# PROGRESS

# DNA $N^6$-methyladenine: a new epigenetic mark in eukaryotes?

*Guan-Zheng Luo, Mario Andres Blanco, Eric Lieberman Greer, Chuan He and Yang Shi*

Abstract | DNA $N^6$-adenine methylation ($N^6$-methyladenine; 6mA) in prokaryotes functions primarily in the host defence system. The prevalence and significance of this modification in eukaryotes had been unclear until recently. Here, we discuss recent publications documenting the presence of 6mA in *Chlamydomonas reinhardtii, Drosophila melanogaster* and *Caenorhabditis elegans*; consider possible roles for this DNA modification in regulating transcription, the activity of transposable elements and transgenerational epigenetic inheritance; and propose 6mA as a new epigenetic mark in eukaryotes.

DNA methylation is a fundamental epigenetic process. The most common DNA modification in eukaryotes is 5-methylcytosine (5mC)[1]. By contrast, $N^6$-methyladenine (6mA)[2] is the most prevalent DNA modification in prokaryotes[3]. Early reports debated the existence and abundance of 6mA in eukaryotes[1,4]. Some unicellular eukaryotes, such as ciliates and green algae, contain both 6mA and 5mC in their genomes, but the biological significance of these modifications in these organisms had remained largely uncharacterized until recently[5].

DNA methylation studies in mammals and plants have focused on 5mC because of its widespread distribution[6]. 5mC has been shown to participate in genomic imprinting, X-chromosome inactivation, transposon suppression, gene regulation and epigenetic memory maintenance[7–9]. By contrast, prokaryotes use 6mA as the main DNA modification to discriminate the host DNA from foreign pathogenic DNA and to protect the host genome via the restriction–modification system. In this system, the host DNA is modified by methyltransferases at restriction sites, conferring the host genome with resistance to digestion. Unmethylated foreign DNA is recognized and cleaved by restriction enzymes[10]. In addition to its crucial role in the restriction–modification system, 6mA also participates in bacterial DNA replication and repair, transposition,

nucleoid segregation and gene regulation[11,12]. Certain bacterial 6mA methyltransferases, such as DNA adenine methylase (Dam)[13] and cell cycle-regulated methylase (CcrM), do not have restriction enzyme counterparts and have been shown to have important roles in multiple cellular processes[14]. So far, the restriction–modification system has not been found in eukaryotes.

In this Progress article, we discuss the recent discoveries of 6mA in the multicellular eukaryotes *Caenorhabditis elegans* and *Drosophila melanogaster*[15,16], and in the unicellular algae *Chlamydomonas reinhardtii*, in which a detailed analysis of 6mA distribution was also reported[17]. We describe the presence of 6mA and methods for detecting it, and discuss the enzymes that are responsible for regulating 6mA and the potential functional roles for 6mA as an epigenetic modification in eukaryotes.

## The presence of 6mA
Most eukaryotic 6mA research has focused on unicellular protists, including the ciliates *Tetrahymena pyriformis*[18] and *Paramecium aurelia*[19], and the green algae *C. reinhardtii*[5]. 6mA accounts for ~0.4–0.8% of the total adenines in these genomes[17,20,21], which is several times lower than the prevalence of 5mC in mammals and plants[22]. Interestingly, there have also been sporadic reports suggesting the presence of 6mA in more

recently evolved organisms, including mosquitoes[23], plants[24,25] and even mammals[26,27]. Recently, 6mA was detected by multiple approaches in the genomic DNA of two metazoans, *C. elegans* (0.01–0.4%)[15] and *D. melanogaster* (0.001–0.07%)[16].

*Detection of 6mA.* In order to monitor the presence and dynamics of 6mA in eukaryotic genomes, sensitive techniques are required (FIG. 1). An antibody against 6mA can detect 6mA in eukaryotic mRNAs[28–30], and this was used to detect and enrich for 6mA in DNA[15–17]. The 6mA antibody can be used to detect methylated DNA by dot blotting and to immunoprecipitate methylated DNA for sequencing. Antibody detection, however, is not quantitative and may be confounded by recognition of other adenine base modifications (such as 1mA or 6mA in RNA). Liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) provides the most unambiguous detection and quantification of modified nucleotides[31]. However, although the high sensitivity of this approach allows for detection of many low-abundance nucleotide modifications, both antibody detection and LC–MS/MS results could be affected by bacterial contamination and therefore need to be cautiously evaluated.

