Global Hypomethylation Identifies Loci Targeted for Hypermethylation in Head and Neck Cancer

Graham M. Poage¹, E. Andres Houseman²,4, Brock C. Christensen²,3, Rondi A. Butler², Michele Avisser-Whiting³, Michael D. McClean⁵, Tim Waterboer⁶, Michael Pawlita⁶, Carmen J. Marsit²,3, and Karl T. Kelsey²,3

Abstract

Purpose: The human epigenome is profoundly altered in cancers, with a characteristic loss of methylation in repetitive regions and concomitant accumulation of gene promoter methylation. The degree to which these processes are coordinated is unclear so we investigated both in head and neck squamous cell carcinomas.

Experimental Design: Global methylation was measured using the luminometric methylation assay (LUMA) and pyrosequencing of LINE-1Hs and AluYb8 repetitive elements in a series of 138 tumors. We also measured methylation of more than 27,000 CpG loci with the Illumina HumanMethylation27 Microarray (n = 91).

Results: LINE-1 methylation was significantly associated with LUMA and Infinium loci methylation (Spearman’s \( r = 0.52/0.56 \), both \( P < 0.001 \)) but not that of AluYb8. Methylation of LINE-1, AluYb8, and Infinium loci differed by tumor site (each Kruskal–Wallis, \( P < 0.05 \)). Also, LINE-1 and LUMA methylation were associated with HPV16 E6 serology (each Mann–Whitney, \( P < 0.05 \)). Comparing LINE-1 methylation to gene-associated methylation, we identified a distinct subset of CpG loci with significant hypermethylation associated with LINE-1 hypomethylation. An investigation of sequence features for these CpG loci revealed that they were significantly less likely to reside in repetitive elements (Gene Set Enrichment Analysis, \( P < 0.02 \)), enriched in CpG islands (\( P < 0.001 \)) and were proximal to transcription factor–binding sites (\( P < 0.05 \)). We validated the top CpG loci that had significant hypermethylation associated with LINE-1 hypomethylation (at EVI2A, IFRD1, KLHL6, and PTPRCAP) by pyrosequencing independent tumors.

Conclusions: These data indicate that global hypomethylation and gene-specific methylation processes are associated in a sequence-dependent manner, and that clinical characteristics and exposures leading to HNSCC may be influencing these processes. Clin Cancer Res; 17(11); 3579–89. ©2011 AACR.

Introduction

Epigenetic regulation is central to the biological function of all cells. Alteration to the patterns of CpG dinucleotide DNA methylation, particularly at sites with the ability to affect gene function (such as gene promoter regions), are commonplace in human diseases. Epigenetic hallmarks of malignant tumors, compared with their analogous normal tissues, include both decreases in global DNA methylation (hypomethylation—commonly assessed in DNA repeat regions) and concomitant increases (hypermethylation) in gene-associated (or locus-specific) methylation. Gene-associated hypermethylation is often associated with phenotypic silencing of tumor suppressor genes in cancer. In fact, gene promoter hypermethylation events during the pathogenesis of neoplasia may be as common, if not more frequent than the well-recognized genetic mechanisms of gene inactivation, like mutation and LOH (1). Head and neck squamous cell carcinomas (HNSCC) are prevalent throughout the world, representing the sixth most common malignancy overall and the eighth most common in U.S. males (excluding nonmelanoma skin cancers; ref. 2). The major etiologic risk factors are alcohol use, tobacco use, and human papillomavirus type 16.
Translational Relevance

Head and neck cancers are a significant health burden worldwide. As we transition toward more personalized medicine, molecular profiling of tumors is necessary to develop prevention measures and to aid oncologists in developing effective treatment plans for this disease. We have identified differences among global methylation indicators (LINE-1, Alu, and LUMA) that independently reflect patient characteristics such as tumor site and HPV16 infection. In addition, by integrating high-resolution methylation microarrays with global methylation data, we propose a novel method to identify loci involved in carcinogenesis that may be useful as biomarkers of disease and/or recurrence.

