

Series: Revisiting discoveries that transformed genetics

Feature Review

DNA methylation: a historical perspective

Alexandra L. Mattei,^{1,2,3,8} Nina Bailly,^{1,4,8} and Alexander Meissner  ^{1,3,5,6,7,*}

In 1925, 5-methylcytosine was first reported in bacteria. However, its biological importance was not intuitive for several decades. After this initial lag, the ubiquitous presence of this methylated base emerged across all domains of life and revealed a range of essential biological functions. Today, we are armed with the knowledge of the key factors that establish, maintain, and remove DNA methylation and have access to a staggering and rapidly growing number of base-resolution methylation maps. Despite this, several fundamental details about the precise role and interpretation of DNA methylation patterns remain under investigation. Here, we review the field of DNA methylation from its beginning to present day, with an emphasis on findings in mammalian systems, and point the reader to select experiments that form the foundation of this field.

DNA methylation: why (still) bother?

A quarter century ago, one of the pioneers in the field of DNA methylation, Rudolf Jaenisch, outlined in the August 1997 issue of *Trends in Genetics* why we should bother caring about DNA methylation and speculated in which developmental contexts it might function [1]. Here, we would like to review why we still bother, what we have learned in nearly a century of research, and what we still need to address in the coming years. Since its initial discovery in bacteria in 1925, DNA methylation has been investigated in a vast range of organisms and is linked to biological topics from gene regulation and genome organization, to reproduction and development, and to disease and aging. It is the most well-studied epigenetic mechanism and is often used as the classical example of epigenetic inheritance, although recent advances have shown this modification to be more dynamic, and hence more complex, than previously thought [2–4].

Despite an ever-growing body of work published on DNA methylation each year, it remains difficult to pinpoint the precise function of most DNA methylation found across the genome. It is also still unresolved why DNA methylation is essential to differentiated, but not pluripotent, cells [5–7] and why it is altered into a distinct landscape across most cancer types [8]. As typical review articles by design focus on summarizing more recent discoveries and advances around their time [9–18], we decided to complement this by providing a systematic review covering the entire history of the field to highlight many foundational discoveries on which our current work is built. As expected, the primary literature is vast, and we apologize for having to omit many elegant experiments as we summarize the emergence and progression of the field of DNA methylation across a century.

1900–1959: From genetics to epigenetics

At the turn of the 20th century, Walter Sutton (1902) and Theodore Boveri (1903) independently proposed the chromosomal theory of inheritance, linking Gregor Mendel's (1866) long overlooked laws on gene behavior and inheritance to their own work on meiosis [19,20]. This initially controversial theory gained credence following a 1910 paper from one of its detractors, Thomas Hunt Morgan, who demonstrated that eye color in *Drosophila melanogaster* is

Highlights

5mC was discovered in mammals and found to have a nonrandom distribution that suggested a possible biological function.

In the early 1980s, DNA methylation within 5' promoter regions, but not elsewhere, was found to inhibit transcription of the associated gene.

Throughout the 1990s and 2000s, mechanisms of gene regulation by DNA methylation were elucidated as well as its relationship with histone modifications and influence on the 3D genome organization uncovered.

Over the past decade, high-throughput sequencing technologies complemented earlier single-gene efforts and ultimately provided a global understanding of DNA methylation and its dynamics in development and disease.

¹Department of Genome Regulation, Max Planck Institute for Molecular Genetics, Berlin, Germany

²Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA

³Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA

⁴Institute of Biology, Freie Universität Berlin, Berlin, Germany

⁵Institute of Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany

⁶Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁷<https://www.molgen.mpg.de/Genome-Regulation/Meissner-lab>

⁸These authors contributed equally to this work

*Correspondence:
meissner@molgen.mpg.de
(A. Meissner).



determined by inheritance of a gene on the X chromosome, which provided the first decisive piece of evidence in support of this theory [21]. Levene and Jacobs' research on nucleic acids revealed that they reside in a polymer chain of nucleotides [22] and the growing interest in the composition of these nucleic acids laid the foundation among others for the field of epigenetics, with DNA methylation as a central actor (Figure 1A).

The discovery of 5-methylcytosine in living cells

In 1925, Johnson and Coghill isolated and crystallized nucleic acids from *Mycobacterium tuberculosis* in an effort to identify its pathogenic determinant. One of their candidates was **5-methylcytosine (5mC)** (see [Glossary](#)), a nucleotide Johnson had postulated might occur naturally in living organisms based on his previous success with its *in vitro* biochemical synthesis [23]. Microscopic examination of their hydrolyzed nucleic acid picrate crystals under polarized light indeed distinguished cytosine from 5mC [24].

Despite this early and seemingly relevant discovery, the next report on 5mC was only published 23 years later. Using recent advances in paper chromatography [25], Hotchkiss observed a faint band near that of cytosine on his chromatograph of calf thymus DNA that behaved like cytosine, yet was slightly shifted in its migration, leading him to suggest it is cytosine but with some modification and therefore labeled it 'epi-cytosine' [26] (Figure 1B). Specifically, he noted that the epi-cytosine relates to cytosine in terms of its absorption spectrum and mobility in the same manner that thymine relates to uracil. As thymine is 5-methyluracil, Hotchkiss inferred that epi-cytosine could possibly be 5mC. Two years later, Wyatt confirmed the presence of 5mC in mammalian, insect, and plant DNA with a broad range of quantities [27,28].

As nucleic acids were confirmed to be the carriers of genetic information [29,30], and the structure of the DNA double helix was reported [31], interest in the field of DNA methylation grew. Sinsheimer subsequently noted that 5mC is not randomly distributed in DNA but is found specifically in the **CpG dinucleotide** context (Figure 1C). Interestingly, the CpG doublet was not found as frequently as expected in eukaryotic DNA [32,33].

Summary

Why did it take so long from its initial biological discovery before research on 5mC started to progress more rapidly? One obvious reason is the historical context of its discovery. In 1925, we did not know yet that polymer chains of nucleic acids carry genetic information. The 1928 transformation experiments by Frederick Griffith [34], the 1944 Avery-MacLeod-McCarty experiment [29], the conclusion of the Second World War, the 1952 Hershey-Chase experiment [30], and the resolution of the DNA double helix [31] helped lay the needed foundation that enabled the exploration of the possible relevance of 5mC in DNA. Two additional factors may have caused some initial hesitation: other groups did not find 5mC in their DNA isolates of *M. tuberculosis* [27,35], and the low abundance of 5mC reported by Hotchkiss and Wyatt seemed disqualifying for 5mC to have a major biological function. As an aside, it is worth mentioning that in parallel to these experimental advances, the developmental biologist Conrad Waddington coined the term 'epigenetics' in 1942 [36] and published his widely used epigenetic landscape in 1957 [37]; however, these concepts were not linked to DNA methylation until its function became clearer over subsequent decades.

1960–1969: Insights into the utility and mechanisms of 5mC in bacteria

The dawn of molecular biology set the stage for a more thorough investigation and appreciation of DNA methylation from plants to mammals. However, essential progress was first made by studying the methylation of nucleic acids in bacteria [38]. As a tractable and abundant model organism,

Glossary

Bisulfite sequencing: a method used to detect 5-methylcytosine at single base resolution. Sodium bisulfite is used to selectively deaminate cytosines into uracils, while methylcytosines remain unchanged. PCR amplification of bisulfite-converted DNA with regular dNTPs converts the U into T. When sequenced and aligned to the reference genome, mapped Cs are inferred to be 5mC, while C to T transitions indicate unmethylated cytosines.

BLAST: basic local alignment search tool, a widely used algorithm developed by Altschul and colleagues in 1990 to align and compare nucleotide or protein sequences.

CCCTC-binding factor (CTCF): a highly conserved, ubiquitously expressed zinc-finger protein and transcription factor known for its function as an insulator and anchor for enhancer-promoter communication that establishes cohesion loops via dimerization.

CpG dinucleotide: cytosine linked to guanine via a phosphate group on the same strand.

Embryonic carcinoma (EC) cells: the first pluripotent cells to be cultured *in vitro* originally isolated in 1964 from a mouse testicular teratocarcinoma, a germ cell tumor comprised of undifferentiated cells as well as cells from all three germ layers.

Heterochromatinization: the process of compacting chromatin, generally associated with transcriptional silencing.

Histone 3 lysine 4 (H3K4)

methylation: methylation of lysine 4 on histone 3 tails marking active or primed regulatory elements, such as promoters and some enhancers, as well as transcribed gene bodies. H3K4 can be mono-, di-, or tri-methylated.

Histone 3 lysine 9 (H3K9)

methylation: methylation of lysine 9 on histone 3 tails marking transcriptionally inactive chromatin, such as silenced repetitive elements. H3K9 can be mono-, di-, or tri-methylated.

5-Hydroxymethylcytosine (5hmC):

the oxidation product of 5-methylcytosine that is the product of TET activity, the enzymes responsible for active DNA demethylation. Subsequent rounds of oxidation can produce 5-formyl- and 5-carboxylcytosine.

Imprinting control region (ICR): a genomic locus involved in controlling the expression of imprinted genes via DNA

bacteria provided major insights into the biology of 5mC in prokaryotes and thereby paved the way for its study in higher organisms (Figure 2A).

The bacterial restriction and modification system

Luria, Bertani, and Weigle first demonstrated that different families of bacteriophage diverge in their ability to infect certain bacterial strains [39,40]. The basis for their strain specificity of viral infection was not due to a phage's differential ability to enter the bacterial strains, but rather because once inside, incompatible phage DNA was found to be degraded in an immune-like response [41]. A key mechanistic advance was the discovery that different bacteria have strain-specific **methyltransferase** activity, which raised the possibility of a role for 5mC in the defense against phages [42]. Thus, Arber proposed the **restriction and modification system (R-M system)** where methylation-sensitive 'restriction enzymes' defend the bacterial host against invading viruses by digesting their DNA. Bacterial DNA is protected from these restriction enzymes due to modifications to their DNA in the form of species-specific DNA methylation [43].

DNA methylation during DNA replication in bacteria

Beyond its role in host protection, a link between bacterial DNA methylation and DNA replication was observed [44]. Billen found that during normal *Escherichia coli* growth, DNA methyltransferase activity was evident behind the replication fork where 5mC was exclusively placed on the unmethylated nascent strand of DNA (Figure 2B). DNA replication in the absence of methionine, the methyl donor, led to the synthesis of an unmethylated nascent strand, which retained the ability to get methylated after S phase when methionine was added back into the media [45]. However, it seemed that the unmethylated nascent strand of DNA cannot serve as template DNA in the subsequent round of replication [46] and strains deficient for the methyl-donor showed DNA degradation [47].

5mC and methyltransferase activity in rat tissues

In their 1964 review on nucleic acid modifications, Srinivasan and Borek speculated that because 5mC plays a defining role in bacteria, similar mechanisms might act in eukaryotes that could underlie their cell type diversity [48]. Four years later, they reported DNA methyltransferase activity in the nuclear extracts of different tissues of embryonic as well as adult rat and tested their ability to methylate DNA from various species. Interestingly, these experiments showed that some extracts, such as from kidney or liver, harbor more potent methyltransferase activity than brain or spleen extracts. Based on these observations, they suggested that different tissues from the same organism might have different 5mC content [49].

Summary

The first biological roles for DNA methylation were gleaned from studies on the basics of bacterial immunity and DNA replication, though it remained unclear whether any of these functions would be conserved in higher organisms. A key advance was based on the discovery that enzymes are responsible for adding the methyl group to cytosines in nucleic acid polymers. This suggested that DNA methylation could be regulated, thus providing a path for specific target modification. In particular, the possible tissue-specific roles of 5mC in rodents were intriguing, but the data were too sparse to draw more meaningful conclusions yet.

1970–1979: DNA methylation in higher organisms

Once it became clear that 5mC, despite its relatively low abundance, does have a biological function in bacteria, the possibility that DNA methylation could also play a more general regulatory role across species gained credibility. As in many fields, important technological advances were needed to enable a thorough and informative investigation of the theoretical models that emerged in this decade.

methylation, where each allele of the locus displays parent-of-origin specific DNA methylation.

Isoschizomer: meaning 'same cut', a restriction enzyme with the same recognition sequence as another restriction enzyme.

5-Methylcytosine (5mC): a pyrimidine resulting from the addition of a methyl group to the fifth carbon of cytosine.

Methyltransferase: an enzyme catalyzing the transfer of a methyl group from a donor, most commonly S-adenosylmethionine (SAM), to a recipient molecule, in the context of this review to DNA or histone proteins.

Moloney murine leukemia virus (M-MuLV): a type C retrovirus first isolated from a sarcoma and propagated in BALB/c mice. The virus causes leukemia within a few months after newborns are infected, though integrated copies are usually methylated and silenced following germline passage.

Restriction and modification system (R-M system): an immune-like system used by bacteria for the defense against foreign invading DNA. Each bacterial strain expresses a specific set of restriction endonuclease (e.g., HpaII) and its cognate methyltransferases (e.g., M.HpaII) that recognize the same DNA sequence. The endonuclease digests foreign, unmethylated DNA while the host's DNA is methylated and thus protected from its own endonucleases.

Satellite repeats: large noncoding tandem DNA repeats that can extend over 2000 kb in length, found primarily at centromeres and the telomeres, but also heterochromatic regions of the genome.

X inactivation: a method of dosage compensation in therian female cells where one of the two X chromosomes is silenced during development.

Broader quantification of DNA methylation

The presence of 5mC in bacteria, plants, and mammals indicated that 5mC is a widespread DNA modification, which led to further exploration of methylation content using mass spectrometry. In the early 1970s, Vanyushin quantified 5mC levels present in different cell types of many animals, including sponges, mollusks, sea urchins, bony fish, amphibians, reptiles, and mammals [50,51]. These analyses showed that while both GC and 5mC content can differ between species, they are often more similar between closely related species and generally comparable between different tissues. Interestingly, Vanyushin later found 5mC in sequence contexts other than the CpG dinucleotide in plants and in varying quantities across different plant species [52].

A role for DNA methylation in gene regulation

As more reports of methylation profiling by mass spectrometry accumulated, several groups speculated about the possible role of 5mC in higher organisms, including that: (i) 5mC may play no role in eukaryotic development [53,54]; (ii) 5mC may guide DNA mutations, which at the time were thought to be required for transcriptional changes [55]; and (iii) 5mC may act as a transcriptional activator [56] (Box 1). In 1975, three notable reviews were published that each

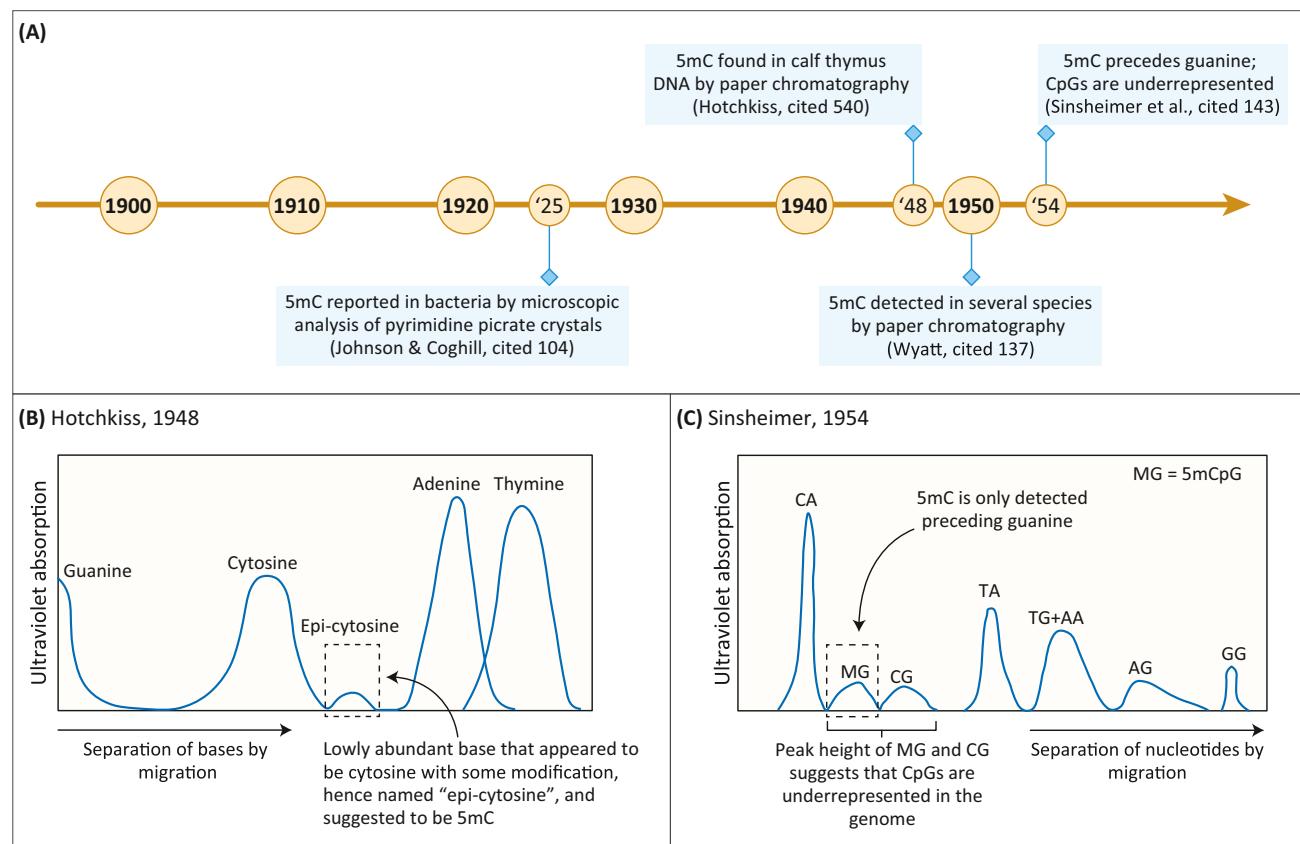


Figure 1. The discovery of 5-methylcytosine. (A) Timeline of early studies that established the field of DNA methylation between 1900 and 1959, including the *in vivo* discovery of 5mC in bacteria and then mammals. Web of Science citation counts as of March 9, 2022 are shown. (B) A schematic illustration of a paper chromatograph shown by Hotchkiss in his 1948 paper that reports the UV absorption of the products of hydrolyzed calf thymus genomic DNA. A base of low abundance believed to be 5-methylcytosine due to its characteristics was labeled by Hotchkiss as 'epi-cytosine' given the uncertainty of its exact identity. (C) Sinsheimer digested calf thymus DNA into dinucleotides that were fractionated by column chromatography. The plot shows the UV absorptions of these nucleotide doublets. MG = 5-methylcytosine followed by guanine. See [24,26,27,33].

