REPORTS

- M. C. Fishman and E. N. Olson, *Cell* **91**, 153 (1997);
 M. C. Fishman and K. R. Chien, *Development* **124**, 2099 (1997).
- R. Zeller, K. D. Bloch, B. S. Williams, R. J. Arceci, C. E. Seidman, Genes Dev. 1, 693 (1987).
- K. R. Chien et al., Annu. Rev. Physiol. 55, 77 (1993);
 G. E. Lyons, Trends Cardiovasc. Med. 3, 184 (1994).
- J. Chen et al., J. Biol. Chem. 273, 1252 (1998); C. M. Pawloski-Dahm et al., Circulation 97, 1508 (1998).
- 6. Chick *Irx4* cDNA was isolated by low-stringency hybridization of an E6-E8 random-primed retinal cDNA library with probes derived from a mouse expressed sequence tag (EST) clone (GenBank accession number R46202). A full-length cDNA was obtained by screening an oligo(dT)-primed whole-embryo cDNA library (stages 12 to 15), with the original *Irx4* clone as a probe. Mouse *Irx4* cDNA was isolated similarly by low-stringency screening with chick *Irx4* as a probe from an E10 mouse heart library (Stratagene). The DNA sequence was determined on both strands by an automated DNA sequencer (Applied Biosystems).
- J. L. Gomez-Skarmeta, R. D. del Corral, E. de la Calle-Mustienes, D. Ferre-Marco, J. Modolell, *Cell* 85, 95 (1996).
- H. McNeill, C. H. Yang, M. Brodsky, J. Ungos, M. A. Simon, *Genes Dev.* **11**, 1073 (1997).
- 9. B. G. Bruneau et al., unpublished data.
- 10. Probe synthesis and in situ hybridization procedures were performed as in (25, 26). Proteinase K was used at 1 $\mu g/ml,$ 2.5 $\mu g/ml,$ and 10 $\mu g/ml$ for embryos of stages 5 to 9, stages 10 to 14, or stage 15 and older, respectively. To avoid cross-hybridization to other Iroquois family members, we generated a template by Bbs I digestion of the *Irx4* cDNA cloned in pBluescript SK(-)for synthesis of an antisense riboprobe (1.2 kb) without the homeobox sequence. In the misexpression experiments, a probe specific for the 5' untranslated region (UTR) of chick Irx4 was used. Probes specific for VMHC1 and AMHC1 were created on the basis of their least conserved 3' sequence including the 3' UTR (12, 13). To prepare templates for making VMHC1 and AMHC1 probes, we performed reverse transcription-polymerase chain reaction (RT-PCR) using total RNA from stage-17 chick embryos (primers for VMHC1: 5'-CTACAAACAC-CAAGCAGAGGAAGC, 3'-GGTTTGCACTTGATATT-TATTTTTGTA; and for AMHC1: 5'-GCGGGTCCAGC-TTCTCCACTCC, 3'-GCACCTTGACACGCCGCTCTGAC-TT).
- 11. Z.-Z. Bao, B. G. Bruneau, J. G. Seidman, C. E. Seidman, C. L. Cepko, unpublished data.
- 12. J. G. Bisaha and D. Bader, *Dev. Biol.* **148**, 355 (1991). 13. K. E. Yutzey, J. T. Rhee, D. Bader, *Development* **120**,
- 871 (1994). 14. G. F. Wang, J. Nikovits, W. M. Schleinitz, F. E. Stock-
- dale, Mol. Cell. Biol. **18**, 6023 (1998). 15. _____, J. Biol. Chem. **271**, 19836 (1996).
- 16. Full-length chick and mouse Irx4 cDNAs were fused in frame with the initiator ATG of the pSLAX-13 shuttle vector (25). The cDNAs were then transferred into the RCASBP(A) retroviral vector. To increase the infection rate, we made pseudotyped viruses with the vesicular stomatitis virus G coat (VSV-G) [M. Logan, S. M. Pagan-Westphal, D. M. Smith, L. Paganessi, C. J. Tabin, Cell 94, 307 (1998)]. The viral titers for RCAS with chick Irx4 were one-tenth to one-twentieth those for RCAS with mouse Irx4, presumably due to a high GC content in the 3' end of the chick Irx4. Because a high infection rate is critical for the misexpression experiments, we used RCAS with the mouse Irx4 insert. Because the mouse and chick Irx4 proteins are 71% homologous and the mouse Irx4 has very similar expression patterns in the developing heart, the mouse Irx4 is likely to have the same effect as the chick Irx4 in the misexpression experiments. A putative dominant negative form of Irx4 (RCAS-H+en^r) was generated by fusion of a truncated chick Irx4 fragment including the homeodomain (from amino acid 109 to 220) with the repressor domain from the Engrailed protein. Control virus RCAS-enr contains the Engrailed repressor domain only. RCAS-H+enr and RCAS-enr viruses were made as in (25), and both titers were about 5×10^8 plaque-forming units per milliliter.
- Fertile White Leghorn chicken eggs (SPAFAS, Norwich, CT) were incubated and staged (27). To visu-