Translational Relevance

 Briefly, samples from incident cases of HNSCC were microscopically examined and histologically confirmed to have more than 70% tumor content by the study pathologist. Patients were enrolled on providing written, informed consent. This study was approved by the Brown University Institutional Review Board. Clinical information was collected and HPV16 status was assessed using a multiplex serology assay to detect antibodies against the HPV16 E6 protein according to previously published methods (14). In total, 138 fresh-frozen tumor specimens from head and neck sites (excluding nasopharyngeal carcinomas) were subjected to methylation analysis. Fresh-frozen normal (nondiseased) head and neck tissues (n = 18) taken at autopsy from the oral cavity, pharynx, and larynx were provided by the National Disease Research Interchange. In addition, peripheral blood, drawn from 213 cancer-free individuals living in New England (enrolled as part of a population-based study of bladder cancer) was studied as a comparison convenience sample; details of this population are described in the work of Wilhelm and colleagues (15).

DNA isolation and genome-wide methylation measurements

DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen) and 1 μg of genomic DNA was sodium bisulfite modified with the EZ DNA Methylation Kit (Zymo Research) as per the manufacturer’s instructions.

LINE-1 methylation was measured in all tumor and blood samples according to the procedures described by Bollati and colleagues (16).

AluYb8 subfamily methylation was measured as outlined by Choi and colleagues (17) with minor modifications. PCR was carried out on 40 ng of modified DNA per 25 μL reaction under the following conditions: denaturation 94°C, 2 minutes/94°C for 30 seconds (cycling); annealing 58°C, 30 seconds; extension 70°C, 30 seconds; 50 cycles. The first 5 of 6 CpG sites were used in analysis of AluYb8 methylation. Pyrosequencing reactions were carried out in triplicate and bisulfite conversion efficiency was monitored using internal non-CpG cytosine residues using the PyroMark Q96 MD System. For LINE-1 and AluYb8, methylation at each CpG position was calculated by taking the percentage of methylated signal divided by the sum of the methylated and unmethylated signals and reported as the mean overall positions.

The LUMA procedure (18) was carried out according to the modifications described in the work of Bjornsson and colleagues (19) with the nucleotide dispensation order (GTGTCACTGTGTG) to compensate for background signal from lower quality DNA. Methylation was estimated from the 1 − (HpaII/MspI) ratios. Because this procedure requires a significant amount of DNA, we selected 50 HNSCCs for which there was a sufficient amount of substrate. Restriction enzymes (HpaII, MspI, and EcoRI-HF) were purchased from New England Biolabs.

Study population/ethics

The study population was previously described (13). Briefly, samples from incident cases of HNSCC were
**Illumina inﬁnium humanmethylatjon27 microarray**

High-resolution methylation analyses of HNSCCs \( n = 91 \), normal tissue \( n = 18 \), and control bloods \( n = 213 \) were conducted on the Illumina Inﬁnium HumanMethylation27 Microarray Platform. This BeadChip assay measures methylation \( (20) \), given as a \( \beta \) value ranging from 0 to 1, at more than 27,000 CpG loci. Arrays were processed at the UCSF Institute for Human Genetics, Genomics Core Facility according to the manufacturer’s protocol. Data were assembled in BeadStudio without normalization, as instructed by Illumina. Array control probes were used to assess sample performance. Speciﬁcally, the multivariate characteristics of array control probes based on ﬁtted mean vector and variance-covariance matrix (Mahalanobis distance) were used to screen outliers. Sex chromosome loci \( n = 1,092 \) were excluded to avoid gender-speciﬁc methylation bias, resulting in a ﬁnal data set that consisted of 26,486 autosomal loci associated with 13,890 genes. Sequence context information, such as CpG island status (as deﬁned by the work of Takai and colleagues; ref. 21) and transcription factor–binding site (TFBS) proximity, was extracted from UCSC Genome Table Browser \( (N C B I 3 6 / b g 1 8 \) assembly) with “repeatmasker v3.2.7” and “TFBS Conserved” (TFBS z-score \( > 2 \) ) tracks.

