CpG islands and genes

Sally H Cross and Adrian P Bird

University of Edinburgh, Edinburgh, UK

Of the estimated 45 000 CpG islands in the human genome, the overwhelming majority are found at the 5′ ends of genes and their identification and cloning are proving very useful for finding and isolating genes. Recent work has shed light on the chromosomal distribution and origin of CpG islands. It has been shown unequivocally that CpG islands are concentrated in the R band chromosomal regions and that intact transcription factor binding sites are required for their maintenance. Cases of methylation of CpG islands and inactivation of the associated genes have been reported which may be important in ageing, tumorigenesis and imprinting.

Current Opinion in Genetics and Development 1995, 5:309–314

Introduction

It is now almost a decade since the existence of CpG islands was first appreciated [1,2]. They were identified as short regions of genomic DNA which contained many sites for the restriction enzyme *Hpa*II and were, therefore, initially called *Hpa*II Tiny Fragment (HTF) islands. CpG island DNA is found in short regions of 1–2 kb, which together account for ~2% of the genome, and has distinctive properties when compared with the DNA in the rest of the genome. It is unmethylated, GC-rich (60–70%) and does not show any suppression of the dinucleotide CpG. By contrast, bulk genomic DNA has a GC content of 40%, is methylated at CpG and has the CpG dinucleotide at only a quarter of the frequency expected from base composition. The depletion of CpGs is thought to be a result of deamination of 5-methylcytosine to thymine leading to mutation of CpGs to TpGs and CpAs. The excess of the latter two dinucleotides in the genome is equivalent to the deficiency of CpGs.

CpG islands have been found at the 5′ ends of all housekeeping genes and of a large proportion of genes with a tissue-restricted pattern of expression. Three examples are the desmin, hypoxanthine phosphoribosyl transferase (*HPRT*), and retinoblastoma genes (Fig. 1). The position of the island can be seen easily in each case by the dense clustering of CpG dinucleotides. Although the genes are of very different sizes, ranging between 8 kb and 177 kb, in all of them the CpG island includes the first exon of the gene. Thus, the CpG island denotes the 5′ end of the associated gene. Within the island, the high GC content and lack of methylation mean that sites for restriction enzymes which cut rarely elsewhere in the genome are clustered (reviewed in [3]). In the past ten years, this property of CpG islands has been exploited in the mapping and isolation of many genes.

Although CpG islands have been used extensively as signposts within the genome, comparatively little has been known until recently about their distribution, origin and maintenance. In this review, we will discuss some recent insights into these aspects of CpG islands. One property that might be important is that the CpG island chromatin is generally underacetylated, lacks histone H1 and has a nucleosome-free region [4]. This 'open' chromatin structure may allow, or be a consequence of, the interaction of transcription factors with gene promoters. Although all vertebrate species examined so far appear to have CpG islands, those from some fish species are markedly different in that they are 'GC-poor' but are still unmethylated and show no suppression of CpG [5]. This indicates that a high GC content is not a prerequisite for the maintenance of CpG islands.

Number of CpG islands and genes

By quantitating the proportion of genomic DNA contained in CpG islands, it has been estimated that 45 000 and 37 000 CpG islands exist in the human and mouse genomes, respectively [6•]. The database was consulted to determine what percentage of genes are associated with CpG islands, allowing the total number of genes in both species to be calculated. The estimate of 80 000 genes in each species falls into the range of 50 000–100 000 genes which is often quoted as an estimate of the number of genes in the human genome. Other recent estimates are 71 000 from genomic sequencing and 64 000 from expressed sequence tag analysis [7]. A survey of genes in the database found that CpG islands are associated with housekeeping genes and with 40% of genes with a tissue-restricted pattern of expression [8]. One conclusion from combining all this data is that half of human CpG island genes are housekeeping

Abbreviations

aprt—adenine phosphoribosyl transferase; ***HPRT***—hypoxanthine phosphoribosyl transferase; **RLGS**—restriction landmark genomic scanning; **Xa**—active X-chromosome; **Xi**—inactive X-chromosome.