Certain restriction enzymes are sensitive to nucleotide methylation and can be used to distinguish between methylated and unmethylated nucleotides in the context of their recognition sequences[32] (FIG. 1). For instance, DpnI digests 5′-G6mATC-3′ sites, whereas DpnII digests unmethylated 5′-GATC-3′ sites[33]. Restriction enzyme digestion-based methods can thus sensitively and accurately determine the presence of 6mA at single-base resolution; however, these enzymes can detect 6mA only in specific sequence contexts[34], which may not be recapitulated in all organisms. Single-molecule real-time (SMRT) sequencing offers another option for detecting DNA modifications[35]. SMRT sequencing both provides accurate sequence reads and determines the kinetics of nucleotide incorporations during synthesis. Different DNA modifications induce unique kinetic signatures that can be used for accurate and sequence-specific detection
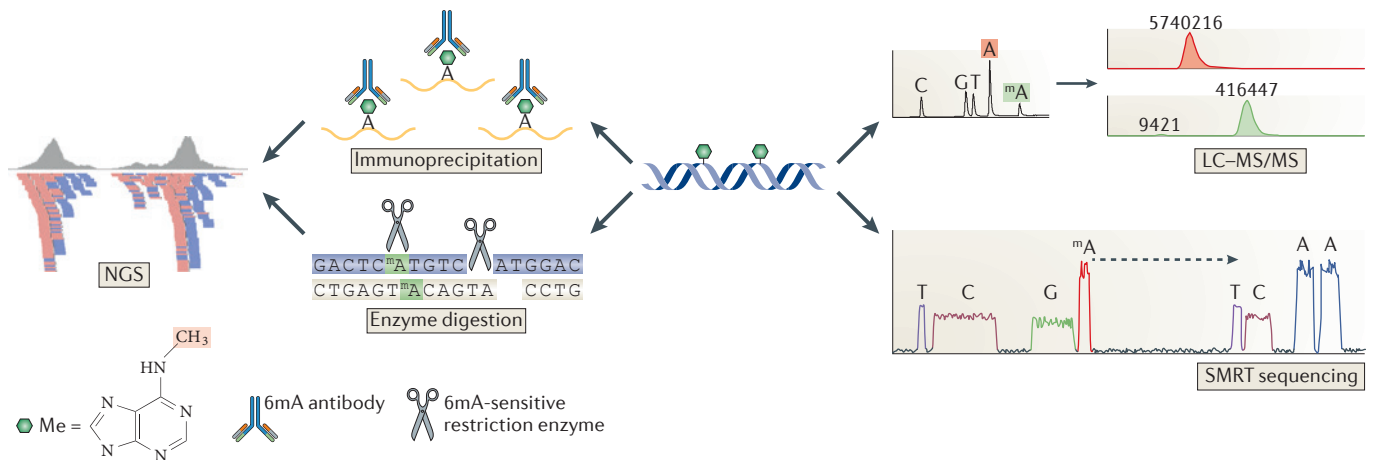
Figure 1 | **Methods to detect *N*[6]-methyladenine (6mA) in genomic DNA.** Left: an antibody against 6mA can sensitively recognize 6mA and enrich 6mA-containing DNA for subsequent next-generation sequencing (NGS); 6mA-sensitive restriction enzymes can specifically recognize either methylated or unmethylated adenines in their recognition-sequence motifs, and this can be captured by sequencing to determine the exact locations of 6mA. Right: liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) can differentiate methylated adenine from unmodified adenine and quantify the 6mA/A ratio by normalization to a standard curve; single-molecule real-time (SMRT) sequencing can detect modified nucleotides by measuring the rate of DNA base incorporation (dashed arrow) during sequencing.

of modifications[35,36]. SMRT sequencing has been used to map 6mA and 5mC simultaneously in *Escherichia coli*[37] and 6mA in *C. elegans*[15]. However, SMRT sequencing cannot distinguish between 6mA and 1mA and therefore must be coupled with LC–MS/MS in order to assess the relative contributions of the two modifications to the observed methylated adenine signals obtained from SMRT sequencing. SMRT sequencing is also still prohibitively expensive, impeding its extensive application to larger eukaryotic genomes.

For genomes with a low abundance of DNA modifications, the sensitivity and specificity of detection methods need to be balanced. High sensitivity with low specificity may generate false positive results, whereas high specificity with low sensitivity may miss weak signals. For a rare DNA modification, cross-validation by multiple strategies is indispensable to achieve relative high sensitivity and high specificity.

*Genomic distributions and sequence motifs.* Mapping genomic distributions can be an effective first step for uncovering potential functions of DNA modifications[7,9,38]. In the latest studies[15–17], immunoprecipitation of 6mA-containing DNA for subsequent high-throughput sequencing (6mA-IP-seq) or SMRT sequencing was used to identify the genomic distribution patterns of 6mA in *C. elegans*, *D. melanogaster* and *C. reinhardtii*, which were found to differ among these organisms: 6mA is broadly and evenly distributed across the genome of *C. elegans*, whereas in *D. melanogaster*, 6mA is enriched

at transposable elements. By contrast, 6mA is highly enriched around transcription start sites (TSSs) in the *C. reinhardtii* genome. Reminiscent of the genomic distribution of 5mC, 6mA localization patterns are thus species-specific, probably reflecting different functions and mechanisms of biogenesis[6]. By extension, the distinctive distribution patterns of 6mA in these three organisms may suggest diverse functional roles.

In prokaryotes, most of the 6mA sites are located within palindromic sequences[39]. Similar 6mA sites have also been identified in the genome of *Tetrahymena thermophila*[40]. Nearest-neighbour analyses, assisted by ^{3}H labelling, revealed that the methylation in *T. thermophila* occurs at the sequence 5′-NAT-3′ (REF. 41), which is similar to the 6mA motif found in *C. reinhardtii*[17]. Restriction enzyme analysis followed by next-generation sequencing (FIG. 1) was used to determine the methylation status of each CATG and GATC site in *C. reinhardtii*. Approximately 1/5–1/3 of the 6mAs were estimated to be located within these two motifs, and a majority of them (over 90%) were methylated in almost all DNA samples that were analysed. Others most probably occur at 5′-NATN-3′ sequences, which were also predicted to be 6mA motifs by a photo-crosslinking-based profiling method (6mA-CLIP-exo) that provides higher resolution information than does standard immunoprecipitation-based profiling[17].