**Statistical analysis**

Methylation data were analyzed in R statistical software package v2.8.1 \( (h t t p : / / w w w . r - p r o j e c t . o r g ) \). Global methylation comparisons were conducted by computing Spearman’s rank correlation among the tumors measured for LINE-1, AluYb8, LUMA, and mean per-sampled Infinium array methylation. An omnibus \( P \) value was assigned by comparing the observed correlation estimate with the corresponding null distribution obtained by permutation \( (10,000 \) permutations). We tested for univariate associations with epidemiologic factors as predictors of global methylation using a permutation test \( (10,000 \) permutations) with the Kruskal–Wallis or Mann–Whitney \( U \) test statistic for categorical predictors and Spearman correlations for continuous data. Multivariable regression models were used to control for the HNSCC risk factors age, HPV16 serology, site, stage, drinking, and smoking.

LINE-1 methylation was ﬁrst measured in a core set of 138 HNSCCs. Microarray analysis was next conducted on 91 HNSCCs (one 96-well plate; limited by resources) which had the most complete clinical/exposure covariate information available. Among these 91 samples, 5 failed AluYb8 pyrosequencing, leaving 86 tumors. Finally, as each LUMA assay requires a substantial amount of DNA, sufﬁcient substrate was only available to allow for analysis of 50 tumors. In multivariable models, the number of samples reﬂects those with complete data for all variables in the model.

To discern the relationships between LINE-1 and array CpGs with coordinate methylation, we used recursively partitioned mixture modeling (RPMM), a computationally efﬁcient likelihood-based method of hierarchical clustering \( (22) \), to cluster autosomal CpGs into 16 classes based on their methylation pattern. Although metric (nonparametric) hierarchical clustering is a well-characterized method, it does not scale to tens of thousands of observations and seems to have poorer clustering consistency in this context (see Supplementary Text). Individual CpG–speciﬁc array values \((\beta’s)\) were subsequently averaged together, within each class, to form 16 average methylation values (per individual, 1 for each class). This dimensionality reduction approach combines data from CpGs with similar patterns, thereby attenuating biochemical noise and reducing multiple comparisons \( (23) \). We subsequently determined the methylation class–speciﬁc Spearman correlation coefﬁcients with LINE-1 methylation after stratifying the tumors by site. The null distribution of the multivariate 16-dimensional correlation statistic was obtained by randomly permuting the LINE-1 values with respect to their corresponding methylation proﬁles \( 10,000 \) times. An omnibus test of signiﬁcance was conducted by comparing the observed maximum absolute value over the 16 correlations with the corresponding quantity over the permutation distribution.

To investigate the CpG–speciﬁc correlations with LINE-1 methylation, we stratified the tumors by site and calculated Spearman correlations at all autosomal array loci. \( P \) values were corrected for multiple comparisons (using the \( q \) value package in R) and signiﬁcant \((q < 0.05)\) loci were grouped by correlation sign (±). To interrogate these loci for their representation among genomic sequence elements, we utilized Gene Set Enrichment Analysis (GSEA; ref. 24). GSEA allows for evaluation of overrepresentation among loci associated with genomic sequence elements among the rank-ordered list of CpGs correlated with LINE-1 methylation. The enrichment analysis uses a Kolmogorov–Smirnov–like statistic to estimate the proportion of loci at the top of the ranked list compared with all loci examined \( (n = 26,486) \). For this implementation, the procedure was implemented in R using the R function obtained from the Broad Institute GSEA web site, with Pearson correlations and a permutation null distribution generated with 1,000 permutations.

**Array validation**

Pyrosequencing assays for the top 5 CpG loci negatively associated with LINE-1 were designed, and 4 of these 5 assays (associated with the genes EVI2A, IFRD1, KHL6, and PTPRCA) were amenable to sequencing. Pyrosequencing was conducted both on our original oral/pharyngeal tumors \( n = 59 \) and in an independent set of 48 paraffin-embedded HNSCCs. Sample collection and extraction of substrate is described elsewhere \( (3) \). Spearman correlations were again calculated and signiﬁcance determined via permutation test. Primer sequences are listed in Supplementary Table S1.
Global measures of methylation are correlated

