nature genetics

Review article

https://doi.org/10.1038/s41588-025-02303-1

Non-CG DNA methylation in animal genomes

Received: 26 March 2024

Accepted: 21 July 2025

Published online: 11 September 2025

Check for updates

Thirsa Brethouwer [1], Alex de Mendoza [2,3] ✉ & Ozren Bogdanovic [1,4] ✉

Cytosine DNA methylation is widespread in animal genomes and occurs predominantly at CG dinucleotides (mCG). While the roles of mCG, such as in genomic imprinting and genome stability, are well established, non-CG DNA methylation (mCH) remains poorly understood. In most vertebrate tissues, roughly 80% of CGs are methylated, whereas mCH levels are generally low, typically ranging from 1% to 3%. In vertebrates, mCH is most prevalent in neural tissue, oocytes and embryonic stem cells and has been linked to neurodevelopmental disorders. Moreover, mCH appears to have a conserved role in regulating vertebrate neural genomes, and recent studies suggest that it has functions in the embryogenesis of teleost fish. Overall, mCH represents an intriguing emerging aspect of gene regulation with potential implications for cellular identity, repeat silencing and neural function. In this Review, we provide a critical overview of the patterning, mechanisms and functional implications of mCH in animals.

DNA methylation (5-methylcytosine (5mC)) is a major covalent modification of eukaryotic genomes mostly associated with transcriptional repression[1]. The deposition of 5mC in animals is catalyzed by DNA methyltransferases (DNMTs), which are broadly classified into maintenance methyltransferases (DNMT1), de novo methyltransferases (DNMT3A and DNMT3B) and tRNA methyltransferases (DNMT2)[2]. Moreover, 5mC can be read by methyl-CpG-binding proteins (MBDs)[3] or transcription factors[4] and removed either passively through replication-dependent dilution or actively via iterative oxidation catalyzed by ten-eleven translocation (TET) enzymes[5–7]. 5mC is required for cellular differentiation and animal embryonic development, and deletions or mutations in DNMTs[8,9], MBDs[10,11] and TETs[12–14] frequently result in severe developmental phenotypes or embryonic death. Besides its roles in differentiation, 5mC has been implicated in a myriad of diseases, including cancers[15,16], neurodevelopmental disorders[17,18], immune disorders[19,20] and aging[21]. In animals, 5mC predominantly occurs at CG dinucleotides (mCG), with ~80% of CG dyads methylated in vertebrate genomes[22]. By contrast, nonvertebrate species tend to have more sparsely methylated genomes, with mCG typically ranging from 0% to 12%[23]. Noncanonical DNA methylation (mCH, where H = A, C or T) is a common phenomenon in plants, where it exists in either the CHG or CHH contexts and is mostly associated with repetitive elements[24,25] (Box 1). In mammals, mCH was already observed decades ago[26,27]; however, its function remained disputed for a long time, mostly because of the low genomic levels in most tissues and its colocalization with mCG, which are suggestive of DNMT off-targeting effects or bisulfite conversion artifacts[28,29]. These low levels stem from the differences in the maintenance of each context; mCG is faithfully maintained by DNMT1, whereas mCH requires constant remethylation after each cell division (Fig. 1a). The advent of massively parallel sequencing technologies and whole-genome bisulfite sequencing (WGBS)[30,31] allowed for an unprecedented view of sequence contexts of 5mC and their classification based on cell types and genomic features. In addition to confirming previous findings from plant genomes, these studies uncovered substantial mCH enrichment in the CAG context within gene bodies of human embryonic stem cells (hESCs)[32] and germ cells[33,34], as well as prominent mCH in the CAC context in mammalian neurons[35] (Fig. 1b,c). In addition to identifying the sequence and cell-type specificity of mCH, WGBS enabled precise quantification of this modification, which is commonly represented either as the proportion of methylated cytosines in the mCH context (for example, pie charts in Fig. 1c) or as global mCH levels, calculated as the fraction of reads supporting methylation at CH sites over the total reads covering CH positions genome-wide (Fig. 1c and Supplementary Table 1). Since then, studies have extended to other animals from both vertebrate and nonvertebrate lineages, providing important insights into the function and genomic patterning of mCH[36–41]. Nevertheless, the extent to which mCH contributes to genome regulation remains unclear. In mammals, investigating this is challenging due to the shared mechanism of mCG and mCH deposition, as both are deposited by the DNMT3 family of enzymes. Consequently, generating a knockout

[1]Centro Andaluz de Biología del Desarrollo, CSIC-UPO-JA, Seville, Spain. [2]School of Biological and Behavioural Sciences, Queen Mary University of London, London, UK. [3]Centre for Epigenetics, Queen Mary University of London, London, UK. [4]School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, New South Wales, Australia. ✉e-mail: a.demendozasoler@qmul.ac.uk; o.bogdanovic@csic.es

https://doi.org/10.1038/s41588-025-02303-1

BOX 1

mCH in other eukaryotes

Although this Review focuses on the presence and function of mCH in animals, mCH is also highly abundant and well-studied in other eukaryotes, particularly in plants[152]. In plants, mCH tends to be enriched at repetitive elements in CHH and CHG contexts. Repeat-associated mCH is primarily linked to transposon silencing, although in some species, it is also associated with gene silencing[111,153–155]. While there is some redundancy between mCH and mCG in silencing transposons and variation between plant species, mCH appears to be the predominant factor mediating CG-poor transposon silencing. In plants, mCHH is deposited through de novo RNA-directed DNA methylation targeting domain-rearranged methylases, early plant orthologs of DNMT3 (ref. 153) or by the plant-specific chromomethylase (CMT) CMT2 (ref. 156). mCHG, in the symmetrical context, is maintained by plant-specific CMTs, which functionally resemble DNMT1 in their role as maintenance methyltransferases, although they are structurally distinct and specific to plants[24,157]. Notably, mCH in CHH and CHG contexts occurs via distinct pathways associated with different functional roles—mCHG contributes to stable transposon silencing, whereas mCHH is more dynamic and often guided by small RNA pathways[24]. Beyond plants, some fungi also exhibit relatively high levels of mCH, predominantly found in repetitive sequences and transposable elements, where it is similarly associated with their silencing[152,158,159]. Fungi possess several methyltransferases, including DNMT1 and DNMT5 homologs, and the repeat-induced point mutation defective and Masc1 methyltransferases, which are believed to mediate de novo methylation in both CG and CH contexts, although it is hard to predict the preferred methylation context based exclusively on DNMT repertoires[158,160–163]. Other divergent eukaryotic lineages, including the ameba *Naegleria* or dinoflagellates, also use CH methylation to mark silenced transposable elements[23,164]. By contrast, many unicellular eukaryotes have exclusive mCG, including many algal lineages or protists closely related to animals[165–167]. In summary, mCH has been recurrently deployed to transposon silencing across diverse eukaryotic lineages, with lineage-specific methyltransferases and regulatory pathways adapting this mechanism to their genomic contexts. These similarities underscore the evolutionary importance of mCH beyond the animal kingdom.

model that selectively affects mCH is challenging because DNMT3 loss disrupts both marks. At the same time, recent advances in Oxford Nanopore Technologies (ONT) have enabled more accurate detection of DNA modifications, including mCH, particularly across long and repetitive regions that were previously inaccessible with short-read sequencing[42,43]. In this Review, we summarize current knowledge on noncanonical DNA methylation in animal genomes, focusing on its patterning, mechanisms and functional implications in vertebrates and beyond. We also discuss the remaining challenges and how technological advancements will enhance the study of this gene-regulatory mark.

mCH during pluripotency and germline formation

Although mCH is very low in most mammalian tissues and cell types[44], substantial mCH levels have been detected in female germ cells, hESCs and induced pluripotent stem cells (iPSCs)[45] (Supplementary Table 1). In hESCs, mCH, mainly in the CAG context, is enriched in gene bodies (Fig. 2a) but depleted from transcription factor-binding sites and distal regulatory elements[32,46]. A similar distribution was observed in iPSCs, although differences in mCH content and localization exist between iPSCs and hESCs[29,47]. For example, iPSCs exhibit large differentially methylated regions (DMRs) of hypomethylated CH that overlap with regions of differential CG methylation[47]. These hypomethylated CH DMRs are found in regions enriched for histone 3 lysine 9 trimethylation (H3K9me3), typically associated with transcriptional silencing, indicating broader deficiencies in epigenome reprogramming characteristic of the iPSC state (Fig. 2a). In general, primed iPSCs, generated using Yamanaka factors[48], retain features of somatic cell memory, which can be defined as the epigenetic resemblance to their tissue of origin[49–51]. However, alternative culture methods have been developed to produce naive iPSCs, which exhibit global DNA hypomethylation, reminiscent of natural early embryonic stages, including at imprinted regions[52–55]. A hybrid approach combining naive and primed culture conditions, named transient-naive-treatment reprogramming, has been shown to generate iPSCs more closely resembling hESCs, while also improving consistency and reducing heterogeneity during reprogramming[56] (Fig. 2a). Notably, human[57] and mouse embryos[58] both display negligible mCH levels, whereas in mouse ESCs, mCH was only found associated with major satellite repeat (mSat) classes and not in gene bodies as in hESCs[59]. Overall, mCH patterns appear to correlate with in vitro pluripotency status and may serve as a biomarker for assessing reprogramming efficiency and stem cell identity[60,61].

In mammalian germ cells, mCH is mainly present in the female germline, where it accumulates during oocyte growth. By contrast, male germ cells exhibit peak mCH levels in gonocytes and neonatal prospermatogonia, but these levels drop sharply in mature sperm, where mCH is largely absent[62,63]. Unlike sperm, oocytes maintain relatively high levels of mCH, which increase with maturation, particularly at CAG and CAC trinucleotides[33,64,65]. Interestingly, in mouse oocytes, ~65% of all methylated cytosines are found at mCH sites[33,66]. As in hESCs and iPSCs, mCH in oocytes largely overlaps with mCG and is enriched in gene bodies and repetitive regions[33,64,67]. Moreover, a study in mice revealed that oocytes from older animals exhibit a global reduction in mCG, accompanied by an increase in mCH. As age is associated with reduced oocyte viability, mCH may thus serve as a suitable marker to assess oocyte quality[68].