Surprisingly, the sequence motifs identified in *C. elegans* are completely different from those in unicellular eukaryotes and bacteria. Using SMRT sequencing, two

motifs, AGAA and GAGG, were identified[15]. Sites with high abundance of 6mA are most strongly associated with GAGG, whereas lower-abundance 6mA sites are most strongly associated with AGAA. However, these motifs represent only a fraction (~10%) of the total methylated adenines, suggesting that additional factors beyond DNA sequence determine whether adenines are methylated in *C. elegans*. In *D. melanogaster*, conserved sequence motifs containing 6mA are yet to be reported[16]. The different 6mA motifs found in *C. elegans* and *C. reinhardtii* suggest potential diverse biological functions in distant organisms.
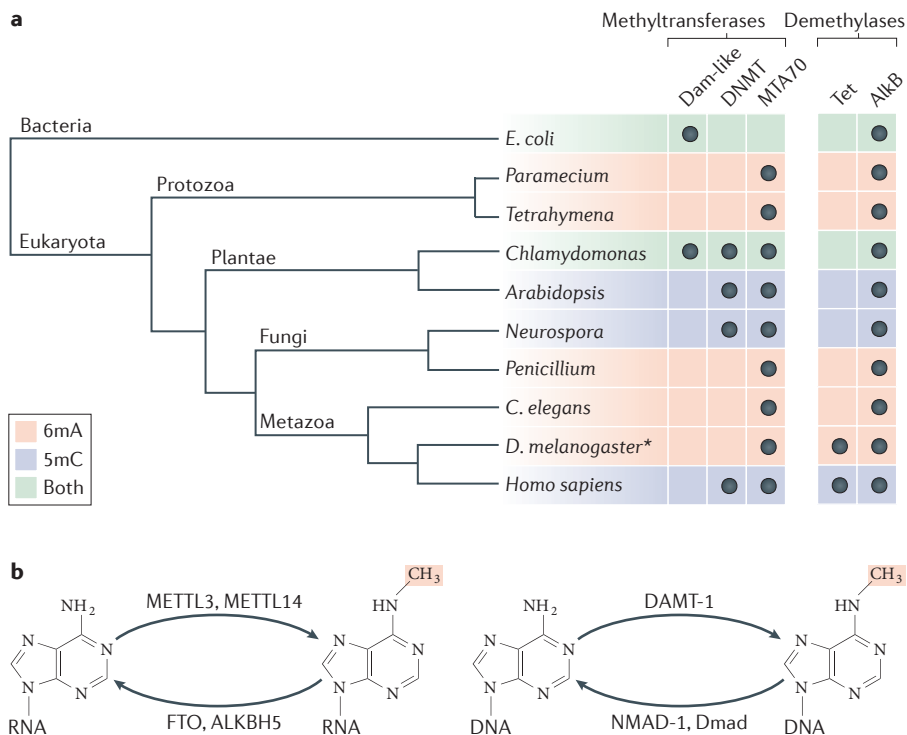
**Methyltransferases and demethylases**
6mA is the most abundant internal modification in mRNAs and is widely conserved in eukaryotes. Methylation of mRNA is catalysed by a methyltransferase complex and reversed by demethylases[42–45]. The enzymes that are responsible for DNA 6mA in eukaryotes were previously unknown. Computation-based sequence analysis of several eukaryotic organisms (FIG. 2a) predicted the existence of potential 6mA methyltransferases similar to the bacterial DNA adenine methyltransferases Dam and m.MunI[46]. However, other DNA and RNA methyltransferase and demethylase proteins could also have evolved to catalyse DNA 6mA in eukaryotic species[37].

*6mA methyltransferases.* Members of the DNA methyltransferase (DNMT) family function as 5mC methyltransferases in animals[47]. Although widely conserved, these proteins have been lost in certain

species, including in *C. elegans* and *D. melanogaster* (FIG. 2a). A family of enzymes containing an MT-A70 domain has evolved from the m.MunI-like 6mA methyltransferase of bacteria[46]. This family includes yeast and mammalian mRNA methyltransferases (inducer of meiosis 4 (Ime4) and karyogamy 4 (Kar4) in yeast; methyltransferase-like protein 3 (METTL3) and METTL14 in humans)[44,46,48]. In humans, the MT-A70 domain is the *S*-adenosylmethionine (SAM) binding subunit that catalyses mRNA 6mA[48]. *C. elegans* encodes a member of this family, DAMT-1. Overexpression of DAMT-1 in insect cells led to elevated 6mA in genomic DNA, whereas the expression of DAMT-1 with a mutated catalytic domain did not affect 6mA levels. In *C. elegans,* knockdown of *damt-1* decreased 6mA levels in genomic DNA[15]. These data suggest that DAMT-1 is probably a 6mA methyltransferase in *C. elegans* (FIG. 2b; TABLE 1), although direct biochemical evidence is still needed to confirm this possibility. In humans, METTL3 and METTL14 have weak methylation activity on DNA[44]. Another mammalian MTA70-type protein, METTL4, is the closest predicted orthologue of DAMT-1 in humans; however, its function has not been tested. The high level of evolutionary conservation and the broad distribution of the MTA70 family proteins raises the possibility that 6mA might be present in other eukaryotes, including mammals.

*6mA demethylases.* Proteins of the AlkB family catalyse the demethylation of various methylated DNA and RNA nucleotides[49]. For instance, homologues of the AlkB family members α-ketoglutarate-dependent dioxygenase FTO and RNA demethylase ALKBH5 have been shown to demethylate mRNA 6mA in eukaryotes[43,45]. *C. elegans* encodes five AlkB family members. Deletion of one member, *nmad-1,* caused elevated DNA 6mA levels *in vivo*, and purified NMAD-1 catalyses $N^6$-deoxyadenine demethylation *in vitro*[15] (FIG. 2b; TABLE 1). Together, these results demonstrated that NMAD-1 is a 6mA demethylase in *C. elegans*.