alize the embryo, we introduced 1 to 10 μ l of India ink, diluted 1:50 in Ringer's buffer, beneath the blastoderm. Concentrated retroviral stocks were pressure injected into the cardiogenic mesoderm at stage 7–8 with a glass micropipette. The injected embryos were returned to the incubator until further analyses.

- P. Badiani, P. Corbella, D. Kioussis, J. Marvel, K. Weston, *Genes Dev.* 8, 770 (1994); F. L. Conlon, S. G. Sedgwick, K. M. Weston, J. C. Smith, *Development* 122, 2427 (1996); T. Furukawa, E. M. Morrow, C. L. Cepko, *Cell* 91, 531 (1997).
- 19. D. Yelon and D. Stainier, unpublished data.
- K. Yutzey, M. Gannon, D. Bader, *Dev. Biol.* **170**, 531 (1995);
 T. Mikawa and D. A. Fischman, *Annu. Rev. Physiol.* **58**, 509 (1996);
 D. Y. Stainier, R. K. Lee, M. C. Fishman, *Development* **119**, 31 (1993).
- J. Satin, S. Fujii, R. DeHaan, Dev. Biol. 129, 103 (1988).
- P. J. Gruber, S. W. Kubalak, K. R. Chien, *Development* 125, 4427 (1998).
- E. J. Bellefroid *et al.*, *EMBO J.* **17**, 191 (1998); J. L. Gomez-Skarmeta, A. Glavic, E. de la Calle-Mustienes, J. Modolell, R. Mayor, *ibid.*, p. 181.

- 24. J. L. Gomez-Skarmeta and J. Modolell, *Genes Dev.* **10**, 2935 (1996).
- R. D. Riddle, R. L. Johnson, E. Laufer, C. Tabin, *Cell* 75, 1401 (1993).
- Z. Z. Bao and C. L. Cepko, J. Neurosci. 17, 1425 (1997); L. Tessarollo, L. Nagarajan, L. F. Parada, Development 115, 11 (1992).
- 27. V. Hamburger and H. L. Hamilton, J. Morphol. 88, 49 (1951).
- Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; F, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.
- 29. We thank K.-H. Lee for his help with histology; D. Schulte and D. Smith for control viruses; D. Schulte, J. Lin, A. Lassar, K.-H. Lee, M. Marvin, and M. Logan for critical review of the manuscript; and D. Yelon and D. Stainier for sharing unpublished results. Supported by the Howard Hughes Medical Institute, NIH (Z.-Z.B.), the American Heart Association, and the Medical Research Council of Canada (B.G.B.).

23 November 1998; accepted 22 January 1999

Nonmethylated Transposable Elements and Methylated Genes in a Chordate Genome

Martin W. Simmen,¹* Sabine Leitgeb,¹* Jillian Charlton,¹ Steven J. M. Jones,² Barbara R. Harris,² Victoria H. Clark,¹ Adrian Bird¹†

The genome of the invertebrate chordate *Ciona intestinalis* was found to be a stable mosaic of methylated and nonmethylated domains. Multiple copies of an apparently active long terminal repeat retrotransposon and a long interspersed element are nonmethylated and a large fraction of abundant short interspersed elements are also methylation free. Genes, by contrast, are predominantly methylated. These data are incompatible with the genome defense model, which proposes that DNA methylation in animals is primarily targeted to endogenous transposable elements. Cytosine methylation in this urochordate may be preferentially directed to genes.