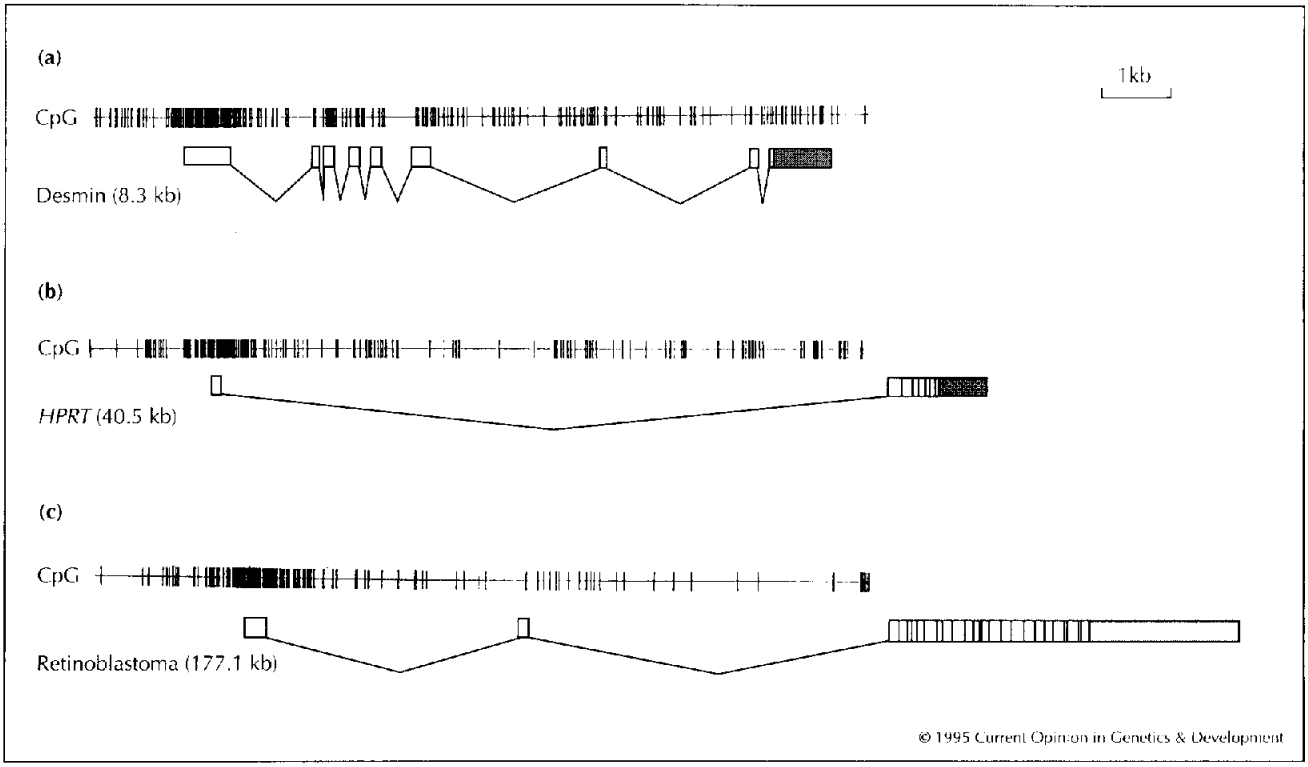

Fig. 1. Diagram showing the structure of three human CpG island genes of different sizes. Vertical lines show the positions of CpGs in the first 10 kb of **(a)** the desmin (EMBL hsdes01), **(b)** hypoxanthine phosphoribosyl transferase (*HPRT*; EMBL hshprt8a) and **(c)** retinoblastoma (EMBL L11910) genes. The locations of the exons are shown by boxes. Open and tinted portions denote translated and untranslated regions, respectively. Any exons not present in the first 10 kb of genomic DNA are shown fused together to the right. The total genomic length of each gene (in kb) is given in brackets.

genes and half have a tissue-restricted pattern of expression. Although the total number of genes is the same in the two species, fewer CpG island genes are found in mouse than in human [6•,9•]. Mouse CpG islands are also consistently less GC-rich than those in the human genome [9•]. In agreement with this, the number of unmethylated *Not*I sites (virtually all of which are in CpG islands) have been estimated in the human and mouse genomes to be 3240 and 2380, respectively [10]. The lower number in mouse reflects both the lower GC content and the smaller number of mouse CpG islands. It is thought that CpG islands associated with tissue-restricted genes are being 'lost' from the mouse genome, presumably by a process of methylation of the CpG island in the germline followed by deamination and mutation. Evidence exists for CpG islands being lost in the human genome as well, but the process is, for some reason, not as rapid as in the mouse.