*D. melanogaster* nuclear extracts have DNA 6mA-demethylation activity. Interestingly, these nuclear extracts have the highest 6mA-demethylation activity when extracted at the time point at which 6mA levels are the lowest[16]. The homologue of the 5mC demethylase Tet[50] (renamed Dmad), when added to nuclear extracts, increased the 6mA-demethylating activity[16]. *In vitro* assays showed that the nuclear extract from *dmad* mutants lost demethylation activity, and the



Figure 2 | **DNA methyltransferases and demethylases. a** | Three families of methyltransferases and two families of DNA demethylases in ten different model organisms are shown in a simplified phylogenetic tree. Colour codes represent the reported presence of $N^6$-methyladenine (6mA) and/or 5-methylcytosine (5mC) in the corresponding organism (not the activity of the enzymes mentioned). The MTA70 family proteins are widely conserved in eukaryotes but are missing in *Escherichia coli*. They function as the RNA methyltransferases in plants and animals but appear to serve as the DNA 6mA methyltransferase in *Caenorhabditis elegans*. The Tet family proteins are DNA cytosine demethylases but have been proposed to demethylate adenines in *Drosophila melanogaster*. The AlkB family proteins are conserved in all of the organisms shown and are potential DNA 6mA demethylases in *C. elegans*. The black dot indicates that at least one member of each protein family exists in the corresponding organism. *The *D. melanogaster* genome contains a low level of 5mC (~0.03% of total cytosines)[71]. **b** | Two members — methyltransferase-like protein 3 (METTL3) and METTL14 — of the MTA70 family are known to catalyse 6mA in mammalian mRNA, whereas FTO and ALKBH5, which are members of the AlkB family, have been found to demethylate 6mA in mRNA. DAMT-1 is an MTA70 family member that potentially mediates DNA 6mA methylation in *C. elegans*. An AlkB family member, NMAD-1, was identified as a demethylase in *C. elegans*. In *D. melanogaster*, the potential 6mA demethylase Dmad is a member of the Tet family of proteins.

addition of purified Dmad recovered the demethylation activity. Dmad is expressed at low levels in early embryonic stages (45 minutes after fertilization) but is induced at later stages, indicating that it has a role in removing 6mA during embryogenesis[16]. Genomic DNA isolated from brains of *dmad* mutants has ~100-fold higher levels of 6mA than that from wild-type flies[16]. Taken together, these data suggest that Dmad is a 6mA demethylase in *D. melanogaster* (FIG. 2b). This is somewhat surprising, as the Tet proteins are evolutionarily conserved DNA cytosine demethylases rather than adenine demethylases. The available crystal structures of Tet catalytic domains, unlike those of the AlkB family, revealed an active site that might not accommodate a purine base[51,52]. However, as 5mC levels in

*D. melanogaster* are quite low, Dmad could have evolved as a 6mA demethylase instead of oxidizing 5mC; further biochemical and structural investigations will provide additional insights. Interestingly, *D. melanogaster* has an orthologue of NMAD-1, CG4036, and it will be interesting to determine whether CG4036 mediates demethylation of 6mA, in addition to Dmad.

## Functions of 6mA
Although DNA 6mA has been well studied in prokaryotes, its eukaryotic biological functions remain elusive[21]. There is no known eukaryotic equivalent of the bacterial restriction–modification system, ruling out a possible role for 6mA in this context. The addition of a methyl group at

Table 1 | **Potential *N*[6]-adenine methyltransferases and demethylases**

| Protein | Family | Catalytic domain | Substrate | Organism | Evidences | Refs |
|---|---|---|---|---|---|---|
| Dam* | Dam-like | SAM-dependent MTases | DNA | *E. coli* | *In vitro* and *in vivo* | 13 |
| DNMT1* | DNMT | SAM-dependent MTases C5-type | DNA | *H. sapiens* | *In vitro* and *in vivo* | 47 |
| METTL3, METTL14* | MT-A70 | SAM-dependent MTases | RNA | *H. sapiens* | *In vitro* and *in vivo* | 44 |
| DAMT-1* | MT-A70 | SAM-dependent MTases | DNA | *C. elegans* | *In vivo* | 15 |
| FTO, ALKBH5[‡] | AlkB | 2OG-FeII oxygenase superfamily | RNA | *H. sapiens* | *In vitro* and *in vivo* | 43,45 |
| NMAD-1[‡] | AlkB | 2OG-FeII oxygenase superfamily | DNA | *C. elegans* | *In vitro* and *in vivo* | 15 |
| Tet1, Tet2, Tet3[‡] | Tet | CD (Cys-rich insert and DSBH domain) | DNA | *H. sapiens* | *In vitro* and *in vivo* | 50,52 |
| DMAD[‡] | Tet | CD (Cys-rich insert and DSBH domain) | DNA | *D. melanogaster* | *In vivo* | 16 |

Protein family and catalytic domain information was derived from the Pfam and Uniprot databases. *C. elegans*, *Caenorhabditis elegans*; Dam, DNA adenine methylase; *D. melanogaster*, *Drosophila melanogaster*; DNMT, DNA methyltransferase; DSBH, double-stranded β-helix; *E. coli*, *Escherichia coli*; *H. sapiens*, *Homo sapiens*; METTL; methyltransferase-like protein; MTases, methyltransferases; SAM, *S*-adenosylmethionine. *indicates methyltransferases, ‡indicates demethylases.

the *N*[6]-position of adenine slightly reduces base-pairing energy[53] and could enhance or interfere with protein–DNA interactions[54–56]. Treatment of cancer cells with the nucleoside *N*[6]-methyldeoxyadenosine induces their differentiation[57]. Transgenic incorporation of bacterial Dam into tobacco plants induced 6mA at most GATC sites and changed the leaf and florescence morphology[58]. Collectively, these phenotypes suggest that DNA adenine methylation may have a role in regulating multiple biological processes.

*Transcription.* Methylation of adenines could affect transcription by modifying transcription factor binding or altering chromatin structure. In bacteria, 6mA has been shown to regulate transcription[59,60], raising the possibility that a similar function has been retained in eukaryotes. Interestingly, 6mA has been shown to have the opposite effects on transcription factor binding in mammals and in plants. Transfected DNA that was methylated at the *N*[6] position of adenines decreased the DNA-binding affinity of transcription factors in mammalian cells[54,56]. In plants, the binding affinity of the zinc-finger protein arabinogalactan protein 1 (AGP1) to DNA can be enhanced by 6mA of the target sequence[55]. Additionally, using a transient expression system in barley, transcription was found to increase in 6mA-modified reporter plasmids, whereas 5mC had little or no effect on transcription efficiency[61].

