Regulation of transposable elements by DNA modifications

Özgen Deniz[1], *Jennifer M. Frost*[1] *and Miguel R. Branco*[1*]

[1]**Blizard Institute, Barts and The London School of Medicine and Dentistry, QMUL, London E1 2AT, UK**

***e-mail: m.branco@qmul.ac.uk**

Abstract | Maintenance of genome stability requires control over the expression of transposable elements (TEs), whose activity can have substantial deleterious effects on the host. Chemical modification of DNA is a commonly used strategy to achieve this, and it has long been argued that the emergence of 5-methylcytosine (5mC) in many species was driven by the requirement to silence TEs. Potential roles in TE regulation have also been suggested for other DNA modifications, such as N6-methyladenine and oxidation derivatives of 5mC, although the underlying mechanistic relationships are poorly understood. Here, we discuss current evidence implicating DNA modifications and DNA modifying enzymes in TE regulation across different species.

[H1] Introduction

Organismal complexity does not correlate with genome size, which varies across species by far more than the number of genes it harbours, in what is known as the C-value paradox[1]. The main contributors to this variation are transposable elements (TEs), mobile genetic entities whose abundance in the genome differs immensely between species (Figure 1)[2,3]. TEs are considered 'selfish' genetic elements, whose evolution is primarily determined by selective pressures that only affect their own survival within the genome, with generally minimal effects on host phenotype. Although TE insertions can lead to adaptive effects through the generation of new genes or *cis*-regulatory elements[4-6], most extant TE insertions are neutral or only mildly deleterious to the host, and have become fixed through genetic drift **[G]** [7]. The fast evolution of the TE landscape is further fuelled by horizontal propagation **[G]** of TEs across species and viruses[8]. The myriad origins and evolutionary paths of TEs have generated thousands of distinct families that underlie remarkable differences in TE composition across species (Figure 1b)[2,3]. This diversity presents a challenge for categorizing TEs into groups with distinct structural and molecular properties (Figure 1a)[9-11]. The top of the classification hierarchy divides TEs into retrotransposons (class I TEs), which transpose via an RNA intermediate, and DNA transposons (class II TEs). DNA transposons are widespread and active across many bacterial, archaeal and eukaryotic species, but have become inactive in most mammals[12,13]. Retrotransposons can be broadly divided into long terminal repeat (LTR) and non-LTR elements, with the former being predominant in the TE landscape of plants, including in the relatively TE-poor genome of *Arabidopsis thaliana*, where several families retain the capacity for mobility[14]. Whereas LTR elements are also numerous among animal species, long interspersed nuclear elements (LINEs) are more abundant in many animal genomes, with nearly 20% of the mouse and human genomes being made up of LINE-1 (also known as L1) elements (Figure 1b)[15,16]. The mouse genome hosts thousands of functional LINE-1 copies with the potential to generate new heritable insertions in germ cells and during early embryogenesis[17]. In humans, a mere 80–100 LINE-1 elements account for virtually all of the transposition activity observed today[18], including retrotransposition of non-autonomous short interspersed nuclear elements (SINEs), which depend on proteins encoded by LINE-1 elements[19].

Although many TEs are neutral in their effect on the host, some insertions can disrupt gene function or lead to harmful chromosomal rearrangements, as demonstrated by over 120 disease-causing TE insertions in humans[20]. Additionally, exacerbated TE expression in the germline can lead to sterility in mice and fruit flies[21,22]. The resulting selective pressure has driven the evolution of numerous transcriptional and post-transcriptional host defence mechanisms that repress TE expression (Figure 2), which have been recently reviewed by Molaro and Malik[22]. Small RNAs, including PIWI-interacting RNAs **[G]** (piRNAs), are the

primary mechanism of TE silencing in nematodes, flies and yeast, of *de novo* silencing in flowering plants, and during germ cell development in vertebrates (Figure 2a,b,d)[21,23]. Small RNAs can act post-transcriptionally via targeted RNA degradation, as well as transcriptionally, inducing the deposition of repressive epigenetic modifications at TE loci. A key silencing pathway in mammals that exemplifies the rapid evolution of host defence mechanisms involve the large family of Krüppel-associated box (KRAB) zinc-finger proteins (ZFPs), which are thought to have co-evolved with TEs[24,25]. KRAB-ZFPs binding to TEs recruits KRAB-associated protein 1 (KAP1; also known as TRIM28), leading to the formation of repressive chromatin via multiple interacting partners (Figure 2c). ATP-dependent chromatin remodellers are also important players in TE regulation in both mammals and plants, providing access for methyltransferases to confer repressive DNA and chromatin modifications (Figure 2e)[26-28]. However, DNA methylation, more specifically 5-methylcytosine (5mC), is arguably the most widely adopted strategy by higher eukaryotes for maintaining TEs in a repressive state. Deposition of 5mC often depends on, and affects, other TE-targeting pathways such as those mentioned above (Figure 2). It has been argued that the need to maintain TE silencing drove the evolution of DNA methylation as a repressive mechanism[29], and later co-opted to act in other contexts, such as gene imprinting.

Although 5mC has been the most studied DNA modification to date, there are other modifications that have also been linked to TEs in different species. Apart from 5mC, the most common types of enzyme-catalysed DNA modifications include *N*4-methylcytosine (4mC) and *N*6-methyladenine (6mA), which are widespread across bacteria. Notably, 6mA is also found in varying amounts in eukaryotes[30] and has recently been implicated in TE regulation in *Drosophila melanogaster* and mice[31,32], although the presence and functional relevance of 6mA in higher eukaryotes remains controversial[33-35]. Moreover, 5mC can be oxidized by ten-eleven translocation (TET) enzymes to 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC), as part of a replication-independent pathway to DNA demethylation (Figure 3). Whether through their demethylating action or the direct impact of the modifications they generate, TET enzymes are also emerging as important regulators of TE activity[36-40].

Here we review the evidence implicating DNA modifications in the regulation of TE expression, focusing on 5mC, its oxidation derivatives, and 6mA. We discuss their distribution and associated molecular mechanisms across species and tissues, with particular emphasis on mammalian embryogenesis, germline development and cancer. Modifications to RNA have also been implicated in TE regulation (Box 1) but are not comprehensively covered in this Review. A wider view of the effects of DNA modifications on genome function and their interplay with other aspects of gene regulation can be gained from many excellent reviews (For example, REFs[33,41-46]).

[H1] 5-methylcytosine

5mC is a widespread form of DNA methylation that is present in a large number of bacterial, archaeal and eukaryotic genomes, including fungi, plants and animals. Yet many eukaryotes lack detectable levels of 5mC, including *D. melanogaster*, *Saccharomyces cerevisiae* and *Caenorhabditis elegans*[47]. This suggests that selective pressures to adopt 5mC as a silencing mechanism have varied between different branches of evolution. One potential explanation for these differences is that species with large population sizes are able to efficiently eliminate weakly deleterious TE insertions and are therefore not dependent on DNA methylation[48].

[H2] Evidence for 5mC-mediated TE silencing across species. The distribution of 5mC across genomes is tied to the evolution of DNA methyltransferase (DNMT) enzymes, whose genes have undergone duplication and/or loss since the last common eukaryotic ancestor[47]. The dynamic nature of *Dnmt* gene evolution is particularly notable in the nematode lineage, where relatively closely related species display different numbers of *Dnmt* genes, including species where these genes are altogether absent, such as in *C. elegans*[49]. Interestingly, *Dnmt*-containing nematodes display an enrichment of 5mC at TEs, further suggesting a link between the emergence of 5mC and TE silencing[49]. Eukaryotic DNMTs are broadly classified into DNMT1 and DNMT3 families (Figure 3). DNMT1 has a preference for hemi-methylated CpG dinucleotides, which enables replication-coupled maintenance of 5mC across cell division, whereas DNMT3-type enzymes mainly catalyse *de novo* deposition of 5mC, including in non-CpG contexts. Non-CpG methylation is particularly prominent in plants and is deposited by DNMT3-like enzymes (DRM1 and DRM2), as well as by members

of a separate chromomethylase family (CMT2 and CMT3)[23]. Notably, unlike other regions in the *A. thaliana* genome, TEs are heavily methylated in all cytosine contexts, consistent with a primary role of 5mC in TE silencing[23]. Indeed, *A. thaliana* mutants defective in DNA methylation (e.g., *met1* and *ddm1*) display massive upregulation of TE expression and mobilization of certain TE families, such as *copia*-like LTR elements and *CACTA* DNA TEs[14,50]. One notable group of TEs in *A. thaliana* are *VANDAL* elements, which have evolved a mechanism to evade 5mC-mediated silencing[51]. Multiple *VANDAL* TE families encode for VANC proteins, which induce highly sequence-specific hypomethylation of these TEs, enabling their expression and propagation[51].

Other species where definitive associations between DNMT action and TE silencing have been established include the fungus *Neurospora crassa*[52] and zebrafish[53], which seemingly use 5mC primarily for this purpose. In mice, a role for 5mC in TE repression was first revealed in mid-gestation embryos lacking DNMT1, which display a marked upregulation of intracisternal A particle (IAP) endogenous retroviruses (ERVs)[54]. However, since then surprisingly little has been found about 5mC-mediated TE regulation in differentiated tissues *in vivo*. The use of *Dnmt* conditional knockout lines has the potential to bring deeper insights into this question, but analyses of these models commonly overlook the expression of TEs, or are limited to a specific family[55]. Nonetheless, tissue culture models have provided additional clues about the target- and tissue-specific action of 5mC at TEs. Namely, 5mC is largely dispensible for LINE-1 silencing in cultured embryonic fibroblasts, which contrasts with the pronounced effects that hypomethylation has on IAP elements[27,56]. However, in embryonic stem cells (ESCs) and neural progenitor cells, IAPs and many other ERVs are kept repressed through largely DNMT-independent mechanisms, and instead depend mainly on KAP1-mediated recruitment of the H3K9 methyltransferase SET domain bifurcated 1 (SETDB1)[57-61]. Although KAP1- and DNMT-dependent pathways are mechanistically intertwined, they play largely complementary roles in TE silencing, displaying distinct preferences for different TE families[59,62].

[H2] TE regulation during epigenetic reprogramming in mice. During development, mammals undergo two major waves of epigenetic reprogramming where the genome is rapidly demethylated, first to form the pluripotent cells of the conceptus during pre-implantation development **[G]** , and second to produce gametes (Figure 4). In mice, DNA hypomethylation coincides with the transient upregulation of several TEs during both reprogramming periods[63,64], including LINE-1 elements (Figure 4), although a causal link between DNA hypomethylation and TE expresion has not been firmly established.

In these critical developmental stages, particularly in the germline**,** the risk and consequences of deleterious *de novo* insertions of TEs are high. To minimize the associated mutagenic risks, the host uses numerous complementary strategies to restrict TE mobility. In the male germline, TE expression following from 5mC erasure leads rapidly to the activation of the piRNA pathway (Figure 2a). piRNAs drive TE re-methylation in a sequence-specific manner, thus minimizing exposure of the germline to mobile elements[65]. Interestingly, it has been suggested that demethylation also drives the expression of several genes encoding proteins involved in the piRNA pathway (e.g., *Mov10l1* and *Mili*) during reprogramming[66]. The importance of 5mC for TE repression can be observed in *Dnmt3l* or *Miwi2* knockout mice during spermatogenesis, where LINE-1s and IAPs are reactivated[67,68]. *Dnmt3l* knockout mice are also phenocopied by mutants of a newly discovered DNMT3C enzyme that specifically methylates young TEs in mouse male germ cells[69,70]. The specificity of this enzyme highlights the selective pressure to ensure TE silencing during germ cell development, and its evolution may have been driven by high rates of retrotransposon invasion in the muroid lineage[69,70]. Notably, *Dnmt3l* mutation is linked to germ cell developmental arrest and infertility, which is not caused by retrotransposition, but via chromatin changes at active TE loci that lead to aberrant meiotic hotspot formation[71]. Similar to the male germline, oocytes make use of small RNA-based mechanisms to reduce the TE mutational load, with DICER-dependent endogenous small interfering RNAs (endosiRNAs **[G]**) targeting MT elements in particular, and piRNAs playing an important role in IAP repression[72-74]. Re-methylation of the oocyte genome occurs only after birth, providing a large window of opportunity for TE expression (Figure 4). Indeed, oocytes display high levels of LINE-1 expression, which is linked to fetal oocyte attrition[75]. It has been proposed that LINE-1 mediated DNA damage may provide a mechanism to only allow survival of oocytes with low LINE-1 expression[75].

A fraction of the genome is partially resistant to demethylation during epigenetic reprogramming, which notably includes highly mutagenic IAP elements (specifically the IAPEz family), as well as other ERV1 and

ERVK families[76-79]. In both preimplantation embryos and primordial germ cells **[G]** (PGCs), enrichment for H3K9me3 at demethylation-resistant TEs implies a role for this mark in preventing the expression of particularly aggressive TEs[80]. Indeed, depletion of SETDB1 in PGCs drives a reduction in 5mC levels at IAP LTRs and reactivation of these TEs[80]. Further insights into a potential mechanism underlying DNA methylation resistance to reprogramming have been gained from ESCs during culture condition-mediated transition to naïve pluripotency **[G]** , which recapitulates many of the key features of *in vivo* epigenetic reprogramming, including DNA demethylation[81]. Importantly, demethylation-resistant ERVs, including IAP elements, are also enriched for H3K9me3 in naïve ESCs[81]. It was found that this selective resistance is linked to the specific recruitment of UHRF1 (ubiquitin like with PHD and RING finger domains 1, also known as NP95) to those loci by H3K9me2/3, leading to localised maintenance of 5mC levels (Figure 2f)[82-84]. Thus, the resistance of a subset of TEs to DNA demethylation may be underpinned by limiting levels of UHRF1 during replication, which is seemingly achieved in both preimplantation embryos and PGCs by largely excluding UHRF1 from cell nuclei[76,85]. Importantly, UHRF1 is essential for maintenance of 5mC levels at IAPs in preimplantation embryos[85].

