Integrated Systems and Technologies

Cancer Research

Identification of an Epigenetic Profile Classifier That Is Associated with Survival in Head and Neck Cancer

Graham M. Poage[1,2], Rondi A. Butler[3], E. Andrés Houseman[3,5], Michael D. McClean[6], Heather H. Nelson[7,8], Brock C. Christensen[9,10], Carmen J. Marsit[9,10], and Karl T. Kelsey[3,4]

Abstract

Panels of prognostic biomarkers selected using candidate approaches often do not validate in independent populations, so additional strategies are needed to identify reliable classifiers. In this study, we used an array-based approach to measure DNA methylation and applied a novel method for grouping CpG dinucleotides according to well-characterized genomic sequence features. A hypermethylation profile among 13 CpG loci, characterized by polycomb group target genes, mammalian interspersed repeats, and transcription factor–binding sites (PcG/MIR/TFBS), was associated with reduced survival (HR, 3.98; $P = 0.001$) in patients with head and neck squamous cell carcinoma. This association was driven by CpGs associated with the *TAP1* and *ALDH3A1* genes, findings that were validated in an independent patient group (HR, 2.86; $P = 0.04$). Together, the data not only elucidate new potential targets for therapeutic intervention in head and neck cancer but also may aid in the identification of poor prognosis patients who may require more aggressive treatment regimens. *Cancer Res; 72(11); 2728–37. ©2012 AACR.*

Introduction

Head and neck squamous cell carcinoma (HNSCC) arises from the mucosal lining of the larynx, pharynx, and oral cavity. More than 40,000 new cases develop annually in the United States, and despite advances in treatment regimens, 5-year survival rates near 50% have declined only modestly over the past 30 years (1). The major determinants of patient prognosis include patient age, tumor stage and site, lifetime duration of alcohol/tobacco exposure, and oncogenic human papillomavirus 16 (HPV16) infection, although these do not fully inform patient outcome and are used conservatively in treatment planning (2). Identification of additional prognostic molecular biomarkers is needed to provide clinicians with better therapeutic decision-making tools. Candidate gene approaches have been used in past studies of head and neck cancer to identify additional biomarkers of prognosis, although their results have almost uniformly been poorly replicated in independent studies (3–5). Thus, using discovery-based approaches coupled with appropriate validation may provide novel markers of greater use.

Epigenetic silencing of genes via promoter hypermethylation is a critical event in the development of many cancers, including HNSCCs (6–9). Because of the inherent stability of 5-methylcytosine, DNA methylation is an attractive epigenetic indicator for measurement. Consistent with studies of candidate genetic markers, candidate studies using DNA methylation as a predictive tool have uncovered novel biomarkers of survival (10). However, many prior studies suffer from a limited ability to detect strong overall survival associations and these identified associations may vary widely among different populations (11–13). Heterogeneous patterns of somatic alterations across tumors can decrease the use of single candidate genes as biomarkers, and molecular profile–based markers that address this challenge are increasingly pursued.

Previous studies by our group and others have shown that profiles of cancer-related epigenetic alterations are dependent upon the local genomic architecture (11, 14–16). In fact, Lienert and colleagues (17) recently described methylation-determining regions of short DNA fragments whose sequence content [CpG density, transcription factor–binding sites (TFBS)] largely defines methylation patterning in mouse embryonic stem cells. Therefore, we compared DNA methylation grouped by specific sets of phenotypically well-understood DNA sequence elements to ask whether methylation profiles of these sets were associated with the survival times of patients with HNSCCs. We measured methylation genome-wide, at more than 26,377 CpG

Authors' Affiliations: [1]Department of Clinical Cancer Prevention, The University of Texas MD Anderson Cancer Center, Houston, Texas; Departments of [2]Molecular Pharmacology and Physiology [3]Epidemiology, and [4]Pathology and Laboratory Medicine, Division of Biology and Medicine, Brown University, Providence, Rhode Island; [5]Department of Public Health, Oregon State University, Corvallis, Oregon; [6]Department of Environmental Health, Boston University, Boston, Massachusetts; [7]Department of Epidemiology, [8]Masonic Cancer Center, University of Minnesota, Minneapolis, Minnesota; and Departments of [9]Pharmacology and Toxicology and [10]Community and Family Medicine, Giesel School of Medicine, Dartmouth College, Hanover, New Hampshire

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Microarray data from this study have been contributed to the NCBI Gene Expression Omnibus under the accession number GSE25093 (http://www.ncbi.nlm.nih.gov/geo).

Corresponding Author: Karl T. Kelsey, Brown University, 70 Ship Street, Box G-E4, Providence, RI 02903. Phone: 401-863-6420; Fax: 401-863-9008; E-mail: Karl_Kelsey@Brown.edu

doi: 10.1158/0008-5472.CAN-11-4121-T

loci, and defined novel groups of loci on the basis of their proximity to functional sequence elements.

Materials and Methods

Samples and study population

The initial discovery phase study population has been previously described (18) and an independent validation population was drawn from Boston-area hospitals between 2003 and 2007, in an extension of the original case–control study of HNSCCs. Tumor samples (encompassing all head and neck sites except nasopharyngeal carcinomas) from incident cases were microscopically examined by the study pathologist and histologically confirmed to have more than 70% tumor content. Patients were enrolled after providing written, informed consent, as approved by all of the participating Institutional Review Boards. Clinical information was collected and HPV16 status assessed using short-fragment PCR to amplify a region of the *L1* gene of HPV16, according to previously published methods (19). In total, 91 fresh-frozen tumor specimens from the discovery patient population and 101 formalin-fixed, paraffin-embedded (FFPE) tumors from the validation population were subjected to methylation analysis. All-cause patient survival data were obtained from the National Death Index and survival was tracked for 8 years after diagnosis when possible.

DNA isolation and methylation array procedures

DNA was extracted from fresh-frozen tumors and 18 clinically normal head and neck tissues sourced from the National Disease Research Interchange using the DNeasy Blood and Tissue Kit (Qiagen) or from 20-μm FFPE sections according to previously published procedures (20). One microgram of genomic DNA was sodium bisulfite–modified using the EZ DNA Methylation Kit (Zymo Research) as per the manufacturer's instructions. DNA methylation was measured in the discovery patient population on a genome-wide scale using the Illumina Infinium HumanMethylation27 Microarray Platform (Illumina). This BeadChip assay measures methylation (21), given as a β-value ranging from 0 to 1, at more than 27,000 CpG loci. Arrays were processed at the University of California San Francisco (San Francisco, CA) Institute for Human Genetics, Genomics Core Facility, according to the manufacturer's protocol. Data were assembled in BeadStudio without normalization, as recommended by the manufacturer. 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 for outlying samples (although none were found). Sex chromosomal loci ($n = 1{,}092$) were excluded to avoid gender-specific methylation bias. As single-nucleotide polymorphisms near the interrogated CpG site are known to induce spurious signals in the Illumina platforms, we excluded an additional 109 probes, which resulted in a final data set of 26,377 autosomal loci associated with 13,856 genes.

