Chapter 1

CpG Islands: A Historical Perspective

Francisco Antequera and Adrian Bird

Abstract

The discovery of CpG islands (CGIs) and the study of their structure and properties run parallel to the development of molecular biology in the last two decades of the twentieth century and to the development of high-throughput genomic technologies at the turn of the millennium. First identified as discrete G + C-rich regions of unmethylated DNA in several vertebrates, CGIs were soon found to display additional distinctive chromatin features from the rest of the genome in terms of accessibility and of the epigenetic modifications of their histones. These features, together with their colocalization with promoters and with origins of DNA replication in mammals, highlighted their relevance in the regulation of genomic processes. Recent approaches have shown with unprecedented detail the dynamics and diversity of the epigenetic landscape of CGIs during normal development and under pathological conditions. Also, comparative analyses across species have started revealing how CGIs evolve and contribute to the evolution of the vertebrate genome.

Key words CpG islands, DNA mehylation, Chromatin, Transcription, Evolution

1 DNA Methylation Patterns

5-methyl-cytosine (5mC) was first detected as a natural component of DNA by Hotchkiss [1] when trying to determine the purity of DNA preparations in calf thymus. Soon afterward, it was also found in wheat [2] and subsequently in other types of eukaryotes. Technology at that time only permitted 5mC to be recognized as a minor fraction of all cytosine in DNA, leaving many of the questions concerning its genomic distribution and possible biological significance unanswered. Despite these limitations, chemical and enzymatic hydrolysis of DNA and biochemical methods such as column or paper chromatography, mass spectrometry, or thin-layer chromatography, established that more than 90% of all 5mC was found in CpG dinucleotides [3–5], and that a large fraction of them were methylated in the mammalian genome [6, 7]. Early studies also showed that nuclease-resistant regions and repetitive sequences were enriched in 5mC [8, 9], while actively transcribed regions were hypomethylated relative to the genome average [7],

Tanya Vavouri and Miguel A. Peinado (eds.), *CpG Islands: Methods and Protocols*, Methods in Molecular Biology, vol. 1766, https://doi.org/10.1007/978-1-4939-7768-0_1,

providing the first indications that 5mCpG dinucleotides were not uniformly distributed within the genome.

The advent of molecular biology techniques around the middle of 1970s provided the unprecedented opportunity to dissect the structure, and hence elucidate the function of specific genomic regions, including individual genes. In the case of DNA methylation, the use of restriction enzymes to analyze CpG in their recognition sites allowed methylated and nonmethylated versions of the same sites to be distinguished and were used to generate the first maps of DNA methylation [10–14]. In particular, the isoschizomers Hpa II and Msp I were widely used, since both recognize the 5′-CCGG sequence, but Hpa II does not cleave if the internal CpG is methylated. Molecular hybridization, coupled with the newly developed Southern blot analysis [15], detected enrichment of nonmethylated CpGs in the promoter region of some genes in expressing tissues, which gave rise to the first hints that DNA methylation might repress transcription and suggested a possible role in the regulation of gene expression [16–21]. This notion was reinforced by the finding that gene promoters in the inactive X chromosome were methylated and could be reactivated by inducing demethylation with 5-azacytidine [22, 23].

A second and more straightforward application of the Hpa II/Msp I isochizomer pair was the analysis of the restriction patterns they generated using genomic DNA and agarose gels. A pioneer analysis in sea urchin revealed that approximately 40% of its genome was resistant to Hpa II and other 5mCpG-sensitive enzymes, but not to Msp I. This fraction was present in different tissues and remained stable during development, although small changes would have gone unnoticed at the level of resolution available at the time [24]. This pattern of interspersed methylated and nonmethylated domains was initially called “echinoderm-type” and was later found in many other invertebrates and some fungi [25–27]. Despite its widespread phylogenetic distribution, this organization was not universal, however, as two additional global genomic methylation patterns were detected. One was the “insect-type pattern,” as exemplified by *Drosophila*, which showed undetectable differences between the Msp I and Hpa II digestions, suggesting a widespread lack of genomic methylation. The second pattern was found in mice and humans, as well as birds, reptiles, amphibia, and fish. This “vertebrate-type” pattern was opposite to that of *Drosophila*, as the genome was poorly digested by 5mCpG-sensitive enzymes, indicating widespread genomic methylation. In particular, there was no evidence of the methylated and nonmethylated compartments seen in organisms with the echinoderm-type pattern.