ESC-based models have presented many other clues about the control of TEs during early embryo and PGC development. The transition of ESCs to naïve pluripotency, their differentiation into PGC-like cells, as well as the conditional deletion of key enzymes, have all provided a window into the early events following DNA demethylation and how it affects TE control[86-89]. These studies have suggested that 5mC regulates many more TEs than can be appreciated from static knockout models, even if over time its silencing role is compensated by other mechanisms, such as the deposition of repressive histone marks[86] or the generation of endosiRNAs[89]. Such insights will be useful to guide future studies into the mechanisms underlying 5mC-coupled TE control *in vivo*.

[H2] Loss of TE 5mC in cancer. One of the hallmarks of cancer cells is global DNA hypomethylation, which has been proposed to promote oncogenesis[90] and has been generally associated with TE reactivation[91-93]. In particular, expression of LINE-1 proteins is a common feature of multiple human cancers, which increases the risk of insertional mutagenesis driven by LINE-1-mediated retrotransposition[94]. High rates of somatic LINE-1 retrotransposition have been mostly reported in epithelial tumours (e.g., colorectal, oesophageal, hepatocellular, lung and ovarian cancers), whereas haematological malignancies and gliomas have little evidence of retrotransposition[95,96]. Apart from increasing retrotransposition rates, 5mC loss in cancer activates dormant TE-encoded promoters[97]. Namely, activation of an antisense promoter within LINE-1 elements commonly generates chimeric transcripts **[G]** with host genes, such as in the case of the *MET* proto-oncogene[98,99]. LTR elements are also a source of oncogenic promoters, generating additional chimeric transcripts in various cancer types[97]. Treatment of cells with DNMT inhibitors (DNMTi), which are clinically used as therapeutic drugs in hematopoietic malignancies, leads to the emergence of numerous new TE chimeric transcripts that have the potential to affect cellular function[100].

In contrast to the developmental contexts discussed above, it would first appear that cancer cell populations are under no selective pressure to compensate for a loss of 5mC at TEs. Provided that TE insertions that compromise cancer growth are rare, tumours will easily tolerate (and may sometimes benefit from) TE mobility. However, a broad range of cancer types display overexpression of the ERV-repressing enzymes SETDB1 and LSD1, which appears to be the result of a clonal selection **[G]** process[101,102]. Depletion of either of these enzymes leads to the generation of ERV-derived double-stranded RNAs (dsRNAs), which are thought to activate antiviral response pathways that eventually lead to the apoptosis of cancer cells[101,102]. ERV silencing is therefore seemingly essential for cancer survival, although it remains to be established whether 5mC loss is a driver for the clonal evolution of alternative ERV-silencing mechanisms. Interestingly, DNMTi treatment also leads to an interferon response that has been associated with ERV derepression[103,104]. Activation of TEs may therefore ironically underlie the therapeutic efficacy of DNMTi treatment, which has opened exciting new therapeutic avenues that look to explore the pathways that induce an antiviral response through ERV activation[105].

[H1] Oxidation derivatives of 5mC

Replicating cells can efficiently remove 5mC through a passive mechanism that involves uncoupling of the 5mC maintenance machinery during DNA replication. However, 5mC can also be removed in an active, replication-independent manner, which in plants involves DNA glycosylases capable of removing the 5mC base directly[106] (Figure 3). By contrast, active DNA methylation in vertebrates first involves conversion of 5mC to 5hmC by an oxidation reaction (Figure 3). 5hmC was first identified in T-even bacteriophages as part of a mechanism to protect viral DNA from degradation by host restriction enzymes[107], and was later found in a number of vertebrate tissues[108]. However, it was the eventual discovery that TET enzymes oxidize 5mC to 5hmC[109], as well as to 5fC and 5caC[110,111], that drew widespread interest in this pathway as a mechanism for active DNA demethylation (Figure 3), although the generation of 5hmC also promotes replication-dependent demethylation[112,113]. Quantification of 5hmC levels in mice, amphibians and zebrafish have demonstrated that 5hmC abundance is highly variable among different tissues, and that strong enrichment of 5hmC in the central nervous system is conserved between species[114-116]. Although it is tempting to suggest a link between 5hmC abundance in the brain and the high activity of LINE-1 observed therein (Figure 4)[117], this remains to be tested. High levels of 5hmC are also found in ESCs and dramatically decrease upon differentiation[118]. Furthermore, 5hmC levels are inversely correlated with cell proliferation and are substantially lower in cancer cells compared to healthy tissues (Figure 4)[119]. Less is known about the distribution of 5caC and 5fC, which are 10–100 times less abundant than 5hmC. For example, in mouse ESCs 5hmC levels are around 0.4% of all cytosines (~10% of all 5mC) and only ~1% of those bases are further converted to 5fC/5caC, although there is substantial inter-tissue variation[110,120].

[H2] TETs as regulators of mammalian TE expression. TET enzymes emerged from a common metazoan ancestor and are part of the larger TET–JBP (J-binding protein) family that is also represented in several other species, including basidiomycete fungi[121]. Interestingly, some TET–JBP genes in fungi are encoded within KDZ (Kyakuja, Dileera and Zisupton) DNA transposons, which may have contributed to gene duplication and/or play a role in protein function[122]. These TET–JBP genes oxidize 5mC at repetitive elements, including Ty–*gypsy*-like TEs and, remarkably, the KDZ elements themselves[123]. Surprisingly, Ty3–*gypsy* retrotransposons were also found marked by 5hmC in rice cultivars, despite the absence of a known TET–JBP protein in plants[124].

In mammalian cells, TET enzymes bind *cis*-regulatory regions, such as gene promoters and enhancers, where their action implicates TETs in cell differentiation, neuronal function and oncogenesis[45]. However, accumulated evidence suggests that TET proteins also work as regulators of mammalian TE expression. Early profiling efforts in mouse ESCs showed that 5hmC is enriched at the 5′ untranslated region (UTR) of LINE-1 elements, and suggested that these TEs underwent TET-dependent removal of 5mC[125,126]. This potentially implicated TET enzymes in the demethylation of TEs during mouse preimplantation development. Notably, in zygotes **[G]** 5hmC occurs asymmetrically with respect to the two parental genomes in a mirror image to concomitant 5mC placement. Shortly after fertilization, the paternal genome undergoes rapid global loss of 5mC and gain of 5hmC, 5fC and 5caC, whereas 5mC levels on the maternal genome are largely maintained during the same period, and no accumulation of 5hmC is observed[127-129]. However, whereas 5hmC deposition on the paternal genome is dependent on oocyte-derived TET3, global loss of 5mC does not require its activity[128,129]. The demethylating role of TET3 appears to instead be restricted to a few genomic regions. Notably, TET3 depletion leads to an increase in 5mC levels at LINEs, SINEs and DNA transposons in the paternal pronucleus[128], whereas its action is less prominent at LTRs, especially in ERV1 and IAP classes[130]. This pattern coincides with the known preferred targets of demethylation in the zygote[77,131], which are associated with differences in H3K9me2/3 deposition, as discussed above. Despite the preferential activity of TET3 on the paternal pronucleus, TET3 also facilitates maternal DNA demethylation at repetitive regions, albeit to a lesser extent[130].

The above observations suggest that TET3-dependent demethylation underlies TE activation during epigenetic reprogramming. However, TET3 is dispensable for the activation of TEs (e.g., LINE-1, IAP, ERV1 and ERVL elements) in preimplantation embryos, irrespective of their dependency on this enzyme for demethylation[132]. Whereas simultaneous deletion of *Tet1* and *Tet3* is associated with a decrease in IAP expression in blastocysts, LINE-1 elements remain unaffected, despite visible increases in 5mC levels[36]. Similar observations have been made in ESCs, where depletion of the main TET enzymes expressed therein (TET1 and TET2) drives an increase in 5mC levels at evolutionarily young LINE-1 elements but does not lead to their silencing[37], despite the fact that these elements are relatively well expressed in ESCs when

compared with differentiated cells. These findings suggest that additional 5mC-independent mechanisms can couteract the effects of 5mC alterations at LINE-1 elements. Notably, TET enzymes can also serve as transcriptional co-activators or co-repressors in a catalytic-independent manner through interactions with transcriptional regulators, such as *O*-glcNAC transferase (OGT)[133-135], the SIN3A complex[136], and Polycomb repressive complex 2 (PRC2)[137] (Figure 5). Interestingly, TET1 recruits the SIN3A co-repressor complex to LINE-1 elements in ESCs, thus directly coupling active DNA demethylation to a repressive mechanism that ensures LINE-1 silencing[37]. Another candidate group of repressors that may counteract TET1 action at LINE-1s are methyl-CpG binding domain (MBD) proteins[38]. MERVL repression is also controlled by TET2-dependent recuitment of histone deacetylases (HDACs) and, remarkably, by oxidation of RNA 5mC (Figure 5; Box 1)[138].

Given that TET enzymes are expressed during epigenetic reprogramming, are specific mechanisms required to prevent active demethylation of highly mutagenic TEs, such as IAPs? One hypothesis is that the KAP1–SETDB1 pathway of H3K9me3 deposition protects certain TEs from TET activity. Consistent with this model, KAP1 binding and 5hmC deposition at LINE-1s appears to be mutually exclusive in both mouse and human ESCs, where 5hmC is enriched at evolutionarily young subfamilies, and KAP1 binds older elements[37,62]. However, SETDB1 depletion in mouse ESCs leads only to a subtle increase in 5hmC levels (and concomitant decrease in 5mC) at demethylation-resistant TEs, including IAPs[139]. Although the absence of SETDB1 enables TET2 to activate IAP expression in a catalytic-dependent manner, this appears to be the result of indirect effects[39]. These findings argue against a major role for H3K9me3 in protecting IAPs from TET-mediated demethylation. It is unclear whether other mechanisms minimize the activity and impact of TETs on reprogramming-resistant loci, and to what extent the findings in ESCs reflect the more dynamic *in vivo* state.

[H2] A direct role for 5hmC, 5fC or 5caC in TE regulation? Although oxidized forms of 5mC are normally considered simple intermediates of DNA demethylation, at least 5hmC and 5fC are predominantly stable *in vivo* and may play direct roles in transcriptional regulation[140,141]. One important consideration is whether the levels of each of these modifications are high enough to effect functional and phenotypic outcomes. Nevertheless, a localized accumulation of oxidized forms of 5mC could affect the accessibility of DNA-binding proteins by changing the physical properties of DNA. For instance, binding of the transcription factors Wilms tumour 1 (WT1) and early growth response protein 1 (EGR1) is impaired or even abolished by the presence of oxidized derivatives of 5mC (Figure 5)[142]. Furthermore, 5fC and 5caC impede transcription by binding to RNA polymerase II (Pol II), inducing Pol II pausing and delaying transcriptional elongation[143], although it remains unclear to what extent this impacts gene expression *in vivo*. Oxidized 5mC bases can also be recognized by specific 'readers' that subsequently direct chromatin organization and remodelling (Figure 5). DNA glycosylases (MPG and NEIL3), helicases (RECQ1) and the SALL4A transcription factor are among the 5hmC-specific readers, while 5fC and 5caC recruit a large number of DNA repair proteins, TP53, chromatin remodelling factors, and forkhead box transcription factors[144-146]. Despite the enticing enrichment of oxidized forms of 5mC at specific TE classes, it remains to be seen whether any of these mechanisms are involved in their transcriptional regulation.

[H1] *N*6-methyladenine

6mA is present across all kingdoms of life[30] and has recently received renewed attention due to a number of quantification, characterization and profiling efforts in several eukaryotic species, including mice and humans[31,32,147-150]. With 6mA levels varying widely across species (Figure 6), sensitive assays have been required to allow the detection of extremely low levels of this modification. The presence of such low 6mA levels has not only been questioned from a technical perspective[34,35], but also triggered a discussion about the potential functional importance of 6mA in these cases. Nevertheless, a number of different roles have been suggested for 6mA, with TE regulation being a recurrent theme across several species.

In bacteria, 6mA plays a key role as part of antiviral restriction-modification systems, but it has also been implicated in other processes, including transcriptional regulation[151]. Interestingly, 6mA regulates the Tn*10* DNA transposon, whose activity is dramatically increased in *Escherichia coli* strains mutant for the 6mA methyltransferase *dam*[152]. Several unicellular eukaryotes have also long been known to contain high levels of

6mA, including *Tetrahymena thermophila* and *Chlamydomonas reinhardtii*, wherein 6mA exists within an ApT-containing motif and is thought to play a role in nucleosome positioning at transcription start sites[147,153]. Species lacking an ApT motif at 6mA sites generally contain lower levels of this modification, and diverge with respect to its genomic distribution and putative roles (Figure 6). The evolution of different 6mA methyltransferases and demethylases (Figure 3) may have therefore conferred species-specific roles to this DNA modification. Notably, two plant pathogens (*Phytophthora infestans* and *Phytophthora sojae*) display an enrichment of 6mA at TEs and low-expressing genes, suggesting that it plays a role in TE silencing[154]. *A. thaliana* also displays 6mA enrichment at TEs, but in this case the mark correlates with host gene activation[155]. Determining whether 6mA plays a functional role in either or both of these contexts will be the first step to understanding how this apparent duality is achieved.