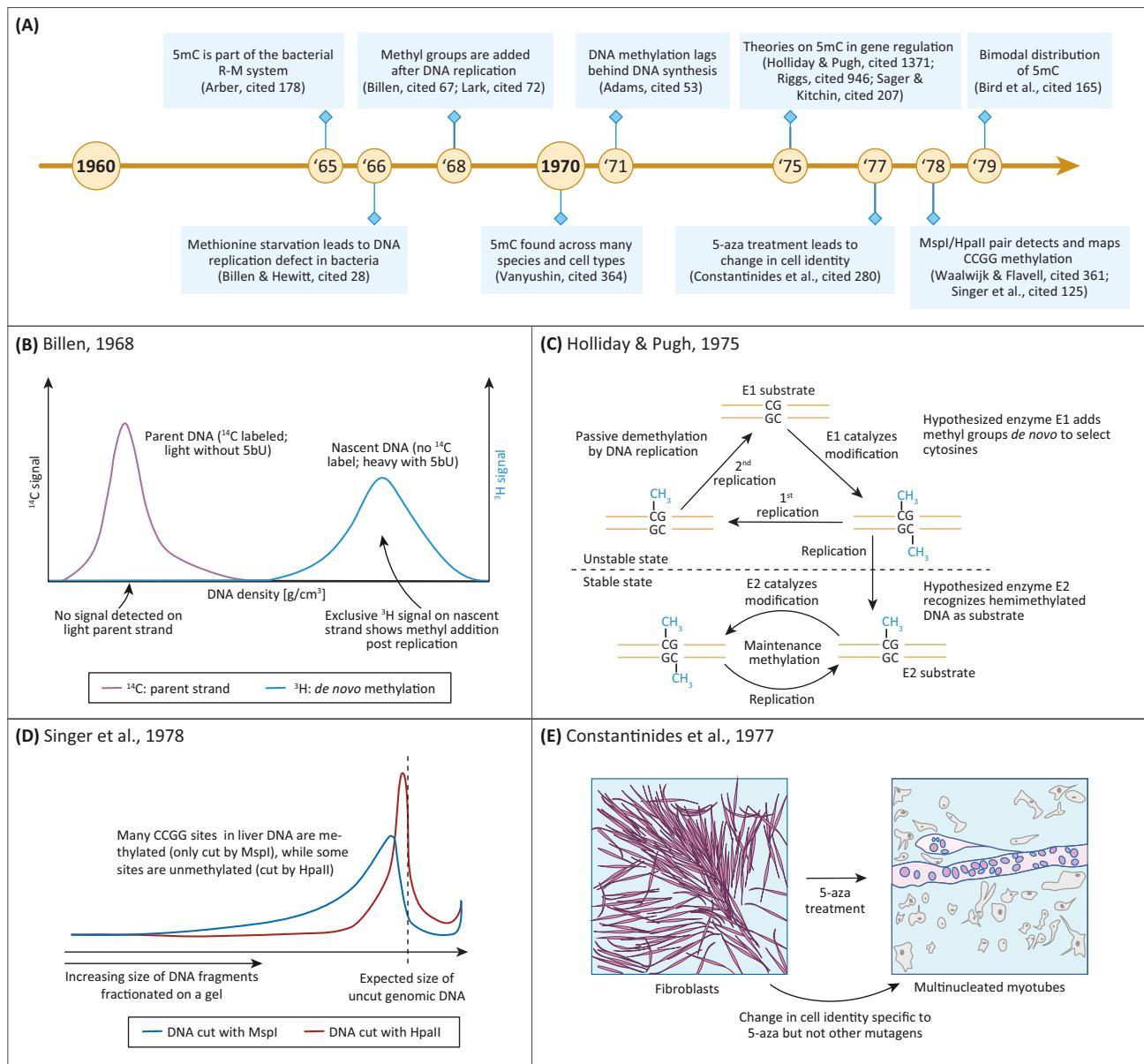


Figure 2. Propagation of stable methylation patterns and first theories on their functions. (A) In the 1960s and 1970s, the first major roles for DNA methylation were discovered and explored. (B) To determine where methyl groups are added after DNA synthesis, Billen cultivated bacteria first in the presence of ^{14}C -labeled thymine for one cycle of replication followed by 5-bromouracil (5bU) present for the second cycle, to label the parent and nascent strand, respectively. In the second cycle, ^3H -labeled methionine was added to track the placement of methyl groups. As shown, the ^3H -labeled methyl groups were found solely on nascent DNA (blue line) and not parent DNA (purple line) when separated by density. (C) Schematic of Holliday and Pugh's model for the generation and stable propagation of DNA methylation. Enzyme E1 methylates unmodified DNA at target sites, but this activity alone is unstable due to the expected loss through DNA replication (top: Unstable state). Thus, a second enzyme E2 is required to enable maintenance of the methylation pattern by specifically recognizing hemimethylated DNA and complementing methylation at the symmetrical nascent CpG (bottom: Stable state). E1 and E2 were not known at the time and were hypothetical enzymes. (D) Densitometer tracing of gel-fractionated mouse DNA, digested with Mspl or its methylation-sensitive isoschizomeric HpaII, by Singer *et al.* The abundance of smaller Mspl fragments compared with HpaII fragments indicates that the genome-wide cut sites are predominantly methylated. (E) Schematics of C3H 10T1/2 C18 fibroblasts by phase-contrast microscopy in Constantinides *et al.* revealed a striking change in morphology after treatment with the mutagen and methylation inhibitor 5-aza for 9–10 days, indicating differentiation into myotubes. See [43–46,50,53,57–59,64,65,79,94]. Abbreviations: 5-aza, 5-azacytidine; 5mC, 5-methylcytosine; R-M system, restriction and modification system.

Box 1. Early theories on the function of DNA methylation

Work prior to the 1970s led several scientists to propose formal hypotheses about the function of DNA methylation in eukaryotes. In the late 1960s, Scarano and colleagues observed that 90% of 5mC in sea urchin DNA is found in the CG context and is thus not randomly distributed in DNA, leading them to speculate about a role for 5mC in differentiation [277,278]. In 1971, Scarano proposed that spontaneous deamination of 5mC, which generates a C→T conversion, could lead to heritable changes in the DNA sequence. A popular theory from the 1960s up until the early 1980s was that sequence mutations in genes direct differentiation. Scarano therefore speculated that 5mC-guided mutation could direct cellular differentiation during embryogenesis [55].

In the same year, Adams' work on 5mC patterning following DNA replication in mouse fibroblasts revealed that early replicating DNA is methylated quickly, while late replicating DNA takes several hours to become fully methylated. The observation that active DNA methylation occurs predominantly in S-phase led Adams to conclude that 5mC must not play a role in controlling transcription [53]. His conclusions might also have been influenced by the bacterial studies by Billen and Lark that implicated 5mC as a regulator of DNA replication [45,46]. Interestingly, in 1972 Comings came to a different conclusion looking at Chinese hamster ovarian cells, where he found that late replicating AT-rich DNA is undermethylated to a greater extent than would be expected from its base composition, while early replicating GC-rich DNA is highly methylated. Comings speculated that if DNA methylation is needed in high amounts in euchromatic DNA where it might play a role in active transcription, then spontaneous deamination of 5mC leading to CG→TA mutations would be actively selected against in euchromatin [56]. To Comings, the idea that 5mC is enriched in active DNA regions in eukaryotes implicated it as a transcriptional activator.

In 1973, Adams demonstrated that sea urchin DNA is twice as methylated at the pluteus stage than at the morula stage [54], in agreement with earlier findings suggesting that methylation in sea urchins does not occur until gastrulation [278,279]. Adams' report was the first to quantify such changes at each developmental stage, which led him to revise his previous theory that DNA methylation does not regulate gene expression and to instead propose the new idea that DNA methylation could function to 'switch off' genes after contributing to their specific function in early development.

provided unique frameworks for contemplating and investigating the biological effects of DNA methylation [57–59]. While each review differed in its specific, well rationalized mechanisms, they all fundamentally agreed that 5mC would play a role in regulating gene expression and orchestrating development.

Holliday and Pugh proposed a model for switching genes on and off that utilizes two different specialized enzymes for *de novo* methylation and maintenance methylation (Figure 2C) [57]. They argued that DNA methylation seemed a better candidate than sequence mutation to reversibly control gene expression, considering experiments by Gurdon and colleagues that showed transplanting a somatic cell nucleus into an enucleated oocyte can reprogram it to totipotency [60].

Riggs detailed the theoretical connection between DNA methylation and the process of **X inactivation** and proposed that the evolutionarily ancient 5mC could serve as its initiator in a two-step process carried out by one enzyme [58]. Interestingly, Riggs argued that the rules governing his X inactivation model could be applied more broadly to the control of gene expression, with a high barrier to transcriptional change in either direction that, once triggered, could respond quickly. Specifically, inspired by the mechanism of the bacterial R-M system, he suggested that 5mC found in gene regulatory sequences could influence the ability of proteins to bind there, but when 5mC occurs outside of regulatory sequences, such as at **satellite repeats**, it may serve other roles [58].

Sager and Kitchin also took inspiration from the work in bacteria to extrapolate a parsimonious role for 5mC as an R-M system in all organisms harboring genomic methylation [59]. Broadly speaking, the proposed system would generate small, inherited modifications to DNA that enable its later 'restriction' by degradation or **heterochromatinization**, therefore leading to inactivation. In eukaryotes, they proposed that the insertions or deletions resulting from double strand breaks

created by restriction endonuclease activity could provide the heritable sequence mutations thought to be needed to stably regulate gene expression changes during differentiation. Changes in DNA methylation throughout development could direct those cut sites and their subsequent mutation.

The restriction enzyme toolbox

Shortly after Arber postulated their existence [43], restriction enzymes were isolated for the first time from *Escherichia coli* and *Haemophilus influenzae* [61–63]. Their potential as invaluable tools was quickly realized and commercialization led to their widespread use in all of molecular biology. Arguably one of the most important restriction nucleases used in the field of DNA methylation was discovered in a strain of *Haemophilus parainfluenzae* grown at Cold Spring Harbor and then sent to New England Biolabs for the production and sale of what was thought to be the previously characterized enzyme Hpall. In 1978, two groups simultaneously published their independent observation that this new Hpall behaved differently than the Hpall they had used before and instead turned out to be Hpall's **isoschizomer** Mspl [64,65]. These papers illustrated how a methylation-insensitive (for the inner C) Mspl digestion can locate all CCGG sites in the genome and the complementary methylation-sensitive Hpall digestion can reveal the CpG methylation status (Figure 2D). Together, this pair of enzymes finally enabled researchers to determine the methylation status of specific CpGs within their local sequence contexts.

DNA methylation and gene expression

Combining Hpall/Mspl digestions with Southern blots, many research groups began to investigate the methylation status of genes with tissue-specific expression patterns, such as *β-globin* and *albumin* [66–68]. Comparing the methylation of individual genes across different cell types of the same organism quickly revealed sites that were always, never, or variably methylated. These variably methylated sites were much less methylated in cell types where their associated gene was active, meaning that the methylation status of this gene was anticorrelated with its expression [69,70]. This observation, which was verified by many others, led to the wider consensus that 5mC acts as a repressive DNA modification.

5-Azacytidine (5-aza) and loss of methylation

Research on a compound called 5-aza intersected with the field of DNA methylation in the late 1970s and eventually provided another important tool for the field [71–74]. 5-aza is a nucleoside analog of cytosine that can be incorporated into DNA and RNA [75] and was widely used in the late 1960s and early 1970s as a mutagen [76] and anticancer drug [77]. While studying the mutagenic effects of 5-aza on cells *in vitro*, Jones and colleagues observed that treatment was accompanied by morphological changes indicating differentiation [78,79]. Constantinides *et al.* showed that 5-aza-treated mouse fibroblasts became tubular and multinucleated, which suggested their differentiation into muscle cells [79] (Figure 2E). Similarly, Taylor and Jones treated mouse fibroblasts with a selection of known mutagens to test whether the induced mutations could cause any differentiation *in vitro* and found that only 5-aza had this specific effect [80]. In an important next step, Friedman found that 5-aza treatment in *E. coli* decreased 5mC but not 6mA levels, demonstrating that in addition to its mutagenic properties, 5-aza may act as a specific inhibitor of 5mC methyltransferases [81]. A year later, Jones and Taylor extended these findings to eukaryotes and reported that 5-aza reduces 5mC, and speculated it may impede the advance of the presumably progressive methyltransferase activity along the DNA [82].

Summary

The 1970s saw notable advances in both research and hypothetical models about the role of DNA methylation in gene regulation. Improved 5mC detection methods enabled locus-specific

methylation analysis, serving as a major accelerator for the field. Careful reading of these early papers reveals additional points to highlight, including the global differences in genome methylation between invertebrates and vertebrates [50], though other results such as hypomethylation of vertebrate sperm did not hold true when repeated by others [83]. At the end of this decade, the field had a good appreciation of DNA methylation within and across species and it became generally accepted that DNA methylation acts as a repressor of gene expression.

1980–1989: Gene and genome regulation by 5mC

Several advances in *in vitro* and *in vivo* cloning [84–88] as well as transgenic techniques from the mid-1970s to mid-1980s [89–93] enabled a suite of new experiments to explore the sequence context and functional role of DNA methylation. For instance, the possibility of inserting both unmethylated and methylated DNA constructs into living cells served as the foundation of many studies in this decade (Figure 3A).

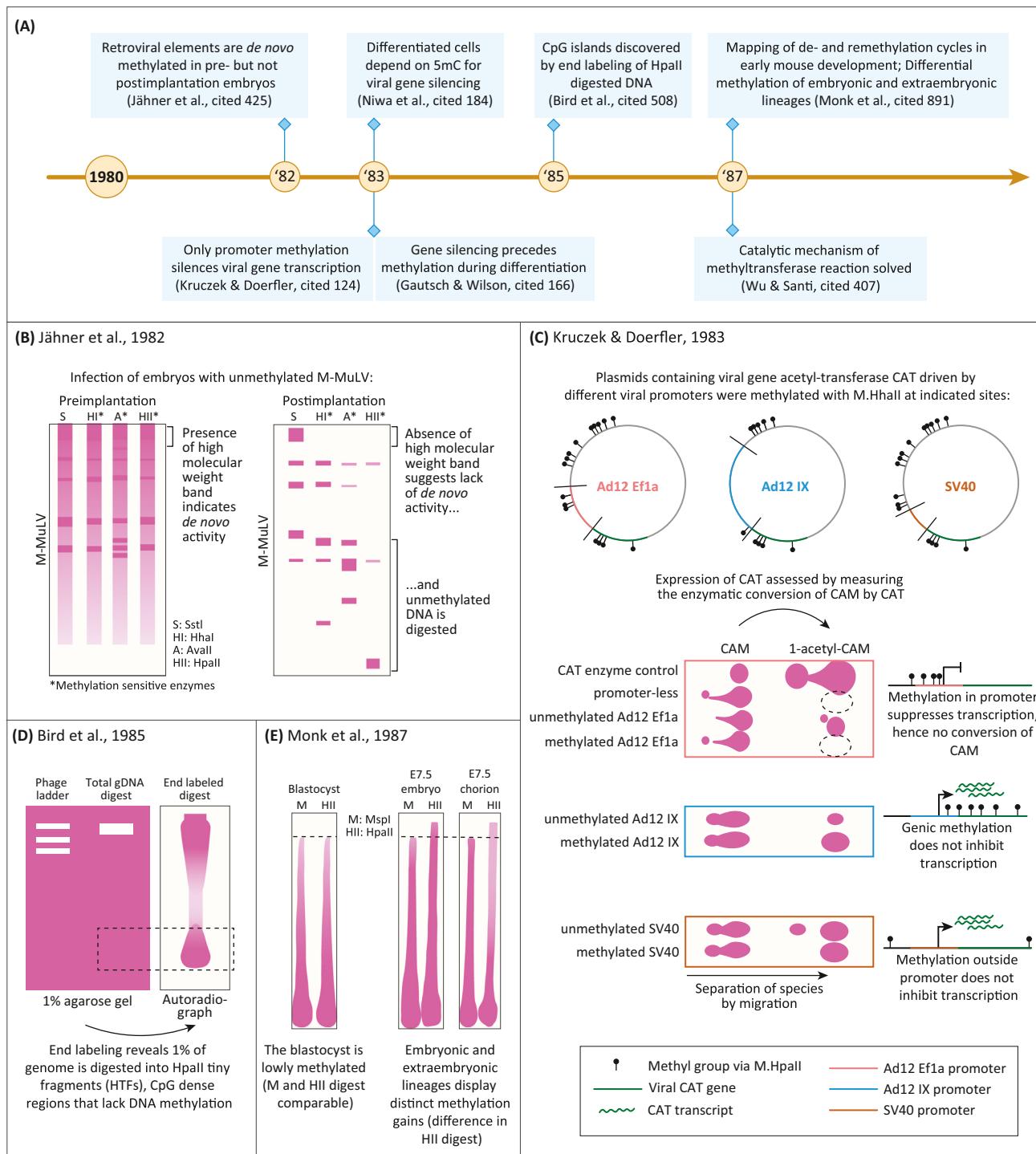
DNA methylation and gene regulation

The 1980s began with the observation by Bird and colleagues that 5mC seemed to be distributed in vertebrate and invertebrate genomes in an ‘all or none’ fashion, with stretches of methylated DNA punctuated by stretches of unmethylated DNA, each occupying different chromatin fractions. They speculated that methylation distinguishes transcriptionally inert DNA from transcribable DNA [94,95]. This idea was in line with the earlier 5-aza experiments that demonstrated demethylation could lead to the activation of genes [82]. However, it was not yet clear to what extent the relative quantity of DNA methylation at a gene or the location of DNA methylation within and around this gene mattered for its expression. Nevertheless, the observation of transcriptional inactivity of some relatively unmethylated genes already suggested that loss or lack of methylation alone does not always lead to gene activation [96–98]. A few examples were also reported where tissue-specific genes were active while methylation was high across the gene [99,100] and, in another case, methylation was noted to be lost upstream of the 5' end of the gene, but not elsewhere, upon its activation [97]. Due to these various observations and the as of yet incomplete information on this topic, questions remained about the general rules for when and where methylation acts as a gene repressor and how its removal may lead to gene activation.

In addition, while DNA replication was one possible cellular mechanism of passive methylation loss, it was expected, but not known, whether an enzyme or other biological process could more specifically demethylate certain genes during *in vivo* development. In 1982, Gjerset and Martin published initial evidence of a demethylating activity found in the nucleoplasm isolated from mouse erythroleukemia cells. The nuclear extract was shown to have proteinase K sensitive CpG demethylation capabilities in the absence of DNA replication [101]. However, there was no further purification or characterization of the putative demethylating enzyme.

