Published in final edited form as:

Clin Cancer Res. 2011 April 1; 17(7): 2014–2023. doi:10.1158/1078-0432.CCR-10-3026.

The A/G Allele of Rs16906252 Predicts for *MGMT* Methylation and Is Selectively Silenced in Premalignant Lesions from Smokers and in Lung Adenocarcinomas

Shuguang Leng¹, Amanda M. Bernauer¹, Chibo Hong², Kieu C. Do¹, Christin M. Yingling¹, Kristina G. Flores³, Mathewos Tessema¹, Carmen S. Tellez¹, Randall P. Willink¹, Elizabeth A. Burki¹, Maria A. Picchi¹, Christine A. Stidley³, Michael D. Prados², Joseph F. Costello², Frank D. Gilliland⁴, Richard E. Crowell⁵, and Steven A. Belinsky^{1,*}

¹Lung Cancer Program, Lovelace Respiratory Research Institute, Albuquerque, New Mexico

²Brain Tumor Research Center, Department of Neurological Surgery, University of California-San Francisco, San Francisco, California

³Department of Internal Medicine, University of New Mexico, Albuquerque, New Mexico

⁴Norris Cancer Center and the Keck School of Medicine, University of Southern California, Los Angeles, California

⁵New Mexico VA Health Care System, Albuquerque, New Mexico

Abstract

Purpose—To address the association between sequence variants within the *MGMT* promoter-enhancer region and methylation of *MGMT* in premalignant lesions from smokers and lung adenocarcinomas, their biological effects on gene regulation, and targeting *MGMT* for therapy.

Experimental Design—SNPs identified through sequencing a 1.9kb fragment 5' of *MGMT* were examined in relation to *MGMT* methylation in 169 lung adenocarcinomas and 1731 sputum samples from smokers. The effect of promoter haplotypes on *MGMT* expression was tested using a luciferase reporter assay and cDNA expression analysis along with allele-specific sequencing for methylation. The response of *MGMT* methylated lung cancer cell lines to the alkylating agent temozolomide was assessed.

Results—The A allele of rs16906252 and the haplotype containing this SNP were strongly associated with increased risk for *MGMT* methylation in adenocarcinomas (ORs ≥ 94). This association was observed to a lesser extent in sputum samples in both smoker cohorts. The A allele was selectively methylated in primary lung tumors and cell lines heterozygous for rs16906252. With the most common haplotype as the reference, a 20–41% reduction in promoter activity was seen for the haplotype carrying the A allele that correlated with lower *MGMT* expression. The sensitivity of lung cancer cell lines to temozolamide was strongly correlated with levels of *MGMT* methylation and expression.

*To whom all correspondence should be addressed. Lung Cancer Program, Lovelace Respiratory Research Institute, Albuquerque, New Mexico 87108. Phone: 505-348-9465; Fax: 505-348-4990; sbelinsk@LRRRI.org.

Conflicts of Interest Statement: S.A. Belinsky is a consultant to Oncomethylome Sciences. Under a licensing agreement between Lovelace Respiratory Research Institute and Oncomethylome Sciences, nested MSP was licensed to Oncomethylome Sciences, and the author is entitled to a share of the royalties received by the Institute from sales of the licensed technology. The Institute, in accordance with its conflict-of-interest policies, is managing the terms of these arrangements.

Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

Conclusions—These studies provide strong evidence that the A allele of a *MGMT* promoter-enhancer SNP is a key determinant for *MGMT* methylation in lung carcinogenesis. Moreover, temozolamide treatment may benefit a subset of lung cancer patients methylated for *MGMT*.

Keywords

MGMT; allele specific methylation; single nucleotide polymorphism; sputum; lung cancer

Introduction

The fact that detection of gene promoter methylation of tumor suppressor genes in sputum that contains exfoliated cells from premalignant lesions from smokers' lungs is a promising marker for early stage lung cancer makes identifying factors that influence the propensity for this epigenetic process throughout the respiratory epithelium a high priority. Such knowledge could impact strategies for screening to improve risk models, and identify persons who would benefit most from chemoprevention. The genes and the pathways they modulate underlying the inter-individual susceptibility to DNA methylation remain largely elusive. Accumulating evidence from our group and others implicate DNA damage as an important step in the acquisition of *de novo* methylation (1–4). This premise was supported *in vivo* through a community-based study in which we showed a significant association between DNA repair capacity and germ line sequence variants within DNA repair genes and gene promoter methylation in sputum from smokers (5). The impact of this finding for prevention was addressed by asking whether dietary nutrients and vitamin supplements that may influence DNA damage and repair would modify the extent of gene methylation in the aerodigestive tract. Significant protection against methylation was observed in subjects consuming a diet rich in leafy green vegetables, or folate, or taking a multi-vitamin supporting the concept that diet and supplements can affect re-programming of the epigenome and thus, have the potential to modify risk for lung cancer (6).