Distribution of CpG islands

How are CpG islands distributed in the genome? An association of CpG islands with R bands has been noted for some time [11,12]. Chromosomes appear to made up of two sorts of domain which have reciprocal properties and are interspersed along the chromosome arms. The two can be visualized using a variety of staining techniques as either R bands or G bands (see the review by K Gardiner in this issue [pp 315–322]). R bands correspond to those chromosomal domains that are early-replicating, comparatively GC-rich and less condensed than those in G bands. A sub-set of R bands, the so-called T bands, have the highest GC content and are often found at telomeric locations. Of mapped genes, ~65% are found in T bands [12]. Evidence exists for chromosomal domains being made up of isochores (long regions of homogeneous base composition) [13]. There appears to be preferential hybridization of genes to the most GC-rich sort of isochore, which accounts for only 3.5% of the genome, and this maps principally to T bands [14]. Craig and Bickmore [15••] used a combination of preparative pulsed-field gel electrophoresis and fluorescence *in situ* hybridization to elegantly show that CpG islands are indeed found in the R band regions of the genome and that they are concentrated in the T bands. Fig. 2 shows the strikingly uneven distribution of CpG islands both along and between chromosomes; some chromosomes appear to be very rich in CpG islands (e.g. chromosomes 19 and 22), whereas others appear to have few islands (e.g. chromosome 18). These differences parallel the differences in the numbers of genes which have been assigned to different chromosomes and seem to highlight those regions which should prove most fruitful in gene searches [16]. The banding patterns seen on metaphase chromosomes may reflect functional domains which have linked activities and characteristics.

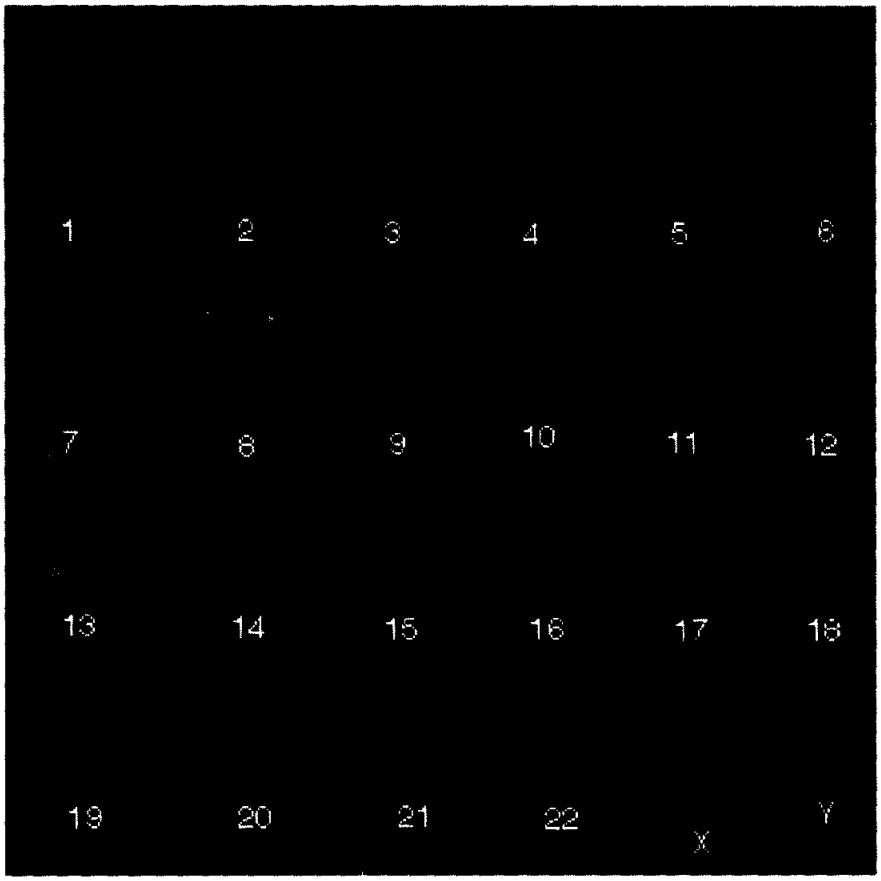

Fig. 2. Distribution of CpG islands in the human genome. Karyotype analysis of fluorescence *in situ* hybridization on human metaphase chromosomes hybridized with a biotinylated CpG island fraction prepared as in [28••] (Texas red). Bromodeoxyuridine (FITC/green) is incorporated into late-replicating DNA. The distribution of CpG islands both along and between chromosomes is strikingly uneven. Some chromosomes appear to be very rich in CpG islands (e.g. chromosomes 19 and 22), whereas others appear to have few islands (e.g. chromosome 18). The low incidence of yellow (overlapping red and green) signal indicates that there is very little overlap between CpG islands and late-replicating DNA. Picture courtesy of WA Bickmore.