To examine the nature and distribution of global methylation in a case series of HNSCCs (summarized in Table 1), we bisulfite pyrosequenced LINE-1 and AluYb8 consensus repetitive element sequences; measured CCGG simple sequence repeat methylation using the luminometric methylation assay (LUMA) and measured locus-specific DNA methylation with Illumina’s Infinium microarray (26,486 autosomal CpG loci). The data for each global methylation indicator are shown in Figure 1. Specifically, the tumor-to-tumor variability in methylation was greatest for LINE-1 [median: 60%, interquartile range (IQR): 18.3%] and LUMA (median: 62%, IQR: 11.4%). Conversely, AluYb8 (median: 85%, IQR: 3.0%) and mean Infinium methylation (median: 27%, IQR: 3.3%) values were much more tightly distributed.

Next, the association between markers of global methylation was investigated. All correlations between markers of global methylation (except between AluYb8 and LUMA) were statistically significant (Supplementary Table S2 and Supplementary Fig. S1). AluYb8 methylation was least associated with all other indicators of global methylation. Despite the differences in measurement procedures and nature of the loci examined, the strongest correlation was between LINE-1 methylation and mean array CpG methylation (Spearman correlation $\rho = 0.56$). In addition, there were clear relationships between LUMA and LINE-1 methylation as well as between LUMA and Infinium methylation, suggestive of a link between global- and gene-targeted methylation.

Global methylation measures are associated with tumor site and HPV16 seropositivity

Because LINE-1 methylation has previously been shown to be modified by environmental factors in a number of diseases (11, 16) and we observed associations between LINE-1 and other measures of global methylation, we hypothesized that clinical and exposure characteristics may alter global methylation indicators. Univariate analysis revealed a significant relationship between tumor site and both LINE-1 methylation and average array CpG methylation (each Kruskal–Wallis, $P < 0.03$) but not with other global methylation measures.

### Table 1. Clinicopathologic characteristics of study participants

<table>
<thead>
<tr>
<th></th>
<th>All cases ($N = 138$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, $n$ (%)</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>38 (28)</td>
</tr>
<tr>
<td>Male</td>
<td>100 (72)</td>
</tr>
<tr>
<td>Age at diagnosis, y</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>32–91</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>61 (12.0)</td>
</tr>
<tr>
<td>HPV16 E6 serology, $n$ (%)$^a$</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>12 (12)</td>
</tr>
<tr>
<td>Negative</td>
<td>92 (88)</td>
</tr>
<tr>
<td>Tumor site, $n$ (%)$^b$</td>
<td></td>
</tr>
<tr>
<td>Oral</td>
<td>76 (64)</td>
</tr>
<tr>
<td>Pharynx</td>
<td>24 (20)</td>
</tr>
<tr>
<td>Larynx</td>
<td>19 (16)</td>
</tr>
<tr>
<td>Clinical stage, $n$ (%)$^c$</td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>9 (7)</td>
</tr>
<tr>
<td>II</td>
<td>27 (21)</td>
</tr>
<tr>
<td>III</td>
<td>24 (19)</td>
</tr>
<tr>
<td>IV</td>
<td>67 (53)</td>
</tr>
<tr>
<td>Lifetime drink-years of consumption, $n$ ($^d$)</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0–61</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>26 (20)</td>
</tr>
<tr>
<td>Never-drinkers, $n$</td>
<td></td>
</tr>
<tr>
<td>Lifetime pack-years smoked, $n$ ($^e$)</td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>0–135</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>35 (31)</td>
</tr>
<tr>
<td>Never-smokers, $n$</td>
<td></td>
</tr>
</tbody>
</table>

$^a$Thirty-four samples missing serology data.

$^b$Nineteen samples missing site data.

$^c$Eleven tumors missing stage data.

$^d$Twenty-two patients missing self-reported drinking data.

$^e$Seventeen patients missing self-reported smoking data.
In a multivariable regression model controlled for potential confounders, LINE-1, AluYb8, and average array CpG methylation were significantly associated with tumor site (all 3 values of \( P < 0.05 \)). For each global methylation measure, laryngeal tumors had decreased methylation compared with pharyngeal or oral tumors (Table 2).

