosine is recognised by a hydrophobic pocket in the catalytic domain of DNMT1 and that both the 5mC and the corresponding G in the target DNA strand are recognized accurately[32]. This observation also explains the specificity of DNMT1 towards CpG sites over non-CpG sites (see below).

In vertebrate genomes, cytosine methylation is thought to be largely restricted to CpG sites, for which the inheritance through cell division is well established; however, recent studies revealed the presence of non-CpG methylation in several cell types and tissues, both in mouse and in humans. Still, the molecular function, as well as the mechanisms of its establishment and maintenance are yet unknown.

The original DNA methylation pattern is set by DNMT3A and DNMT3B enzymes, which are classically regarded as *de novo* MTases, as they can methylate DNA regardless of the DNA methylation status at the other DNA strand. Although both enzymes methylate cytosine residues preferentially in the context of CpG dinucleotides, biochemical studies provided evidence that they can also introduce methylation in a non-CpG context with an apparent preference for CA >> CT > CC[65, 66]. In addition, knockout of DNMT3 enzymes in ES cells or ectopic expression of DNMT3A in *Drosophila melanogaster* (which lacks DNA methylation) provided direct evidence that DNMT3 proteins can introduce methylation in non-CpG context also in vivo[67]. Interestingly, the methylation rates of DNMT3 enzymes at non-CpG sites can be as high as half of the rates at CpG sites[65-67] and the non-CpG methylation reaches levels similar to that of mCpG in some human cells[68, 69]. In contrast, the non-CpG methylation rates of DNMT1 are very low[29, 31], indicating that non-CpG methylation cannot be propagated by DNMT1 and would be lost through cellular division in the absence of DNMT3 enzymes. Therefore, non-CpG methylation can serve as a direct imprint of DNMT3 enzyme expression or/and activity in the cells. Consistently, methylated non-CpG sites are widespread in embryonic stem cells, induced pluripotent cells, oocytes and postnatal brain, where DNMT3A and DNMT3B are highly expressed, but absent in most somatic tissues and in cells with low expression of these enzymes[68-72]. Unexpectedly, a recent deep-sequencing survey of 18 human tissues revealed presence of methylation at non-CpG methylation in almost all tissues[73], indicating that DNMT3 activity is widespread and contributes to the overall DNA methylation patterns.

These observations raised important questions regarding the functional relevance of non-CpG methylation. It has been considered a by-product of the hyperactivity and low specificity of DNMT3 enzymes[10, 70]. Depending on the experimental system, there is evidence of its potential role in gene repression[69, 74, 75], but also in gene expression [76]. Most insights about the potential role of non-CpG methylation came from studies on brain (reviewed in[76, 77]), where non-CpG methylation occurs at high levels and contributes to neuronal maturation and specification of brain cells, processes crucial for brain development[68]. First insights into the mechanistic understanding of non-CpG methylation signalling were provided recently with the observation that non-CpG methylation can recruit MeCP2[69, 78, 79], an important epigenetic factor, mutation of which leads to Rett syndrome in humans, and that this process contributes to the repression of long genes in the brain[78]. However, further studies are required to elucidate the exact biological function of the non-CpG methylation.

C Flanking sequence preferences

Although in comparison to prokaryotic methyltransferases DNMT3A and DNMT3B do not seem to have strong sequence specificity beyond CpG sites, both enzymes are very sensitive to the sequences flanking their target sites. DNMT3A prefers purine bases at the 5' end of the CpG sites, whereas pyrimidines are favoured at their 3' end[80-82]. These so called flanking sequence preferences of the methyltransferases might have strong impact on the generation of methylation patterns, as CpG sites embedded unfavourable flanking sequence context cannot be methylated by DNMT3A at all[81]. Interestingly, experimental flanking sequence preferences of DNMT3A and DNMT3B correlate with the statistical data on the methylation level of CpG sites found in the human genome[80, 83], suggesting that the inherent sequence preferences of *de novo* enzymes contribute to the selection of genomic regions that undergo methylation.

Since both DNA strands of a CpG site are embedded in different flanking sequence context, they will vary strongly in the ability to undergo methylation by DNMT3. This leads to the preferential methylation of one strand, generating hemimethylated products. Consistently, the presence of hemimethylated sites was demonstrated in ES cells[84]. As hemimethylated sites are preferred substrates for DNMT1, the cooperation of both enzymes might promote efficient *de novo* methylation of unmethylated DNA[85]. The exact mechanistic understanding of the flanking sequence preferences and specificity of DNMT3 awaits availability of the structure with bound substrate DNA.

B Genome-wide distribution of DNA methylation

Rapid development of the next-generation-based deep sequencing technologies enabled genome-wide interrogation of cytosine methylation at single-base resolution. First comprehensive methylome maps provided invaluable insights into the frequency and genomic distribution of 5mCs, as well as into the interplay between DNA methylation and other epigenetic mechanisms[72, 86, 87]. In addition, global methylation maps of early developmental stages, germ line development and differentiation of progenitor cells shed lights into the dynamics of DNA methylation during global reprogramming and cellular specialization (reviewed in[10, 88-90]).

DNA methylation occurs predominantly in the context of CpG sites and decorates most of the cytosines (70-80%) throughout the entire human genome. However, methylated CpGs are not equally distributed in the genome[10, 91]. Most methylation is found in repetitive sequences, gene bodies and intergenic regions[72, 92, 93]. In turn, CpG islands (CGI), which are regions of higher than expected density of CpG sites, are mostly unmethylated, especially when located in promoters of active genes[94-96]. Conversely, CpG-poor promoters are usually methylated when not active. While CGIs are found throughout the genome, they are often associated with promoter regions; with around 60% of annotated genes having CGI related promoters[97, 98]. Only a fraction of CGI promoters that control imprinted and tissue-specific genes become methylated[10, 99].

Similarly, distal regulatory regions, encompassing enhancers, display reduced levels of DNA methylation when they are active and occupied by transcription factors[91, 100, 101]. The greatest variation in DNA methylation across different cell types is thought to occur in regions located near CpG islands (within 2 kb), termed CpG shores that acquire tissue- and disease-specific methylation changes[102]. Interestingly, a new feature has been added to the human epigenetic landscape with identification of very large regions with low average methylation, called methylation canyons[103] or methylation valleys[104]. These domains include highly conserved, developmentally important genes that might be associated with cancer[103].

B Recruitment of DNMT enzymes

Despite more than two decades of intensive research on the targeting and regulation of DNA methyltransferases in cells, the major question in the field, namely, understanding how specific DNA methylation patterns are established, remains only partially answered. Several synergistic models have been proposed, including both the inherent properties of DNMT enzymes, as well as the contribution of other epigenetic marks and protein partners (Figure 3). The most important ones will be summarised below.

C Recruitment of DNMT3s by chromatin marks

Direct recognition of specific chromatin marks has been proposed as a general mechanism involved in the recruitment of DNA methyltransferases to specific genomic regions. All DNMT3 proteins possess specific domains in their N-termini (ADD domain and PWWP domain) that directly sense the modification state of histone H3 tail in chromatin and could therefore recruit the MTases to the nucleosomes containing unmethylated H3K4 and (or) trimethylated H3K36. Through their ADD domains, DNMT3 proteins interact specifically with H3 tails unmethylated at K4 and this interaction is blocked by the methylation of K4 (H3K4me3)[36, 37, 105, 106]. Interestingly, binding to H3 tails allosterically activates DNMT3A[38, 39] and stimulates methylation of chromatin-bound DNA by DNMT3A *in vitro*[106]. Because methylation of H3 at K4 is associated with active genes, the lack of this modification in specific regions could be interpreted as a signal for their inactivation, whereas its presence would consequently repel DNA methyltransferases. Several genome-wide studies support this hypothesis, as strong inverse correlation of DNA methylation and H3K4me3 modification was observed[83, 86, 94, 107]. This targeting mechanism was provided recently by two elegant studies. Morselli and colleagues showed that the introduction of DNMT3B in yeast cells, which lack DNA methylation, leads to the generation of methylation in regions devoid of H3K4me3[108]. Finally, engineering of the ADD domain of DNMT3A led to aberrant DNA methylation patterns in cells[105], directly demonstrating the crucial role of this domain in enzyme targeting.

Specific recognition of H3 tails tri-methylated at K36 (H3K36me3) is mediated by the PWWP domains of DNMT3A and DNMT3B. As in the case of ADD domain-H3 interaction, several lines of evidence support a model, in which PWWP-H3K36me3 interaction might directly contribute to the recruitment of DNMT3 enzymes to specific genomic regions, including gene bodies and pericentromeric chromatin. Accumulation of H3K36me3 and DNA methylation, and their strong correlation was observed in euchromatin in the body of active genes and at exon-intron boundaries, with exons showing increased levels of both marks[44, 92, 93, 107, 109-113]. Functional role of gene body methylation has been revealed recently by an elegant study, which showed that H3K36me3-dependent intragenic DNA methylation by DNMT3B protects the gene bodies from spurious RNA polymerase II entry and cryptic transcription initiation[42]. Furthermore, a subset of heterochromatic repeats is strongly enriched in H3K36me3[114], which can explain the role of the DNMT3A PWWP domain in the heterochromatic localization of the enzyme[40, 41]. The central role of H3K36me3 recognition in targeting of DNA methylation has also been recently experimentally confirmed in a variety of cellular systems[42, 44, 108]. In addition to H3 binding, the PWWP domains of DNMT3A and DNMT3B interact with DNA[115, 116]. Recently, a model for methylation of nucleosomal DNA by DNMT3A has been proposed[117]. It suggested that the targeting of DNMT3A occurs through a specific binding of H3K36me3 by the PWWP domain, which is followed by an activation of the catalytic domain through the binding of H3 tails unmodified at K4 to the ADD domain, resulting in the methylation of nearby cytosines.

The general picture that emerges from all these observations is that the multivalent interaction of the DNMT3 enzymes with chromatin plays a crucial role in the generation of the genomic DNA methylation pattern.

C Recruitment of DNMT1 to replicating chromatin

Several targeting mechanisms contribute to the proper localization of DNMT1 to replicating DNA. The main ones involve PCNA and UHRF1. PCNA, a component of the replication machinery, interacts and co-localizes with DNMT1 *in vivo*[118], indicating that it might recruit the methyltransferase to the replication fork and load it onto DNA. This interaction contributes to the efficiency of DNA re- methylation, but it is not essential for this process[23]. UHRF1 is an epigenetic factor essential for the maintenance of DNA methylation patterns in mammals, as emphasized by the phenotype of UHRF1 knockout[119, 120]. UHRF1 specifically binds to hemimethylated DNA via its SRA domain[120-123] and recognizes the N-terminal tails of histone H3 di- and tri-methylated at lysine 9 (H3K9me2/me3) via combined binding of its tandem Tudor domain (TTD) and its plant homeodomain (PHD)[124-127]. These specific chromatin interactions of UHRF1 are necessary for the recruitment of DNMT1 to replicating chromatin and DNA methylation maintenance, since mutations preventing histone binding in any of the domains abolished DNA methylation by DNMT1 in cells[125, 127, 128]. In addition to its role in targeting of DNMT1, UHRF1 was also shown to stimulate the catalytic activity of DNMT1 through direct interaction[129, 130].

First evidence for direct binding of histone marks by DNMT1 has been provided with the observation that the methyltransferase preferentially associates with H3 tails ubiquitinated at K18 and K23[131, 132]. This interaction is mediated by the replication foci-targeting domain of DNMT1 and leads to the activation of the enzyme and its recruitment to newly replicated DNA. The ubiquitination of the H3 tail is introduced by the RING domain of UHRF1 and is stimulated by UHRF1 binding to hemimethylated DNA[133]. Ubiquitinated H3 accumulates during S-phase, behind replicating DNA polymerase, leading to DNMT1 recruitment to newly replicated DNA[131-133]. These data indicate

an important additional connection between DNMT1 and UHRF1 chromatin interactions, which is essential for an efficient recruitment of DNMT1 and maintenance methylation.