While both oocytes and hESCs display mCH enrichment in similar sequence contexts and at similar levels, how mCH is deposited in these two systems is somewhat different. Both DNMT3A and DNMT3B contribute to mCH in hESCs (Fig. 2a); however, DNMT3B preferentially does so in the CAG context, whereas DNMT3A shows a higher preference for CAC[69]. The preference for a G or a C in the third base downstream of the methylated C is also evident for DNMT3A- and DNMT3B-mediated mCG deposition patterns[70]. In hESCs, mCAG is associated with transcriptionally active genes[32], which is potentially due to the ability of DNMT3B to interact with the active histone mark H3K36me3, via its Pro–Trp–Trp–Pro (PWWP) domain[71,72] (Fig. 2a). By contrast, in oocytes, DNMT3A and DNMT3L appear to be largely responsible for the deposition of mCH[33,66], although some evidence supports a role for DNMT3B in human oocytes as well[67]. In mammals, DNMT3A and DNMT3L are recruited to genomic regions by their ATRX–DNMT3–DNMT3L (ADD) domains, which recognize H3K4me0 (unmethylated H3K4)[73]. Interestingly, double knockout (dKO) mice with nonfunctional ADD domains in both DNMT3A and DNMT3L exhibited a global loss of mCG, but a surprising accumulation of mCH[66]. Although overall mCG levels were reduced, the regions showing mCH accumulation overlapped with a small number of loci that also exhibited increased mCG in the dKOs, despite the global loss of mCG. This raises the question of whether mCH in mammals can be deposited independently of mCG and to what extent mCH may have functions distinct from those of mCG. Notably, none of the histone marks analyzed in the dKO study, including H3K36me2 and H3K36me3, were detected at mCH-enriched loci. This suggests that the PWWP domain of DNMT3A, which normally recognizes H3K36me2/H3K36me3, does not have a major role in chromatin binding related

https://doi.org/10.1038/s41588-025-02303-1

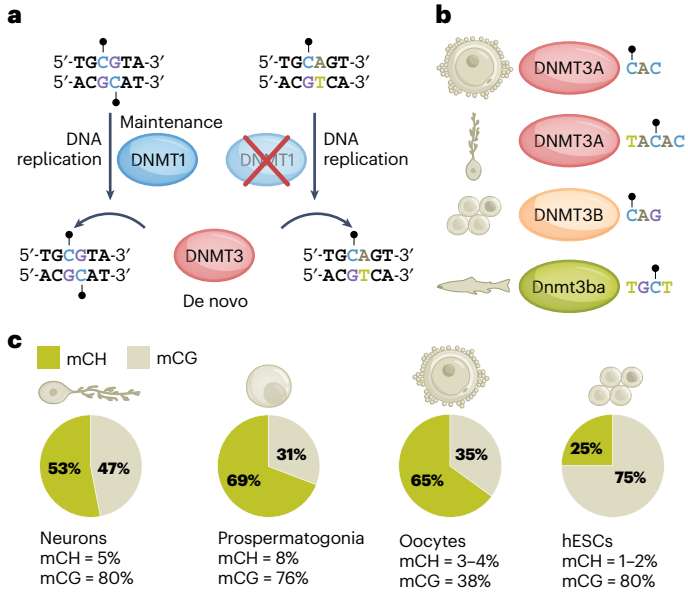

Fig. 1 | Mechanisms of mCH deposition and genomic mCH content across diverse cell types. a, De novo methyltransferases (DNMT3A/DNMT3B) and maintenance methyltransferase (DNMT1) both contribute to the patterning of mCG; most notably, DNMT1 ensures faithful propagation of mCG during DNA replication. mCH is not maintained by DNMT1 during cell division and is dependent exclusively on de novo deposition by DNMT3 enzymes. Black circles indicate methylated cytosines. **b**, Examples of vertebrate DNMT3 enzymes and their preferred substrates for mCH deposition. Mammalian DNMT3A enzymes may display differences in sequence preferences depending on the tissue of action (that is, oocyte CAC and brain TACAC). A teleost-specific enzyme Dnmt3ba recognizes clustered TGCT sequences within MoSAT. **c**, mCH and mCG expressed as a percentage of total genomic 5mC (pie charts), and global mCH levels expressed as mCH percentage (mCH/CH). Although global levels of mCH are far lower than those of mCG, when considering the proportion of methylated cytosines, mCH is more abundant than mCG in a number of tissues, including human neurons, mouse neonatal prospermatogonia and mouse oocytes. A considerable proportion of mCH (~25%) can be found in hESCs, although mCG is more prevalent in this cell type.

to mCH deposition in this dKO context[66]. These findings raise the possibility that sequence context or underlying DNA structure are key determinants of oocyte-specific mCH deposition. Collectively, the data support the notion that mCH is a conserved feature across both pluripotent and germline states, although established through distinct mechanisms. Moreover, the observed correlation between mCH and developmental state highlights the potential of mCH as a valuable biomarker for evaluating pluripotency, reprogramming fidelity and oocyte quality.

Readers, writers, regulatory impact and phenotypes of neural mCH

In mammalian brains, mCH, mostly found in the CAC context, is the predominant form of 5mC, even more abundant than mCG[35]. Brain mCH is mainly present in gene bodies, and gene-body mCH levels are strongly anticorrelated with gene expression (Fig. 2a). In both humans and mice, mCH is absent at birth but rapidly accumulates postnatally, particularly in the nervous system[35,74]. mCH is mostly found in neurons, with much lower levels observed in glial cells. Nevertheless, even in glia, mCH contributes to gene silencing, particularly at neuronal genes that are often hypomethylated in neurons[35]. Functionally, mCH is linked to gene silencing across the vertebrate lineage[36], with a notable preference for repressing long genes and those involved in neural development[41,75]. The main writer of neural mCH in vertebrates is DNMT3A[36], which is recruited to genomic regions marked by H3K36me2 and excluded from regions enriched in active histone marks, such as H3K36me3 and H3K4me2/H3K4me3 (ref. 76). Notably, mCH is typically absent in regions with low chromatin accessibility, likely because DNMT3A cannot access these areas; however, in more accessible regions, mCH levels correlate with chromatin openness[35]. Building on this, recent findings support a model in which H3K36me2 is deposited broadly across topologically associating domains, establishing regional mCH set points in the postnatal brain[76]. Gene expression then promotes the local conversion of H3K36me2 to H3K36me3, particularly in gene bodies and intragenic enhancers, leading to local depletion of mCH due to reduced DNMT3A affinity. This two-step mechanism, which involves regional patterning by H3K36me2, followed by expression-driven modification, accounts for both the megabase-scale mCH profiles and gene-specific variation in mCH levels.

Although most extensively studied in mammals, neural mCH is also found in other vertebrates[36,41,77] and is primarily present in the CAC context, which appears highly conserved throughout vertebrate evolution. Despite the general association between mCH and gene repression, some species, such as zebrafish, opossum, elephant shark and lamprey, were not found to exhibit a clear negative correlation between mCH and gene expression in our comparative study[36]. This discrepancy may stem from technical limitations, including sampling issues in species with smaller brains, which could result in a relatively higher proportion of glial cells and thus lower apparent mCH levels. Indeed, our more recent study, conducted on zebrafish brains, revealed a strong anticorrelation between mCH and gene expression, as well as mCH preference for long genes, as observed in mammals[41]. Zebrafish neural mCH was found to be deposited in the CAC context by two teleost paralogues of DNMT3A, Dnmt3aa and Dnmt3ab, which argues for a deeply conserved gene-regulatory mechanism important for nervous system development in divergent vertebrate species[41].

Currently, MeCP2 is the only known reader of mCH; however, it binds to both mCH and mCG[78,79] (Fig. 2a). MeCP2 in neurons is expressed at near histone-octamer levels and is characterized by broad genome-binding patterns that have hindered the precise identification of its genomic binding sites[80,81]. Mutations in *MECP2* cause Rett syndrome (RTT), a severe neurodevelopmental disorder affecting brain development and function[82]. RTT primarily affects female individuals due to its X-linked dominant inheritance pattern, with male hemizygosity usually being lethal. *MECP2* mutations vary widely in severity, with some mutations resulting in milder phenotypes that are not classified as RTT; phenotypes range from mild intellectual disability to profound neurological impairment[82]. MeCP2 is thought to be recruited to mCH-enriched regions, where it facilitates the recruitment of the NCoR corepressor complex[75], an interaction that is directly compromised by RTT-associated mutations[83]. Although MeCP2 binds both mCH and mCG, elegant experiments using a chimeric version of MeCP2 that can bind only mCG, but not mCH, still resulted in Rett-like symptoms in mice[84]. This strongly suggests that mCH is the primary regulatory signal for MeCP2 in the brain. Supporting this, the loss of neural mCH in mice, obtained by conditional disruption of DNMT3A in postnatal brains, leads to phenotypes resembling those observed in RTT[85,86]. While compensatory mechanisms may exist, such as PRC2-mediated H3K27me3 accumulation at some mCH target genes in DNMT3A-deficient mice, these do not fully restore normal gene regulation[86]. It is also possible that noncatalytic functions of DNMT3A may have an important role in postnatal neurons beyond mCH deposition[87]. Overall, transcriptomic studies across diverse RTT models show that MeCP2 loss results in widespread, albeit modest, dysregulation of hundreds to thousands of genes[88], many of which are involved in neural function[89], with some appearing to be conserved across vertebrates[90].

Beyond the broad differences between neurons and glia, recent studies have explored mCH heterogeneity across neuronal subtypes. Neurons from different brain regions display global differences in mCH levels[74,91]. Despite this variation, a conserved set of MeCP2-repressed

