Mapping Patterns of CpG Island Methylation in Normal and Neoplastic Cells Implicates Both Upstream and Downstream Regions in de Novo Methylation*

(Received for publication, April 11, 1997, and in revised form, June 6, 1997)

Jeremy R. Graff‡§, James G. Herman‡, Sanna Myöhänen¶, Stephen B. Baylin¶, and Paula M. Vertino***

From the Oncology Center and Department of Medicine, The Johns Hopkins University, Baltimore, Maryland 21231

Promoter region CpG island methylation is associated with tumor suppressor gene silencing in neoplasia. GenBank sequence analyses revealed that a number of CpG islands are juxtaposed to multiple Alu repeats, which have been proposed as "de novo methylation centers." These islands also contain multiple Sp1 elements located upstream and downstream of transcription start, which have been shown to protect CpG islands from methylation. We mapped the methylation patterns of the E-cadherin (E-cad) and von Hippel-Lindau (VHL) tumor suppressor gene CpG island regions in normal and neoplastic cells. Although unmethylated in normal tissue, these islands were embedded between densely methylated flanks containing multiple Alu repeats. These methylated flanks were segregated from the unmethylated, island CpG sites by Sp1-rich boundary regions. Finally, in human fibroblasts overexpressing DNA methyltransferase, de novo methylation of the E-cad CpG island initially involved sequences at both ends of the island and the adjacent, flanking regions and progressed with time to encompass the entire CpG island region. Together, these data suggest that boundaries exist at both ends of a CpG island to maintain the unmethylated state in normal tissue and that these boundaries may be progressively overridden, eliciting the de novo methylation associated with tumor suppressor gene silencing in neoplasia.

The CpG dinucleotide is underrepresented in the mammalian genome except in clusters known as CpG islands. CpG islands are generally unmethylated in normal tissues, irrespective of gene transcription status, while non-island CpG dinucleotides in bulk chromatin are often methylated (1). In normal tissues, extensive methylation of promoter region CpG islands is exclusively associated with transcriptional silencing of imprinted alleles and genes on the inactive X chromosome (reviewed in Ref. 2). In neoplasia, these patterns of methylation are often altered. Non-island CpGs in bulk chromatin may become hypomethylated, while CpG islands can become densely methylated (1–4). Indeed, aberrant DNA methylation of promoter region CpG islands can serve as an alternative to coding region mutation for the inactivation of tumor suppressor genes, including the retinoblastoma gene (Rb), the von Hippel-Lindau gene (VHL), p16INK4A, p15INK4B, and E-cadherin (E-cad) (5–14).

The establishment of regional methylation patterns in normal and neoplastic tissue has not been clearly defined. Recent work has demonstrated that the promoter region CpG island of the adenine phosphoribosyltransferase gene (aprt) is normally protected from methylation by a cluster of Sp1 elements located upstream of the transcription start site (15–18). Disruption of these Sp1 elements facilitates de novo methylation of the aprt promoter (15, 16, 18), which may spread from an upstream "methylation center" comprising B1 repetitive elements (17, 18). Therefore, the reord aprt CpG island remains unmethylated as a function of protective, cis-acting elements (e.g. Sp1 elements or G/C boxes) despite the influence of normally methylated, repetitive elements immediately 5' to the CpG island. It is not clear whether other CpG islands are similarly juxtaposed to normally methylated, repetitive elements (i.e. putative methylation centers) and whether such CpG islands may be protected from methylation by cis-acting elements (e.g. Sp1 elements).

To gain insight into how CpG islands remain methylation-free in normal tissue and how these islands become aberrantly methylated in neoplasia, we analyzed a number of CpG island sequences and found that proximity to Alu repetitive elements and the position of multiple Sp1 elements both upstream and downstream of transcription start are common features of many CpG islands. We have combined methylation-specific PCR and bisulfite-modified genomic sequencing to provide a detailed map of the methylation patterns in and around the E-cad and VHL tumor suppressor gene CpG islands in normal tissue and tumor cell lines. Finally, we examined the time-dependent, de novo methylation of the endogenous E-cad CpG island in fibroblasts engineered to overexpress human DNA MTase.