[H2] TEs and 6mA in metazoans. The levels of 6mA in metazoans are substantially lower than those seen in species such as *E. coli* and *T. thermophila* (Figure 6), thus placing a premium on having both sensitive and accurate detection systems for sequencing and biochemical detection methodologies[33]. This is an ongoing technical challenge, as a recent ultrasensitive mass spectrometry approach could not confirm the presence of 6mA in mouse ESCs[34]. Genome-wide mapping of 6mA using antibody-based enrichment techniques is also particularly prone to artefacts due to the high background resulting from the affinity of immunoglobulin G (IgG) for simple repeats[35].

With these caveats in mind, the identification and manipulation of 6mA methyltransferases and/or demethylases in some species have nonetheless provided clues to the putative roles of 6mA therein. In *C. elegans*, deletion of *N*6-methyl adenine demethylase 1 (*nmad-1*) leads to exacerbation of a transgenerationally-coupled infertility phenotype seen in *spr-5* (a H3K4me2 demethylase) mutants, although the underying mechanisms are unclear[148]. Interestingly, *D. melanogaster* contains a TET orthologue **[G]** that acts as a DNA 6mA demethylase (thus renamed *DMAD*), the deletion of which leads to a clear derepression of TEs[31]. *DMAD* is essential for *D. melanogaster* embryonic development and promotes germ cell differentiation[31]. Accordingly, levels of 6mA are developmentally regulated, with early embryonic stages displaying a 6mA content of ~0.07% of all adenines, which then declines to 0.001% in later stages, and a similar pattern is seen during early germ cell development. In ovaries, 6mA is found enriched at multiple TE families and this is more pronounced in *DMAD* mutants, where TE activation was observed. TE families that were both 6mA enriched and derepressed in *DMAD* mutants included LTR (*Idefix*, *Copia*) and non-LTR (*Het-A*, *Tart*) TEs[31]. Although it is tempting to implicate 6mA in fruitfly TE activation, further experiments will be needed to demonstrate that the action of *DMAD* at TEs is both direct and dependent on its catalytic activity.

A potential role of 6mA during development has also been suggested in pigs and zebrafish where, similar to *D. melanogaster*, high levels of 6mA have been reported during preimplantation stages of embryogenesis (with 6mA content peaking at 0.1–0.2% of all adenines)[156]. In zebrafish this is coupled to an enrichment of 6mA at repetitive elements, including LINE-1, LTR and DNA TEs[156]. 6mA has also been associated with TE regulation in the mouse, where knockout of the 6mA demethylase *Alkbh1* in ESCs leads to the accumulation of 6mA at evolutionarily young LINE-1 elements, concomitant with their silencing[32]. Similarly, 6mA levels at LINE-1s are raised in the prefrontal cortex of mice subjected to chronic stress, and this is correlated with LINE-1 silencing[157]. These findings contrast with those in fruitflies — particularly that 6mA is associated with TE repression in mice but TE activation in *D. melanogaster* — highlighting key differences in the associated mechansims across species, which remain unclear. Human lymphoblastoid cells have also been reported to display an enrichment of 6mA at the 5′ UTR of young LINE-1 elements[158], suggesting a potentially conserved function of this modification between mice and humans. Interestingly, cancer types that have been associated with high levels of LINE-1 retrotransposition, such as gastric and liver cancers, undergo 6mA loss[150]. Conversely, 6mA is markedly upregulated in glioblastoma[159], where LINE-1 activity has not been detected, raising the possibility that 6mA levels are linked to LINE-1 regulation in cancer. It will therefore be important to establish whether functional links between 6mA and TE deregulation exist in cancer that could drive LINE-1 mobility therein.

[H2] Relationship between 6mA and 5mC. Most species have one predominant form of DNA methylation (Figure 6), with this mutual exclusivity being particularly notable amongst early diverging fungi[33,160]. In the

special case of *C. reinhardtii*, which contains relatively high levels of both 5mC and 6mA, there is an interesting spatial compartmentalization between the two marks, with 5mC covering TEs and gene bodies, and 6mA marking the transcription start sites of active genes[147]. Given these observations, could 6mA have emerged in species that lack 5mC to provide the host genome with a comparable TE silencing tool? This could be the case for *Phytophthora* species, for example, although a functional role for 6mA in TE silencing has not been formally tested therein. However, the diversity in the distribution and apparent roles of 6mA across different species argue against such a simplistic model. For example, in *T. thermophila* 6mA is associated with the promoters of active genes[153], and in *D. melanogaster* increasing levels of 6mA correlate with TE activation[31]. In vertebrate genomes, given the high abundance of 5mC, it may seem unlikely that the putative silencing capabilities of 6mA have been harnessed to complement the action of 5mC. However, it has been proposed that during developmental epigenetic reprogramming in vertebrates, 6mA could compensate for the loss of 5mC and contribute to the maintenance of TE silencing and/or transcriptional regulation of genes[46]. Indeed, it is striking that the peak of 6mA deposition during preimplantation development of pigs and zebrafish coincides with a period of 5mC hypomethylation[156]. In this respect, it is also interesting to note that 6mA reportedly targets mainly young, highly active LINE-1s in mouse ESCs[32], which are the same subfamilies that undergo TET-mediated 5mC demethylation in these cells[37]. Thus, it is possible that 5mC and 6mA bear mechanistic links during epigenetic reprogramming that ensure an anti-correlation between the two marks. However, in *Alkbh1* knockout ESCs, increasing 6mA levels at LINE-1 elements are associated with a concomitant increase in 5mC[32], which does not support the hypothesis. It remains to be seen whether this relationship differs during epigenetic reprogramming.

[H1] Conclusions and perspectives

Advances in genomics and epigenomics have brought in a new era in the study of TE regulation and its impact on host genomes. This includes investigating the roles of an expanding repertoire of DNA modifications that are potentially far more widespread across species than previously thought. Using the regulatory signals provided by DNA modifications to control TEs seems to have been a commonly adopted strategy throughout evolution, albeit displaying intriguingly high variation across even closely related species. Comparative genomics and epigenomics efforts will continue to provide clues into the intricate relationships between TE evolution and that of DNA modifying enzymes. However, are DNA modifications themselves instructive for TE regulation? Can we uncouple the role of DNA modifying enzymes from that of the DNA modifications? Although the case for 5mC appears robust, it is also clear that 5mC-mediated TE silencing is context dependent. For other modifications the picture is far less clear, namely because in many genomes the levels of these modifications are so much lower than those of 5mC. Additionally, manipulating the expression of DNA modifying enzymes may lead to indirect effects on TEs, as we found recently for IAP regulation by TET2[39]. Known non-catalytic functions of DNMTs and TETs can also confound results and are rarely tested for. These considerations seem particularly relevant for the more recently described 6mA-modifying enzymes, whose effects on TEs may be unrelated to their ability to modulate 6mA levels. The advent of epigenetic editing tools offers the opportunity to tackle these questions by altering the levels of DNA modifications at specific loci while also controlling for catalytic effects[161].

It is now clear that the impact of TEs on genomes is dictated to a large extent by the regulatory activities that target them, including the action of DNA modifying enzymes. Dissecting the mechanisms underlying TE regulation is therefore pivotal to understanding how TEs contribute to genome evolution, development and disease. Encouragingly, our increased understanding of TE regulation will enable the exploration of novel therapeutic avenues for cancer that, for example, aim to activate an ERV-mediated antiviral response. As the role of modifications other than 5mC are further uncovered, the opportunities to explore these mechanisms for technological and clinical benefit will increase.

Box 1 | **The other nucleic acids: impact of RNA modifications on TE expression.**

To date more than 100 types of chemical modifications in RNA have been characterized, which are present in almost all types of RNA including mRNA, tRNA, rRNA, long non-coding RNA (lncRNA) and small nuclear RNA (snRNA)[162]. There is mounting evidence that functional roles of RNA modifications are as complex as those of DNA and chromatin, including associations with transposable element (TE) regulation.

[b1] *N*6-methyladenine

*N*6-methyladenine (6mA; commonly termed m^6A when in RNA) is the most abundant modification known on mRNA and lncRNA, and has been implicated in various aspects of mRNA metabolism including nuclear export, RNA structure, RNA stability and mRNA translation[163]. In *Arabidopsis thaliana* TE transcripts display relatively high abundance of 6mA, which appear to be linked to the maintenance of low levels of TE expression[164]. Using computational methods to map non-unique reads to specific TE classes, Zhang et al demonstrated that *Alu* elements are significantly enriched for 6mA[165], which may be involved in regulating *Alu* RNA conformation[166]. Circular RNAs (circRNAs), which form a covalently closed continuous loop, are also enriched with 6mA modifications[167]. Interestingly, TEs are significantly enriched in the flanking regions of 6mA-enriched circRNAs in both human embyonic stem cells and HeLa cells, suggesting that TE density may be associated with 6mA deposition in circRNAs[167].

[b1] Ribose methylation

2′-O-methylated nucleosides, which protect RNA from degradation, are found in all major classes of eukaryotic RNA and are the most abundant modifications of rRNA[162]. PIWI-interacting RNAs (piRNAs), which are master regulators of TEs in the germline, also undergo 2′-O-methylation by the HEN1 methyltransferase[168,169]. In the germline of *Drosophila melanogaster* and zebrafish, loss of HEN1 results in decreased piRNA levels and mild derepression of TEs, showing that piRNA methylation is crucial for its stabilization and transposon silencing[168,170].

[b1] Wobble uridine modifications in tRNA

The wobble uridine residues of tRNAs (U_{34}) are generally modified in all species[171]. Mutant forms of Elongator complex, which is involved in U_{34} modifications, lead to substantial downregulation of Ty1 retrotransposons and a small number of genes in the proximity of Ty elements in budding yeast[131]. However, overexpression of its target tRNAs does not rescue the effect of the mutation on expression of nearby genes, suggesting the involvement of tRNA modifications other than wobble nucleotide modifications[132].

[b1] 5mC and 5hmC

5-methylcytosine (5mC; commonly termed m^5C when in RNA) is highly abundant in tRNA and rRNAs and is catalysed by DNMT2 and NSUN proteins[163]. In *D. melanogaster* and *Dictyolstelium discoideum*, DNMT2 was shown to have a role in TE regulation[172,173]. Although this effect was attributed to DNA methylation rather than RNA methylation, others did not reproduce these observations[174]. Instead, a recent study suggested that the TE activation seen in *D. melanogaster Dnmt2* and *Nsun2* mutants is related to reduced tRNA stability, indicating a potential link between tRNA modifiers and TE regulation[175]. A role in TE regulation has also been proposed for TET-mediated oxidation of RNA 5mC into 5-hydroxymethylcytosine (5hmC)[138]. In mouse embryonic stem cells, TET2 does not modify MERVL DNA but it does increase 5hmC levels at MERVL transcripts, which correlates with their destablilization[138].

Figure legends

Figure 1 | **Transposable element classification and species distribution.** **a** | Transposable elements (TEs) are classified into retrotransposons (class I), which are defined by their use of a reverse transcribed RNA intermediate, and DNA transposons (class II). TEs can be further subdivided into orders (e.g., long terminal repeat (LTR), long interspersed nuclear element (LINE) and short interspersed nuclear element (SINE)) and superfamilies (e.g., endogenous retrovirus (ERV), *Copia* and LINE-1), following the classification system proposed by Wicker et al[9]. The TE order is defined based on key differences in the transposition mechanism. For example, LTRs use an integrase to insert TE cDNA into the genome, whereas

LINE elements use an endonuclease to perform target-primed reverse transcription at the insertion locus. **b** | The genomic abundance and composition of TEs (pie charts; annotation from RepeatMasker) varies widely across species. This is also reflected in the extent and diversity of TE activity, for which examples of transposition-competent families are provided. For example, LINEs and SINEs are highly abundant in human and mouse, where they are also the main active TE orders, despite a large fraction of the genome also containing LTR and DNA transposons. By contrast, DNA transposons are the main TE class in *Caenorhabditis elegans*, where they remain active. *A. thaliana*, *Arabidopsis thaliana*; APE, apurinic endonuclease; ATP, packaging ATPase; C-INT, c-integrase; CYP, cysteine protease; *D. melanogaster*, *Drosophila melanogaster*; DIRS, *Dictyostelium* intermediate repeat sequence; EN, endonuclease; EVD, Evadé; *H. sapiens*, *Homo sapiens*; HEL, helicase; IAP, intracisternal A particle; ITR, inverted terminal repeat; *M. musculus*, *Mus musculus*; MLV, murine leukaemia virus; ORF, open reading frame; POL B, DNA polymerase B; REP, replication initiator protein; RH, RNase H; RT, reverse transcriptase; RTE, retrotransposabe element; SVA, an element formed of fragments of SINEs, variable number tandem repeats (VNTRs) and Alu elements; TIR, terminal inverted repeat; TPase, transposase; TR, terminal repeat (variable structure); YR, tyrosine recombinase.