Specific sequence patterns within gene promoters and embryonic targets of polycomb-repressive complex 2 are predictive for identifying specific genes methylated in many cancers, but cannot discriminate individual susceptibility for gene silencing (7–10). However, other local sequence features that are polymorphic among people may contribute to the susceptibility of a specific CpG island to becoming methylated (11). A genome-wide survey in normal tissues found differential DNA methylation and expression between alleles at multiple non-imprinted loci that was sequence dependent (12). An additional study assessing 16 CpG rich gene promoters of chromosome 21 in leukocyte DNA identified four genes with allele-specific methylation (ASM) associated with single nucleotide polymorphisms (SNPs) (13). Therefore, sequence dependent ASM may represent an important mechanism connecting genetic polymorphisms and phenotypic variability to impact cancer susceptibility.

O⁶-methylguanine-DNA methyltransferase (*MGMT*) is a DNA repair enzyme that protects cells from the carcinogenic effects of alkylating agents by removing adducts from the O⁶ position of guanine (14). The protective effect of *MGMT* activity against the cytotoxic effects of alkylating agents has been demonstrated in human cell lines, xenograft models, and *MGMT* transgenic mice (14). The predominant mechanism for *MGMT* gene inactivation is methylation of a CpG island within its promoter-enhancer region that leads to transcriptional silencing in brain, colorectal, lung, lymphoma, and head and neck cancers (15). Silencing of this gene in cancer is associated with an increased prevalence for transition mutations in the *K-ras* and *p53* genes, consistent with the important role for this gene in removing alkyl DNA adducts (16,17). Methylation of *MGMT* is also a promising biomarker for lung cancer detection and its silencing is prognostic for response of

glioblastoma patients to the alkylating agent temozolamide (TMZ) (18,19). Transcriptional regulation of *MGMT* occurs largely in a 1.2 kb fragment 5' of the gene which contains the first untranslated exon (20). A 59 bp enhancer required for efficient *MGMT* promoter function was identified between the exon 1 and intron 1 boundary along with a SNP (rs16906252) located at the boundary between the enhancer and exon 1 (20,21). Colorectal cancer patients heterozygous or homozygous for the A allele (versus G allele) of this SNP had a 32-fold increased odds for methylation of the *MGMT* gene in their tumors, a finding replicated in a second, independent study (21,22).

The purpose of this study was to address the association between sequence variants within the *MGMT* promoter-enhancer region and methylation of *MGMT* in lung adenocarcinomas and exfoliated epithelial cells within sputum from smokers, and their biological effects on gene regulation. Genetic and epigenetic variations that affect *MGMT* expression may impact personalized therapy for lung cancer. Therefore, the potential of TMZ as a therapeutic option for lung cancer was evaluated in lung cancer cell lines with different levels of methylation and expression of *MGMT*.

Materials and Methods

Study populations and sample collection

The Lovelace Smokers Cohort (LSC) began recruitment of female smokers in 2001 and expanded to include male smokers in 2004 (5). Enrollment is restricted to current and former smokers age 40 to 74 y with a minimum of 15 pack-years of smoking. At study entry all cohort members completed a detailed questionnaire that collected information on medical, smoking, exposure history, socioeconomic status, and quality of life. Pulmonary function was tested to assess lung function, blood was drawn for collection of plasma and lymphocytes, and sputum was collected by induction. The Veterans Smokers Cohort (VSC) began recruitment of smokers in 2000. Similar enrollment criteria and procedures as used in the LSC were followed. Most participants in the VSC are males and have over 20 pack-years smoking history. In total, 1731 lung cancer-free subjects (1256 subjects from LSC and 475 subjects from VSC) were included in this study.