Statistical analysis

Methylation data were analyzed in R statistical software v2.8.1 (http://www.r-project.org). Differential methylation (hypermethylation/hypomethylation) was determined by comparing the distributions of the methylation β-values for each locus between tumor samples and normal tissue. A nonparametric Cox–Wilcoxon rank-sum test was used to determine significance (false discovery rate, $q < 0.05$), and a threshold of $|\Delta\beta| > 0.2$ was imposed to identify potentially meaningful biologic changes. All 26,377 autosomal array CpG loci were clustered into mutually exclusive groups on the basis of the following genomic functional sequence element criteria: CpG island status (ref. 22; obtained from the array annotation file), polycomb group (PcG) target status of the gene associated with the CpG [i.e., gene was described as a PcG target in at least one of the study of Bracken and colleagues (23), Lee and colleagues (24), Schlesinger and colleagues (25), or Squazzo and colleagues (26)], presence within 1 kb of, at least, one of 258 computationally predicted TFBS [sequences obtained from the *tfbsConsSites* track of the UCSC Genomes Table Browser (NCBI36/hg18 assembly, TFBS *z*-score > 2)], and presence within any of the following types of repetitive element as defined by the *Repeatmasker* v3.2.7 track within Genomes Browser: *Alu*, *LINE-1*, *LINE-2*, and MIR sequences. Methylation array β-values were averaged across all CpG loci within each bioinformatic class to generate an aggregate methylation value and tumors were grouped into high/intermediate/low categories depending on whether their specific methylation level was in the highest, middle, or lowest tertile across the population for each class.

Cox proportional hazards models were conducted to determine class-specific (and locus-specific) associations between methylation groups and prognosis and were controlled for patient age at diagnosis, anatomic site, combined American Joint Committee on Cancer (AJCC) stage, and HPV16 status. The null distribution of the 41-dimensional Cox *z*-statistic was obtained by randomly permuting aggregate methylation values with respect to the survival variables 10,000 times. Note here that the response variable is survival and the predictor is that of tumor methylation. An omnibus test of significance, analogous to a Kolmogorov–Smirnov test, was conducted by comparing the observed maximum for 41 *z*-scores with the corresponding quantity over the permutation distribution. We tested for univariate associations between epidemiologic factors and high/intermediate/low categorical tumor methylation using nonparametric Kruskal–Wallis ANOVA tests for continuous variables and a 10,000-permutation χ^2 test for categorical variables. An association was considered significant where $P < 0.05$.

Locus specificity and array validation

To determine the specificity of individual CpG locus methylation associations with patient outcomes, we assessed methylation at sites proximal to the index CpG loci (*ALDH3A1* and *TAP1*). CpG loci in different bioinformatic groups or in nearby genes were compared, using Spearman rho, for their concordance with the index CpG, using the array CpG methylation values. Pyrosequencing assays were designed for the top 2 CpG loci (corresponding to *ALDH3A1* and *TAP1* genes), which were associated with survival. Pyrosequencing reactions were conducted in triplicate according to the procedures described by

Bollati and colleagues (27), with the following modifications: the annealing temperatures were 62°C-*ALDH3A1*/55.5°C-*TAP1* and 47 (rather than 45) cycles were used in the case of *ALDH3A1*. Primers for this assay are provided in Supplementary Table S1. Sodium bisulfite conversion efficiency was monitored using internal non-CpG cytosine residues using the PyroMark Q96MD System. DNA methylation at each locus was calculated by taking the percentage of methylated signal divided by the sum of the methylated and unmethylated signals for each CpG. Nonparametric Spearman correlations were calculated between array CpG methylation and bisulfite pyrosequencing values. Tertile groups of methylation were determined across all tumors (within each data set) for each CpG and tumors were stratified into low-, intermediate-, or high-methylation groups, depending on their tertile membership. Survival was compared among groups using a log-rank test.

DNA methylation of *TAP1* and *ALDH3A1* loci were assessed in the independent validation patient set of 101 FFPE tumors. Pyrosequencing values of both loci were aggregated together (averaged) for each tumor, and patient survival was compared across methylation tertiles. Cox models were used to adjust associations for clinical and demographic factors.

Results

Bioinformatically clustered CpG loci are associated with patient survival

We determined the methylation state of 91 discovery phase HNSCCs at all 26,377 autosomal loci using the Illumina Infinium HumanMethylation27 Microarray. The methylation β-distribution of this data set is shown in Supplementary Fig. S1, where each CpG locus is averaged across all 91 tumors. Given the enrichment of promoter and early exonic CpG loci in this array, a large proportion of loci in this array are unmethylated (45% of all CpGs where $\beta < 0.10$) whereas a small proportion are highly methylated (0.5% of CpGs where $\beta > 0.90$). To identify CpG loci whose altered methylation is associated with prognosis, we used a supervised approach that leverages genomic architecture to inform locus selection. Array CpG loci were bioinformatically clustered on the basis of phenotypically well-understood DNA sequence elements into 41 mutually exclusive groups. Specifically, CpGs were clustered on the basis of their presence within *LINE-1*, *LINE-2*, *Alu*, or MIR repeats, CpG islands, genes that are known targets of PcG proteins, or within 1,000 bp of TFBS. Individual CpGs were only allowed to be members of a single bioinformatically derived cluster. We hypothesized that this approach would attenuate biochemical and biologic noise and reduce false discovery, thus providing increased power to detect significant associations with survival. Table 1 lists the DNA sequence elements that define each of the 41 CpG clusters and provides the number of CpGs in each cluster. Indicative of a preference for gene-rich regions in the array design, a single group of CpGs defined by being located in a CpG island and near a TFBS (CpGI|TFBS) contains nearly half of all autosomal array CpG loci. The remaining CpG loci were spread unevenly across the other bioinformatically derived clusters and 6% of CpG loci were not associated with any sequence elements that defined our clusters.