Figure 2 | **Interplay between key transposon silencing mechanisms and DNA methylation.** **a** | In the animal germline, transposable element (TE) mRNAs are cleaved to PIWI-interacting RNAs (piRNAs) by MILI and MIWI in mouse, PIWI and Aubergine (AUB) in *Drosophila melanogaster* and PRG1 in *Caenorhabditis elegans*, followed by post-transcriptional gene silencing **[G]** (PTGS). Transcriptional gene silencing **[G]** (TGS) of TEs via epigenetic modifications is achieved by piRNA-loaded MIWI2 (mouse) conferring *de novo* 5-methylcytosine (5mC)[65,68], and piRNA-loaded PIWI in *D. melanogaster* (WAGO and CSR1 in *C. elegans*), resulting in H3K9me2/3 enrichment (reviewed in REF[21]). **b** | In flowering plants, RNA-dependent DNA methylation **[G]** (RdDM) begins by RDR6 making transcribed TE RNAs double stranded, followed by DCL2- and DCL4-mediated cleavage to 21–22-nucleotide small RNAs (sRNAs). DRM1 and DRM2 generate 5mC at TE loci, resulting in silencing[23]. RdDM is accompanied by PTGS mediated by Argonaute 1 (AGO1) and AGO2 loaded with sRNAs. In rice and *Arabidopsis thaliana* germlines, DEMETER family glycosylases remove DNA methylation from TEs in gamete companion cells[176] leading to TE expression. At least in pollen, this leads to RdDM-mediated TE silencing in the adjacent sperm cell[176] **c** | In early embryogenesis, Krüppel-associated box (KRAB) proteins containing zinc fingers (ZNFs) recognize TEs in a sequence-specific manner[25], recruiting the KAP1 (also known as TRIM28) cofactor. KAP1 provides a scaffold for the SET domain bifurcated 1 (SETDB1) H3K9 methyltransferase and heterochromatin protein 1 (HP1), forming a repressive chromatin structure [58,60], followed by *de novo* DNA methylation[177]. **d** | TE DNA methylation in plants is reinforced by RdDM, similarly to part **b**, except that RNA polymerase IV (POL IV) and POL V can transcribe methylated DNA. POL IV transcripts are processed to 24-nucleotide sRNAs by RDR2 and DCL3 and are loaded onto AGO4 and AGO6. POL V TE transcription provides a scaffold for AGO4 and AGO6 binding, which then recruits DRM1 and DRM2, leading to DNA methylation in all sequence contexts[23]. **e** | Asymmetric DNA methylation is maintained by chromomethylase 2 (CMT2), and at heterochromatin, DDM1 remodels chromatin to allow methyltransferase access[28] **f** | TE methylation in mammals is maintained following replication by DNA methyltransferase 1 (DNMT1) and UHRF1 action on hemimethylated DNA. During reprogramming, H3K9me3-enriched intracisternal A particles (IAPs) and long interspersed nuclear element 1 (LINE-1) elements are protected from passive demethylation, probably by UHRF1 recruitment[84].

Figure 3 | **DNA modifications and DNA modifying enzymes.** Chemical structure of unmodified and modified DNA bases and their respective DNA modifiers in different species. See also Figure 6 for the levels of 5-methylcytosine (5mC) and *N*6-methyladenine (6mA) in different species, which are associated with the evolution of DNA modifying enzymes. Cytosine is methylated by DNA methyltransferase (DNMT) enzymes, which fall into different major families (shown are DNMT1, DNMT3 and chromomethylase (CMT) families) [47]. In *Arabidopsis thaliana,* 5mC can be directly removed by glycosylases, followed by repair via the base excision repair (BER) pathway[106]. In metazoans, 5mC can be oxidized into 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) via the TET–JBP family of enzymes[109-111]. Thymine DNA glycosylase (TDG) can remove 5fC and 5caC bases, triggering base excision repair (BER). All modified cytosine bases can also be removed passively through replication-coupled demethylation[112,113].

6mA methyltransferases (MT) and demethylases (DM) have started to be identified in higher eukaryotes[31,32,150]. *C. elegans*, *Caenorhabditis elegans*; DME/ROS1, DEMETER and REPRESSOR OF SILENCING 1 family of 5mC DNA glycosylases; *D. melanogaster*, *Drosophila melanogaster*; *H. sapiens*, *Homo sapiens*; *M. musculus*, *Mus musculus*.

Figure 4 | **Dynamics of 5mC, 5hmC and LINE-1 expression during mouse development.** During mouse development, the genome undergoes two waves of epigenetic reprogramming, leading to rapid loss of 5-methylcytosine (5mC) during primordial germ cell (PGC) migration, and immediately following fertilization. During these periods, class-dependent DNA methylation is observed at transposable elements (TEs): whereas long interspersed nuclear element 1 (LINE-1) elements undergo a complete erasure of 5mC, certain endogenous retroviruses (ERVs) such as intracisternal A particles (IAPs) are resistant to DNA demethylation[63,64]. After PGC reprogramming, the establishment of DNA methylation takes place at different times in males and females: in male PGCs, methylation initiates in prospermatogonia, whereas re-methylation of the oocyte genome only occurs after birth, opening a window of opportunity for TE activation. Indeed, LINE-1s and mammalian apparent LTR retrotransposons (MaLRs) are activated in mouse oocytes[75]. The second wave of 5mC loss, during preimplantation, coincides with transient accumulation of 5-hydroxymethylcytosine (5hmC) levels and is characterized by high TE expression levels, including LINE-1. High 5hmC levels are also observed in the brain, where LINE-1s are also expressed[117], but most other somatic tissues are largely depleted for 5hmC and LINE-1 activity. Finally, in cancer cells, global DNA hypomethylation is a ubiquitous feature and 5mC loss at TEs is highly correlated with their activation in cancer[91-93].

Figure 5 | **Potential mechanisms for TET-mediated regulation of TEs.** Biochemical studies and cell culture experiments (mainly on embryonic stem cells) have revealed different modes by which TET proteins can regulate transcription, thus potentially affecting transposable element (TE) expression. The top half of the figure refers to mechanisms with documented examples in the regulation of TE expression, whereas the bottom half presents mechanisms known to act at other loci but with no reported role in TE regulation. First and foremost, TETs are thought of as DNA demethylases, and have been shown to decrease 5-methylcytosine (5mC) levels at long interspersed nuclear element 1 (LINE-1) elements, intracisternal A particles (IAPs), short interspersed nuclear elements (SINEs) and DNA transposons[36-38,128]. TETs can also modify RNA and destabilize TE transcripts such as murine endogenous retrovirus L (*MERVL*)[138]. Additionally, TET proteins play non-catalytic roles in TE regulation through the recruitment of repressor complexes such as SIN3A, MBD and HDAC[37,38 138], as well as potential co-activators such as OGT[37]. The modifications generated by TETs (5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC)) can themselves affect downstream pathways, by either preventing binding of particular proteins, such as transcription factors (TFs)[142], or by recruiting modification-specific readers[144-146] — it is unknown whether these mechanisms act at TEs.

Figure 6 | **Variation in 6mA abundance and relationship with TEs across species.** Levels of *N*6-methyladenine (6mA) vary widely across species and also show spatial and temporal differences, such as between the micronucleus (MIC) and macronucleus (MAC) of *Tetrahymena thermophila*, and the vegetative and reproductive phases of *Arabidopsis thaliana* development. Generally, high levels of 6mA correlate with low levels of 5mC and are associated with a palindromic ApT motif that enables replication-coupled maintenance of the modification. Documented relationships with transposable elements (TE) are described, including the enrichment of 6mA at TEs, or more direct evidence involving the manipulation of 6mA-modifying enzymes, as in the case of *Escherichia coli*, *Drosophila melanogaster* and mouse (*Mus musculus*). For some species a potential role in TE regulation has not been formally investigated. *C. anguillulae*, *Catenaria anguillulae*; *C. elegans*, *Caenorhabditis elegans*; *C. reinhardtii*, *Chlamydomonas reinhardtii*; *D. rerio*, *Danio rerio*; *H. sapiens*, *Homo sapiens*; *H. vesiculosa*, *Hesseltinella vesiculosa*; *P. infestans*, *Phytophthora infestans*; *X. laevis*, *Xenopus laevis*.

Glossary

Genetic drift
Changes in the frequency of a given allele in a population due to random sampling. Genetic drift can lead to the fixation of a particular allelic variant in a population without any selective pressure.

Horizontal propagation
Better known as horizontal gene transfer, it entails the transfer of genetic material between organisms. It contrasts with vertical transfer, which occurs from parents to offspring via the germline.

PIWI-interacting RNAs
(piRNAs). A class of 26–30-nucleotide small, single-stranded RNAs that interact with the PIWI family of proteins.

Pre-implantation development
The first phase of embryonic development that begins after fertilization and ends upon implantation of the blastocyst into the uterus.

endosiRNAs
Small RNAs (20–23 nucleotides) generated from double-stranded RNAs, including sense–antisense transcript hybrids.

Primordial germ cells
(PGCs). The precursor cells of mammalian gametes that are specified at around embryonic day 6.25 in mice, and that differentiate into oocytes or sperm.

Naïve pluripotency
A stem cell state that resembles that of the inner cell mass of the blastocyst.

Chimeric transcripts
In the context of this Review, chimeric transcripts are RNA molecules that involve a fusion between a transposable element acting as a transcriptional promoter and a host gene.

Clonal selection
In the context of cancer evolution, clonal selection entails the selective expansion of a particular cell due to genetic and/or epigenetic changes that confer a growth advantage.

Zygotes
One-cell embryos resulting from the fusion of sperm with an oocyte, i.e., fertilization.

Orthologue
Orthologues are genes from different species that have evolved from a common ancestor.

Post-transcriptional gene silencing
(PTGS). The process of silencing a gene after it has been transcribed, e.g., by cleavage of its nascent RNA.

Transcriptional gene silencing
(TGS). Silencing a gene at the transcriptional level, i.e., by preventing the transcriptional process, often by epigenetic modification of the locus to a less open conformation, disfavouring binding of RNA Pol II.

RNA-dependent DNA methylation
(RdDM). One of the key strategies for *de novo* and maintenance DNA methylation in *Arabidopsis thaliana*, whereby RNA molecules from expressed loci direct DNA methylation in a sequence-dependent manner.

References

1. Gregory, T. R. Coincidence, coevolution, or causation? DNA content, cell size, and the C-

value enigma. *Biol Rev Camb Philos Soc* **76,** 65–101 (2001).
2. Jurka, J., Bao, W. & Kojima, K. K. Families of transposable elements, population structure and the origin of species. *Biol. Direct* **6,** 44 (2011).
3. Sotero-Caio, C. G., Platt, R. N., Suh, A. & Ray, D. A. Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol Evol* **9,** 161–177 (2017).
4. Feschotte, C. & Betrán, E. Transposable Element Domestication As an Adaptation to Evolutionary Conflicts. *Trends Genet* **33,** 817–831 (2017).
5. Joly-Lopez, Z. & Bureau, T. E. Exaptation of transposable element coding sequences. *Current Opinion in Genetics & Development* **49,** 34–42 (2018).
6. Chuong, E. B., Elde, N. C. & Feschotte, C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* **18,** 71–86 (2017).
7. Arkhipova, I. R. Neutral Theory, Transposable Elements, and Eukaryotic Genome Evolution. *Molecular Biology and Evolution* **35,** 1332–1337 (2018).
8. Gilbert, C. & Feschotte, C. Horizontal acquisition of transposable elements and viral sequences: patterns and consequences. *Current Opinion in Genetics & Development* **49,** 15–24 (2018).
9. Wicker, T. *et al.* A unified classification system for eukaryotic transposable elements. *Nat Rev Genet* **8,** 973–982 (2007).

A comprehensive description of TE classification and nomenclature, based on a combination of TE sequence structure, phylogeny and mechanisms of transposition.

10. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA* **6,** 11 (2015).
11. Kojima, K. K. Human transposable elements in Repbase: genomic footprints from fish to humans. *Mob DNA* **9,** 2 (2018).
12. Feschotte, C. & Pritham, E. J. DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41,** 331–368 (2007).
13. Rodriguez-Terrones, D. & Torres-Padilla, M.-E. Nimble and Ready to Mingle: Transposon Outbursts of Early Development. *Trends Genet* **34,** 806–820 (2018).
14. Tsukahara, S. *et al.* Bursts of retrotransposition reproduced in Arabidopsis. *Nature* **461,** 423–426 (2009).
15. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409,** 860–921 (2001).
16. Mouse Genome Sequencing Consortium *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520–562 (2002).
17. Richardson, S. R. *et al.* Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Research* **27,** 1395–1405 (2017).
18. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc Natl Acad Sci USA* **100,** 5280–5285 (2003).
19. Dewannieux, M., Esnault, C. & Heidmann, T. LINE-mediated retrotransposition of marked Alu sequences. *Nat Genet* **35,** 41–48 (2003).
20. Hancks, D. C. & Kazazian, H. H. Roles for retrotransposon insertions in human disease. *Mob DNA* **7,** 9 (2016).
21. Czech, B. & Hannon, G. J. One Loop to Rule Them All: The Ping-Pong Cycle and piRNA-Guided Silencing. *Trends in Biochemical Sciences* **41,** 324–337 (2016).
22. Molaro, A. & Malik, H. S. Hide and seek: how chromatin-based pathways silence retroelements in the mammalian germline. *Current Opinion in Genetics & Development* **37,** 51–58 (2016).
23. Kim, M. Y. & Zilberman, D. DNA methylation as a system of plant genomic immunity.

Trends Plant Sci. **19,** 320–326 (2014).
24. Jacobs, F. M. J. *et al.* An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* **516,** 242–245 (2014).
25. Imbeault, M., Helleboid, P.-Y. & Trono, D. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543,** 550–554 (2017).
26. Rowe, H. M. & Trono, D. Dynamic control of endogenous retroviruses during development. *Virology* **411,** 273–287 (2011).
27. Dunican, D. S. *et al.* Lsh regulates LTR retrotransposon repression independently of Dnmt3b function. *Genome Biol.* **14,** R146 (2013).
28. Zemach, A. *et al.* The Arabidopsis nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153,** 193–205 (2013).
29. Yoder, J. A., Walsh, C. P. & Bestor, T. H. Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet* **13,** 335–340 (1997).
30. Ratel, D., Ravanat, J.-L., Berger, F. & Wion, D. N6-methyladenine: the other methylated base of DNA. *Bioessays* **28,** 309–315 (2006).
31. Zhang, G. *et al.* N6-methyladenine DNA modification in Drosophila. *Cell* **161,** 893–906 (2015).