2 Discovery of the CpG Islands

To enhance the resolution of the Msp I/Hpa II comparison, Adrian Bird and his collaborators in Edinburgh used the simple strategy of end-labeling the Hpa II fragments with ^{32}P before separating them in an agarose gel. The aim was to separate the fragments of vertebrate DNA according their molecular weight, visualized by ethidium bromide staining, but at the same time to assess fragment number through direct autoradiography of the gel [28]. The difference between the two resulting patterns was astonishing: the ethidium bromide pattern consisted of very high molecular weight fragments, indicative of poorly digested vertebrate DNA (Fig. 1c, d), whereas autoradiography of the same gel revealed a prominent set of small fragments, which greatly exceeded the number of Hpa II fragments at the top of the gel (Fig. 1g, j–o). Most were less than 500 base pairs (bp) in length, the average being approximately 120 bp. Thus the experiment revealed two asymmetric compartments: a small fraction, amounting to less than 2% of the genome and rich in closely spaced and nonmethylated sites for CpG enzymes, and the rest of the genome, which was highly methylated. Given its large size, the methylated fraction was likely to include satellite DNA, repeated elements, nontranscribed regions, and genes.

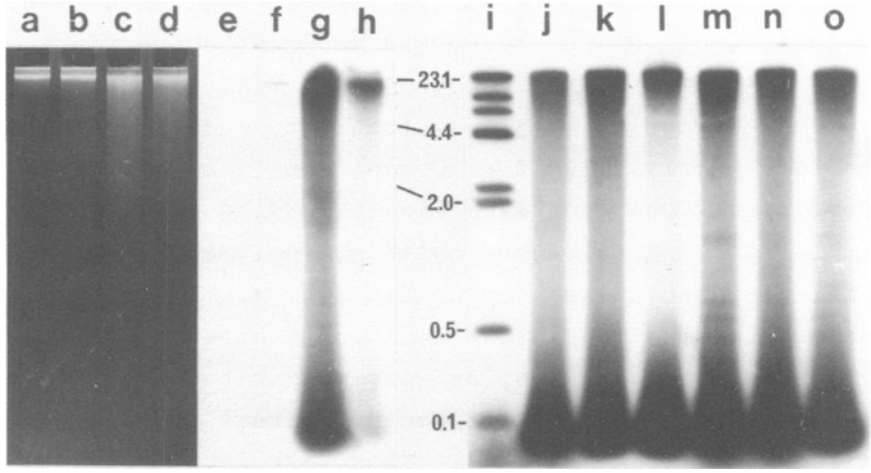

Fig. 1 End-labeling of Hpa II fragments of chicken DNA. Lanes (a–d), agarose gel of end-labeled chicken kidney DNA stained with ethidium bromide. Lanes (e–h), autoradiograph of the same gel. Samples (a) and (e), undigested DNA labeled with (α-^{32}P) dCTP; (b) and (f), undigested DNA labeled with (α-^{32}P) dTTP; (c) and (g), Hpa II-digested DNA labeled with (α-^{32}P) dCTP. (d) and (h), Hpa II-digested DNA labeled with (α-^{32}P) dTTP. Lane (i), bacteriophage lambda DNA digested with Hind III and end labeled with (α-^{32}P) dCTP. Lanes (j–o), autoradiograph of DNA from various chicken tissues after Hpa II digestion and end-labeling with (α-^{32}P) dCTP. Sample (j), blood cell nuclei; (k), whole blood; (l), sperm, (m), liver, (n), kidney; (o), brain. Gels were 1.2% agarose. Fragment lengths are given in kilobase pairs (Reproduced from [28], with permission)

A similar nonmethylated fraction was found in many other vertebrates and was initially called the "HTF fraction" (for Hpa II Tiny Fragments) [29], although it was later renamed as the CpG island fraction (CGI) [30]. Its discovery immediately provided a framework for the explanation of several other reports concerning asymmetric base composition along some gene sequences. For example, McClelland and Ivarie [31] found that the average CpG frequency along 15 mammalian genes was higher at the 5′ end than at the 3′ end. Regions enriched in CpGs were also found in the polymorphic exons of some MHC class I and II genes, and in the 5′ end of some genes from chicken and mice [32]. Since a few nonmethylated CpG sites had been mapped by Southern hybridization to these regions [21, 33], it was likely that they contributed to the CpG island fraction.