A Removing DNA methylation

For decades, 5mC was considered as a stable modification, due to the chemical nature of the C-C bond, therefore DNA demethylation was believed to occur through replication-dependent dilution due to the absence or inhibition of the maintenance methylation machinery. This notion changed in the year 2000, when global loss of the methylation mark was detected in mouse zygotes, in a manner independent of DNA replication[134, 135]. During mammalian development DNA demethylation was observed at two stages. The first wave of demethylation occurs during early embryogenesis in the paternal genome, following fertilization and preceding DNA replication and confers totipotency to the developing embryo. DNA methylation pattern is then re-established in the preimplantation stages[134, 135]. The second occurrence is during the germ cell specification that includes demethylation of imprinting genes[136, 137]. Furthermore, active DNA demethylation has also been observed at specific loci in T cells, neurons and other cells[138, 139].

Despite the discovery of biological processes where DNA demethylation occurs in the absence of DNA replication, enzymatic machinery and molecular explanation of the demethylation process was at least controversial and spoiled by lack of reproducibility of the findings by independent laboratories (reviewed in[140]). Initial reports on the involvement of base excision repair (BER) in excising 5mC in chicken embryos[141] and the demonstration of DNA demethylation through direct excision of the methyl group containing base by the DEMETER/ROS1 family of DNA glycosylases in plants had raised the possibility of the presence of similar pathways in mammals (reviewed in[142]). However, the search for an orthologous glycosylase, which could excise the methyl group in mammals, was not fruitful, as the enzymes suspected to have the glycosylase activity on 5mC, such as TDG and MBD4, had a stronger activity on T/G mismatch repair and around a 30-40 fold weaker activity on 5mC *in vitro* than on T:G mismatches[143, 144]. Thus, the enzyme responsible for active DNA demethylation in mammals remained enigmatic for a long time. A major breakthrough came in 2009, when a group of enzymes called Ten-Eleven Translocation (TET) was shown to oxidize 5mC to 5-hydroxymethylcytosine (5hmC) both in vitro and in mES cells, in which 5hmC constitutes about 0.03% of total nucleotides[145]. In parallel, an independent group identified the presence of higher levels of 5hmC (about 0.6% and 0.2% of total nucleotides) in mouse Purkinje neurons and granule cells respectively[146]. These two seminal discoveries uncovered a possible pathway for active DNA demethylation and give rise to a new dynamic field of research.

B Mammalian TET enzymes

Although the gene coding for TET1 (also known as CXXC6) was known to be a fusion partner of MLL involved in acute myeloid leukaemia (AML)[147], its function was not characterized until 2009. The catalytic activity of TET enzymes was first predicted based on the computational search for DNA modifying enzymes using as bait the sequence of the dioxygenase domain of JBP1 and JBP2, which oxidize the methyl group of thymine to 5-hydroxymethyluracil in trypanosomes[148, 149]. The TET enzymes belong to the iron and α-ketoglutarate dependent dioxygenase family (Fe2+/αKG-DO). Bioinformatics searches further revealed that TET enzymes are distributed across the metazoans that have DNA methylation marks and are also present in fungi and algae[148].

The oxidized base 5hmC was first described in 1952, when it was identified in the genomes of T-even bacteriophages (T2 and T4) as a modified base, which gets further glucosylated and provides protection against cleavage by bacterial restriction enzymes[150]. Later, Penn and colleagues demonstrated that 5hmC was also found in adult rats, mice and frogs and that it accounted for ~ 15% of total cytosines[151]. Yet this finding was disregarded by the scientific community, as it could not be reproduced by another group[152]. Thereafter, the formation of 5hmC in mammalian cells was thought to result from oxidative damage until its rediscovery in 2009. Later studies showed that 5hmC is present in different mouse tissues, such as heart, kidney, lung, muscle and the highest level is found in the brain and ES cells[153].

Furthermore, it was demonstrated that similarly like in the thymidine salvage pathway (Smiley et al. 2005), TET enzymes can further oxidize 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)[154, 155] and that these higher oxidized bases (5fC and 5caC) are recognized and excised by the thymine DNA glycosylase (TDG), which trigger the base excision repair pathway (BER) to replace the abasic site by an unmodified cytosine[154], thereby completing the DNA demethylation pathway (Figure 2).

B Structural organization of TET enzymes

The mammalian TET family comprises three members that share similar domain architecture, namely, TET1, TET2 and TET3 (Figure 1). These are large proteins harbouring the catalytic domain at their C-terminus, which is composed of a double-stranded β helix domain (DSBH) characteristic for Fe^{2+}/αKG dioxygenases and a cysteine-rich region preceding the DSBH. In metazoan TETs, the DSBH is interrupted by a large unstructured region, which is less conserved than the DSBH domain and is believed to engage in protein-protein interactions. Both TET1 and TET3 contain a CXXC domain at their amino terminus. The TET2 protein lacks the CXXC domain, which was lost during evolution after gene duplication and inversion, and is now coded as separate protein IDAX (inhibition of the dvl and axin complex)[148, 156].

The core catalytic domain (DSBH) forms the characteristics dioxygenase domain and contains the binding sites for Fe^{2+} and αKG, which are both essential for the catalytic activity. The amino acids that are crucial for Fe^{2+} and αKG binding in the DSBH and Zn^{2+} binding in the Cys-rich domain are conserved among the family members and all the three enzymes are shown to be catalytically active in vitro[145, 148] and in vivo[155].

The crystal structure of the human TET2 catalytic domain and nTet (TET homolog from an amoeboflagellate *Naegleria gruberi*) was solved recently[157, 158]. The structure of the TET2 catalytic domain (PDB ID: 4NM6) in complex with 5mC containing DNA revealed that the catalytic domain containing a DSBH core forms a globular structure, which is stabilized by the

flanking region of DSBH and the Cys-rich region. Unlike the Cys-rich region in other proteins, in TET this region does not form an independent domain, but wraps around the DSBH core stabilizing the overall structure of the enzyme and are crucial for catalysis[158]. DNA is bound above the DSBH core, which is enriched in basic and hydrophobic amino acids. Similar to DNA MTases and DNA repair enzymes[159, 160], TET enzymes utilize a base flipping mechanism to position the target base in the catalytic pocket for the oxidation reaction. Once the methyl group is located in the catalytic pocket, it is oriented towards the catalytic iron and α-KG, which facilitate the catalytic turnover[157, 158].

Analysis of the interaction of human TET2 with DNA indicates that besides the target 5mC within a CpG dinucleotide context, the enzyme does not interact with the bases flanking the CpG site[158]. Intriguingly, the enzyme also does not make contact with the methyl group of the target cytosine, suggesting that this would allow TET2 to generate higher oxidation of 5hmC to 5fC and 5caC[158]. Additionally, TET enzymes have been shown to oxidize the methyl group of thymine (T) to 5-hydroxymethyl uracil (5hmU)[161], however the physiological relevance still needs to be uncovered.

B TET-mediated DNA demethylation pathways

Identification of TET enzymes and their reaction products has paved the way for DNA demethylation through a direct enzymatic action on the methyl groups. Since their discovery, numerous plausible DNA demethylation pathways involving TETs have been investigated, both *in vivo* and *in vitro*. Intriguingly, the formation of 5hmC via TET enzymes was shown to facilitate the passive dilution of modified bases, as DNMT1 is less active on hemi-5hmC containing DNA (> 60-fold) *in vitro*[162, 163]. This contributes to the replication-dependent loss of methylation induced by 5hmC formation. However, this observation has been challenged by recent reports. In vitro studies showed that the DNMT1 interaction partner UHRF1 binds 5hmC[164-166] thereby targeting DNMT1 to hemi-5hmC containing DNA. Secondly, unlike DNMT1, DNMT3A and DNMT3B are not sensitive to hemi-5hmC DNA and can re-methylate hemi-5hmC containing DNA[162, 167]. These observations argue against 5hmC-mediated passive dilution and require further investigation.

The other suggested pathway involves TET-TDG-BER mediated DNA demethylation, where the higher oxidation products 5fC and 5caC generated by TET enzymes are excised by TDG, followed by the generation of an abasic site and lesion repair with an unmodified cytosine by the BER machinery [154, 155]. Accumulated evidence suggested that this is the main DNA demethylation pathway triggered by TET enzymes[168]. This is supported by the observations that the knockdown of TDG in mES cells results in an up to 10-fold increase of 5fC and 5caC[169, 170], whereas the overexpression of TDG leads to the opposite effect with no significant changes in the level of 5mC or 5hmC[171]. Moreover, in vitro investigation of TDG demonstrated that TDG removes 5fC and 5caC in CpG dinucleotide context more efficiently than the mismatch repair T:G[172, 173], suggesting that the main function of TDG is the excision of the oxidized bases over T:G removal. Nevertheless, TDG-BER mediated demethylation cannot account for the genome-wide demethylation, as TDG is not highly expressed in the zygote and loss of TDG does not affect the demethylation in zygote[174]. Furthermore, TDG and the BER mechanism may compromise the genome stability by introducing multiple nicks and single or double strand breaks in the DNA, while processing multiple methylation sites on the DNA molecule. This indicates that either other DNA glycosylases are involved in this process or other TET-independent demethylation mechanisms are responsible for the majority of observed demethylation[175]. Besides BER, nucleotide excision repair (NER) protein GADD45A has also been implicated in DNA demethylation and it was shown to interact with TET[176-178]. These observations suggest that both NER and BER may cooperate with TET enzymes in processing the oxidized base, but further experimental evidence is needed.

Another possible demethylation pathway involves the AID/APOBEC mediated deamination of 5hmC to 5hmU, which is removed by DNA glycosylases such as SMUG, MBD4, NEIL or TDG[170, 179, 180]. Supporting this view, a study conducted in the mouse brain (which has the highest level of 5hmC) reported that TET-mediated formation of 5hmC could be processed further to 5hmU by AID/APOBEC. Moreover, both enzymes work synergistically at the locus specific DNA demethylation of neuronal activity induced genes in mouse dentate gyrus[181]. Alternatively, the capacity of TET enzymes to oxidize thymine to 5hmU was shown to trigger DNA demethylation through DNA glycosylases and BER[161]. However, this pathway of DNA demethylation still remains controversial due to the inconsistencies in the reported results and needs to be clarified in further studies.

Another very elegant and biochemically plausible demethylation pathway that has been proposed involves direct removal of the carboxyl group from 5caC by a putative 5caC decarboxylase. This hypothetical enzyme should work similarly to orotate decarboxylase, as observed in the thymidine salvage pathway[182], which very efficiently catalyzes the loss of carboxyl group linked to C6 of the pyrimidine ring. However, despite intensive effort of multiple research groups, no such decarboxylase specific for 5caC has been identified so far in mammals.

Interestingly, DNA demethylation involving both DNMTs and TETs has also been observed *in vitro*. In the absence of AdoMet, the methyltransferase DNMT1 was shown to remove 5hmC as formaldehyde[183] and a similar result was shown for DNMT3A and DNMT3B, however it required non-physiological concentration of H_2O_2[184]. Moreover, DNMT3A was reported to convert 5caC to C in the absence of AdoMet[185]. It should be noted that these conclusions were drawn based on *in vitro* experiments and need to be validated in vivo. In addition, such activity *in vivo* would require prevention of the AdoMet binding by the methyltransferase, which could occur through a post-translational modification of the enzyme that would block the binding site.

B Recruitment of TET enzymes

Despite a recent progress in understanding the physiological relevance of TET enzymes and their reaction products, there is a rather limited progress in understanding of the mechanisms of recruitment and regulation of TET enzymes. The presence of the CXXC domain on the N-terminus of TET1 and TET3 is believed to be partly responsible for targeting of these enzymes to the CpG containing regions, as the CXXC domain has been shown to recruit DNMT1, MLL1, CFP1 to unmethylated CpG sites[186, 187]. Consistently, DNA binding studies showed that TET1

is able to bind to CpG-rich DNA irrespective of its modification state (C, 5mC or 5hmC)[188, 189], whereas the *Xenopus* TET3-CXXC domain binds unmodified C in both CpG and non-CpG context, with a slightly higher preference for CpG[190, 191]. Another interesting study demonstrated that CXXC domain of TET3 can bind 5caCpG and that TET3-FL preferentially binds to the TSS of genes involved in base excision repair mechanism[191]. This suggests that TET3 may be specifically targeted to these loci through the CXXC domain or by other interacting proteins[191] and oxidize 5mC at these transcriptional start sites (TSSs).