Repression of viral DNA elements

Some mouse strains carrying germline copies of **Moloney murine leukemia virus (M-MuLV)** show virus activation at different stages of development and therefore provided an opportunity to study the potential of viral genes to be regulated by DNA methylation [102]. In line with the findings that endogenous genes can become activated by loss of methylation, Jaenisch and colleagues found that cloned, and thus unmethylated, copies of M-MuLV were active and infectious when transfected into fibroblasts, while their endogenous and methylated counterparts were not [103]. Injecting mouse embryos at different developmental stages with viral DNA revealed that in addition to germline silencing, preimplantation embryos quickly *de novo* methylated and inactivated the viral sequences, while post-implantation embryos failed to do so [104]



Trends In Genetics

Figure 3. Promoter methylation and the discovery of CpG islands (CGIs). (A) Advances made in the 1980s made key contributions to our understanding of how 5-methylcytosine (5mC) affects gene expression. (B) Jaenisch and colleagues studied the ability of mouse pre- and postimplantation embryos to methylate newly integrated copies of murine leukemia virus (M-MuLV). Methylation-sensitive restriction digestions show that the M-MuLV sequence becomes methylated when introduced into

(Figure legend continued at the bottom of the next page.)

(Figure 3B). Therefore, Jaenisch and colleagues reasoned that the early mouse embryo might be uniquely capable of retroviral *de novo* methylation due to the existence of a specific methyltransferase activity that would be absent or less active later in development.

Injections of various constructs into *Xenopus* oocytes together with *in vitro* cell culture transfection experiments revealed that mammalian and viral gene expression is sensitive to methylation at the 5' end of the genes. These promoter regions are rich in CpGs and overlap with a nucleosome-free gap [105–107] (Figure 3C), where methylation was proposed to directly influence protein binding, nucleosome positioning, and thus transcription [105,108]. Further work suggested that only certain CpGs within these regions are responsible for controlling a gene's expression [107,109,110]. Intriguingly, two studies that infected **embryonic carcinoma (EC) cells** with M-MuLV showed that proviral sequences are quickly silenced upon infection, yet only become methylated days later, demonstrating that DNA methylation is not the primary silencer of these genes [111,112]. Furthermore, Niwa and colleagues showed that removing methylation using 5-aza on EC cells did not reactivate the virus; however, if the cells were first differentiated before 5-aza treatment, methylation removal was sufficient to activate M-MuLV. The authors therefore hypothesized that there could be two distinct mechanisms regulating gene expression during different stages of development, with early development being independent of DNA methylation and late development being dependent on it [112].

The discovery of CpG islands (CGIs)

Although sperm is overall highly methylated, undermethylated stretches like those found in early metazoan embryos were also seen in sperm cells at constitutively expressed genes [94,95,113–115]. Bird and colleagues then showed that these unmethylated regions are comprised of CpG-rich DNA and are preserved unmethylated across many mouse tissues. Given their sensitivity to HpaII nuclease digestion, they were initially named HpaII tiny fragments (Figure 3D) [116] and later popularized as CGIs. Bird reasoned that the structure of CGIs at housekeeping genes could reflect their continuous occupancy by proteins protecting them from methyltransferases and, thus, from the accompanying mutational loss of CpGs by spontaneous deamination of 5mC [117] (Box 2). CGIs, with their genomic rarity but genic abundance, may present unique platforms for protein-DNA interaction that could be modulated by DNA methylation [110,118]. The possibility of a more complex translation of the methylation signal on DNA emerged through experiments initially aimed at identifying proteins that exclusively bind the characteristic stretches of unmethylated, CpG-rich DNA. However, in these experiments, Bird and colleagues unexpectedly found the MeCP1 protein complex that formed not on unmethylated DNA, but rather on methylated probes or highly methylated genomic DNA [119].

5mC dynamics from gametes through early development

In 1987, Monk *et al.* applied a novel low-input DNA end-labeling technique to assess DNA methylation in germ cells, zygotes, and early developing mouse embryos [120]. They found oocyte

preimplantation but not postimplantation embryos, suggesting that only the preimplantation embryo is capable of *de novo* methylation. (C) Kruczak and Doerfler tested the positional effect of 5mC on the expression of viral chloramphenicol acetyl-transferase (CAT) by cloning viral promoters with varying CpG distribution and content into a vector they then methylated with M.HhaI *in vitro* before measuring its expression capabilities. CAT expression was assessed by detecting the acetylation of chloramphenicol (CAM) using thin-layer chromatography. Methylation within a promoter sequence inhibited transcription, while methylation flanking the promoter or within the gene did not. (D) Bird and colleagues used end-labeling to visualize the low quantity of DNA resulting from HpaII digestion of sea urchin genomic DNA into tiny fragments. These HpaII tiny fragments (HTF) are the result of the presence of regions containing many HpaII motifs (5' C-C-G-G 3') that are unmethylated, regions now known as CpG islands (CGIs). (E) Monk and colleagues investigated methylation changes during mouse embryonic development using end-labeling of digested genomic DNA. The varying amounts of large HpaII fragments indicate a differential gain of 5mC in the embryonic and extraembryonic tissues. The strong signal at the bottom of the lanes stems from the sensitive detection of canonically unmethylated CGIs. See [104,107,111,112,116,120,126].

Box 2. Mutagenic effects of 5mC

Several studies in the 1950s and 1960s demonstrated that animal and plant DNA is relatively AT rich and GC poor and that the CpG dinucleotide specifically is under-represented. This raised questions about the biological function and evolution of 5mC, yielding the hypothesis that 5mC may direct DNA sequence mutations controlling gene expression [32,33,48,277,278,280]. Evidence supporting this hypothesis first came from Coulondre and colleagues who found that mutational hotspots in an *E. coli* gene were located at 5mC residues but disappeared when this gene was demethylated by being introduced into bacterial strains deficient in DNA methyltransferases [281]. The authors proposed that spontaneous deamination of 5mC to thymine is the basis of its increased mutagenesis. When cytosine is deaminated it produces uracil, which results in a mismatch that is easily recognized as an unnatural base in DNA, excised, and repaired by DNA-uracil glycosylase. However, deamination of 5mC produces thymine. Coulondre *et al.* found that the resulting mismatch is often not properly corrected; instead, the guanine on the opposite strand may be excised and repaired as adenine, thus generating the frequent CG→TA mutations [281]. These findings were corroborated by an analysis looking at the potential correlation of CpG frequency with DNA methylation levels in animals. Bird found that organisms with the lowest CpG frequency had higher levels of DNA methylation and an excess of CpTs and CpAs, whereas lowly methylated genomes displayed no deficiency in CpGs and no excess of CpTs and CpAs [282].

In 1989, Cooper and Krawczak sought to determine the rate of deamination of 5mC *in vivo*, however, the process is too slow for accurate measurements. They therefore measured the rate of 5mC deamination using controlled *in vitro* conditions and used these values in a mathematical model to estimate the time span over which the ‘CpG suppression’ observed in vertebrate genomes would have occurred [283]. They estimated the length of time needed to achieve the GC:AT composition of the vertebrate genome today to be 450 million years. The authors proposed that the adaptive radiation of vertebrates might have coincided with the evolution of a heavily methylated genome [283]. More recent estimates predict CpGs to be about 18 times more likely to mutate than non-CpGs in the human germline and the CpG context is over-represented among germline mutations, suggesting that the hypermutability of 5mC may be an important contributor to human genetic diseases [284,285]. Given that the cost of 5mC in the genome is a slow erosion of CpGs with potentially deleterious consequences for germline mutation, tolerance of these negative effects further supported the likely functional importance of 5mC in the genomes of vertebrates [282,285].

DNA was relatively undermethylated, while sperm DNA was highly methylated. Following zygote formation, embryonic methylation decreased steadily to its lowest point around the onset of implantation. Next, the extraembryonic lineages and embryo proper were progressively and distinctly *de novo* methylated through gastrulation (Figure 3E) [121].

Cloning the first DNA methyltransferases

Cloning and protein sequencing of bacterial methyltransferases provided important insights about the evolutionary conservation and functionality of their catalytic domains [122,123]. In particular, elucidation of the chemistry of the methyl-transfer reaction, as well as the enzymology of bacterial methyltransferase sequence specificity, came from studies of the bacterial enzyme M.Hhal [124–126]. Building on this momentum, a series of studies that purified and characterized DNA methyltransferase activities from mammalian tissues [127–130] led to the successful cloning of murine *DNA methyltransferase 1* (*Dnmt1*), the first mammalian DNA methyltransferase, by Bestor and colleagues in 1988 [131].

Summary

The 1980s provided crucial insights into the overall distribution and function of DNA methylation, including the discovery of CGIs and a much-improved understanding of the role of 5mC in gene regulation. The successful cloning and characterization of over a dozen bacterial and the first mammalian methyltransferases represented another milestone. Although a general consensus emerged that 5' promoter methylation suppresses transcription, some exceptions to this rule highlighted that further studies were still needed to appreciate and interpret the context-specific role of DNA methylation [97,99,100]. Despite these key advances in mammalian systems, it is worth noting that the dominant model organisms at the time included *Saccharomyces cerevisiae*, *Drosophila melanogaster*, and *Caenorhabditis elegans*, which do not have DNA methylation; this may explain some remaining concerns in the community about the wider significance of 5mC.

1990–1999: The eukaryotic methyltransferases

The 1990s saw major progress in our molecular understanding of phenomena such as genomic imprinting and X chromosome inactivation that we are unable to cover here in detail but have been reviewed extensively elsewhere [132–134] (Box 3). Beyond that, this decade was particularly noteworthy for a series of studies that identified, cloned, and knocked out mammalian methyltransferases, revealing their essential roles during development. In parallel, *Arabidopsis thaliana* gained traction as a powerful model organism for epigenetic studies, including DNA methylation (Figure 4A).

Methylation artifacts of *in vitro* culture

As the Bird lab continued their investigation of CGIs, in 1990 they noted a propensity of certain CGIs to become aberrantly methylated during *in vitro* culture (Figure 4B). The mechanism that protects these *in vivo* and the reason for their susceptibility in cell culture remained unclear. CGIs typically lie in open, accessible chromatin and, hence, they speculated, would have to be actively protected from methylation during normal development and tissue homeostasis [135].

The essential role of DNA methylation in mammalian development

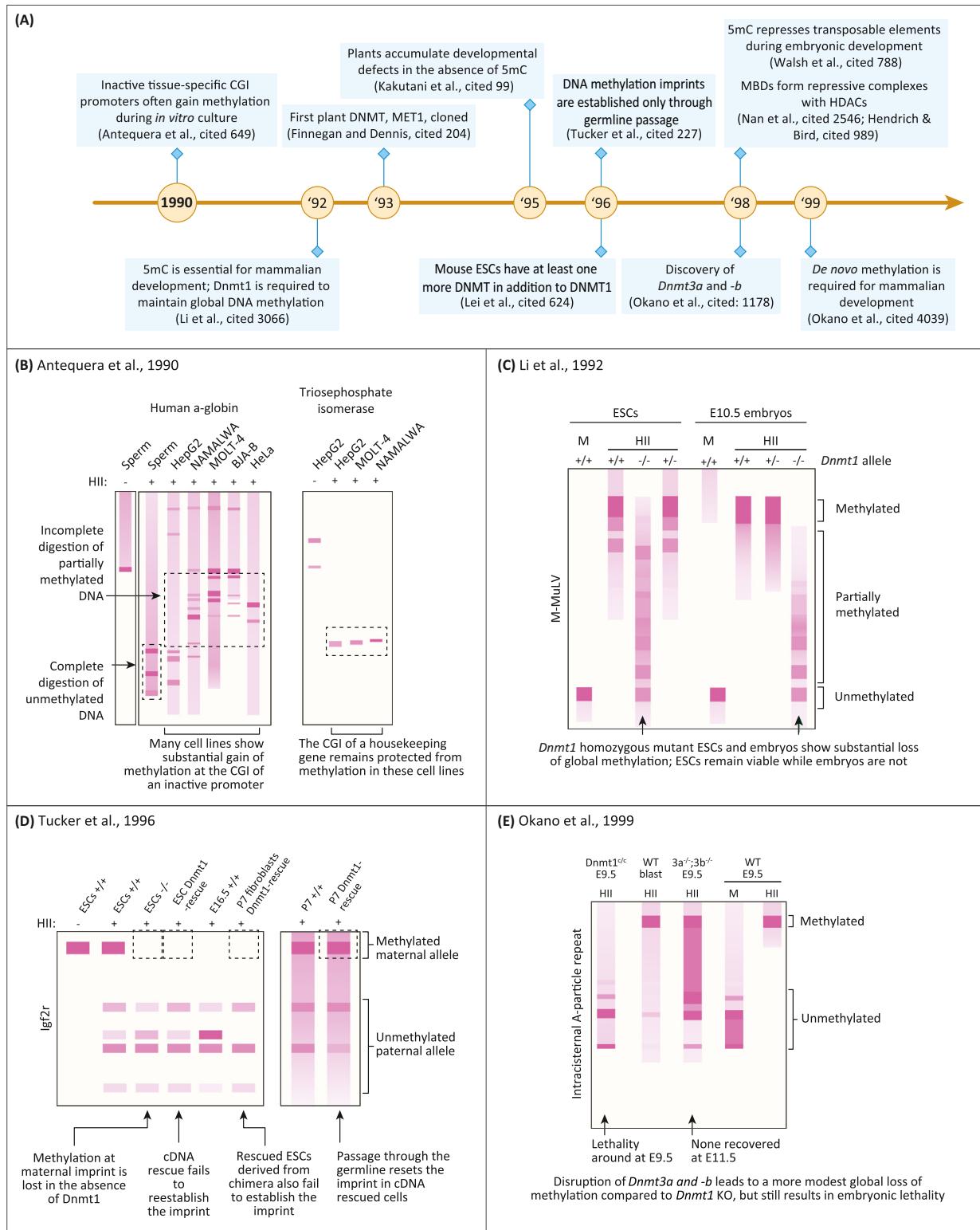
In 1992, the Jaenisch Lab disrupted mouse *Dnmt1* activity that resulted in global loss of most, but not all, DNA methylation *in vitro* and *in vivo* (Figure 4C) [136]. Undifferentiated *Dnmt1* knockout (KO) mouse embryonic stem cells (ESCs) remained viable, while differentiated ESCs arrested or died. Similarly *Dnmt1* KO mice displayed developmental delays and lethality around embryonic day E8.5–E9.5 [136–138]. Reintroducing a *Dnmt1* cDNA into the endogenous locus of KO cells restored global DNA methylation but failed to rescue imprinted methylation [139], which could only be achieved through germline passage [140] (Figure 4D). Combined, these studies

Box 3. Imprinting and X inactivation

The term ‘imprint’ was coined by Helen Crouse in 1960 following her experiments on sex determination in *Sciarra*, a species of fungus gnat, that loses the paternal X chromosome during differentiation of the germline. Chromosomal translocations revealed that a specific region of the X chromosome carries an ‘imprint’ established during germline development that determines which X chromosome is lost [286]. In the mid-1980s, several studies demonstrated that the maternal and paternal genomes are both needed for mammalian embryonic development, which led to the discovery of genes whose expression was determined by the parent that gene was inherited from [287,288]. The idea that DNA methylation may play a role in imprinted gene expression arose following studies showing that germ cells display sex-specific methylation patterns [120,289,290] and that DNA methylation plays a role in the imprinted expression of single-copy transgenes [291–293].

During the first half of the 1990s, many imprinted genes, such as *Igf2*, *Igfr2*, *H19*, *Snrpn*, and *Peg1* were mapped and further characterized in mouse and human [294–298]. To date, 25 clusters of ICRs regulating multiple imprinted genes have been discovered [134]. Maternally or paternally imprinted alleles are methylated at the imprinting control regions (ICRs) in the oocyte and sperm, respectively, by DNMT3A and DNMT3L [192,194]. Though the oocyte and sperm methylomes differ quite extensively such that thousands of genes show differential DNA methylation in the early embryo, only a subset resist preimplantation DNA demethylation and maintain their parent-of-origin specific DNA methylation [299,300]. The zinc-finger protein ZFP57 recruits the KAP1-complex and DNMT1, maintaining methylation at these select sites [301–303]. In addition to DNA methylation imprints, a small number of noncanonical, maternal H3K27me3 imprints were found in the extraembryonic ectoderm [304], and more recent work has described the different imprinting mechanisms at play in the embryonic and extraembryonic lineages [305].

Prior to Crouse’s studies on sex determination, research on the regulation of X chromosomes in female cells had already been underway for a decade following the 1949 description of the Barr body, by Barr and Bertram, as a nucleolar satellite found in female cat neurons that they speculated to be ‘sex chromatin’ [306]. Ten years later, Ohno confirmed that the Barr body is indeed the X chromosome, reporting that one mammalian female X is condensed and heterochromatinized, while the other is euchromatic [307]. In 1961, Lyon’s observations about the coat color of female mice provided critical evidence that one female X is inactive while the other is active, and she postulated that the condensed X is genetically inactive [308]. In 1975, Riggs was the first to propose the X chromosome inactivation center and that the process of X inactivation as a whole might be controlled by DNA methylation [58]. While this idea was investigated by several groups throughout the early 1980s [309–312], DNA methylation was ultimately ruled out as the primary initiator of X inactivation, although it, along with other epigenetic modifications, plays a role in stable gene silencing on the X chromosome [313].



Trends In Genetics

(See figure legend at the bottom of the next page.)

provided crucial insights and highlighted the essential role that DNA methylation plays in normal development.

The first crystal structure of a DNA methyltransferase

In 1993, the Roberts lab published the first crystal structure of a DNA methyltransferase, M.Hhal, in a complex with the methyl donor SAM at 2.5 Angstrom resolution [141]. The core structure of M.Hhal, being one of the smallest methyltransferases, was found to consist largely of the sequence motifs conserved across all 5mC methyltransferases. The mechanism of a methyl group transfer to DNA as determined by Wu and Santi in 1987 [126] revealed that the nucleophilic attack on the targeted cytosine central to this mechanism results in a covalent intermediate between DNA and the methyltransferase [142]. This intermediate can be trapped by substituting the hydrogen at the C5 position with a fluorine [143]. Using this technique, the crystal structure of the M.Hhal-SAM complex trapped in its covalent intermediate revealed that both methyltransferase and DNA change conformation. B-form DNA is bound in a cleft formed by the two domains of M.Hhal with the target cytosine base flipped completely out of the DNA helix and into the active site of M.Hhal [144].

Discovery and characterization of the mammalian *de novo* Dnmts

The residual methylation present in *Dnmt1* KO cells provided an indication of the existence of additional methyltransferases [136,137]. Previous work demonstrated that the catalytic domains of different bacterial cytosine methyltransferases exhibit a great degree of homology with each other as well as DNMT1 [123,126,131,145–147]. Therefore, different groups took advantage of homology-based **BLAST** searches and independently identified additional eukaryotic DNA methyltransferases [148–151]. *Dnmt2* was subsequently ruled out as a candidate *de novo* methyltransferase, given its apparent lack of 5mC methyltransferase activity in ESCs [150]. *Dnmt3a* and *Dnmt3b*, however, were able to methylate CpGs of both hemimethylated and fully unmethylated substrates with equivalent efficiencies *in vitro* while showing high expression in mouse ESCs and generally lower expression in somatic cells [151]. Single and double KO of *Dnmt3a* and -*b* in mouse ESCs and their corresponding KO mouse models demonstrated the essential and partially overlapping roles of these enzymes during embryonic development (Figure 4E) [152].