In *C. reinhardtii*, 6mA is enriched around the TSSs of more than 14,000 genes, most of which are actively transcribed, and most of these 6mA sites are methylated in almost every cell analysed[17]. By contrast, silent genes have lower 6mA levels around their TSS. Therefore, it appears that 6mA is a general mark of active genes, although whether it plays a part in the dynamic regulation of gene expression requires additional

investigation (FIG. 3a). 5mC also exists in the genome of *C. reinhardtii,* but is located at gene bodies and is correlated with transcriptional repression. It is unclear what role, if any, 6mA has during transcription in *C. elegans*. Interestingly, 6mA levels were elevated in mutants with increased levels of histone H3 Lys 4 dimethylation (H3K4me2), which is a mark associated with active transcription[15]. This suggests that 6mA may also mark active genes in *C. elegans* as it does in *C. reinhardtii*[17] (FIG. 3b).

In *D. melanogaster*, 6mA was proposed to promote transposon expression. 6mA-IP-seq assays revealed enrichment of 6mA on transposons, and loss of the putative demethylase Dmad led to increased transposon expression[16] (FIG. 3c). Thus, the latest discoveries in three evolutionarily distant organisms suggest a correlation between 6mA and elevated gene expression, although it remains to be seen how general this mechanism is and whether 6mA has evolved to have different functions in different species and/or contexts. Important future experiments will involve directing DNA methyltransferases or demethylases to specific loci and examining the effects on transcription of the addition and the removal of 6mA, respectively.

*Methylation and nucleosome positioning.* Previous studies in *T. thermophila* showed that 6mA levels were lower in nucleosomal DNA than in linker DNA[62]. Neither *C. elegans* nor *D. melanogaster* revealed a nucleosome-positioning bias for 6mA[15,16]. However, the *C. elegans* study was performed on mixed tissues, and specific genomic patterns might only emerge in a more detailed tissue-specific examination. In *C. reinhardtii*, conversely, 6mA was found to be preferentially located at TSS-proximal linker DNA (FIG. 3a). This could be explained by the methylation machinery favouring linker DNA, potentially owing to

physical ease of access[40], or by regulation of nucleosome positioning by 6mA. Interestingly, methylation in CG repeats in certain algae species occurs at extreme densities, which disfavours nucleosome assembly and thus dictates nucleosome positioning[63]. Therefore, DNA methylation could provide a general mechanism for controlling nucleosome positioning in unicellular eukaryotes.

*6mA as an epigenetic mark.* In prokaryotes, 6mA methyltransferases methylate adenine in the context of specific sequence motifs[11]; however, eukaryotic 6mA does not seem to be as strongly dependent on sequence motif recognition. In *C. reinhardtii*, 6mA occurs in multiple sequences, which are mainly located at ApT dinucleotides, resembling the 5mC methylation at CpG dinucleotides in mammals[17]. During multiple fission cell cycles, the 6mA levels are stably maintained. Single-base maps indicated that most of the individual 6mA sites are faithfully conserved under varying culture conditions, reinforcing the speculation regarding 6mA as a heritable epigenetic mark[17]. Although a 6mA methyltransferase has yet to be characterized in *C. reinhardtii*, there is very likely to be a mechanism by which 6mA signatures are inherited from parent cells to daughter cells. In *C. elegans*, deletion of the H3K4me2 demethylase suppressor of presenilin defect 5 (*spr-5*) causes a transgenerational progressive loss of fertility[64]. This fertility defect coincides with the progressive accumulation of H3K4me2 and a decline in the levels of the repressive mark H3K9me3 (REF. 65). In *spr-5*-mutant worms, 6mA also increases across generations (FIG. 3b); deletion of the 6mA demethylase *nmad-1* accelerates the progressive fertility defect of *spr-5*-mutant worms, whereas deletion of the potential 6mA methyltransferase *damt-1* suppresses it. Additionally, deleting *damt-1* suppressed the

accumulation of H3K4me2, suggesting that 6mA and H3K4me2 might be co-regulated and reinforce each other[15]. Although it remains to be determined whether 6mA or H3K4me2 can themselves transmit epigenetic information across generations, these modifications accumulate when epigenetic information is improperly inherited. These discoveries raise the possibility that 6mA may function as an epigenetic mark that carries heritable epigenetic information in eukaryotes.

## Summary and perspective

Both 5mC and 6mA were discovered in eukaryotic genomes decades ago[1,18]. Most research has focused on characterizing 5mC, owing to its abundance in mammals and plants. Multiple functions of 5mC have been revealed, whereas studies of 6mA have been very limited in eukaryotes. It is well known that 5mC can be spontaneously deaminated, which results in a C-to-T mutation[66]; thus, organisms possessing 5mC tend to lose CpG dinucleotides owing to deamination of the methylated cytosines[9]. Conversely, 6mA does not tend towards spontaneous mutations, which obviously could be advantageous for genomic stability.