Unlike TET1 and TET3, TET2 lacks the CXXC domain and may depend on other proteins or TFs for locus specific recruitment. Supporting this idea, TET2 has been shown to interact with the transcription factor Wilms tumor (WT) and Early B cell factor 1 (EBF1) which modulate the TET2 activity and target gene expression[192-194].

Likewise, NANOG-dependent recruitment of TET1 and TET2 has also been suggested to promote expression of genes involved in reprogramming and lineage commitment[195]. Furthermore, a study by Perera and colleagues in mouse retinal cells demonstrated that RE1-silencing transcription factor (REST) recruits a TET3 isoform lacking the CXXC domain, which then interact with the histone methyltransferase NSD3 and activates its target genes[196]. TET enzymes were shown to interact with proteins involved in base excision repair pathway such as TDG, PARP1, MBD4, NEIL etc[197]. Furthermore, all three TET enzymes are reported to associate with O-linked ß-D-N-acetylglucosamine (O-GlcNAc) transferse (OGT). It has been suggested that TETs recruit OGT to the chromatin and that TET-OGT interaction promotes the OGT activity[198-200]. In summary, it is increasingly clear that TET enzymes do not function alone, and they interact with multiple other proteins in a contextual manner and through this cooperation modulate gene expression.

B Genome-wide distribution of TETs and their reaction products

Initial quantification of the level of 5hmC has revealed that it is detectable in most of the tissue tested. However, unlike 5mC, level of which is relatively constant across different somatic cells constituting ~4% of total cytosines[201], the level of 5hmC is lower than of 5mC and varies between different tissue types. It is most abundant in ES cells and in brain, where it constitutes between 0.4 and 0.7% of total cytosines, and present at lower levels in other tissues[146, 202-205]. Interestingly, cancer cells often contain lower levels of 5hmC than the surrounding untransformed tissues, which has become a hallmark of certain types of cancers, for example, melanoma[206].

In contrast to 5hmC, initial attempts to quantify the levels of 5fC and 5caC were unsuccessful due to their very low abundance as they are rapidly excised from the DNA. However, development of more sensitive techniques has enabled quantifying the levels of the oxidized bases, showing that they are 10-100 fold less abundant than 5hmC (0.02-0.002%)[207, 208]. Interestingly, despite their low abundance (especially of 5fC), both 5hmC and 5fC could be detected after several cell divisions[208, 209], suggesting that they might be stable marks.

Genome-wide studies showed that in both ES cells and in brain tissues, 5hmC is enriched in euchromatic regions, especially at the transcriptional start sites (TSSs), moderate and low CG content promoters, and gene bodies[189, 210-215]. 5hmC on gene bodies is positively correlated with gene expression in both brain and ES cells[213]. In ES cells, 5hmC is mostly prevalent in developmentally regulated genes marked with bivalency (PRC2 target gene), TF binding regions, active enhancers and CTCF binding sites, but not on housekeeping genes[212, 216, 217], whereas in brain, 5hmC is present in higher levels on poised enhancers primed for activation than on active enhancers[218]. The overall level of 5hmC observed around TSSs and gene promoters in neuronal cells are lower compared to ES cells[210, 213, 219], indicating that 5hmC in somatic cells has cell type specific roles.

Similar to 5hmC, in ES cells 5fC is enriched at CpG island (CGI) containing promoters, TSSs (marked with H3K4me3), exons and gene regulatory elements, especially at the poised enhancers [169, 220, 221]. However, 5fC is enriched at exons and enhancers of actively expressed genes in a tissue type specific fashion[222, 223]. Interestingly, 5fC-enriched sites show increased binding of p300[220] and 5fC containing promoters are positively correlated to gene expression[169], indicating that the 5fC mark can recruit chromatin factors and may exhibit a distinct regulatory function.

Genome-wide mapping of TET1 occupancy showed that there is a significant overlap between the co-occurrence of both 5hmC and TET1 in the genome. In ES cells, TET1 is localised in 60% of bivalent genes and on the promoters of PRC2 occupied genes carrying H3K27me3, but not on the promoters harbouring H3K4me3 alone[211, 224]. TET1 is present on GC-rich promoters with intermediate and low CpG content (like NANOG, ESRRB, TCL1, KLF4[223]), but not with high CpG density [189]. This suggests that TET1 regulates both the developmentally regulated genes and the TFs inducing pluripotency[225]. TET enzymes also co-localise with Sin3A independent of 5hmC[212].

B TET intrinsic DNA sequence specificity

Most of the studies on TET enzymes were focused on elucidating their biological role and their reaction products; however, the intrinsic biochemical properties of TET enzymes that govern their function remain not well understood. In particular, little is known about how TET enzymes choose the target sequence. Do TET enzymes exhibit any flanking sequence preference (in the context of CpG sites)? How specific are they towards CpG sites? Do they show preference for oxidation of different modified base (5mC, 5hmC and 5fC)? How do they catalyse the stepwise oxidation on one site (5mC to 5caC) and how do they oxidize multiple 5mC (also 5hmC or 5fC) substrates on a single DNA strand (from one 5mC/5hmC/5fC substrate to another), in a distributive or in a processive manner?

Although the solved crystal structures of hTET2-CD and nTet in complex with DNA (both 5mC and 5hmC) provided some insight into the behaviour of TET enzymes[157, 158, 226], detailed biochemical evidence is still lacking. This is especially important considering recent reports showing that TET-dependent demethylation in zygotes represents only a small fraction of all demethylation events observed[174, 175] and that TET-associated demethylation seems to be locus specific, suggesting potential more stringent sequence preference than the ubiquitous CpGs. Interestingly, fine mapping of 5hmC genomic locations using SCL-exo protocol showed that 5hmC is not randomly distributed, but rather highly enriched within defined sequence context[227]. It is still unknown what is the molecular reason granting this sequence preference.

C CpG vs non-CpG hydroxymethylation

Both DNMTs and TET enzymes recognize and modify CpG dinucleotides, yet DNMT3 enzymes were shown to efficiently modify non-CpG sites. Surprisingly, little work was contributed to investigate the intrinsic preference of TET enzymes towards non-CpG sites. In the initial report that identified TET enzymes as 5mC hydroxylases, the authors showed that TET enzymes are capable of oxidation of 5mC embedded in a CpG site, yet non-CpG substrates were not tested. Hu and colleagues used three DNA duplexes containing single 5mCpX sites and showed that 5mCpA and 5mCpC sites were poor substrates for TET, with conversion efficiencies of <2% and <5%, respectively, as opposed to >85% for 5mCG sites in the same sequence context[158]. Another report using a different experimental approach came to a similar conclusion[228]. These authors used a library of dsDNA substrates where 5mC was embedded in randomized sequence context, treated DNA *in vitro* with TET1, enriched the converted, 5hmC-containing products with an anti-5hmC specific antibody and analysed with high-throughput sequencing. Similarly as TET2, the TET1 was shown to preferentially oxidize 5mCpG with some incidence of oxidation of 5mCpC sites.

B Processive oxidation on the site and lateral processivity

Processivity of TET enzymes can be regarded in two different means, first "on site processivity" is the serial oxidation of 5mC to 5hmC, 5fC and 5caC on a single CpG site without the enzyme dissociating from the site and second is the consecutive oxidation of numerous CpGs on a single DNA molecule, which could be called "lateral processivity". The group of Yanhui Xu reported that the catalytic domain of human TET2 efficiently oxidizes 5mC to 5hmC, but the further oxidations are inefficient leading to stalling of the reaction at the 5hmC state[226]. However, the same group reported before that the TET2 could convert 5mC all the way to 5caC[158]. Numerous reports from other groups also showed that TET enzymes are capable to efficiently convert 5mC to 5caC[229-231]. Surprisingly, the two studies that investigated the mouse TET2 "on site" processivity came to contradictory conclusions, which could potentially be explained by differences in the reaction conditions and experimental setup[229, 231].

An unexpected discovery that TET3 CXXC preferentially binds 5caCpG led to proposition that this interaction can stimulate processive activity of the enzyme leading to spreading of the 5caC from the first oxidized CpG site. In the proposed model, the first 5mCpG site that is oxidized to 5caCpG gets bound by the CXXC domain of the enzyme, therefore keeping the catalytic domain in close proximity and promoting oxidation of nearby 5mCpGs[191]. It is a very interesting hypothesis, which still requires experimental proof.

C Flanking sequence preference

The non-random distribution of 5hmC as discussed above suggests that TET enzymes can have a more stringent sequence preference and/or be targeted to these locations by other protein factors. In the TET2/DNA co-crystal structure no protein-base specific contacts outside of the CpG site were observed, suggesting that the enzyme has weak or no flanking sequence specificity[158, 226]. Nevertheless, one can note that the bound DNA is strongly bent and distorted, giving the possibility of indirect readout of DNA sequence as observed with numerous other DNA binding proteins, restriction enzymes and bacterial MTases[232, 233]. Whether TET enzymes use indirect readout for sequence recognition still remains to be addressed.

Perspectives

DNA methylation contributes an important mechanism to epigenetic regulation of cellular differentiation. Intensive research of the past 2 decades elucidated the distribution of DNA methylation in human genome, as well as contributed to the understanding of molecular mechanisms involved in DNA methylation pattern establishment and maintenance. Important mechanisms responsible for the recruitment of the DNA methyltransferases to specific genomic regions have been identified, including interaction of DNMTs with modified histone marks, as well contribution of inherent properties of the MTases, like sequence preference, processivity or oligomerisation (Figure 3).

Considering its role in epigenetic inheritance, DNA CpG methylation has traditionally been considered as a very stable mark that could only be lost via its dilution over several replication cycles in the absence of the maintenance methyltransferase activity. This paradigm was challenged by multiple observations of active demethylation events occurring independently from DNA replication. After the discovery of TET enzymes and elucidation of enzymatic activities contributing to further processing of the 5mC intermediates, the methylation-demethylation cycle has been closed. With the follow up genetic and functional studies it become increasingly clear that active DNA demethylation, conferring reversibility of DNA methylation pattern, contributes an important mechanism to the epigenetic regulation. Based on the seminal research of the past decade, novel view emerges, in which a delicate balance of ongoing methylation and demethylation shapes final epigenetic methylation pattern of the cell. Despite the considerable progress in understanding how cell type specific DNA methylation patterns are established and maintained we are still far from understanding a complete picture.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

The publication of this manuscript was supported by DFG SPP1784 (T.P.J).

Notes and references

1. M. Ehrlich and R. Y. Wang, *Science*, 1981, **212**, 1350-1357.
2. A. Razin and A. D. Riggs, *Science*, 1980, **210**, 604-610.
3. T. H. Bestor, *Gene*, 1988, **74**, 9-12.
4. T. H. Bestor and V. M. Ingram, *Proceedings of the National Academy of Sciences of the United States of America*, 1983, **80**, 5559-5563.
5. M. Okano, S. Xie and E. Li, *Nature genetics*, 1998, **19**, 219-220.
6. J. Barau, A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Herault, F. Guillou and D. Bourc'his, *Science*, 2016, **354**, 909-912.