DNA methylation in the dinucleotide sequence 5'-CpG can silence transcription. The genome defense model (1) posits that the primary role of methylation in animal genomes is to repress potentially damaging transposition of endogenous elements. By analogy with fungal systems (2), the elements are hypothesized to be targets for methylation because of their repetition in the genome (3). The hypothesis is difficult to test through analysis of the globally methylated mammalian genome. We therefore studied the specificity of methylation in a fractionally methylated genome belonging to the sea squirt Ciona intestinalis, an invertebrate member of the chordate phylum. Like most invertebrate genomes, that of C. intestinalis contains comparable amounts of methylated and nonmethylated DNA (4). We reasoned that any bias in the distribution of transposons or genes between the two fractions should therefore be readily detectable.

Three cosmids containing C. intestinalis genomic DNA were studied in detail (5). The cosmids were sequenced, and the locations of likely protein-coding regions were determined by GENEFINDER exon-prediction software (6) and database homology searches (Fig. 1, A through C). Putative proteins encoded by 10 of 13 potential genes showed similarity to known proteins. A systematic search for repetitive elements among the cosmid sequences and 1486 short random genomic sequences (5) identified four transposable elements belonging to recognizable families (7): a gypsy/Ty3-like long terminal repeat (LTR) retrotransposon (Cigr-1), a long interspersed element (LINE)-like element (Cili-1), a miniature inverted repeat (Cimi-1), and a composite short interspersed element (SINE) (Cics-1) with consensus RNA po-

¹Institute of Cell and Molecular Biology, University of Edinburgh, The King's Buildings, Edinburgh EH9 3JR, UK. ²The Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK.

^{*}These authors contributed equally to this work. †To whom correspondence should be addressed. Email: A.Bird@ed.ac.uk

lymerase III promoter A and B box sites (Table 1) (8). Cigr-1, Cimi-1, and Cics-1 elements occurred in the cosmids, the latter two being very abundant (Fig. 1, A through C).

Mosaic methylation of invertebrate genomes has been inferred previously (4, 9), but domain maps have not been reported. DNA methylation was analyzed in genomic DNA from C. intestinalis by means of methylationsensitive restriction endonucleases (10). The 39-kb cicos2 and cicos46 inserts each contained a domain of several kilobase pairs that is methvlated in the genome and flanked by nonmethylated DNA (Fig. 1, A and B). Most of the cicos41 insert was methylated in the genome, but the right extremity was nonmethylated (Fig. 1C). Methylation patterns in sperm and carcass DNA were similar. Projection of the methylation patterns onto a map of CpG frequency showed striking coincidence between methylated domains and domains of CpG deficiency (Fig. 1, A through C). Given strong evidence that CpG deficiency is due to hypermutability of methyl-CpG (11), the data suggest that the observed methylation patterns are evolutionarily stable. The *C. intestinalis* genome appears to be equally partitioned between methylated and nonmethylated domains, based on the separation of methylated and nonmethylated DNA after restriction endonuclease treatment (4, 8). The presence of methylation at about 25% of genomic CpGs (4) is compatible with this estimate, as the CpG density of methylated domains is about half that of nonmethylated domains.