HPV16 infection is an established risk factor for HNSCC and was evaluated for association with global methylation measures. In the adjusted regression model, HPV16 E6 antibody seropositivity was significantly associated with an increase in LINE-1 (\( P < 0.05 \)) and LUMA methylation (\( P < 0.03 \); Table 2). Although, the prevalence of HPV-positive tumors differed by site (highest in pharyngeal tumors), we did not have sufficient data for an analysis stratified by additional tumor subsites (Supplementary Table S3). There were no significant associations between any of the global methylation measures and gender, race, patient age at diagnosis, combined TNM (tumor, node, metastasis) stage, body mass index, or smoking intensity/duration. Because the number of samples for each model differed due to missing covariate data, we conducted a subset analysis and determined that all tumor subsets were representative of the overall HNSCC population (data not shown).

**CpG cluster methylation patterns are correlated with global methylation measures in a sequence-dependent manner**

As the data suggest a general correlation between global measures of methylation, we next focused on the relationship between LINE-1 methylation and array CpG methylation using RPMM. RPMM has been previously described in detail (22) and applied in numerous DNA methylation array analyses (3, 23, 25). In contrast to computationally inefficient hierarchical clustering methods, this approach robustly classifies loci (or samples) into biologically informed groups whose methylation patterns can be examined for associations with covariates. All array autosomal CpG loci were clustered into 16 methylation classes (median group: \( n = 1,763 \) CpGs, range = 763–2,821 CpGs). For each class, we computed the Spearman correlation between average class methylation and LINE-1 methylation (Fig. 3A). Among the 16 CpG methylation classes, mean methylation was significantly positively correlated with LINE-1 methylation in the 7 classes that had a lower CpG island locus prevalence (Supplementary Fig. S2). No negative correlations between CpG class methylation and LINE-1 methylation were observed by class. To determine whether the observed positive correlations are unique to HNSCCs, an identical analysis was conducted in normal (Fig. 2).
Table 2. Multivariable regression for global measures of methylation on patient characteristics

<table>
<thead>
<tr>
<th>Predictors</th>
<th>L1Hs (n = 81)</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>AluYb8 (n = 39)</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>95% CI</td>
<td>P</td>
<td>Estimate</td>
<td>95% CI</td>
<td>P</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.2</td>
<td>(0.0, 0.5)</td>
<td>0.06</td>
<td>−0.01</td>
<td>(−0.08, 0.06)</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HPV16 E6 serology</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Negative</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>9.8</td>
<td>(0.8, 18.8)</td>
<td>0.04</td>
<td>1.00</td>
<td>(−1.2, 3.3)</td>
<td>0.38</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor site</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oral</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pharynx</td>
<td>−6</td>
<td>(−13.6, 1.7)</td>
<td>0.13</td>
<td>−0.29</td>
<td>(−2.0, 1.4)</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larynx</td>
<td>−14.2</td>
<td>(−22.5, −5.8)</td>
<td>&lt;0.01</td>
<td>−1.98</td>
<td>(−3.7, −0.3)</td>
<td>0.03</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I and II</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td>ref</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III and IV</td>
<td>−5.7</td>
<td>(−11.8, 0.5)</td>
<td>0.07</td>
<td>−0.23</td>
<td>(−1.8, 1.3)</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lifetime drink-years</td>
<td>0.05</td>
<td>(−0.01, 0.21)</td>
<td>0.50</td>
<td>0.04</td>
<td>(−0.02, 0.09)</td>
<td>0.21</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>smoked</td>
<td>−0.01</td>
<td>(−0.10, 0.08)</td>
<td>0.80</td>
<td>−0.01</td>
<td>(−0.03, 0.02)</td>
<td>0.56</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NOTE: Models controlled for all variables listed. Results in bold are considered significant where the 95% CI does not cross zero.