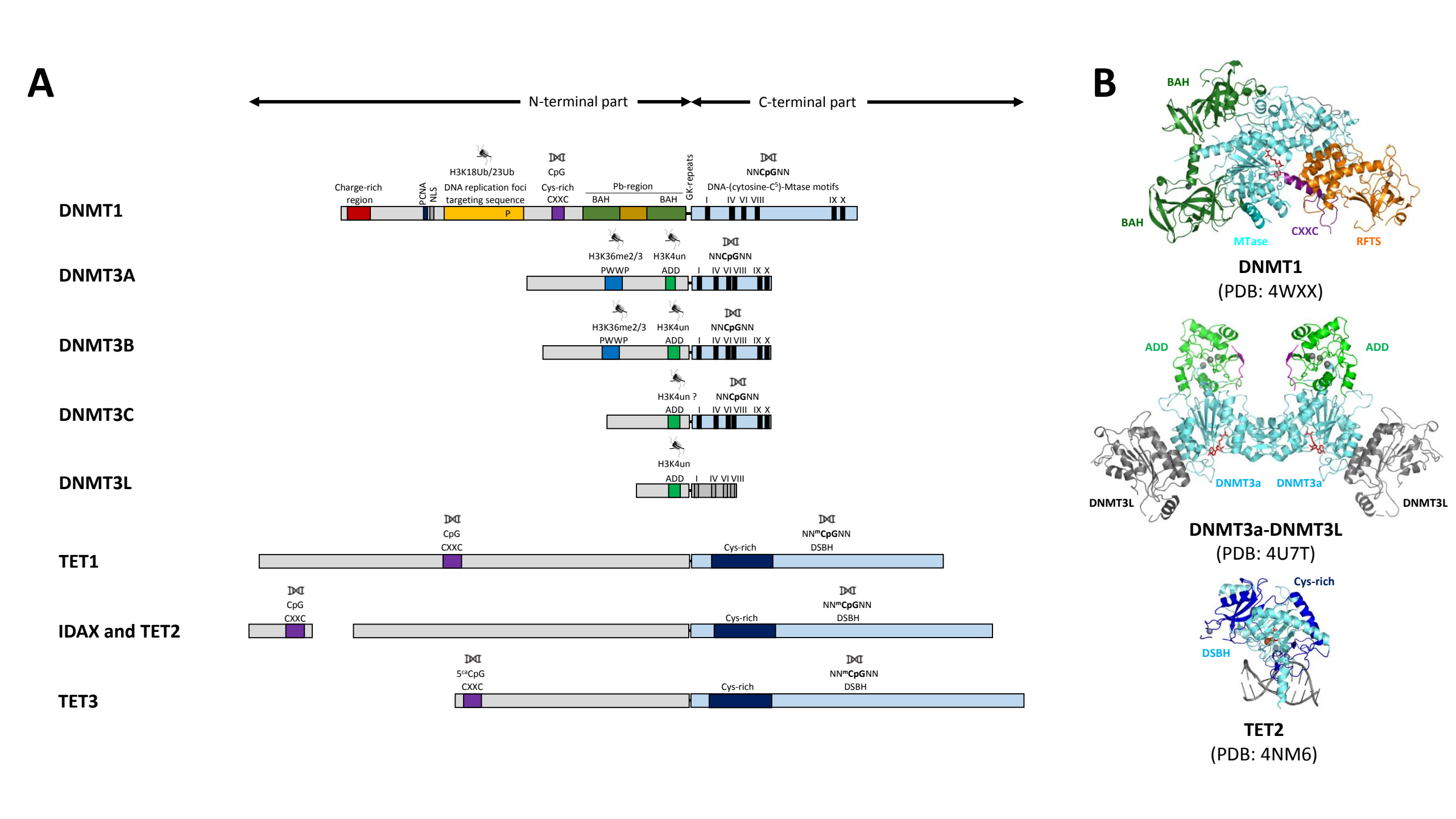
7. K. Hata, M. Okano, H. Lei and E. Li, *Development*, 2002, **129**, 1983-1993.
8. A. Jeltsch and R. Z. Jurkowska, *Trends in biochemical sciences*, 2014, **39**, 310-318.
9. A. Jeltsch, *Epigenetics*, 2006, **1**, 63-66.
10. Z. D. Smith and A. Meissner, *Nature reviews. Genetics*, 2013, **14**, 204-220.
11. O. Bogdanovic and R. Lister, *Current opinion in genetics & development*, 2017, **46**, 9-14.
12. M. L. Suva, N. Riggi and B. E. Bernstein, *Science*, 2013, **339**, 1567-1570.
13. T. Hamidi, A. K. Singh and T. Chen, *Epigenomics*, 2015, **7**, 247-265.
14. Y. Bergman and H. Cedar, *Nature structural & molecular biology*, 2013, **20**, 274-281.
15. S. B. Baylin and P. A. Jones, *Nature reviews. Cancer*, 2011, **11**, 726-734.
16. P. A. Jones, J. P. Issa and S. Baylin, *Nature reviews. Genetics*, 2016, **17**, 630-641.
17. N. Ahuja, A. R. Sharma and S. B. Baylin, *Annual review of medicine*, 2016, **67**, 73-89.
18. S. Tajima, I. Suetake, K. Takeshita, A. Nakagawa and H. Kimura, *Advances in experimental medicine and biology*, 2016, **945**, 63-86.
19. R. Z. Jurkowska, T. P. Jurkowski and A. Jeltsch, *Chembiochem : a European journal of chemical biology*, 2011, **12**, 206-222.
20. R. Z. Jurkowska and A. Jeltsch, *Advances in experimental medicine and biology*, 2016, **945**, 87-122.
21. A. Jeltsch and R. Z. Jurkowska, *Nucleic acids research*, 2016, **44**, 8556-8575.
22. M. R. Rountree, K. E. Bachman and S. B. Baylin, *Nature genetics*, 2000, **25**, 269-277.
23. G. Egger, S. Jeong, S. G. Escobar, C. C. Cortez, T. W. Li, Y. Saito, C. B. Yoo, P. A. Jones and G. Liang, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 14080-14085.
24. L. S. Chuang, H. I. Ian, T. W. Koh, H. H. Ng, G. Xu and B. F. Li, *Science*, 1997, **277**, 1996-2000.
25. H. Leonhardt, A. W. Page, H. U. Weier and T. H. Bestor, *Cell*, 1992, **71**, 865-873.
26. H. P. Easwaran, L. Schermelleh, H. Leonhardt and M. C. Cardoso, *EMBO reports*, 2004, **5**, 1181-1186.
27. J. Song, O. Rechkoblit, T. H. Bestor and D. J. Patel, *Science*, 2011, **331**, 1036-1040.
28. M. Pradhan, P. O. Esteve, H. G. Chin, M. Samaranayke, G. D. Kim and S. Pradhan, *Biochemistry*, 2008, **47**, 10000-10009.
29. P. Bashtrykov, G. Jankevicius, A. Smarandache, R. Z. Jurkowska, S. Ragozin and A. Jeltsch, *Chemistry & biology*, 2012, **19**, 572-578.
30. J. B. Margot, A. E. Ehrenhofer-Murray and H. Leonhardt, *BMC molecular biology*, 2003, **4**, 7.
31. M. Fatemi, A. Hermann, S. Pradhan and A. Jeltsch, *Journal of molecular biology*, 2001, **309**, 1189-1199.
32. J. Song, M. Teplova, S. Ishibe-Murakami and D. J. Patel, *Science*, 2012, **335**, 709-712.
33. F. Syeda, R. L. Fagan, M. Wean, G. V. Avvakumov, J. R. Walker, S. Xue, S. Dhe-Paganon and C. Brenner, *The Journal of biological chemistry*, 2011, **286**, 15344-15351.
34. K. Takeshita, I. Suetake, E. Yamashita, M. Suga, H. Narita, A. Nakagawa and S. Tajima, *Proceedings of the National Academy of Sciences of the United States of America*, 2011, **108**, 9055-9059.
35. S. Ito, A. C. D'Alessio, O. V. Taranova, K. Hong, L. C. Sowers and Y. Zhang, *Nature*, 2010, **466**, 1129-1133.
36. J. Otani, T. Nankumo, K. Arita, S. Inamoto, M. Ariyoshi and M. Shirakawa, *EMBO reports*, 2009, **10**, 1235-1241.
37. S. K. Ooi, C. Qiu, E. Bernstein, K. Li, D. Jia, Z. Yang, H. Erdjument-Bromage, P. Tempst, S. P. Lin, C. D. Allis, X. Cheng and T. H. Bestor, *Nature*, 2007, **448**, 714-717.
38. X. Guo, L. Wang, J. Li, Z. Ding, J. Xiao, X. Yin, S. He, P. Shi, L. Dong, G. Li, C. Tian, J. Wang, Y. Cong and Y. Xu, *Nature*, 2015, **517**, 640-644.
39. B. Z. Li, Z. Huang, Q. Y. Cui, X. H. Song, L. Du, A. Jeltsch, P. Chen, G. Li, E. Li and G. L. Xu, *Cell research*, 2011, **21**, 1172-1181.
40. Y. Z. Ge, M. T. Pu, H. Gowher, H. P. Wu, J. P. Ding, A. Jeltsch and G. L. Xu, *The Journal of biological chemistry*, 2004, **279**, 25447-25454.
41. T. Chen, N. Tsujimoto and E. Li, *Molecular and cellular biology*, 2004, **24**, 9048-9058.
42. F. Neri, S. Rapelli, A. Krepelova, D. Incarnato, C. Parlato, G. Basile, M. Maldotti, F. Anselmi and S. Oliviero, *Nature*, 2017, **543**, 72-77.
43. A. Dhayalan, A. Rajavelu, P. Rathert, R. Tamas, R. Z. Jurkowska, S. Ragozin and A. Jeltsch, *The Journal of biological chemistry*, 2010, **285**, 26114-26120.
44. T. Baubec, D. F. Colombo, C. Wirbelauer, J. Schmidt, L. Burger, A. R. Krebs, A. Akalin and D. Schubeler, *Nature*, 2015, **520**, 243-247.
45. I. Suetake, Y. Mishima, H. Kimura, Y. H. Lee, Y. Goto, H. Takeshima, T. Ikegami and S. Tajima, *The Biochemical journal*, 2011, **437**, 141-148.
46. S. Jeong, G. Liang, S. Sharma, J. C. Lin, S. H. Choi, H. Han, C. B. Yoo, G. Egger, A. S. Yang and P. A. Jones, *Molecular and cellular biology*, 2009, **29**, 5366-5376.
47. H. Gowher and A. Jeltsch, *The Journal of biological chemistry*, 2002, **277**, 20409-20414.
48. R. J. Klose and A. P. Bird, *Trends in biochemical sciences*, 2006, **31**, 89-97.
49. J. R. Edwards, O. Yarychkivska, M. Boulard and T. H. Bestor, *Epigenetics & chromatin*, 2017, **10**, 23.
50. X. Cheng and R. M. Blumenthal, *Structure*, 2008, **16**, 341-350.
51. S. Hong and X. Cheng, *Advances in experimental medicine and biology*, 2016, **945**, 321-341.
52. X. Cheng and R. J. Roberts, *Nucleic acids research*, 2001, **29**, 3784-3795.
53. G. Vilkaitis, I. Suetake, S. Klimasauskas and S. Tajima, *The Journal of biological chemistry*, 2005, **280**, 64-72.
54. A. Hermann, R. Goyal and A. Jeltsch, *The Journal of biological chemistry*, 2004, **279**, 48350-48359.
55. R. Goyal, R. Reinhardt and A. Jeltsch, *Nucleic acids research*, 2006, **34**, 1182-1188.
56. A. B. Norvil, C. J. Petell, L. Alabdi, L. Wu, S. Rossie and H. Gowher, *Biochemistry*, 2016, DOI: 10.1021/acs.biochem.6b00964.
57. A. Rajavelu, R. Z. Jurkowska, J. Fritz and A. Jeltsch, *Nucleic acids research*, 2012, **40**, 569-580.
58. R. Z. Jurkowska, N. Anspach, C. Urbanke, D. Jia, R. Reinhardt, W. Nellen, X. Cheng and A. Jeltsch, *Nucleic acids research*, 2008, **36**, 6656-6663.
59. D. Jia, R. Z. Jurkowska, X. Zhang, A. Jeltsch and X. Cheng, *Nature*, 2007, **449**, 248-251.
60. A. Jeltsch and R. Z. Jurkowska, *Progress in molecular biology and translational science*, 2013, **117**, 445-464.
61. M. Emperle, A. Rajavelu, R. Reinhardt, R. Z. Jurkowska and A. Jeltsch, *The Journal of biological chemistry*, 2014, **289**, 29602-29613.
62. P. Stepper, G. Kungulovski, R. Z. Jurkowska, T. Chandra, F. Krueger, R. Reinhardt, W. Reik, A. Jeltsch and T. P. Jurkowski, *Nucleic acids research*, 2017, **45**, 1703-1713.
63. C. Holz-Schietinger and N. O. Reich, *The Journal of biological chemistry*, 2010, **285**, 29091-29100.
64. P. Bashtrykov, S. Ragozin and A. Jeltsch, *FEBS letters*, 2012, **586**, 1821-1823.
65. H. Gowher and A. Jeltsch, *Journal of molecular biology*, 2001, **309**, 1201-1208.
66. A. Aoki, I. Suetake, J. Miyagawa, T. Fujio, T. Chijiwa, H. Sasaki and S. Tajima, *Nucleic acids research*, 2001, **29**, 3506-3512.
67. B. H. Ramsahoye, D. Biniszkiewicz, F. Lyko, V. Clark, A. P. Bird and R. Jaenisch, *Proceedings of the National Academy of Sciences of the United States of America*, 2000, **97**, 5237-5242.
68. R. Lister, E. A. Mukamel, J. R. Nery, M. Urich, C. A. Puddifoot, N. D. Johnson, J. Lucero, Y. Huang, A. J. Dwork, M. D. Schultz, M. Yu, J. Tonti-Filippini, H. Heyn, S. Hu, J. C. Wu, A. Rao, M. Esteller, C. He, F. G. Haghighi, T. J. Sejnowski, M. M. Behrens and J. R. Ecker, *Science*, 2013, **341**, 1237905.
69. J. U. Guo, Y. Su, J. H. Shin, J. Shin, H. Li, B. Xie, C. Zhong, S. Hu, T. Le, G. Fan, H. Zhu, Q. Chang, Y. Gao, G. L. Ming and H. Song, *Nature neuroscience*, 2014, **17**, 215-222.
70. M. J. Ziller, F. Muller, J. Liao, Y. Zhang, H. Gu, C. Bock, P. Boyle, C. B. Epstein, B. E. Bernstein, T. Lengauer, A. Gnirke and A. Meissner, *PLoS genetics*, 2011, **7**, e1002389.