Adenine methylation in eukaryotic mRNA was recently shown to have profound effects on gene expression[42]. The potential roles of 6mA DNA in shaping gene expression remain largely unknown. The methyl group of RNA 6mA can destabilize the Watson–Crick base-pairing by ~1.0 kcal/-mol[67,68], and this was shown to induce an 'm6A-switch' mechanism that alters RNA structure and thus binding by proteins[69]. The same property in DNA could affect transcription, replication and other processes that require strand separation or DNA bending. Furthermore, the methyl group may facilitate or inhibit recognition and binding by DNA-binding proteins to regulate gene expression. In mammalian cells, FTO-catalysed demethylation of RNA 6mA generates two newly discovered intermediates, $N^6$-hydroxymethyladenosine (6hmA or hm6A) and $N^6$-formyladenosine (6fA or f6A), in non-coding RNA and in mRNA[70]. As 6hmA and 6fA are stable modifications, it is possible that they are not merely intermediates but have regulatory functions. It remains to be seen whether similar modifications are produced during oxidative demethylation reactions of DNA 6mA and, if they exist, whether they have any biological role. Such modifications could further destabilize the DNA duplex or produce new protein-binding sites.
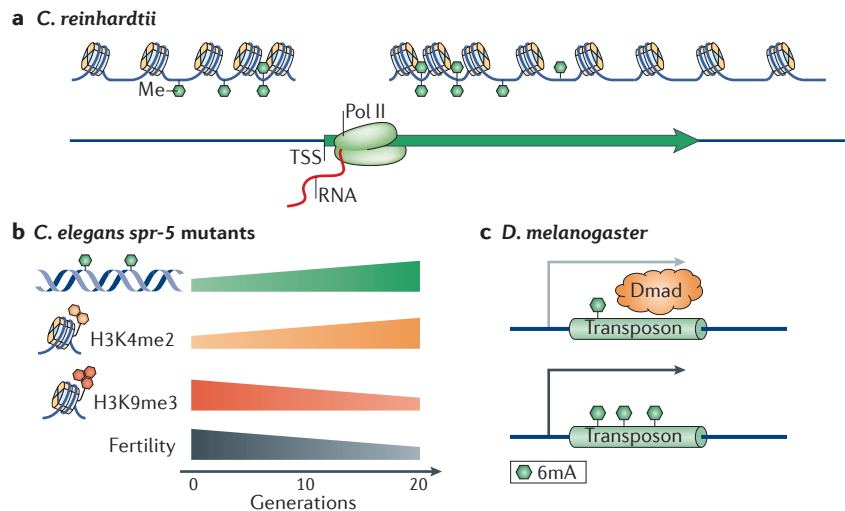


**Figure 3 | $N^6$-methyladenine (6mA) genomic distribution and function. a** | Single-base-resolution maps of 6mA in *Chlamydomonas reinhardtii* reveal a periodic distribution pattern, with depletion at active transcription start sites (TSSs) and enrichment around them. The nucleosomes are mutually exclusive with 6mA sites, which are found in the linker regions. The phasing level of nucleosomes varies from highly phased near the TSS to randomly positioned ~1 kb from the TSS, where the 6mA-methylation level also drops. **b** | In *Caenorhabditis elegans* suppressor of presenilin defect 5 (*spr-5*) mutants, 6mA and histone H3 Lys 4 dimethylation (H3K4me2) levels increase progressively from generation to generation, in conjunction with a decline in H3K9me3 levels and fertility. **c** | DNA 6mA is correlated with increased transposon expression in *Drosophila melanogaster*. Loss of the demethylase Dmad leads to elevated 6mA levels and increased expression of transposons. Pol II, RNA polymerase II.

It is interesting to note that *C. elegans* and *D. melanogaster* possess little or no 5mC in their genomic DNA, similar to some other 6mA-containing organisms such as *P. aurelia*[19] and *T. thermophila*[62]. *C. reinhardtii,* however, contains both 6mA and 5mC in its genome, but the relatively overall low level of 5mC combined with its unusual enrichment at exons suggests different functions for 5mC in this organism compared to more recently evolved plants and animals, in which it is mostly a mark of gene silencing[6]. It is possible that the relative change in 6mA and 5mC abundance and function could have had diverse evolutionary consequences. 6mA could be a major DNA-methylation mechanism that affects gene expression in certain eukaryotes, as revealed by recent studies[15–17]. In other systems, it may have a complementary role to that of the functionally more dominant 5mC DNA methylation. The presence of 6mA-methyltransferase homologues in distinct species raises the possibility that 6mA may be present in many more organisms, including mammals. Whether the relative biological importance of 6mA has declined in conjunction with its lower abundance in the more recently evolved organisms remains to be seen. We speculate that 6mA could have evolved a more specialized function in these species. With the advent of new, sensitive detection techniques, we are now poised to probe in detail the function and precise genomic distribution of this DNA modification throughout the tree of life.

*Guan-Zheng Luo is at the Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, and the Howard Hughes Medical Institute, 929 East 57th Street, Chicago, Illinois 60637, USA.*

*Mario Andres Blanco is at the Department of Cell Biology, Harvard Medical School, and the Division of Newborn Medicine, Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.*

*Eric Lieberman Greer is at the Division of Newborn Medicine, Children's Hospital Boston, 300 Longwood Avenue, Boston, and the Departments of Pediatrics and Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.*

*Chuan He is at the Department of Chemistry and Institute for Biophysical Dynamics, University of Chicago, and the Howard Hughes Medical Institute, 929 East 57th Street, Chicago, Illinois 60637, USA.*

*Yang Shi is at the Department of Cell Biology, Harvard Medical School, and the Division of Newborn Medicine, Children's Hospital Boston, 300 Longwood Avenue, Boston, Massachusetts 02115, USA.*

*Correspondence to C.H. and Y.S. e-mails: chuanhe@uchicago.edu; yang_shi@hms.harvard.edu*