This study describes 6mA dynamics during *D. melanogaster* embryogenesis and reports a correlation between 6mA demethylation and TE suppression.

32. Wu, T. P. *et al.* DNA methylation on N(6)-adenine in mammalian embryonic stem cells. *Nature* **532,** 329–333 (2016).

The first paper to find 6mA in mammalian genomes, identifying both 6mA, and its associated demethylase in mouse ESCs, which when removed led to 6mA enrichment at young LINE-1 elements.

33. O'Brown, Z. K. & Greer, E. L. N6-Methyladenine: A Conserved and Dynamic DNA Mark. *Adv. Exp. Med. Biol.* **945,** 213–246 (2016).
34. Schiffers, S. *et al.* Quantitative LC-MS Provides No Evidence for m6dA or m4dC in the Genome of Mouse Embryonic Stem Cells and Tissues. *Angewandte Chemie (International ed in English)* **56,** 11268–11271 (2017).
35. Lentini, A. *et al.* A reassessment of DNA-immunoprecipitation-based genomic profiling. *Nat. Methods* **15,** 499–504 (2018).
36. Kang, J. *et al.* Simultaneous deletion of the methylcytosine oxidases Tet1 and Tet3 increases transcriptome variability in early embryogenesis. *Proc Natl Acad Sci USA* **112,** E4236–45 (2015).
37. la Rica, de, L. *et al.* TET-dependent regulation of retrotransposable elements in mouse embryonic stem cells. *Genome Biol.* **17,** 234 (2016).

In this paper the authors show that TET enzymes demethylate LINE-1 elements in ESCs, but also recruit the co-repressor SIN3A to ensure LINE-1 silencing.

38. Zhang, P. *et al.* L1 retrotransposition is activated by Ten-eleven-translocation protein 1 and repressed by methyl-CpG binding proteins. *Nucleus* **8,** 548–562 (2017).
39. Deniz, O., la Rica, de, L., Cheng, K. C. L., Spensberger, D. & Branco, M. R. SETDB1 prevents TET2-dependent activation of IAP retroelements in naïve embryonic stem cells. *Genome Biol.* **19,** 6 (2018).
40. Coluccio, A. *et al.* Individual retrotransposon integrants are differentially controlled by KZFP/KAP1-dependent histone methylation, DNA methylation and TET-mediated hydroxymethylation in naïve embryonic stem cells. *Epigenetics & chromatin* **11,** 7 (2018).
41. Du, J., Johnson, L. M., Jacobsen, S. E. & Patel, D. J. DNA methylation pathways and their crosstalk with histone methylation. *Nat Rev Mol Cell Biol* **16,** 519–532 (2015).
42. Schübeler, D. Function and information content of DNA methylation. *Nature* **517,** 321–326 (2015).

43. Allis, C. D. & Jenuwein, T. The molecular hallmarks of epigenetic control. *Nat Rev Genet* **17,** 487–500 (2016).
44. Rasmussen, K. D. & Helin, K. Role of TET enzymes in DNA methylation, development, and cancer. *Genes & Development* **30,** 733–750 (2016).
45. Wu, X. & Zhang, Y. TET-mediated active DNA demethylation: mechanism, function and beyond. *Nat Rev Genet* **18,** 517–534 (2017).
46. Luo, G.-Z. & He, C. DNA N6-methyladenine in metazoans: functional epigenetic mark or bystander? *Nat Struct Mol Biol* **24,** 503–506 (2017).
47. Jeltsch, A. Molecular biology. Phylogeny of methylomes. *Science* **328,** 837–838 (2010).
48. Lechner, M. *et al.* The correlation of genome size and DNA methylation rate in metazoans. *Theory Biosci.* **132,** 47–60 (2013).
49. Rošić, S. *et al.* Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. *Nat Genet* **50,** 452–459 (2018).
50. Lippman, Z., May, B., Yordan, C., Singer, T. & Martienssen, R. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *Plos Biol* **1,** E67 (2003).
51. Hosaka, A. *et al.* Evolution of sequence-specific anti-silencing systems in Arabidopsis. *Nature Communications* **8,** 2161 (2017).
52. Zhou, Y., Cambareri, E. B. & Kinsey, J. A. DNA methylation inhibits expression and transposition of the Neurospora Tad retrotransposon. *Mol. Genet. Genomics* **265,** 748–754 (2001).
53. Chernyavskaya, Y. *et al.* Loss of DNA methylation in zebrafish embryos activates retrotransposons to trigger antiviral signaling. *Development* **144,** 2925–2939 (2017).
54. Walsh, C. P., Chaillet, J. R. & Bestor, T. H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20,** 116–117 (1998).

This study is the first to demonstrate the role of DNA methylation in the silencing of TEs (IAPs) in mouse development.

55. Hutnick, L. K. *et al.* DNA hypomethylation restricted to the murine forebrain induces cortical degeneration and impairs postnatal neuronal maturation. *Human Molecular Genetics* **18,** 2875–2888 (2009).
56. Jackson-Grusby, L. *et al.* Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat Genet* **27,** 31–39 (2001).
57. Hutnick, L. K., Huang, X., Loo, T.-C., Ma, Z. & Fan, G. Repression of retrotransposal elements in mouse embryonic stem cells is primarily mediated by a DNA methylation-independent mechanism. *J Biol Chem* **285,** 21082–21091 (2010).
58. Matsui, T. *et al.* Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464,** 927–931 (2010).
59. Karimi, M. M. *et al.* DNA Methylation and SETDB1/H3K9me3 Regulate Predominantly Distinct Sets of Genes, Retroelements, and Chimeric Transcripts in mESCs. *Cell Stem Cell* **8,** 676–687 (2011).

This paper demonstrates that DNA methylation and H3K9me3 are targeted to different loci, and that SETDB1-mediated H3K9me3 enrichment contributes to silencing of certain ERVs in mouse ESCs.

60. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463,** 237–240 (2010).
61. Fasching, L. *et al.* TRIM28 represses transcription of endogenous retroviruses in neural progenitor cells. *CellReports* **10,** 20–28 (2015).
62. Castro-Diaz, N. *et al.* Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes & Development* **28,** 1397–1409 (2014).
63. Molaro, A. *et al.* Two waves of de novo methylation during mouse germ cell development.

Genes & Development **28,** 1544–1549 (2014).

64. Fadloun, A. *et al.* Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat Struct Mol Biol* **20,** 332–338 (2013).

This study reveals the dynamic nature of TE expression during mouse preimplantation underlining a transient expression of LINE-1s during this period.

65. Aravin, A. A. *et al.* A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol. Cell* **31,** 785–799 (2008).

66. Hackett, J. A. *et al.* Promoter DNA methylation couples genome-defence mechanisms to epigenetic reprogramming in the mouse germline. *Development* **139,** 3623–3632 (2012).

67. Bourc'his, D. & Bestor, T. H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96–99 (2004).

A seminal paper showing *in vivo* that DNA methylation is required for transposon silencing during spermatogenesis in mice.

68. Manakov, S. A. *et al.* MIWI2 and MILI Have Differential Effects on piRNA Biogenesis and DNA Methylation. *CellReports* **12,** 1234–1243 (2015).

69. Barau, J. *et al.* The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* **354,** 909–912 (2016).

This study discovers DNMT3C, a fourth DNA methyltransferase enzyme that specifically methylates young TEs in the male germline.

70. Jain, D. *et al.* rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline. *PLoS Genetics* **13,** e1006964 (2017).

71. Zamudio, N. *et al.* DNA methylation restrains transposons from adopting a chromatin signature permissive for meiotic recombination. *Genes & Development* **29,** 1256–1270 (2015).

In this paper, the authors show that TE silencing during spermatogenesis is required during meiosis due to an aberrant chromatin structure formed at expressed TE loci, which form meiotic hotspots.

72. Murchison, E. P. *et al.* Critical roles for Dicer in the female germline. *Genes & Development* **21,** 682–693 (2007).

73. Tam, O. H. *et al.* Pseudogene-derived small interfering RNAs regulate gene expression in mouse oocytes. *Nature* **453,** 534–538 (2008).

74. Kabayama, Y. *et al.* Roles of MIWI, MILI and PLD6 in small RNA regulation in mouse growing oocytes. *Nucleic Acids Research* **45,** 5387–5398 (2017).

75. Malki, S., van der Heijden, G. W., O'Donnell, K. A., Martin, S. L. & Bortvin, A. A Role for Retrotransposon LINE-1 in Fetal Oocyte Attrition in Mice. *Developmental Cell* **29,** 521–533 (2014).

76. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48,** 849–862 (2012).

In this paper, the authors describe global DNA methylation dynamics in mouse PGCs, featuring DNA methylation resistant genomic regions, including IAPs, ERV1 and ERVK families.

77. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *genesis* **35,** 88–93 (2003).

78. Kobayashi, H. *et al.* High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Research* **23,** 616–627 (2013).

79. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339–344 (2012).

80. Liu, S. *et al.* Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes & Development* **28,**

2041–2055 (2014).
This study identifies SETDB1 as responsible for silencing of DNA-demethylation-resistant TEs in PGCs.
81. Habibi, E. *et al.* Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. *Cell Stem Cell* **13,** 360–369 (2013).
82. Meyenn, von, F. *et al.* Impairment of DNA Methylation Maintenance Is the Main Cause of Global Demethylation in Naive Embryonic Stem Cells. *Mol. Cell* **62,** 848–861 (2016).
This study shows that replication-dependent passive demethylation is the dominant process during the remodelling of ESC to a naïve state. The authors also link H3K9me2 enrichment with UHRF1 recruitment
83. Rothbart, S. B. *et al.* Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* **19,** 1155–1160 (2012).
84. Liu, X. *et al.* UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nature Communications* **4,** 1563 (2013).
85. Maenohara, S. *et al.* Role of UHRF1 in de novo DNA methylation in oocytes and maintenance methylation in preimplantation embryos. *PLoS Genet.* **13,** e1007042 (2017).
86. Walter, M., Teissandier, A., Pérez-Palacios, R. & Bourc'his, D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* **5,** (2016).
87. Meyenn, von, F. *et al.* Comparative Principles of DNA Methylation Reprogramming during Human and Mouse In Vitro Primordial Germ Cell Specification. *Developmental Cell* **39,** 104–115 (2016).
88. Sharif, J. *et al.* Activation of Endogenous Retroviruses in Dnmt1(-/-) ESCs Involves Disruption of SETDB1-Mediated Repression by NP95 Binding to Hemimethylated DNA. *Cell Stem Cell* **19,** 81–94 (2016).
89. Berrens, R. V. *et al.* An endosiRNA-Based Repression Mechanism Counteracts Transposon Activation during Global DNA Demethylation in Embryonic Stem Cells. *Cell Stem Cell* **21,** 694–703.e7 (2017).
90. Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300,** 489–492 (2003).
91. Iskow, R. C. *et al.* Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* **141,** 1253–1261 (2010).
92. Schauer, S. N. *et al.* L1 retrotransposition is a common feature of mammalian hepatocarcinogenesis. *Genome Research* **28,** 639–653 (2018).
93. Nguyen, T. H. M. *et al.* L1 Retrotransposon Heterogeneity in Ovarian Tumor Cell Evolution. *CellReports* **23,** 3730–3740 (2018).
94. Rodić, N. *et al.* Long interspersed element-1 protein expression is a hallmark of many human cancers. *Am. J. Pathol.* **184,** 1280–1286 (2014).
95. Lee, E. *et al.* Landscape of somatic retrotransposition in human cancers. *Science* **337,** 967–971 (2012).
This study provides a detailed overview of somatic TE retrotransposition activity in different types of cancer.
96. Burns, K. H. Transposable elements in cancer. *Nat Rev Genet* **17,** 415–424 (2017).
97. Babaian, A. & Mager, D. L. Endogenous retroviral promoter exaptation in human cancer. *Mob DNA* **7,** 24 (2016).
98. Weber, B., Kimhi, S., Howard, G., Eden, A. & Lyko, F. Demethylation of a LINE-1 antisense promoter in the cMet locus impairs Met signalling through induction of illegitimate transcription. *Oncogene* **29,** 5775–5784 (2010).
99. Cruickshanks, H. A. & Tufarelli, C. Isolation of cancer-specific chimeric transcripts induced by hypomethylation of the LINE-1 antisense promoter. *Genomics* **94,** 397–406 (2009).

