Analysis of Promoter CpG Island Hypermethylation in Cancer:
Location, Location, Location!
Iris J.H. van Vlodrop1, Hanneke E.C. Niessen1, Sarah Derks2, Marcella M.L.L. Baldewijns1,
Wim van Criekinge3, James G. Herman4, and Manon van Engeland1

Abstract
The genetic and epigenetic alterations that underlie cancer pathogenesis are rapidly being identified. This provides novel insights in tumor biology as well as in potential cancer biomarkers. The somatic mutations in cancer genes that have been implemented in clinical practice are well defined and very specific. For epigenetic alterations, and more specifically aberrant methylation of promoter CpG islands, evidence is emerging that these markers could be used for the early detection of cancer as well as prediction of prognosis and response to therapy. However, the exact location of biologically and clinically relevant hypermethylation has not been identified for the majority of methylation markers. The most widely used approaches to analyze DNA methylation are based on primer- and probe-based assays that provide information for a limited number of CpG dinucleotides and thus for only part of the information available in a given CpG island. Validation of the current data and implementation of hypermethylation markers in clinical practice require a more comprehensive and critical evaluation of DNA methylation and limitations of the techniques currently used in methylation marker research. Here, we discuss the emerging evidence on the importance of the location of CpG dinucleotide hypermethylation in relation to gene expression and associations with clinicopathologic characteristics in cancer. Clin Cancer Res; 17(13); 4225–31. ©2011 AACR.

Introduction
DNA methylation is involved in regulating gene expression in normal physiology (e.g., by managing imprinting, X-chromosome activation, and tissue-specific gene expression) and disease (e.g., neurodevelopmental and degenerative disorders, autoimmune diseases, and cancer; ref. 1). DNA hypermethylation–induced silencing of tumor suppressor and DNA repair genes is a frequent phenomenon affecting the hallmarks of cancer (2, 3). Aberrant DNA methylation often occurs around the transcription start site (TSS) within a CpG island, and as was recently shown, even outside of the traditionally defined islands (4, 5). These hypermethylation markers are promising tools to detect cancer cells in tissue and body fluids (6, 7) with the use of simple PCR technology (8–10). Proof of principle for the clinical value of methylation markers has been reported for early detection and classification of cancer (11–21), risk assessment and prognosis (19, 22–24), and prediction of therapy response (25–27), with some already having shown their importance in (pre)clinical practice. Thus, the promise of methylation changes to become a powerful diagnostic and predictive tool (6) is becoming a reality.

Nevertheless, the clinical value of biomarkers depends on the accuracy and prognostic or predictive value of the marker. The CpG islands of a variety of cancer-associated genes have been evaluated for methylation, and positive, negative, and null associations with gene expression and clinical characteristics are reported. Here, we discuss emerging evidence on the importance of the location of aberrant CpG dinucleotide methylation in relation to gene expression and its clinical value in cancer.

Location of Biologically Relevant Methylation in Promoter CpG Islands

The dogma that promoter CpG island methylation generally induces gene silencing is currently being specified. Specific regions within the promoter CpG islands, designated as core regions crucial for regulating gene expression, are rapidly being identified. As illustrated in Fig. 1, these regions are often situated around the TSS within a CpG island but can also be observed more upstream or downstream of the TSS.

One of the first studies to show that hypermethylation of a specific locus is critical for transcriptional repression was conducted in the human bladder cancer cell line T24.
and clustered hypermethylation more downstream of the cellular carcinoma cell lines revealed one unmethylated cell upstream of this region did not influence colorectal cancer cell lines (32), whereas hypermethylation been identified to regulate expression by methylation in 24 TSS region. A small region proximal to the TSS region initially analyzed (upstream of TSS; ref. 29). A more detailed analysis of the promoter CpG island region around the TSS revealed that silencing of hTERT expression was associated with dense methylation at, or in close proximity to, the TSS and is independent of methylation more upstream of TSS (29). Similar observations have been reported for the TGF-β signaling target RUNX3 (30) and the T-cell differentiation protein MAL (31) in gastric cancer cell lines and primary gastric cancers.