1. Vanyushin, B. F., Tkacheva, S. G. & Belozersky, A. N. Rare bases in animal DNA. *Nature* **225**, 948–949 (1970).
2. Dunn, D. B. & Smith, J. D. Occurrence of a new base in the deoxyribonucleic acid of a strain of *Bacterium coli*. *Nature* **175**, 336–337 (1955).
3. Vanyushin, B. F., Belozersky, A. N., Kokurina, N. A. & Kadirova, D. X. 5-methylcytosine and 6-methylamino-purine in bacterial DNA. *Nature* **218**, 1066–1067 (1968).
4. Unger, G. & Venner, H. Remarks on minor bases in spermatic desoxyribonucleic acid. *Hoppe Seylers Z. Physiol. Chem.* **344**, 280–283 (1966).
5. Hattman, S., Kenny, C., Berger, L. & Pratt, K. Comparative study of DNA methylation in three unicellular eucaryotes. *J. Bacteriol.* **135**, 1156–1157 (1978).
6. Feng, S. H. *et al.* Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl Acad. Sci. USA* **107**, 8689–8694 (2010).
7. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
8. Jones, P. A. & Takai, D. The role of DNA methylation in mammalian epigenetics. *Science* **293**, 1068–1070 (2001).
9. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).
10. Naito, T., Kusano, K. & Kobayashi, I. Selfish behavior of restriction–modification systems. *Science* **267**, 897–899 (1995).
11. Wion, D. & Casadesus, J. $N^6$-methyl-adenine: an epigenetic signal for DNA–protein interactions. *Nat. Rev. Microbiol.* **4**, 183–192 (2006).
12. Vasu, K. & Nagaraja, V. Diverse functions of restriction–modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.* **77**, 53–72 (2013).
13. Urieli-Shoval, S., Gruenbaum, Y. & Razin, A. Sequence and substrate specificity of isolated DNA methylases from *Escherichia coli* C. *J. Bacteriol.* **153**, 274–280 (1983).
14. Wright, R., Stephens, C. & Shapiro, L. The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. *J. Bacteriol.* **179**, 5869–5877 (1997).
15. Greer, E. L. *et al.* DNA methylation on $N^6$-adenine in *C. elegans*. *Cell* **161**, 868–878 (2015).
16. Zhang, G. *et al.* $N^6$-methyladenine DNA modification in *Drosophila*. *Cell* **161**, 893–906 (2015).
17. Fu, Y. *et al.* $N^6$-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* **161**, 879–892 (2015).
18. Gorovsky, M. A., Hattman, S. & Pleger, G. L. [6N] methyl adenine in the nuclear DNA of a eucaryote, *Tetrahymena pyriformis*. *J. Cell Biol.* **56**, 697–701 (1973).
19. Cummings, D. J., Tait, A. & Goddard, J. M. Methylated bases in DNA from *Paramecium aurelia*. *Biochim. Biophys. Acta* **374**, 1–11 (1974).
20. Hattman, S. DNA-[adenine] methylation in lower eukaryotes. *Biochemistry* **70**, 550–558 (2005).
21. Ratel, D., Ravanat, J. L., Berger, F. & Wion, D. $N^6$-methyladenine: the other methylated base of DNA. *Bioessays* **28**, 309–315 (2006).
22. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* **9**, 465–476 (2008).
23. Adams, R. L., McKay, E. L., Craig, L. M. & Burdon, R. H. Methylation of mosquito DNA. *Biochim. Biophys. Acta* **563**, 72–81 (1979).
24. Ashapkin, V. V., Kutueva, L. I. & Vanyushin, B. F. The gene for domains rearranged methyltransferase (DRM2) in *Arabidopsis thaliana* plants is methylated at both cytosine and adenine residues. *FEBS Lett.* **532**, 367–372 (2002).
25. Vanyushin, B. F., Alexandrushkina, N. I. & Kirnos, M. D. $N^6$-methyladenine in mitochondrial-DNA of higher-plants. *Febs Lett.* **233**, 397–399 (1988).
26. Kay, P. H. *et al.* Evidence for adenine methylation within the mouse myogenic gene *Myo-D1*. *Gene* **151**, 89–95 (1994).
27. Huang, W. *et al.* Determination of DNA adenine methylation in genomes of mammals and plants by liquid chromatography/mass spectrometry. *RSC Adv.* **5**, 64046–64054 (2015).
28. Jia, G. *et al.* $N^6$-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
29. Dominissini, D. *et al.* Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* **485**, 201–206 (2012).
30. Luo, G. Z. *et al.* Unique features of the m6A methylome in *Arabidopsis thaliana*. *Nat. Commun.* **5**, 5630 (2014).
31. Frelon, S. *et al.* High-performance liquid chromatography—tandem mass spectrometry measurement of radiation-induced base damage to isolated and cellular DNA. *Chem. Res. Toxicol.* **13**, 1002–1010 (2000).
32. Roberts, R. J. & Macelis, D. REBASE — restriction enzymes and methylases. *Nucleic Acids Res.* **29**, 268–269 (2001).
33. Lacks, S. & Greenberg, B. Complementary specificity of restriction endonucleases of *Diplococcus pneumoniae* with respect to DNA methylation. *J. Mol. Biol.* **114**, 153–168 (1977).
34. Laird, P. W. Principles and challenges of genomewide DNA methylation analysis. *Nat. Rev. Genet.* **11**, 191–203 (2010).
35. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* **7**, 461–465 (2010).
36. Schadt, E. E. *et al.* Modeling kinetic rate variation in third generation DNA sequencing data to detect putative modifications to DNA bases. *Genome Res.* **23**, 129–141 (2013).
37. Fang, G. *et al.* Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* **30**, 1232–1239 (2012).
38. Jones, P. L. *et al.* Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19**, 187–191 (1998).
39. Touzain, F., Petit, M. A., Schbath, S. & El Karoui, M. DNA motifs that sculpt the bacterial chromosome. *Nat. Rev. Microbiol.* **9**, 15–26 (2011).
40. Karrer, K. M. & VanNuland, T. A. Methylation of adenine in the nuclear DNA of *Tetrahymena* is internucleosomal and independent of histone H1. *Nucleic Acids Res.* **30**, 1364–1370 (2002).