DNA methylation readers

The Bird lab showed that the methylated versions of four different promoters were each transcribed equivalently to their unmethylated versions when MeCP1 binding was diluted by adding methylated competitor DNA to the *in vitro* reaction [153]. Likewise, extracts from cells deficient for MeCP1 used in analogous assays were unable to efficiently repress transcription of methylated DNA. Similar results obtained from the transient transfection of methylated DNA and competitor DNA suggested these mechanistic principles might also apply *in vivo*. These experiments led the

Figure 4. Discovery and characterization of the eukaryotic DNA methyltransferases. (A) Key advances in the field of DNA methylation between 1990 and 1999 include the cloning and characterization of plant and mammalian DNMTs. (B) Antequera and colleagues tested the methylation status of CpG islands (CGIs) at housekeeping and tissue-specific genes in a variety of cell lines using a methylation-sensitive HpaII digestion. They detected varying degrees of aberrant methylation at CGIs of tissue-specific genes that are not active in the respective cell lines, while CGIs of housekeeping genes remained methylation free. (C) Li *et al.* assayed genomic 5-methylcytosine (5mC) levels by methylation-sensitive HpaII digestion in wild type (wt) (+/+) and *Dnmt1* mutant (+/–, –/–) embryonic stem cells (ESCs) and E10.5 embryos. *Dnmt1*–/– cells were extensively demethylated and, while these hypomethylated ESCs were viable, *Dnmt1*–/– derived embryos were not. (D) Tucker *et al.* stably integrated a *Dnmt1* cDNA into the endogenous locus of *Dnmt1*–/– cells. While global methylation was restored, methylation of imprints, exemplified here by methylation-sensitive digestion of the *Igf2r* locus, was not restored until the rescued cells were passaged through the germline. (E) This digestion by Okano *et al.* visualized the methylation status of the intracisternal A-particle (IAP) repeat of *Dnmt1* knockout (KO) (*Dnmt1c/c*), *Dnmt3a/Dnmt3b* double KO, and wt embryos, as well as wt blastocysts. KO of *Dnmt3a* and -*b* resulted in widespread demethylation of this repeat, indicated by the presence of small and medium-sized fragments; however, the loss of methylation was not as extensive as in *Dnmt1* KO embryos. See [135,136,138,140,150,152,157,159,167,171,180]. Abbreviations: Blast, Blastocyst; ESCs, embryonic stem cells; HpaII; M, MspI; MBD, methyl-CpG binding domain.

Bird lab to hypothesize that it is not the direct effect of DNA methylation that inhibits transcription, but rather an indirect effect involving MeCP1 binding of promoter proximal methylated CpGs [153]. The next year, a second methyl-CpG binding protein, MeCP2, was found to be capable of binding a single symmetrically methylated CpG. It localized predominantly, but not exclusively, to pericentromeric heterochromatin, which in the mouse genome covers approximately 40% of all 5mC and is therefore readily visible by immunofluorescence [154]. MeCP2 also repressed transcription from the DNA templates it bound *in vitro* [155]. While MeCP2 is not necessary for ESC viability, the mouse KO of MeCP2 by chimera formation from mutant ESCs showed a variety of mutant phenotypes scaling with the degree of contribution [156]. In a series of pulldown experiments, it was found that the transcriptional repression domain of MeCP2 binds another repressive complex containing a histone deacetylase and the repressor mSin3A. To test the mechanism implied by the association of these complexes, cells were treated with a histone deacetylase inhibitor, which resulted in transcriptional derepression [157], findings in line with the previously shown repressive nature of histone deacetylation in yeast [158] (Box 4). Several other proteins containing methyl-CpG binding domains (MBDs) were identified in mouse and human with both overlapping and unique expression patterns. All MBDs except MBD3 appear to bind methylated, but not unmethylated, DNA [159], and MBD2 was found to belong to the MeCP1 histone deacetylase complex [160].

DNA methylation in *Arabidopsis thaliana*

Some of the earliest studies on DNA methylation had observed high 5mC content in plant DNA [27,32,52] and noted that both CpG and CpXpG (X = A, C, or T) methylation occur [52,161]. While the genomes of plants such as wheat were an asset for obtaining large quantities of DNA, this property proved difficult for mapping and cloning genes. In the 1980s, *Arabidopsis thaliana* became a popular genetic model, in part due to its smaller genome size [162] as well as advances in transformation protocols [163–165] that made it amenable to mutational studies [166].

In 1993, Finnegan and Dennis identified and cloned the first plant DNA methyltransferase, *MET1*, using sequence homology to murine and bacterial methyltransferases [167]. That same year, a mutational screen looking for genes involved in DNA methylation in *Arabidopsis* identified the

Box 4. Histone modifications in gene regulation

As the field of DNA methylation grew to cover ever broader disciplines of study, parallel work on the organization of DNA helped show that gene regulation did not just occur at the DNA surface but also involved histone proteins. As early as 1950, histones were first proposed to be possible gene repressors [314] and, by 1964, histone acetylation was found to be associated with active transcription [315]. This implied that DNA did not need to be stripped of all nucleosomes to be used for active transcription, contrary to previous thinking about the mechanics of transcription. Until 1984, when the first crystal structure of the nucleosome was published [316,317], it was not known in what manner a histone might be modified. These structures revealed that histones have amino-terminal tails that extend beyond the DNA wrapped octamer and that are free to be chemically modified. In 1991, a series of studies from Grunstein and colleagues led to the important finding that modification of H4 histone tails in yeast is required for activation of promoters [318]. By 1996, important cross species comparisons of a yeast transcriptional activator and repressor to the *Tetrahymena* histone acetyltransferase [319] and mammalian histone deacetylase [158], respectively, confirmed a role for histone deacetylation in the repression of transcription in yeast. As histone modifications are essential not only in organisms lacking DNA methylation, such as yeast, but also in the highly methylated mammalian genome, a critical question in the 1990s was how DNA methylation might interface with histone modifications to regulate gene expression and control access to the DNA. By the year 2000, the histone code hypothesis postulated that combinations of histone tail modifications provide a code for genome regulation [320]. The next year, several groups demonstrated that H3K9 methylation placed by SUV39 creates heterochromatin protein 1 (HP1) binding sites, which in turn recruits more SUV39. This process establishes a propagative cycle of silencing by heterochromatinization [321–323]. Though clear mechanistic connections between the different layers of epigenetic modifications have been elucidated, it remains to be investigated how and when DNA methylation acts in concert with or independently from histone modifications to coordinate gene expression, and what role methylation-sensitive protein binding plays in this process.

DDM1 locus, disruption of which resulted in the demethylation of mainly repeat elements [168,169]. Despite seeming to be nonessential for plant development, DNA methylation was demonstrated to be involved in a number of physiological processes, such as the vernalization required to induce flowering in *Arabidopsis*, where both vernalization and 5-aza treatment resulted in similar degrees of DNA demethylation. Burn and colleagues therefore hypothesized that demethylation of genes involved in flowering may be required for the initiation of this process [170]. Further studies revealed that the developmental defects associated with hypomethylation increase in severity with several rounds of self-fertilization [171]. In contrast to mice, methylation levels in *Arabidopsis* were only slowly restored when functional alleles of *DDM1* and *MET1* were reintroduced through crossing, implying that plants do not reset their DNA methylation between generations and do not undergo phases of strong *de novo* methylation during development [168,172]. Finnegan and colleagues demonstrated that demethylation results in the aberrant expression of floral homeotic genes in leaves, further linking DNA methylation to transcriptional regulation in plants [172].

Dnmt classes in plants

Loss of *MET1* was found to reduce 5mC at CpG dinucleotides only, while loss of *DDM1* reduced 5mC within both the CpG and CpXpG contexts [168,172]. Biochemical studies of pea *MET1* supported the idea that *MET1* is closely related to the mammalian maintenance methyltransferase DNMT1, as it also preferentially acts on hemimethylated CpGs [173]. These findings pointed towards the existence of further putative Dnmcts in plants. In addition, in 1997, Jacobsen and Meyerowitz showed that a phenotype observed in plants with reduced *MET1* activity is caused by the hypermethylation of a specific regulatory region [174].

The next year, chromomethylase 1 (CMT1) was identified in *Arabidopsis* [175] and more genes encoding *MET1*-like and CMT1-like methyltransferases were detected [176]. By 1999, *DDM1* was shown to encode a SWI2/SNF2-type chromatin remodeler required for the methylation of repeat elements as well as high-fidelity maintenance methylation of low-copy sequences [177]. Linking their findings to the recently uncovered interplay between 5mC and histone deacetylation, Jeddeloh and colleagues proposed that chromatin remodeling to alter histone modifications or to increase DNA accessibility is required for proper DNA methylation [177].

Finally, a third class of DNA methyltransferases was identified. Using the sequence of the recently identified Dnmt3 enzymes in mammals [151], Jacobsen and colleagues discovered homologous genes in maize, soy, and *Arabidopsis* [178]. However, in contrast to all previously identified classes of DNA methyltransferases in plants and mammals, the conserved sequence motifs in the catalytic domain of these novel enzymes were found to be arranged in a different order, leading to the name domains rearranged methyltransferase (DRM) in *Arabidopsis* [178].

Biological function remained controversial

In 1997, the Bestor lab published a review article in *Trends in Genetics*, the central thesis of which was the idea that the primary function of DNA methylation may be to suppress parasitic sequence elements in the mammalian genome [179]. Bestor and colleagues noted that a subset of genes were subject to allele-specific expression via imprinting or part of X chromosome inactivation. Beyond that, it seemed less clear that methylation of silent genes in somatic tissues prevents their transcription and that promoter demethylation would result in gene activation *in vivo*. Although differentiation of cells upon 5-aza treatment pointed to a demethylation mechanism of gene activation [79], more recent work had shown that the CGI promoters of such genes are not normally methylated *in vivo*, but frequently become so *in vitro* [135]. Thus, their activation by demethylation, at least in this context, was not a true developmental mechanism. However, they argued

that the detection of transposon transcripts during lowly methylated developmental stages was good evidence of their regulation by DNA methylation. The Bestor lab continued to study the regulation of retrotransposons and a year later showed the massive upregulation of intracisternal A-particles (IAPs) in Dnmt1-deficient mouse embryos [180].

In a response article in the same journal, Bird argued strongly against this hypothesis, stating that there was little to no evidence for the silencing of transposable elements in germ cells and stem cells by DNA methylation. In contrast, he pointed to evidence consistent with the silencing of transposons by DNA methylation in somatic cells, where transposon activity would be more detrimental than in germ and stem cells [112]. Why, Bird asked, would an organism fail to protect its germline from these parasitic DNA elements while it can do so effectively in somatic cells? Furthermore, given the relatively low number of individual loci studied in great detail, it was too soon to conclude that none of the many genes with promoters that lie within highly methylated regions rely on DNA methylation for silencing [181].

Summary

The 1990s were clearly defined by the functional evaluation of the mammalian and plant DNA methyltransferases and 5mC readers having methyl-binding domains, as well as investigations of transcriptional repression. The essential, though incompletely understood, role for genomic methylation helped create a broader interest in DNA methylation within the scientific community. Unfortunately, despite much excitement, the question of whether and how active DNA demethylation occurs, remained unresolved at the end of this decade, which yielded no clear mechanism despite many attempts and publications.

2000–2009: Genome-wide DNA methylation landscapes

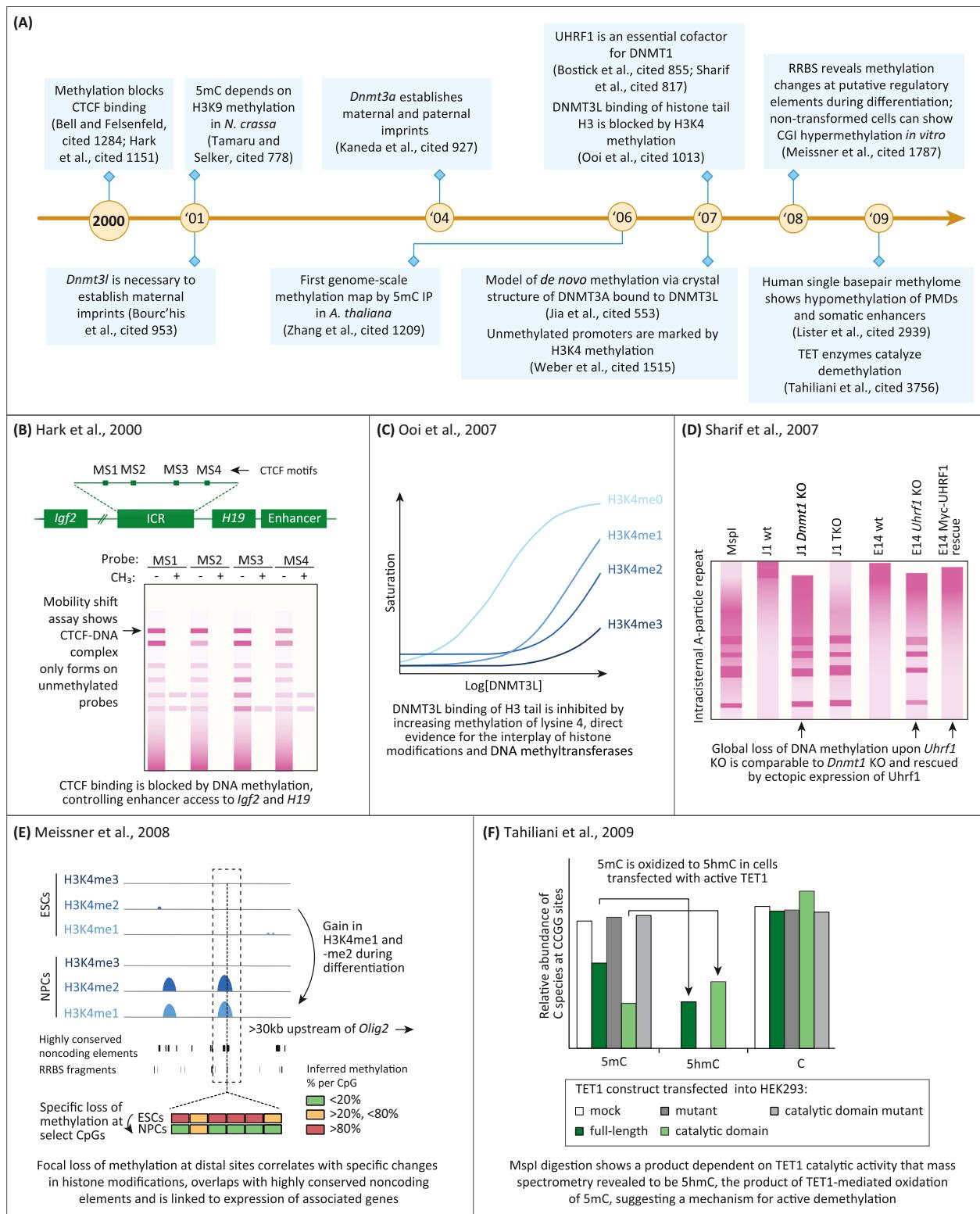
The early 2000s saw a major transition in the biological sciences with the completion of several draft genome assemblies, including *Arabidopsis*, mouse, and human [182–184]. These genome maps in turn enabled investigation of additional layers of the epigenome and hence served as the basis that allowed inferences first made about the regulation of individual genes to be expanded to whole genome scales by the end of the decade. Computational tools and more cost-effective sequencing methods emerged in the second half of this decade. A technique called **bisulfite sequencing**, first reported in 1992 [185], became the gold standard for quantifying and mapping methylation. These tools were rapidly applied to generate the first genome-scale methylation maps (methylomes). In parallel, major advances in the fundamental biology of methylation readers, writers, and erasers were reported (Figure 5A).

Impact on enhancer–promoter loops

The study of the role of 5mC in imprinted gene expression led to the finding that *Igf2* and *H19* have reciprocal expression controlled by the same **imprinting control region (ICR)**, raising the question of how methylation can lead to the transcription of one gene and the repression of the other. Two studies identified four **CCCTC-binding factor (CTCF)** binding motifs in the *Igf2/H19* ICR. The methylation or deletion of these binding sites, which then prevented CTCF binding, reduced the ICR's enhancer blocking activity [186,187] (Figure 5B). This suggested that methylation of this ICR may determine whether an enhancer contacts the *Igf2* or *H19* promoter through the methylation-sensitive binding of CTCF and that methylation outside of promoters can have regulatory activity that influences looping of the DNA.

Linking DNMTs to histone modifications

In 2001, Tamaru and Selker found the *Neurospora crassa* gene *dim-5* to be required for normal DNA methylation patterns and showed it encodes an enzyme catalyzing **histone 3 lysine 9**



Trends In Genetics
(See figure legend at the bottom of the next page.)

(H3K9) methylation [188]. Partially replacing H3K9 with H3L9 or H3R9 led to reduced 5mC levels, demonstrating that H3K9 methylation is required upstream of some DNA methylation. This led to the idea that eukaryotic Dnmts may depend on specific histone modifications as co-factors guiding their activity [188]. Jackson and colleagues next demonstrated that this principle holds in plants, reporting that the *Arabidopsis* H3K9 methyltransferase KRYPTONITE is required for CpXpG methylation placed by the DNA methyltransferase CMT3 [189,190]. In addition, they showed that LHP1, an *Arabidopsis* homolog of heterochromatin protein 1 (HP1) and reader of H3K9 methylation, interacts with CMT3, and proposed a mechanistic link that extends across animals and plants that have homologs of HP1 (Box 4).

Structure and function of Dnmts with their cofactors

The discovery and subsequent KO of the catalytically inactive cofactor DNMT3L highlighted its role in imprinting [191–193]. Kaneda and colleagues later used germline conditional KOs to show that both *Dnmt3a* and *Dnmt3l* are essential for the establishment of maternal and paternal imprints [194]. Protein biochemistry of DNMT3L revealed that it binds to DNMT3A, DNMT3B, and histone H3, but this interaction with H3 is inhibited by **histone 3 lysine 4 (H3K4) methylation**. This provided experimental evidence that DNMT activity can be directly blocked by a specific histone modification [195] (Figure 5C). Further structural insights into the mechanisms guiding *de novo* methylation were provided by the first crystal structure of a mammalian DNMT published in 2007 [196]. The structure comprised the C terminal domains of DNMT3A and DNMT3L in complex with each other. Later that year, an essential cofactor for DNMT1, ubiquitin like with PHD and ring finger domains (UHRF1 or NP95), was discovered and shown to bind the methylated cytosine of hemimethylated DNA. Its depletion led to global loss of CpG methylation that largely phenocopied the loss of DNMT1, demonstrating that UHRF1 is required for the recruitment or activity of DNMT1 [197,198] (Figure 5D).