Lung adenocarcinomas (n=169) were obtained from the tumor banks at Johns Hopkins and the Mayo Clinic (23). Demographic characteristics of the patients have been described (23). Stage I lung cancer patients (n=98) enrolled in the New Mexico lung cancer cohort were also used in this study because of the availability of frozen tissue for assessing *MGMT* expression. A detailed description of the enrollment and characteristics of the cohort members has been provided previously (24). All samples were collected with informed consent and the institutional review boards of the Lovelace Respiratory Research Institute (Western IRB) and New Mexico VA Health Care System approved all investigations using human tissues and clinical data.

Cell lines

Normal human bronchial epithelial cells (NHBE) isolated from bronchoscopy of cancer-free smokers (n = 20), nineteen lung cancer-derived cell lines (A549, Calu-3, Calu-6, HCC4006, HCC827-1, H23, H358, H460, H522, H1299, H1435, H1568, H1795, H1993, H2009, H2023, H2085, H441, and H2228), one immortalized normal human bronchial epithelial cell line (HBEC2), and two glioblastoma cell lines (U251 and T98G) were used in this study. HBEC2 was received from Drs. Shay and Minna, Southwestern Medical Center, Dallas, TX. All the other cell lines were from American Type Culture Collection (Rockville, MD, USA). Cells were maintained in recommended medias in a 37°C incubator supplied

with 5% CO₂. All experiments were conducted using cell lines which were passaged in the lab for fewer than six months after resuscitation.

Sputum processing and MGMT promoter methylation

A detailed procedure for sputum processing has been introduced (18). Briefly, sputum samples were stored in Saccomanno's fixative. Sputum adequacy defined as the presence of deep lung macrophages or Curschmann's spiral (18) was assessed by cytotechnologists and subjects with cancer cells detected in sputum slides were removed from this study. An aliquot from sputum samples for each study subject was used for DNA isolation. Promoter methylation of *MGMT* was studied in NHBECs and cell lines using Combined Bisulfite Modification and Restriction Analysis (COBRA) and methylation-specific PCR (MSP) as described (25). Promoter methylation of *MGMT* in primary adenocarcinoma and sputum samples was evaluated using MSP and nested MSP, respectively (25). The COBRA and MSP assays are designed to interrogate the methylation status of CpGs around the transcriptional start site of *MGMT*. COBRA and MSP primer sequences are listed in Table S1, respectively.

SNP discovery and haplotype-tagging SNPs (htSNPs) selection

A 1,871 bp fragment of the *MGMT* gene that contains the well-characterized maximal promoter, the entire CpG island, and part of intron 1 was carefully selected for SNP discovery because several major regulatory elements for *MGMT* transcription were identified to locate in this region (Supplemental Figure) (20,26). Primers are listed in Table S1. DNA samples from 49 cancer-free non-Hispanic Whites randomly selected from the LSC were used for discovery of SNPs. An additional 11 samples were selected for sequencing based on their rs16906252 genotypes (3 A/Gs and 8 A/As) to improve the probability of finding a SNP in perfect linkage disequilibrium (LD) with rs16906252. PCR products were gel-purified and were commercially sequenced to identify variants (Sequetech, Mountain View, CA). Pairwise r^2 between any two SNPs was calculated to identify a SNP in perfect LD with rs16906252 ($r^2=1$). A Bayesian statistical method implemented in the program PHASE (Version 2.1) was used to construct the haplotypes using the variants discovered in the 60 subjects (27,28). HtSNPs were chosen using the TagSNPs program (29) based on the haplotype alleles and frequencies generated by the PHASE program in 49 randomly selected samples. In this process, the squared correlation (R_h^2) between the true haplotypes (h) and their estimates was calculated. An expectation maximization algorithm approach was then implemented to find the minimum set of SNPs (within the promoter-enhancer region of *MGMT*) that would have $R_h^2 > 0.80$ for all common haplotypes.