41. Bromberg, S., Pratt, K. & Hattman, S. Sequence specificity of DNA adenine methylase in the protozoan *Tetrahymena thermophila*. *J. Bacteriol.* **150**, 993–996 (1982).
42. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m6A RNA methylation. *Nat. Rev. Genet.* **15**, 293–306 (2014).
43. Jia, G. F. *et al.* N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* **7**, 885–887 (2011).
44. Liu, J. *et al.* A METTL3–METTL14 complex mediates mammalian nuclear RNA $N^6$-adenosine methylation. *Nat. Chem. Biol.* **10**, 93–95 (2014).
45. Zheng, G. Q. *et al.* ALKBH5 is a mammalian RNA demethylase that impacts RNA metabolism and mouse fertility. *Mol. Cell* **49**, 18–29 (2013).
46. Iyer, L. M., Abhiman, S. & Aravind, L. Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* **101**, 25–104 (2011).
47. Song, J., Rechkoblit, O., Bestor, T. H. & Patel, D. J. Structure of DNMT1–DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* **331**, 1036–1040 (2011).
48. Bokar, J. A., Shambaugh, M. E., Polayes, D., Matera, A. G. & Rottman, F. M. Purification and cDNA cloning of the AdoMet-binding subunit of the human mRNA ($N^6$-adenosine)-methyltransferase. *RNA* **3**, 1233–1247 (1997).
49. Trewick, S. C., Henshaw, T. F., Hausinger, R. P., Lindahl, T. & Sedgwick, B. Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* **419**, 174–178 (2002).
50. Tahiliani, M. *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930–935 (2009).
51. Yang, C. G. *et al.* Crystal structures of DNA/RNA repair enzymes AlkB and ABH2 bound to dsDNA. *Nature* **452**, 961–965 (2008).
52. Hu, L. *et al.* Crystal structure of TET2–DNA complex: insight into TET-mediated 5mC oxidation. *Cell* **155**, 1545–1555 (2013).
53. Bang, J., Bae, S. H., Park, C. J., Lee, J. H. & Choi, B. S. Structural and dynamics study of DNA dodecamer duplexes that contain un-, hemi-, or fully methylated GATC sites. *J. Am. Chem. Soc.* **130**, 17688–17696 (2008).
54. Tronche, F., Rollier, A., Bach, I., Weiss, M. C. & Yaniv, M. The rat albumin promoter: cooperation with upstream elements is required when binding of APF/HNF1 to the proximal element is partially impaired by mutation or bacterial methylation. *Mol. Cell. Biol.* **9**, 4759–4766 (1989).
55. Sugimoto, K., Takeda, S. & Hirochika, H. Transcriptional activation mediated by binding of a plant GATA-type zinc finger protein AGP1 to the AG-motif (AGATCCAA) of the wound-inducible Myb gene *NtMyb2*. *Plant J.* **36**, 550–564 (2003).
56. Lichtsteiner, S. & Schibler, U. A glycosylated liver-specific transcription factor stimulates transcription of the albumin gene. *Cell* **57**, 1179–1187 (1989).
57. Ratel, D. *et al.* The bacterial nucleoside $N^6$-methyldeoxyadenosine induces the differentiation of mammalian tumor cells. *Biochem. Biophys. Res. Commun.* **285**, 800–805 (2001).
58. van Blokland, R., Ross, S., Corrado, G., Scollan, C. & Meyer, P. Developmental abnormalities associated with deoxyadenosine methylation in transgenic tobacco. *Plant J.* **15**, 543–551 (1998).
59. Roberts, D., Hoopes, B. C., Mcclure, W. R. & Kleckner, N. Is10 transposition is regulated by DNA adenine methylation. *Cell* **43**, 117–130 (1985).
60. Hernday, A., Krabbe, M., Braaten, B. & Low, D. Self-perpetuating epigenetic pili switches in bacteria. *Proc. Natl Acad. Sci. USA* **99**, 16470–16476 (2002).
61. Rogers, J. C. & Rogers, S. W. Comparison of the effects of $N^6$-methyldeoxyadenosine and $N^5$-methyldeoxycytosine on transcription from nuclear gene promoters in barley. *Plant J.* **7**, 221–233 (1995).
62. Pratt, K. & Hattman, S. Deoxyribonucleic-acid methylation and chromatin organization in *Tetrahymena thermophila*. *Mol. Cell. Biol.* **1**, 600–608 (1981).
63. Huff, J. T & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* **156**, 1286–1297 (2014).
64. Katz, D. J., Edwards, T. M., Reinke, V. & Kelly, W. G. A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* **137**, 308–320 (2009).
65. Greer, E. L. *et al.* A histone methylation network regulates transgenerational epigenetic memory in *C. elegans*. *Cell Rep.* **7**, 113–126 (2014).
66. Poole, A., Penny, D. & Sjoberg, B. M. Confounded cytosine! Tinkering and the evolution of DNA. *Nat. Rev. Mol. Cell Biol.* **2**, 147–151 (2001).
67. Kierzek, E. & Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing $N^6$-alkyladenosines and 2-methylthio-$N^6$-alkyladenosines. *Nucleic Acids Res.* **31**, 4472–4480 (2003).
68. Roost, C. *et al.* Structure and thermodynamics of $N^6$-methyladenosine in RNA: a spring-loaded base modification. *J. Am. Chem. Soc.* **137**, 2107–2115 (2015).
69. Liu, N. *et al.* $N^6$-methyladenosine-dependent RNA structural switches regulate RNA–protein interactions. *Nature* **518**, 560–564 (2015).
70. Fu, Y. *et al.* FTO-mediated formation of $N^6$-hydroxymethyladenosine and $N^6$-formyladenosine in mammalian RNA. *Nat. Commun.* **4**, 1798 (2013).
71. Capuano, F., Mulleder, M., Kok, R., Blom, H. J. & Ralser, M. Cytosine DNA methylation is found in *Drosophila melanogaster* but absent in *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and other yeast species. *Anal. Chem.* **86**, 3697–3702 (2014).

### Competing interests statement

The authors declare competing interests: see Web version for details.