Core regions have also been observed outside the direct TSS region. A small region proximal to the MLH1 TSS has been identified to regulate expression by methylation in 24 colorectal cancer cell lines (32), whereas hypermethylation upstream of this region did not influence MLH1 expression (32) and was later suggested to be age-related (33). The same correlation was observed in 64 primary colorectal cancers (34) as well as in 123 patients with colorectal cancer in an independent study (35). Similarly, mapping of WIF-1 promoter CpG island hypermethylation reveals regional methylation just proximal to the TSS that correlates with transcriptional silencing, whereas other more upstream regions do not (36).

CpG island methylation analyses of SOCS1 in hepatocellular carcinoma cell lines revealed one unmethylated cell line without SOCS1 expression. More detailed analyses by bisulfite sequencing of a larger region revealed regional and clustered hypermethylation more downstream of the initially analyzed region, indicating a silencing effect by methylation in this critical 3'-TSS region (37).

Interestingly, a recent study showed transcriptional silencing of TTP in liver cancer by hypermethylation of a specific single CpG site. One specific CpG dinucleotide, located at the 5'-boundary of the CpG island, was exclusively hypermethylated in transcriptionally silenced cell lines (38). This observation narrows down the core region for hypermethylation-induced silencing of TTP to just one CpG dinucleotide.

These studies show that transcriptional silencing does not require hypermethylation of the entire CpG island, but that methylation of a few gene-specific core CpG dinucleotides, most likely associated with transcription, may be sufficient. It is important to realize that data obtained solely in cell lines can be biased, as they exhibit significantly more CpG island methylation than the primary tumors they represent (39) and thus proof of principle in primary tumors is required. Identifying the core regions regulating gene expression is essential for evaluation of the clinical value of DNA hypermethylation (Fig. 1). For example, two regions within the MAL promoter were analyzed for methylation in gastric cancer samples. Hypermethylation of both regions occurred in 71% and 80%, respectively; however, only methylation at the region closest to the TSS was correlated with a better disease-free survival (31). In addition, increased expression of MAL in serous ovarian cancer patients with a poor prognosis is associated with decreased methylation of a specific region of the MAL promoter (40).

We recently described a region in the promoter CpG island of GREM1 that was specifically associated with poor prognosis in clear cell renal cell carcinoma. Three regions were analyzed for hypermethylation, but only one was correlated with poor survival (23). This indicates that location of hypermethylation is also important for marker discovery.

These studies clearly indicate that the biological and clinical consequences of promoter CpG island hypermethylation are strongly dependent on silencing of expression-regulating core regions in the CpG island. Although clinically relevant hypermethylation of a specific locus is not always perfectly associated with gene expression and might serve as a surrogate marker for functional hypermethylation of another locus, we expect that the best validated markers will be those for which good correlations between DNA methylation and gene expression exist.

Hypermethylation outside core regions is frequently observed in cancer cells but sometimes also in normal cells (41–44) and is correlated with aging and chronic inflammation (41–43, 45, 46). This is hypothesized to progress toward the core region, initiating gene silencing (47). For example, a demarcation has been observed between RASSF1A hypermethylation in exon 1 and in its immediate upstream promoter region. In normal breast tissue, exon 1 is methylated without affecting gene expression, whereas in breast cancer samples, hypermethylation is observed in both exon 1 and its immediate upstream promoter region that is associated with RASSF1A silencing. A progressive spreading from exon 1 upstream is proposed, which can
occur early in breast tumorigenesis (48). Additional evidence for spreading of hypermethylation in the promoter CpG island region has been observed for CDKN2B in leukemia (49), CDKN2A (50), MGMT (51), and NDRG4 (11) in colorectal cancer, and for RUNX3 (30) in gastric cancer.

Spreading of DNA methylation is often consistent with increasing density of methylation, but whether density itself or spreading toward (expression regulating) specific regions is correlated with gene silencing is currently not clear.