Isolation of CpG islands and genes

One important approach in both the search for genes and the construction of long-range physical maps has been the isolation of CpG island DNA. The DNA adjacent to rare-cutter sites, which are concentrated in CpG islands, have been cloned from a number of regions of the genome [17–21]. Analysis of such clones has shown that they are derived from CpG islands. The clones detect expressed sequences and often contain di-, tri- or tetra-nucleotide repeats. They have been used successfully to identify a number of novel genes in large cloned genomic regions and to generate long-range physical maps of such regions [22–25]. The enzyme used most commonly in this sort of experiment is *Not*I. As sites for this enzyme are found in only a small sub-set of CpG islands, many CpG islands, and therefore genes, will be missed. An alternative strategy using computer analysis of the DNA immediately adjacent to *Eag*I and *Sac*II sites, both of which are found far more commonly in CpG islands, seems to be fairly successful in identifying potential CpG islands and coding regions [26]. A PCR-based method has been used to amplify fragments that flank rare-cutter sites in order to select transcribed sequences from large genomic clones [27].

All of the above methods rely on the use of enzymes that cut within islands, with the consequence that only fragments of CpG islands are cloned. Recently, a method was developed for cloning intact CpG islands from genomic DNA [28••]. Libraries of CpG islands contain sequences which can be used for both mapping studies and for the isolation of full-length cDNAs, thus providing a simple link between genomic DNA and gene transcripts. One advantage of CpG island libraries is that all clones in the library are represented equally, in contrast to cDNA libraries, in which representation depends on the transcriptional characteristics of the associated gene. Interestingly, the CpG island clones appear to contain very low numbers of dispersed repeats, which makes them excellent hybridization probes.

Origin and maintenance of CpG islands

With the exception of a few special cases, CpG islands are always unmethylated irrespective of the level or pattern of transcription of their associated genes. Why do they remain free of methylation in an otherwise heavily methylated genome? Three mechanisms have been suggested. One suggestion is that the DNA methyltransferase methylates GC-rich DNA poorly [29]. A second proposal is that CpG islands are methylated *de novo*, but the methylation is removed by an island-specific demethylating activity [30]. This is supported by evidence that partially methylated CpG islands lose methylation on transfection into fertilized mouse eggs or embryonic cells [30,31]. A third possibility is that factors are bound to CpG islands which deny access to DNA methyltransferase. Genomic footprinting across

the mouse adenine phosphoribosyl transferase (*aprt*) gene CpG island has showed that Sp1 sites positioned on one edge of the island are occupied by bound factors [32••]. In transgenic mouse experiments, intact *aprt* islands remained methylation-free, but if the Sp1 sites were deleted or mutated, the island became *de novo* methylated. A similar role of Sp1 sites has been found to be played at the hamster *aprt* gene CpG island using a transgenic mouse assay [33••]. The mechanism by which factors bound at peripherally located sites in CpG islands prevent CpG islands becoming methylated is not clear. Perhaps the factors antagonize DNA methyltransferase at a crucial developmental stage when the global methylation pattern is established and maintenance methylation retains that pattern thereafter.

Methylation of CpG islands

Some important exceptions are known to the general rule that CpG islands are unmethylated, the classic case being CpG islands on the inactive X-chromosome (Xi). When a CpG island becomes methylated, the associated gene is inactivated, probably through a combination of direct inhibition of transcription factor binding and because methylation directs chromatin into an inactive structure. This latter step may be mediated by methylated DNA binding proteins [34]. On the Xi, CpG island hypermethylation appears to stabilize the inactive state [35]. Examination of X-linked CpG islands in human and mouse has shown that different methylation patterns are found at CpG islands on the Xi, but in all cases, CpG islands on the active X-chromosome (Xa) are unmethylated. On the human Xi, the *PGK-1* CpG island is methylated at 60 of a possible 61 CpG sites [36], whereas the *HPRT* CpG island is heavily methylated except in the GC-box region, where either unmethylated or only partially methylated CpGs are found [37]. This GC-box region contains potential Sp1-binding sites, and *in vivo* footprints are found over this region on the Xa. For both *PGK-1* and *HPRT*, however, no *in vivo* footprints are found over the inactive CpG island. In the case of the *HPRT* island, it is possible that factors bound to the GC-box region protect that region from becoming methylated during X-inactivation. In contrast to the human *HPRT* island, methylation patterns at the mouse island have been found to be heterogeneous within both a tissue and clonal cell populations. The highest density of methylation coincides with transcriptional start sites [38]. The evidence suggests that although methylation appears to stabilize the inactive chromatin structure of Xi CpG islands, there is considerable heterogeneity in the density of methylation that is present.