REPORTS

The *Cigr-1* element was identified in cicos41, where it is flanked by methylated DNA (Fig. 1C). Methylation of the element (Fig. 2A) was investigated by Southern (DNA) blotting and bisulfite sequencing (12). We estimated that there were approximately 75 copies of Cigr-1 per genome, based on Southern blots (Fig. 2C) and element frequency in a database of C. intestinalis DNA sequences (Table 1). All tested CpG restriction sites were nonmethylated in most Cigr-1 copies (Fig. 2, B and D). Because transposition depends on transcription from the LTR promoter (13), we tested methylation of all 20 CpGs in this region and found that 8 out of 10 sampled elements were methylation free, with 2 showing low-level methylation (Fig. 2B). Comparison of the genomic contexts of Cigr-1 elements in four individuals showed a high degree of heterogeneity, which suggests that Cigr-1 is transpositionally active (Fig. 2C). The presence of an internal 1.9-kb Pst I fragment confirmed that the probes detect bona fide Cigr-1 elements and that heterogeneity is not due to incomplete digestion with Pst I. A Cigr-1 transcript was detected by reverse transcriptase polymerase chain reac-



Fig. 1. Genes, repetitive elements, and endogenous methylation patterns in three cosmids containing *C. intestinalis* genomic DNA. (**A** through **C**) Diagrams of cosmids cicos2, cicos46, and cicos41, showing (from the top down) (i) repeats: the repetitive element families *Cigr-1* (yellow), *Cimi-1* (red), and *Cics-1* (blue). (ii) Genes: predicted coding sequences (linked exon boxes) on 5' (above the line) and 3' (below the line) strands. C2.3 (A) and C46.3 (B) correspond to homologs of AND-1 (EMBL accession number X98884) and a zinc-finger protein gene, respectively. (iii) Nonmethylated sites for a range of methyl-CpG-sensitive restriction enzymes project above the line, and





Fig. 2. Nonmethylated transposable elements in the *C. intestinalis* genome. (**A**) Diagram of the *Cigr-1* element, showing LTRs, an open reading frame (ORF), and hybridization probes (RT1 and RT2). (**B**) The upper map shows the absence of methylation at sites in *Cigr-1* for CpG methylationsensitive restriction endonucleases (open lollipops). Hpa II sites detected by probing with RT1 are labeled H. The lower maps show the analysis of methylation at all CpGs in the LTR promoter region (nucleotides 19 through 405) of *Cigr-1* by bisulfite sequencing (12). Solid lollipops indicate methylated CpGs. (**C**) Variable genomic locations of *Cigr-1* in four *C. intestinalis* individuals collected from the same location. Pst I digests were probed with RT2 to detect fragments extending downstream from the rightmost Pst I site in *Cigr-1* to the next site in flanking genomic DNA, or with RT1 to show the internal 1.9-kb band. (**D**) The absence of methylation at Hpa II sites in *Cigr-1*. DNA from carcass and sperm were digested with no enzyme (-), Msp I (M), and Hpa II (H), and blots were probed with RT1. (**E**) Identical Hpa II and Msp I fragment patterns in genomic DNA probed with part of the LINE-like element *Cili-1*. An absence of methylation is also seen at Hha I (Hh) sites. (**F**) Analysis of methylated and nonmethylated sites at seven copies of the miniature inverted repeat *Cimi-1* by bisulfite sequencing. Missing CpG sites are due to sequence heterogeneity. Solid lollipops indicate methylated sites; open lollipops indicate nonmethylated sites.

tion (8). Thus, an apparently active multicopy transposon escapes DNA methylation in *C. intestinalis.*

Similar results were obtained with a probe against *Cili-1*. Hpa II and Msp I patterns were identical and the element was susceptible to

Hha I, indicating the absence of methylation (Fig. 2E). The short element *Cimi-1* is abundant in both methylated and nonmethylated domains of the cosmids (Fig. 1, A through C). Bisulfite sequencing confirmed that many copies are nonmethylated at all CpGs but that methylated

Table 1. Major repetitive elements in the genome of *C. intestinalis.* Elements were identified according to (7). The copy number estimate is based on extrapolation from the number of copies found in 0.88 Mbp of genomic sequence in 1486 fragments of mean size 592 bp, assuming a haploid genome size of 162 Mbp.