EXPERIMENTAL PROCEDURES

DNA and Cell Lines—DNA from normal breast epithelial tissue was kindly provided by Drs. Rena G. Lapidus, Nancy E. Davidson, Helene Smith, and Sigmond Weitzman. The renal carcinoma cell line, RFX393, was kindly provided by Dr. Michael Lerman. Genomic DNA was isolated from cell lines as described previously (14, 19). The generation and characterization of SV40-immortalized, IMR90 fetal lung fibroblasts expressing 40–50-fold increased DNA MTase activity (HMT clones) and the neomycin-resistant transfection controls (Neo clones) have been described (19). The DNA used in this study was isolated from the Neo.1 and Neo.20 clones at cell passages 6, 20, and 39, from the HMT.19 clone.

† The abbreviations used are: PCR, polymerase chain reaction; MTase, DNA methyltransferase; MSP, methylation-specific PCR analysis.

§ To whom correspondence should be addressed: The Johns Hopkins University Oncology Center, 424 N. Bond St., Rm. 132, Baltimore, MD 21231.

¶ Recipient of an award from the Academy of Finland.

*** Current address: Dept. of Radiation Oncology, Emory University School of Medicine, Atlanta, GA 30333.

This paper is available on line at http://www.jbc.org
at cell passages 6, 20, and 37 and from the HMT.1E1 clone at cell passages 6, 27, and 34.

Bisulfite Modification and Methylation-specific PCR Analysis—Genomic DNA was modified by bisulfite treatment as detailed previously (20, 21). All primer pairs (Table I) for methylation-specific PCR (MSP) of the E-cad and VHL CpG island regions were purchased from Life Technologies Inc. Each unmethylated/methylated primer pair set was routinely engineered to assess the methylation status of four to six CpG dinucleotides with at least one CpG dinucleotide positioned at the 3' end of each primer to facilitate maximal discrimination between methylated and unmethylated alleles following bisulfite modification (20). MSP primers for the Alu repeats upstream

**FIG. 1.** Sequence arrangement of multiple promoter region CpG islands. Alu sequences are depicted with large arrows. Alu repeats that have been defined as densely methylated are shaded. Sp1 elements are denoted by starbursts with white centers. The bold lines indicate the CpG island sequences where CpG density is greater than 6%. The dashed lines represent the non-island flanking sequences of each region. The smaller, closed arrows represent transcription start sites.

**FIG. 2.** Methylation patterns of the E-cad CpG island region in breast tissue. A, diagram of the E-cadherin 5' CpG island and flanking sequences. The sequences of the 5' regulatory region of the E-cad gene were compiled from GenBank accession numbers L34545, L36526, and L34937 (23). ThaI, SacII, and EagI restriction enzyme recognition sites are depicted. Sp1 elements (GGCCGG or CCGGCC) are denoted by starbursts with white centers. The open arrows represent the position and orientation of Alu sequences, which were assigned by homology search of sequence L34545 with the Alu repeat data base of the National Center for Biotechnology Information using the BLAST algorithm. The smaller, closed arrow denotes the transcription start site. The amplified products for the individual primer sets are depicted by bold lines. B, summary of the MSP results for breast tumor cell lines and normal breast epithelia. Each "lollipop" summarizes the methylation data scored by each primer set represented in A. Completely methylated alleles are depicted by black ovals, predominantly methylated alleles by black ovals with white centers, both unmethylated and methylated alleles by striped ovals, predominantly unmethylated alleles by white ovals with black centers, and completely unmethylated alleles by white ovals.
and downstream of the VHL CpG island were designed with a common antisense primer in a region devoid of CpG sites to assess the methylation status of CpG sites located only within the Alu sequences (Table I).