The *Xist* gene, which is expressed exclusively from Xi, may be responsible for X inactivation. This gene has a CpG island which is methylated at the silent allele and unmethylated at the expressed allele [39•]. One interesting aspect of X inactivation is the preferential inactivation of the paternal X-chromosome, and therefore expression of the paternal *Xist* allele, in some mouse tissues, which may be an 'imprinting' effect [40]. Several autosomal genes have been found that have an imprinted pattern of expression (i.e. are expressed from only one parental chromosome) [41]. Methylation seems to play a role in imprinting, because mice deficient for DNA methyltransferase activity are unable to maintain the imprint at several genes [42]. Some imprinted genes have CpG islands that are methylated at the inactive allele, although it is still not clear whether this modification establishes the imprint or is involved only in its maintenance. Support for the latter idea comes from examination of the *U2af1-rs1* gene, imprinted expression of which exclusively from the maternal allele precedes methylation of the paternal CpG island [43•].

CpG islands associated with non-essential genes in established tissue culture cell lines have been found to be methylated [44] and it is possible that cells in which accidental methylation of islands led to gene inactivation may have been put at a selective advantage. Several lines of evidence suggest that an analogous mechanism can lead to inactivation of tumour-suppressor genes during tumorigenesis. This epigenetic silencing was first found for the retinoblastoma-susceptibility gene (*Rb-1*) [45] and the same mechanism has been found recently to be important for the inactivation of both the *VHL* tumour-suppressor gene [46] and the *bcr–abl* locus in chronic myelogenous leukaemia [47]. Inactivation of the oestrogen receptor gene by methylation of its CpG island in colorectal tumours has also been found, and re-introduction of the gene into carcinoma cell lines resulted in growth suppression [48•]. Interestingly, a direct correlation was also found between oestrogen receptor CpG island methylation and ageing. Methylation of this island in ageing colorectal cells could be a first step in the development of colorectal tumours.

Some CpG islands contain arrays of CCG repeats, and such CpG islands have been found at folate-sensitive fragile sites (see the review by GR Sutherland and RI Richards in this issue [pp 323–327]). In affected individuals, expansion of the CCG trinucleotide repeat array to >200 copies accompanied by methylation of the CpG island seems to be important for the expression of the fragile site [49]. Two fragile sites associated with mental retardation, *FRAXA* and *FRAXE*, are found on the distal long arm of the X-chromosome. Recently, two more fragile sites of this type have been isolated, *FRAXF*, which is from the same part of the X-chromosome [50,51], and an autosomal fragile site, *FRA16A* [52]. It is intriguing that all these fragile sites appear to arise by similar mechanisms. It will now be interesting to see whether others are found which also have expansions of methylatable trinucleotide arrays.

Use of the restriction landmark genomic scanning (RLGS) technique, which shows methylation profiles, suggests that the number of CpG island genes in which imprinted or developmentally regulated methylation occurs might be as high as 2% [53]. An imprinted gene from mouse chromosome 11 has been isolated using this technique [54•,55•]. This raises the possibility that methylation of CpG islands occurs *in vivo* other than

just on the Xi. Although only a small minority of CpG islands become methylated, this methylation may have important developmental consequences, because inactivation of the associated gene almost certainly results.

Conclusions

The majority of genes in the human genome are associated with CpG islands. These patches of GC-rich methylation-free DNA may be a consequence of chromatin structure at the 5′ ends of the genes. It is now clear that transcription factor binding is crucial for their maintenance and we look forward to the elucidation of the mechanism by which these sequences are kept methylation free. The structure of CpG islands allows them to be separated from the rest of the genome, facilitating the isolation of their associated genes. CpG island libraries are likely to have a significant impact on gene discovery in the next few years. Methylation of CpG islands appears to be important in gene silencing, in such processes as X-inactivation, imprinting and, perhaps, cancer.

Acknowledgements

We are grateful to the Wellcome Trust, Imperial Cancer Research Fund and the Howard Hughes Medical Institute for research support and we would like to thank Ian Jackson, Richard Meehan, Donald Macleod and Susan Tweedie for helpful comments on the manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Tykocinski ML, Max EE: **CG dinucleotide clusters in MHC genes and in 5′ demethylated genes.** *Nucleic Acids Res* 1984, **12**:4385–4396.

2. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D: **A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA.** *Cell* 1985, **40**:91–99.

3. Bickmore WA, Bird AP: **Use of restriction endonucleases to detect and isolate genes from mammalian cells.** *Methods Enzymol* 1992, **216**:224–244.