Genome-wide methylation maps: the first methylomes

The first genome-scale methylation maps were produced in the mid-2000s for the *Arabidopsis* genome using enzyme- and antibody-based enrichment of 5mC followed by microarray hybridization [199,200]. While previous work in *Arabidopsis* had identified DNA methylation at repeats and transposable elements [201], genome-scale maps both confirmed and extended these findings. They showed that the 5' ends of genes are largely devoid of CpG methylation, while bodies of constitutively expressed genes are extensively methylated and relatively CpG depleted. Henikoff and Jacobsen hypothesized that genic methylation might serve to silence cryptic intragenic promoters [199]. Extensive methylation at regions producing small interfering RNAs and intergenic noncoding RNAs suggested that these elements may also be controlled by DNA methylation [200]. In parallel with these plant studies, the Bestor and Schübeler labs used similar enrichment-based techniques to create the first comprehensive maps of human DNA methylation [202,203]. These maps strengthened previous work supporting the idea that promoter CGIs are

Figure 5. Interplay of histone modifications with 5-methylcytosine (5mC) and active demethylation by the TET enzymes. (A) In the 2000s, global surveys of 5mC and the discovery of Dnmt cofactors underscored the connection of 5mC to other epigenetic modifications. (B) This binding assay by Hark and colleagues determined that CTCF cannot bind its methylated motifs (MS1–4) within the imprinting control region (ICR) of the *H19/Igf2* locus when they are methylated, as it is the case on the paternal allele. (C) Ooi *et al.* characterized the dissociation behavior of H3 tail peptides from recombinant human DNMT3L using fluorescence polarization. This assay showed that DNMT3L efficiently binds unmodified H3 peptides (H3K4me0), but hardly binds peptides methylated at lysine 4, even at high DNMT3L concentrations. (D) Methylation-sensitive restriction digestion of genomic DNA probed for the intracisternal A-particle (IAP) sequence by Sharif and colleagues indicated extensive demethylation of both *Dnmt1* and *Uhrf1* knockout (KO) embryonic stem cells (ESCs). (E) Reduced representation bisulfite sequencing (RRBS) data generated by Meissner *et al.* showed a loss of methylation at highly conserved noncoding sequences, indicative of regulatory elements, upon differentiation of ESCs to neural progenitor cells (NPCs). The alignment with ChIP data revealed a concomitant gain of H3K4 methylation at these regions. (F) Tahiliani and colleagues transfected HEK293 cells with various constructs to determine the effect of TET1 on 5mC levels. This included the catalytic domain of TET1 or full-length TET1, either in its wild type (wt) or a catalytically inactive mutant form. Genomic DNA was digested with MspI and the fragments were resolved by thin-layer chromatography and then quantified, revealing the presence of a novel nucleotide, subsequently confirmed to be 5-hydroxymethylcytosine. See [186–188,192,194–198,200,206,208,211,214]. Abbreviations: CGI, CpG islands; PMD, partially methylated domain.

predominantly unmethylated in normal human cells; however, the Schübeler lab found hypermethylation at CGIs in one colon cancer cell line, consistent with the earlier observations of the Bird lab [135]. Interestingly, this cell line also showed notable hypomethylation across large gene-poor domains and the inactive X chromosome [202].

Further advances in sequencing technologies then enabled the first single-base resolution methylation maps in *Arabidopsis*, mouse, and human ESCs as well as somatic tissues [204,205]. Reduced representation bisulfite sequencing (RRBS) showed limited dynamic methylation over the majority of promoters and uncovered specific focal methylation changes at more distal, highly conserved putative regulatory sites [206] (Figure 5E). Paired with histone modification maps [207] these studies highlighted clear correlations between changes in 5mC and changes in H3K4 methylation during differentiation as well as the relationship to other modifications, including H3K27me3 [206,208]. In addition, maps of 5mC during neural differentiation showed that aberrant culture-induced CGI hypermethylation can also occur in nontransformed cells [206]. Ecker and colleagues showed that CpG methylation in immortalized cells, in contrast to human pluripotent stem cells, is reduced across large areas of the genome, with about 40% of the genome now being covered by partially methylated domains (PMDs) [208]. Furthermore, both mouse and human maps demonstrated the hypomethylation of active enhancers, which may be a cause or consequence of protein–DNA interactions at these regulatory sites [206,208].

Promoter architecture and function

Saxonov and colleagues performed computational analysis of the human genome that revealed two classes of promoters: low CpG (LCG, 30% of promoters) and high CpG (HCG, 70% of promoters) density promoters [209]. Genes associated with LCG promoters tended to have transcriptional potential independent of their promoter methylation and most of these promoters were typically methylated. In contrast, most HCG promoters were found to be unmethylated, even when their associated gene was not being transcribed. In specific contexts where these HCG promoters do get methylated, such as germline genes in somatic cells, this methylation is sufficient to prevent transcription [203]. These studies, together with previous work [210], further highlighted that most endogenous promoters are not primarily regulated by methylation and that unmethylated promoters are enriched for H3K4 methylation [211]. Combined with the DNMT structure and biochemistry papers from 2007 [195,196], these studies helped build a consensus that H3K4 methylation may shield these promoters from DNMTs.

Active DNA demethylation

Immunostainings of 5mC in early mouse preimplantation embryos showed a rapid loss of 5mC within the paternal pronucleus before the first cleavage division [212]. In contrast, 5mC signal in the maternal pronucleus was largely stable up to the two-cell stage. Importantly, the paternal loss of 5mC was independent of DNA replication, meaning that an active mechanism must be at work in the first hours following fertilization. In the decade before and after this finding, many papers had been published that claimed to have identified this demethylation activity, but none led to robust and reproducible mechanisms [213]. Finally, a computational search in 2009 pointed towards TET1, TET2, and TET3 and their orthologs [214]. Specifically, TET1 was then shown to catalyze the conversion of 5mC to **5-hydroxymethylcytosine (5hmC)** in mammalian cells (Figure 5F). Published back-to-back, Kriaucionis and Heintz also demonstrated that 5hmC is abundant in neurons, further strengthening the role of 5hmC in the active demethylation pathway [215]. Despite the general conservation of the TET enzymes across metazoans, fungi, and algae, plants appear to rely more on DNA glycosylases for the removal of methylation, particularly at genes near highly methylated repetitive sequences [216].

Summary

This decade provided key insights into the genome-wide distribution of DNA methylation that, together with mechanistic and structural advances, established a more holistic view of DNA methylation. These efforts allowed the principles that had been established individually and at lower resolution in the prior decades to be generalized across the genome, for example, by showing the sufficiency but not necessity of 5mC to repress transcription from CpG-dense promoters. However, the finding that most promoters are in fact not regulated by 5mC prompted a paradigm shift away from a promoter-centric view to a more comprehensive, global approach to understanding DNA methylation. Importantly, the interconnectedness of 5mC with other epigenetic modifications, such as lysine 4 methylation of H3 peptides, also became increasingly clear. The long-sought discovery of an actual demethylating enzyme was a major milestone to explain the focal DNA methylation dynamics that occur during differentiation and in normal development.

2010–2019: DNA methylation is dynamic

Several comprehensive reviews have been published over the past years covering the more recent discoveries from this decade [14, 16, 17, 217–219]. To complement this, we selected a subset of experimental advances that further expanded the fundamental principles of 5mC and its role in genome regulation. Increased sensitivity, throughput, and affordability in sequencing technologies enabled a range of new mapping studies that provided the basis for a number of novel insights (Figure 6A).

Crosstalk of DNA methylation and transcription factors (TFs)

As DNA methylomes were measured with increasing resolution for various cell types, it became clear that lowly methylated CpGs are often concentrated in short regions that do not overlap with CGIs [220–222]. Schübeler and colleagues took a closer look at lowly methylated regions (LMRs) in mouse ESCs and neural progenitors and showed that they comprise about 4% of all CpGs and adopt an intermediate level of methylation between 10% and 50% [222]. The majority of LMRs are lost or gained *de novo* during differentiation and are enriched for the motifs of TFs expressed in their respective cell types [221, 222]. Earlier studies had shown that the binding of individual TFs is directly affected by DNA methylation [110, 118, 223]. However, DNA methylation also changes chromatin organization and accessibility, which in turn may influence TF binding. Computational predictions suggest a significant effect of 5mC on DNA shape, which may impact TF access to target sequences [224]. To quantify this effect, high-throughput *in vitro* screens were developed [225, 226], finding that the affinity of 90% of the studied TFs containing a CpG in their motif was influenced by DNA methylation [226], but only few studies have gone on to demonstrate this *in vivo* [227, 228]. Nonetheless, it is also worth noting that the binding of CTCF and RE1-silencing transcription factor (REST) can induce the formation of LMRs [222]. Moreover, classic pioneer factors, such as FOXA2, also possess the ability to trigger targeted loss of DNA methylation [229]. Elevated levels of 5hmC have been found at LMRs, suggesting a role for active demethylation at LMRs [222].

DNA methylation valleys (DMVs)

Closer inspection of whole genome bisulfite sequencing (WGBS) data from mouse hematopoietic stem cells and human *in vitro* differentiation experiments revealed another characteristic feature: DMVs or canyons, long stretches of unmethylated DNA that often cover multiple CGIs [230, 231] (Figure 6B). These are generally conserved across tissues and species, are near developmental genes, and have a median length of several kilobases [230, 231] (Figure 6C). Follow-up work showed that DMV borders seem to be dynamically regulated in mouse ESCs by the dual targeting of DNMT3A and TET1 [232].

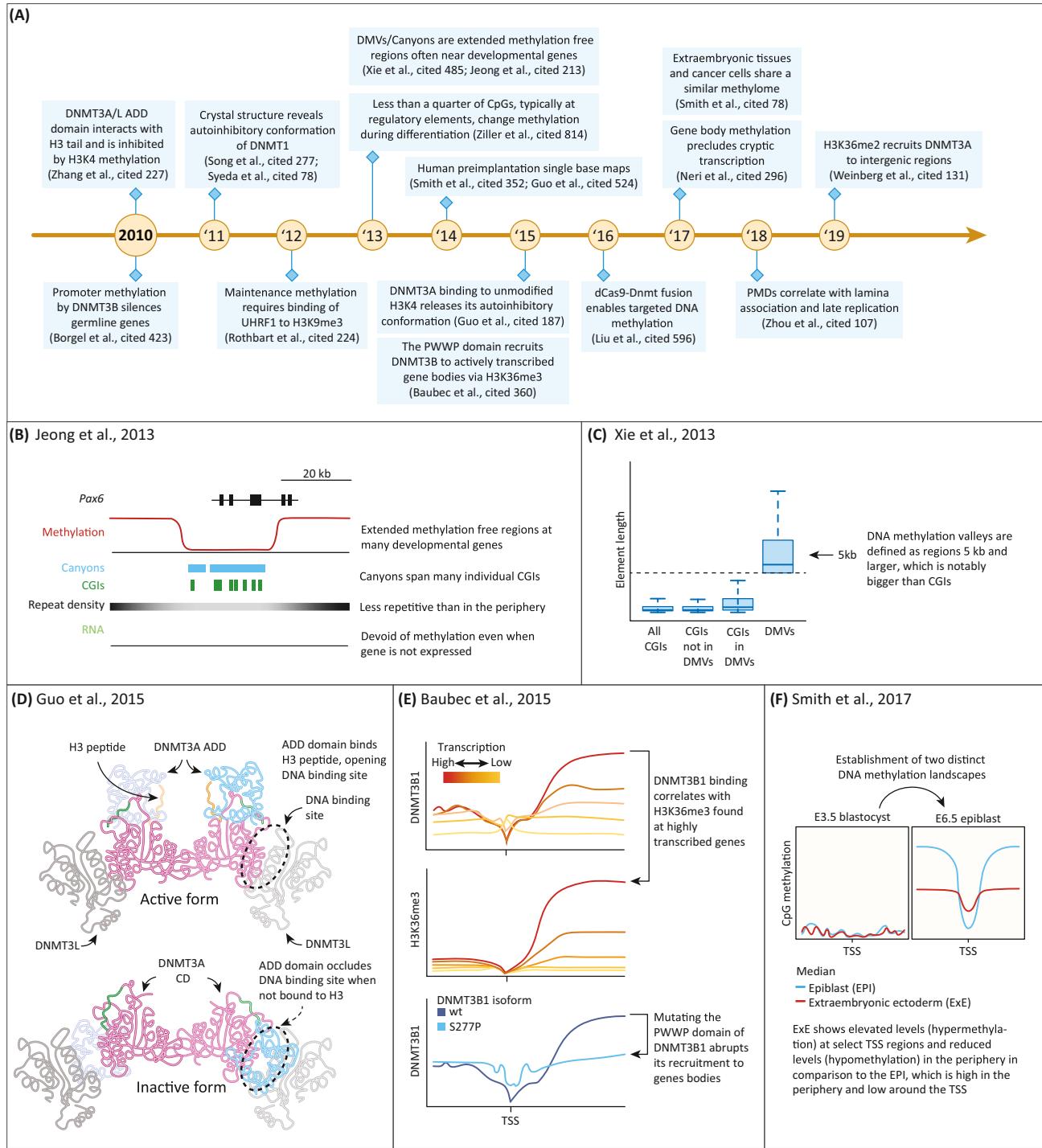


Figure 6. Recruitment mechanisms of the DNMTs and global methylation architecture. (A) Comprehensive mapping studies in the last decade enabled new insights into the recruitment and regulation of DNMTs and highlighted several novel elements of global methylation architecture. (B) Simplified schematic of a DNA methylation canyon using whole genome bisulfite sequencing (WGBS) data from mouse hematopoietic stem cells by Jeong and colleagues. (C) Interrogation of WGBS data showing DNA methylation valleys. (D) Molecular mechanism of DNMT3A recruitment. (E) Genomic distribution of DNMT3B1 and H3K36me3. (F) Establishment of DNA methylation landscapes in E3.5 blastocyst and E6.5 epiblast.

(Figure legend continued at the bottom of the next page.)

Allosteric regulation of the DNMTs

Several structural and biochemical studies showed that all of the DNMT3 enzymes bind to the H3 peptide via their highly conserved ADD domain, which is sensitive to the methylation status of H3K4 [233]. DNMT3A mutated to be insensitive to inhibition by H3K4me3 aberrantly methylated a subset of H3K4me3-marked promoters in mouse ESCs, leading to the downregulation of their associated genes [234]. The mechanism of how methylated H3K4 inhibits DNA methylation was revealed when Guo and colleagues resolved the crystal structure of DNMT3A in conformations with and without bound H3 peptide (Figure 6D). It was found that unbound DNMT3A adopts an autoinhibitory conformation where the ADD domain occludes the DNA binding site within the catalytic domain. When DNMT3A is bound to unmethylated H3, the ADD-catalytic domain interaction is disrupted and allows for DNA binding and its subsequent methylation [235]. This study proposed a model for DNMT regulation where the *de novo* methyltransferases adopt the autoinhibitory conformation until they are locally activated by binding H3 tails. Additional studies showed that this regulatory principle not only applies to the canonical *de novo* methyltransferases, but also to DNMT1. In this case, binding of the replication foci targeting (RFT) domain to allosteric activators such as UHRF1 is required to expose the catalytic site [236–238]. UHRF1, in turn, has been shown to adopt an autoinhibited conformation that can be relieved by binding to hemimethylated DNA and H3K9me3 [239,240].

Recruitment and targeting of *de novo* DNMTs

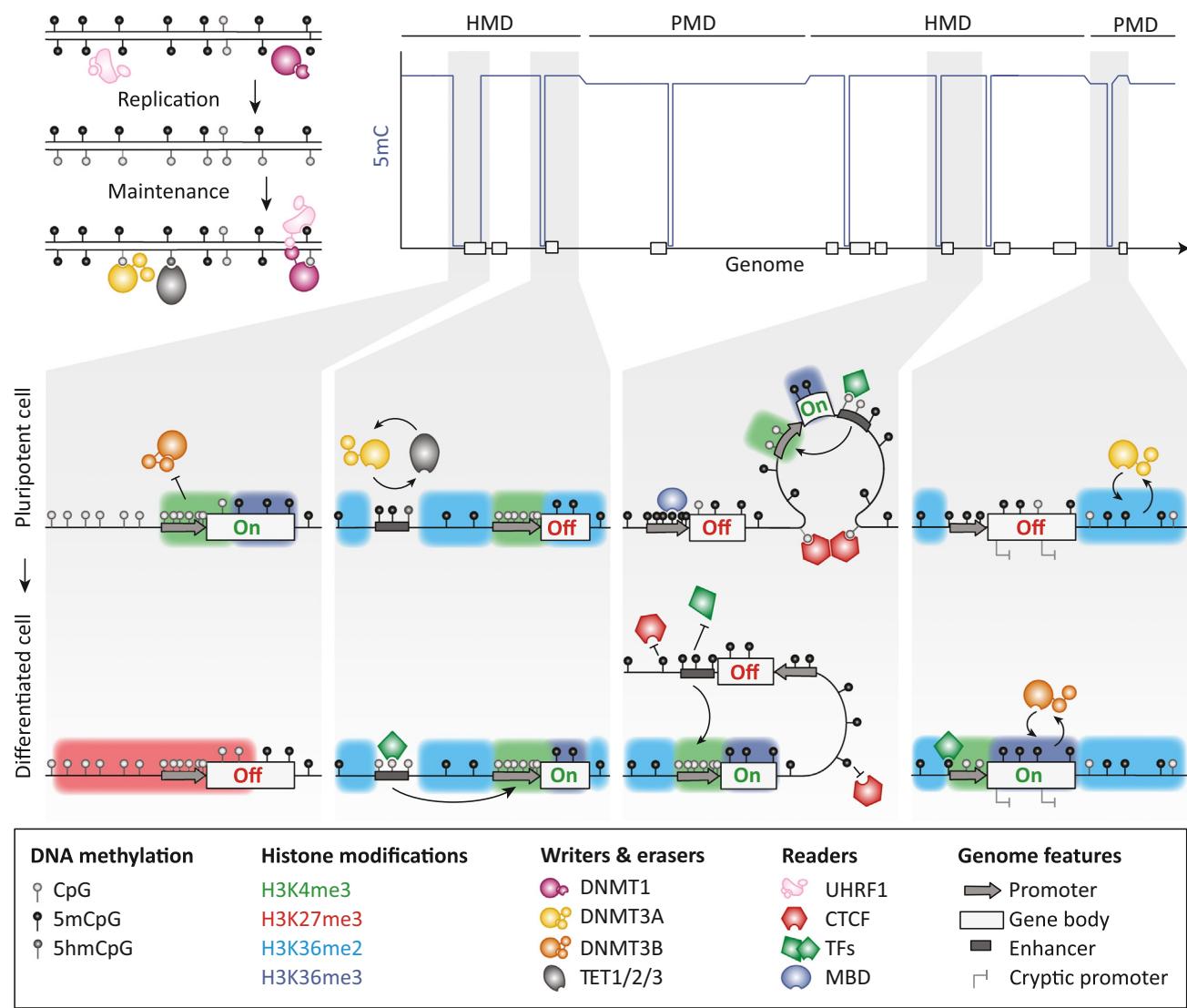
In contrast to the shielding effect of H3K4 methylation, several studies found that methylation of H3K36 recruits DNMTs. In 2010, the PWWP domain conserved in the *de novo* methyltransferases was found to interact with H3K36 di- and trimethylation *in vitro* and this interaction was required for the proper localization of DNMT3A in human cells [241]. A few years later, the Schübeler lab showed that while both DNMT3A and -B are generally recruited to CpG-dense regions, DNMT3B is specifically recruited to the bodies of actively transcribed genes through its PWWP domain interacting with H3K36me3 [242] (Figure 6E). Weinberg and colleagues then showed that DNMT3A is specifically recruited to intergenic regions by H3K36me2. Removing H3K36me2 from mouse ESCs redistributed DNMT3A binding to H3K36me3 marked regions [243]. Another study reported that H3K36 methyltransferases, NSD1 and SETD2 play key roles in shaping the DNA methylation landscape of male and female germ cells in mice [244,245]. Together, these studies showed that the *de novo* methyltransferases typically bind CpG-rich regions, unless protected by H3K4 methylation, while H3K36 methylation recruits *de novo* DNMTs to additional targets such as actively transcribed gene bodies. Combined with earlier work in plants, these observations strengthened the hypothesis that gene body methylation might serve to regulate the use of intergenic promoters [199]. In support of this, Neri *et al.* uncovered a significant increase of spurious transcripts originating from cryptic intragenic promoters in *Dnmt3b* KO and *Setd2* KO mouse ESCs [246].