To identify the genomic origin of small nonmethylated fragments, several mouse genomic clones were isolated through the hybridization of individual HTF fragments to genomic libraries. Analysis of their sequences revealed that they were derived from regions approximately 1000 bp long with a base composition of 65% G + C, which was significantly higher than the genomic average of 40% G + C [29]. Based on this size and on the percentage of the genome represented by the HTF fraction, it was estimated that there were approximately 30,000 CpG islands in the mouse genome [29], a figure that is very close to more recent estimates based of genomic methylation analyses [34]. A striking feature of these regions was the absence of CpG suppression found in bulk DNA. In the methylated majority of the genome, CpG dinucleotides are present at 20% of the expected frequency due to the frequent conversion of 5mCpG to TpG caused by deamination of 5mC [35, 36]. This fact, together with their cleavage by 5mC-sensitive endonucleases, raised the possibility that CpG islands regions could be nonmethylated along their entire length in all of the tissues tested, including sperm. This expectation was proven correct by the development of methods capable of detecting methylation of every CpG by ligation-mediated PCR [37, 38], and later by the bisulphite treatment of DNA. The latter technique, which chemically converts C, but not mC, to T, has become the gold-standard method for methylation analyses at single nucleotide resolution [39], and is now widely used coupled with next generation sequencing [40].

3 Functional Properties of CpG Islands

An essential question that arose after their discovery concerned the connection between CGIs and genes. Northern hybridization of several CGI clones isolated at random from the mouse genome detected transcripts in several tissues, suggesting that many CGis

could be associated with the promoters of widely expressed genes [29]. This association, together with the high frequency of sites for "rare cutter" enzymes (those with a high G + C content in their recognition sites) at CGIs, opened the unprecedented possibility of directly identifying the 5′ region of human genes in cosmid libraries and in genomic DNA [41, 42]. The first CGI promoter characterized in detail sustained bidirectional transcription [43], which was later found to be a general property of most CGI promoters [44–46]. Many studies of specific genes over the years and more recent genome-wide analyses have generalized the association between CGIs and genes to the extent that even those CGIs initially considered "orphans" have been found to colocalize with sites of transcription initiation [34]. That lack of methylation of CGIs was essential for the activity of the associated genes was initially shown by the transcriptional silencing caused by the artificial methylation of the 5′ CGI region of the hamster APRT gene in mouse cells, but not by methylation of the body of the gene [20, 47].

Methylation of the CGIs at the 5′ end of genes on the inactive X chromosomes was also associated with transcriptional silencing [48–51] and later studies showed that the repressive effect of methylation depended on the density of CpGs within the promoter regions, such that transcriptional silencing from fully methylated CGIs could not be reactivated even by strong enhancers [52, 53]. These findings were more broadly relevant, since methylated CGIs were also found in autosomal genes in tumours, for example within a region of human chromosome 11 [54], and also at tissue-specific genes in cultured cell lines [55, 56]. The affected CGIs were nonmethylated in normal tissues. Since those early studies, hundreds of examples have been documented in a variety of tumours, making unscheduled CGI methylation one of the hallmarks of cancer [57]. The finding of aberrant CGI methylation in cultured cells and in cancer highlighted the possibility that DNA methylation could be a natural regulator of CGI promoter activity during development. Indeed it emerged that, in addition to its role in X chromosome inactivation, DNA methylation is also functionally relevant for the monoallelic expression of imprinted genes (reviewed in [58]) and in the silencing of endogenous retrotransposons [59]. Despite these examples, it is unlikely that DNA methylation of CGI promoters is widely used to dynamically regulate gene expression during development, as the majority of CGIs remain in a nonmethylated state throughout [60]. There are also many examples where demethylation of non CGi promoters is not accompanied by their activation in tissues where these genes are normally silenced [59, 61–63].

Another important property of CGIs was the finding that replication origins colocalized with them at the CGI promoter region of the human *c-MYC* [64], *TIMM13* [65, 66], and *HSP70*

genes [67]. As in the case of their promoter function, the colocalization of CGIs with ORIs was later generalized to include many CGIs, which colocalize with active ORIs in different human and mouse cell types [68–71].