Bisulfite-modified Genomic Sequencing—Genomic sequencing of bisulfite-modified DNAs (22) was performed by solid phase DNA sequencing (21). The first amplification of sequences at the 5' end of the CpG island was performed for 35 cycles using the following primer position (product size) GenBank accession no. Annealing Temperature

TABLE I

| Primer Pair | Methylation set (5'→3') upstream/downstream | Unmethylated set (5'→3') upstream/downstream | Primer position (product size) GenBank accession no. | Annealing Temperature |
|-------------|---------------------------------------------|----------------------------------------------|--------------------------------------------------|----------------------|
| Alu1        | TTTTTTATGTTGTTAATGTTTGA                    | TTTCCTCAAAATCTCCTACCC                      | CH3, pos. 24–196 (172 bp) L34545                  | 57 °C                |
|             | TATACGCAACCACCCACCAA                      | CTTACCAACTACCTACCACTACCC                  | UN, pos. 25–193 (168 bp) L34545                   |                      |
| Alu2        | CTTTAAATGTTTTTATGTTTGA                    | TATACGCAACCACCCACCAA                      | CH3, pos. 473–645 (172 bp) L34545                 | 55 °C                |
|             | ATTTACCTATACCTATACCC                      | CTTACCAACTACCTACCACTACCC                  | UN, pos. 473–644 (171 bp) L34545                   |                      |
| Island 1    | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 844–960 (116 bp) L34545                 | 57 °C (CH3) V53 (Un) |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 843–940 (97 bp) L34545                    |                      |
| Island 2    | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 881–1053 (172 bp) L34545                 | 57 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 881–1055 (174 bp) L34545                  |                      |
| Island 3    | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 945–1149 (204 bp) L34545                 | 57 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 941–1152 (211 bp) L34545                  |                      |
| Island 4    | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 254–376 (144 bp) L34545                  | 60 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 251–376 (147 bp) L34545                    |                      |
| Island 5    | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 612–797 (185 bp) L34526                  | 60 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 610–795 (183 bp) L34526                    |                      |
| Island 6    | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 41–135 (95 bp) L34526                    | 55 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 41–136 (95 bp) L34526                      |                      |
| VHL-Alu1A   | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 88–466 (378 bp) L34526                   | 57 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 80–466 (387 bp) L34526                      |                      |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | U19763                                             |                      |
| VHL-Alu1B   | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 280–466 (186 bp) L34526                   | 56 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 274–468 (192 bp) L34526                    |                      |
| VHL-TS      | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 532–699 (158 bp) L34526                   | 60 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 530–695 (165 bp) L34526                    |                      |
| VHL-Exon 1  | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | CH3, pos. 8–602 (610 bp) L34526                     | 55 °C                |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | UN, pos. 8–602 (612 bp) L34526                      |                      |
| VHL-Alu2A   | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | L15409/U68055                                      |                      |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | L15409/U68055                                      |                      |
| VHL-Alu2B   | TTTTCTAACCCATCCTACCC                      | CTTACCAACTACCTACCACTACCC                  | L15409/U68055                                      |                      |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | U68055                                             |                      |
|             | CTTACCAACTACCTACCACTACCC                  | CTTACCAACTACCTACCACTACCC                  | U68055                                             |                      |

RESULTS

Sequence Arrangement of Multiple 5' CpG Island Regions—Multiple Sp1 elements at the 5' edge of the rodent apt CpG island protect the island from methylation (15–18), which spreads from normally methylated, upstream tandem B1 repetitive elements (17, 18). To determine whether other CpG island regions may be similarly constituted, we examined the GenBank sequences for a number of 5' CpG island regions, especially those in the promoter regions of genes involved in neoplasia. Maps of the 5' CpG island regions of the human APRT (accession no. U09817), E-cadherin (accession nos. L34545, L36526, and L34937), glutathione S-transferase π (accession no. X08058), tissue inhibitor of metalloprotease II (TIMP-2; accession no. U44381), neurofibromatosis-1 (NF-1; accession nos. U19763, L15409, and U68055) and von Hippel-Lindau (VHL; accession nos. U19763, L15409, and U68055) are represented in Fig. 1. For each of these CpG islands, multiple Alu repeats were located immediately 5' to the CpG island with the most proximal Alu located within ~1 kilobase upstream of transcription start. Furthermore, each CpG island contained multiple Sp1 elements located both upstream and downstream of transcription start (Fig. 1). Genomic regions 3' to these CpG islands did not contain enough deposited sequence to evaluate the presence of Alu repeats, except for the VHL CpG island region, which has Alu repeats both upstream and downstream of the CpG island (Fig. 1). The promoter region CpG islands of the Bβ and estrogen receptor genes are devoid of any proximal Alu repetitive elements (data not shown). Hence, the proximity of multiple Alu repeats may be common to many CpG islands, but is not universal.