SNP genotyping

Genotyping was conducted in DNA samples isolated from peripheral lymphocytes in subjects from LSC and VSC. DNA isolated from tumor tissues was used for genotyping in 169 lung adenocarcinomas because no blood was collected from these patients. It is highly unlikely that DNA recovered from tumor tissue will generate false calls for germ line SNPs because stromal and inflammatory cells also comprise the tumor specimen. Primers amenable to detecting a PCR-restriction fragment length polymorphism (RFLP) were designed to genotype the SNPs in DNA samples isolated from the lung adenocarcinomas (n=169), subjects (n=475) from the VSC, NHBECs (n=20), and cell lines (n=22). The primers and restriction digestion enzymes are listed in Table S1. The 60 DNA samples used for SNP discovery were regenotyped by PCR-RFLP and 100% concordance of genotypes was observed between this method and sequencing. Two DNA samples with wild homozygote and variant homozygote were included in each batch of PCR-RFLP assay for quality control. The Illumina Golden Gate Assay was used to genotype rs16906252 in

subjects from the LSC (n=1256). Intra- and inter-plate duplicates and one parent-child trio were included in each 96-well plate for quality control.

Selection of subjects and construction of MGMT promoter constructs

A 1,528-bp fragment covering the entire maximal promoter and CpG island of *MGMT* with different haplotypes was amplified by PCR from homozygotes, or heterozygotes when a homozygote for a haplotype was not present in the 60 subjects used in discovery of the SNPs (Supplemental Figure). The primers are listed in Table S1. The promoter-enhancer fragment was directionally cloned into the pGL2-basic Luciferase Reporter Vector (Promega, Madison, WI) upstream of the luciferase coding sequence. Four to six clones from each person were commercially sequenced to identify the correct construct with each specific haplotype (Sequetech, Mountain View, CA).

Transient transfection and reporter gene assays

Transient transfections were conducted in Calu-6, H1299, and HBEC2. Cells ($0.8\text{--}1.2 \times 10^6$) were plated into 6-well plates and transfected the following day. Plasmid DNA (1 μg) and the pSV- β -Galactosidase control vector (0.2 μg ; Promega, Madison, WI) were co-transfected into cells with Lipofectamine 2000 transfection reagent (Invitrogen, Carlsbad, CA) at a Lipofectamine 2000: DNA ratio of 3:1. A promoter-less pGL2-basic vector and the pGL2-control vector that contains the SV40 promoter were used as a negative and positive control, respectively. Forty-eight hours after transfection, cells were harvested and lysed. Immediately after lysing, cell extracts were assayed in the Lumineskan Ascent luminometer (Thermo Electron, Milford, MA) for luciferase activity using the Luciferase Assay System (Promega, Madison, WI). β -galactosidase activity in cell lysates was measured using the Galacto-Star Reporter Gene Assay System (Tropix, Bedford, MA). Promoter activity was calculated as the ratio of activities of luciferase and β -galactosidase. Transfections were done in triplicate in three independent experiments.

Cytotoxicity of TMZ in tumor cell lines

The stock TMZ (Sigma-Aldrich, St. Louis, MO) solution (129 mM) was prepared by dissolving 100 mg TMZ in 4 ml DMSO and was frozen at -80°C in aliquots. Before treatment, TMZ stock was thawed and added to the culture medium at a series of final concentrations (0, 10, 30, 90, 270, 810, 2500 μM). Cells (5×10^3 /well) were seeded into 96-well plates and were incubated with TMZ in the culture medium for 72 hrs. The MTT assay measured cytotoxicity and an IC50 was calculated. Three independent experiments were conducted for each cell line and the average IC50 was presented.

Real time PCR for measuring MGMT expression

Cells were harvested in TRI reagent (Sigma-Aldrich, St. Louis, MO) and mRNA was isolated following TRI reagent instructions. cDNA was synthesized using the high capacity cDNA reverse transcription kit (Applied Biosystems, Foster City, CA). A real time PCR using SYBR® Green as the fluorescent dye was developed to measure the MGMT expression level using the ΔCT method. The primers are listed in Table S1.

Statistical analysis

The two-sample t-test, Wilcoxon rank sum test, and χ^2 test were used to compare demographic variables and genotypes between *MGMT* methylated and unmethylated groups as appropriate. Hardy-Weinberg equilibrium (HWE) was tested for each SNP in the LSC, VSC, and the lung adenocarcinomas separately and no deviation from HWE was identified (not shown). The minor allele frequency (MAF) for rs16906252 is 7.6% in the 1900 samples. This study is well powered to investigate an association between a SNP with MAF

of 0.076 and risk for *MGMT* methylation in three different cohorts. Under the additive inheritance model that assumes that the effect of the A/A genotype is twice that of the G/A genotype, we have 99% power to detect an OR of 2.3, 1.9, and 1.6 in lung adenocarcinoma samples, VA cohort and Lovelace cohort, respectively (30). A logistic regression model was used to calculate the ORs and 95% CIs for each individual SNP with adjustment for age, sex, ethnicity, and smoking history selected *a priori*.