https://doi.org/10.1038/s41588-025-02303-1

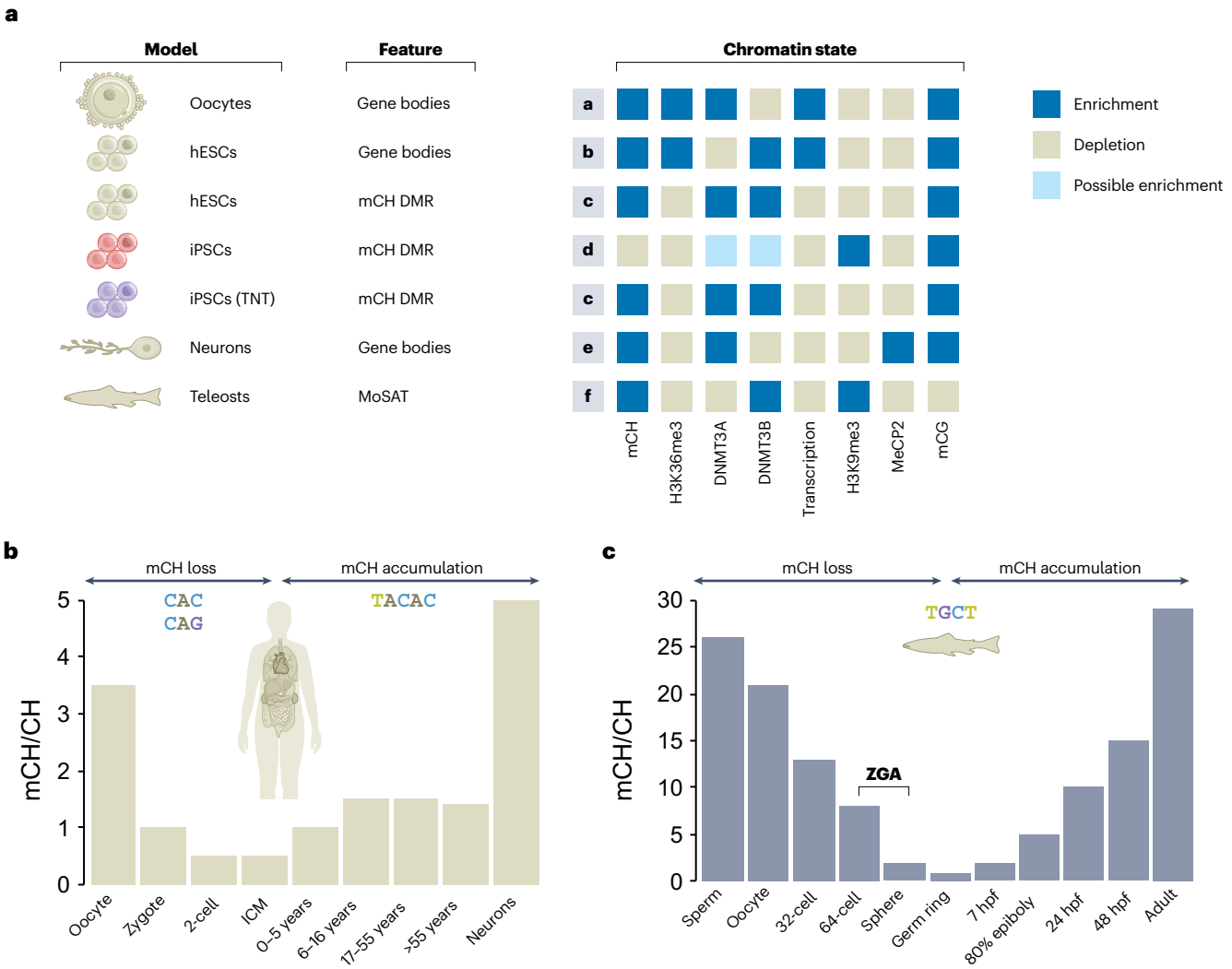

Fig. 2 | Chromatin context and developmental dynamics of mCH in vertebrates. a, Schematic illustration of various chromatin states associated with mCH. From top to bottom—in both oocytes and hESCs, DNMT3A and DNMT3B deposit mCH and mCG, respectively, across gene bodies that are enriched for H3K36me3 and are transcriptionally active (states a and b). Megabase-scale mCH-DMRs, likely established by a combination of DNMT3A and DNMT3B enzymes, are present near centromeres and telomeres in hESCs (state c). Standard iPSC reprogramming methods lead to the loss of this methylation and accumulation of H3K9me3 (state d), whereas a new iPSC reprogramming approach (transient-naive-treatment reprogramming; TNT) restores these regions to the canonical hESC state (state c). In vertebrate neurons, mCH is deposited by DNMT3A, associated with transcriptionally silent genes, and bound by MeCP2 (state e). In teleosts, MoSATs are enriched in mCH in the TGCT sequence context and H3K9me3. MoSAT mCH is deposited by the DNMT3B orthologue Dnmt3ba (state f). Notably, these repeats are depleted of mCG due to the absence of the CG dinucleotide from MoSAT sequences. **b**, mCH levels throughout the human life cycle. Human oocytes and early embryos show relatively high levels of mCH in both CAG and CAC contexts, which are lost during embryonic development. After birth, mCH is re-established, primarily in neuronal tissues and specifically in the CAC context (for example, TACAC). Age brackets (0–5, 6–16, 27–55 and >55 years) correspond to human frontal cortex samples. ICM, inner cell mass. **c**, In zebrafish embryos, MoSAT (TGCT) mCH is high in germ cells and decreases after fertilization during early cleavage stages (32-cell and 64-cell stages), reaching a minimum around ZGA. From gastrulation (germ ring to 80% epiboly), mCH levels increase sharply and remain high in adult tissues derived from all three embryonic germ layers. hpf, hours post fertilization.

mCH target genes appears across all neuron types[91]. Additionally, each neuronal subtype shows specific mCH depletion at certain genes. It has been proposed that high expression of these genes during early postnatal development may prevent DNMT3A from depositing mCH, as active histone marks, known to negatively correlate with mCH, are enriched at highly transcribed loci[76]. The use of single-cell bisulfite sequencing has greatly enhanced our understanding of mCH heterogeneity and cell-type specificity. Two recent large-scale single-cell studies in mouse and human brains confirmed distinct mCH profiles across neuron types; in fact, mCH appears to be a more effective marker for distinguishing neuron types than mCG[92,93]. Both studies also reported an interplay between three-dimensional (3D) chromatin structure and mCH. Notably, long genes enriched with mCH were less likely to form active chromatin loops, while both mCH and mCG within gene bodies showed a negative correlation with loop boundaries at transcription start and end sites[93]. This suggests that both mCH and MeCP2 act to suppress loop formation and maintain gene silencing during differentiation. Altogether, current evidence indicates that while mCH is not required to initiate neuronal differentiation, it has a crucial role in stabilizing or 'locking in' the differentiation programs as they unravel.

mCH in embryonic development and evolution

In contrast to in vitro cultured ESCs, where the presence and levels of mCH are highly dependent on the species of origin and specific to cell culture conditions[32,56,69], mCH is most abundant in mammalian tissues, such as prospermatogonia, oocytes, skeletal muscle and the

brain[33–35,44,67,94]. These are nondividing, postmitotic cell types that must maintain cellular integrity over extended periods of time, which in humans often means decades. In these long-lived cells, mCH thus likely accumulates to support stable, long-term gene regulation that is driven by cell-type-specific activity of DNMT3 enzymes and reinforced by the absence of DNA replication, allowing for gradual mCH buildup without dilution (Fig. 2b). Consequently, the absence of DNMT3 activity leads to rapid loss of mCH, a phenomenon particularly evident during early embryonic development in both mice and humans[57,58]. In mammals, mCH is primarily inherited from the oocyte, as sperm is largely devoid of mCH. Following fertilization, mCH levels decline with each cell division; in mice, approximately 3% of cytosines in oocytes are methylated in the mCH context, which drops to around 1% at the two-cell stage and is virtually absent in the inner cell mass[58]. Similar levels and developmental dynamics have been observed in humans[57]. Thus, mCH is lost during the very early stages of embryogenesis and re-emerges later during the development of the nervous system, particularly in neurons, as well as during oogenesis (Fig. 2b).

In zebrafish, we observed high levels of mCH at mosaic satellite repeats (MoSAT) within the TGCT sequence context (Fig. 2a). These repeats are located in the introns of long genes and in intergenic regions and are generally enriched in the constitutive heterochromatin histone modification H3K9me3 (ref. 37). Notably, these regions lack mCG due to the absence of CG dinucleotides from MoSAT sequences. MoSAT mCH patterns are established by a teleost-specific DNMT3 ortholog, Dnmt3ba, which is unique in that it contains a calponin-homology domain[95]. Calponin-homology domains are typically associated with actin binding[96], making their function in a DNMT protein unclear. While it is possible that calponin-homology domains contribute to mCH deposition, this hypothesis remains to be experimentally validated. In zebrafish, mCH is enriched in the gametes, with both oocytes and sperm exhibiting high mCH levels (Fig. 2c). During early embryonic development, MoSAT mCH is diluted, reaching its lowest point around the time of zygotic genome activation (ZGA). Following ZGA, mCH levels increase; however, unlike in mammals, MoSAT mCH in zebrafish reaccumulates not only in the brain but also in adult tissues derived from all three germ layers[37] (Fig. 2c). In medaka, a distantly related teleost that diverged from zebrafish approximately 150–200 million years ago, MoSAT mCH is present exclusively in oocytes but not in sperm and is lost after fertilization[38]. This pattern mirrors the dynamics observed in zebrafish, where mCH levels reach their lowest point at the time of ZGA before being re-established in differentiating embryos[38]. One possible explanation for the absence of MoSAT mCH in medaka sperm, when compared to zebrafish, might be the differences in sperm chromatin structure between the two species. Zebrafish sperm chromatin is entirely nucleosome based and lacks protamines[97], whereas medaka sperm is protamine packed[98]. This suggests that the targeting or maintenance of MoSAT-associated mCH by Dnmt3ba may require nucleosome-based chromatin as a substrate. Notably, we found that in medaka–zebrafish hybrids (medaka—male and zebrafish—female), mCH patterning, including its loss around ZGA and reestablishment post-ZGA, was maintained, indicative of a compatibility between the Dnmt3ba enzymes and their target sequences in these distantly related teleosts[38].

Outside the vertebrate lineage, overall 5mC levels in both mCG and mCH contexts are generally lower[23]. While accumulating evidence suggests that mCH may be present in various species and have essential regulatory functions, these findings require cautious interpretation due to several confounding factors. One prominent example comes from eusocial insects, such as honeybees and ants. In honeybees, mCH, most notably in the CA context, has been detected in adult heads, with a small but observable difference in mCH levels between queen and worker heads, and this difference has been associated with *dnmt3* expression levels[99]. mCH in honeybees is predominantly found in exons; however, its relationship with transcription and potentially mRNA splicing in this system requires further investigation[99,100]. It is important to note, however, that the mCH levels reported in honeybees are very low (~0.2% mCH) and that they lack a particular motif such as the one identified in vertebrates[99]. Nevertheless, based on previous observations in mammals and other vertebrates regarding the accumulation of mCH in the nervous system[35,36,41], it is plausible that mCH in insects may either represent a deeply conserved feature of neural development or be the result of off-target mCH accumulation that persists due to the absence of postmitotic mCH removal. Beyond honeybees, mCH has been detected in a variety of nonvertebrate lineages, including the ants *Camponotus floridanus* and *Harpegnathos saltator*[39], the lepidopteran *Helicoverpa armigera*[101], pearl oyster *Pinctada fucata martensii*[102], sea cucumber *Apostichopus japonicus*[103] and the beetle *Tribolium castaneum*[104], among others. Notably, in the beetle, methylation in the CHH context (primarily CA) was the predominant form, accounting for 75% of methylated cytosines and occurring mainly in intergenic regions and introns[104]. However, it is well established that *T. castaneum* lacks DNMT3 enzymes and shows mCG levels that are indistinguishable from false positives; therefore, these claims are controversial[105]. Moreover, beetle mCHH was found in intergenic and intronic regions, while in most of the other nonvertebrates outlined above, mCH levels tended to mirror those of mCG and were primarily found within gene bodies. It is important to note that most of these studies do not account for the presence of single-nucleotide polymorphisms (SNPs) or specifically compare dinucleotide levels against a negative spike-in control. For example, some positions may be misclassified as mCH because the reference genome shows a CH dinucleotide at that position, while the sequenced DNA contains a CG SNP that is susceptible to methylation and should be labeled as mCG. By contrast, in our carefully controlled study accounting for genotype correction and false-positive rates, neural tissues from honeybee, octopus and amphioxus did not show mCH levels that convincingly exceeded those of the negative controls[36]. In summary, while 5mC is generally detected at lower levels beyond the vertebrate lineage, mCH may exist across a range of animal species and have regulatory roles in development and behavior; however, carefully designed studies will be required to target this area. The degree to which mCH is tissue and stage specific, as it is in vertebrates, and how it is associated with gene expression remain to be fully explored.

mCH as a regulator of the repetitive genome

Repetitive sequences are estimated to comprise approximately 50–70% of the human genome[106], with this proportion being closer to 50% in zebrafish[107] and significantly higher in some vertebrates, such as lungfish (up to ~90%)[108]. Repetitive elements in vertebrates are predominantly silenced by high levels of mCG and other repressive epigenetic marks, such as H3K9me3, and are typically organized into constitutive heterochromatin[109]. In plant genomes, repetitive regions are marked by mCG and, in many cases, also by mCHH and mCHG, which are maintained by distinct enzymes and mechanisms (mCHG through maintenance and mCHH through de novo methylation)[24,31,110]. These three marks often colocalize and appear to function redundantly, although mCH is less dispensable for transposon silencing[111]. The repetitive elements silenced in plants are primarily transposable elements[24], and their repression is essential for maintaining genome integrity[112]. As in plants, mCH in vertebrates is predominantly enriched at short interspersed nuclear elements and long interspersed nuclear elements (LINEs)[29,113] (Supplementary Table 2); however, the exact role of mCH in regulating repetitive DNA in vertebrates remains underexplored. Supporting the idea that mCH mediates repeat silencing in vertebrates, a study in birds reported a slight but statistically significant negative correlation between mCH levels and transposon expression[114]. Similarly, in zebrafish brains, TDR and TC1DR3 elements, members of the Tc1-mariner DNA transposon superfamily found across eukaryotes, as well as other transposable elements that contain potential MeCP2-binding sites (5mCAC), are also highly methylated at CH sites[41].