4. Tazi J, Bird A: **Alternative chromatin structure at CpG islands.** *Cell* 1990, **60**:909–920.

5. Cross S, Kovarik P, Schmidtke J, Bird A: **Nonmethylated islands in fish genomes are GC-poor.** *Nucleic Acids Res* 1991, **19**:1469–1474.

6. • Antequera F, Bird A: **Number of CpG islands and genes in human and mouse.** *Proc Natl Acad Sci USA* 1993, **90**:11995–11999.

This paper describes an attempt to estimate gene number in two mammals by measuring the fraction of CpG island DNA. Although mouse and man both appear to have 80 000 genes, the number of CpG islands in the mouse is significantly fewer.

7. Fields C, Adams MD, White O, Venter JC: **How many genes in the human genome?** *Nature Genet* 1994, **7**:345–346.

8. Larsen F, Gundersen G, Lopez R, Prydz H: **CpG islands as gene markers in the human genome.** *Genomics* 1992, **13**:1095–1107.

9. • Matsuo K, Clay O, Takahashi T, Silke J, Schaffner W: **Evidence for erosion of mouse CpG islands during mammalian evolution.** *Somatic Cell Mol Genet* 1993, **19**:543–555.

A thorough computer analysis estimated that fewer CpG islands exist in mouse than in man, and that mouse CpG islands are less GC rich.

10. Imoto H, Hirotsune S, Muramatsu M, Okuda K, Sugimoto O, Chapman VM, Hayashizaki Y: **Direct determination of *Not*I cleavage sites in the genomic DNA of adult mouse kidney and human trophoblast using whole-range restriction landmark genomic scanning.** *DNA Res* 1994, **1**:239–243.

11. Bickmore WA, Sumner AT: **Mammalian chromosome banding—an expression of genome organization.** *Trends Genet* 1989, **5**:144–148.

12. Holmquist GP: **Review article: chromosome bands, their chromatin flavors, and their functional features.** *Am J Hum Genet* 1992, **51**:17–37.

13. Bernardi G: **The isochore organization of the human genome.** *Annu Rev Genet* 1989, **23**:637–661.

14. Saccone S, De Sario A, Della Valle G, Bernardi G: **The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes.** *Proc Natl Acad Sci USA* 1992, **89**:4913–4917.

15. •• Craig JM, Bickmore WA: **The distribution of CpG islands in mammalian chromosomes.** *Nature Genet* 1994, **7**:376–382.

This paper presents a conclusive demonstration that CpG islands are located in R bands and are particularly enriched in the subset of these known as T bands. This study also shows up dramatically different numbers of CpG islands, and therefore of genes, on different chromosomes (see Fig. 2).

16. Antonarakis SE: **Genome linkage scanning: systematic or intelligent?** *Nature Genet* 1994, **8**:211–212.

17. Allikmets RL, Kashuba VI, Pettersson B, Gizatullin R, Lebedeva T, Kholodnyuk ID, Bannikov VM, Petrov N, Zakharyev VM, Winberg G *et al.*: ***Not*I linking clones as a tool for joining physical and genetic maps of the human genome.** *Genomics* 1994, **19**:303–309.

18. Zhu YM, Cantor CR, Smith CL: **DNA sequence analysis of human chromosome 21 *Not*I linking clones.** *Genomics* 1993, **18**:199–205.

19. Ten Hoeve J, Morris C, Poustka A, Groffen J, Heisterkamp N: **Isolation of *Not*I sites from chromosome 22q11.** *Genomics* 1993, **18**:588–597.

20. Tribioli C, Tamanini F, Patrosso C, Milanesi L, Villa A, Pergolizzi R, Maestrini E, Rivella S, Bione S, Mancini M *et al.*: **Methylation and sequence analysis around *Eag*I sites: identification of 28 new CpG islands in Xq24–Xq28.** *Nucleic Acids Res* 1992, **20**:727–733.

21. Wang J-Y, Zabarovsky ER, Talmadge C, Berglund P, Chan KWK, Pokrovskaya ES, Kashuba VI, Zhen D-K, Boldog F, Zabarovskaya VI *et al.*: **Somatic cell hybrid panel and *Not*I linking clones for physical mapping of human chromosome 3.** *Genomics* 1994, **20**:105–113.

22. Gessler M, Poustka A, Cavenee W, Neve RL, Orkin SH, Bruns GAP: **Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping.** *Nature* 1990, **343**:774–778.

23. Palmieri G, Romano G, Ciccodicola A, Casamassimi A, Campanile C, Esposito T, Cappa V, Lania A, Johnson S, Reinbold R *et al.*: **YAC contig organisation and CpG island analysis in Xq28.** *Genomics* 1994, **24**:149–158.