Name	Description	Size (bp)	Copy no.
Cigr-1	Gypsy/Ty-3–like LTR retrotransposon	4226	75
Cimi-1	Miniature terminal inverted repeat SINE	193	17,000
Cics-1	Composite tRNA-derived SINF	302–333	40,000
Cili-1	LINE-like element	7000*	50

*No complete copies were found in the cosmids, so length could not be determined. The quoted size is that of a typical mammalian LINE.

copies also occur (Fig. 2F). The abundant element *Cics-1* occurs in methylated and nonmethylated domains of the cosmids (Fig. 1, A through C). Correspondingly, five out of eight genomic copies of *Cics-1* were nonmethylated by bisulfite analysis (8). In the sequenced cosmids, six tested CpGs were within *Cics-1* elements and all conformed to the methylation status of the surrounding domain. Four of the elements were intergenic (Fig. 1, A through C) and one was within an intron (see Fig. 3B). These findings suggest that the methylated copies of these elements are those that happen to lie within methylated domains.

Most of the C. intestinalis genes found in the three cosmids (9 out of 13) occur in methylated domains (Fig. 1, A through C). The methylated domain in cicos2 covers the central region of the AND1-related gene but does not extend to either the promoter or the last exon (Fig. 1A, gene C2.3). Similarly, the downstream transcribed region and upstream flanks of the zinc finger protein gene in cicos46 are methylated (Fig. 1B, gene C46.3). The 5' end of the gene is within a nonmethylated, CpGrich sequence patch that resembles a mammalian CpG island. All seven genes in cicos41 are within heavily methylated DNA (Fig. 1C). The fact that exons of genes can be heavily methylated was confirmed by bisulfite sequencing of exon 6 of a guanylate-cyclase-like gene (Fig. 3B). We also examined methylation of 10 C. intestinalis cDNAs with strong homology to known genes (Fig. 3A) (8). Eight hybridized to highly methylated sequences on the basis of Hpa II/Msp I susceptibility and two were nonmethylated. The two independent estimates suggest that about three-quarters of C. intestinalis genes, including many housekeeping genes, lie within methylated DNA.

The origin and significance of mosaic methylation patterns in invertebrate genomes are uncertain. No correlation between methylation and expression of genes is yet evident, although





Fig. 3. Most C. intestinalis genes lie within methylated DNA. (A) Comparison of Msp I (M) and Hpa II (H) digestion patterns for five cDNAs that have good matches to known gene transcripts. The matches are to translation initiation factor EIF-5A, isoleucyl-tRNA synthetase (ile.synth), CPI anchor

biosynthesis protein PIG-A, epidermal surface antigen (epi.s.a.), and chymotrypsinogen (chymo). All but the chymotrypsinogen probes hybridize to Hpa II–resistant DNA sequences. (B) Methylation at CpGs across exon 6 of a guanylate cyclase–like gene identified in the cosmid cicos1 (EMBL accession number Z80904), as determined by bisulfite sequencing (12). Solid and open lollipops represent methylated and nonmethylated sites, respectively. The site labeled by an asterisk lies within a *Cics-1* element.

promoter methylation has not been specifically studied. Methylated genes of several invertebrates are clearly transcribed, whereas nonmethylated genes can be silent (4). If the primary function of methylation in these organisms is to repress transcription, then a role in reducing transcriptional interference (14) is compatible with the data. Active gene domains often generate incorrectly initiated transcripts whose synthesis may interfere with authentic transcription (15). Methylation of genes might reduce interference and thereby focus initiation on the genuine promoter. Why then are some genes nonmethylated? Perhaps genes that are highly expressed in a few cell types, which appear to be overrepresented among the nonmethylated class (4), have robust promoters that are less susceptible to transcriptional interference.

Genome defense can explain the role of DNA methylation in preventing sequence duplication in certain fungi (2). The mechanism is not universal, however, as the nonmethylated genomes of both Drosophila and Caenorhabditis elegans harbor transposons. The present study shows that where methylation is present in an invertebrate genome, it is not necessarily targeted to transposable elements. DNA sequence repetition does not provoke methylation in C. intestinalis, as multicopy transposons and genes [for example, ribosomal RNA genes (4)] often are nonmethylated, whereas single-copy genes are methylated. Thus, for invertebrates, the idea that transposition is controlled by DNA methylation lacks supporting evidence [see also (16)]. In the human genome, transposon sequences are heavily methylated (3), but methylation affects most genomic DNA (about 85% of those CpGs that are not within CpG islands), including exons of genes (17). Also, several retrotransposons are reportedly nonmethylated for many cell generations in mammalian germ cells and early embryos, where protection against transposition might be thought to be important (18). In mammals, therefore, the evidence for a relationship between DNA methylation and the restraint of endogenous elements remains inconclusive.