To investigate the relationship between these profiles of DNA methylation alterations and patient prognosis, we determined the average methylation value across all CpGs (aggregate methylation) among bioinformatic clusters for each of the tumor samples. For every bioinformatic cluster, tumors were sorted into high, intermediate, or low methylation subgroups, depending on whether their methylation level (among member CpG loci) fell into the highest, middle, or lowest tertile, respectively. An omnibus test for significance indicated that, across bioinformatic clusters, methylation was significantly

Table 1. Description of mutually exclusive bioinformatic clusters of methylation array loci

Cluster definition	Loci (*n*)	%Total	Cluster definition	Loci (*n*)	%Total	Cluster definition	Loci (*n*)	%Total
Not in any elements	1,618	6	CpGI|MIR|TFBS	253	1	PcG|LINE2|TFBS	6	<1
TFBS	4,436	17	CpGI|ALU	120	<1	PcG|LINE1	2	<1
MIR	98	<1	CpGI|ALU|TFBS	173	1	PcG|LINE1|TFBS	1	<1
MIR|TFBS	183	1	CpGI|LINE2	32	<1	CpGI|PcG	329	1
ALU	68	<1	CpGI|LINE2|TFBS	130	<1	CpGI|PcG|TFBS	2,687	10
ALU|TFBS	79	<1	CpGI|LINE1	24	<1	CpGI|PcG|MIR	4	<1
LINE2	88	<1	CpGI|LINE1|TFBS	35	<1	CpGI|PcG|MIR|TFBS	33	<1
LINE2|TFBS	103	<1	PcG	64	<1	CpGI|PcG|ALU	10	<1
LINE2|MIR|TFBS	1	<1	PcG|TFBS	313	1	CpGI|PcG|ALU|TFBS	21	<1
LINE1	73	<1	PcG|MIR	8	<1	CpGI|PcG|LINE2	4	<1
LINE1|TFBS	45	<1	PcG|MIR|TFBS	13	<1	CpGI|PcG|LINE2|TFBS	7	<1
CpGI	2,341	9	PcG|ALU	3	<1	CpGI|PcG|LINE1	5	<1
CpGI|TFBS	12,931	49	PcG|ALU|TFBS	1	<1	CpGI|PcG|LINE1|TFBS	1	<1
CpGI|MIR	71	<1	PcG|LINE2	3	<1			

NOTE: A total of 26,377 autosomal loci were clustered. The shaded box indicates the cluster whose methylation is associated with survival.

Abbreviations: CpGI, CpG island; LINE, long interspersed nuclear element; MIR, mammalian interspersed repeat.

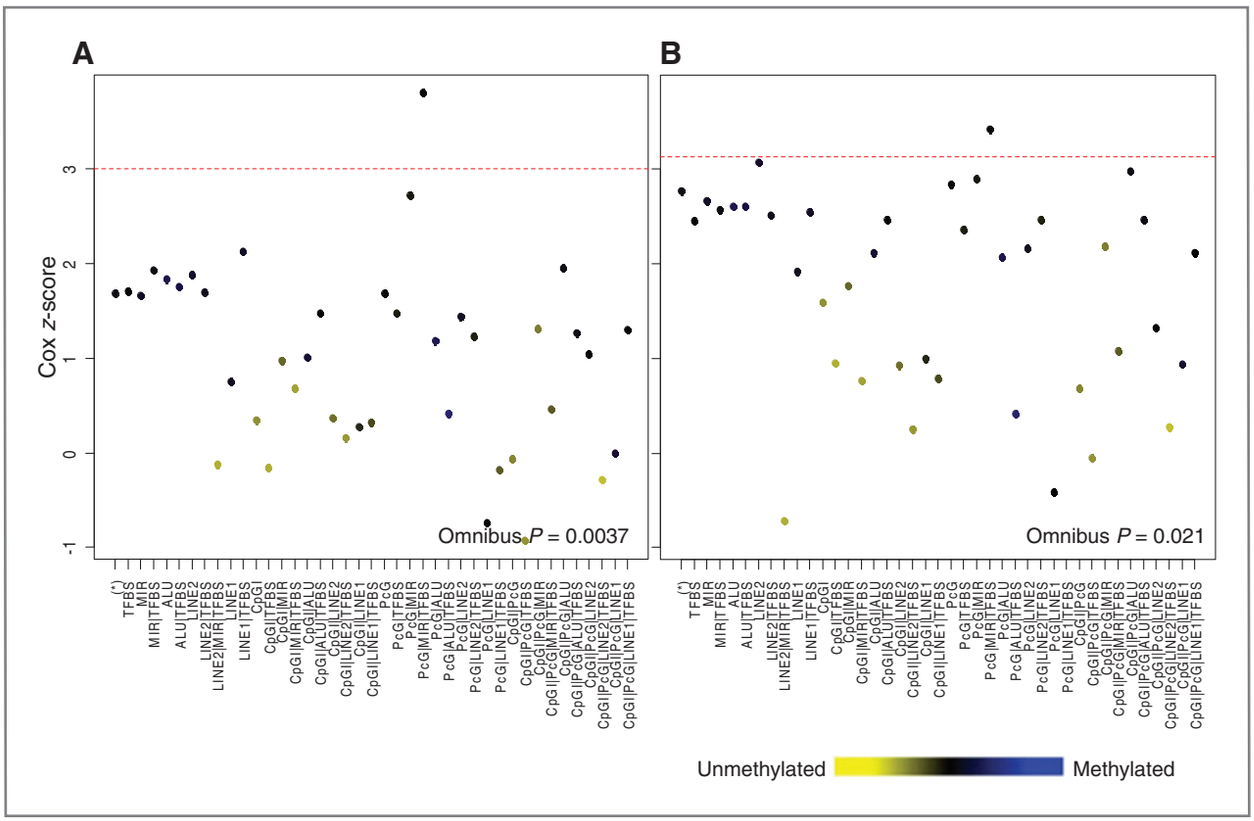

Figure 1. Cox survival scores for all bioinformatically derived clusters. Infinium array CpG loci were clustered into 41 groups on the basis of local sequence context. The cluster denoted by (*) contains 1,636 loci that are not proximal to any of the functional elements used for clustering. For each cluster, the average methylation across all member loci for all tumors was calculated (represented by point color). Tumors were stratified into high, intermediate, and low methylation depending on each tumor's methylation within each cluster, and the dichotomous high/low group status was used as the predictor in a Cox proportional hazards model. Cluster-specific Cox scores are plotted for the unadjusted model ($n = 90$ with complete survival data; A) and Cox model adjusted for age, site, stage, and HPV16 status ($n = 70$; B). Dashed red lines represent the upper boundary for the 95% confidence limit of the null permutation distribution. The omnibus P value indicates the overall significance level of the association between survival and methylation over all bioinformatic clusters of CpG loci. Associations where $P < 0.05$ are considered significant.

associated with patient survival (permutation test, $P = 0.0037$; Fig. 1A). This association was robust to potential confounding variables such as age at diagnosis, anatomic site, combined AJCC stage, and HPV16 status (permutation, $P = 0.021$; Fig. 1B). In the latter model, the power to detect significant differences in survival times, based on methylation of locus clusters, was diminished because of incomplete clinical data, however the effect remained significant.