24. Lee W-C, Salido E, Yen PH: **Isolation of a new gene GS2 (DX1283E) from a CpG island between *STS* and *KAL1* on Xp22.3.** *Genomics* 1994, **22**:372–376.

25. Tribioli C, Mancini M, Plassart E, Bione S, Rivella S, Sala C, Torri G, Toniolo D: **Isolation of new genes in distal Xq28: transcriptional map and identification of a human homologue of the ARD1 N-acetyl transferase of *Saccharomyces cerevisiae*.** *Hum Mol Genet* 1994, **3**:1061–1067.

26. John RM, Robbins CA, Myers RM: **Identification of genes within CpG-enriched DNA from human chromosome 4p16.3.** *Hum Mol Genet* 1994, **3**:1611–1616.

27. Valdes JM, Tagle DA, Collins FS: **Island rescue PCR: a rapid and efficient method for isolating transcribed sequences from yeast artificial chromosomes and cosmids.** *Proc Natl Acad Sci USA* 1994, **91**:5377–5381.

28. •• Cross SH, Charlton JA, Nan X, Bird AP: **Purification of CpG islands using a methylated DNA binding column.** *Nature Genet* 1994, **6**:236–244.
Describes a novel procedure for the preparation of intact CpG islands using a methylated DNA binding column. The resulting clones detected transcribed sequences and were remarkably free of repeats.

29. Carotti D, Palitti F, Lavia P, Strom R: ***In vitro* methylation of CpG-rich islands.** *Nucleic Acids Res* 1989, **17**:9219–9229.

30. Frank D, Keshet I, Shani M, Levine A, Razin A, Cedar H: **Demethylation of CpG islands in embryonic cells.** *Nature* 1991, **351**:239–241.

31. Choi Y-C, Chae C-B: **Demethylation of somatic and testis-specific histone *H2A* and *H2B* genes in F9 embryonal carcinoma cells.** *Mol Cell Biol* 1993, **13**:5538–5548.

32. •• Macleod D, Charlton J, Mullins J, Bird AP: **Sp1 sites in the mouse *aprt* gene promoter are required to prevent methylation of the CpG island.** *Genes Dev* 1994, **8**:2282–2292.
A thorough analysis of the mouse *aprt* gene CpG island by genomic sequencing and *in vivo* footprinting established that bound factors are confined to the 5′ edge of the island. Mutation of the factor-binding sites abolished the methylation-free state of the island in a transgenic assay.

33. •• Brandeis M, Frank D, Keshet I, Siegfried Z, Mendelsohn M, Nemes A, Temper V, Razin A, Cedar H: **Sp1 elements protect a CpG island from *de novo* methylation.** *Nature* 1994, **371**:435–438.
Identification of a crucial role for intact Sp1 sites in maintaining the methylation-free status of the hamster *aprt* gene. Sp1 sites were able to reduce methylation in their vicinity when introduced into other sequence contexts.

34. Meehan R, Lewis J, Cross S, Nan X, Jeppesen P, Bird A: **Transcriptional repression by methylation of CpG.** *J Cell Sci* 1992, **16(suppl)**:9–14.

35. Riggs AD, Pfeifer GP: **X-chromosome inactivation and cell memory.** *Trends Genet* 1992, **8**:169–174.

36. Pfeifer GP, Tanguay RL, Steigerwald SD, Riggs AD: ***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* 1990, **4**:1277–1287.

37. Hornstra IK, Yang TP: **High-resolution methylation analysis of the human hypoxanthine phosphoribosyltransferase gene 5′ region on the active and inactive X chromosomes: correlation with binding sites for transcription factors.** *Mol Cell Biol* 1994, **14**:1419–1430.

38. Park J-G, Chapman VM: **CpG island promoter region methylation patterns of the inactive-X-chromosome hypoxantine phosphoribosyltransferase (*Hprt*) gene.** *Mol Cell Biol* 1994, **14**:7975–7983.

39. • Norris DP, Patel D, Kay GF, Penny GD, Brockdorff N, Sheardown SA, Rastan S: **Evidence that random and imprinted *Xist* expression is controlled by preemptive methylation.** *Cell* 1994, **77**:41–51.
A provocative study suggesting that X-chromosome inactivation is imprinted by methylation of the presumptive Xi at the *Xist* CpG island. This study resurrects the possibility that DNA methylation has a primary causal role in the X-inactivation process.

40. Kay GF, Barton SC, Surani MA, Rastan S: **Imprinted and X-chromosome counting mechanisms determine *Xist* expression in early mouse development.** *Cell* 1994, **77**:639–650.