References and Notes

- J. A. Yoder, C. P. Walsh, T. H. Bestor, *Trends Genet.* 13, 335 (1997).
- E. U. Selker, *ibid.*, p. 296; E. U. Selker and P. W. Garrett, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6870 (1988);
 J.-L. Rossignol and G. Faugeron, *Experientia* **50**, 307 (1994);
 T. Freedman and P. J. Pukkila, *Genetics* **135**, 357 (1993);
 P. A. Whittaker, A. McLachlan, N. Hardman, *Nucleic Acids Res.* **9**, 801 (1981);
 C. Goyon, J. L. Rossignol, G. Faugeron, *ibid.* **24**, 3348 (1996).
- M. C. Kricker, J. W. Drake, M. Radman, *Proc. Natl. Acad. Sci. U.S.A.* 89, 1075 (1992); T. H. Bestor and B. Tycko, *Nature Genet.* 12, 363 (1996).
- S. Tweedie, J. Charlton, V. Clark, A. Bird, Mol. Cell Biol. 17, 1469 (1997).
- Specimens of C. intestinalis were obtained from Plymouth Marine Biology Labs, UK, and from Glasgow University Marine Biological Station at Millport, UK. The construction and sequencing of the cosmid and 1486 plasmid clones containing C. intestinalis genomic DNA has been described [M. W. Simmen, S. Leitgeb, V. H. Clark, S. J. M. Jones, A. Bird, Proc. Natl. Acad. Sci. U.S.A. 95, 4437 (1998)]. The cosmid European Molecular Biology Laboratory (EMBL/GenBank entry names and accession numbers are as follows: cicos2 (Z79640), ci cos41 (Z83760), and cicos46 (Z83861). The short genomic fragments have EMBL/GenBank accession numbers ranging from AJ226133 to AJ227618.
- 6. P. Green and L. Hillier, unpublished software. 7. Repetitive elements were identified by two methods. Known elements were detected by BLAST [S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, J. Mol. Biol. 215, 403 (1990)] searches against the nonredundant nucleotide and protein databases of the National Center for Biotechnology Information. To detect novel elements, the 1486 C. intestinalis genomic fragments were used to construct a BLAST database; querying the cosmids against this database revealed regions possessing multiple hits, which are indicative of repetitive elements. The resulting elements were classified on the basis of detailed sequence analysis (M. W. Simmen, in preparation). The Cigr-1 sequence is given by the reverse complement of bases 30449 through 34674 of cicos41. Representative Cimi-1 and Cics-1 elements are found in bases 34761 through 34957 of cicos2 and bases 331 through 646 of AJ226662, respectively. Sections of

the putative *Cili-1* element lie in sequences AJ226190, AJ226259, and AJ226870. To plot the location of the repeats on the cosmids, the above sequences (or an appropriate consensus sequence) were used to form a small BLAST database. Cosmid hits to this database passing an MSPcrunch [E. L. L. Sonnhammer and R. Durbin, *Comput. Appl. Biosci.* **10**, 301 (1994)] screen requiring more than 70% sequence identity were accepted.