A question raised by the initial characterization of CGIs was whether their distinctive properties would be paralleled by an equally distinctive chromatin organization. Histone fractionation of CGI chromatin in acid-urea gels showed that histone H1 was undetectable at CGIs and histones H3 and H4 were hyperacetylated [72]. DNAse I and micrococcal nuclease digestion revealed that CGIs included sites devoid of nucleosomes close to the site of transcription initiation. Although the extensive catalogue of histone modifications that are known today was missing in those days, these features potentially endowed CGIs with properties of "accessible" or open chromatin. How, then, did they manage to avoid the layer of methylation that covers the rest of the genome? The current view is that several mechanisms contribute to preventing methylation of CGIs. These include the binding of proteins to nonmethylated CpGs through the zinc finger CxxC domain such as the H3K36 histone demethylases KDM2A/KDM2B [73] and the histone H3K4 methyltransferase CFP1 [74]. Importantly, the histone modification H3K4me3, which frequently coincides with CGIs, prevents the DNA methyltransferase subunit Dnmt3L from accessing and methylating the associated DNA [75]. In addition, unscheduled methylation of CGIs could be actively removed by the TET proteins, which are also members of the CxxC protein family [76].

Paradoxically, it was the search for proteins that could shield CGIs from DNA methyltransferases that led to the identification of two protein activities, MeCP1 and MeCP2, which behaved in exactly the opposite way, as they were able to bind to methylated CpGs without sequence specificity for their flanking sequences. Those proteins included a methylated DNA binding domain (MBD) [77–79], which was later found in more proteins that bind to methylated DNA in vivo [80, 81]. The MBD family of proteins interprets the information encoded in the DNA methylation patterns through the interaction with transcriptional corepressors, histone epigenetic modifiers, and chromatin remodelers, and has overlapping but not redundant functions as shown by the different phenotypes generated by deletion of their coding genes (review of [82]).

4 Closing Remarks: From Maniatis to Methylomes

The basic structure and functional properties of CGIs were established following mostly the instructions and protocols described in the legendary first edition of the "Maniatis" manual [83]. This

single volume was devoutly read by a whole generation of molecular biologists who laid the foundations for our current understanding of many genetic and developmental processes. In the field of DNA methylation, the one-step-at-a-time approach typical of classic molecular biology meant that several days of work involving DNA digestion, Southern blotting and hybridization were needed before the exciting moment arrived of developing the autoradiography to spot the differences—often corresponding to just one or two CpGs—between the Msp I and Hpa II banding patterns. At that time, the possibility of mapping the methylation status of every CpG in the genome in a single experiment or even in single cells [40] was beyond the reach of anyone's imagination.

The advent of high-throughput genomic technologies has provided a more nuanced picture of the methylation landscape than the vertebrate nonmethylated and methylated compartments initially described. For example, we now know that CGI-like regions can occasionally be methylated in different tissues or developmental stages [34, 84–87], and also that small nonmethylated regions, often encompassing one or a few CpGs, result from the binding of proteins to DNA [88–90]. The increasing availability of whole genome methylomes allows direct comparison between genomes within the same species at different developmental stages or physiological conditions, and makes the computational prediction of CGIs—that vary widely depending on small differences in base composition and CpG frequency [34, 84, 91]—unnecessary. These genome-wide analyses have confirmed previous findings that orthologous CGIs differ greatly in terms of base composition and CpG frequency among vertebrate species, although they maintain their nonmethylated condition [84, 92–94]. In fact CGI promoters evolve faster than CpG-poor promoters [95], which could account for the fact that CGIs are single-copy sequences in the genome, although they can share some binding sites for transcription factors. Several evolutionary scenarios for the emergence of CGIs have been proposed which, despite their colocalization with promoters and replication origins, suggest that CGIs have evolved under very little selective pressure [96].

The work of hundreds of laboratories, after almost 35 years since CGIs were first time visualized ([28], Fig. 1), has established CGIs as distinctive regulatory regions within the vast excess of genomic DNA sequence. Taking advantage of the breathtaking pace with which molecular techniques have improved our ability to scrutinize the dynamics of the genome, CGIs will doubtless continue to contribute to our understanding of the regulation, pathology, and evolution of the vertebrate genome for many years to come.