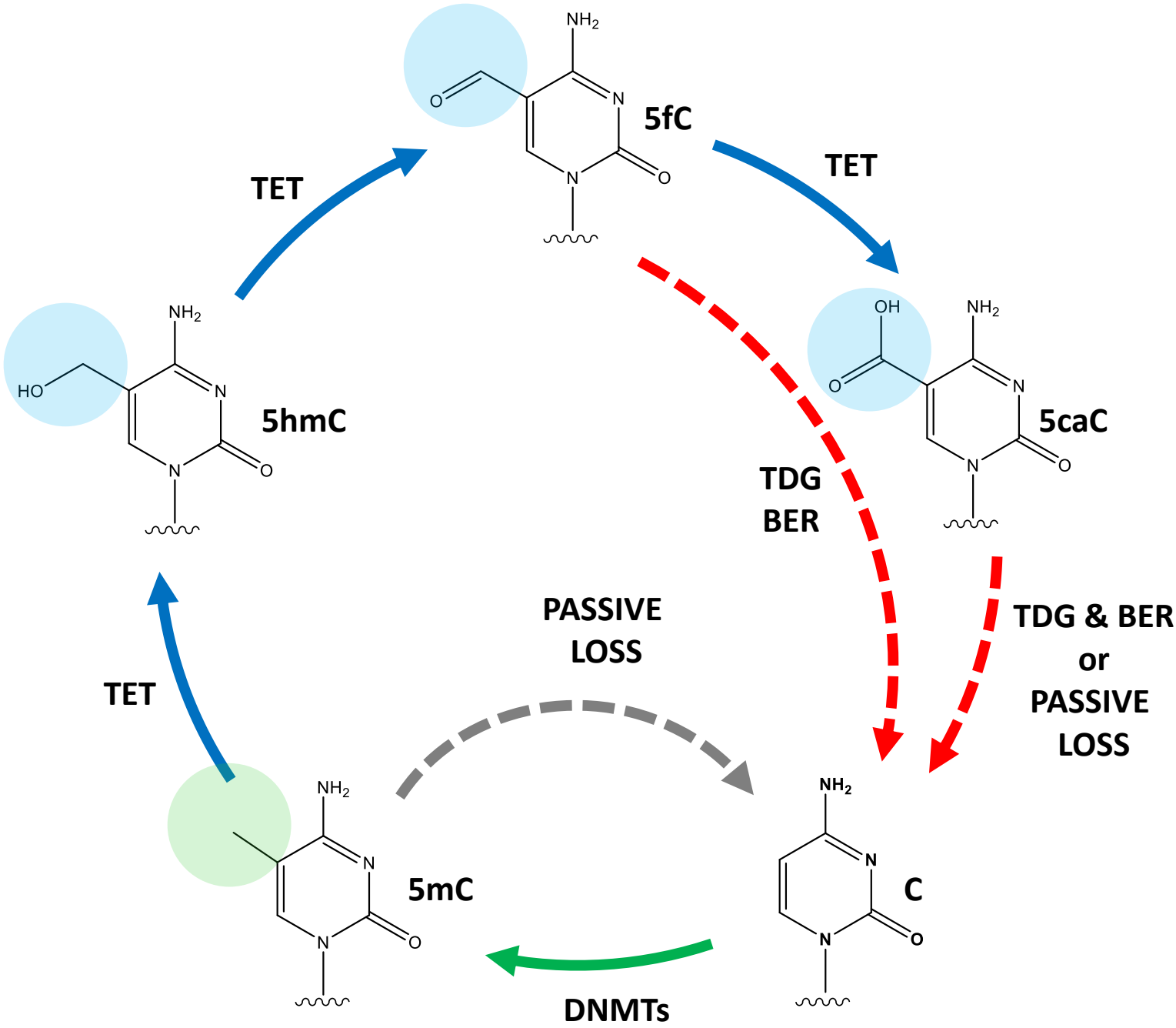
71. K. E. Varley, J. Gertz, K. M. Bowling, S. L. Parker, T. E. Reddy, F. Pauli-Behn, M. K. Cross, B. A. Williams, J. A. Stamatoyannopoulos, G. E. Crawford, D. M. Absher, B. J. Wold and R. M. Myers, *Genome research*, 2013, **23**, 555-567.
72. R. Lister, M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren and J. R. Ecker, *Nature*, 2009, **462**, 315-322.
73. M. D. Schultz, Y. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang and J. R. Ecker, *Nature*, 2015, **523**, 212-216.
74. S. Inoue and M. Oishi, *Gene*, 2005, **348**, 123-134.
75. R. Barres, M. E. Osler, J. Yan, A. Rune, T. Fritz, K. Caidahl, A. Krook and J. R. Zierath, *Cell metabolism*, 2009, **10**, 189-198.
76. H. S. Jang, W. J. Shin, J. E. Lee and J. T. Do, *Genes*, 2017, **8**.
77. B. Kinde, H. W. Gabel, C. S. Gilbert, E. C. Griffith and M. E. Greenberg, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, 6800-6806.
78. H. W. Gabel, B. Kinde, H. Stroud, C. S. Gilbert, D. A. Harmin, N. R. Kastan, M. Hemberg, D. H. Ebert and M. E. Greenberg, *Nature*, 2015, **522**, 89-93.
79. L. Chen, K. Chen, L. A. Lavery, S. A. Baker, C. A. Shaw, W. Li and H. Y. Zoghbi, *Proceedings of the National Academy of Sciences of the United States of America*, 2015, **112**, 5509-5514.
80. V. Handa and A. Jeltsch, *Journal of molecular biology*, 2005, **348**, 1103-1112.
81. R. Z. Jurkowska, A. N. Siddique, T. P. Jurkowski and A. Jeltsch, *Chembiochem : a European journal of chemical biology*, 2011, **12**, 1589-1594.
82. I. G. Lin, L. Han, A. Taghva, L. E. O'Brien and C. L. Hsieh, *Molecular and cellular biology*, 2002, **22**, 704-723.
83. Y. Zhang, C. Rohde, S. Tierling, T. P. Jurkowski, C. Bock, D. Santacruz, S. Ragozin, R. Reinhardt, M. Groth, J. Walter and A. Jeltsch, *PLoS genetics*, 2009, **5**, e1000438.
84. J. Arand, D. Spieler, T. Karius, M. R. Branco, D. Meilinger, A. Meissner, T. Jenuwein, G. Xu, H. Leonhardt, V. Wolf and J. Walter, *PLoS genetics*, 2012, **8**, e1002750.
85. M. Fatemi, A. Hermann, H. Gowher and A. Jeltsch, *European journal of biochemistry*, 2002, **269**, 4981-4984.
86. A. Meissner, T. S. Mikkelsen, H. Gu, M. Wernig, J. Hanna, A. Sivachenko, X. Zhang, B. E. Bernstein, C. Nusbaum, D. B. Jaffe, A. Gnirke, R. Jaenisch and E. S. Lander, *Nature*, 2008, **454**, 766-770.
87. A. L. Brunner, D. S. Johnson, S. W. Kim, A. Valouev, T. E. Reddy, N. F. Neff, E. Anton, C. Medina, L. Nguyen, E. Chiao, C. B. Oyolu, G. P. Schroth, D. M. Absher, J. C. Baker and R. M. Myers, *Genome research*, 2009, **19**, 1044-1056.
88. C. Ambrosi, M. Manzo and T. Baubec, *Journal of molecular biology*, 2017, **429**, 1459-1475.
89. D. B. Lipka, Q. Wang, N. Cabezas-Wallscheid, D. Klimmeck, D. Weichenhan, C. Herrmann, A. Lier, D. Brocks, L. von Paleske, S. Renders, P. Wunsche, P. Zeisberger, L. Gu, S. Haas, M. A. Essers, B. Brors, R. Eils, A. Trumpp, M. D. Milsom and C. Plass, *Cell Cycle*, 2014, **13**, 3476-3487.
90. A. Meissner, *Nature biotechnology*, 2010, **28**, 1079-1088.
91. D. Schubeler, *Nature*, 2015, **517**, 321-326.
92. M. P. Ball, J. B. Li, Y. Gao, J. H. Lee, E. M. LeProust, I. H. Park, B. Xie, G. Q. Daley and G. M. Church, *Nature biotechnology*, 2009, **27**, 361-368.
93. A. Hellman and A. Chess, *Science*, 2007, **315**, 1141-1143.
94. M. Weber, I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan and D. Schubeler, *Nature genetics*, 2007, **39**, 457-466.
95. A. M. Deaton, S. Webb, A. R. Kerr, R. S. Illingworth, J. Guy, R. Andrews and A. Bird, *Genome research*, 2011, **21**, 1074-1086.
96. A. Bird, M. Taggart, M. Frommer, O. J. Miller and D. Macleod, *Cell*, 1985, **40**, 91-99.
97. S. Saxonov, P. Berg and D. L. Brutlag, *Proceedings of the National Academy of Sciences of the United States of America*, 2006, **103**, 1412-1417.
98. F. Larsen, G. Gundersen, R. Lopez and H. Prydz, *Genomics*, 1992, **13**, 1095-1107.
99. E. Li, C. Beard and R. Jaenisch, *Nature*, 1993, **366**, 362-365.
100. M. B. Stadler, R. Murr, L. Burger, R. Ivanek, F. Lienert, A. Scholer, E. van Nimwegen, C. Wirbelauer, E. J. Oakeley, D. Gaidatzis, V. K. Tiwari and D. Schubeler, *Nature*, 2011, **480**, 490-495.
101. G. C. Hon, N. Rajagopal, Y. Shen, D. F. McCleary, F. Yue, M. D. Dang and B. Ren, *Nature genetics*, 2013, **45**, 1198-1206.
102. R. A. Irizarry, C. Ladd-Acosta, B. Wen, Z. Wu, C. Montano, P. Onyango, H. Cui, K. Gabo, M. Rongione, M. Webster, H. Ji, J. Potash, S. Sabunciyan and A. P. Feinberg, *Nature genetics*, 2009, **41**, 178-186.
103. M. Jeong, D. Sun, M. Luo, Y. Huang, G. A. Challen, B. Rodriguez, X. Zhang, L. Chavez, H. Wang, R. Hannah, S. B. Kim, L. Yang, M. Ko, R. Chen, B. Gottgens, J. S. Lee, P. Gunaratne, L. A. Godley, G. J. Darlington, A. Rao, W. Li and M. A. Goodell, *Nature genetics*, 2014, **46**, 17-23.
104. W. Xie, M. D. Schultz, R. Lister, Z. Hou, N. Rajagopal, P. Ray, J. W. Whitaker, S. Tian, R. D. Hawkins, D. Leung, H. Yang, T. Wang, A. Y. Lee, S. A. Swanson, J. Zhang, Y. Zhu, A. Kim, J. R. Nery, M. A. Urich, S. Kuan, C. A. Yen, S. Klugman, P. Yu, K. Suknuntha, N. E. Propson, H. Chen, L. E. Edsall, U. Wagner, Y. Li, Z. Ye, A. Kulkarni, Z. Xuan, W. Y. Chung, N. C. Chi, J. E. Antosiewicz-Bourget, I. Slukvin, R. Stewart, M. Q. Zhang, W. Wang, J. A. Thomson, J. R. Ecker and B. Ren, *Cell*, 2013, **153**, 1134-1148.
105. K. M. Noh, H. Wang, H. R. Kim, W. Wenderski, F. Fang, C. H. Li, S. Dewell, S. H. Hughes, A. M. Melnick, D. J. Patel, H. Li and C. D. Allis, *Molecular cell*, 2015, **59**, 89-103.
106. Y. Zhang, R. Jurkowska, S. Soeroes, A. Rajavelu, A. Dhayalan, I. Bock, P. Rathert, O. Brandt, R. Reinhardt, W. Fischle and A. Jeltsch, *Nucleic acids research*, 2010, **38**, 4246-4253.
107. E. Hodges, A. D. Smith, J. Kendall, Z. Xuan, K. Ravi, M. Rooks, M. Q. Zhang, K. Ye, A. Bhattacharjee, L. Brizuela, W. R. McCombie, M. Wigler, G. J. Hannon and J. B. Hicks, *Genome research*, 2009, **19**, 1593-1605.
108. M. Morselli, W. A. Pastor, B. Montanini, K. Nee, R. Ferrari, K. Fu, G. Bonora, L. Rubbi, A. T. Clark, S. Ottonello, S. E. Jacobsen and M. Pellegrini, *eLife*, 2015, **4**, e06205.
109. C. R. Vakoc, M. M. Sachdeva, H. Wang and G. A. Blobel, *Molecular and cellular biology*, 2006, **26**, 9185-9195.
110. P. Kolasinska-Zwierz, T. Down, I. Latorre, T. Liu, X. S. Liu and J. Ahringer, *Nature genetics*, 2009, **41**, 376-381.
111. J. W. Edmunds, L. C. Mahadevan and A. L. Clayton, *The EMBO journal*, 2008, **27**, 406-420.
112. A. Barski, S. Cuddapah, K. Cui, T. Y. Roh, D. E. Schones, Z. Wang, G. Wei, I. Chepelev and K. Zhao, *Cell*, 2007, **129**, 823-837.
113. M. G. Guenther, S. S. Levine, L. A. Boyer, R. Jaenisch and R. A. Young, *Cell*, 2007, **130**, 77-88.
114. J. Ernst, P. Kheradpour, T. S. Mikkelsen, N. Shoresh, L. D. Ward, C. B. Epstein, X. Zhang, L. Wang, R. Issner, M. Coyne, M. Ku, T. Durham, M. Kellis and B. E. Bernstein, *Nature*, 2011, **473**, 43-49.
115. C. Qiu, K. Sawada, X. Zhang and X. Cheng, *Nature structural biology*, 2002, **9**, 217-224.
116. M. M. Purdy, C. Holz-Schietinger and N. O. Reich, *Archives of biochemistry and biophysics*, 2010, **498**, 13-22.
117. G. Rondelet, T. Dal Maso, L. Willems and J. Wouters, *Journal of structural biology*, 2016, **194**, 357-367.
118. T. Iida, I. Suetake, S. Tajima, H. Morioka, S. Ohta, C. Obuse and T. Tsurimoto, *Genes to cells : devoted to molecular & cellular mechanisms*, 2002, **7**, 997-1007.
119. J. Sharif, M. Muto, S. Takebayashi, I. Suetake, A. Iwamatsu, T. A. Endo, J. Shinga, Y. Mizutani-Koseki, T. Toyoda, K. Okamura, S. Tajima, K. Mitsuya, M. Okano and H. Koseki, *Nature*, 2007, **450**, 908-912.
120. M. Bostick, J. K. Kim, P. O. Esteve, A. Clark, S. Pradhan and S. E. Jacobsen, *Science*, 2007, **317**, 1760-1764.
121. H. Hashimoto, J. R. Horton, X. Zhang, M. Bostick, S. E. Jacobsen and X. Cheng, *Nature*, 2008, **455**, 826-829.
122. G. V. Avvakumov, J. R. Walker, S. Xue, Y. Li, S. Duan, C. Bronner, C. H. Arrowsmith and S. Dhe-Paganon, *Nature*, 2008, **455**, 822-825.
123. K. Arita, M. Ariyoshi, H. Tochio, Y. Nakamura and M. Shirakawa, *Nature*, 2008, **455**, 818-821.
124. C. Wang, J. Shen, Z. Yang, P. Chen, B. Zhao, W. Hu, W. Lan, X. Tong, H. Wu, G. Li and C. Cao, *Cell research*, 2011, **21**, 1379-1382.