Location of DNA Methylation Initiation

It might be speculated that the core region for which hypermethylation is associated with gene silencing and clinical consequences has specific (sequence) characteristics. To study this hypothesis, Feltus and colleagues applied DNA pattern recognition techniques in a DNA cytosine-5-methyltransferase 1 (DNMT1) overexpressing human cell culture model and showed that methylation-prone and methylation-resistant CpG islands can be distinguished by an underlying sequence signature based on 13 DNA motifs (52, 53). These motifs were proposed to represent protein-binding sites involved in the susceptibility to or prevention of DNA methylation. Although the methylation-prone motifs do not obviously resemble a transcription factor consensus sequence or protein-binding site, transcription factors PML-RAR (54) and c-Myc (55) have shown the ability to initiate DNA hypermethylation by the recruitment of DNA methyltransferase enzymes (DNMT)

Figure 1. Location of biologically relevant methylation in promoter CpG islands. Promoter regions from −1,000 to +1,000 bp are depicted relative to the TSS (at 0 in red) with CpG islands in blue. Vertical lines represent CpG sites, and gray boxes show the relevant regions (core regions) for expression or progression. All genes are presented in a forward fashion and are grouped by core region position relative to the TSS: top genes are around TSS, middle genes are pre-TSS, and bottom genes are post-TSS. Top, core region of cell-cycle regulating gene CDKN2A was identified from −121 to +123 relative to the TSS. The catalytic subunit of telomerase, hTERT, showed a core region from −150 to +150, relative to TSS. The important region in the human runt-related transcription factor 1, RUNX1, was reported from −194 to +451 relative to the TSS. Methylation at MAL promoter from −92 to +7 relative to the first ATG correlated with expression and survival (B). Another region was found at −452 to −266 relative to the TSS, which showed a correlation with worse prognosis (A). Middle, the region −248 to −178 relative to the TSS of the mismatch repair gene MLH1 was identified as the core region. For the Wnt-pathway antagonist Wnt inhibitory factor-1 (WIF-1), the core region is reported to be proximal to the TSS from −295 to −95. TTP, a negative post-transcriptional regulator of c-Myc, uniquely showed one CpG at −500 bp at the 5'-boundary of the CpG island, as the core dinucleotide. Bottom, for the bone morphogenetic protein (BMP) pathway antagonist GREM1, the region +311 to +471 relative to the TSS showed clinical correlations when methylated. The core region of a negative regulator of the Janus-activated kinase (JAK)/STAT pathway SOCS1 was identified at +901 to +924 (relative to new TSS). Note: differences in location as compared with Yoshikawa and colleagues (ref. 37) are due to a repositioning of the predicted TSS after publishing (680 bp more upstream according to the previous predicted TSS position). Original reported core region at −221 to +244.
to specific loci. The opposite is observed for the presence of Alu elements (52) and Sp1-binding sites (56, 57), as well as binding of the insulin protein CTIF (58), which are all associated with resistance to DNA hypermethylation. Subsequent studies have shown that genes with a methylation-prone sequence motif and genes characterized by Polycomb group (PcG) protein occupancy in embryonic stem cells are strongly related (59). PcG proteins have been shown to mark target genes in the progenitor or stem cell state by targeting H3K27 histone methylation. Several observations indicate that there could be a functional link between PcG protein binding and CpG island hypermethylation. First, the reported percentage of PcG-binding sites that correspond to CpG islands ranges from 50% to 88% (60, 61). Second, direct interactions have been described between PcG proteins and DNMTs (62, 63). Third, PcG target genes are up to 12 times more likely to have cancer-specific promoter hypermethylation than non-PcG targets (64–66). These observations make it tempting to speculate that PcG proteins recruit DNMTs to their target genes and thereby induce aberrant transcriptional silencing of promoter CpG islands by DNA hypermethylation.

**Location of Methylation Outside of Classical Promoter CpG Islands**

DNA methylation studies in cancer initially focused on gene promoter CpG island hypermethylation. However, recent research revealed novel insights on the location of DNA hypermethylation. Hypermethylation of intra- and intergenic CpG dinucleotides might contribute to regulating gene expression by functioning as alternative promoters (5). For example, in-depth investigation of the human SHANK3 locus (~60 kb) showed hypermethylation-regulated intragenic promoter activity, expressing alternative transcripts in a tissue-(brain) and cell-type (primary cortical astrocyte)–specific manner (5). In addition, other gene-regulating regions such as enhancers, which are cis-regulatory DNA sequences that increase transcription independent of their orientation and distance relative to the TSS, can be regulated by hypermethylation (67, 68). For example, hypermethylation-dependent enhancer-like activity, located at a CpG island in EGR2 intron 1, is suggested to regulate transcription (69).