References

1. Hotchkiss RD (1948) The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. J Biol Chem 175:315–332
2. Wyatt GR (1951) Recognition and estimation of 5-methylcytosine in nucleic acids. Biochem J 48:581–584
3. Doskocil J, Sorm F (1962) Distribution of 5-methylcytosine in pyrimidine sequences of deoxyribonucleic acids. Biochim Biophys Acta 55:953–959
4. Grippo P, Iaccarino M, Parisi E, Scarano E (1968) Methylation of DNA in developing sea urchin embryos. J Mol Biol 36:195–208
5. Sinsheimer RL (1955) The action of pancreatic deoxyribonuclease. II Isomeric dinucleotides. J Biol Chem 215:579–583
6. Gruenbaum Y, Stein R, Cedar H, Razin A (1981) Methylation of CpG sequences in eukaryotic DNA. FEBS Lett 124:67–71
7. Naveh-Many T, Cedar H (1981) Active gene sequences are undermethylated. Proc Natl Acad Sci U S A 78:4246–4250
8. Razin A, Cedar H (1977) Distribution of 5-methylcytosine in chromatin. Proc Natl Acad Sci U S A 74:2725–2728
9. Solage A, Cedar H (1978) Organization of 5-methylcytosine in chromosomal DNA. Biochemistry 17:2934–2938
10. Bird AP (1978) Use of restriction enzymes to study eukaryotic DNA methylation: II The symmetry of methylated sites supports semiconservative copying of the methylation pattern. J Mol Biol 118:49–60
11. Bird AP, Southern EM (1978) Use of restriction enzymes to study eukaryotic DNA methylation: I The methylation pattern in ribosomal DNA from Xenopus laevis. J Mol Biol 118:27–47
12. Gautier F, Bunemann H, Grotjahn L (1977) Analysis of calf-thymus satellite DNA: evidence for specific methylation of cytosine in C-G sequences. Eur J Biochem 80:175–183
13. Mandel JL, 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
14. Waalwijk C, Flavell RA (1978) DNA methylation at a CCGG sequence in the large intron of the rabbit beta-globin gene: tissue-specific variations. Nucleic Acids Res 5:4631–4634
15. Southern EM (1975) Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517
16. Busslinger M, Hurst J, Flavell RA (1983) DNA methylation and the regulation of globin gene expression. Cell 34:197–206
17. Kruczek I, Doerfler W (1982) The unmethylated state of the promoter/leader and 5′-regions of integrated adenovirus genes correlates with gene expression. EMBO J 1:409–414
18. Ott MO, Sperling L, Cassio D, Levilliers J, Sala-Trepat J, Weiss MC (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 30:825–833
19. Shen CK, Maniatis T (1980) Tissue-specific DNA methylation in a cluster of rabbit beta-like globin genes. Proc Natl Acad Sci U S A 77:6634–6638
20. Stein R, Razin A, Cedar H (1982) In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. Proc Natl Acad Sci U S A 79:3418–3422
21. Stein R, Sciaky-Gallili N, Razin A, Cedar H (1983) Pattern of methylation of two genes coding for housekeeping functions. Proc Natl Acad Sci U S A 80:2422–2426
22. Mohandas T, Sparkes RS, Shapiro LJ (1981) Reactivation of an inactive human X chromosome: evidence for X inactivation by DNA methylation. Science 211:393–396
23. Venolia L, Gartler SM, Wassman ER, Yen P, Mohandas T, Shapiro LJ (1982) Transformation with DNA from 5-azacytidine-reactivated X chromosomes. Proc Natl Acad Sci U S A 79:2352–2354
24. Bird AP, Taggart MH, Smith BA (1979) Methylated and unmethylated DNA compartments in the sea urchin genome. Cell 17:889–901
25. Antequera F, Tamame M, Villanueva JR, Santos T (1984) DNA methylation in the fungi. J Biol Chem 259:8033–8036
26. Bird AP, Taggart MH (1980) Variable patterns of total DNA and rDNA methylation in animals. Nucleic Acids Res 8:1485–1497
27. Whittaker PA, Hardman N (1980) Methylation of nuclear DNA in Physarum polycephalum. Biochem J 191:859–862