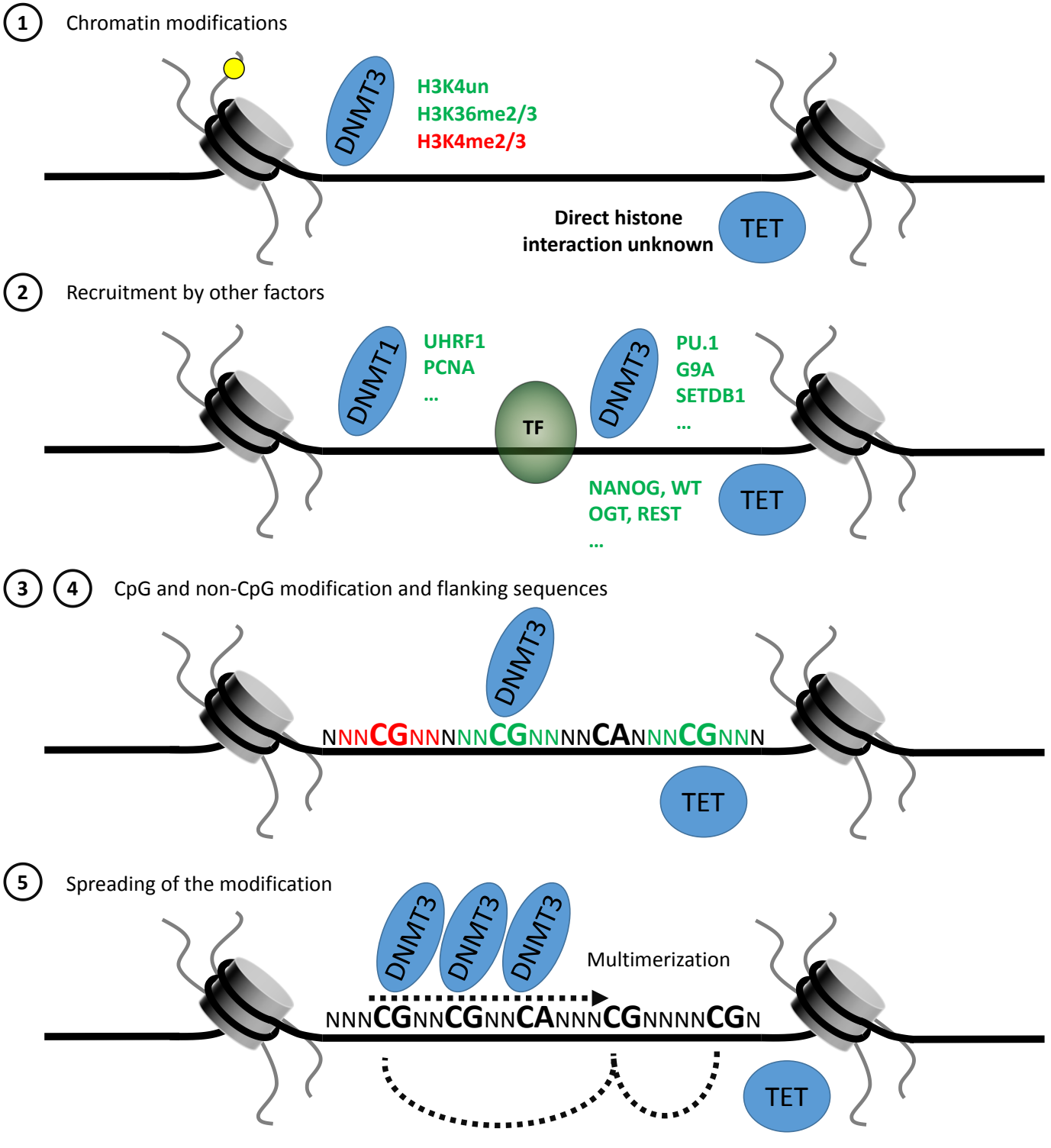
125. S. B. Rothbart, K. Krajewski, N. Nady, W. Tempel, S. Xue, A. I. Badeaux, D. Barsyte-Lovejoy, J. Y. Martinez, M. T. Bedford, S. M. Fuchs, C. H. Arrowsmith and B. D. Strahl, *Nature structural & molecular biology*, 2012, **19**, 1155-1160.
126. E. Rajakumara, Z. Wang, H. Ma, L. Hu, H. Chen, Y. Lin, R. Guo, F. Wu, H. Li, F. Lan, Y. G. Shi, Y. Xu, D. J. Patel and Y. Shi, *Molecular cell*, 2011, **43**, 275-284.
127. N. Nady, A. Lemak, J. R. Walker, G. V. Avvakumov, M. S. Kareta, M. Achour, S. Xue, S. Duan, A. Allali-Hassani, X. Zuo, Y. X. Wang, C. Bronner, F. Chedin, C. H. Arrowsmith and S. Dhe-Paganon, *The Journal of biological chemistry*, 2011, **286**, 24300-24311.
128. S. Qi, Z. Wang, P. Li, Q. Wu, T. Shi, J. Li and J. Wong, *The Journal of biological chemistry*, 2015, **290**, 14181-14191.
129. A. C. Berkyurek, I. Suetake, K. Arita, K. Takeshita, A. Nakagawa, M. Shirakawa and S. Tajima, *The Journal of biological chemistry*, 2014, **289**, 379-386.
130. P. Bashtrykov, G. Jankevicius, R. Z. Jurkowska, S. Ragozin and A. Jeltsch, *The Journal of biological chemistry*, 2014, **289**, 4106-4115.
131. W. Qin, P. Wolf, N. Liu, S. Link, M. Smets, F. La Mastra, I. Forne, G. Pichler, D. Horl, K. Fellinger, F. Spada, I. M. Bonapace, A. Imhof, H. Harz and H. Leonhardt, *Cell research*, 2015, **25**, 911-929.
132. A. Nishiyama, L. Yamaguchi, J. Sharif, Y. Johmura, T. Kawamura, K. Nakanishi, S. Shimamura, K. Arita, T. Kodama, F. Ishikawa, H. Koseki and M. Nakanishi, *Nature*, 2013, **502**, 249-253.
133. J. S. Harrison, E. M. Cornett, D. Goldfarb, P. A. DaRosa, Z. M. Li, F. Yan, B. M. Dickson, A. H. Guo, D. V. Cantu, L. Kaustov, P. J. Brown, C. H. Arrowsmith, D. A. Erie, M. B. Major, R. E. Klevit, K. Krajewski, B. Kuhlman, B. D. Strahl and S. B. Rothbart, *eLife*, 2016, **5**.
134. W. Mayer, A. Niveleau, J. Walter, R. Fundele and T. Haaf, *Nature*, 2000, **403**, 501-502.
135. J. Oswald, S. Engemann, N. Lane, W. Mayer, A. Olek, R. Fundele, W. Dean, W. Reik and J. Walter, *Current biology : CB*, 2000, **10**, 475-478.
136. P. Hajkova, S. Erhardt, N. Lane, T. Haaf, O. El-Maarri, W. Reik, J. Walter and M. A. Surani, *Mechanisms of development*, 2002, **117**, 15-23.
137. Y. Yamazaki, M. R. Mann, S. S. Lee, J. Marh, J. R. McCarrey, R. Yanagimachi and M. S. Bartolomei, *Proceedings of the National Academy of Sciences of the United States of America*, 2003, **100**, 12207-12212.
138. D. Bruniquel and R. H. Schwartz, *Nature immunology*, 2003, **4**, 235-240.
139. K. Martinowich, D. Hattori, H. Wu, S. Fouse, F. He, Y. Hu, G. Fan and Y. E. Sun, *Science*, 2003, **302**, 890-893.
140. S. K. Ooi and T. H. Bestor, *Cell*, 2008, **133**, 1145-1148.
141. J. P. Jost, *Proceedings of the National Academy of Sciences of the United States of America*, 1993, **90**, 4684-4688.
142. J. K. Zhu, *Annual review of genetics*, 2009, **43**, 143-166.
143. B. Zhu, Y. Zheng, H. Angliker, S. Schwarz, S. Thiry, M. Siegmann and J. P. Jost, *Nucleic acids research*, 2000, **28**, 4157-4165.
144. B. Zhu, Y. Zheng, D. Hess, H. Angliker, S. Schwarz, M. Siegmann, S. Thiry and J. P. Jost, *Proceedings of the National Academy of Sciences of the United States of America*, 2000, **97**, 5135-5139.
145. M. Tahiliani, K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind and A. Rao, *Science*, 2009, **324**, 930-935.
146. S. Kriaucionis and N. Heintz, *Science*, 2009, **324**, 929-930.
147. R. Ono, T. Taki, T. Taketani, M. Taniwaki, H. Kobayashi and Y. Hayashi, *Cancer research*, 2002, **62**, 4075-4080.
148. L. M. Iyer, M. Tahiliani, A. Rao and L. Aravind, *Cell Cycle*, 2009, **8**, 1698-1710.
149. L. J. Cliffe, R. Kieft, T. Southern, S. R. Birkeland, M. Marshall, K. Sweeney and R. Sabatini, *Nucleic acids research*, 2009, **37**, 1452-1462.
150. G. R. Wyatt and S. S. Cohen, *Nature*, 1952, **170**, 1072-1073.
151. N. W. Penn, R. Suwalski, C. O'Riley, K. Bojanowski and R. Yura, *The Biochemical journal*, 1972, **126**, 781-790.
152. R. M. Kothari and V. Shankar, *Journal of molecular evolution*, 1976, **7**, 325-329.
153. D. Globisch, M. Munzel, M. Muller, S. Michalakis, M. Wagner, S. Koch, T. Bruckl, M. Biel and T. Carell, *PloS one*, 2010, **5**, e15367.
154. Y. F. He, B. Z. Li, Z. Li, P. Liu, Y. Wang, Q. Tang, J. Ding, Y. Jia, Z. Chen, L. Li, Y. Sun, X. Li, Q. Dai, C. X. Song, K. Zhang, C. He and G. L. Xu, *Science*, 2011, **333**, 1303-1307.
155. S. Ito, L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He and Y. Zhang, *Science*, 2011, **333**, 1300-1303.
156. W. A. Pastor, L. Aravind and A. Rao, *Nature reviews. Molecular cell biology*, 2013, **14**, 341-356.
157. H. Hashimoto, J. E. Pais, X. Zhang, L. Saleh, Z. Q. Fu, N. Dai, I. R. Correa, Jr., Y. Zheng and X. Cheng, *Nature*, 2014, **506**, 391-395.
158. L. Hu, Z. Li, J. Cheng, Q. Rao, W. Gong, M. Liu, Y. G. Shi, J. Zhu, P. Wang and Y. Xu, *Cell*, 2013, **155**, 1545-1555.
159. S. Klimasauskas, S. Kumar, R. J. Roberts and X. Cheng, *Cell*, 1994, **76**, 357-369.
160. C. G. Yang, C. Yi, E. M. Duguid, C. T. Sullivan, X. Jian, P. A. Rice and C. He, *Nature*, 2008, **452**, 961-965.
161. T. Pfaffeneder, F. Spada, M. Wagner, C. Brandmayr, S. K. Laube, D. Eisen, M. Truss, J. Steinbacher, B. Hackner, O. Kotljarova, D. Schuermann, S. Michalakis, O. Kosmatchev, S. Schiesser, B. Steigenberger, N. Raddaoui, G. Kashiwazaki, U. Muller, C. G. Spruijt, M. Vermeulen, H. Leonhardt, P. Schar, M. Muller and T. Carell, *Nature chemical biology*, 2014, **10**, 574-581.
162. H. Hashimoto, Y. Liu, A. K. Upadhyay, Y. Chang, S. B. Howerton, P. M. Vertino, X. Zhang and X. Cheng, *Nucleic acids research*, 2012, **40**, 4841-4849.
163. V. Valinluck, H. H. Tsai, D. K. Rogstad, A. Burdzy, A. Bird and L. C. Sowers, *Nucleic acids research*, 2004, **32**, 4100-4108.
164. C. G. Spruijt, F. Gnerlich, A. H. Smits, T. Pfaffeneder, P. W. Jansen, C. Bauer, M. Munzel, M. Wagner, M. Muller, F. Khan, H. C. Eberl, A. Mensinga, A. B. Brinkman, K. Lephikov, U. Muller, J. Walter, R. Boelens, H. van Ingen, H. Leonhardt, T. Carell and M. Vermeulen, *Cell*, 2013, **152**, 1146-1159.
165. M. Iurlaro, G. Ficz, D. Oxley, E. A. Raiber, M. Bachman, M. J. Booth, S. Andrews, S. Balasubramanian and W. Reik, *Genome biology*, 2013, **14**, R119.
166. C. Frauer, T. Hoffmann, S. Bultmann, V. Casa, M. C. Cardoso, I. Antes and H. Leonhardt, *PloS one*, 2011, **6**, e21306.
167. D. Ji, K. Lin, J. Song and Y. Wang, *Molecular bioSystems*, 2014, **10**, 1749-1752.
168. M. Bochtler, A. Kolano and G. L. Xu, *BioEssays : news and reviews in molecular, cellular and developmental biology*, 2017, **39**, 1-13.
169. E. A. Raiber, D. Beraldi, G. Ficz, H. E. Burgess, M. R. Branco, P. Murat, D. Oxley, M. J. Booth, W. Reik and S. Balasubramanian, *Genome biology*, 2012, **13**, R69.
170. S. Cortellino, J. Xu, M. Sannai, R. Moore, E. Caretti, A. Cigliano, M. Le Coz, K. Devarajan, A. Wessels, D. Soprano, L. K. Abramowitz, M. S. Bartolomei, F. Rambow, M. R. Bassi, T. Bruno, M. Fanciulli, C. Renner, A. J. Klein-Szanto, Y. Matsumoto, D. Kobi, I. Davidson, C. Alberti, L. Larue and A. Bellacosa, *Cell*, 2011, **146**, 67-79.
171. C. S. Nabel, H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang and R. M. Kohli, *Nature chemical biology*, 2012, **8**, 751-758.
172. H. Hashimoto, S. Hong, A. S. Bhagwat, X. Zhang and X. Cheng, *Nucleic acids research*, 2012, **40**, 10203-10214.
173. A. Maiti and A. C. Drohat, *The Journal of biological chemistry*, 2011, **286**, 35334-35338.
174. F. Guo, X. Li, D. Liang, T. Li, P. Zhu, H. Guo, X. Wu, L. Wen, T. P. Gu, B. Hu, C. P. Walsh, J. Li, F. Tang and G. L. Xu, *Cell stem cell*, 2014, **15**, 447-459.
175. F. von Meyenn, M. Iurlaro, E. Habibi, N. Q. Liu, A. Salehzadeh-Yazdi, F. Santos, E. Petrini, I. Milagre, M. Yu, Z. Xie, L. I. Kroeze, T. B. Nesterova, J. H. Jansen, H. Xie, C. He, W. Reik and H. G. Stunnenberg, *Molecular cell*, 2016, **62**, 983.
176. Z. Li, T. P. Gu, A. R. Weber, J. Z. Shen, B. Z. Li, Z. G. Xie, R. Yin, F. Guo, X. Liu, F. Tang, H. Wang, P. Schar and G. L. Xu, *Nucleic acids research*, 2015, **43**, 3986-3997.
177. S. Kienhofer, M. U. Musheev, U. Stapf, M. Helm, L. Schomacher, C. Niehrs and A. Schafer, *Differentiation; research in biological diversity*, 2015, **90**, 59-68.
178. K. Arab, Y. J. Park, A. M. Lindroth, A. Schafer, C. Oakes, D. Weichenhan, A. Lukanova, E. Lundin, A. Risch, M. Meister, H. Dienemann, G. Dyckhoff, C. Herold-Mende, I. Grummt, C. Niehrs and C. Plass, *Molecular cell*, 2014, **55**, 604-614.