Evidence is accumulating that CpG island hypermethylation in bidirectional promoters is correlated with silencing of both genes, thereby possibly accelerating tumorigenesis, for example, in gene pairs WNT9A/CD558500, CIDTSPL/BC040563, KCNK15/BF195580, and MLH1/EPM2AIP1 (70, 71). Even in promoters without a classical CpG island (low CpG density), hypermethylation still can regulate expression, as has been shown for Maspin in breast cancer (72).

Although the impact of DNA hypermethylation has been studied mainly in CpG islands located at TSSs (73), Trizarry and colleagues recently introduced the term “CpG island shores” (4), regions with a relatively low CpG density located within 2 kb of traditional CpG islands. Aberrant methylation in these shores was reported to segregate tissue subtypes and cancerous tissue from matched normal tissues (4). These observations change the current focus from exclusively CpG islands in promoter regions to much larger regions of interest, which potentially possess regulatory regions previously characterized.

The biological relevance of hypermethylation throughout the gene locus by means of long-range interactions with the promoter region has recently been shown by Tiwari and colleagues (74). They reported DNA methylation at 6 of 7 CpG islands, including the island spanning the TSS, throughout the GATA4 gene. Chromatin looping can enable long-range interactions of these islands around a single gene. This can cluster aberrant methylation of CpG islands and other epigenetic markers, thereby facilitating and enhancing transcriptional repression (74). These findings demand mapping of DNA hypermethylation of genes in higher-order chromatin structures as there might be an additional role for chromatin looping in mediating gene expression.

Frigola and colleagues showed for the first time that clusters of genes could be coordinately repressed by epigenetic mechanisms, a concept termed long-range epigenetic silencing (LRES; ref. 75). They identified an epigenetically repressed 4-Mb spanning region of chromosome 2q14.2. Genes located in this cytogenetic region are affected by hypermethylation of clusters of neighboring CpG islands and coordinately inactivated by chromatin remodeling. Similar LRES mechanisms have been observed by others in chromosomal regions 3q22 (76) and 5q35.2 (77). Recently, 47 LRES regions were identified in prostate cancer, typically spanning about 2 Mb and harboring approximately 12 genes (78). Global gene silencing by LRES is comparable with genetic deletions by LOH, as large regions become simultaneously inactivated. Therefore, LRES provides an efficient silencing mechanism in cancer development.

Furthermore, nucleosome organization, location, and dynamics are critical for gene regulation. Lin and colleagues studied MLH1 silencing by hypermethylation and nucleosomal occupancy in cancer (70). They showed nucleosome depletion just upstream of each start site on the active MLH1 promoter in normal cells, whereas 3 nucleosomes were present on the hypermethylated, inactive promoter. Moreover, gene reactivation induced by the demethylating agent 5-aza-2′-deoxycytidine involved promoter nucleosome removal, suggesting that epigenetic silencing may involve the (reversible) movement of nucleosomes into previously vacant positions (70). Changes in nucleosomal occupancy not only occur at TSS regions but also at enhancers acting at variable distances from the start site (79).

**Conclusions and Perspectives: Reflect on Location**

The above-described location-related complexities of gene expression regulation by aberrant DNA methylation can all, separately or combined, result in unexpected or
misinterpreted information on the associations among DNA hypermethylation, gene expression, and clinical parameters. Promoter CpG islands of genes have often been reported as "unmethylated" or "hypermethylated," based on the data of only a small number of CpG dinucleotides independent of location or the assays that have been used. Because it now has become clear that the location of core regions and the density of methylation required for gene silencing can vary per gene, a broader view than just the classical dogma of promoter CpG island methylation and gene silencing is needed to interpret data on DNA hypermethylation, gene expression, and clinicopathologic associations. Unexpected results do not per se contradict this dogma regarding the complexity and the number of parameters involved in epigenetic silencing. In addition, the above-mentioned phenomena might be tissue, cell-type, cancer type, genomic region, or gene specific, thereby complicating data analysis, interpretation, and validation of results and conception of the literature.