This hypothesis is further supported by evidence that, in mammalian brains, MeCP2 is involved in the repression of LINE-1 retrotransposons, suggesting that mCH may contribute to transposable element silencing by facilitating MeCP2 binding[115]. In mammals, mCH is also found at other types of repetitive elements, such as mSat, which exhibit high levels of mCH in mouse ESCs[59]. At these sites, mCH co-occurs with mCG and is mediated by DNMT3A, DNMT3B and DNMT3L[59]. Interestingly, mSat is highly transcribed in mESCs and is associated with proper heterochromatin organization[116]. Similarly, in teleosts, mCH is highly abundant at MoSAT in pre-ZGA embryos and adult tissues, where its primary function may not necessarily be linked to transcriptional repression[37,38]. These findings suggest that mCH may not always act as a silencing mark but could also have roles in gene activation or chromatin organization. This aligns with the observed positive correlation between gene-body mCH and transcriptional activity in hESCs, although it contrasts with the generally repressive role of mCH in neurons[32,35]. Finally, a comparative study examining repeat-associated mCH across various human and mouse cell types and tissues revealed that mCH, particularly at CA and CT sites, is cell-type specific and likely conserved among mammals[117]. Together, these findings underscore the complexity of mCH patterns at repetitive elements in vertebrates, indicating that mCH enrichment at repeats may reflect context-dependent deposition influenced by factors such as cell type or the specific repeat class.

Challenges, pitfalls and emerging technologies for mCH analysis and detection

Although it is now possible to detect 5mC at single-base resolution through WGBS and other emerging technologies (Box 2 and Fig. 3a), several challenges and pitfalls remain when investigating mCH. Experimentally, protocols based on bisulfite conversion remain the gold standard for detecting methylated cytosines; however, this chemical reaction, while highly efficient, is also harsh, leading to substantial DNA degradation in the process[118]. This degradation is not a major issue when working with bulk samples, but it becomes problematic in low-input or single-cell applications, where it can reduce coverage and limit the number of cytosines that can be assessed on a genome-wide scale. Enzymatic conversion methods, such as enzymatic methyl sequencing[119] (EM-seq; Fig. 3a), TET-assisted pyridine borane sequencing[120] and others[121,122], have recently emerged as alternatives; here, in particular, EM-seq offers a strong alternative due to its low false-positive rates in the mCH context[119,123]. ONT can detect 5mC directly by measuring changes in electrical current as DNA strands pass through a nanopore, as methylated bases produce distinct signal patterns[124]. Machine-learning-based tools analyze the changes in signal intensities to distinguish between modified and unmodified bases, thus enabling base-resolution methylation calling without chemical conversion[124,125]. Nevertheless, while mCG detection by ONT is fairly accurate and its major limitation remains input quantity, reliable detection of mCH remains challenging with this platform for modifications with low abundance, such as brain mCH[126–128]. However, plant mCHH or mCHG patterns can be confidently determined from ONT signals using specialized tools[129,130], which is consistent with these methylation contexts being highly abundant in plant genomes.

Another key consideration when studying mCH is the nonconversion rate of unmethylated cytosines[118]. Both bisulfite and enzymatic conversion protocols can fail to convert some unmethylated cytosines (Fig. 3b), leading to false-positive methylation calls[131]. This error can propagate across the genome either uniformly or with sequence composition biases. In species where 5mC is naturally very low or absent, calculating the total number of 5mC calls per sequence context may misleadingly suggest that mCH is the predominant methylation context. Thus, while false positives are less of an issue for animal species with high mCG levels, they pose a major challenge for mCH analysis in species or samples with low overall methylation, such as nonvertebrates. To estimate nonconversion rates, mitochondrial DNA can sometimes serve as a proxy for an unmethylated bisulfite conversion control[132,133]; however, the methylation status of animal mitochondrial DNA remains a topic of intense debate, in part because the 3D configuration of circular genomes might affect bisulfite conversion[134,135]. Therefore, unmethylated spike-in controls are the best option for assessing nonconversion artifacts. The most widely used control to date is unmethylated λ phage DNA[118,136], which enables a reliable, independent assessment of nonconversion rates, thus making it a critical control for ensuring experimental accuracy when studying mCH. Additionally, orthogonal approaches such as EM-seq, ONT or mass spectrometry-based methods such as liquid chromatography–mass spectrometry might be needed depending on the biological question, sequence context and overall mCG/mCH levels in the studied organism.

Beyond experimental considerations, there are also notable pitfalls in the computational analysis of mCH (Fig. 3). While many tools

BOX 2

Methods for mCH profiling

There are several sequencing-based methods available to detect mCH at base-level resolution. The gold standard remains WGBS[30,31]. In this method, sodium bisulfite treatment of DNA converts unmethylated cytosines (C) into uracils (U), which are then amplified as thymines (T) during PCR[168]. Methylated cytosines, however, are protected from conversion. As a result, sequencing reads show unmethylated cytosines as T and methylated cytosines as C. Although nonconversion rates are generally very low, the harsh chemical treatment can lead to substantial DNA degradation[118]. More recently, EM-seq has been developed as an alternative[119]. Like WGBS, EM-seq detects 5mC and 5-hydroxymethylcytosine (5hmC) but uses enzymatic rather than chemical conversion for 5mC/5hmC detection. First, TET2 oxidizes 5mC to 5-carboxylcytosine (5caC) in a reaction where 5hmC is protected from oxidation by a glucose cap added by the T4 phage β-glucosyltransferase (T4-BGT). Then, the cytidine deaminase APOBEC3A converts unmodified cytosines to uracils, whereas modified cytosines (originally 5mC and 5hmC) are protected. As with WGBS, methylated cytosines are detected as C and unmethylated ones as T. The key advantage of EM-seq is that enzymatic treatment causes far less DNA degradation, making it more suitable for low-input or single-cell samples. PacBio sequencing, also known as single-molecule real-time sequencing, enables the sequencing of long reads. Additionally, this technology enables the direct identification of methylated cytosines, eliminating the need for conversion protocols[43,169]. However, PacBio 5mC detection is currently restricted to the CG context. Nevertheless, treating long DNA fragments with EM-seq and then sequencing with PacBio could be an option for more efficient mCH studies. Finally, ONT allows for the direct detection of methylated cytosines without the need for chemical or enzymatic conversion. This approach uses protein nanopores embedded in an artificial membrane. As the DNA strand passes through the pore, it disrupts the electronic current in a sequence- and modification-specific manner[170]. Because different bases and their chemical modifications, including 5mC, have distinct shapes and sizes, they generate unique electrical signatures that can be computationally identified. ONT's main advantages are long-read lengths and the absence of conversion steps, which improve mappability. While mCG calling is well established, detecting mCH remains more challenging in low-abundance cases[171]. Nonetheless, this technology is rapidly improving and is expected to become more sensitive for mCH detection.