To characterize the methylation state of the tumors in each patient subgroup, we compared the number of CpG loci that were differentially methylated with normal head and neck tissues (Supplementary Fig. S2) and found that the ratios of hypomethylated/hypermethylated CpGs were significantly different across patient subgroups (permutation, $P < 0.001$). The top differentially methylated sites within each group are described in Supplementary Table S2.

Methylation patterns of a 13-CpG locus cluster identify a patient subgroup with poor prognosis

Methylation of the CpG loci associated with the cluster: PcG target genes, mammalian interspersed repeats, and TFBS (PcG|MIR|TFBS; Fig. 1) was consistently associated with prognosis in these aggregate models. The aggregate methylation value for this group ($\beta = 0.52$) was significantly higher than the mean across all clusters (average = 0.43, range = 0.11–0.71). Many of the genes associated with the 13 CpG loci in this cluster (described in Table 2) are involved in the maintenance of cellular homeostasis or possess tumor-suppressive function. To assess possible bias or confounding arising from the characteristics of individual patients, we compared clinical and demographic factors of the patients in the low-, intermediate-, and high-methylation groups (Table 3). Notably, no significant differences in age at diagnosis, gender, HPV16 status, tumor site, tumor stage, drinking, or smoking exposures were observed.

We next asked whether the DNA methylation level within this cluster of 13 PcG|MIR|TFBS-member loci was associated with patient survival times. Kaplan–Meier analysis (Fig. 2) revealed a significant difference in overall survival among the patient methylation groups by taking into account all loci in this class (log-rank; $P = 0.0003$). The median survival time of patients in the high-methylation group was 1.2 years with 13%

Table 2. Gene-associated loci in bioinformatic cluster PcG|MIR|TFBS

Locus name	Chromosome	Genomic position	Entrez gene ID	Symbol	Protein name
cg01091565	15	88,095,766	55897	*MESP1*	Mesoderm posterior 1
cg01861509	10	73,519,632	9806	*SPOCK2*	Sparc/osteonectin; cwcv and kazal-like domains proteoglycan (testican) 2
cg03021690	14	102,661,717	7127	*TNFAIP2*	Tumor necrosis factor; alpha-induced protein 2
cg03389164	14	93,493,625	51676	*ASB2*	Ankyrin repeat and SOCS box-containing protein 2
cg03465320	6	32,931,056	6890	*TAP1*	Transporter 1; ATP-binding cassette; sub-family B
cg16853860	6	32,931,094	6890	*TAP1*	Transporter 1; ATP-binding cassette; sub-family B
cg03714916	6	36,753,864	1026	*CDKN1A*	Cyclin-dependent kinase inhibitor 1A
cg07617246	15	63,503,004	57722	*NOPE*	DDM36
cg09358725	11	33,870,664	4005	*LMO2*	LIM domain only 2
cg11258532	16	65,435,374	766	*CA7*	Carbonic anhydrase VII isoform 1
cg15796819	17	19,591,782	218	*ALDH3A1*	Aldehyde dehydrogenase 3 family; member A1
cg22340747	15	43,470,502	2628	*GATM*	Glycine amidinotransferase (L-arginine: glycine amidinotransferase)
cg24387380	15	24,741,904	2558	*GABRA5*	Gamma-aminobutyric acid (GABA) A receptor; alpha 5 precursor

of patients alive at 8 years, as compared with a 50% 8-year survivorship and a median survival time of 5.9 years in the low-methylation group. In addition, the patient group with intermediate-methylation showed intermediate survival times. A Cox proportional hazards analysis, adjusted for age at diagnosis, tumor site, tumor stage, and HPV16 status, showed that patients in the high DNA methylation group had significantly decreased survival times [HR, 3.98; 95% confidence interval (CI), 1.8–8.9; Table 4]. Because of potential differences in treatment regimens of low- versus high-stage disease that may influence survival times within and each patient methylation subgroup, we conducted a stratified subset analysis and found that survival times did not significantly vary by stage within any of the high-, intermediate-, or low-methylation subgroups (log-rank; $P = 0.43$, 0.33, 0.98, respectively).

Specificity and validation of survival association by bisulfite pyrosequencing

To determine whether specific CpGs in the PcG|MIR|TFBS cluster were driving the observed survival association, individual Cox proportional hazards models were constructed for each of the 13 PcG|MIR|TFBS CpG loci (Supplementary Table S3). Methylation levels of *TAP1* and *ALDH3A1* CpGs were most highly associated with survival time (HRs, 2.42 and 1.84; $P < 0.001$ and $P < 0.02$); we therefore selected these loci for confirmation by pyrosequencing all of the tumors of the discovery set with available DNA (88 of 91). Methylation values determined by pyrosequencing were highly concordant with array measurements at both loci (Spearman correlation = 0.88 and 0.86) and survival associations were validated (log-rank; $P = 0.003$ and 0.001; Supplementary Fig. S3). To determine the specificity of this survival association for CpG loci in the PcG|MIR|TFBS cluster, array methylation values for *ALDH3A1* and *LMO2* CpGs were compared with those of loci within the same genes but clustered into a separate PcG|TFBS group (*LMO2* is examined here as there were no additional *TAP1* CpGs with which to compare methylation in any other bioinformatic groups). Despite a close proximity of the *ALDH3A1* CpG pairs and *LMO2* CpGs pairs (~100 bp for each gene), methylation concordance was poor (Spearman correlation = 0.29 and 0.62), and proximal CpG methylation levels were not associated with survival in either case (log-rank; $P = 0.53$ and 0.97; Supplementary Fig. S4). Similarly, concordance of *TAP1* and *ALDH3A1* with neighboring gene loci (in *ULK2* and *PSMB9*, respectively) was low (Spearman correlation = 0.14 and −0.14), and methylation of these CpGs were not associated with survival (log-rank; $P = 0.17$ and 0.57; Supplementary Fig. S5).