179. R. Olinski, M. Starczak and D. Gackowski, *Mutation research. Reviews in mutation research*, 2016, **767**, 59-66.
180. D. Banerjee, S. M. Mandal, A. Das, M. L. Hegde, S. Das, K. K. Bhakat, I. Boldogh, P. S. Sarkar, S. Mitra and T. K. Hazra, *The Journal of biological chemistry*, 2011, **286**, 6006-6016.
181. J. U. Guo, D. K. Ma, H. Mo, M. P. Ball, M. H. Jang, M. A. Bonaguidi, J. A. Balazer, H. L. Eaves, B. Xie, E. Ford, K. Zhang, G. L. Ming, Y. Gao and H. Song, *Nature neuroscience*, 2011, **14**, 1345-1351.
182. J. A. Smiley, M. Kundracik, D. A. Landfried, V. R. Barnes, Sr. and A. A. Axhemi, *Biochimica et biophysica acta*, 2005, **1723**, 256-264.
183. Z. Liutkeviciute, G. Lukinavicius, V. Masevicius, D. Daujotyte and S. Klimasauskas, *Nature chemical biology*, 2009, **5**, 400-402.
184. C. C. Chen, K. Y. Wang and C. K. Shen, *The Journal of biological chemistry*, 2012, **287**, 33116-33121.
185. Z. Liutkeviciute, E. Kriukiene, J. Licyte, M. Rudyte, G. Urbanaviciute and S. Klimasauskas, *Journal of the American Chemical Society*, 2014, **136**, 5884-5887.
186. J. P. Thomson, P. J. Skene, J. Selfridge, T. Clouaire, J. Guy, S. Webb, A. R. Kerr, A. Deaton, R. Andrews, K. D. James, D. J. Turner, R. Illingworth and A. Bird, *Nature*, 2010, **464**, 1082-1086.
187. N. P. Blackledge, J. P. Thomson and P. J. Skene, *Cold Spring Harbor perspectives in biology*, 2013, **5**, a018648.
188. H. Zhang, X. Zhang, E. Clark, M. Mulcahey, S. Huang and Y. G. Shi, *Cell research*, 2010, **20**, 1390-1393.
189. Y. Xu, F. Wu, L. Tan, L. Kong, L. Xiong, J. Deng, A. J. Barbera, L. Zheng, H. Zhang, S. Huang, J. Min, T. Nicholson, T. Chen, G. Xu, Y. Shi, K. Zhang and Y. G. Shi, *Molecular cell*, 2011, **42**, 451-464.
190. Y. Xu, C. Xu, A. Kato, W. Tempel, J. G. Abreu, C. Bian, Y. Hu, D. Hu, B. Zhao, T. Cerovina, J. Diao, F. Wu, H. H. He, Q. Cui, E. Clark, C. Ma, A. Barbara, G. J. Veenstra, G. Xu, U. B. Kaiser, X. S. Liu, S. P. Sugrue, X. He, J. Min, Y. Kato and Y. G. Shi, *Cell*, 2012, **151**, 1200-1213.
191. S. G. Jin, Z. M. Zhang, T. L. Dunwell, M. R. Harter, X. Wu, J. Johnson, Z. Li, J. Liu, P. E. Szabo, Q. Lu, G. L. Xu, J. Song and G. P. Pfeifer, *Cell reports*, 2016, **14**, 493-505.
192. Y. Wang, M. Xiao, X. Chen, L. Chen, Y. Xu, L. Lv, P. Wang, H. Yang, S. Ma, H. Lin, B. Jiao, R. Ren, D. Ye, K. L. Guan and Y. Xiong, *Molecular cell*, 2015, **57**, 662-673.
193. R. Rampal, A. Alkalin, J. Madzo, A. Vasanthakumar, E. Pronier, J. Patel, Y. Li, J. Ahn, O. Abdel-Wahab, A. Shih, C. Lu, P. S. Ward, J. J. Tsai, T. Hricik, V. Tosello, J. E. Tallman, X. Zhao, D. Daniels, Q. Dai, L. Ciminio, I. Aifantis, C. He, F. Fuks, M. S. Tallman, A. Ferrando, S. Nimer, E. Paietta, C. B. Thompson, J. D. Licht, C. E. Mason, L. A. Godley, A. Melnick, M. E. Figueroa and R. L. Levine, *Cell reports*, 2014, **9**, 1841-1855.
194. P. Guilhamon, M. Eskandarpour, D. Halai, G. A. Wilson, A. Feber, A. E. Teschendorff, V. Gomez, A. Hergovich, R. Tirabosco, M. Fernanda Amary, D. Baumhoer, G. Jundt, M. T. Ross, A. M. Flanagan and S. Beck, *Nature communications*, 2013, **4**, 2166.
195. Y. Costa, J. Ding, T. W. Theunissen, F. Faiola, T. A. Hore, P. V. Shliaha, M. Fidalgo, A. Saunders, M. Lawrence, S. Dietmann, S. Das, D. N. Levasseur, Z. Li, M. Xu, W. Reik, J. C. Silva and J. Wang, *Nature*, 2013, **495**, 370-374.
196. A. Perera, D. Eisen, M. Wagner, S. K. Laube, A. F. Kunzel, S. Koch, J. Steinbacher, E. Schulze, V. Splith, N. Mittermeier, M. Muller, M. Biel, T. Carell and S. Michalakis, *Cell reports*, 2015, **11**, 283-294.
197. U. Muller, C. Bauer, M. Siegl, A. Rottach and H. Leonhardt, *Nucleic acids research*, 2014, **42**, 8592-8604.
198. P. Vella, A. Scelfo, S. Jammula, F. Chiacchiera, K. Williams, A. Cuomo, A. Roberto, J. Christensen, T. Bonaldi, K. Helin and D. Pasini, *Molecular cell*, 2013, **49**, 645-656.
199. R. Fujiki, W. Hashiba, H. Sekine, A. Yokoyama, T. Chikanishi, S. Ito, Y. Imai, J. Kim, H. H. He, K. Igarashi, J. Kanno, F. Ohtake, H. Kitagawa, R. G. Roeder, M. Brown and S. Kato, *Nature*, 2011, **480**, 557-560.
200. Q. Chen, Y. Chen, C. Bian, R. Fujiki and X. Yu, *Nature*, 2013, **493**, 561-564.
201. M. Ehrlich, M. A. Gama-Sosa, L. H. Huang, R. M. Midgett, K. C. Kuo, R. A. McCune and C. Gehrke, *Nucleic acids research*, 1982, **10**, 2709-2721.
202. C. E. Nestor, R. Ottaviano, J. Reddington, D. Sproul, D. Reinhardt, D. Dunican, E. Katz, J. M. Dixon, D. J. Harrison and R. R. Meehan, *Genome research*, 2012, **22**, 467-477.
203. M. Munzel, D. Globisch and T. Carell, *Angew Chem Int Ed Engl*, 2011, **50**, 6460-6468.
204. W. Li and M. Liu, *Journal of nucleic acids*, 2011, **2011**, 870726.
205. S. M. Kinney, H. G. Chin, R. Vaisvila, J. Bitinaite, Y. Zheng, P. O. Esteve, S. Feng, H. Stroud, S. E. Jacobsen and S. Pradhan, *The Journal of biological chemistry*, 2011, **286**, 24685-24693.
206. C. G. Lian, Y. Xu, C. Ceol, F. Wu, A. Larson, K. Dresser, W. Xu, L. Tan, Y. Hu, Q. Zhan, C. W. Lee, D. Hu, B. Q. Lian, S. Kleffel, Y. Yang, J. Neiswender, A. J. Khorasani, R. Fang, C. Lezcano, L. M. Duncan, R. A. Scolyer, J. F. Thompson, H. Kakavand, Y. Houvras, L. I. Zon, M. C. Mihm, Jr., U. B. Kaiser, T. Schatton, B. A. Woda, G. F. Murphy and Y. G. Shi, *Cell*, 2012, **150**, 1135-1146.
207. T. Pfaffeneder, B. Hackner, M. Truss, M. Munzel, M. Muller, C. A. Deiml, C. Hagemeier and T. Carell, *Angew Chem Int Ed Engl*, 2011, **50**, 7008-7012.
208. M. Bachman, S. Uribe-Lewis, X. Yang, H. E. Burgess, M. Iurlaro, W. Reik, A. Murrell and S. Balasubramanian, *Nature chemical biology*, 2015, **11**, 555-557.
209. M. Bachman, S. Uribe-Lewis, X. Yang, M. Williams, A. Murrell and S. Balasubramanian, *Nature chemistry*, 2014, **6**, 1049-1055.
210. K. E. Szulwach, X. Li, Y. Li, C. X. Song, J. W. Han, S. Kim, S. Namburi, K. Hermetz, J. J. Kim, M. K. Rudd, Y. S. Yoon, B. Ren, C. He and P. Jin, *PLoS genetics*, 2011, **7**, e1002154.
211. W. A. Pastor, U. J. Pape, Y. Huang, H. R. Henderson, R. Lister, M. Ko, E. M. McLoughlin, Y. Brudno, S. Mahapatra, P. Kapranov, M. Tahiliani, G. Q. Daley, X. S. Liu, J. R. Ecker, P. M. Milos, S. Agarwal and A. Rao, *Nature*, 2011, **473**, 394-397.
212. F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, A. Pagnani, R. Zecchina, C. Parlato and S. Oliviero, *Genome biology*, 2013, **14**, R91.
213. S. G. Jin, X. Wu, A. X. Li and G. P. Pfeifer, *Nucleic acids research*, 2011, **39**, 5015-5024.
214. G. Ficz, M. R. Branco, S. Seisenberger, F. Santos, F. Krueger, T. A. Hore, C. J. Marques, S. Andrews and W. Reik, *Nature*, 2011, **473**, 398-402.
215. M. J. Booth, M. R. Branco, G. Ficz, D. Oxley, F. Krueger, W. Reik and S. Balasubramanian, *Science*, 2012, **336**, 934-937.
216. M. Yu, G. C. Hon, K. E. Szulwach, C. X. Song, L. Zhang, A. Kim, X. Li, Q. Dai, Y. Shen, B. Park, J. H. Min, P. Jin, B. Ren and C. He, *Cell*, 2012, **149**, 1368-1380.
217. A. B. Robertson, J. A. Dahl, C. B. Vagbo, P. Tripathi, H. E. Krokan and A. Klungland, *Nucleic acids research*, 2011, **39**, e55.
218. L. Wen, X. Li, L. Yan, Y. Tan, R. Li, Y. Zhao, Y. Wang, J. Xie, Y. Zhang, C. Song, M. Yu, X. Liu, P. Zhu, Y. Hou, H. Guo, X. Wu, C. He, F. Tang and J. Qiao, *Genome biology*, 2014, **15**, R49.
219. C. X. Song, K. E. Szulwach, Y. Fu, Q. Dai, C. Yi, X. Li, Y. Li, C. H. Chen, W. Zhang, X. Jian, J. Wang, L. Zhang, T. J. Looney, B. Zhang, L. A. Godley, L. M. Hicks, B. T. Lahn, P. Jin and C. He, *Nature biotechnology*, 2011, **29**, 68-72.
220. C. X. Song, K. E. Szulwach, Q. Dai, Y. Fu, S. Q. Mao, L. Lin, C. Street, Y. Li, M. Poidevin, H. Wu, J. Gao, P. Liu, L. Li, G. L. Xu, P. Jin and C. He, *Cell*, 2013, **153**, 678-691.
221. F. Neri, D. Incarnato, A. Krepelova, S. Rapelli, F. Anselmi, C. Parlato, C. Medana, F. Dal Bello and S. Oliviero, *Cell reports*, 2015, DOI: 10.1016/j.celrep.2015.01.008.
222. M. Iurlaro, G. R. McInroy, H. E. Burgess, W. Dean, E. A. Raiber, M. Bachman, D. Beraldi, S. Balasubramanian and W. Reik, *Genome biology*, 2016, **17**, 141.
223. H. Wu, X. Wu, L. Shen and Y. Zhang, *Nature biotechnology*, 2014, **32**, 1231-1240.
224. K. Williams, J. Christensen, M. T. Pedersen, J. V. Johansen, P. A. Cloos, J. Rappsilber and K. Helin, *Nature*, 2011, **473**, 343-348.
225. H. Wu, A. C. D'Alessio, S. Ito, K. Xia, Z. Wang, K. Cui, K. Zhao, Y. E. Sun and Y. Zhang, *Nature*, 2011, **473**, 389-393.
226. L. Hu, J. Lu, J. Cheng, Q. Rao, Z. Li, H. Hou, Z. Lou, L. Zhang, W. Li, W. Gong, M. Liu, C. Sun, X. Yin, J. Li, X. Tan, P. Wang, Y. Wang, D. Fang, Q. Cui, P. Yang, C. He, H. Jiang, C. Luo and Y. Xu, *Nature*, 2015, **527**, 118-122.
227. A. A. Serandour, S. Avner, E. A. Mahe, T. Madigou, S. Guibert, M. Weber and G. Salbert, *Genome biology*, 2016, **17**, 56.
228. S. Kizaki, A. Chandran and H. Sugiyama, *Chembiochem : a European journal of chemical biology*, 2016, **17**, 403-406.