https://doi.org/10.1038/s41588-025-02303-1

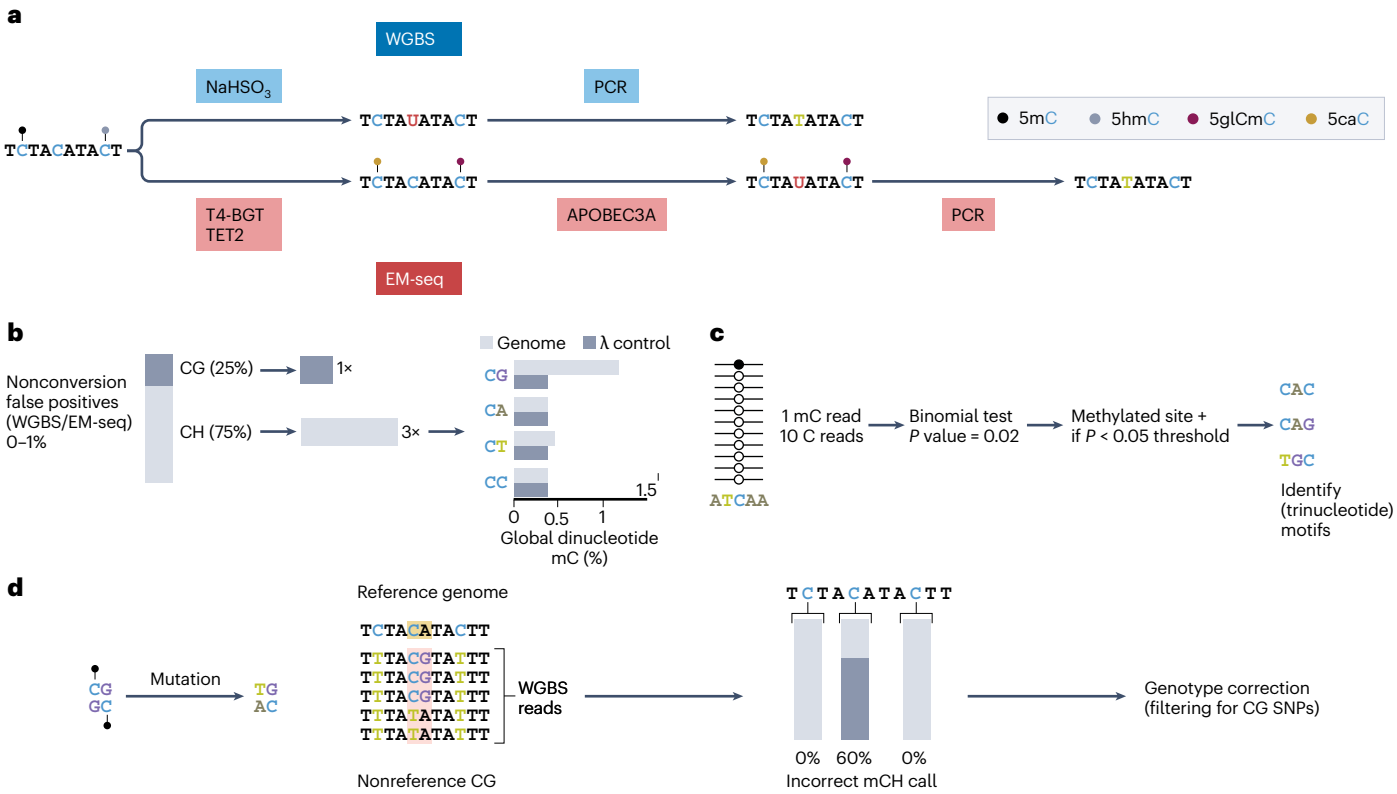

Fig. 3 | Challenges and pitfalls in mCH detection and quantification. **a**, Schematic overview of conversion protocols to detect methylated cytosines at base resolution. In WGBS, during bisulfite ($NaHSO_3$) treatment, unmethylated cytosines are converted to uracils (U), whereas 5mC and 5hmC remain as cytosines. Upon PCR amplification, uracils are replaced with thymines (T); thus, in the final sequencing readout, a cytosine (C) corresponds to a methylated position and a thymine (T) to an unmethylated one. In EM-seq, TET2 oxidizes 5mC to 5caC, while 5hmC is protected from oxidation by the glucose moiety (5glcmC) via the T4-BGT enzyme. Next, APOBEC3A deaminates unmodified cytosines, converting them to uracils, whereas 5caC and 5gmC remain unchanged. During PCR, uracils are replaced by thymines, while 5caC and 5gmC remain as cytosines, thereby yielding the same sequence output as WGBS. **b**, An inherent bias of WGBS and EM-seq is the incomplete conversion of unmethylated cytosines to uracils, resulting in acceptable nonconversion rates of approximately 0–1%. Nevertheless, these artifacts can become problematic in species with low 5mC. Because the CH sequence context is more abundant than CG in most eukaryotic genomes, incomplete conversion can affect mCH calling, leading to false positives. Including controls such as unmethylated λ phage DNA and assessing nonconversion rates in specific sequence contexts (for example, CA, CC and CT) can help mitigate such biases. **c**, While methods such as the binomial test may classify a site as methylated based on a single supporting read, that read could still result from nonconversion, leading to a false-positive 5mC call. Methylation calling should thus be followed by an analysis of the sequence context of the putative 5mC at the trinucleotide level to determine whether methylation occurs in a known DNMT3 context (for example, CAC, CAG or TGC). **d**, Due to 5mC mutability, many CG sites mutate to TG/CA over evolutionary time and are missing from reference assemblies. If genotype information is not considered, methylated CG sites present in the sample but absent from the reference can be misclassified as mCH. Filtering CG SNPs before methylation calling helps prevent such miscalls.

and pipelines exist for analyzing 5mC, most are optimized for mCG[137]. In fact, many pipelines still report mCHH and mCHG as default values to accommodate plant methylation patterns, yet these contexts are not representative of vertebrate mCH (for example, mCAC, mCAG and TGmCT; Fig. 3c). Moreover, a major limitation of traditional short-read sequencing technologies is their reduced effectiveness in repetitive genomic regions[138,139]. This is not a major concern when studying mCH in coding or repeat-free regions, but mCH is also frequently enriched in repetitive elements[29,41,113], which are difficult to map reliably using short reads, especially after conversion of most Cs to Ts. As a result, coverage in these regions may be low, making an accurate assessment of mCH levels challenging. Long-read sequencing technologies have substantially improved repeat-region mapping[140,141], but they are not yet standard due to higher error rates, although these can be mitigated to some extent[142,143]. When investigating repeat-associated mCH, it may be valuable to incorporate long-read approaches, either independently or as a validation tool alongside short-read sequencing. A further computational challenge related to mCH analyses involves accurate methylation calling (Fig. 3d), as CH positions that overlap with CGs and cytosines in ambiguous sequence contexts might result in biased or overestimated mCH[36]. This is particularly important because 5mC is highly mutagenic and often deaminates into thymine. As a result, many individual genomes may contain CGs at positions where the reference genome shows CH (most frequently, CA). This challenge is even greater in nonvertebrates, where heterozygosity tends to be higher than in mammals. To reduce this bias, mCH calls originating from SNP-containing regions or confirmed as CGs in WGBS or EM-seq data should be filtered out.

Finally, identifying DMRs in the CH context remains particularly difficult. While mCG often occurs in dense clusters, enabling detection of discrete DMRs, mCH tends to be more sparsely distributed. As a result, CH DMRs may span much larger genomic regions[32,56], for which conventional DMR calling tools are generally not optimized[144]. For targeted analyses, such as mCH in specific gene bodies or repeat elements, it is possible to manually compute methylation levels across annotated regions. Binning the genome and comparing methylation levels between bins is currently one of the most reliable approaches for detecting CH DMR[56,145]. Although some tools implement this strategy, the development of dedicated, standardized tools for mCH analysis remains essential to advance the field[144]. In summary, while single-base resolution assessment of mCH is feasible, both experimental and computational challenges must be carefully considered. Robust

https://doi.org/10.1038/s41588-025-02303-1

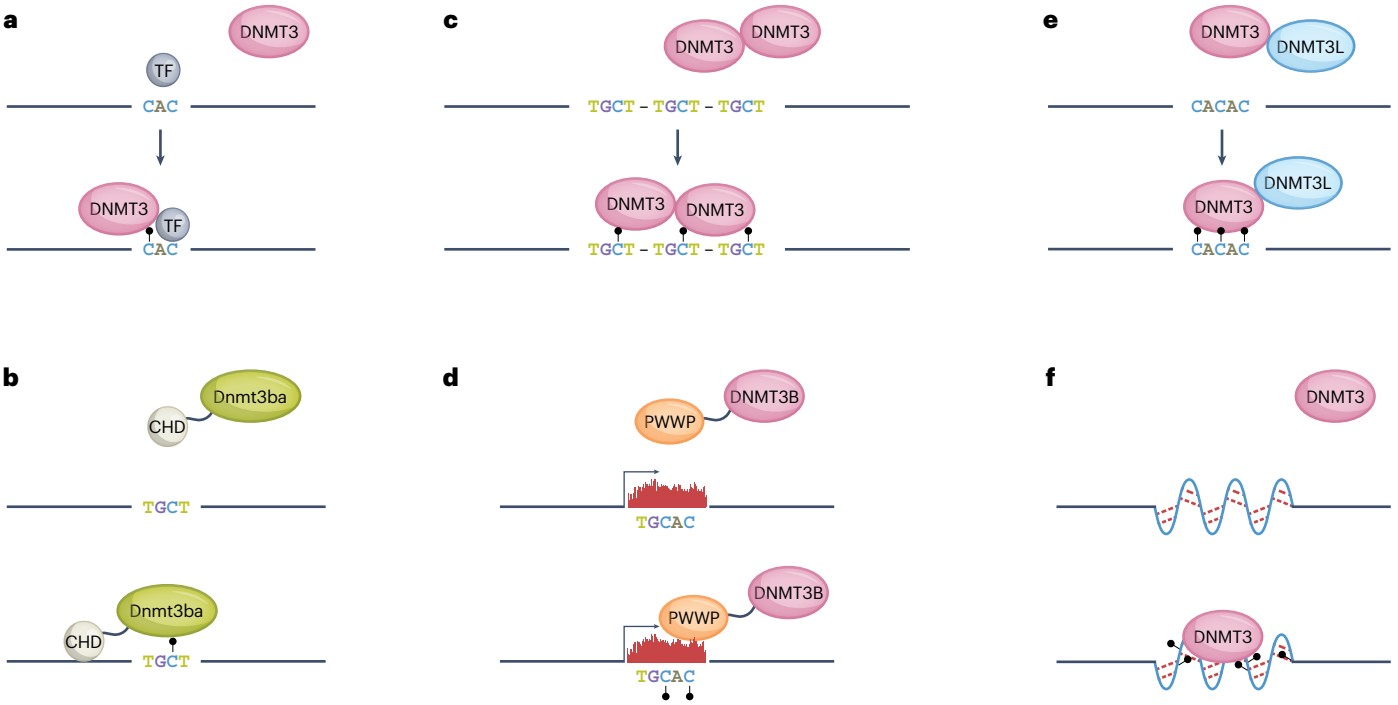

Fig. 4 | Potential mechanisms of mCH deposition. a, Recruitment by sequence-specific transcription factors (TFs). A cofactor, for example, a transcription factor, binds to a given region and recruits DNMT3 to methylate proximal cytosines in the CH context. **b**, Recruitment by specialized DNMT domains. A specific domain of DNMT3 may regulate its targeting to DNA, such as the calponin-homology domain (CHD) specifically found in zebrafish Dnmt3ba, the enzyme responsible for mCH deposition at MoSAT. **c**, Recruitment by sequence density. DNMT3 recognizes specific DNA motifs; here DNMT3 recruitment and subsequent mCH deposition are regulated by motif density. **d**, Recruitment by histone modifications (for example, H3K36me3). Histone modifications such as H3K36me3 can recruit DNMT3B through its PWWP domains, thereby linking mCH deposition to active transcription. **e**, Recruitment by DNMT cooperativity (for example, DNMT3L–DNMT3 interaction). DNMT3 interactions with cofactors such as DNMT3L can stabilize recruitment and promote mCH methylation. **f**, Recruitment by secondary structure. The formation of secondary DNA structures, resulting from DNA interactions (red lines), formed at repetitive elements may further facilitate DNMT3 targeting.

experimental controls, thoughtful selection of detection methods and tailored computational analyses are crucial for obtaining reliable insights into mCH biology.

Future prospects

Although the function of mCH is still not fully understood, it is now widely accepted that this DNA modification has a role in gene regulation across a range of biological contexts. Emerging evidence suggests that mCH may also be present and functional outside of vertebrates. However, rigorously controlled experiments are still needed to determine whether mCH represents a deeply conserved regulatory mechanism, a product of convergent evolution in distinct lineages, or simply a byproduct of DNMT activity with limited, context-dependent functionality. Arguments suggesting that mCH is a byproduct of mCG deposition include the shared enzymatic machinery[45], the frequent genome-wide colocalization of mCH and mCG[29] and functional experiments showing that deletion of the DNMT3A and DNMT3L ADD domains in mice leads to increased mCH at loci where mCG is also upregulated, even in the context of global mCG loss[66]. However, several lines of evidence also argue against this hypothesis. For instance, in mammals, different cell types exhibit markedly different levels of mCH. Neurons and oocytes show relatively high levels, while most adult somatic tissues have virtually none[33,35,44]. If mCH were merely a byproduct of mCG deposition, such variation would be unexpected, particularly given that global mCG levels are relatively stable across cell types. In teleosts, the Dnmt3ba enzyme specifically targets mCH to CG-poor MoSAT repeats[37,38], a phenomenon reminiscent of mCH-mediated silencing of CG-poor repeats in plants. Moreover, in both mammals and teleosts, mCH displays developmental dynamics distinct from those of mCG across multiple tissues[35,37]. However, the clearest insights into potential mCH functions come from the mammalian brain, where mCH likely contributes to gene silencing via the repressive activity of MeCP2, the only known reader of mCH to date[35,75]. MeCP2, and mCH more broadly, may also be involved in silencing mCH-rich repetitive elements in the brain, an underexplored aspect of neurological disorders such as RTT (caused by mutations in *MECP2*)[82] and Tatton–Brown–Rahman syndrome (linked to *DNMT3A* mutations)[146]. Historically, repeats have been difficult to study, but long-read sequencing technologies now enable more accurate characterization of their sequences and methylation landscapes[42]. Regarding the complex relationships between mCH and repeats, many open questions remain, such as to what extent are different repeats methylated? How heterogeneous is this phenomenon across cells and tissues? Is mCH involved solely in silencing, or might it also contribute to repeat activation or other, yet unidentified, functions? Interestingly, the published body of work suggests that mCH-mediated repeat regulation in animals and plants may be more similar than initially thought, with mCH participating in the regulation of CG-poor repeats[29,41,117]. Nonetheless, these hypotheses remain to be rigorously tested to determine whether they reflect true cases of functional convergence across kingdoms.