41. Efstratiadis A: **Parental imprinting of autosomal mammalian genes.** *Curr Opin Genet Dev* 1994, **4**:265–280.

42. Li E, Beard C, Jaenisch R: **Role for DNA methylation in genomic imprinting.** *Nature* 1993, **366**:362–365.

43. • Hatada I, Kitagawa K, Yamaoka T, Wang XD, Arai Y, Hashido H, Ohishi S, Masuda J, Ogata J, Mukai T: **Allele-specific methylation and expression of an imprinted *U2af1-rs1* (*SP2*) gene.** *Nucleic Acids Res* 1995, **23**:36–41.
A study of the transcriptional and methylation status of an imprinted gene shows that imprinted expression precedes methylation differences. In this case, therefore, methylation would appear to be a secondary consequence of imprinting.

44. Antequera F, Boyes J, Bird A: **High levels of *de novo* methylation and altered chromatin structure at CpG islands in cell-lines.** *Cell* 1990, **62**:503–514.

45. Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B: **Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma.** *Hum Genet* 1989, **83**:155–158.

46. Herman JG, Latif F, Weng Y, Lerman MI, Zbar B, Liu S, Samid D, Duan D-SR, Gnarra JR, Linehan WM, Baylin SB: **Silencing of the *VHL* tumor-suppressor gene by DNA methylation in renal carcinoma.** *Proc Natl Acad Sci USA* 1994, **91**:9700–9704.

47. Zion M, Ben-Yehuda D, Avraham A, Cohen O, Wetzler M, Melloul D, Ben Neriah Y: **Progressive *de novo* methylation at the *bcr–abl* locus in the course of chronic myelogenous leukemia.** *Proc Natl Acad Sci USA* 1994, **91**:10722–10726.

48. • Issa J-PJ, Ottaviano YL, Celano P, Hamilton SR, Davidson NE, Baylin SB: **Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon.** *Nature Genet* 1994, **7**:536–540.
Describes progressive methylation of a normally unmethylated CpG island during ageing and tumorigenesis. Some evidence is presented that loss of expression of the oestrogen receptor gene accelerates cell growth.

49. Willems PJ: **Dynamic mutations hit double figures.** *Nature Genet* 1994, **8**:213–215.

50. Parrish JE, Oostra BA, Verkerk AJMH, Richards CS, Reynolds J, Spikes AS, Shaffer LG, Nelson DL: **Isolation of a GCC repeat showing expansion in FRAXF, a fragile site distal to FRAXA and FRAXE.** *Nature Genet* 1994, **8**:229–235.

51. Ritchie RJ, Knight SJL, Hirst MC, Grewal PK, Bobrow M, Cross GS, Davies KE: **The cloning of *FRAXF*: trinucleotide repeat expansion and methylation at a third fragile site in distal Xqter.** *Hum Mol Genet* 1994, **3**:2115–2121.

52. Nancarrow JK, Kremer E, Holman K, Eyre H, Doggett NA, Le Paslier D, Callen DF, Sutherland GR, Richards RI: **Implications of *FRA16A* structure for the mechanism of chromosomal fragile site genesis.** *Science* 1994, **264**:1938–1941.

53. Kawai J, Hirotsune S, Hirose K, Fushiki S, Watanabe S, Hayashizaki Y: **Methylation profiles of genomic DNA of mouse developmental brain detected by restriction landmark genomic scanning (RLGS) method.** *Nucleic Acids Res* 1993, **21**:5604–5608.

54. • Hatada I, Sugama T, Mukai T: **A new imprinted gene cloned by a methylation-sensitive genome scanning method.** *Nucleic Acids Res* 1993, **21**:5577–5582.
Describes the use of the ingenious RLGS method to isolate a novel imprinted gene by detection of differential methylation. This is the first time that DNA methylation has been used to find an imprinted gene.

55. • Hayashizaki Y, Shibata H, Hirotsune S, Sugino H, Okazaki Y, Sasaki N, Hirose K, Imoto H, Okuizumi H, Muramatsu M *et al.*: **Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method.** *Nature Genet* 1994, **6**:33–40.
Using the RLGS technique, an imprinted gene was identified and cloned. Surprisingly, the gene is the same as that cloned using the same method by another group (see annotation [54•]).

SH Cross and AP Bird, Institute of Cell and Molecular Biology, University of Edinburgh, King's Buildings, Edinburgh EH9 3JR, Scotland, UK.

Author for correspondence: SH Cross. E-mail: Sally.Cross@ed.ac.uk