The PHASE program was used to reconstruct the haplotypes and calculate their estimated probabilities from the htSNPs data in the 5' region of *MGMT* in lung adenocarcinomas (27,28). All possible haplotypes with corresponding estimated probabilities (> 0.01) for each individual were generated. The probabilities of the common haplotypes for each individual were used as explanatory variables in the logistic regression model with adjustment for non-genetic factors to assess the association between the haplotypes and the risk for *MGMT* methylation. Haplotypes with frequency $< 5\%$ were combined into one group. The effect for each common haplotype was estimated using the most common haplotype as the reference group. The effects for the common haplotypes were also analyzed using all other haplotypes as the reference group and similar results were obtained. All statistical analyses were conducted in Statistical Analysis System 9.2.

Results

Association of SNPs to *MGMT* Methylation

Several SNPs were discovered in the promoter-enhancer region of *MGMT* in a European population; however, the LD between those SNPs has not been characterized (31). Sequencing of a 1.9 kb fragment of *MGMT* identified 9 SNPs with one SNP detected for the first time (Table S2). No two SNPs were in perfect LD (Figure 1). The r^2 value between any two SNPs in this area was ≤ 0.32 except for rs1625649 and rs12356501 ($r^2=0.71$). Four SNPs had MAF > 0.04 . A total of four haplotypes with frequency > 0.04 were constructed based on the four common SNPs. Three SNPs including rs1711646, rs1625649, and rs16906252 were selected as the htSNPs (29) that accounted for 87% of the haplotype variation in the promoter-enhancer region of *MGMT*.

The association between these three htSNPs and *MGMT* methylation was first tested in 169 lung adenocarcinomas. Rs1625649 was modestly associated with *MGMT* methylation (Table 1). In contrast, a striking and highly significant association was seen between rs16906252 and increased risk for *MGMT* methylation (OR=97.0, $P<0.0001$). Only rs16906252 remained significant after Bonferroni correction. The association between the genetic variation of the entire promoter-enhancer region and *MGMT* methylation was then analyzed using a haplotype-based approach that integrated the genetic information of multiple SNPs. A total of 7 haplotypes were reconstructed based on the genotype for the 3 htSNPs including rs1711646, rs1625649, and rs16906252 in the 169 lung adenocarcinomas (27,28). Four of the 7 haplotypes including HAP1 (GCG), HAP3 (GAG), HAP4 (GAA), and HAP7 (TCG) had frequency $> 5\%$ with their cumulative frequency of 94% (Table S3). Haplotype analysis found that only HAP4 defined solely by rs16906252 was associated with *MGMT* methylation (Table S3). *MGMT* is one of a panel of genes whose concomitant methylation detected in sputum collected from moderate and heavy smokers associates with increased risk for lung cancer in a high-risk cohort (18,32). Therefore, the association between rs16906252 and *MGMT* methylation was assessed in sputum samples collected in two smoker cohorts. The variant homozygotes and heterozygotes were associated with 12.4- and 6.3-fold increased risk for *MGMT* methylation, respectively in the LSC (Table 1). The strong association between rs16906252 and *MGMT* methylation was also replicated in smokers from the VSC (Table 1).

The commonality for promoter sequence variants to affect other frequently methylated genes was addressed by studying the *p16* gene that is silenced by promoter methylation in > 50% of lung tumors (25). Only one common (MAF > 0.05) SNP (rs3814960) was identified within the *p16* basal promoter and CpG island by searching the Environmental Genome Project database (<http://egp.gs.washington.edu/>). No significant association ($P > 0.10$) was found between rs3814960 and *p16* methylation in adenocarcinomas or sputum samples from the LSC indicating that not all genes are silenced by sequence dependent ASM.