Methylation Patterns of the E-cadherin 5' CpG Island and
Flanking Sequences—To understand how these sequences may be related to patterns of methylation in normal tissue and in neoplasia, we mapped the methylation status of the 2.2-kilobase region encompassing the entire E-cad CpG island and flanking, non-island sequences (Fig. 2A) in normal breast epithelia and breast tumor cell lines. To identify the critical areas within this region that may participate in establishing normal and aberrant patterns of methylation, we employed the re-}

Fig. 3. Methylation-specific PCR analysis of the E-cadherin CpG island region. **A**, representative mixing experiment with E-cad island primer set 3. After bisulfite modification, MCF-7 DNA, in which the E-cad CpG island is unmethylated, was mixed with MDA-MB-231 DNA, in which the E-cad CpG island is extensively methylated, at the indicated ratios. The sizes of the methylated (M)/unmethylated (U) products are indicated. B, representative MSP reactions with island primer sets 1–6 from E-cad-positive (MCF-7, T47D, and ZR-75) and E-cad-negative breast tumor cell lines (231, Hs578t, HBL100, MCF-7ADR, and 435) and from normal breast epithelia samples (NBR). MSP products using primers that specifically amplify only unmethylated DNA are indicated by U while products amplified by primers specific for methylated DNA are indicated by M. The sizes of amplified products (base pairs) of the methylated/unmethylated reactions are indicated. C, Representative MSP reactions with the Alu1 and Alu2 primer sets in normal breast epithelia (NBR) and breast tumor cell lines. NBR3 did not amplify with the Alu1 primer set in this particular set of reactions. Lanes designated H2O represent the control reactions containing no template DNA.

**Flanking Sequences**—To understand how these sequences may be related to patterns of methylation in normal tissue and in neoplasia, we mapped the methylation status of the 2.2-kilobase region encompassing the entire E-cad CpG island and flanking, non-island sequences (Fig. 2A) in normal breast epithelia and breast tumor cell lines. To identify the critical areas within this region that may participate in establishing normal and aberrant patterns of methylation, we employed the recently developed MSP, which can readily identify methylated alleles comprising as little as 0.1% of the total sample (20). Once identified, these targeted areas were examined in greater detail by bisulfite-modified, genomic sequencing in select nor-
The primer sets used to examine the E-cad CpG island region covered 33 of 138 CpG sites throughout the region (listed in Table I, depicted in Fig. 2A). Since unmethylated and methylated primers for these MSP primer sets were directed to the same region of the genome, amplify similarly sized products, and, in most cases, were designed with identical annealing temperatures, the unmethylated and methylated primer sets of each pair amplify with similar efficiency (20). To demonstrate this, we assayed mixtures of methylated and unmethylated DNA with various primer sets directed to the area of highest CpG density within the E-cad CpG island (Fig. 1). The results for island set 3 (Fig. 3A) exemplify that MSP has the resolution to define a population of alleles as completely methylated or unmethylated, predominantly methylated or unmethylated, or comprising both methylated and unmethylated alleles.

In normal breast epithelia and E-cad-expressing breast cancer cell lines (MCF-7, T47D, and ZR-75; Ref. 24), CpG sites within the island were completely unmethylated (island primer sets 1–4, Fig. 3). The flanking, non-island CpG sites in exon 2 were also extensively methylated in the E-cad-expressing cell lines and in normal breast epithelia (Island primer set 5, Fig. 3B), which also displayed methylation at the 3' edge of the CpG island (island set 5, Fig. 3B). By contrast, the E-cad-negative breast tumor cell lines (MDA-MB-231, HS775, MDA-MB-435, MCF7ADR, and HBL100; Ref. 24) showed extensive methylation of the upstream Alu repeats (Fig. 3C), and virtually all CpG sites examined within the island (representative data for these cell lines are depicted in Fig. 3B), particularly those in the region of highest CpG density near the transcription start site (island set 3, Fig. 3B). Therefore, in normal tissue and E-cad-positive tumor cells, the CpG island is unmethylated but is embedded between regions of dense methylation, while the entire region is densely methylated in the E-cad-negative breast tumor cell lines (summarized in Fig. 2B).