Methylation of *TAP1* and *ALDH3A1* independently confirms survival association in an HNSCC validation population

To validate the observed association with survival, we further examined an additional 101 tumors drawn from an independent tumor collection period. Bisulfite pyrosequencing analysis of these tumors at the *TAP1* and *ALDH3A1* loci was conducted, and aggregate methylation values were generated for each tumor. Membership in the patient subgroup with the highest methylation was independently poorly prognostic (HR, 2.86; 95% CI, 1.02–8.11; Fig. 3 and Table 4), validating the initial discovery phase observation. In addition, patients who

Table 3. Clinicopathologic characteristics of study participants stratified according to methylation status across 13 CpG loci

	Low-methylation (N = 30)	Intermediate-methylation (N = 30)	High-methylation (N = 31)	*P*
Age at diagnosis, y				0.87[e]
Range	32–84	41–87	34–80	
Mean (SD)	58 (11.1)	63 (13.5)	59.5 (12.7)	
Gender, *n* (%)				0.22[f]
Female	6 (20)	11 (37)	6 (19)	
Male	24 (80)	19 (63)	25 (81)	
HPV16 status, *n* (%)				0.30[f]
Positive	8 (27)	11 (37)	6 (19)	
Negative	22 (73)	19 (63)	25 (81)	
Tumor site, *n* (%)[a]				0.45[f]
Oral	12 (46)	17 (65)	17 (71)	
Pharynx	7 (27)	5 (19)	3 (12)	
Larynx	7 (27)	4 (15)	4 (17)	
Tumor stage, *n* (%)[b]				0.71[f]
I	0 (0)	0 (0)	2 (7)	
II	8 (30)	5 (19)	8 (29)	
III	5 (19)	4 (15)	4 (14)	
IV	14 (52)	18 (67)	14 (50)	
Lifetime drink-years of consumption, *n*[c]				0.84[e]
Range	0–51	0–58	0–58	
Mean (SD)	27 (19)	22 (21.7)	29 (19)	
Never-drinkers, *n*	5	4	5	
Lifetime pack-years smoked, *n*[d]				0.32[e]
Range	0–105	0–100	0–125	
Mean (SD)	36 (29)	30 (28)	29 (31)	
Never-smokers, *n*	5	6	7	

[a]Fifteen samples missing site data.
[b]Nine tumors missing stage data.
[c]Eighteen patients missing self-reported drinking data.
[d]Thirteen patients missing self-reported smoking data.
[e]Kruskal–Wallis ANOVA.
[f]Permutation χ^2 test.

belonged to the intermediate-methylation subpopulation had decreased survival times compared with the low-methylation subgroup, although this association did not reach significance in multivariable analyses (HR, 2.82; 95% CI, 0.99–8.02; Fig. 3 and Table 4).

Discussion

Epigenetic alterations, reflected in changes to the 5-methylcytosine content at specific genomic loci, are required for normal cellular function and development. Candidate gene approaches for identification of prognostic biomarkers are widely used by researchers, and studies have found that DNA methylation alterations are associated with patient survival in a number of cancers (11, 28–36), including HNSCCs. Head and neck malignancies are often aggressive tumors with a poor probability of survival, and there are few tools currently available to assist the oncologist in determining long-term outcome in these patients. Here, we used a genome-wide array to measure DNA methylation in HNSCCs, identified markers of patient survival time with a novel approach by clustering CpGs on the basis of their association with local sequence features, and validated the identified markers in an independent set of tumors.

Recently, investigations have revealed that, in addition to aging and environment, the architecture of the genome itself may predispose certain CpG sites to DNA methylation, both in normal cells and in cancer (16, 37, 38). Therefore, methylation profiling using genomically informed methodologies is an attractive approach for identifying novel biomarkers. Here, we have developed a unique method for classifying CpG loci that allows functional sequence elements to dictate the clustering. This technique has proven useful in defining a novel

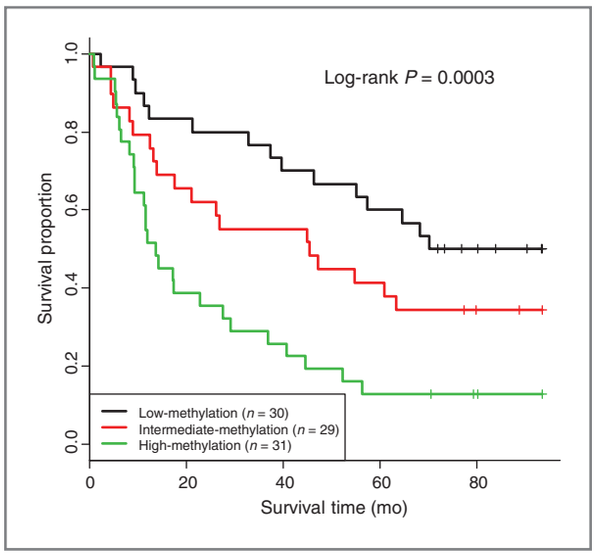

Figure 2. Kaplan–Meier analysis reveals an association between DNA methylation and overall patient survival. Patients with high-, intermediate-, and low-methylation levels across the 13 PcG|MIR|TFBS CpG locus cluster in a discovery population ($n = 90$) are shown. Vertical tick marks represent censored observations. Log-rank test P value is considered significant where $P < 0.05$.

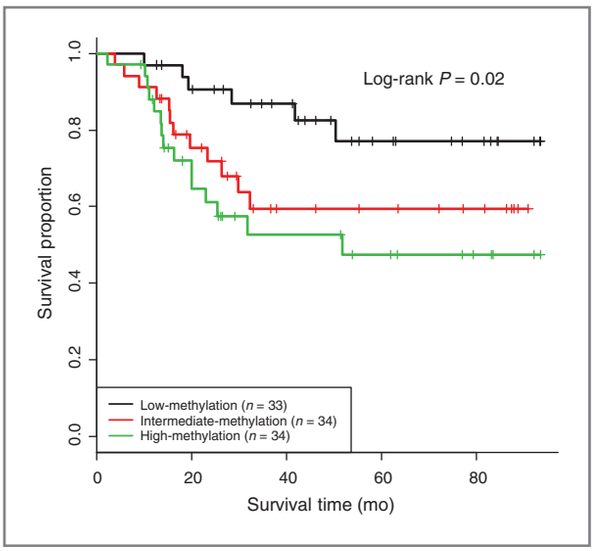

Figure 3. Combined methylation of *TAP1* and *ALDH3A1* loci is associated with overall survival in an HNSCC validation population. Aggregate methylation values for these top loci (of the 13 PcG|MIR|TFBS CpGs) were determined for each of the 101 tumors in the validation patient population and the methylation groupings were compared. Vertical tick marks represent censored observations. Log-rank test P value is considered significant where $P < 0.05$.

association between methylation of specific locus groups and patient prognosis.