With the advent of Cas9-based gene-editing tools [247,248], the ability of Cas9 to recognize specific genomic sequences was harnessed to build tools for the targeted recruitment of various

data of the human ESC line H1 by Xie *et al.* uncovered long stretches (>5 kb) of lowly methylated DNA, termed DNA methylation valleys (DMVs). Canyons and DMVs are describing the same type of DNA methylation architecture that is largely preserved across tissue types and species. (D) Crystal structures of DNMT3A (ADD and catalytic domain) and DNMT3L (catalytic domain-like) by Guo *et al.* demonstrating the autoinhibitory conformation the complex adopts when not bound to the unmodified H3 tail. (E) Baubec *et al.* overexpressed DNMT3B1 in mouse ES cells and determined its genome-wide binding using ChIP-seq. Average binding profiles centered around the transcription start site (TSS) and binned by transcriptional activity of the gene revealed an enrichment of DNMT3B1 at highly transcribed, H3K36me3-marked gene bodies (top and middle). This strong binding was not observed when the PWWP domain of DNMT3B1 was mutated (bottom). (F) Smith *et al.* profiled DNA methylation during early mouse embryonic development separately in the embryonic and extraembryonic tissues using WGBS. The levels of 5-methylcytosine (5mC) and the difference between the epiblast and the extraembryonic ectoderm are displayed for a subset of TSSs. While methylation is similar in both precursor lineages in the blastocyst, upon implantation the extraembryonic ectoderm displays hypermethylation at this subset of TSSs and hypomethylation of the flanking regions relative to the epiblast. See [210,230,231,233,235,237,238,240,242,243,246,250,255,256,259,260,266]. Abbreviations: CGIs, CpG islands; PMD, partially methylated domain.

Key figure

Key concepts of mammalian gene regulation by DNA methylation



Trends In Genetics

Figure 7. (Top left) 5-methylcytosine (5mC) occurs at symmetrical CpG dinucleotides. DNA replication produces two hemimethylated molecules. The hemimethylated DNA is recognized by UHRF1, which in turn recruits DNMT1 to add methylation onto the nascent strand. Proofreading by DNMT3A/B increases the fidelity of maintenance methylation by filling sites omitted by DNMT1 or counteracting active demethylation by the TETs. (Top right) In essentially all somatic cell types, the methylome adopts a bimodal distribution. About half of the genome is covered by megabase-size domains, termed partially methylated domains (PMDs), that exhibit slightly lower methylation and alternate with highly methylated domains (HMDs). This high level of methylation is punctuated by unmethylated DNA methylation valleys (DMVs) spanning several kilobases and shorter CpG-dense stretches called CpG islands (CGIs). (Bottom left) DMVs or canyons demarcate extended regions devoid of 5mC, often near developmental genes. When active, these promoters are marked by H3K4me3, which ensure DNMT3A and -B remain in their autoinhibitory conformation. DMVs remain free of methylation even when the associated genes are silenced by H3K27me3. (Bottom center left) A subset of enhancers in pluripotent cells is highly methylated but undergoes constant turnover of 5mC through the counteracting activities of DNMT3A/B and TETs. These enhancers can stay methylated or become hypomethylated upon differentiation. CpG-dense promoters are generally unmethylated, irrespective of transcriptional status. (Bottom center right) If

(Figure legend continued at the bottom of the next page.)

regulators [249]. The first generation of epi-editing tools included fusion proteins of catalytically inactive Cas9 (dCas9) and the catalytic domain of DNMT3A or TET1. These systems were designed to control transcription of reporter genes by methylating or demethylating their promoters [250]. Fusing multiple DNMT3A or TET1 units to dCas9 using the SunTag system further improved their efficiency [251,252]. However, as expected, extensive off-target effects were observed and further improvements to on-target specificity are needed to restrict catalytic activity to the desired targets only [253].

Escaping targeted DNA methylation

Genome-scale methylation maps of early development helped highlight other more specific features of the mammalian genome that are targeted for methylation during discrete windows of development. For example, many but not all repetitive elements escape the global erasure during preimplantation development [254–256]. Among the human long interspersed nuclear elements (LINEs), the younger L1PA subfamilies are the only LINEs that continue to be active and, together with L1HS, account for nearly all LINE transcription in human preimplantation embryos. These select LINE subfamilies were also found to be demethylated early in development, while the older subfamilies maintained higher methylation during this period. Notably, the demethylated and active subfamilies shared a deletion in their 5' UTRs compared with their nontransposing ancestors that appeared to co-occur with the evolutionary emergence of LINEs that evade a sequence-based repressive mechanism [255]. Shortly thereafter, it was revealed that this sequence contains the Krab zinc finger ZNF93 binding site that would normally have repressed these elements by binding and recruitment of TRIM28/KAP1 [257].

Global methylation alterations in development and disease

More comprehensive mapping of 5mC improved the description of the distinct global architecture of highly methylated domains (HMDs) and PMDs [208]. While PMDs can be detected in most healthy cell types, aged, cancerous, and extraembryonic tissues all display a further reduction in methylation levels within PMDs [258–261]. Zhou *et al.* suggested that the degree of PMD hypomethylation relates to the number of mitotic cell divisions. Incomplete methylation in these late replicating regions may result in a continued loss of DNA methylation during an organism's life, a process that may be accelerated during tumor development and progression [259].

As noted, the early extraembryonic lineages and placenta, as well as somatic cancers, share the reduction in PMD methylation, but also exhibit specific gain of methylation over CpG-rich domains that overlap with H3K27me3 repressed developmental genes [260,262–265] (Figure 6F), which often lie within DMVs. Once the DNA methylation landscape is set up in the early epiblast, methylation appears static at most features, with the exception of the subset of CpGs typically found outside of promoter regions that shows dynamic methylation changes during differentiation [206,208,222,266].

Summary

The use of sequencing technology to map methylomes of hundreds of cell types and developmental stages from different organisms has refined insights from the prior decades, closed some important knowledge gaps, and provided a comprehensive overview of the distribution

present in sufficient density, 5mC silences associated promoters and is bound by methyl-CpG binding domain proteins (MBDs). Some transcription factors (TFs) are repelled by 5mC at their motifs. Promoter–enhancer contacts involve binding of methylation-sensitive CTCF. (Bottom right) Intergenic regions are marked by H3K36me2, attracting DNMT3A. Transcribed genes are marked by H3K36me3, recruiting DNMT3B to methylate the gene body. Gene body methylation does not impede transcriptional elongation but suppresses cryptic promoter activity. Binding of some TFs may require 5mC at their motifs.

and dynamics of 5mC. Combined with structural studies, this led to a detailed understanding of the recruitment and activity of DNMTs as well as the impact of DNA methylation on gene regulation. How some of that is mechanistically translated and what role it plays in other parts of the genome remains to be explored.

Concluding remarks

In this review, we have provided a chronological overview of the field of DNA methylation from its start to the present day. This historical perspective highlights key experiments that document the innovation and tremendous progress the field has made. It also serves to contextualize former hypotheses that have withstood or been challenged over time. Many of the core concepts of the field were established by the 1990s, however, recent work continues to provide insights, including unexpected discoveries that add complexity and nuances to these well-established views (Figure 7, Key figure). For example, another mammalian methyltransferase, *Dnmt3c*, was discovered in male germ cells in 2016 [267] and several papers have reported functional *de novo* methylation capabilities of the canonical maintenance methyltransferase DNMT1 [268–271]. Another example is the discovery of active turnover of 5mC at highly methylated somatic enhancers in pluripotent stem cells, a finding that changes the assumption that DNA methylation is, once established, a stably propagated modification that remains static over successive cell divisions [2–4]. In addition, several new regulators, including QSER1 and BANP, have been reported to play a role in shaping mammalian methylation landscapes and their interpretation [272,273]. In addition, two recent publications provide evidence that DNMT1 may be able to directly interact with H3K9me3 [274] as well as H4K20me3, a modification frequently co-occurring with H3K9me3 in ESCs [275]. Furthermore, DNMT3A was found to be recruited to regions marked by ubiquitinated H2AK119 placed by the Polycomb repressive complex 1, an interaction usually masked by the stronger recruitment to H3K36me2 [276]. As we are approaching the end of a century since the discovery of 5mC in bacteria [24], the field has matured and DNA methylation is now frequently used as a tool to study other biological processes and disease phenotypes. Nonetheless, questions remain that need to be addressed in the coming years before we can claim to have a complete understanding of this small but essential and impactful chemical modification to DNA (see Outstanding questions).

Acknowledgments

We thank S. Hetzel and H. Kretzmer as well as all members of the Meissner Lab for helpful comments and discussions about the figures and text. We thank S. Pradhan for helpful information about the history of Mspl. We also thank the anonymous reviewers for their helpful feedback and perspective. A.L.M is supported by the National Science Foundation Graduate Research Fellowship and the Frederick Sheldon Traveling Fellowship. N.B. is supported by the doctoral scholarship of the German Academic Scholarship Foundation and the Max Planck Society. A.M is supported by the Max Planck Society.

Declaration of interests

No interests are declared.

References

- Jaenisch, R. (1997) DNA methylation and imprinting: why bother? *Trends Genet.* 13, 323–329
- Ginno, P.A. *et al.* (2020) A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat. Commun.* 11, 2680
- Charlton, J. *et al.* (2020) TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat. Genet.* 52, 819–827
- Spada, F. *et al.* (2020) Active turnover of genomic methylcytosine in pluripotent cells. *Nat. Chem. Biol.* 16, 1411–1419
- Jackson-Grusby, L. *et al.* (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* 27, 31–39
- Chen, T. *et al.* (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell. Biol.* 23, 5594–5605
- Tsumura, A. *et al.* (2006) Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* 11, 805–814
- Baylin, S.B. and Jones, P.A. (2016) Epigenetic determinants of cancer. *Cold Spring Harb. Perspect. Biol.* 8, a019505
- Jones, P.A. and Laird, P.W. (1999) Cancer epigenetics comes of age. *Nat. Genet.* 21, 163–167
- Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21

Outstanding questions

The past decade has highlighted that DNA methylation is much more dynamic than classically portrayed, but how elastic is DNA methylation really and why?

Most promoters are not regulated by DNA methylation and, while it correlates with active enhancers, its functional role there is not well understood. We therefore still need to define what its precise function in cell and lineage regulation is.

What role does the intergenic recruitment (through H3K36me2) of DNA methylation play?

Does methylation have a function in parts of the genome that have been ignored so far and become mappable in new T2T assemblies?

How are ubiquitous enzymes like the DNMTs and TETs targeted or recruited to their specific substrates and how does their regulation and activity change during differentiation? Is there a biological role for the oxidation products of TETs?

What are the mechanistic underpinnings of aging clocks? What are the changes and why do they occur at defined rates over time?

Given the mutational burden of 5mC and the gradual depletion of CpGs from the genome, will vertebrate genomes have to evolve a new mechanism for genome regulation in the future?

Why do virtually all cancer cells converge on a common DNA methylation pattern and how does this pattern affect development and progression of disease? What is the function and disease relevance of the intermediate CGI hypermethylation and genome-wide hypomethylation typical of cancer (and aging)?

What allows pluripotent cells to proliferate in the absence of DNA methylation and, in contrast, what function underlies the somatic cell requirement for DNA methylation?

Related to that, why do DNMT deficient embryos die? This is a question that has remained unanswered since the initial *Dnmt1* disruption in 1992.

11. Suzuki, M.M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nat. Rev. Genet.* 9, 465–476
12. Jones, P.A. and Liang, G. (2009) Rethinking how DNA methylation patterns are maintained. *Nat. Rev. Genet.* 10, 805–811
13. Law, J.A. and Jacobsen, S.E. (2010) Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220
14. Smith, Z.D. and Meissner, A. (2013) DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220
15. Du, J. *et al.* (2015) DNA methylation pathways and their crosstalk with histone methylation. *Nat. Rev. Mol. Cell Biol.* 16, 519–532
16. Lyko, F. (2018) The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat. Rev. Genet.* 19, 81–92
17. Greenberg, M.V.C. and Bourc'his, D. (2019) The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* 20, 590–607
18. Parry, A. *et al.* (2021) Active turnover of DNA methylation during cell fate decisions. *Nat. Rev. Genet.* 22, 59–66
19. Sutton, W.S. (1902) On the morphology of the chromosome group in *Brachystola magna*. *Biol. Bull.* 4, 24–39
20. Sutton, W.S. (1903) The chromosomes in heredity. *Biol. Bull.* 4, 231–251
21. Morgan, T.H. (1910) Sex limited inheritance in *Drosophila*. *Science* 32, 120–122
22. Levene, P.A. and Jacobs, W.A. (1909) Über die hefe-nucleinsäure. *Ber. Dtsch. Chem. Ges.* 42, 2474–2478
23. Wheeler, H.L. (1904) Researches on pyrimidine derivatives. *J. Am. Chem. Soc.* 31, 591–606
24. Johnson, T.B. and Coghill, R.D. (1925) Researches on pyrimidines. C111. The discovery of 5-methyl-cytosine in tuberculinic acid, the nucleic acid of the tubercle bacillus. *J. Am. Chem. Soc.* 47, 2838–2844
25. Vischer, E. and Chargaff, E. (1947) The separation and characterization of purines in minute amounts of nucleic acid hydrolysates. *J. Biol. Chem.* 175, 168, 781
26. Hotchkiss, R.D. (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* 175, 315–332
27. Wyatt, G.R. (1950) Occurrence of 5-methyl-cytosine in nucleic acids. *Nature* 166, 237–238
28. Wyatt, G.R. (1951) Recognition and estimation of 5-methylcytosine in nucleic acids. *Biochem. J.* 48, 581–584
29. Avery, O.T. *et al.* (1914) Studies on the chemical nature of the substance inducing transformation of pneumococcal types. *J. Exp. Med.* 79, 137–158
30. Hershey, B.A.D. and Chase, M. (1952) Independent functions of viral protein and nucleic acid in growth of bacteriophage. *J. Gen. Physiol.* 36, 39–56
31. Watson, J.D. and Crick, F.H.C. (1953) Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* 171, 737–738
32. Smith, J.D. and Markham, R. (1952) Polynucleotides from deoxyribonucleic acids. *Nature* 170, 120–121
33. Sinsheimer, R.L. *et al.* (1954) The action of pancreatic desoxyribonuclease. I. Isolation of mono- and dinucleotides. *J. Biol. Chem.* 208, 445–459
34. Griffith, F. (1928) The significance of pneumococcal types. *J. Hyg.* 27, 113–159
35. Vischer, Ernst *et al.* (1949) Microbial nucleic acids: the desoxypentose nucleic acids of avian tubercle bacilli and yeast. *J. Biol. Chem.* 177, 429–438
36. Waddington, C.H. (1942) The epigenotype. *Endeavour* 1, 18–20
37. Waddington, C.H. (1957) *The Strategy of the Genes; A Discussion of Some Aspects of Theoretical Biology*; Allen & Unwin
38. Borek, E. and Srinivasan, P.R. (1966) The methylation of nucleic acids. *Annu. Rev. Biochem.* 35, 275–298
39. Luria, S.E. (1945) Mutations of bacterial viruses affecting their host range. *Genetics* 30, 84–99
40. Bertani, G. and Weigle, J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.* 65, 113–121
41. Lederberg, S. (1957) Suppression of the multiplication of heterologous bacteriophages in lysogenic bacteria. *Virology* 3, 496–513
42. Gold, M. *et al.* (1963) The enzymatic methylation of RNA and DNA. II. On the species specificity of the methylation enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 50, 164–169
43. Arber, W. (1965) Host-controlled modification of bacteriophage. *Annu. Rev. Microbiol.* 19, 365–378
44. Billen, D. and Hewitt, R. (1966) Influence of starvation for methionine and other amino acids on subsequent bacterial deoxyribonucleic acid replication. *J. Bacteriol.* 92, 609–617
45. Billen, D. (1968) Methylation of the bacterial chromosome: an event at the "replication point"? *J. Mol. Biol.* 31, 477–486
46. Lark, C. (1968) Studies on the *in vivo* methylation of DNA in *Escherichia coli* 15T. *J. Mol. Biol.* 31, 389–399
47. Lark, C. and Arber, W. (1970) Host specificity of DNA produced by *Escherichia coli*. *J. Mol. Biol.* 52, 337–348
48. Srinivasan, P.R. and Borek, E. (1964) Enzymatic alteration of nucleic acid structure: enzymes put finishing touches characteristic of each species on RNA and DNA by insertion of methyl groups. *Science* 145, 548–553
49. Sheid, B. *et al.* (1968) Deoxyribonucleic acid methylase of mammalian tissues. *Biochemistry* 7, 280–285
50. Vanyushin, B.F. *et al.* (1970) Rare bases in animal DNA. *Nature* 225, 948–949
51. Vanyushin, B.F. *et al.* (1973) The content of 5-methylcytosine in animal DNA: the species and tissue specificity. *Biochim. Biophys. Acta* 299, 397–403
52. Guseinov, V.A. and Vanyushin, B.F. (1975) Content and localisation of 5-methylcytosine in DNA of healthy and wilt-infected cotton plants. *Biochim. Biophys. Acta (BBA) - Nucleic Acids and Protein Synth.* 395, 229–238
53. Adams, R.L.P. (1971) The relationship between synthesis and methylation of DNA in mouse fibroblasts. *Biochim. Biophys. Acta* 254, 205–212
54. Adams, R.L.P. (1973) Delayed methylation of DNA in developing sea urchin embryos. *Nat. New Biol.* 244, 27–29
55. Scarano, E. (1971) The control of gene function in cell differentiation and in embryogenesis. *Adv. Cytopharmacol.* 1, 13–24
56. Comings, D.E. (1972) Methylation of euchromatic and heterochromatic DNA. *Exp. Cell Res.* 74, 383–390
57. Holliday, R. and Pugh, J.E. (1975) DNA modification mechanisms and gene activity during development. *Science* 187, 226–232
58. Riggs, A.D. (1975) X inactivation, differentiation, and DNA methylation. *Cytogenet. Cell Genet.* 14, 9–25
59. Sager, R. and Kitchin, R. (1975) Selective silencing of eukaryotic DNA. *Science* 189, 426–433
60. Gurdon, J.B. and Uehlinger, V. (1966) "Fertile" intestine nuclei. *Nature* 210, 1240–1241
61. Linn, S. and Arber, W. (1968) Host specificity of DNA produced by *Escherichia coli*. X. In vitro restriction of phage fd replicative form. *Proc. Natl. Acad. Sci. U. S. A.* 59, 1300–1306
62. Meselson, M. and Yuan, R. (1968) DNA restriction enzyme from *E. coli*. *Nature* 217, 1110–1114
63. Smith, H.O. and Wilcox, K.W. (1970) A restriction enzyme from *Haemophilus influenzae*: I. Purification and general properties. *J. Mol. Biol.* 51, 379–391
64. Singer, J. *et al.* (1979) Methylation of mouse liver DNA studied by means of the restriction enzymes Msp I and Hpa II. *Science* 203, 1019–1021
65. Waalwijk, C. and Flavell, R.A. (1978) MspI, an isoschizomer of *hpa*I which cleaves both unmethylated and methylated *hpa*I sites. *Nucleic Acids Res.* 5, 3231–3236
66. Southern, E.M. (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98, 503–511
67. Waalwijk, C. and Flavell, R.A. (1978) DNA methylation at a CGGG sequence in the large intron of the rabbit β -globin gene: tissue-specific variations. *Nucleic Acids Res.* 5, 4631–4641
68. Mandel, J.L. and Chambon, P. (1979) DNA methylation: organ specific variations in the methylation pattern within and around ovalbumin and other chicken genes. *Nucleic Acids Res.* 7, 2081–2103
69. Kuo, M.T. *et al.* (1979) DNA methylation: correlation with DNase I sensitivity of chicken ovalbumin and conalbumin chromatin. *Nucleic Acids Res.* 7, 2105–2113