28. Cooper DN, Taggart MH, Bird AP (1983) Unmethylated domains in vertebrate DNA. Nucleic Acids Res 11:647–658
29. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D (1985) A fraction of the mouse genome that is derived from islands of non-methylated, CpG-rich DNA. Cell 40:91–99
30. Bird A (1987) CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3:342–347
31. McClelland M, Ivarie R (1982) Asymmetrical distribution of CpG in an "average" mammalian gene. Nucleic Acids Res 10:7865–7877
32. Tykocinski ML, Max EE (1984) CG dinucleotide clusters in MHC genes and in 5′ demethylated genes. Nucleic Acids Res 12:4385–4396
33. McKeon C, Ohkubo H, Pastan I, de Crombrugghe B (1982) Unusual methylation pattern of the alpha 2 (I) collagen gene. Cell 29:203–210
34. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP (2010) Orphan CpG islands identify numerous conserved promoters in the mammalian genome. PLoS Genet 6:e1001134
35. Coulondre C, Miller JH, Farabough PJ, Gilbert W (1978) Molecular basis of base substitution hotspots in *Escherichia coli*. Nature 274:775–780
36. Bird A (1980) DNA methylation and the frequency of CpG in animal DNA. Nucleic Acids Res 8:1499–1504
37. Nick H, Bowen B, Ferl RJ, Gilbert W (1986) Detection of cytosine methylation in the maize alcohol dehydrogenase gene by genomic sequencing. Nature 319:243–246
38. Pfeifer GP, Steigerwald SD, Mueller PR, Wold B, Riggs AD (1989) Genomic sequencing and methylation analysis by ligation mediated PCR. Science 246:810–813
39. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, Molloy PL, Paul CL (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
40. Yong WS, Hsu FM, Chen PY (2016) Profiling genome-wide DNA methylation. Epigenetics Chromatin 9:26
41. Brown WR, Bird AP (1986) Long-range restriction site mapping of mammalian genomic DNA. Nature 322:477–481
42. Lindsay S, Bird AP (1987) Use of restriction enzymes to detect potential gene sequences in mammalian DNA. Nature 327:336–338
43. Lavia P, Macleod D, Bird A (1987) Coincident start sites for divergent transcripts at a randomly selected CpG islands as gene markers in the vertebrate nucleus. Trends Genet 3: 342–347
44. Adachi N, Lieber MR (2002) Bidirectional gene organization: a common architectural feature of the human genome. Cell 109:807–809
45. Core LJ, Waterfall JJ, Lis JT (2008) Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science 322:1845–1848
46. Seila AC, Calabrese JM, Levine SS, Yeo GW, Rahl PB, Flynn RA, Young RA, Sharp PA (2008) Divergent transcription from active promoters. Science 322:1849–1851
47. Keshet I, Yisraeli J, Cedar H (1985) Effect of regional DNA methylation on gene expression. Proc Natl Acad Sci U S A 82:2560–2564
48. Pfeifer GP, Tanguay RL, Steigerwald SD, Riggs AD (1990) In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. Genes Dev 4:1277–1287
49. Toniolo D, Martini G, Migeon BR, Dono R (1988) Expression of the G6PD locus on the human X chromosome is associated with demethylation of three CpG islands within 100 kb of DNA. EMBO J 7:401–406
50. Wolf SF, Jolly DJ, Lunnen KD, Friedmann T, Migeon BR (1984) Methylation of the hypoxanthine phosphoribosyltransferase locus on the human X chromosome: implications for X-chromosome inactivation. Proc Natl Acad Sci U S A 81:2806–2810
51. Yen PH, Patel P, Chinault AC, Mohandas T, Shapiro LJ (1984) Differential methylation of hypoxanthine phosphoribosyltransferase genes on active and inactive human X chromosomes. Proc Natl Acad Sci U S A 81:1759–1763
52. Boyes J, Bird A (1992) Repression of genes by DNA methylation depends on CpG density and promoter strength: evidence for involvement of a methyl-CpG binding protein. EMBO J 11:327–333
53. Hsieh CL (1994) Dependence of transcriptional repression on CpG methylation density. Mol Cell Biol 14:5487–5494
54. de Bustros A, Nelkin BD, Silverman A, Ehrlich G, Poiesz B, Baylin SB (1988) The short arm of chromosome 11 is a "hot spot" for hypermethylation in human neoplasia. Proc Natl Acad Sci U S A 85:5693–5697
55. Antequera F, Boyes J, Bird A (1990) High levels of de novo methylation and altered