Importantly, our analysis shows that patients with HNSCCs with tumors hypermethylated at a group of 13-CpG loci (defined by proximity to polycomb gene targets, mammalian interspersed repetitive elements, and TFBS) have a significantly reduced survival time independent of HPV16 infection status. This suggests that methylation alterations at these sites may determine the phenotype or therapeutic response of this disease. While more work is necessary to precisely determine how the nexus of these 3 functional sequence element types defines an observable phenotype, it is possible that methylation of these sites, particularly PcG target gene promoters and TFBSs, may potentiate the transformation into (or represent a

Table 4. Multivariable cox proportional hazards analysis in discovery and validation patient populations

	Discovery		Validation	
Predictors	**HR (95% CI)**	***P***	**HR (95% CI)**	***P***
Methylation group status				
Low-methylation	Reference		Reference	
Intermediate-methylation	**2.47 (1.08–5.65)**	**0.03**	2.82 (0.99–8.02)	0.05
High-methylation	**3.98 (1.77–8.93)**	**0.001**	**2.86 (1.02–8.11)**	**0.04**
HPV16 detection				
Negative	Reference		Reference	
Positive	**0.40 (0.17–0.89)**	**0.02**	0.47 (0.13–1.72)	0.25
Age (per decade)	1.28 (0.98–1.65)	0.07	1.08 (0.77–1.53)	0.64
Tumor site				
Oral cavity	Reference		Reference	
Pharynx	1.03 (0.46–2.32)	0.93	0.81 (0.36–1.87)	0.63
Larynx	1.85 (0.83–4.12)	0.13	1.51 (0.50–4.52)	0.46
Combined AJCC stage				
I and II	Reference		Reference	
III and IV	1.51 (0.76–2.96)	0.24	3.25 (0.75–14.1)	0.25

NOTE: Significant results in bold.

product of) a stem cell–like tumor. Indeed, there is accumulating evidence that DNA hypermethylation is observed in many cancers at the sites of polycomb-mediated gene repression in embryonic cells, which become relaxed during differentiation, and this methylation correlates with stem cell characteristics (39). In addition, a number of transcription factors have been shown to be involved in the recruitment of DNA methyltransferases to PcG target sites (39). Furthermore, studies of aging-dependent methylation have shown that PcG marking and frequency of coincident retrotransposable elements are both correlated and complementary (14). All of these observations further support our result that the combination of these sequence elements is critical in determining tumor behavior.

Focusing on the individual gene members of the PcG|MIR|TFBS cluster, we see that many of these are highly relevant to head and neck disease. Chief among them is *ALDH3A1*, which is expressed in the oral mucosa (40) and is a member of the aldehyde dehydrogenase family of enzymes that convert the carcinogenic intermediate of ethanol metabolism, acetaldehyde, into nontoxic acetic acid. As alcohol is a major risk factor for the development of HNSCCs, one might expect alterations of *ALDH3A1* to play a role in the initiation and promotion of malignancy. At the same time, this gene has been shown to inhibit epithelial cell proliferation (41) and variant *ALDH3A1* risk alleles are prevalent among breast cancer patients (42). In addition, it is well known that aldehyde dehydrogenase is involved in normal stem cell biology and is a functional marker for epithelial cells with enhanced tumorigenic potential. A recent publication showed enrichment for *ALDH3* in the stem cell populations of mammalian oral tissues (43). This supports our hypothesis that alterations at PCG|MIR|TFBS-associated genes define a more stem-like constitution of tumor cells that engender more aggressive HNSCCs.

Downregulation of another gene represented in our PcG|MIR|TFBS prognostic cluster, *TAP1*, allows HNSCC cells to avoid immune surveillance by cytotoxic T lymphocytes (44) and a lack of protein expression has recently been shown to confer a negative prognostic risk in HNSCCs (45) as well as in many other cancers. This study, however, did not include HPV status in the survival analysis, so the question remained whether downregulation was truly independently prognostic. Another group reported that ectopic expression of *TAP1* in xenograft assays significantly prolonged mouse survival time and increased immune infiltrate (46). In addition, functional investigations have revealed that this gene is downregulated in primary HNSCCs (47, 48), metastasis (45), and HNSCC cell lines (49), although promoter methylation was not previously described in this setting. Together, these data suggest that *TAP1* may be a candidate for therapy in human HNSCCs.

Another gene marked by a CpG in the PcG|MIR|TFBS cluster that was associated with HNSCC survival is the ankyrin-repeat SOCS box-containing protein 2 (*ASB2*), which appears to be a modulator of Notch signaling (50) and inhibits growth of leukemic cells (51). However, its potential status as a tumor suppressor in the head and neck has yet to be described. A prominent tumor suppressor in the PcG|MIR|TFBS cluster is *CDKN1A*, encoding the p21 (WAF1) protein that signals G_1 cell-cycle arrest or senescence. Two additional genes identified in our analysis (*SPOCK2* and *NOPE*) are known to be methylated as potential biomarkers in cancer (52, 53). Consistent with our observations, the genes represented by CpGs in the PcG|MIR|TFBS group are primarily tumor suppressors, and one would therefore expect that their inactivation through a combination of hypermethylation and additional somatic alterations would result in a poorer prognosis.

In summary, we have developed a novel technique to identify clinical characteristics of HNSCCs using the genomic information of CpG loci coupled with epigenetic content at those sites. In 2 independent populations, we show that DNA methylation profile markers may be used to identify those at greatest risk of death, irrespective of HPV status. The identification of specific DNA methylation biomarkers, such as those presented here, may assist in selecting patients who are most likely to benefit from tissue-sparing procedures and targeted therapies. It will be important to validate this in other populations in an effort to move these biomarkers into clinical practice.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interests were disclosed.

Authors' Contributions

Conception and design: G.M. Poage, B.C. Christensen, K.T. Kelsey
Development of methodology: G.M. Poage
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): G.M. Poage, M.D. McClean, K.T. Kelsey
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): G.M. Poage, E.A. Houseman, H.H. Nelson, B.C. Christensen, C.J. Marsit, K.T. Kelsey
Writing, review, and/or revision of the manuscript: G.M. Poage, E.A. Houseman, M.D. McClean, H.H. Nelson, B.C. Christensen, C.J. Marsit, K.T. Kelsey
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): G.M. Poage, R.A. Butler, M.D. McClean, K.T. Kelsey
Study supervision: M.D. McClean, K.T. Kelsey

Grant Support

This study was funded by the Flight Attendant Medical Research Institute (C.J. Marsit) and the NIH (grants CA078609, CA100679 to 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 December 21, 2011; revised March 26, 2012; accepted April 9, 2012; published OnlineFirst April 16, 2012.