Influence of the A/G Genotype on MGMT Transcription, Expression, and Methylation

The effect of sequence variants in the promoter-enhancer region on transcription of *MGMT* was assessed in HBEC2, Calu-6, and H1299. The lowest promoter activity was seen in HAP4 carrying the A allele. With the most common haplotype (HAP1) as the reference, a 20 – 41% reduction in promoter activity associated with HAP4 was seen (Figure 2A). In addition, expression of the A allele was reduced 31–39% compared with the G allele in two NHBECs from cancer-free smokers heterozygous for rs16906252 and unmethylated for *MGMT* (Figure 2B). In contrast, a previous study focused on a small fragment of the *MGMT* promoter (between SmaI and SstI in Supplemental Figure) and showed a 16 – 64% increase of *MGMT* promoter-enhancer activity associated with A allele (33). However that small promoter construct generated only 13% promoter activity compared with that seen with the 1.2 kb *MGMT* promoter tested by Harris et al (20) and replicated in our studies using a slightly larger construct. Furthermore, our findings that are based on the haplotype of the entire maximal promoter should be more reflective of the *in vivo* situation by allowing the effect of a SNP along with the complex interaction of the multiple transcription factors binding to the maximal *MGMT* promoter and the LD of multiple SNPs in this area to influence promoter activity (34).

Reduced gene transcription has been proposed as one mechanism predisposing a gene for epigenetic silencing via promoter methylation (35). Therefore, studies were conducted to test whether cell lines with the A/G genotype have preferential methylation of the A allele that in turn should lead to monoallelic expression of *MGMT*. Calu-6 and H1993 were heterozygous for rs16906252 and hemi-methylated for *MGMT*. Bisulfite sequencing found that only the A allele is densely methylated and selectively silenced in Calu-6 and H1993 (Figure 2B, Figure 3A; not shown). The ASM of *MGMT* associated with rs16906252 was further confirmed in primary lung tumors. An MSP assay targeting the reverse strand of the *MGMT* promoter that distinguishes the G versus the A allele was developed. Six A/G tumors methylated for *MGMT* were sequenced and only the A allele was methylated (Figure 3B).

Targeting MGMT for Lung Cancer Therapy

TMZ is an alkylating chemotherapeutic agent that does not cause severe systemic toxicity and significantly extends survival in glioblastoma patients with methylation of *MGMT* (19). If a similar response to TMZ dependent on *MGMT* methylation status could be replicated in lung cancer, this agent could potentially benefit a subset of patients. This premise was tested by determining if lung tumor cell lines with differential *MGMT* methylation and expression would have different sensitivity toward TMZ treatment. Expression of *MGMT* was quantified in nine G/G lung cancer cell lines with unmethylated *MGMT* and compared to the average expression seen in six G/G NHBEC lines. Two unmethylated cell lines with high *MGMT* expression were selected to serve as controls (Table 2). All three of the G/G cancer cell lines with *MGMT* methylation showed hemi-methylation based on the COBRA assay (not shown). Two G/G methylated cell lines and one A/G cell line were selected and expression levels of *MGMT* were quantified. Two cell lines derived from glioblastoma, one G/G fully methylated for *MGMT* showing no expression (U251) and the second (T98G) G/G was sparsely methylated showing no reduction in expression compared to NHBECs were

included for comparison of sensitivity to TMZ. The IC50 for TMZ was similar in the unmethylated glioblastoma and lung cancer cell lines (1590 – 2500 μ M). As expected the fully methylated glioblastoma cell line showed heightened sensitivity to TMZ with an 83% reduction in IC50 compared to unmethylated cell lines (Table 2). The response of the hemi-methylated lung cancer cell lines was intermediate between fully methylated and unmethylated tumor cell lines. A high correlation between TMZ IC50s and MGMT expression was observed for the lung cancer cell lines ($r=0.99$) (Table 2 and Figure 4).

Discussion

This study provides strong evidence that the A allele of a *MGMT* promoter-enhancer SNP (rs16906252) is a key determinant in the acquisition of *MGMT* methylation in lung carcinogenesis. This association was found in primary adenocarcinoma and in exfoliated cells in sputum from cancer-free smokers, a finding that reflects the ongoing field cancerization within the aerodigestive tract. The magnitude of the association between rs16906252 and *MGMT* methylation in premalignant lesions from lung cancer-free smokers was much smaller than observed from the lung adenocarcinoma patients consistent with the increase in prevalence of *MGMT* methylation of 25% in premalignant lesions from smokers to 43% in stage I lung cancer and 61% in stage II-IV lung cancer (36). The reduced transcription and expression seen for the A allele of *MGMT* may be one predisposing factor for methylation. This hypothesis is supported by the selective methylation and accompanied silencing of the A allele in tumors from A/G heterozygotes. Thus, some of the *MGMT* unmethylated lung cancer-free smokers who carry the A allele should be more prone to become methylated as their field cancerization progresses over time (32).