To fully understand and functionally manipulate mCH, it is essential to gain deeper insight into the mechanisms by which DNMT3 family enzymes mediate mCH deposition. Multiple pathways likely contribute to the recruitment of DNMT3 enzymes to mCH target sites (Fig. 4). Sequence-specific recognition by transcription factors[78] or specialized DNMT domains, such as the Dnmt3ba calponin-homology domain in teleosts[37], may direct DNMT3 activity to specific sequence motifs. Additionally, regions with a high local density of such sequence motifs,

including tandem repeats, may enhance DNMT3 recruitment. The chromatin environment likely has an important role as well, with histone modifications such as H3K36me3 facilitating recruitment of DNMT3 enzymes through reader domains such as PWWP, which recognize both H3K36me2 and H3K36me3. This provides a mechanistic link between mCH deposition and transcriptionally active regions of the genome[46,71]. Protein–protein interactions, including DNMT3 homomeric associations and interactions with cofactors such as DNMT3L, can stabilize recruitment and promote efficient methylation, as previously demonstrated for the example of satellite repeats in mESCs[59]. Finally, the formation of secondary DNA structures at repetitive elements may further facilitate DNMT3 targeting[147]. Together, these data suggest that mCH deposition is likely orchestrated by an interplay of DNA sequence, chromatin landscape and enzymatic cooperativity. Furthermore, beyond MBDs, many transcription factors are known to be sensitive to methylation in the CG context, raising the possibility that mCH may similarly influence the binding of factors that do not specifically recognize CG sites[4,148–150].

Mammalian models have been central to our understanding of mCH function, but studies show that disrupting mCH reading in the brain also affects a substantial proportion (10–20%) of mCG sites, complicating efforts to dissect the specific effects of mCH[85,86]. In this regard, teleosts such as zebrafish, which are potentially equipped with either specialized DNMTs that preferentially deposit mCH or a larger repertoire of DNMTs overall, may offer powerful systems for studying mCH biology in isolation from mCG effects. The presence of multiple DNMTs in the teleost genome[95] may help buffer the loss of mCG in *Dnmt3ba*-knockout models; however, this redundancy appears insufficient to compensate for the loss of mCH, highlighting its distinct regulatory role[37,41]. The combination of long-read sequencing and single-cell technologies provides unprecedented resolution for tackling these questions. Additionally, better ONT detection methods or better mCH pattern recognition can aid in uncovering new patterns and generating testable hypotheses. While these models will likely need to be trained in a context-specific manner and tailored to distinct tissues and cell types, they hold promise for revealing previously unnoticed regulatory phenomena. Furthermore, epigenome engineering approaches could be developed to directly target mCH to specific loci, thereby directly interrogating its function[151]. Overall, while mCH is now established as an emerging gene-regulatory mark in neural tissues, its full range of functions, particularly in gene regulation and repeat element control, remains elusive. It also remains crucial to determine under which conditions mCH serves as a functional signal or when it instead represents a byproduct of DNMT3 binding or mCG deposition. Advances in molecular and computational tools will be key to addressing these gaps, and it is likely that the next few decades will yield major insights into the biology of this intriguing epigenetic modification.

References

1. Schubeler, D. Function and information content of DNA methylation. *Nature* **517**, 321–326 (2015).
2. Jurkowska, R. Z., Jurkowski, T. P. & Jeltsch, A. Structure and function of mammalian DNA methyltransferases. *ChemBioChem* **12**, 206–222 (2011).
3. Bogdanovic, O. & Veenstra, G. J. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* **118**, 549–565 (2009).
4. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
5. Ross, S. E. & Bogdanovic, O. TET enzymes, DNA demethylation and pluripotency. *Biochem. Soc. Trans.* **47**, 875–885 (2019).
6. Pastor, W. A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* **14**, 341–356 (2013).
7. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–479 (2013).
8. Li, E., Bestor, T. H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915–926 (1992).
9. Xu, G. L. et al. Chromosome instability and immunodeficiency syndrome caused by mutations in a DNA methyltransferase gene. *Nature* **402**, 187–191 (1999).
10. Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A. & Bird, A. Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* **15**, 710–723 (2001).
11. Guy, J., Hendrich, B., Holmes, M., Martin, J. E. & Bird, A. A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27**, 322–326 (2001).
12. Dai, H. Q. et al. TET-mediated DNA demethylation controls gastrulation by regulating Lefty–Nodal signalling. *Nature* **538**, 528–532 (2016).
13. Bogdanovic, O. et al. Active DNA demethylation at enhancers during the vertebrate phylotypic period. *Nat. Genet.* **48**, 417–26 (2016).
14. Li, C. et al. Overlapping requirements for Tet2 and Tet3 in normal development and hematopoietic stem cell emergence. *Cell Rep.* **12**, 1133–1143 (2015).
15. Robertson, K. D. DNA methylation, methyltransferases, and cancer. *Oncogene* **20**, 3139–3155 (2001).
16. Nishiyama, A. & Nakanishi, M. Navigating the DNA methylation landscape of cancer. *Trends Genet.* **37**, 1012–1027 (2021).
17. Reichard, J. & Zimmer-Bensch, G. The epigenome in neurodevelopmental disorders. *Front. Neurosci.* **15**, 776809 (2021).
18. Ciptasari, U. & van Bokhoven, H. The phenomenal epigenome in neurodevelopmental disorders. *Hum. Mol. Genet.* **29**, R42–R50 (2020).
19. Ballestar, E., Sawalha, A. H. & Lu, Q. Clinical value of DNA methylation markers in autoimmune rheumatic diseases. *Nat. Rev. Rheumatol.* **16**, 514–524 (2020).
20. Mazzone, R. et al. The emerging role of epigenetics in human autoimmune disorders. *Clin. Epigenetics* **11**, 34 (2019).
21. Horvath, S. & Raj, K. DNA methylation-based biomarkers and the epigenetic clock theory of ageing. *Nat. Rev. Genet.* **19**, 371–384 (2018).
22. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
23. de Mendoza, A., Lister, R. & Bogdanovic, O. Evolution of DNA methylome diversity in eukaryotes. *J. Mol. Biol.* **432**, 1687–1705 (2019).
24. Zhang, H., Lang, Z. & Zhu, J. K. Dynamics and function of DNA methylation in plants. *Nat. Rev. Mol. Cell Biol.* **19**, 489–506 (2018).
25. Kumar, S. & Mohapatra, T. Dynamics of DNA methylation and its functions in plant growth and development. *Front. Plant Sci.* **12**, 596236 (2021).
26. Salomon, R. & Kaye, A. M. Methylation of mouse DNA in vivo: di- and tripyrimidine sequences containing 5-methylcytosine. *Biochim. Biophys. Acta* **204**, 340–351 (1970).
27. Grafstrom, R. H., Yuan, R. & Hamilton, D. L. The characteristics of DNA methylation in an in vitro DNA synthesizing system from mouse fibroblasts. *Nucleic Acids Res.* **13**, 2827–2842 (1985).
28. Patil, V., Ward, R. L. & Hesson, L. B. The evidence for functional non-CpG methylation in mammalian cells. *Epigenetics* **9**, 823–828 (2014).
29. Ziller, M. J. et al. Genomic distribution and inter-sample variation of non-CpG methylation across human cell types. *PLoS Genet.* **7**, e1002389 (2011).
30. Cokus, S. J. et al. Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* **452**, 215–219 (2008).

31. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* **133**, 523–536 (2008).
32. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
33. Shirane, K. et al. Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* **9**, e1003439 (2013).
34. Kubo, N. et al. DNA methylation and gene expression dynamics during spermatogonial stem cell differentiation in the early postnatal mouse testis. *BMC Genomics* **16**, 624 (2015).
35. Lister, R. et al. Global epigenomic reconfiguration during mammalian brain development. *Science* **341**, 1237905 (2013).
36. De Mendoza, A. et al. The emergence of the brain non-CpG methylation system in vertebrates. *Nat. Ecol. Evol.* **5**, 369–378 (2021).
37. Ross, S. E., Angeloni, A., Geng, F. S., de Mendoza, A. & Bogdanovic, O. Developmental remodelling of non-CG methylation at satellite DNA repeats. *Nucleic Acids Res.* **48**, 12675–12688 (2020).
38. Ross, S. E. et al. Evolutionary conservation of embryonic DNA methylome remodelling in distantly related teleost species. *Nucleic Acids Res.* **51**, 9658–9671 (2023).
39. Bonasio, R. et al. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr. Biol.* **22**, 1755–1764 (2012).
40. Klughammer, J. et al. Comparative analysis of genome-scale, base-resolution DNA methylation profiles across 580 animal species. *Nat. Commun.* **14**, 232 (2023).
41. Ross, S. E., Hesselson, D. & Bogdanovic, O. Developmental accumulation of gene body and transposon non-CpG methylation in the zebrafish brain. *Front. Cell Dev. Biol.* **9**, 643603 (2021).
42. Fu, Y., Timp, W. & Sedlazeck, F. J. Computational analysis of DNA methylation from long-read sequencing. *Nat. Rev. Genet.* **26**, 620–634 (2025).
43. Liu, T. & Conesa, A. Profiling the epigenome using long-read sequencing. *Nat. Genet.* **57**, 27–41 (2025).
44. Schultz, M. D. et al. Human body epigenome maps reveal noncanonical DNA methylation variation. *Nature* **523**, 212–216 (2015).
45. He, Y. & Ecker, J. R. Non-CG methylation in the human genome. *Annu Rev. Genomics Hum. Genet.* **16**, 55–77 (2015).
46. Tan, H. K. et al. DNMT3B shapes the mCA landscape and regulates mCG for promoter bivalency in human embryonic stem cells. *Nucleic Acids Res.* **47**, 7460–7475 (2019).
47. Lister, R. et al. Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471**, 68–73 (2011).
48. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
49. Kim, K. et al. Donor cell type can influence the epigenome and differentiation potential of human induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 1117–1119 (2011).
50. Ohi, Y. et al. Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. *Nat. Cell Biol.* **13**, 541–549 (2011).
51. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
52. Liu, X. et al. Comprehensive characterization of distinct states of human naive pluripotency generated by reprogramming. *Nat. Methods* **14**, 1055–1062 (2017).
53. Giulitti, S. et al. Direct generation of human naive induced pluripotent stem cells from somatic cells in microfluidics. *Nat. Cell Biol.* **21**, 275–286 (2019).
54. Pastor, W. A. et al. Naive human pluripotent cells feature a methylation landscape devoid of blastocyst or germline memory. *Cell Stem Cell* **18**, 323–329 (2016).
55. Wang, Y. et al. Unique molecular events during reprogramming of human somatic cells to induced pluripotent stem cells (iPSCs) at naive state. *eLife* **7**, e29518 (2018).
56. Buckberry, S. et al. Transient naive reprogramming corrects hiPS cells functionally and epigenetically. *Nature* **620**, 863–872 (2023).
57. Guo, H. et al. The DNA methylation landscape of human early embryos. *Nature* **511**, 606–610 (2014).
58. Wang, L. et al. Programming and inheritance of parental DNA methylomes in mammals. *Cell* **157**, 979–991 (2014).
59. Arand, J. et al. In vivo control of CpG and non-CpG DNA methylation by DNA methyltransferases. *PLoS Genet.* **8**, e1002750 (2012).
60. Butcher, L. M. et al. Non-CG DNA methylation is a biomarker for assessing endodermal differentiation capacity in pluripotent stem cells. *Nat. Commun.* **7**, 10458 (2016).
61. Joe, S. & Nam, H. Prediction model construction of mouse stem cell pluripotency using CpG and non-CpG DNA methylation markers. *BMC Bioinformatics* **21**, 175 (2020).
62. Ichiyanagi, T., Ichiyanagi, K., Miyake, M. & Sasaki, H. Accumulation and loss of asymmetric non-CpG methylation during male germ-cell development. *Nucleic Acids Res.* **41**, 738–745 (2013).
63. Kobayashi, H. et al. High-resolution DNA methylome analysis of primordial germ cells identifies gender-specific reprogramming in mice. *Genome Res.* **23**, 616–627 (2013).
64. Tomizawa, S. et al. Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138**, 811–820 (2011).
65. Demond, H., Khan, S., Castillo-Fernandez, J., Hanna, C. W. & Kelsey, G. Transcriptome and DNA methylation profiling during the NSN to SN transition in mouse oocytes. *BMC Mol. Cell Biol.* **26**, 2 (2025).
66. Kubo, N. et al. Combined and differential roles of ADD domains of DNMT3A and DNMT3L on DNA methylation landscapes in mouse germ cells. *Nat. Commun.* **15**, 3266 (2024).
67. Yu, B. et al. Single-cell analysis of transcriptome and DNA methylome in human oocyte maturation. *PLoS ONE* **15**, e0241698 (2020).
68. Castillo-Fernandez, J. et al. Increased transcriptome variation and localised DNA methylation changes in oocytes from aged mice revealed by parallel single-cell analysis. *Aging Cell* **19**, e13278 (2020).
69. Lee, J. H., Park, S. J. & Nakai, K. Differential landscape of non-CpG methylation in embryonic stem cells and neurons caused by DNMT3s. *Sci. Rep.* **7**, 11295 (2017).
70. Jeltsch, A., Adam, S., Dukatz, M., Emperle, M. & Bashtrykov, P. Deep enzymology studies on DNA methyltransferases reveal novel connections between flanking sequences and enzyme activity. *J. Mol. Biol.* **433**, 167186 (2021).
71. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).
72. Neri, F. et al. Intragenic DNA methylation prevents spurious transcription initiation. *Nature* **543**, 72–77 (2017).
73. Otani, J. et al. Structural basis for recognition of H3K4 methylation status by the DNA methyltransferase 3A ATRX–DNMT3–DNMT3L domain. *EMBO Rep.* **10**, 1235–1241 (2009).
74. Mo, A. et al. Epigenomic signatures of neuronal diversity in the mammalian brain. *Neuron* **86**, 1369–1384 (2015).
75. Boxer, L. D. et al. MeCP2 represses the rate of transcriptional initiation of highly methylated long genes. *Mol. Cell* **77**, 294–309 (2020).

76. Hamagami, N. et al. NSD1 deposits histone H3 lysine 36 dimethylation to pattern non-CG DNA methylation in neurons. *Mol. Cell* **83**, 1412–1428 (2023).
77. Angeloni, A. et al. Extensive DNA methylome rearrangement during early lamprey embryogenesis. *Nat. Commun.* **15**, 1977 (2024).
78. Lagger, S. et al. MeCP2 recognizes cytosine methylated tri-nucleotide and di-nucleotide sequences to tune transcription in the mammalian brain. *PLoS Genet.* **13**, e1006793 (2017).
79. Tillotson, R. & Bird, A. The molecular basis of MeCP2 function in the brain. *J. Mol. Biol.* **432**, 1602–1623 (2019).
80. Skene, P. J. et al. Neuronal MeCP2 is expressed at near histone-octamer levels and globally alters the chromatin state. *Mol. Cell* **37**, 457–468 (2010).
81. Rube, H. T. et al. Sequence features accurately predict genome-wide MeCP2 binding in vivo. *Nat. Commun.* **7**, 11025 (2016).
82. Amir, R. E. et al. Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. *Nat. Genet.* **23**, 185–188 (1999).
83. Lyst, M. J. et al. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat. Neurosci.* **16**, 898–902 (2013).
84. Tillotson, R. et al. Neuronal non-CG methylation is an essential target for MeCP2 function. *Mol. Cell* **81**, 1260–1275 (2021).
85. Lavery, L. A. et al. Losing Dnmt3a dependent methylation in inhibitory neurons impairs neural function by a mechanism impacting Rett syndrome. *eLife* **9**, e52981 (2020).
86. Li, J. et al. Dnmt3a knockout in excitatory neurons impairs postnatal synapse maturation and increases the repressive histone modification H3K27me3. *eLife* **11**, e66909 (2022).
87. Fuks, F., Burgers, W. A., Godin, N., Kasai, M. & Kouzarides, T. Dnmt3a binds deacetylases and is recruited by a sequence-specific repressor to silence transcription. *EMBO J.* **20**, 2536–2544 (2001).
88. Liu, Y. et al. Exploring the complexity of MECP2 function in Rett syndrome. *Nat. Rev. Neurosci.* **26**, 379–398 (2025).
89. Renthal, W. et al. Characterization of human mosaic Rett syndrome brain tissue by single-nucleus RNA sequencing. *Nat. Neurosci.* **21**, 1670–1679 (2018).
90. Santistevan, N. J., Ford, C. T., Gilsdorf, C. S. & Grinblat, Y. Behavioral and transcriptomic analyses of mecp2 function in zebrafish. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **195**, e32981 (2024).
91. Moore, J. R. et al. MeCP2 and non-CG DNA methylation stabilize the expression of long genes that distinguish closely related neuron types. *Nat. Neurosci.* **28**, 1185–1198 (2025).
92. Tian, W. et al. Single-cell DNA methylation and 3D genome architecture in the human brain. *Science* **382**, eadf5357 (2023).
93. Liu, H. et al. Single-cell DNA methylome and 3D multi-omic atlas of the adult mouse brain. *Nature* **624**, 366–377 (2023).
94. Zhou, J. et al. Human body single-cell atlas of 3D genome organization and DNA methylation. Preprint at *bioRxiv* https://doi.org/10.1101/2025.03.23.644697 (2025).
95. Goll, M. G. & Halpern, M. E. DNA methylation in zebrafish. *Prog. Mol. Biol. Transl. Sci.* **101**, 193–218 (2011).
96. Yin, L. M., Schnoor, M. & Jun, C. D. Structural characteristics, binding partners and related diseases of the calponin homology (CH) domain. *Front. Cell Dev. Biol.* **8**, 342 (2020).
97. Wu, S. F., Zhang, H. & Cairns, B. R. Genes for embryo development are packaged in blocks of multivalent chromatin in zebrafish sperm. *Genome Res.* **21**, 578–589 (2011).
98. Hong, Y. et al. Establishment of a normal medakafish spermatogonial cell line capable of sperm production in vitro. *Proc. Natl Acad. Sci. USA* **101**, 8011–8016 (2004).
99. Harris, K. D., Lloyd, J. P. B., Domb, K., Zilberman, D. & Zemach, A. DNA methylation is maintained with high fidelity in the honey bee germline and exhibits global non-functional fluctuations during somatic development. *Epigenetics Chromatin* **12**, 62 (2019).
100. Cingolani, P. et al. Intronic non-CG DNA hydroxymethylation and alternative mRNA splicing in honey bees. *BMC Genomics* **14**, 666 (2013).
101. Royle, J. W., Hurwood, D., Sadowski, P. & Dudley, K. J. Non-CG DNA methylation marks the transition from pupa to adult in *Helicoverpa armigera*. *Insect Mol. Biol.* **33**, 493–502 (2024).
102. Gu, Z. et al. Whole-genome bisulfite sequencing reveals the function of DNA methylation in the allotransplantation immunity of pearl oysters. *Front. Immunol.* **14**, 1247544 (2023).
103. Yang, Y., Zheng, Y., Sun, L. & Chen, M. Genome-wide DNA methylation signatures of sea cucumber *Apostichopus japonicus* during environmental induced aestivation. *Genes (Basel)* **11**, 1020 (2020).
104. Song, X. et al. Genome-wide DNA methylomes from discrete developmental stages reveal the predominance of non-CpG methylation in *Tribolium castaneum*. *DNA Res.* **24**, 445–457 (2017).
105. Schulz, N. K. E. et al. Dnmt1 has an essential function despite the absence of CpG DNA methylation in the red flour beetle *Tribolium castaneum*. *Sci. Rep.* **8**, 16462 (2018).
106. De Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A. & Pollock, D. D. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* **7**, e1002384 (2011).
107. Howe, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**, 498–503 (2013).
108. Schartl, M. et al. The genomes of all lungfish inform on genome expansion and tetrapod evolution. *Nature* **634**, 96–103 (2024).
109. 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).
110. Gouil, Q. & Baulcombe, D. C. DNA methylation signatures of the plant chromomethyltransferases. *PLoS Genet.* **12**, e1006526 (2016).
111. Stroud, H. et al. Non-CG methylation patterns shape the epigenetic landscape in *Arabidopsis*. *Nat. Struct. Mol. Biol.* **21**, 64–72 (2014).
112. Kazazian, H. H. Jr Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
113. Tooley, K. B. et al. Differential usage of DNA modifications in neurons, astrocytes, and microglia. *Epigenetics Chromatin* **16**, 45 (2023).
114. Derks, M. F. et al. Gene and transposable element methylation in great tit (*Parus major*) brain and blood. *BMC Genomics* **17**, 332 (2016).
115. Muotri, A. R. et al. L1 retrotransposition in neurons is modulated by MeCP2. *Nature* **468**, 443–446 (2010).
116. Novo, C. L. et al. Satellite repeat transcripts modulate heterochromatin condensates and safeguard chromosome stability in mouse embryonic stem cells. *Nat. Commun.* **13**, 3525 (2022).
117. Guo, W., Zhang, M. Q. & Wu, H. Mammalian non-CG methylations are conserved and cell-type specific and may have been involved in the evolution of transposon elements. *Sci. Rep.* **6**, 32207 (2016).
118. Olova, N. et al. Comparison of whole-genome bisulfite sequencing library preparation strategies identifies sources of biases affecting DNA methylation data. *Genome Biol.* **19**, 33 (2018).
119. Vaisvila, R. et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res.* **31**, 1280–1289 (2021).
120. Liu, Y. et al. Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution. *Nat. Biotechnol.* **37**, 424–429 (2019).