These considerations underscore the importance of detailed analysis of CpG dinucleotide analysis and careful data analysis, with regard to diverse techniques and/or primer and probe design. Results of analyses at the same region are dependent on the detection method, that is, primer (design), reagents, detectors, equipment, and protocols, which all influence sensitivity and specificity. Frequently used technologies are restriction enzyme- and/or bisulfite-based analyses, the results of which are highly dependent on primer- and/or probe/microarray design-like methylated-specific PCR (MSP; ref. 80), methylated DNA immunoprecipitation (MeDIP; ref. 81), methylated-CpG island recovery assay (MIRA; ref. 82), and Illumina Infinium methylation assay (83). Limitations of these techniques can introduce bias; for example, MSP only assesses 2 to 4 CpG dinucleotides per oligo and thus needs to perfectly cover the core region of interest. Methylation-sensitive restriction enzyme digestion can introduce recognition site bias and is prone to false-positive results because of incomplete digestion. Techniques using DNA hybridization to microarrays introduce ascertainment bias (for an extensive overview of the resolution and limitations of the most widely used techniques to analyze DNA methylation, see ref. 84). The recently developed novel technologies that enable (semi)epigenome-wide analyses such as bisulfite deep sequencing (85) or methyl-binding protein domain (MBD)-sequencing (86) are promising in this respect. However the technical limitations (such as sensitivity/specificity and resolution, but also bisulfite conversion, CpG coverage, number of methylated CpG sites, choice of region analyses, etc.) have to be considered, especially when reporting methylation data. Furthermore, hydroxymethylcytosine (hmC) has been discovered recently, but its role is as yet unknown. It is hypothesized that the presence of hmC in DNA can inhibit methyl-binding proteins, enzymatic functions, and gene expression. Enzymatic- or bisulfite-based approaches cannot discriminate between hydroxymethylcytosine and 5-methylcytosine because of structural similarity (87). Therefore, the possible presence of hmC should be considered in future methylation assay design.

The future discovery of clinically relevant hypermethylation markers would preferably be genome-wide and location- and CpG density-independent. In contrast, subsequent sequence-specific methylation analyses would need to be core-region specific. Careful and thorough experiment and assay design will lead to the development of sensitive and specific hypermethylation markers that can be used for early detection of cancer and prediction of prognosis and response to anticancer therapy. These methods enable independent validation by studying the same core regions, accurate identification of the biologically relevant location of hypermethylation, and translation of data into an accurate biomarker assay.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Grant Support

This work was supported by the Center for Translational Molecular Medicine (grant 03O-101).

Received December 23, 2010; revised March 14, 2011; accepted April 22, 2011; published OnlineFirst May 10, 2011.

References


37. Yoshikawa H, Matsubara K, Gian GS, Jackson P, Groopman JD, Manning JE, et al. SOCS-1, a negative regulator of the JAK-STAT pathway, is silenced by methylation in human hepatocellular carcino-


40. Lee PS, Teaberry VS, Bland AE, Huang Z, Whittaker RS, Baba T, et al. Elevated MAL expression is accompanied by promoter hypermethyla-


47. Graft JR, Herman JG, Myohanen S, Baylin SB, Vertino PM. Mapping patterns of CpG island methylation in normal and neoplastic cells implicates both upstream and downstream regions in de novo methyla-


Clinical Cancer Research

Analysis of Promoter CpG Island Hypermethylation in Cancer: Location, Location, Location!


Clin Cancer Res 2011;17:4225-4231. Published OnlineFirst May 10, 2011.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-10-3394

Cited articles
This article cites 85 articles, 37 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/17/13/4225.full.html#ref-list-1

Citing articles
This article has been cited by 13 HighWire-hosted articles. Access the articles at:
/content/17/13/4225.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.