References

1. Edwards BK, Ward E, Kohler BA, Eheman C, Zauber AG, Anderson RN, et al. Annual report to the nation on the status of cancer, 1975–2006, featuring colorectal cancer trends and impact of interventions (risk factors, screening, and treatment) to reduce future rates. Cancer 2010;116:544–73.

2. Adelstein DJ, Ridge JA, Gillison ML, Chaturvedi AK, D'Souza G, Gravitt PE, et al. Head and neck squamous cell cancer and the human papillomavirus: summary of a National Cancer Institute State of the Science Meeting, November. 9–10, 2008, Washington, D.C. Head Neck 2009;31:1393–422.

3. Avissar M, McClean MD, Kelsey KT, Marsit CJ. MicroRNA expression in head and neck cancer associates with alcohol consumption and survival. Carcinogenesis 2009;30:2059–63.
4. Vlatkovic N, El-Fert A, Devling T, Ray-Sinha A, Gore DM, Rubbi CP, et al. Loss of MTBP expression is associated with reduced survival in a biomarker-defined subset of patients with squamous cell carcinoma of the head and neck. Cancer 2011;13:2939–50.
5. Chang KP, Wu CC, Chen HC, Chen SJ, Peng PH, Tsang NM, et al. Identification of candidate nasopharyngeal carcinoma serum biomarkers by cancer cell secretome and tissue transcriptome analysis: potential usage of cystatin A for predicting nodal stage and poor prognosis. Proteomics 2010;10:2644–60.
6. Rosas SL, Koch W, da Costa Carvalho MG, Wu L, Califano J, Westra W, et al. Promoter hypermethylation patterns of p16, O6-methylguanine-DNA-methyltransferase, and death-associated protein kinase in tumors and saliva of head and neck cancer patients. Cancer Res 2001;61:939–42.
7. Ha PK, Califano JA. Promoter methylation and inactivation of tumour-suppressor genes in oral squamous-cell carcinoma. Lancet Oncol 2006;7:77–82.
8. Hasegawa M, Nelson HH, Peters E, Ringstrom E, Posner M, Kelsey KT. Patterns of gene promoter methylation in squamous cell cancer of the head and neck. Oncogene 2002;21:4231–6.
9. Bennett KL, Romigh T, Eng C. Disruption of transforming growth factor-beta signaling by five frequently methylated genes leads to head and neck squamous cell carcinoma pathogenesis. Cancer Res 2009;69:9301–5.
10. Langevin SM, Stone RA, Bunker CH, Lyons-Weiler MA, Laframboise WA, Kelly L, et al. MicroRNA-137 promoter methylation is associated with poorer overall survival in patients with squamous cell carcinoma of the head and neck. Cancer 2011;117:1454–62.
11. Misawa K, Ueda Y, Kanazawa T, Misawa Y, Jang I, Brenner JC, et al. Epigenetic inactivation of galanin receptor. 1 in head and neck cancer. Clin Cancer Res 2008;14:7604–13.
12. Estilo CL, O-Charoenrat P, Ngai I, Patel SG, Reddy PG, Dao S, et al. The role of novel oncogenes squamous cell carcinoma-related oncogene and phosphatidylinositol. 3-kinase p110alpha in squamous cell carcinoma of the oral tongue. Clin Cancer Res 2003;9:2300–6.
13. Calmon MF, Rodrigues RV, Kaneto CM, Moura RP, Silva SD, Mota LD, et al. Epigenetic silencing of CRABP2 and MX1 in head and neck tumors. Neoplasia 2009;11:1329–39.
14. Estecio MR, Gallegos J, Vallot C, Castoro RJ, Chung W, Maegawa S, et al. Genome architecture marked by retrotransposons modulates predisposition to DNA methylation in cancer. Genome Res 2010;20:1369–82.
15. Wolff EM, Chihara Y, Pan F, Weisenberger DJ, Siegmund KD, Sugano K, et al. Unique DNA methylation patterns distinguish noninvasive and invasive urothelial cancers and establish an epigenetic field defect in premalignant tissue. Cancer Res 2010;70:8169–78.
16. Poage GM, Houseman EA, Christensen BC, Butler RA, Avissar-Whiting M, McClean MD, et al. Global hypomethylation identifies loci targeted for hypermethylation in head and neck cancer. Clin Cancer Res 2011;17:3579–89.
17. Lienert F, Wirbelauer C, Som I, Dean A, Mohn F, Schubeler D. Identification of genetic elements that autonomously determine DNA methylation states. Nat Genet 2011;43:1091–7.
18. Marsit CJ, McClean MD, Furniss CS, Kelsey KT. Epigenetic inactivation of the SFRP genes is associated with drinking, smoking and HPV in head and neck squamous cell carcinoma. Int J Cancer 2006;119:1761–6.
19. Furniss CS, McClean MD, Smith JF, Bryan J, Nelson HH, Peters ES, et al. Human papillomavirus. 16 and head and neck squamous cell carcinoma. Int J Cancer 2007;120:2386–92.
20. Marsit CJ, Christensen BC, Houseman EA, Karagas MR, Wrensch MR, Yeh RF, et al. Epigenetic profiling reveals etiologically distinct patterns of DNA methylation in head and neck squamous cell carcinoma. Carcinogenesis 2009;30:416–22.
21. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium® assay. Epigenomics 2009;1:177–200.
22. Takai D, Jones PA. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. Proc Natl Acad Sci U S A 2002;99:3740–5.
23. Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. Genes Dev 2006;20:1123–36.
24. Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 2006;125:301–13.
25. Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, et al. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat Genet 2007;39:232–6.
26. Squazzo SL, O'Geen H, Komashko VM, Krig SR, Jin VX, Jang SW, et al. Suz12 binds to silenced regions of the genome in a cell-type-specific manner. Genome Res 2006;16:890–900.
27. Bollati V, Baccarelli A, Hou L, Bonzini M, Fustinoni S, Cavallo D, et al. Changes in DNA methylation patterns in subjects exposed to low-dose benzene. Cancer Res 2007;67:876–80.
28. Kong KL, Kwong DL, Fu L, Chan TH, Chen L, Liu H, et al. Characterization of a candidate tumor suppressor gene uroplakin 1A in esophageal squamous cell carcinoma. Cancer Res 2010;70:8832–41.
29. Hoque MO, Begum S, Brait M, Jeronimo C, Zahurak M, Ostrow KL, et al. Tissue inhibitor of metalloproteinases-3 promoter methylation is an independent prognostic factor for bladder cancer. J Urol 2008;179:743–7.
30. Wang J, Lee JJ, Wang L, Liu DD, Lu C, Fan YH, et al. Value of p16INK4a and RASSF1A promoter hypermethylation in prognosis of patients with resectable non-small cell lung cancer. Clin Cancer Res 2004;10:6119–25.
31. Marsit CJ, Liu M, Nelson HH, Posner M, Suzuki M, Kelsey KT. Inactivation of the Fanconi anemia/BRCA pathway in lung and oral cancers: implications for treatment and survival. Oncogene 2004;23:1000–4.
32. Ghosh A, Ghosh S, Maiti GP, Sabbir MG, Zabarovsky ER, Roy A, et al. Frequent alterations of the candidate genes hMLH1, ITGA9 and RBSP3 in early dysplastic lesions of head and neck: clinical and prognostic significance. Cancer Sci 2010;101:1511–20.
33. Marsit CJ, Posner MR, McClean MD, Kelsey KT. Hypermethylation of E-cadherin is an independent predictor of improved survival in head and neck squamous cell carcinoma. Cancer 2008;113:1566–71.
34. Huang KH, Huang SF, Chen IH, Liao CT, Wang HM, Hsieh LL. Methylation of RASSF1A, RASSF2A, and HIN-1 is associated with poor outcome after radiotherapy, but not surgery, in oral squamous cell carcinoma. Clin Cancer Res 2009;15:4174–80.
35. Long NK, Kato K, Yamashita T, Makita H, Toida M, Hatakeyama D, et al. Hypermethylation of the RECK gene predicts poor prognosis in oral squamous cell carcinomas. Oral Oncol 2008;44:1052–8.
36. Puri SK, Si L, Fan CY, Hanna E. Aberrant promoter hypermethylation of multiple genes in head and neck squamous cell carcinoma. Am J Otolaryngol 2005;26:12–7.
37. Langevin SM, Houseman EA, Christensen BC, Wiencke JK, Nelson HH, Karagas MR, et al. The influence of aging, environmental exposures and local sequence features on the variation of DNA methylation in blood. Epigenetics 2011;6:908–19.
38. Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, et al. Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell 2008;30:755–66.
39. Baylin SB. Stem cells, cancer, and epigenetics (October 31, 2009), StemBook, ed. The Stem Cell Research Community, StemBook, doi/10.3824/stembook.1.50.1, http://www.stembook.org.
40. Crabb DW, Matsumoto M, Chang D, You M. Overview of the role of alcohol dehydrogenase and aldehyde dehydrogenase and their variants in the genesis of alcohol-related pathology. Proc Nutr Soc 2004;63:49–63.
41. Pappa A, Brown D, Koutalos Y, DeGregori J, White C, Vasiliou V. Human aldehyde dehydrogenase 3A1 inhibits proliferation and promotes survival of human corneal epithelial cells. J Biol Chem 2005;280:27998–8006.