Figure 3. Classes of array CpG loci are associated with global methylation levels in tumors but not in normal tissues. RPMM is used to cluster loci with similar methylation β values into 16 methylation classes. Spearman’s correlation coefficients are calculated by comparing the mean methylation of class member loci for each sample to LINE-1 methylation level and plotted by class. Correlation points are colored according to their mean class methylation (as indicated in the color sidebars). Classes were ordered according to the percentage of their member loci mapping to CpG island regions. A separate RPMM model was applied to the data for each tissue type: (A) HNSCCs (tumor), (B) nondiseased head and neck tissues (normal), and (C) normal peripheral blood (blood). Blue shaded regions represent the 95% confidence limits of the observed maximum correlation from 10,000 random permutations (representing the null distribution). Therefore, correlations lying outside of these regions are considered statistically significant. Permutation test omnibus P values are displayed at the bottom of each panel. Dotted blue lines indicate the zero correlation axis.
head and neck tissues \( (n = 18) \) and a convenience sample of DNA extracted from peripheral blood of healthy individuals \( (n = 213) \). In both of these tissue types, no RPMM CpG methylation classes were significantly associated with LINE-1 methylation (Fig. 3B and C). Because LINE-1 retrotransposons are highly methylated in the normal genome, we compared LINE-1 methylation of all 138 HNSCCs with that of the 18 normal head and neck tissues. This analysis revealed that 126 (92%) tumors had less LINE-1 methylation than the least LINE-1-methylated normal sample.

**LINE-1-associated loci are enriched for sequence-dependent elements**

We next sought to identify which individual loci were significantly correlated with LINE-1 hypomethylation. Because LINE-1 methylation varied by tumor site, we stratified the tumors (larynx or oral/pharynx), then calculated Spearman correlations between LINE-1 methylation and all autosomal array CpG loci in a locus-by-locus fashion for each tumor site. Following correction for multiple comparisons, 3,029 CpG loci (11%) and 7,367 CpG loci (28%) were significantly correlated with LINE-1 methylation \( (p < 0.05) \) in the laryngeal and oral/pharyngeal analyses, respectively. Among the top 2,000 most highly associated loci for both sites, about 97% were positively correlated with LINE-1 methylation (hypomethylated CpG loci and hypomethylated LINE-1); a small group of CpGs [37 loci (2%) in the oral cavity/pharynx and 60 loci (3%) in the larynx] were negatively correlated with LINE-1 methylation (hypermethylated CpG and hypomethylated LINE-1; see Supplementary Table S4 as compared with Supplementary Table S5).

Our previous analyses have indicated an interaction between epigenetic and genetic alterations in cancer (26, 27). We therefore investigated the hypothesis that the genomic sequence context influences selection for de novo methylation in the LINE-1 negatively correlated CpG loci using GSEA (i.e., determination of the proportion of LINE-1 negatively correlated CpG loci are found at the top of a ranked correlation scores among all loci). These CpGs were significantly more prevalent in CpG islands for oral/pharyngeal tumors \( (p < 0.02) \), near TFBS (both sites \( p < 0.05 \), Fig. 4), and significantly depleted within repetitive elements \( (p < 0.05 \), Supplementary Fig. S3) among all the loci analyzed by microarray.

Bsulfite pyrosequencing was used for array validation of the top oral/pharyngeal CpG loci (within the EVI2A, IFRD1, KLHL6, and PTPRCA gene regions) negatively correlated with LINE-1 methylation. Tumors from our original fresh-frozen oral/pharyngeal population, where DNA remained \( (n = 59 \) of 100), and tumors from an independent set of HNSCCs \( (n = 48) \), representing all tumor sites, were pyrosequenced. In this validation, all 4 of the CpG loci examined were significantly negatively correlated with LINE-1 methylation (all permutation test \( p < 0.01 \), Supplementary Fig. S4). To ensure the statistical rigor of this novel method for identification of molecularly selected loci, a random CpG (associated with CD22) was pyrosequenced for the independent validation tumor set, and no significant correlation with LINE-1 methylation was observed \( (p = -0.03, \ p = 0.84) \). We also investigated the possibility that clinical stage might be a significant predictor of methylation in these CpGs but no such association was found.

Finally, we determined whether the CpGs with discordant LINE-1 methylation in tumors were also discordant in all normal head and neck tissues and peripheral blood samples using Infinium array CpG methylation at EVI2A, IFRD1, KLHL6, and PTPRCA. Consistent with this being a tumor-specific phenomenon, among all 4 CpGs, no sig-
significant negative correlations were observed in normal head and neck tissues or peripheral blood samples (minimum $\rho = -0.24$, lowest $P = 0.32$).