70. McGhee, J.D. and Ginder, G.D. (1979) Specific DNA methylation sites in the vicinity of the chicken β -globin genes. *Nature* 280, 419–420
71. Doskocil, J. and Sormova, Z. (1965) The occurrence of 5-methylcytosine in bacterial deoxyribonucleic acids. *Biochim. Biophys. Acta* 15, 513–515
72. Doskocil, J. et al. (1967) Inhibition of protein synthesis by 5-azacytidine in *Escherichia coli*. *Biochim. Biophys. Acta* 145, 771–779
73. Levitan, I.B. and Webb, T.E. (1969) Effects of 5-azacytidine on polyribosomes and on the control of tyrosine transaminase activity in rat liver. *Biochim. Biophys. Acta* 182, 491–500
74. Fukic, V. et al. (1970) On the induction of segment extension and chromatin structural changes in *Vicia faba* chromosomes after treatment with 5-azacytidine and 5-azadeoxycytidine. *Mutat. Res.* 9, 599–606
75. Pačes, V. et al. (1968) Incorporation of 5-azacytidine into nucleic acids of *Escherichia coli*. *Biochim. Biophys. Acta* 161, 352–360
76. Cihák, A. (1974) Biological effects of 5-azacytidine in eukaryotes. *Oncology* 3, 405–422
77. von Hoff, D.D. et al. (1976) 5-Azacytidine a new anticancer drug with effectiveness in acute myelogenous leukemia. *Ann. Intern. Med.* 85, 237–245
78. Benedict, W.F. et al. (1977) Induction of morphological transformation in mouse C3H/10T1/2 clone 8 cells and chromosomal damage in hamster A(T1)C1-3 cells by cancer chemotherapeutic agents 1. *Cancer Res.* 37, 2202–2208
79. Constantinides, P.G. et al. (1977) Functional striated muscle cells from non-myoblast precursors following 5-azacytidine treatment. *Nature* 267, 364–366
80. Taylor, S.M. and Jones, P.A. (1979) Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. *Cell* 17, 771–779
81. Friedman, S. (1979) The effect of 5-azacytidine on *E. coli* DNA methylase. *Biochem. Biophys. Res. Commun.* 89, 1328–1333
82. Jones, P.A. and Taylor, S.M. (1980) Cellular differentiation, cytidine analogs and DNA methylation. *Cell* 20, 85–93
83. Pollock, J.M. et al. (1978) Methylation of DNA in early development: 5-methyl cytosine content of DNA in sea urchin sperm and embryos. *Nucleic Acids Res.* 5, 4855–4864
84. Diacumakos, E.G. et al. (1970) A microsurgical methodology for human cells in vitro: evolution and applications. *Proc. Natl. Acad. Sci. U. S. A.* 65, 911–918
85. Jackson, D.A. et al. (1972) Biochemical method for inserting new genetic information into DNA of simian virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 69, 2904–2909
86. Cohen, S.N. et al. (1973) Construction of biologically functional bacterial plasmids in vitro. *Proc. Natl. Acad. Sci. U. S. A.* 70, 3240–3244
87. Bolivar, F. et al. (1977) Construction and characterization of new cloning vehicle. II. A multipurpose cloning system. *Gene* 2, 95–113
88. Yanisch-Perron, C. et al. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mpl8 and pUC19 vectors. *Gene* 33, 103–119
89. Jaenisch, R. and Mintz, B. (1974) Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proc. Natl. Acad. Sci. U. S. A.* 71, 1250–1254
90. Stacey, D.W. and Alfrey, V.G. (1976) Microinjection studies of duck globin messenger RNA translation in human and avian cells. *Cell* 9, 725–732
91. Gordon, J.W. et al. (1980) Genetic transformation of mouse embryos by microinjection of purified DNA. *Genetics* 77, 7380–7384
92. Gordon, J. and Ruddle, F. (1981) Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* 214, 1244–1246
93. Thomas, K.R. and Capecchi, M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51, 503–512
94. Bird, A.P. et al. (1979) Methylated and unmethylated DNA compartments in the sea urchin genome. *Cell* 11, 647–658
95. Bird, A.P. and Taggart, M.H. (1980) Variable patterns of total DNA and rDNA methylation in animals. *Nucleic Acids Res.* 8, 1485–1497
96. van der Ploeg, L.H.T. and Flavell, R.A. (1980) DNA methylation in the human $\gamma\delta$ -globin locus in erythroid and nonerythroid tissues. *Cell* 19, 947–958
97. Wilks, A.F. et al. (1982) Estrogen induces a demethylation at the 5' end region of the chicken vitellogenin gene. *Proc. Natl. Acad. Sci. U. S. A.* 79, 4252–4255
98. Ott, M.-O. et al. (1982) Undermethylation at the 5' end of the albumin gene is necessary but not sufficient for albumin production by rat hepatoma cells in culture. *Cell* 30, 833
99. Cate, R.L. et al. (1983) Comparison of the methylation patterns of the two rat insulin genes. *J. Biol. Chem.* 258, 6645–6652
100. Vedel, M. et al. (1983) Changes in methylation pattern of albumin and alpha-fetoprotein genes in developing rat liver and neoplasia. *Nucleic Acids Res.* 11, 4335–4354
101. Gjerset, R.A. and Martin, D.W. (1982) Presence of a DNA demethylating activity in the nucleus of murine erythroleukemic cells. *J. Biol. Chem.* 257, 8581–8583
102. Stuhlmann, H. et al. (1981) Infectivity and methylation of retroviral genomes is correlated with expression in the animal. *Cell* 26, 221–232
103. Harbers, K. et al. (1981) DNA methylation and gene expression: endogenous retroviral genome becomes infectious after molecular cloning. *Proc. Natl. Acad. Sci. USA* 78, 7609–7613
104. Jähner, D. et al. (1982) De novo methylation and expression of retroviral genomes during mouse embryogenesis. *Nature* 298, 623–628
105. Fradin, A. et al. (1982) Methylation of simian virus 40 Hpa II site affects late, but not early, viral gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 79, 5142–5146
106. Busslinger, M. et al. (1983) DNA methylation and the regulation of globin gene expression. *Prog. Clin. Biol. Res.* 34, 193–203
107. Kruczak, I. and Doerfler, W. (1983) Expression of the chloramphenicol acetyltransferase gene in mammalian cells under the control of adenovirus type 12 promoters: effect of promoter methylation on gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 80, 7586–7590
108. Vardimon, L. et al. (1982) Expression of a cloned adenovirus gene is inhibited by *in vitro* methylation. *Proc. Natl. Acad. Sci. U. S. A.* 79, 1073–1077
109. Langner, K.-D. et al. (1984) DNA methylation of three 5' C-C-G-G 3' sites in the promoter and 5' region inactivate the E2a gene of adenovirus type 2. *Proc. Natl. Acad. Sci. U. S. A.* 81, 2950–2954
110. Watt, F. and Molloy, P.L. (1988) Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter. *Genes Dev.* 2, 1136–1143
111. Gautsch, J.W. and Wilson, M.C. (1983) Delayed de novo methylation in teratocarcinoma suggests additional tissue-specific mechanisms for controlling gene expression. *Nature* 301, 32–37
112. Niwa, O. et al. (1983) Independent mechanisms involved in suppression of the Moloney leukemia virus genome during differentiation of murine teratocarcinoma cells. *Cell* 32, 1105–1113
113. Sobieski, D.A. and Eden, F.C. (1981) Clustering and methylation of repeated DNA: persistence in avian development and evolution. *Nucleic Acids Res.* 9, 6001–6015
114. Eden, F.C. et al. (1981) Clusters of repeated sequences of chicken DNA are extensively methylated but contain specific undermethylated regions. *J. Mol. Biol.* 148, 129–151
115. Cooper, D.N. et al. (1983) Unmethylated domains in vertebrate DNA. *Nucleic Acids Res.* 11, 647–658
116. Bird, A. et al. (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40, 91–99
117. Bird, A.P. (1986) CpG-rich islands and the function of DNA methylation. *Nature* 321, 209–213
118. Becker, P.B. et al. (1987) Genomic footprinting reveals cell type-specific DNA binding of ubiquitous factors. *Cell* 51, 435–443
119. Meehan, R.R. et al. (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 58, 499–507

120. Monk, M. et al. (1987) Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* 99, 371–382
121. Monk, M. (1987) Genomic imprinting. Memories of mother and father. *Nature* 328, 203–204
122. Lauster, R. et al. (1989) Cytosine-specific type II DNA methyltransferases: a conserved enzyme core with variable target-recognizing domains. *J. Mol. Biol.* 206, 305–312
123. Pósfai, J. et al. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* 17, 2421–2435
124. Roberts, R.J. et al. (1976) A specific endonuclease from *Haemophilus haemolyticus*. *J. Mol. Biol.* 103, 199–208
125. Smith, H. (1979) Nucleotide sequence specificity of restriction endonucleases. *Science* 205, 455–462
126. Wu, J.C. and Santi, D.V. (1987) Kinetic and catalytic mechanism of Hhal methyltransferase. *J. Biol. Chem.* 262, 4778–4786
127. Jones, P.A. and Taylor, S.M. (1981) Hemimethylated duplex DNAs prepared from 5-azacytidine-treated cells. *Nucleic Acids Res.* 9, 2933–2947
128. Stein, R. et al. (1982) Clonal inheritance of the pattern of DNA methylation in mouse cells. *Proc. Natl. Acad. Sci. U. S. A.* 79, 61–65
129. Gruenbaum, Y. et al. (1982) Substrate and sequence specificity of a eukaryotic DNA methylase. *Nature* 295, 620–622
130. Bestor, T.H. and Ingram, V.M. (1983) Two DNA methyltransferases from murine erythroleukemia cells: purification, sequence specificity, and mode of interaction with DNA. *Proc. Natl. Acad. Sci. U. S. A.* 80, 5559–5563
131. Bestor, T. et al. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* 203, 971–983
132. Galupa, R. and Heard, E. (2018) Annual review of genetics X-chromosome inactivation: a crossroads between chromosome architecture and gene regulation. 52, 535–566
133. Lee, J.T. and Bartolomei, M.S. (2013) X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* 152, 1308–1323
134. Ferguson-Smith, A.C. and Bourcouis, D. (2018) The discovery and importance of genomic imprinting. *eLife* 7, e42368
135. Antequera, F. et al. (1990) High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines. *Cell* 62, 503–514
136. Li, E. et al. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* 69, 915–926
137. Li, E. et al. (1993) Role for DNA methylation in genomic imprinting. *Nature* 366, 362–365
138. Lei, H. et al. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* 122, 3195–3205
139. Tucker, K.L. et al. (1996) Complementation of methylation deficiency in embryonic stem cells by a DNA methyltransferase minigene. *Proc. Natl. Acad. Sci. U. S. A.* 93, 12920–12925
140. Tucker, K.L. et al. (1996) Germ-line passage is required for establishment of methylation and expression patterns of imprinted but not of nonimprinted genes. *Genes Dev.* 10, 1008–1020
141. Cheng, X. et al. (1993) Crystal structure of the Hhal DNA methyltransferase complexed with S-adenosyl-L-methionine. *Cell* 74, 299–307
142. Santi, D.V. et al. (1984) Covalent bond formation, between a DNA-cytosine methyltransferase and DNA containing 5-azacytosine (5-methylcytosine/5-azacytidine/Hpa II methylase). *Proc. Natl. Acad. Sci. U. S. A.* 81, 6993–6997
143. Osterman, D.G. et al. (1988) 5-Fluorocytosine in DNA is a mechanism-based inhibitor of Hhal methylase. *Biochemistry* 27, 5204–5210
144. Klimasauskas, S. et al. (1994) Hhal methyltransferase flips its target base out of the DNA helix. *Cell* 76, 357–369
145. Balganesh, T.S. et al. (1987) Construction and use of chimeric SPR/phi 3T DNA methyltransferases in the definition of sequence recognizing enzyme regions. *EMBO J.* 6, 3543–3549
146. Wilke, K. et al. (1988) Sequential order of target-recognizing domains in multispecific DNA-methyltransferases. *EMBO J.* 7, 2601–2609
147. Trautner, T.A. et al. (1988) Chimeric multispecific DNA methyltransferases with novel combinations of target recognition. *Nucleic Acids Res.* 16, 6649–6658
148. Altschul, S.F. et al. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410
149. Yoder, J.A. and Bestor, T.H. (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum. Mol. Genet.* 7, 279–284
150. Okano, M. et al. (1998) Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* 26, 2536–2540
151. Okano, M. et al. (1998) Cloning and characterization of a family of novel mammalian DNA (cytosine-5) methyltransferases. *Nat. Genet.* 19, 219–220
152. Okano, M. et al. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 99, 247–257
153. Boyes, J. and Bird, A. (1991) DNA methylation inhibits transcription indirectly via a methyl-CpG binding protein. *Cell* 64, 1123–1134
154. Lewis, J.D. et al. (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. *Cell* 69, 905–914
155. Meehan, R.R. et al. (1992) Characterization of MeCP2, a vertebrate DNA binding protein with affinity for methylated DNA. *Nucleic Acids Res.* 20, 5085–5092
156. Tate, P. et al. (1996) The methyl-CpG binding protein MeCP2 is essential for embryonic development in the mouse. *Nat. Genet.* 12, 205–208
157. Nan, X. et al. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* 393, 386–389
158. Taunton, J. et al. (1996) A mammalian histone deacetylase related to the yeast transcriptional regulator rpd3p. *Science* 272, 408–411
159. Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol. Cell. Biol.* 18, 6538–6547
160. Ng, H.-H. et al. (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nat. Genet.* 23, 58–61
161. Gruenbaum, Y. et al. (1981) Sequence specificity of methylation in higher plant DNA. *Nature* 292, 860–862
162. Leutwiler, L.S. et al. (1984) The DNA of *Arabidopsis thaliana*. *Mol. Gen. Genet. MGG* 194, 15–23
163. Lloyd, A.M. et al. (1986) Transformation of *Arabidopsis thaliana* with *Agrobacterium tumefaciens*. *Science* 234, 464–466
164. Feldmann, K.A. and David Marks, M. (1987) Agrobacterium-mediated transformation of germinating seeds of *Arabidopsis thaliana*: a non-tissue culture approach. *Mol. Gen. Genet. MGG* 208, 1–9
165. Valvekens, D. et al. (1988) Agrobacterium tumefaciens-mediated transformation of *Arabidopsis thaliana* root explants by using kanamycin selection. *Proc. Natl. Acad. Sci. U. S. A.* 85, 5536–5540
166. Feldmann, K.A. et al. (1989) A dwarf mutant of *Arabidopsis* generated by T-DNA insertion mutagenesis. *Science* 243, 1351–1354
167. Finnegan, E.J. and Dennis, E.S. (1993) Isolation and identification by sequence homology of a putative cytosine methyltransferase from *Arabidopsis thaliana*. *Nucleic Acids Res.* 21, 2383–2388
168. Vongs, A. et al. (1993) *Arabidopsis thaliana* DNA methylation mutants. *Science* 260, 1926–1928
169. Ronemus, M.J. et al. (1996) Demethylation-induced developmental pleiotropy in *Arabidopsis*. *Science* 273, 654–657
170. Burn, J.E. et al. (1993) DNA methylation, vernalization, and the initiation of flowering. *Proc. Natl. Acad. Sci. U. S. A.* 90, 287–291
171. Kakutani, T. et al. (1995) Characterization of an *Arabidopsis thaliana* DNA hypomethylation mutant. *Nucleic Acids Res.* 23, 130–137