229. E. Tamanaha, S. Guan, K. Marks and L. Saleh, *Journal of the American Chemical Society*, 2016, **138**, 9345-9348.
230. M. Y. Liu, H. Torabifard, D. J. Crawford, J. E. DeNizio, X. J. Cao, B. A. Garcia, G. A. Cisneros and R. M. Kohli, *Nature chemical biology*, 2017, **13**, 181-187.
231. D. J. Crawford, M. Y. Liu, C. S. Nabel, X. J. Cao, B. A. Garcia and R. M. Kohli, *Journal of the American Chemical Society*, 2016, **138**, 730-733.
232. T. P. Jurkowski, N. Anspach, L. Kulishova, W. Nellen and A. Jeltsch, *The Journal of biological chemistry*, 2007, **282**, 36942-36952.
233. E. J. Little, A. C. Babic and N. C. Horton, *Structure*, 2008, **16**, 1828-1837.


A
N-terminal part
C-terminal part
DNMT1
H3K18Ub/23Ub
CpG
NNCpGNN
Charge-rich region
PCNA
NLS
DNA replication foci targeting sequence
Cys-rich CXXC
Pb-region
BAH
BAH
GK-repeats
DNA-(cytosine-C5)-Mtase motifs
I IV VI VIII IX X
P
DNMT3A
H3K36me2/3
H3K4un
PWWP
ADD
DNMT3B
DNMT3C
H3K4un ?
DNMT3L
TET1
NNmCpGNN
Cys-rich
DSBH
IDAX and TET2
TET3
5caCpG
B
BAH
MTase
CXXC
RFTS
DNMT1
(PDB: 4WXX)
ADD
DNMT3a
DNMT3L
DNMT3a-DNMT3L
(PDB: 4U7T)
DSBH
TET2
(PDB: 4NM6)


5fC
TET
5caC
TDG
BER
TDG & BER
or
PASSIVE
LOSS
C
DNMTs
5mC
PASSIVE
LOSS
TET
5hmC
TET


1
Chromatin modifications
DNMT3
H3K4un
H3K36me2/3
H3K4me2/3
Direct histone
interaction unknown
TET
2
Recruitment by other factors
DNMT1
UHRF1
PCNA
...
TF
DNMT3
PU.1
G9A
SETDB1
...
NANOG, WT
OGT, REST
...
TET
3
4
CpG and non-CpG modification and flanking sequences
DNMT3
NNNCGNNNNNCGNNNNCANNNCGNNN
TET
5
Spreading of the modification
DNMT3
DNMT3
DNMT3
Multimerization
NNNCGNNCGNNCANNNCGNNNNCGN
TET