Discussion

Malignant transformation is a complex process characterized by the accumulation of genetic and epigenetic abnormalities. The patterns of DNA methylation alterations in cancer include both genome-wide hypomethylation and gene-specific hypermethylation. Gene silencing associated with hypermethylation can prevent transcription of critical proteins; however, the consequences of hypomethylation of repetitive DNA sequences are less clear. Furthermore, the nature of the relationship between locus-specific methylation and global DNA methylation alterations is underexplored. Here, we extensively examined both global methylation and gene-specific methylation in HNSCC. We showed that different markers of global methylation are correlated but that LINE-1, LUMA, and $AluYb8$ methylation.

We also investigated the relationship between global and CpG-specific methylation and found that LINE-1 hypomethylation was correlated with significantly increased CpG methylation at only a small number of loci. GSEA analysis indicated that these CpG loci were enriched in CpG islands and TFBSs but depleted in repetitive elements, indicating that genomic context may play a role in this coordination.

Methylation of CCGG repeat DNA (as measured by LUMA), LINE-1, and $AluYb8$ are commonly used as indicators of global methylation. We observed these to be generally correlated, but with different degrees of variance, consistent with previous comparisons between LINE-1 and other global markers in normal blood (16, 28) and various human cancers (9, 10, 17, 29–38). The basis for their differential variance is not well understood, though unequal genomic distributions (39), evolutionary age of retrotransposable elements, and sequence context may all contribute (40). Our observations stress the need to consider interassay differences when comparing and interpreting global methylation measures.

Methylation of LINE-1 by pyrosequencing has previously been measured in HNSCC (38). Univariate analyses in that study identified decreased methylation in HNSCC tumors compared with nondiseased tissue of the aerodigestive tract and tumor-only associations with tobacco, alcohol, and stage. Consistent with this, we identified a (nonsignificant) decrease in methylation for high-stage tumors. Tumor stage was not a significant predictor of LINE-1 methylation in the Smith and colleagues (38) multivariable analysis, but our combined studies suggest a potential association of global hypomethylation with more advanced disease. This association has been noted in other cancers (41).

We have identified hypermethylated CpG loci that are significantly negatively correlated with LINE-1 methylation in HNSCC. The fact that locus-specific methylation increases, whereas global methylation decreases, suggests that these CpGs are actively selected to maintain the tumor phenotype through DNA methylation. We next explored the possibility that genomic sequence context may influence the coordination between global and gene-specific methylation. Highlighting the influence of genome architecture on gene-specific methylation, Estécio and colleagues more generally examined the link between repetitive elements and gene-specific...
methylated in bladder cancer (42). They showed that
the predisposition for de novo methylation of CpG
islands is associated with the absence of transposable
elements. Consistent with their findings, we have shown
that hypermethylated CpGs negatively correlated with
LINE-1 methylation are more likely to be in CpG
islands. Furthermore, these CpGs were less likely to
reside in transposable elements and more likely to be
located near TFBBSs. We did not identify an association
with specific transcription factors prevalent in these
regions, but because of the unique and critical functions
that transcription factors provide, the intersection of
gene-specific methylation, repeat methylation, and
transcription factor presence represents an intriguing
area for further exploration. One study has described
an overall enrichment for DR2 retinoic acid response
elements (RARE) within AluS repetitive elements, yet
this may be explained by the spontaneous deamination
of a single CpG within AluS elements to generate a RARE
(43). Further research using larger tumor populations
and additional cancer types may reveal additional
insight into these interactions.