100. Brocks, D. *et al.* DNMT and HDAC inhibitors induce cryptic transcription start sites encoded in long terminal repeats. *Nat Genet* **49,** 1052–1060 (2017).
101. Cuellar, T. L. *et al.* Silencing of retrotransposons by SETDB1 inhibits the interferon response in acute myeloid leukemia. *The Journal of Cell Biology* **216,** 3535–3549 (2017).
102. Sheng, W. *et al.* LSD1 Ablation Stimulates Anti-tumor Immunity and Enables Checkpoint Blockade. *Cell* **174,** 549–563.e19 (2018).
103. Roulois, D. *et al.* DNA-Demethylating Agents Target Colorectal Cancer Cells by Inducing Viral Mimicry by Endogenous Transcripts. *Cell* **162,** 961–973 (2015).
This is the first study reporting that tumour-suppressive strategies of DNA-demethylating agents are actually via an interferon response associated with ERV activation.
104. Chiappinelli, K. B. *et al.* Inhibiting DNA Methylation Causes an Interferon Response in Cancer via dsRNA Including Endogenous Retroviruses. *Cell* **162,** 974–986 (2015).
105. Ohtani, H., Liu, M., Zhou, W., Liang, G. & Jones, P. A. Switching roles for DNA and histone methylation depend on evolutionary ages of human endogenous retroviruses. *Genome Research* 1–12 (2018). doi:10.1101/gr.234229.118
106. Li, Y., Kumar, S. & Qian, W. Active DNA demethylation: mechanism and role in plant development. *Plant Cell Rep.* **37,** 77–85 (2018).
107. WYATT, G. R. & COHEN, S. S. The bases of the nucleic acids of some bacterial and animal viruses: the occurrence of 5-hydroxymethylcytosine. *Biochem J* **55,** 774–782 (1953).
108. Penn, N. W., Suwalski, R., O'Riley, C., Bojanowski, K. & Yura, R. The presence of 5-hydroxymethylcytosine in animal deoxyribonucleic acid. *Biochem J* **126,** 781–790 (1972).
109. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324,** 930–935 (2009).
This study discovers that TET proteins catalyse the conversion of 5mC to 5hmC by an oxidation reaction.
110. Ito, S. *et al.* Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* **333,** 1300–1303 (2011).
111. He, Y.-F. *et al.* Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* **333,** 1303–1307 (2011).
112. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334,** 194–194 (2011).
113. Hashimoto, H. *et al.* Recognition and potential mechanisms for replication and erasure of cytosine hydroxymethylation. *Nucleic Acids Research* **40,** 4841–4849 (2012).
114. Globisch, D. *et al.* Tissue distribution of 5-hydroxymethylcytosine and search for active demethylation intermediates. *PLoS ONE* **5,** e15367 (2010).
115. Almeida, R. D. *et al.* Semi-quantitative immunohistochemical detection of 5-hydroxymethyl-cytosine reveals conservation of its tissue distribution between amphibians and mammals. *Epigenetics* **7,** 137–140 (2012).
116. Kamstra, J. H., Løken, M., Aleström, P. & Legler, J. Dynamics of DNA hydroxymethylation in zebrafish. *Zebrafish* **12,** 230–237 (2015).
117. Upton, K. R. *et al.* Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* **161,** 228–239 (2015).
118. Szwagierczak, A., Bultmann, S., Schmidt, C. S., Spada, F. & Leonhardt, H. Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Research* **38,** e181 (2010).
119. Jin, S.-G. *et al.* 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res* **71,** 7360–7365 (2011).
120. Pfaffeneder, T. *et al.* The discovery of 5-formylcytosine in embryonic stem cell DNA. *Angewandte Chemie (International ed in English)* **50,** 7008–7012 (2011).

121. Iyer, L. M., Tahiliani, M., Rao, A. & Aravind, L. Prediction of novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids. *Cell Cycle* **8,** 1698–1710 (2009).
122. Iyer, L. M. *et al.* Lineage-specific expansions of TET/JBP genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes. *Proc Natl Acad Sci USA* **111,** 1676–1683 (2014).
123. Chavez, L. *et al.* Simultaneous sequencing of oxidized methylcytosines produced by TET/JBP dioxygenases in Coprinopsis cinerea. *Proc Natl Acad Sci USA* **111,** E5149–58 (2014).
124. Wang, X.-L. *et al.* Genome-wide mapping of 5-hydroxymethylcytosine in three rice cultivars reveals its preferential localization in transcriptionally silent transposable element genes. *J. Exp. Bot.* **66,** 6651–6663 (2015).
125. Ficz, G. *et al.* Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473,** 398–402 (2011).
126. Booth, M. J. *et al.* Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. *Science* **336,** 934–937 (2012).
127. Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res* **21,** 1670–1676 (2011).
128. Gu, T.-P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477,** 606–610 (2011).
129. Amouroux, R. *et al.* De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18,** 225–233 (2016).
130. Shen, L. *et al.* Tet3 and DNA replication mediate demethylation of both the maternal and paternal genomes in mouse zygotes. *Cell Stem Cell* **15,** 459–470 (2014).
131. Kim, S.-H. *et al.* Differential DNA methylation reprogramming of various repetitive sequences in mouse preimplantation embryos. *Biochem. Biophys. Res. Commun.* **324,** 58–63 (2004).
132. Inoue, A., Matoba, S. & Zhang, Y. Transcriptional activation of transposable elements in mouse zygotes is independent of Tet3-mediated 5-methylcytosine oxidation. *Cell Res* **22,** 1640–1649 (2012).
133. Vella, P. *et al.* Tet proteins connect the O-linked N-acetylglucosamine transferase Ogt to chromatin in embryonic stem cells. *Mol. Cell* **49,** 645–656 (2013).
134. Chen, Q., Chen, Y., Bian, C., Fujiki, R. & Yu, X. TET2 promotes histone O-GlcNAcylation during gene transcription. *Nature* **493,** 561–564 (2013).
135. Deplus, R. *et al.* TET2 and TET3 regulate GlcNAcylation and H3K4 methylation through OGT and SET1/COMPASS. *EMBO J* **32,** 645–655 (2013).
136. Williams, K. *et al.* TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* **473,** 343–348 (2011).
137. Neri, F. *et al.* Genome-wide analysis identifies a functional association of Tet1 and Polycomb repressive complex 2 in mouse embryonic stem cells. *Genome Biol.* **14,** R91 (2013).
138. Guallar, D. *et al.* RNA-dependent chromatin targeting of TET2 for endogenous retrovirus control in pluripotent stem cells. *Nat Genet* **30,** 733 (2018).
139. Leung, D. *et al.* Regulation of DNA methylation turnover at LTR retrotransposons and imprinted loci by the histone methyltransferase Setdb1. *Proc Natl Acad Sci USA* **111,** 6690–6695 (2014).
140. Bachman, M. *et al.* 5-Hydroxymethylcytosine is a predominantly stable DNA modification. *Nat Chem* **6,** 1049–1055 (2014).
141. Bachman, M. *et al.* 5-Formylcytosine can be a stable DNA modification in mammals. *Nat.*

Chem. Biol. **11,** 555–557 (2015).
142. Hashimoto, H. *et al.* Wilms tumor protein recognizes 5-carboxylcytosine within a specific DNA sequence. *Genes & Development* **28,** 2304–2313 (2014).
143. Kellinger, M. W. *et al.* 5-formylcytosine and 5-carboxylcytosine reduce the rate and substrate specificity of RNA polymerase II transcription. *Nat Struct Mol Biol* **19,** 831–833 (2012).
144. Iurlaro, M. *et al.* A screen for hydroxymethylcytosine and formylcytosine binding proteins suggests functions in transcription and chromatin regulation. *Genome Biol.* **14,** R119 (2013).
145. Spruijt, C. G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152,** 1146–1159 (2013).
146. Xiong, J. *et al.* Cooperative Action between SALL4A and TET Proteins in Stepwise Oxidation of 5-Methylcytosine. *Mol. Cell* **64,** 913–925 (2016).
147. Fu, Y. *et al.* N6-methyldeoxyadenosine marks active transcription start sites in Chlamydomonas. *Cell* **161,** 879–892 (2015).
148. Greer, E. L. *et al.* DNA Methylation on N6-Adenine in C. elegans. *Cell* **161,** 868–878 (2015).
149. Koziol, M. J. *et al.* Identification of methylated deoxyadenosines in vertebrates reveals diversity in DNA modifications. *Nat Struct Mol Biol* **23,** 24–30 (2016).
150. Xiao, C.-L. *et al.* N6-Methyladenine DNA Modification in the Human Genome. *Mol. Cell* **71,** 306–318.e7 (2018).
151. Sánchez-Romero, M. A., Cota, I. & Casadesús, J. DNA methylation in bacteria: from the methyl group to the methylome. *Curr. Opin. Microbiol.* **25,** 9–16 (2015).
152. Roberts, D., Hoopes, B. C., McClure, W. R. & Kleckner, N. IS10 transposition is regulated by DNA adenine methylation. *Cell* **43,** 117–130 (1985).
***Dam- E.Coli* mutants are used to show that 6mA loss results in increased transcription of the IS10 transposon, and that this leads to transposition.**
153. Wang, Y., Chen, X., Sheng, Y., Liu, Y. & Gao, S. N6-adenine DNA methylation is associated with the linker DNA of H2A.Z-containing well-positioned nucleosomes in Pol II-transcribed genes in Tetrahymena. *Nucleic Acids Research* **45,** 11594–11606 (2017).
154. Chen, H. *et al.* Phytophthora methylomes are modulated by 6mA methyltransferases and associated with adaptive genome regions. *Genome Biol.* **19,** 181 (2018).
155. Liang, Z. *et al.* DNA N6-Adenine Methylation in Arabidopsis thaliana. *Developmental Cell* **45,** 406–416.e3 (2018).
156. Liu, J. *et al.* Abundant DNA 6mA methylation during early embryogenesis of zebrafish and pig. *Nature Communications* **7,** 13052 (2016).
157. Yao, B. *et al.* DNA N6-methyladenine is dynamically regulated in the mouse brain following environmental stress. *Nature Communications* **8,** 1122 (2017).
158. Zhu, S. *et al.* Mapping and characterizing N6-methyladenine in eukaryotic genomes using single-molecule real-time sequencing. *Genome Research* **28,** 1067–1078 (2018).
159. Xie, Q. *et al.* N6-methyladenine DNA Modification in Glioblastoma. *Cell* **175,** 1228–1243.e20 (2018).
160. Mondo, S. J. *et al.* Widespread adenine N6-methylation of active genes in fungi. *Nat Genet* **49,** 964–968 (2017).
161. Brocken, D. J. W., Tark-Dame, M. & Dame, R. T. dCas9: A Versatile Tool for Epigenome Editing. *Curr Issues Mol Biol* **26,** 15–32 (2018).
162. Machnicka, M. A. *et al.* MODOMICS: a database of RNA modification pathways--2013 update. *Nucleic Acids Research* **41,** D262–7 (2013).
163. Zhao, B. S., Roundtree, I. A. & He, C. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* **18,** 31–42 (2017).

164. Wan, Y. *et al.* Transcriptome-wide high-throughput deep m(6)A-seq reveals unique differential m(6)A methylation patterns between three organs in Arabidopsis thaliana. *Genome Biol.* **16,** 272 (2015).
165. Zhang, Z. & Xing, Y. CLIP-seq analysis of multi-mapped reads discovers novel functional RNA regulatory sites in the human transcriptome. *Nucleic Acids Research* **45,** 9260–9271 (2017).
166. Huang, L., Ashraf, S., Wang, J. & Lilley, D. M. Control of box C/D snoRNP assembly by N6-methylation of adenine. *EMBO Rep.* **18,** 1631–1645 (2017).
167. Zhou, C. *et al.* Genome-Wide Maps of m6A circRNAs Identify Widespread and Cell-Type-Specific Methylation Patterns that Are Distinct from mRNAs. *CellReports* **20,** 2262–2276 (2017).
168. Horwich, M. D. *et al.* The Drosophila RNA methyltransferase, DmHen1, modifies germline piRNAs and single-stranded siRNAs in RISC. *Curr Biol* **17,** 1265–1272 (2007).
169. Saito, K. *et al.* Pimet, the Drosophila homolog of HEN1, mediates 2'-O-methylation of Piwi-interacting RNAs at their 3' ends. *Genes & Development* **21,** 1603–1608 (2007).
170. Kamminga, L. M. *et al.* Hen1 is required for oocyte development and piRNA stability in zebrafish. *EMBO J* **29,** 3688–3700 (2010).
171. Jackman, J. E. & Alfonzo, J. D. Transfer RNA modifications: nature's combinatorial chemistry playground. *Wiley Interdiscip Rev RNA* **4,** 35–48 (2013).
172. Phalke, S. *et al.* Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nat Genet* **41,** 696–702 (2009).
173. Kuhlmann, M. *et al.* Silencing of retrotransposons in Dictyostelium by DNA methylation and RNAi. *Nucleic Acids Research* **33,** 6405–6417 (2005).
174. Raddatz, G. *et al.* Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc Natl Acad Sci USA* **110,** 8627–8631 (2013).
175. Genenncher, B. *et al.* Mutations in Cytosine-5 tRNA Methyltransferases Impact Mobile Element Expression and Genome Stability at Specific DNA Repeats. *CellReports* **22,** 1861–1874 (2018).
176. Ibarra, C. A. *et al.* Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337,** 1360–1364 (2012).
This paper provides genome-wide evidence that flowering plants use companion cells to protect their gametes from harmful transposition.
177. Rowe, H. M. *et al.* De novo DNA methylation of endogenous retroviruses is shaped by KRAB-ZFPs/KAP1 and ESET. *Development* **140,** 519–529 (2013).

Competing interests

The authors declare no competing interests.

Acknowledgements

We thank Matthew Lorincz, Richard Meehan and the reviewers for detailed comments on the manuscript, which led to extensive improvements to this Review. We also thank Cédric Feschotte for input on Figure 1. We apologize to colleagues whose work was not cited owing to space limitations. M.R.B. is a Sir Henry Dale Fellow (101225/Z/13/Z), jointly funded by the Wellcome Trust and the Royal Society. Ö.D. has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007-2013) under REA grant agreement number 608765.

Author contributions

All authors contributed to all aspects of the manuscript.

Reviewer information

Nature Reviews Genetics thanks J. Dejardin, M. Gauchier, J. Pontis, D. Trono and the other, anonymous reviewer(s) for their contribution to the peer review of this work.