172. Finnegan, E.J. *et al.* (1996) Reduced DNA methylation in *Arabidopsis thaliana* results in abnormal plant development. *Proc. Natl. Acad. Sci. U. S. A.* 93, 8449–8454
173. Pradhan, S. (1998) Isolation, characterization and baculovirus-mediated expression of the cDNA encoding cytosine DNA methyltransferase from *Pisum sativum*. *Nucleic Acids Res.* 26, 1214–1222
174. Jacobsen, S.E. and Meyerowitz, E.M. (1997) Hypermethylated SUPERMAN epigenetic alleles in *Arabidopsis*. *Science* 277, 1100–1103
175. Henikoff, S. and Comai, L. (1998) A DNA methyltransferase homolog with a chromodomain exists in multiple polymorphic forms in *Arabidopsis*. *Genetics* 149, 307–318
176. Genger, R.K. *et al.* (1999) Multiple DNA methyltransferase genes in *Arabidopsis thaliana*. *Plant Mol. Biol.* 41, 269–278
177. Jeddeloh, J.A. *et al.* (1999) Maintenance of genomic methylation requires a SWI2/SNF2-like protein. *Nat. Genet.* 22, 94–97
178. Cao, X. *et al.* (2000) Conserved plant genes with similarity to mammalian de novo DNA methyltransferases. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4979–4984
179. Yoder, J.A. *et al.* (1997) Cytosine methylation and the ecology of intragenomic parasites. *Trends Genet.* 13, 335–340
180. Walsh, C.P. *et al.* (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* 20, 2
181. Bird, A. (1997) Does DNA methylation control transposition of selfish elements in the germline? *Trends Genet.* 13, 469–470
182. The Arabidopsis Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
183. Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–861
184. Abril, J.F. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562
185. Frommer, M. *et al.* (1992) A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. U. S. A.* 89, 1827–1831
186. Bell, A.C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* 405, 482–485
187. Hark, A.T. *et al.* (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* 405, 486–489
188. Tamaru, H. and Selker, E.U. (2001) A histone H3 methyltransferase controls DNA methylation in *Neurospora crassa*. *Nature* 414, 277–283
189. Lindroth, A.M. *et al.* (2001) Requirement of CHROMOMETHYLASE3 for maintenance of CpXpG methylation. *Science* 292, 2077–2080
190. Jackson, J.P. *et al.* (2002) Control of CpNpG DNA methylation by the KRYPTONITE histone H3 methyltransferase. *Nature* 416, 556–560
191. Apola, U. *et al.* (2000) Isolation and initial characterization of a novel zinc finger gene, DNMT3L, on 21q22.3, related to the cytosine-5-methyltransferase 3 gene family. *Genomics* 65, 293–298
192. Bourc'his, D. *et al.* (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science* 294, 2536–2539
193. Hata, K. *et al.* (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* 129, 1983–1993
194. Kaneda, M. *et al.* (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* 429, 900–903
195. Ooi, S.K.T. *et al.* (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* 448, 714–717
196. Jia, D. *et al.* (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* 449, 248–251
197. Bostick, M. *et al.* (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317, 1760–1764
198. Sharif, J. *et al.* (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450, 908–912
199. Tran, R.K. *et al.* (2005) DNA methylation profiling identifies CG methylation clusters in *Arabidopsis* genes. *Curr. Biol.* 15, 154–159
200. Zhang, X. *et al.* (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in *Arabidopsis*. *Cell* 126, 1189–1201
201. Lippman, Z. *et al.* (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430, 471–476
202. Weber, M. *et al.* (2005) Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37, 853–862
203. Rollins, R.A. *et al.* (2006) Large-scale structure of genomic methylation patterns. *Genome Res.* 16, 157–163
204. Cokus, S.J. *et al.* (2008) Shotgun bisulfite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning. *Nature* 452, 215–219
205. Lister, R. *et al.* (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133, 523–536
206. Meissner, A. *et al.* (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* 454, 766–770
207. Mikkelsen, T.S. *et al.* (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560
208. Lister, R. *et al.* (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462, 315–322
209. Saxonov, S. *et al.* (2006) A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci. U. S. A.* 103, 1412–1417
210. Borgel, J. *et al.* (2010) Targets and dynamics of promoter DNA methylation during early mouse development. *Nat. Genet.* 42
211. Weber, M. *et al.* (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* 39, 457–466
212. Mayer, W. *et al.* (2000) Demethylation of the zygotic paternal genome. *Nature* 403, 501–502
213. Ooi, S.K.T. and Bestor, T.H. (2008) The colorful history of active DNA demethylation. *Cell* 133, 1145–1148
214. Tahiliani, M. *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935
215. Kraucionis, S. and Heintz, N. (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930
216. Penterman, J. *et al.* (2007) Genetic interactions between DNA demethylation and methylation in *Arabidopsis*. *Plant Physiol.* 145, 1549–1557
217. Jeltsch, A. and Jurkowska, R.Z. (2016) Allosteric control of mammalian DNA methyltransferases – a new regulatory paradigm. *Nucleic Acids Res.* 44, 8556–8575
218. Luo, C. *et al.* (2018) Dynamic DNA methylation: in the right place at the right time. *Science* 361, 1336–1340
219. Janssen, S.M. and Lorincz, M.C. (2021) Interplay between chromatin marks in development and disease. *Nat. Rev. Genet.* 23, 137–153
220. Colaneri, A. *et al.* (2011) Expanded methyl-sensitive cut counting reveals hypomethylation as an epigenetic state that highlights functional sequences of the genome. *Proc. Natl. Acad. Sci. U. S. A.* 108, 9715–9720
221. Hodges, E. *et al.* (2011) Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell* 44, 17–28
222. Stadler, M.B. *et al.* (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495
223. Kim, J. (2003) Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. *Hum. Mol. Genet.* 12, 233–245
224. Rao, S. *et al.* (2018) Systematic prediction of DNA shape changes due to CpG methylation explains epigenetic effects on protein-DNA binding. *Epigenetics Chromatin* 11, 6

225. Kribelbauer, J.F. *et al.* (2017) Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. *Cell Rep.* 19, 2383–2395
226. Yin, Y. *et al.* (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356, eaaj2239
227. Domcke, S. *et al.* (2015) Competition between DNA methylation and transcription factors determines binding of NRF1. *Nature* 528, 575–579
228. Wan, J. *et al.* (2017) Methylated cis-regulatory elements mediate KLF4-dependent gene transactivation and cell migration. *eLife* 6, e20068
229. Donaghey, J. *et al.* (2018) Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* 50, 250–258
230. Xie, W. *et al.* (2013) Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153, 1134–1148
231. Jeong, M. *et al.* (2014) Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* 46, 17–23
232. Gu, T. *et al.* (2018) DNMT3A and TET1 cooperate to regulate promoter epigenetic landscapes in mouse embryonic stem cells. *Genome Biol.* 19, 88
233. Zhang, Y. *et al.* (2010) Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail. *Nucleic Acids Res.* 38, 4246–4253
234. Noh, K.-M. *et al.* (2015) Engineering of a histone-recognition domain in Dnmt3a alters the epigenetic landscape and phenotypic features of mouse ESCs. *Mol. Cell* 59, 89–103
235. Guo, X. *et al.* (2015) Structural insight into autoinhibition and histone H3-induced activation of DNMT3A. *Nature* 517, 640–644
236. Takeshita, K. *et al.* (2011) Structural insight into maintenance methylation by mouse DNA methyltransferase 1 (Dnmt1). *Proc. Natl. Acad. Sci. U. S. A.* 108, 9055–9059
237. Syeda, F. *et al.* (2011) The replication focus targeting sequence (RFTS) domain is a DNA-competitive inhibitor of Dnmt1. *J. Biol. Chem.* 286, 15344–15351
238. Song, J. *et al.* (2011) Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation. *Science* (1979) 331, 1036–1040
239. Fang, J. *et al.* (2016) Hemi-methylated DNA opens a closed conformation of UHRF1 to facilitate its histone recognition. *Nat. Commun.* 7, 11197
240. Rothbart, S.B. *et al.* (2012) Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat. Struct. Mol. Biol.* 19, 1155–1160
241. Dhayalan, A. *et al.* (2010) The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* 285, 26114–26120
242. Baubec, T. *et al.* (2015) Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* 520, 243–247
243. Weinberg, D.N. *et al.* (2019) The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* 573, 281–286
244. Xu, Q. *et al.* (2019) SETD2 regulates the maternal epigenome, genomic imprinting and embryonic development. *Nat. Genet.* 51, 844–856
245. Shirane, K. *et al.* (2020) NSD1-deposited H3K36me2 directs de novo methylation in the mouse male germline and counteracts Polycomb-associated silencing. *Nat. Genet.* 52, 1088–1098
246. Neri, F. *et al.* (2017) Intragenic DNA methylation prevents spurious transcription initiation. *Nature* 543, 72–77
247. Jinek, M. *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821
248. Hsu, P.D. *et al.* (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell* 157, 1262–1278
249. Thakore, P.I. *et al.* (2016) Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* 13, 127–137
250. Liu, X.S. *et al.* (2016) Editing DNA methylation in the mammalian genome. *Cell* 167, 233–247
251. Morita, S. *et al.* (2016) Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat. Biotechnol.* 34, 1060–1065
252. Huang, Y.-H. *et al.* (2017) DNA epigenome editing using CRISPR-Cas9 SunTag-directed DNMT3A. *Genome Biol.* 18, 176
253. Galonska, C. *et al.* (2018) Genome-wide tracking of dCas9-methyltransferase footprints. *Nat. Commun.* 9, 597
254. Smith, Z.D. *et al.* (2012) A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* 484, 339–344
255. Smith, Z.D. *et al.* (2014) DNA methylation dynamics of the human preimplantation embryo. *Nature* 511, 611–615
256. Guo, H. *et al.* (2014) The DNA methylation landscape of human early embryos. *Nature* 511
257. Jacobs, F.M.J. *et al.* (2014) An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516, 242–245
258. Berman, B.P. *et al.* (2012) Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat. Genet.* 44, 40–46
259. Zhou, W. *et al.* (2018) DNA methylation loss in late-replicating domains is linked to mitotic cell division. *Nat. Genet.* 50, 591–602
260. Smith, Z.D. *et al.* (2017) Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature* 549, 543–547
261. Salhab, A. *et al.* (2018) A comprehensive analysis of 195 DNA methylomes reveals shared and cell-specific features of partially methylated domains. *Genome Biol.* 19, 150
262. Greger, V. *et al.* (1989) Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum. Genet.* 83, 155–158
263. Herman, J.G. *et al.* (1994) Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl. Acad. Sci. U. S. A.* 91, 9700–9704
264. Schlesinger, Y. *et al.* (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat. Genet.* 39, 232–236
265. Su, J. *et al.* (2018) Homeobox oncogene activation by pan-cancer DNA hypermethylation. *Genome Biol.* 19, 108
266. Ziller, M.J. *et al.* (2013) Charting a dynamic DNA methylation landscape of the human genome. *Nature* 500, 477–481
267. Barau, J. *et al.* (2016) The DNA methyltransferase DNMT3C protects male germ cells from transposon activity. *Science* 354, 909–912
268. Yarychkivska, O. *et al.* (2018) BAH domains and a histone-like motif in DNA methyltransferase 1 (DNMT1) regulate de novo and maintenance methylation in vivo. *J. Biol. Chem.* 293, 19466–19475
269. Li, Y. *et al.* (2018) Stella safeguards the oocyte methylome by preventing de novo methylation mediated by DNMT1. *Nature* 564, 136–140
270. Wang, Q. *et al.* (2020) Imprecise DNMT1 activity coupled with neighbor-guided correction enables robust yet flexible epigenetic inheritance. *Nat. Genet.* 52, 828–839
271. Haggerty, C. *et al.* (2021) Dnmt1 has de novo activity targeted to transposable elements. *Nat. Struct. Mol. Biol.* 28, 594–603
272. Dixon, G. *et al.* (2021) QSER1 protects DNA methylation valleys from de novo methylation. *Science* 372, eabd0875
273. Grand, R.S. *et al.* (2021) BANP opens chromatin and activates CpG-island-regulated genes. *Nature* 596, 133–137
274. Ren, W. *et al.* (2020) Direct readout of heterochromatic H3K9me3 regulates DNMT1-mediated maintenance DNA methylation. *Proc. Natl. Acad. Sci. U. S. A.* 117, 18439–18447
275. Ren, W. *et al.* (2021) DNMT1 reads heterochromatic H4K20me3 to reinforce LINE-1 DNA methylation. *Nat. Commun.* 12, 2490
276. Weinberg, D.N. *et al.* (2021) Two competing mechanisms of DNMT3A recruitment regulate the dynamics of de novo DNA methylation at PRC1-targeted CpG islands. *Nat. Genet.* 53, 794–800
277. Scarano, E. *et al.* (1967) The heterogeneity of thymine methyl group origin in DNA pyrimidine isostichs of developing sea urchin embryos. *Proc. Natl. Acad. Sci. U. S. A.* 57, 1394–1400

278. Grippo, P. *et al.* (1968) Methylation of DNA in developing sea urchin embryos. *J. Mol. Biol.* 36, 195–208
279. Comb, D.G. (1965) Methylation of nucleic acids during sea urchin embryo development. *J. Mol. Biol.* 11, 851–855
280. Swartz, M.N. *et al.* (1962) Enzymatic synthesis of deoxyribonucleic acid. XI. Further studies on nearest neighbor base sequences in deoxyribonucleic acids. *J. Biol. Chem.* 237, 1961–1967
281. Coulondre, C. *et al.* (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. *Nature* 274, 775–780
282. Bird, A.P. (1980) DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Res.* 8, 1499–1504
283. Cooper, D.N. and Krawczak, M. (1989) Cytosine methylation and the fate of CpG dinucleotides in vertebrate genomes. *Hum. Genet.* 83, 181–188
284. Kong, A. *et al.* (2012) Rate of de novo mutations and the importance of father's age to disease risk. *Nature* 488, 471–475
285. Cooper, D.N. and Youssoufian, H. (1988) The CpG dinucleotide and human genetic disease. *Hum. Genet.* 78, 151–155
286. Crouse, H.V. (1960) The controlling element in sex chromosome behavior in *Sciara*. *Genetics* 45, 1429–1443
287. Surani, M.A.H. *et al.* (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* 308, 548–550
288. Cattanach, B.M. and Kirk, M. (1985) Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* 315, 496–498
289. Sanford, J. *et al.* (1984) Methylation patterns of repetitive DNA sequences in germ cells of *Mus musculus*. *Nucleic Acids Res.* 12, 2823–2836
290. Sanford, J.P. *et al.* (1987) Differences in DNA methylation during oogenesis and spermatogenesis and their persistence during early embryogenesis in the mouse. *Genes Dev.* 1, 1039–1046
291. Reik, W. *et al.* (1987) Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* 328, 248–251
292. Sapienza, C. *et al.* (1987) Degree of methylation of transgenes is dependent on gamete of origin. *Nature* 328, 251–254
293. Swain, J.L. *et al.* (1987) Parental legacy determines methylation and expression of an autosomal transgene: a molecular mechanism for parental imprinting. *Cell* 50, 719–727
294. Barlow, D.P. *et al.* (1991) The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus. *Nature* 349, 84–87
295. DeChiara, T.M. *et al.* (1991) Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* 64, 849–859
296. Cattanach, B.M. *et al.* (1992) A candidate mouse model for Prader-Willi syndrome which shows an absence of Shprn expression. *Nat. Genet.* 2, 270–274
297. Ferguson-Smith, A.C. *et al.* (1993) Parental-origin-specific epigenetic modification of the mouse H19 gene. *Nature* 362, 751–755
298. Stoger, R. *et al.* (1993) Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal. *Cell* 73, 61–71
299. Smallwood, S.A. *et al.* (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nat. Genet.* 43, 811–814
300. Xie, W. *et al.* (2012) Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. *Cell* 148, 816–831
301. Hirasawa, R. *et al.* (2008) Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* 22, 1607–1616
302. Borowczyk, E. *et al.* (2009) Identification of a region of the DNMT1 methyltransferase that regulates the maintenance of genomic imprints. *Proc. Natl. Acad. Sci. U. S. A.* 106, 20806–20811
303. Quenneville, S. *et al.* (2011) In embryonic stem cells, ZFP57/KAP1 recognize a methylated hexanucleotide to affect chromatin and DNA methylation of imprinting control regions. *Mol. Cell* 44, 361–372
304. Inoue, A. *et al.* (2017) Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* 547, 419–424
305. Andergassen, D. *et al.* (2021) Diverse epigenetic mechanisms maintain parental imprints within the embryonic and extraembryonic lineages. *Dev. Cell* 56, 2995–3005
306. Barr, M.L. and Bertram, E.G. (1949) A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* 163, 676–677
307. Ohno, S. *et al.* (1959) Formation of the sex chromatin by a single X-chromosome in liver cells of *Rattus norvegicus*. *Exp. Cell Res.* 18, 415–418
308. Lyon, M.F. (1961) Gene Action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* 190, 372–373
309. Mohandas, T. *et al.* (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. *Science* 211, 393–396
310. Jones, P.A. *et al.* (1982) Cell cycle-specific reactivation of an inactive X-chromosome locus by 5-azadeoxycytidine. *Proc. Natl. Acad. Sci. U. S. A.* 79, 1215–1219
311. Venolia, L. and Gartner, S.M. (1983) Comparison of transformation efficiency of human active and inactive X-chromosomal DNA. *Nature* 302, 82–83
312. Venolia, L. *et al.* (1982) Transformation with DNA from 5-azacytidine-reactivated X chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 79, 2352–2354
313. Lock, L.F. *et al.* (1987) Methylation of the Hprt gene on the inactive X occurs after chromosome inactivation. *Cell* 48, 39–46
314. Stedman, E. and Stedman, E. (1950) Cell specificity of histones. *Nature* 166, 780–781
315. Alfrey, V.G. *et al.* (1964) Acetylation and methylation of histones and their possible role in the regulation of RNA synthesis. *Proc. Natl. Acad. Sci. U. S. A.* 51, 786–794
316. Richmond, T.J. *et al.* (1984) Structure of the nucleosome core particle at 7 Å resolution. *Nature* 311, 532–537
317. Luger, K. *et al.* (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260
318. Durrin, L.K. *et al.* (1991) Yeast histone H4 N-terminal sequence is required for promoter activation in vivo. *Cell* 65, 1023–1031
319. Brownell, J.E. *et al.* (1996) *Tetrahymena* histone acetyltransferase A: a homolog to yeast Gcn5p linking histone acetylation to gene activation. *Cell* 84, 843–851
320. Strahl, B.D. *et al.* (2000) The language of covalent histone modifications. *Nature* 403, 41–45
321. Bannister, A.J. *et al.* (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120–124
322. Lachner, M. *et al.* (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410, 116–120
323. Nakayama, J.-I. (2001) Role of histone H3 lysine 9 methylation in epigenetic control of heterochromatin assembly. *Science* 292, 110–113