Among the top CpGs whose methylation was nega-
tively correlated with LINE-1 methylation, many of the
associated genes are frequently dysregulated or epigen-
tically repressed in cancers. Our locus-by-locus analysis
identified CpGs associated with CDKN2A (Supplemen-
tary Table S4), a well-known tumor suppressor gene
commonly silenced in HNSCC (4) and other cancers.
In addition, CpGs associated with 2 genes known to be
hypermethylated in breast tumors were also identified
SIM1 and GHSR (Supplementary Table S4). SIM1
methylation has been identified as a biomarker for
early-stage cancers (44), whereas methylation of the
GHSR gene distinguishes infiltrating ductal carcinoma
from normal and benign tissues with a high sensitivity
and specificity (45). To our knowledge, these genes have
not been explored previously in HNSCC and further
studies are required to clarify their role in cellular
transformation. Among the genes that were validated
by pyrosequencing (commonly altered in the oral cav-
ity/pharynx or also in the larynx), only the tumor
suppressor IFRD1 has a well-defined role, specifically
in lung carcinogenesis (46).

There were a surprisingly small number of CpG loci
whose methylation was negatively correlated with that of
LINE-1 among the top 2,000 significantly altered CpGs. One
potential mechanism to explain this discrepancy
involves a recent modification to the maintenance CpG
methylation theory (in mother to daughter cells). In this
proposed model, DNMT3A/DNMT3B (classically consid-
ered to be de novo methyltransferases) act cooperatively
to maintain DNA methylation marks in CpG islands and
repeat regions, whereas DNMTT1 (the classic mainte-
nance DNA methyltransferase) maintains non-CpG island
methylation (47). A follow-up study by Sharma and col-
leagues (48) showed a self-regulatory mechanism between
global methylation and expression of DNMT3A/DNMT3B,
whereby a reduced global methylation results in a dimin-
ished capacity for aberrant de novo methylation because of
the degradation of DNMT3A/DNMT3B enzymes. Applying
this concept to cancer studies (in the context of genom-
wide hypomethylation), hypermethylation would be
expected to occur primarily at gene loci within dynamic
regulatory trafficking centers where the potential for pro-
tein stabilization is highest such as those inside of TFBBSs,
CpG islands, and outside of repeat elements (Fig. 4 and
Supplementary Fig. S3).

Collectively, our data support the hypothesis that gene
regulation is coordinated through global and locus-specific
methylation at certain sequence elements important for
cellular differentiation and carcinogenesis. Importantly,
our locus-by-locus comparison of CpG-specific methyla-
tion with LINE-1 methylation provides a novel route to
identify selectively hypermethylated genes in cancers.
Further studies in additional tumor types are necessary
to clarify the mechanistic basis for and to expand on these
findings.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank Devin C. Koestler, Charlotte S. Wilhelm-Benartzi, and
William P. Acomando, Jr.

Grant Support

This study was supported by the Flight Attendant Medical Research Institute
(C.J. Marsit) and the NIH grants CA078609, CA100679 (K.T. Kelsey). The
costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received January 6, 2011; revised March 11, 2011; accepted April 5, 2011; published OnlineFirst April 19, 2011.

References

1. Jones PA, Baylin SB. The fundamental role of epigenetic events in
MR, Yeh RF, et al. Epigenetic profiling reveals etiologically distinct
patterns of DNA methylation in head and neck squamous cell carci-
4. Hasegawa M, Nelson HH, Peters E, Ringstrom E, Posner M,
Kelsey KT. Patterns of gene promoter methylation in squa-
mous cell cancer of the head and neck. Oncogene 2002;21:
4231–6.
factors, clinical characteristics, and survival in 235 patients
with laryngeal and hypopharyngeal cancers. Cancer 2007;110:
1745–51.


Clinical Cancer Research

Global Hypomethylation Identifies Loci Targeted for Hypermethylation in Head and Neck Cancer

Graham M. Poage, E. Andres Houseman, Brock C. Christensen, et al.

Clin Cancer Res 2011;17:3579-3589. Published OnlineFirst April 19, 2011.

Updated version
Access the most recent version of this article at:
doi:10.1158/1078-0432.CCR-11-0044

Supplementary Material
Access the most recent supplemental material at:
http://clincancerres.aacrjournals.org/content/suppl/2011/04/21/1078-0432.CCR-11-0044.DC1

Cited articles
This article cites 47 articles, 11 of which you can access for free at:
http://clincancerres.aacrjournals.org/content/17/11/3579.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at:
http://clincancerres.aacrjournals.org/content/17/11/3579.full#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, use this link
http://clincancerres.aacrjournals.org/content/17/11/3579.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.