chromatin structure at CpG islands in cell lines. Cell 62:503–514

56. Jones PA, Wolkowicz MJ, Rideout WM III, Gonzales FA, Marziasz CM, Coetzee GA, Tapscott SJ (1990) De novo methylation of the MyoD1 CpG island during the establishment of immortal cell lines. Proc Natl Acad Sci U S A 87:6117–6121
57. Stirzaker C, Taberlay PC, Statham AL, Clark SJ (2014) Mining cancer methylomes: prospects and challenges. Trends Genet 30:75–84
58. Bartolomei MS, Ferguson-Smith AC (2011) Mammalian genomic imprinting. Cold Spring Harb Perspect Biol 3:a002592
59. Walsh CP, Chaillet JR, Bestor TH (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. Nat Genet 20:116–117
60. Borgel J, Guibert S, Li Y, Chiba H, Schübeler D, Sasaki H, Forné T, Weber M (2010) Targets and dynamics of promoter DNA methylation during early mouse development. Nat Genet 42:1093–1100
61. Bestor TH, Edwards JR, Boulard M (2015) Notes on the role of dynamic DNA methylation in mammalian development. Proc Natl Acad Sci U S A 112:6796–6799
62. Bird AP (1986) CpG-rich islands and the function of DNA methylation. Nature 321:209–213
63. Walsh CP, Bestor TH (1999) Cytosine methylation and mammalian development. Genes Dev 13:26–34
64. Vassilev L, Johnson EM (1990) An initiation zone of chromosomal DNA replication located upstream of the c-myc gene in proliferating HeLa cells. Mol Cell Biol 10:4899–4904
65. Biamonti G, Giacca M, Perini G, Contreas G, Zentilin L, Weighardt F, Guerra M, Della Valle G, Saccone S, Riva S et al (1992) The gene for a novel human lamin maps at a highly transcribed locus of chromosome 19 which replicates at the onset of S-phase. Mol Cell Biol 12:3499–3506
66. Giacca M, Zentilin L, Norio P, Diviacco S, Dimitrova D, Contreas G, Biamonti G, Perini G, Weighardt F, Riva S et al (1994) Fine mapping of a replication origin of human DNA. Proc Natl Acad Sci U S A 91:7119–7123
67. Taira T, Iguchi-Ariga SM, Ariga H (1994) A novel DNA replication origin identified in the human heat shock protein 70 gene promoter. Mol Cell Biol 14:6386–6397
68. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, Marin JM, Lemaitre JM (2012) Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat Struct Mol Biol 19:837–844
69. Cayrou C, Coulombe P, Vigneron A, Stanojcic S, Ganier O, Peiffer I, Rivals E, Puy A, Laurent-Chabalier S, Desprat R, Mechali M (2011) Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. Genome Res 21:1438–1449
70. Delgado S, Gomez M, Bird A, Antequera F (1998) Initiation of DNA replication at CpG islands in mammalian chromosomes. EMBO J 17:2426–2435
71. Sequeira-Mendes J, Diaz-Uriarte R, Apedaile A, Huntley D, Brockdorff N, Gomez M (2009) Transcription initiation activity sets replication origin efficiency in mammalian cells. PLoS Genet 5:e1000446
72. Tazi J, Bird A (1990) Alternative chromatin structure at CpG islands. Cell 60:909–920
73. Blackledge NP, Zhou JC, Tolstorukov MY, Farcas AM, Park PJ, Klose RJ (2010) CpG islands recruit a histone H3 lysine 36 demethylase. Mol Cell 38:179–190
74. Thomson JP, Skene PJ, Selfridge J, Clouaire T, Guy J, Webb S, Kerr AR, Deaton A, Andrews R, James KD, Turner DJ, Illingworth R, Bird A (2010) CpG islands influence chromatin structure via the CpG-binding protein Cfp1. Nature 464:1082–1086
75. Ooi SKT, Qiu C, Bernstein E, Li K, Jia D, Yang Z, Erdjument-Bromage H, Tempst P, Lin SP, Allis CD, Cheng X, Bestor TH (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. Nature 448:714–717
76. Rasmussen KD, Helin K (2016) Role of TET enzymes in DNA methylation, development, and cancer. Genes Dev 30:733–750
77. Lewis JD, Meehan RR, Henzel WJ, Maurer-Fogy I, Jeppesen P, Klein F, Bird A (1992) Purification, sequence, and cellular localization of a novel chromosomal protein that binds to methylated DNA. Cell 69:905–914
78. Meehan RR, Lewis JD, McKay S, Kleiner EL, Bird AP (1989) Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. Cell 58:499–507
79. Nan X, Meehan RR, Bird AP (1993) Dissection of the methyl-CpG binding domain from the chromosomal protein MeCP2. Nucleic Acids Res 21:4886–4892
80. Baubec T, Schubeler D (2014) Genomic patterns and context specific interpretation of