Subject categories:

Biological sciences / Genetics / Epigenetics
[URI /631/208/176]

Biological sciences / Genetics / Genomics / Epigenomics
[URI /631/208/212/177]

Biological sciences / Genetics / Genome / Interspersed repetitive sequences / DNA transposable elements
[URI /631/208/726/2001/1428]

Biological sciences / Evolution / Evolutionary genetics
[URI /631/181/2474]

Biological sciences / Molecular biology / Epigenetics / DNA methylation
[URI /631/337/176/1988]

Biological sciences / Genetics / Genomics / Mobile elements
[URI /631/208/212/2305]

Biological sciences / Genetics / Gene regulation
[URI /631/208/200]

Table of contents blurb

Transposable elements (TEs) need to be tightly regulated in genomes to prevent the detrimental consequences of transposition. In this article, Deniz, Frost and Branco discuss how DNA methylation dynamics play a central role in the multilayered epigenetic mechanisms regulating TEs. Beyond roles for 5-methylcytosine (5mC), they discuss TET-mediated oxidation products of 5mC, as well as ongoing debates about the functional relevance of adenine methylation.

Figure 1

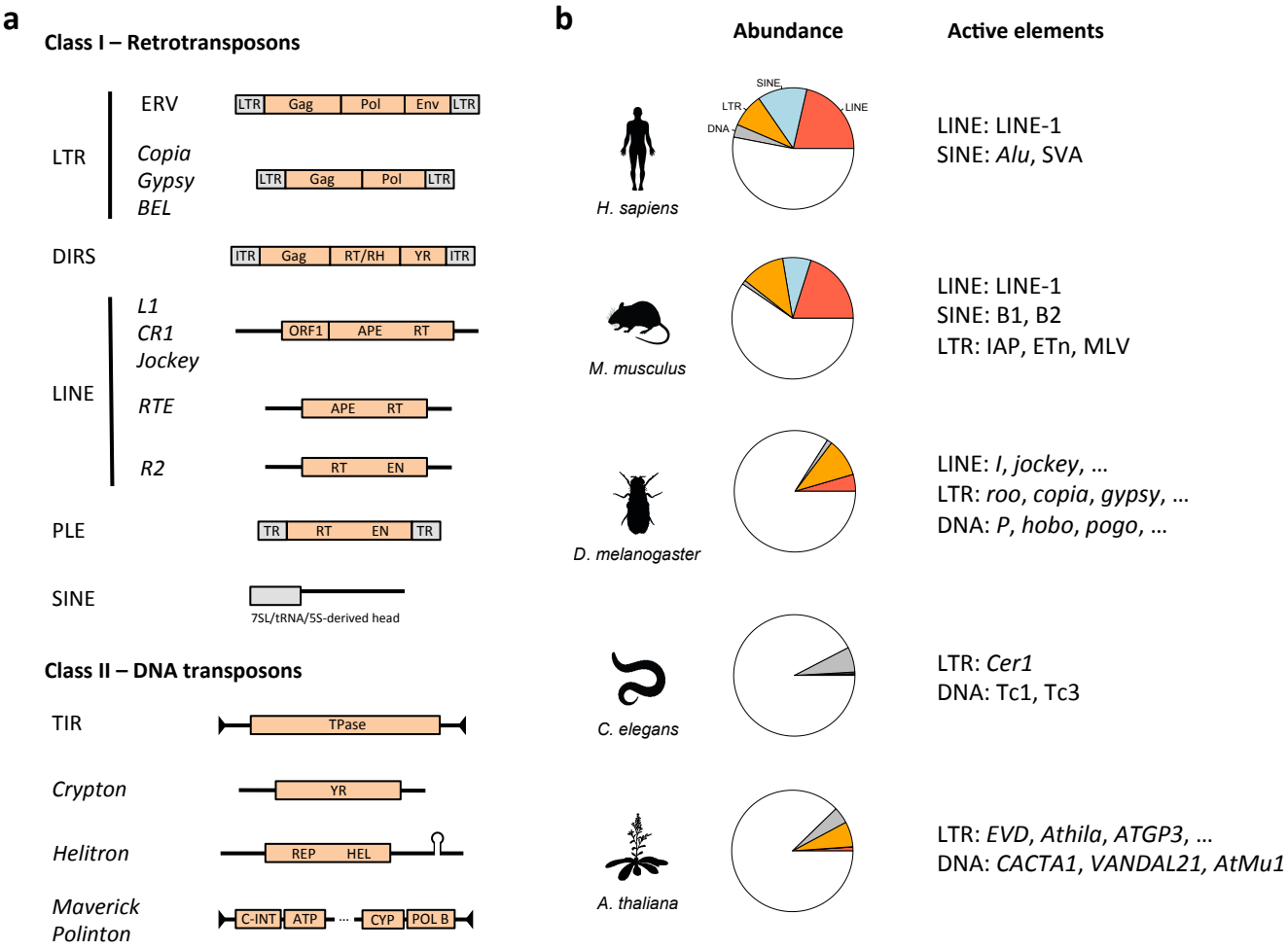

a
Class I – Retrotransposons
LTR
ERV
LTR Gag Pol Env LTR
Copia
Gypsy
BEL
LTR Gag Pol LTR
DIRS
ITR Gag RT/RH YR ITR
LINE
L1
CR1
Jockey
ORF1 APE RT
RTE
APE RT
R2
RT EN
PLE
TR RT EN TR
SINE
7SL/tRNA/5S-derived head
Class II – DNA transposons
TIR
TPase
Crypton
YR
Helitron
REP HEL
Maverick
Polinton
C-INT ATP ... CYP POL B
b
Abundance
Active elements
SINE
LTR
DNA
LINE
H. sapiens
LINE: LINE-1
SINE: Alu, SVA
M. musculus
LINE: LINE-1
SINE: B1, B2
LTR: IAP, ETn, MLV
D. melanogaster
LINE: I, jockey, …
LTR: roo, copia, gypsy, …
DNA: P, hobo, pogo, …
C. elegans
LTR: Cer1
DNA: Tc1, Tc3
A. thaliana
LTR: EVD, Athila, ATGP3, …
DNA: CACTA1, VANDAL21, AtMu1

Figure 2

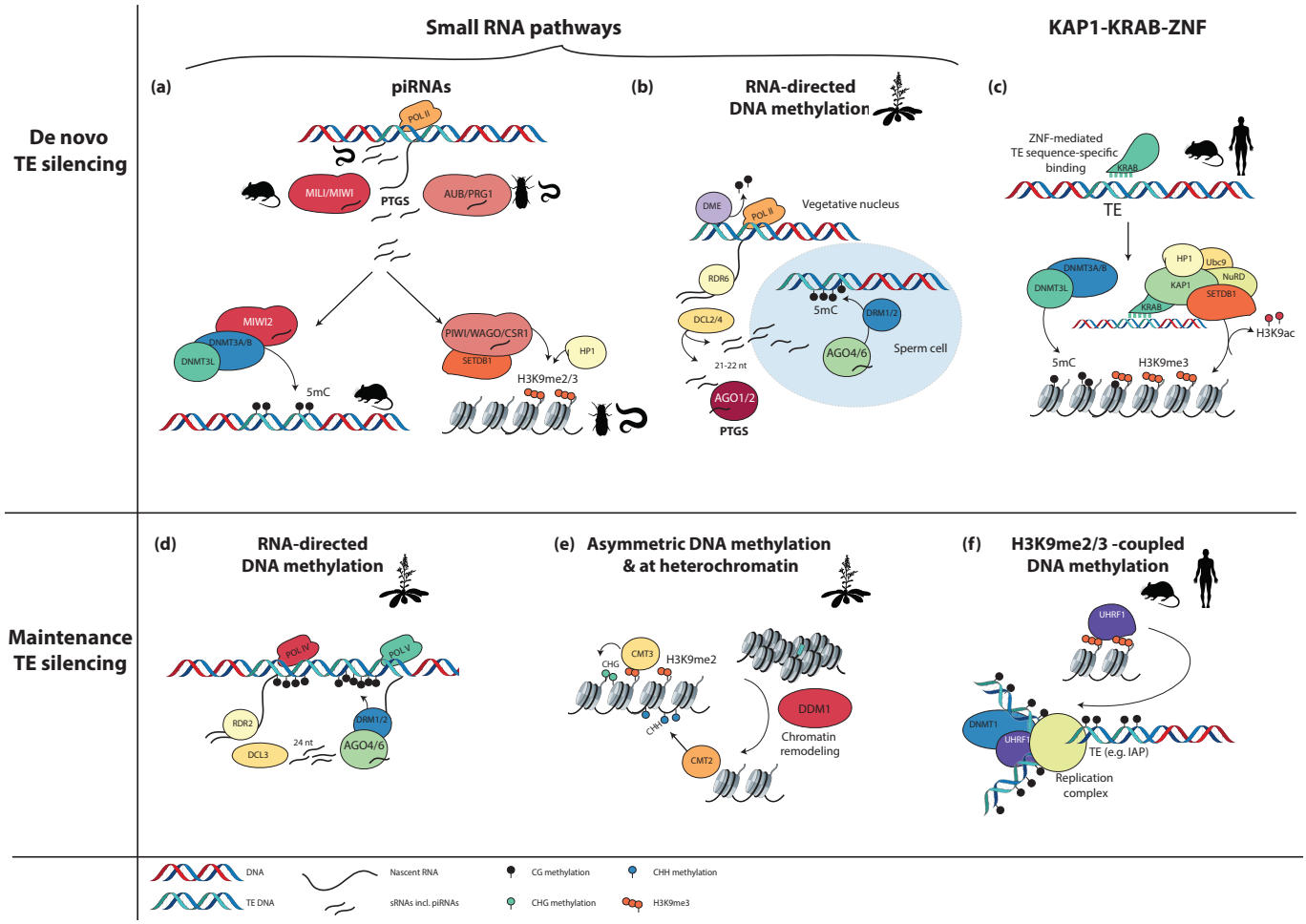

Small RNA pathways
KAP1-KRAB-ZNF
De novo TE silencing
Maintenance TE silencing
(a) piRNAs
(b) RNA-directed DNA methylation
(c)
(d) RNA-directed DNA methylation
(e) Asymmetric DNA methylation & at heterochromatin
(f) H3K9me2/3 -coupled DNA methylation
Pol II
MILI/MIWI
PTGS
AUB/PRG1
MIWI2
DNMT3A/B
DNMT3L
5mC
PIWI/WAGO/CSR1
SETDB1
HP1
H3K9me2/3
DME
Vegetative nucleus
RDR6
DCL2/4
21-22 nt
AGO1/2
DRM1/2
AGO4/6
Sperm cell
ZNF-mediated TE sequence-specific binding
KRAB
TE
Ubc9
NuRD
KAP1
SETDB1
H3K9ac
H3K9me3
Pol IV
Pol V
RDR2
DCL3
24 nt
CMT3
H3K9me2
CHG
CHH
DDM1
Chromatin remodeling
CMT2
UHRF1
DNMT1
TE (e.g. IAP)
Replication complex
DNA
TE DNA
Nascent RNA
sRNAs incl. piRNAs
CG methylation
CHG methylation
CHH methylation
H3K9me3