121. Wang, M. et al. Engineered APOBEC3C sequencing enables bisulfite-free and direct detection of DNA methylation at a single-base resolution. *Anal. Chem.* **95**, 1556–1565 (2023).
122. Wang, T. et al. Bisulfite-free sequencing of 5-hydroxymethylcytosine with APOBEC-coupled epigenetic sequencing (ACE-seq). *Methods Mol. Biol.* **2198**, 349–367 (2021).
123. Han, Y. et al. Comparison of EM-seq and PBAT methylome library methods for low-input DNA. *Epigenetics* **17**, 1195–1204 (2022).
124. Liu, Y. et al. DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biol.* **22**, 295 (2021).
125. Angeloni, A., Ferguson, J. & Bogdanovic, O. Nanopore sequencing and data analysis for base-resolution genome-wide 5-methylcytosine profiling. *Methods Mol. Biol.* **2458**, 75–94 (2022).
126. Goldsmith, C. et al. Low biological fluctuation of mitochondrial CpG and non-CpG methylation at the single-molecule level. *Sci. Rep.* **11**, 8032 (2021).
127. Liau, Y. et al. Low-pass nanopore sequencing for measurement of global methylation levels in plants. *BMC Genomics* **25**, 1235 (2024).
128. Kong, Y. et al. Critical assessment of nanopore sequencing for the detection of multiple forms of DNA modifications. Preprint at *bioRxiv* https://doi.org/10.1101/2024.11.19.624260 (2024).
129. Ni, P. et al. Genome-wide detection of cytosine methylations in plant from Nanopore data using deep learning. *Nat. Commun.* **12**, 5976 (2021).
130. Chen, H. X. et al. Accurate cross-species 5mC detection for Oxford Nanopore sequencing in plants with DeepPlant. *Nat. Commun.* **16**, 3227 (2025).
131. Holmes, E. E. et al. Performance evaluation of kits for bisulfite-conversion of DNA from tissues, cell lines, FFPE tissues, aspirates, lavages, effusions, plasma, serum, and urine. *PLoS ONE* **9**, e93933 (2014).
132. Hong, E. E., Okitsu, C. Y., Smith, A. D. & Hsieh, C. L. Regionally specific and genome-wide analyses conclusively demonstrate the absence of CpG methylation in human mitochondrial DNA. *Mol. Cell. Biol.* **33**, 2683–2690 (2013).
133. Kint, S., De Spiegelaere, W., De Kesel, J., Vandekerckhove, L. & Van Criekinge, W. Evaluation of bisulfite kits for DNA methylation profiling in terms of DNA fragmentation and DNA recovery using digital PCR. *PLoS ONE* **13**, e0199091 (2018).
134. Dou, X. et al. The strand-biased mitochondrial DNA methylome and its regulation by DNMT3A. *Genome Res.* **29**, 1622–1634 (2019).
135. Guitton, R., Nido, G. S. & Tzoulis, C. No evidence of extensive non-CpG methylation in mtDNA. *Nucleic Acids Res.* **50**, 9190–9194 (2022).
136. Urich, M. A., Nery, J. R., Lister, R., Schmitz, R. J. & Ecker, J. R. MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. *Nat. Protoc.* **10**, 475–483 (2015).
137. Gong, W. et al. Benchmarking DNA methylation analysis of 14 alignment algorithms for whole genome bisulfite sequencing in mammals. *Comput Struct. Biotechnol. J.* **20**, 4704–4716 (2022).
138. Treangen, T. J. & Salzberg, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat. Rev. Genet.* **13**, 36–46 (2011).
139. Teissandier, A., Servant, N., Barillot, E. & Bourc'his, D. Tools and best practices for retrotransposon analysis using high-throughput sequencing data. *Mob. DNA* **10**, 52 (2019).
140. Mizuguchi, T. et al. Detecting a long insertion variant in SAMD12 by SMRT sequencing: implications of long-read whole-genome sequencing for repeat expansion diseases. *J. Hum. Genet.* **64**, 191–197 (2019).
141. Stevanovski, I. et al. Comprehensive genetic diagnosis of tandem repeat expansion disorders with programmable targeted nanopore sequencing. *Sci. Adv.* **8**, eabm5386 (2022).
142. Karst, S. M. et al. High-accuracy long-read amplicon sequences using unique molecular identifiers with Nanopore or PacBio sequencing. *Nat. Methods* **18**, 165–169 (2021).
143. Delahaye, C. & Nicolas, J. Sequencing DNA with nanopores: troubles and biases. *PLoS ONE* **16**, e0257521 (2021).
144. Catoni, M., Tsang, J. M., Greco, A. P. & Zabet, N. R. DMRcaller: a versatile R/Bioconductor package for detection and visualization of differentially methylated regions in CpG and non-CpG contexts. *Nucleic Acids Res.* **46**, e114 (2018).
145. Ma, H. et al. Abnormalities in human pluripotent cells due to reprogramming mechanisms. *Nature* **511**, 177–183 (2014).
146. Tatton-Brown, K. et al. Mutations in the DNA methyltransferase gene DNMT3A cause an overgrowth syndrome with intellectual disability. *Nat. Genet.* **46**, 385–388 (2014).
147. Cree, S. L. et al. DNA G-quadruplexes show strong interaction with DNA methyltransferases in vitro. *FEBS Lett.* **590**, 2870–2883 (2016).
148. Jin, J. et al. The effects of cytosine methylation on general transcription factors. *Sci. Rep.* **6**, 29119 (2016).
149. Abhishek, S., Nakarakanti, N. K., Deeksha, W. & Rajakumara, E. Mechanistic insights into recognition of symmetric methylated cytosines in CpG and non-CpG DNA by UHRF1 SRA. *Int. J. Biol. Macromol.* **170**, 514–522 (2021).
150. Spruijt, C. G. et al. Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
151. Roth, G. V., Gengaro, I. R. & Qi, L. S. Precision epigenetic editing: technological advances, enduring challenges, and therapeutic applications. *Cell Chem. Biol.* https://doi.org/10.1016/j.chembiol.2024.07.007 (2024).
152. Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* **328**, 916–919 (2010).
153. Domb, K. et al. DNA methylation mutants in *Physcomitrella patens* elucidate individual roles of CG and non-CG methylation in genome regulation. *Proc. Natl Acad. Sci. USA* **117**, 33700–33710 (2020).
154. Yaari, R. et al. RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. *Nat. Commun.* **10**, 1613 (2019).
155. Ikeda, Y. et al. Loss of CG methylation in *Marchantia polymorpha* causes disorganization of cell division and reveals unique DNA methylation regulatory mechanisms of non-CG methylation. *Plant Cell Physiol.* **59**, 2421–2431 (2018).
156. Zemach, A. et al. The *Arabidopsis* nucleosome remodeler DDM1 allows DNA methyltransferases to access H1-containing heterochromatin. *Cell* **153**, 193–205 (2013).
157. Law, J. A. & Jacobsen, S. E. Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* **11**, 204–220 (2010).
158. Bewick, A. J. et al. Diversity of cytosine methylation across the fungal tree of life. *Nat. Ecol. Evol.* **3**, 479–490 (2019).
159. Shi, J. et al. DNA methylation plays important roles in lifestyle transition of *Arthrobotrys oligospora*. *IET Syst. Biol.* **18**, 92–102 (2024).
160. Nai, Y. S., Huang, Y. C., Yen, M. R. & Chen, P. Y. Diversity of fungal DNA methyltransferases and their association with DNA methylation patterns. *Front. Microbiol.* **11**, 616922 (2020).
161. Chen, Y. Y. et al. DNA methylation-dependent epigenetic regulation of *Verticillium dahliae* virulence in plants. *aBIOTECH* **4**, 185–201 (2023).
162. So, K. K. et al. Global DNA methylation in the chestnut blight fungus *Cryphonectria parasitica* and genome-wide changes in DNA methylation accompanied with sectorization. *Front. Plant Sci.* **9**, 103 (2018).

163. Malagnac, F. et al. A gene essential for de novo methylation and development in *Ascobolus* reveals a novel type of eukaryotic DNA methyltransferase structure. *Cell* **91**, 281–290 (1997).
164. Sarre, L. A., Gastellou Peralta, G. A., Romero Charria, P., Ovchinnikov, V. & de Mendoza, A. Repressive cytosine methylation is a marker of viral gene transfer across divergent eukaryotes. *Mol. Biol. Evol.* **42**, msaf176 (2025).
165. De Mendoza, A. et al. Recurrent acquisition of cytosine methyltransferases into eukaryotic retrotransposons. *Nat. Commun.* **9**, 1341 (2018).
166. Sarre, L. A. et al. DNA methylation enables recurrent endogenization of giant viruses in an animal relative. *Sci. Adv.* **10**, eado6406 (2024).
167. Huff, J. T. & Zilberman, D. Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* **156**, 1286–1297 (2014).
168. Clark, S. J., Harrison, J., Paul, C. L. & Frommer, M. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22**, 2990–2997 (1994).
169. Tse, O. Y. O. et al. Genome-wide detection of cytosine methylation by single molecule real-time sequencing. *Proc. Natl Acad. Sci. USA* **118**, e2019768118 (2021).
170. Wang, Y., Zhao, Y., Bollas, A., Wang, Y. & Au, K. F. Nanopore sequencing technology, bioinformatics and applications. *Nat. Biotechnol.* **39**, 1348–1365 (2021).
171. Kulkarni, O. et al. Comprehensive benchmarking of tools for nanopore-based detection of DNA methylation. Preprint at *bioRxiv* https://doi.org/10.1101/2024.11.09.622763 (2024).

Acknowledgements

This work was supported by the Spanish Ministry of Science (projects CNS2023-144039 and PID2021-128358NA-I00), as well as by funding from the Unit of Excellence María de Maeztu (CEX2020-001088-M to O.B.). A.d.M. was supported by the European Research Council Starting Grant 950230.

Author contributions

O.B. conceptualized the work with the help of A.d.M. T.B. and O.B. wrote the original draft of the manuscript. O.B. T.B. and A.d.M. created Figs. 1–4. All authors critically reviewed and edited the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-025-02303-1.

Correspondence and requests for materials should be addressed to Alex de Mendoza or Ozren Bogdanovic.

Peer review information *Nature Genetics* thanks Maxim Greenberg and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.