DNA methylation. Curr Opin Genet Dev 25:85–92

81. Hendrich B, Bird A (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 18:6538–6547
82. Du Q, Luu PL, Stirzaker C, Clark SJ (2015) Methyl-CpG-binding domain proteins: readers of the epigenome. Epigenomics 7:1051–1073
83. Maniatis T, Fritsch EF, Sambrook J (1982) Molecular cloning. A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
84. Long HK, Sims D, Heger A, Blackledge NP, Kutter C, Wright ML, Grutzner F, Odom DT, Patient R, Ponting CP, Klose RJ (2013) Epigenetic conservation at gene regulatory elements revealed by non-methylated DNA profiling in seven vertebrates. elife 2:e00348
85. Maunakea AK, Nagarajan RP, Bilenky M, Ballinger TJ, D'Souza C, Fouse SD, Johnson BE, Hong C, Nielsen C, Zhao Y, Turecki G, Delaney A, Varhol R, Thiessen N, Shchors K, Heine VM, Rowitch DH, Xing X, Fiore C, Schillebeeckx M, Jones SJ, Haussler D, Marra MA, Hirst M, Wang T, Costello JF (2010) Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature 466:253–257
86. Shen L, Kondo Y, Guo Y, Zhang J, Zhang L, Ahmed S, Shu J, Chen X, Waterland RA, Issa JP (2007) Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. PLoS Genet 3:2023–2036
87. Smallwood SA, Tomizawa S, Krueger F, Ruf N, Carli N, Segonds-Pichon A, Sato S, Hata K, Andrews SR, Kelsey G (2011) Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. Nat Genet 43:811–814
88. Han L, Lin IG, Hsieh CL (2001) Protein binding protects sites on stable episomes and in the chromosome from de novo methylation. Mol Cell Biol 21:3416–3424
89. Lienert F, Mohn F, Tiwari VK, Baubec T, Roloff TC, Gaidatzis D, Stadler MB, Schubeler D (2011) Genomic prevalence of heterochromatic H3K9me2 and transcription do not discriminate pluripotent from terminally differentiated cells. PLoS Genet 7:e1002090
90. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, van Nimwegen E, Wirbelauer C, Oakeley EJ, Gaidatzis D, Tiwari VK, Schubeler D (2011) DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature 480:490–495
91. Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 99:3740–3745
92. Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci U S A 90:11995–11999
93. Cross S, Kovarik P, Schmidtke J, Bird A (1991) Non-methylated islands in fish genomes are GC-poor. Nucleic Acids Res 19:1469–1474
94. Cuadrado M, Sacristan M, Antequera F (2001) Species-specific organization of CpG island promoters at mammalian homologous genes. EMBO Rep 2:586–592
95. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC, Forrest AR, Alkema WB, Tan SL, Plessy C, Kodzius R, Ravasi T, Kasukawa T, Fukuda S, Kanamori-Katayama M, Kitazume Y, Kawaji H, Kai C, Nakamura M, Konno H, Nakano K, Mottagui-Tabar S, Arner P, Chesi A, Gustincich S, Persichetti F, Suzuki H, Grimmond SM, Wells CA, Orlando V, Wahlestedt C, Liu ET, Harbers M, Kawai J, Bajic VB, Hume DA, Hayashizaki Y (2006) Genome-wide analysis of mammalian promoter architecture and evolution. Nat Genet 38:626–635
96. Cohen NM, Kenigsberg E, Tanay A (2011) Primate CpG islands are maintained by heterogeneous evolutionary regimes involving minimal selection. Cell 145:773–786