**Defining the Border Sequences between Methylated and Unmethylated CpG Sites in the E-cad CpG Island**—In normal breast epithelia and E-cad-expressing tumor cell lines, MSP analyses indicated that the E-cad 5' CpG island has both 5' and 3' boundaries delineating methylated, flanking region CpG sites from the unmethylated CpG sites within the island (summarized in Fig. 2B). To define precisely these borders, we used bisulfite-modified genomic sequencing to determine the methylation status of 21 CpG sites within the putative 5' border region and 24 CpG sites within the 3' border region of the E-cad CpG island. A sharp boundary between unmethylated CpG sites and methylated, or partially methylated, CpG sites exists at both the 5' and 3' regions of the island coincident with clusters of Sp1 elements, in the regions of declining CpG density (data from normal breast epithelia are depicted in Fig. 4).

**Methylation Patterns of the VHL 5' CpG Island Region**—To determine whether the patterns of methylation for the E-cad CpG island were shared by other CpG islands, we examined the VHL CpG island region by MSP in normal kidney tissue and a renal carcinoma cell line, RFX393, which does not express VHL (data not shown). The VHL MSP primer sets (Table I) covered 27 of 120 CpG sites throughout the CpG island region (Fig. 5A). In RFX393, the entire VHL CpG island region, including upstream and downstream Alu repeats, was densely methylated. Likewise, in normal kidney, the Alu sequences both 5' and 3' to the island were extensively methylated but the CpG island was completely unmethylated (Fig. 5B). Therefore, like the E-cad CpG island region, the VHL CpG island was unmethylated in normal tissue (summarized in Fig. 5C) but was flanked by regions of dense methylation (i.e. Alu repeats). Similarly, the proximal Alu repeats upstream of the human APRT and TIMP-2 genes were also extensively methylated (illustrated in Fig. 1) while the central region of each island was unmethyl-
Aberrant Methylation of the E-cad and VHL CpG Islands

**DISCUSSION**

In the present study, we have sought to understand how CpG islands may be protected from methylation in normal tissue.

**Fig. 5. Methylation patterns of the VHL 5' CpG island region.**

A. diagram of the VHL 5' CpG island region (GenBank accession numbers U19763, L15409, and U68055; Ref. 25). Hash marks under the island schematic represent individual CpG sites. Open arrow indicate the position of Alu repetitive sequences. The smaller, closed arrow represents the major transcription start site of VHL (24). SP1 elements are represented by starbursts with white centers. The positions of the amplified PCR products used to examine the VHL CpG island region are indicated by bold lines beneath the schematic. The VHL-Alu 1A and VHL-Alu 1B primer sets used the same antisense primer positioned within a region devoid of CpG sites for both unmethylated and methylated reactions. The VHL-Exon 1, VHL-Alu 2A, and VHL-Alu 2B primer sets also utilized a single antisense primer directed to a region devoid of CpG sites for both unmethylated and methylated reactions. B: methylation patterns of the VHL 5' region in normal kidney and a renal cancer cell line, RXF 393. The high molecular weight band for the unmethylated reaction with the Alu 1B primer set is a nonspecific band. C, summary of the methylation patterns of the VHL 5' region in normal kidney and a renal cancer cell line RXF 393. Each lollipop represents the methylation status of the amplified regions depicted in A. Detection of completely methylated alleles is indicated by black ovals, predominantly methylated alleles by black ovals with white centers, both unmethylated and methylated alleles by striped ovals, and completely unmethylated alleles by white ovals.