- 8. M. W. Simmen et al., data not shown.
- A. P. Bird, M. H. Taggart, B. A. Smith, *Cell* **17**, 889 (1979); A. P. Bird and M. H. Taggart, *Nucleic Acids Res.* **8**, 1485 (1980).
- Methylation of genomic DNA corresponding to cosmid inserts was analyzed by digestion of carcass or sperm DNA with CpC methylation-sensitive enzymes (Aci I, Bst UI, Cla I, Hha I, Hpa II, Nru I, Sa I I, Sma I, and Xho I) in combination with other restriction enzymes, followed by probing with a variety of single-copy sequences from each cosmid. Detailed methylation maps of the cosmids are available on request. The coordinates of the probes used in Fig. 1D were as follows: from cicos2, p2.4 (nucleotides 4286-4948; matches the hox gene family), p2.20 (20350-21440), and p2.30 (29050-31461); from cicos41, p41.2 (1730-2400; matches a hypothetical gene found in C. elegans), p41.8 (7679-8908), and p41.16 (16020-16580; matches a membrane transport protein gene).
- A. P. Bird, Nucleic Acids Res. 8, 1499 (1980); P. A. Jones, W. M. Rideout, J.-C. Shen, C. H. Spruck, Y. C. Tsai, BioEssays 14, 33 (1992); D. N. Cooper and H. Youssoufian, Hum. Genet. 78, 151 (1988).
- 12. Bisulfite sequencing [M. Frommer et al., Proc. Natl. Acad. Sci. U.S.A. 89, 1827 (1992)] was carried out using a published protocol [K. D. Tremblay, Technical Tips Online (http://www.elsevier.com/locate/tto), tip number p01242 (1998)]. Primers to amplify bases 19 through 405 of the Cigr-1 element after bisulfite modification were: 5'-TATAAAACATTCCACAAA and 5'-AG-GAAGAATTTTGATTTTTA. Primers used to amplify bases 44 through 160 of the modified Cimi-1 element were: 5'-ATAAAACACCTTTAACACATATATCC and 5'-GGGGGAAGATGGGATATTTTTTA. Primers used for analysis of methylation across exon 6 of the guanylate cyclase–like gene after bisulfite treatment were: 5'-AATTATAAATTCAAACTCTATACTACTACTACC and 5'-TCTTGTTATTAATATAAATATTGTATTAATA.
- J. D. Boeke and J. P. Stoye, in *Retroviruses*, J. M. Coffin, S. H. Hughes, H. Varmus, Eds. (Cold Spring Harbor Laboratory Press, Plainview, NY, 1997), pp. 343–435.
- S. Adhya and M. Gottesman, *Cell* **29**, 939 (1982); B. R. Cullen, P. T. Lomedico, G. Ju, *Nature* **307**, 241 (1984);
 V. Corbin and T. Maniatis, *ibid*. **337**, 279 (1989); S. J. Elledge and R. W. Davis, *Genes Dev*. **3**, 185 (1989);
 I. H. Greger, F. Demarchi, M. Giacca, N. J. Proudfoot, *Nucleic Acids Res*. **26**, 1294 (1998); N. J. Proudfoot, *Nature* **322**, 562 (1986).
- H. L. Ashe, J. Monks, M. Wijgerde, P. Fraser, N. J. Proudfoot, *Genes Dev.* **11**, 2494 (1997); F. Broders, A. Zahraoui, K. Scherrer, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 503 (1990); T. J. Ley and A. W. Nienhuis, *Biochem. Biophys. Res. Commun.* **112**, 1041 (1983); M. Allan, W. G. Lanyon, J. Paul, *Cell* **35**, 187 (1983); F. Calabi and M. S. Neuberger, *EMBO J.* **4**, 667 (1985); T. Moore *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 12509 (1997).
- A. Regev, M. J. Lamb, E. Jablonka, *Mol. Biol. Evol.* 15, 880 (1998).
- For example, see S. Tornaletti and G. P. Pfeifer, Oncogene 10, 1493 (1995).
- A. Bird, *Trends Genet.* **13**, 469 (1997); C. P. Walsh, J. R. Chaillet, T. H. Bestor, *Nature Genet.* **20**, 116 (1998).
- 19. We thank M. Jones, R. Dobson, N. Lennard, and J. Kershaw (The Sanger Centre, Cambridge) for cosmid sequencing; J. Davidson and A. Greig (Edinburgh) for technical assistance; R. Britten for samples of C. *intestinalis* sperm DNA; and S. Tweedie, D. Finnegan, and E. Selker for comments on the manuscript. Supported by The Wellcome Trust and The Medical Research Training Fellowship in Mathematical Biology from the Wellcome Trust.

15 October 1998; accepted 20 January 1999