Figure 3

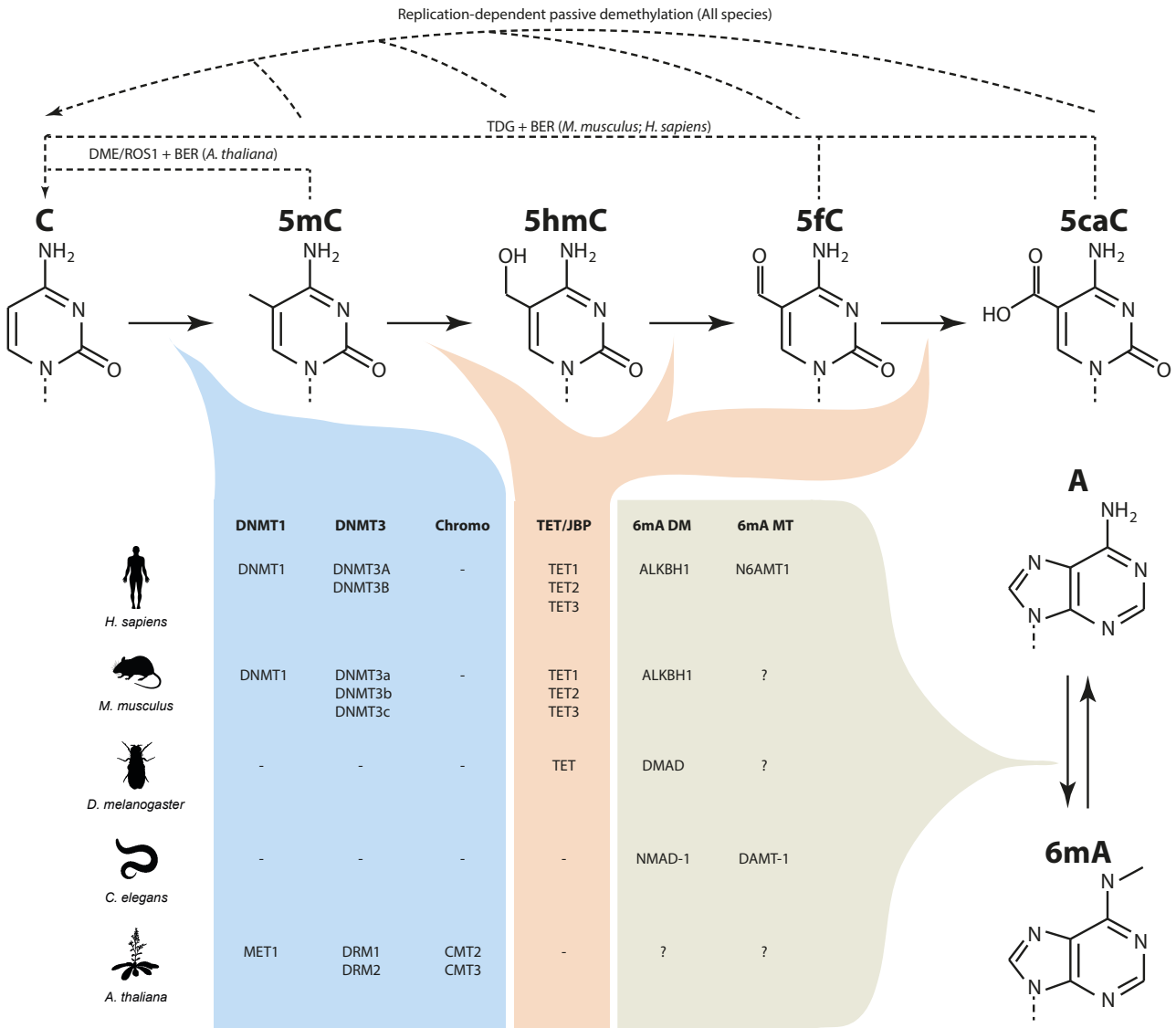

Replication-dependent passive demethylation (All species)
TDG + BER (M. musculus; H. sapiens)
DME/ROS1 + BER (A. thaliana)
C
5mC
5hmC
5fC
5caC
A
6mA
DNMT1
DNMT3
Chromo
TET/JBP
6mA DM
6mA MT
H. sapiens
DNMT1
DNMT3A
DNMT3B
-
TET1
TET2
TET3
ALKBH1
N6AMT1
M. musculus
DNMT1
DNMT3a
DNMT3b
DNMT3c
-
TET1
TET2
TET3
ALKBH1
?
D. melanogaster
-
-
-
TET
DMAD
?
C. elegans
-
-
-
-
NMAD-1
DAMT-1
A. thaliana
MET1
DRM1
DRM2
CMT2
CMT3
-
?
?

Figure 4

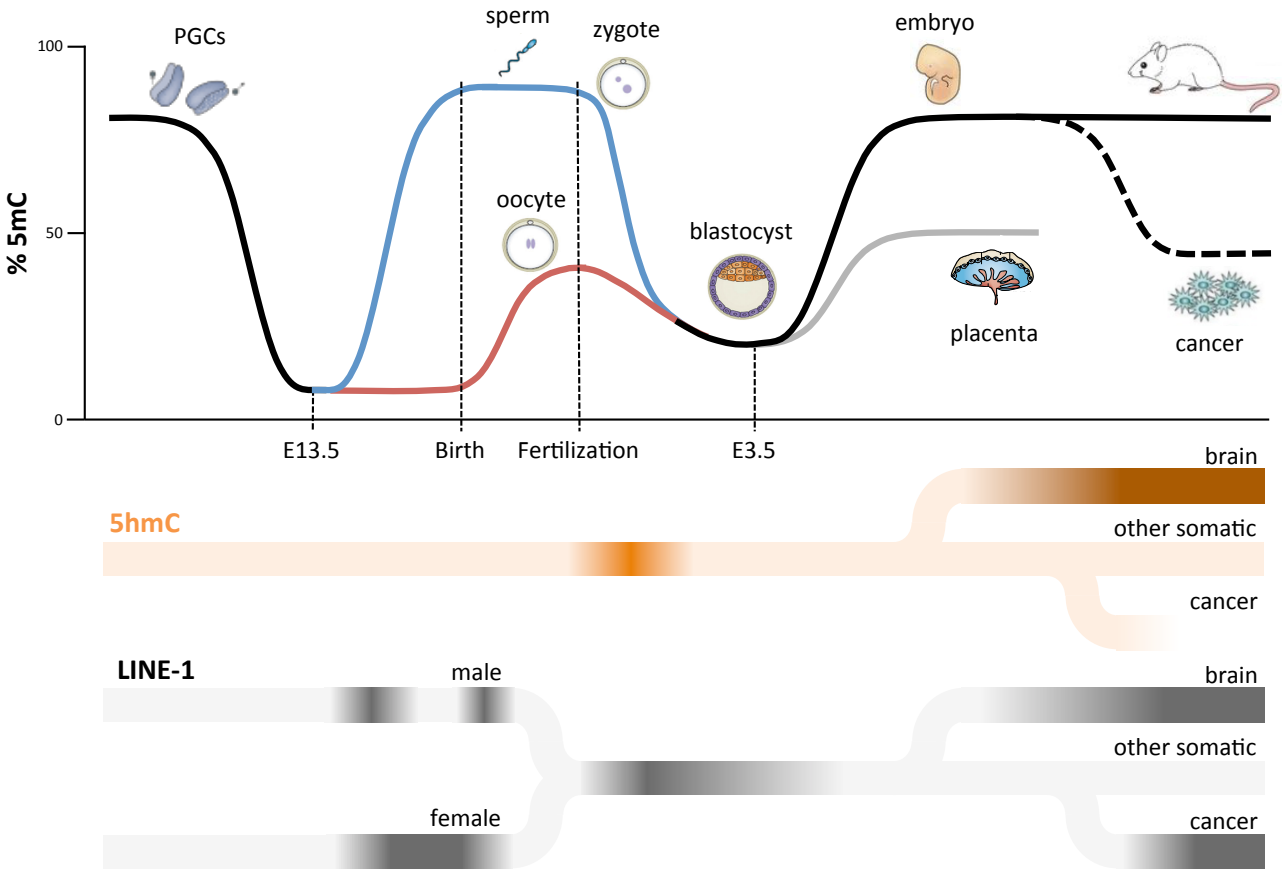
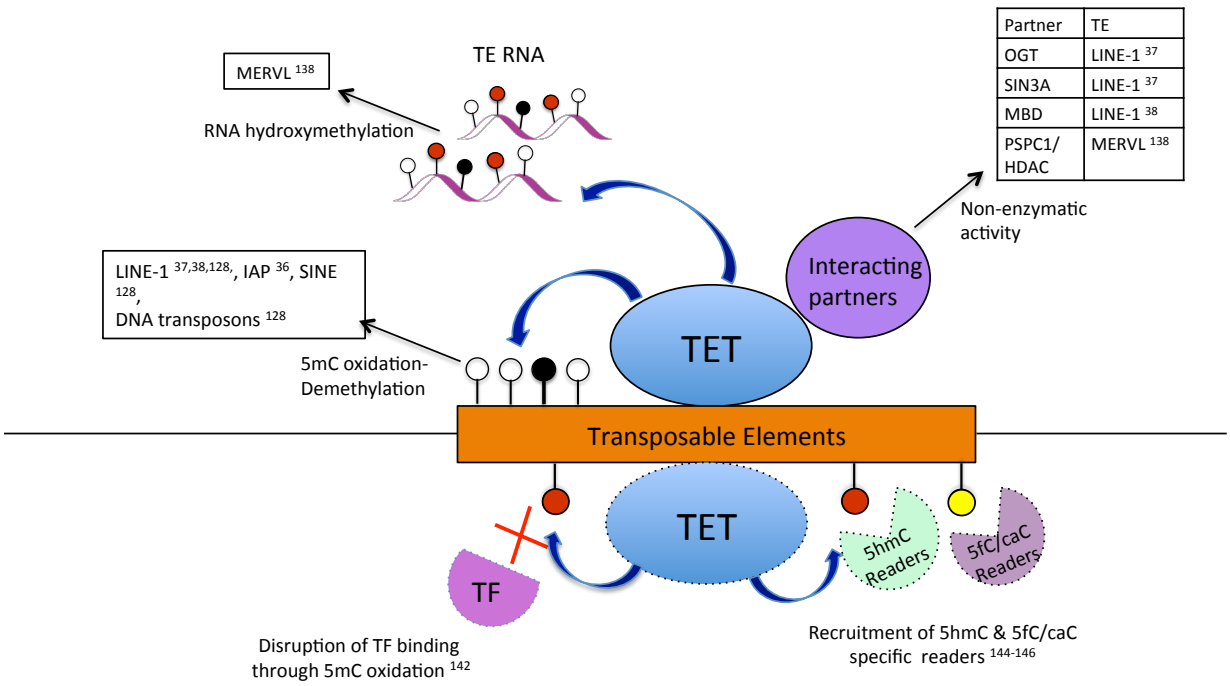

% 5mC
100
50
0
PGCs
sperm
zygote
oocyte
blastocyst
embryo
placenta
cancer
E13.5
Birth
Fertilization
E3.5
5hmC
brain
other somatic
cancer
LINE-1
male
female
brain
other somatic
cancer

Figure 5

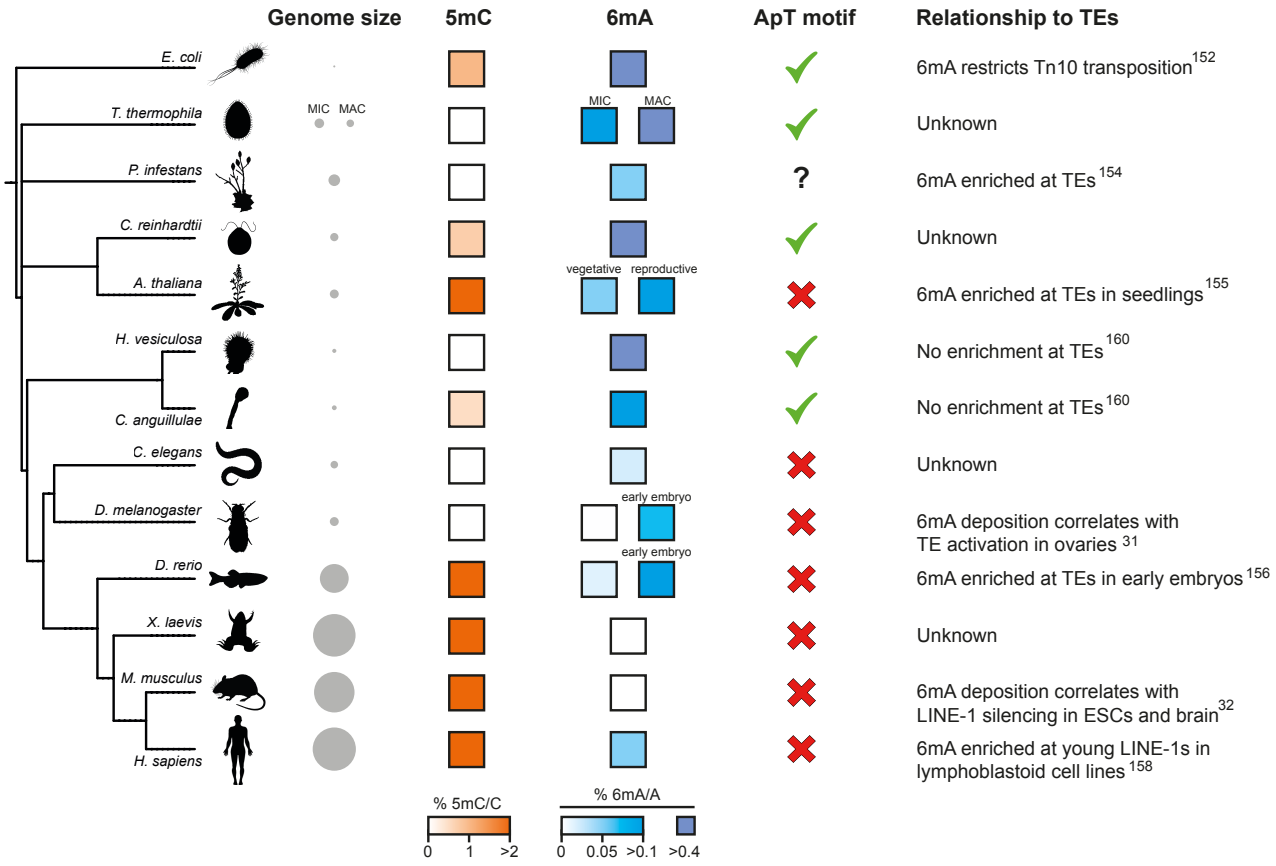
Mechanisms reported to act at TEs



Partner	TE
OGT	LINE-1 [37]
SIN3A	LINE-1 [37]
MBD	LINE-1 [38]
PSPC1/ HDAC	MERVL [138]

Mechanisms potentially acting at TEs

Figure 6


Genome size
5mC
6mA
ApT motif
Relationship to TEs
E. coli
T. thermophila
MIC MAC
P. infestans
C. reinhardtii
A. thaliana
H. vesiculosa
C. anguillulae
C. elegans
D. melanogaster
D. rerio
X. laevis
M. musculus
H. sapiens
MIC MAC
vegetative reproductive
early embryo
early embryo
?
6mA restricts Tn10 transposition[152]
Unknown
6mA enriched at TEs[154]
Unknown
6mA enriched at TEs in seedlings[155]
No enrichment at TEs[160]
No enrichment at TEs[160]
Unknown
6mA deposition correlates with TE activation in ovaries[31]
6mA enriched at TEs in early embryos[156]
Unknown
6mA deposition correlates with LINE-1 silencing in ESCs and brain[32]
6mA enriched at young LINE-1s in lymphoblastoid cell lines[158]
% 5mC/C
0 1 >2
% 6mA/A
0 0.05 >0.1 >0.4