**Fig. 6. Methylation patterns of the E-cad CpG island region.**

A. diagram of the E-cad CpG island region (GenBank accession numbers U19763, L15409, and U68055; Ref. 25). Hash marks under the island schematic represent individual CpG sites. Open arrow indicate the position of Alu repetitive sequences. The smaller, closed arrow represents the major transcription start site of VHL (24). SP1 elements are represented by starbursts with white centers. The positions of the amplified PCR products used to examine the VHL CpG island region are indicated by bold lines beneath the schematic. The VHL-Alu 1A and VHL-Alu 1B primer sets used the same antisense primer positioned within a region devoid of CpG sites for both unmethylated and methylated reactions. The VHL-Exon 1, VHL-Alu 2A, and VHL-Alu 2B primer sets also utilized a single antisense primer directed to a region devoid of CpG sites for both unmethylated and methylated reactions. B: methylation patterns of the E-cad CpG island in normal breast tumor cell lines (data not shown).

The Evolution of Aberrant Methylation in the E-cad CpG Island—There is currently little information regarding the evolution of aberrant, de novo methylation of endogenous CpG islands. We have previously demonstrated that overexpression of DNA MTase in SV40-immortalized, IMR-90 fibroblasts can drive the de novo methylation of certain CpG islands, including the E-cad CpG island (19). To examine the time-dependent evolution of aberrant methylation of the E-cad CpG island, we compared the methylation status of the E-cad CpG island and flanking sequences in two Neo control clones and two HMT clones over the course of 40 cell passages.

As in normal breast epithelia, the non-island CpG sites upstream (Alu) and downstream of the E-cad CpG island (island set 6, exon 2) were extensively methylated in the Neo and HMT clones (Fig. 6). In the Neo clones, the regions within the CpG island were unmethylated in the early passage samples (island sets 1–5; Fig. 6). By mid-passage, these regions generally remained unmethylated in the Neo clones, although some methylation was evident within the 3' edge of the island in both clones (island set 5, Fig. 6) and at the 5' edge of the island in Neo.20 (island set 1, Fig. 6). These patterns of methylation remained relatively constant between passages 20 and 39 with the region of greatest CpG density remaining unmethylated (island sets 2, 3, and 4), despite methylation at the fringes of the CpG island (island sets 1 and 5, Fig. 6) and dense methylation within the flanking, non-island sequences (summarized in Fig. 7).

By contrast, the E-cad CpG island became progressively more methylated with time in two independent cell clones that overexpress human DNA MTase (HMT.19 and HMT.1E1). In the early passage HMT clones, methylation was evident in all regions examined throughout the island except in the heart of the island near the transcription start site (island set 3, Fig. 6). By mid-passage, methylation within the CpG island had become more prominent (island sets 1–5, Fig. 6) and, for HMT.1E1, even the area around the transcription start site was almost completely methylated (island set 3, Fig. 6). By late passage, all of the regions examined within the island were predominantly methylated in both HMT clones, including the region spanning the transcription start site (Fig. 6). These data indicate that aberrant methylation first involved both the 5' and 3' edges of the E-cad CpG island and the flanking, non-island sequences and, with time, progressively extended throughout the entire CpG island to include the central area of highest CpG density near the transcription start site (summarized in Fig. 7).
and how aberrant CpG island methylation may develop in neoplastic cells. We show that the CpG islands and flanking sequences of E-cad and VHL are extensively methylated in tumor cell lines for which gene expression is extinguished. In normal tissue, both CpG islands are unmethylated but are immediately flanked by regions of dense methylation, containing Alu repeats. Additionally, for E-cad, the regions that mark the boundaries between unmethylated, island CpG sites and the methylated, flanking region CpG sites contain multiple Sp1 elements located at the 5' and 3' edges of the island, where CpG density declines. GenBank sequence analysis revealed that other CpG islands are also positioned immediately downstream from Alu-rich, non-island flanking regions and have multiple Sp1 elements located both 5' and 3' to transcription.

**FIG. 6. The dynamics of aberrant methylation of the E-cad CpG island.** The methylation status of the E-cad 5' CpG island was assessed by MSP analysis in fibroblast clones that overexpress human DNA MTase (HMT.19 and HMT.1E1) and in the transfection control clones (Neo.1 and Neo.20) over approximately 40 cell passages. MSP products using primers that specifically amplify only unmethylated DNA are indicated by U, while products amplified by primers specific for methylated DNA are indicated by M. The primer sets used to examine the E-cad CpG island region are listed to the left of each gel and refer to the positions indicated in Fig. 2.