42. Afsar NA, Haenisch S, Mateen A, Usman A, Ufer M, Ahmed KZ, et al. Genotype frequencies of selected drug metabolizing enzymes and ABC drug transporters among breast cancer patients on FAC chemotherapy. Basic Clin Pharmacol Toxicol 2010;107:570–6.
43. Banh A, Xiao N, Cao H, Chen CH, Kuo P, Krakow T, et al. A novel aldehyde dehydrogenase-3 activator leads to adult salivary stem cell enrichment in vivo. Clin Cancer Res 2011;17:7265–72.
44. Leibowitz MS, Andrade Filho PA, Ferrone S, Ferris RL. Deficiency of activated STAT1 in head and neck cancer cells mediates TAP1-dependent escape from cytotoxic T lymphocytes. Cancer Immunol Immunother 2011;60:525–35.
45. Bandoh N, Ogino T, Katayama A, Takahara M, Katada A, Hayashi T, et al. HLA class I antigen and transporter associated with antigen processing downregulation in metastatic lesions of head and neck squamous cell carcinoma as a marker of poor prognosis. Oncol Rep 2010;23:933–9.
46. Lou Y, Vitalis TZ, Basha G, Cai B, Chen SS, Choi KB, et al. Restoration of the expression of transporters associated with antigen processing in lung carcinoma increases tumor-specific immune responses and survival. Cancer Res 2005;65:7926–33.
47. Meissner M, Reichert TE, Kunkel M, Gooding W, Whiteside TL, Ferrone S, et al. Defects in the human leukocyte antigen class I antigen processing machinery in head and neck squamous cell carcinoma: association with clinical outcome. Clin Cancer Res 2005;11:2552–60.
48. Ogino T, Shigyo H, Ishii H, Katayama A, Miyokawa N, Harabuchi Y, et al. HLA class I antigen down-regulation in primary laryngeal squamous cell carcinoma lesions as a poor prognostic marker. Cancer Res 2006;66:9281–9.
49. Lopez-Albaitero A, Nayak JV, Ogino T, Machandia A, Gooding W, DeLeo AB, et al. Role of antigen-processing machinery in the in vitro resistance of squamous cell carcinoma of the head and neck cells to recognition by CTL. J Immunol 2006;176:3402–9.
50. Nie L, Zhao Y, Wu W, Yang YZ, Wang HC, Sun XH. Notch-induced Asb2 expression promotes protein ubiquitination by forming non-canonical E3 ligase complexes. Cell Res 2011;21:754–69.
51. Guibal FC, Moog-Lutz C, Smolewski P, Di Gioia Y, Darzynkiewicz Z, Lutz PG, et al. ASB-2 inhibits growth and promotes commitment in myeloid leukemia cells. J Biol Chem 2002;277:218–24.
52. Chung W, Kwabi-Addo B, Ittmann M, Jelinek J, Shen L, Yu Y, et al. Identification of novel tumor markers in prostate, colon and breast cancer by unbiased methylation profiling. PLoS One 2008;3:e2079.
53. Taylor KH, Pena-Hernandez KE, Davis JW, Arthur GL, Duff DJ, Shi H, et al. Large-scale CpG methylation analysis identifies novel candidate genes and reveals methylation hotspots in acute lymphoblastic leukemia. Cancer Res 2007;67:2617–25.