**FIG. 7. Summary of the time-dependent de novo methylation of the E-cad CpG island.** A, E-cad CpG island map compiled from GenBank accession numbers L34545, L36526, and L34937. Each region examined by MSP primer sets is depicted by bold lines beneath the map. Sp1 elements are represented by starbursts with white centers. B, summary of the methylation patterns in the Neo and HMT clones over approximately 40 cell passages. Each lollipop summarizes the methylation status for each primer set depicted in A. Completely methylated alleles are depicted by black ovals, predominantly methylated alleles by black ovals with white centers, both unmethylated and methylated alleles by striped ovals, predominantly unmethylated alleles by white ovals with black centers, and completely unmethylated alleles by white ovals.
start (Fig. 1). It seems plausible that these sequence characteristics may participate in establishing both normal and aberrant patterns of methylation in these CpG island regions.

In the rodent aprt CpG island, disruption of a cluster of Sp1 sites (or G/C boxes) located at the 5’ end of the island elicits de novo methylation of the aprt CpG island (15, 16), even when Sp1-mediated transcription is not disrupted (18). The de novo methylation of the mouse aprt CpG island appears to originate within, and spread from, normally methylated B1 repetitive elements located immediately 5’ to the island (17, 18). Turker and Bestor (26) have proposed that such repetitive sequences, which are frequently methylated in normal tissue (27, 28), may act as de novo methylation centers, i.e. cis-acting elements methylated in normal tissue from which methylation may spread bidirectionally into adjacent sequences. Such spreading of premethylated methylation patterns into adjacent sequences has been documented (29). Furthermore, an Alu element in intron 6 of the human p53 gene has been shown to act as such a methylation center, directing the ubiquitous methylation of the CpG site in codon 248 (30). B1 and B2 repetitive elements, the rodent equivalent to the human Alu family of repetitive sequences (31), may also act as a methylation center directing the de novo methylation of the rat α-fetoprotein gene CpG island (32).

Our data show that the Alu repeats within the CpG island regions of E-cad and VHL are extensively methylated in normal tissue and tumor cell lines, regardless of gene expression status or the tissue of origin. Additionally, we show that multiple Sp1 sites exist within both the 5’ and 3’ edges of the E-cad and VHL CpG islands marking the boundary between the unmethylated island CpG sites and the methylated, non-island CpG sites. Furthermore, our multi-gene sequence analysis (Fig. 1) shows that the Sp1 elements present in most CpG islands (33) are often located both upstream and downstream of transcription start, perhaps protecting these islands from the spread of methylation originating in either flanking region. Consistent with this possibility, we show that de novo methylation of the E-cad CpG island in fibroblasts overexpressing DNA MTase begins within both flanking regions and in sequences at both edges of the island, progressing with cell passage to the central region of highest CpG density within the island (summarized in Fig. 7).

Several scenarios may explain these results. For instance, these data are consistent with the hypothesis that the normally methylated, sequences (e.g. Alu repetitive elements) flanking the E-cad CpG island may act as methylation centers directing the spread of methylation toward the island. Alternatively, since sequences with highest CpG content may be inherently more resistant to the action of DNA MTase (34), it is possible that the lateral regions of the island, which are the least CpG-rich, may be better substrates for DNA MTase than the area of highest CpG density. Therefore, methylation may accumulate more readily in these border regions, independent of the methylation status of the flanking, non-island sequences. While we cannot currently distinguish between these two possibilities, it is clear that the region of highest CpG density remained most resistant to methylation.

In summary, the data in this report, in conjunction with previous reports on the APRT CpG island (15–19, 26), suggest that CpG islands may commonly be juxtaposed with densely methylated, Alu-rich regions and may be protected from the influence of these methylated flanking sequences by clusters of Sp1 elements at both the 5’ and 3’ sides of the island.

Tumor suppression. The protection mediated by these Sp1-rich bar-