The lack of an association between rs16906252 with risk for lung cancer based on the similar percentage of subjects carrying the A allele in lung adenocarcinoma cases and cancer-free controls in the Lovelace and Veteran Cohorts (14.2%, 15.6%, and 12.8%, respectively) is not surprising. Our previous studies demonstrate that concomitant methylation of a panel of tumor suppressor genes including *MGMT* in sputum could prospectively predict risk for lung cancer in moderate and heavy smokers (18,32). *MGMT* methylation alone in sputum only contributes to a portion of the predicted increased cancer risk. Thus, a large case-control study would be needed to more precisely estimate the association of this SNP to risk for lung cancer. Moreover, approximately 20% of subjects with G/G genotype also have *MGMT* methylated in sputum, suggesting that other genetic or environmental factors may also contribute to *MGMT* methylation in lung carcinogenesis, albeit probably to a lesser degree than rs16906252. Thus, the prediction for lung cancer risk by *MGMT* methylation in sputum could also vary by rs16906252 genotype. Testing this hypothesis will require long-term follow-up in our smoker cohorts to accumulate incident lung cancer cases with sputum samples collected before cancer diagnosis.

The sequence dependent ASM of *MGMT* is not lung tumor-specific because similar associations were also reported in colorectal cancer (21,22). Studies in glioblastoma did not identify an association between rs16906252 and risk for methylation (37). However, the primers used for genotyping had several incorrect bases that likely influenced the accuracy for identifying the variant allele (37). Recently, methylation of *GSTP1*, *RIL*, and *MSH2* were found to be allele-specific in several tumor types (11,38,39). Low-level methylation of *MGMT* was also seen in normal colon mucosa and peripheral lymphocytes from heterozygote subjects, corroborating a likely important role for this SNP in the susceptibility of the A allele of *MGMT* for silencing (22,40). Zhang *et al.* studying methylation of gene promoters on chromosome 21 estimated that ASM may affect 10% of human genes in leukocytes (13). Thus, genetic differences can influence inter-individual variation for DNA methylation of many genes in normal tissues and tumors.

Evidence is accumulating that reduced transcription may be one trigger for initiation and spreading of heterochromatin and methylation along a gene promoter (34). Song *et al.* and Stirzaker *et al.* (41,42) showed that reduced transcription and seeds of methylation within the *GSTP1* promoter served as a stimulus for the spread of methylation and chromatin remodeling. *In vivo* studies found that breast cancer patients carrying the *GSTP1* promoter haplotype E comprised of four SNPs and the shortest ATAAA repeat had increased methylation (38). The polymorphisms within haplotype E disrupted binding of c-Myb leading to reduced transcription. Our studies also support reduced transcription as a mechanism for *de novo* promoter methylation based on the strong association between reduced promoter activity associated with the A allele in NHBEC and the targeting of this allele for silencing in lung tumors from heterozygote patients. Other mechanisms independent of effects on gene transcription may also influence *de novo* promoter methylation. For example, a 12 bp insertion polymorphism in the *RIL* promoter creates an Sp1/Sp3 binding site that protects against methylation in cancer and does not change expression in lymphoblastoid cell lines (11). The Sp1/Sp3 binding site created by this polymorphism may through protein-DNA interactions prevent spreading of DNA methylation and heterochromatin from neighboring “methylation centers”. *In silico* prediction of transcription factor-binding sites using TFSEARCH and MATCH programs did not identify any transcription factors with core motifs binding rs16906252. In addition, resequencing the 1.9 Kb 5' region of the *MGMT* gene did not identify any SNP in perfect LD with rs16906252. Therefore, rs16906252 may influence the binding of an unknown transcription factor or unlisted transcription factor in these two databases. Recently, Hawkins *et al.* examined six gene-associated CpG islands flanking *MGMT* in colorectal tumors and found that the *EBF3* gene located 500 kb downstream of *MGMT* had a similar methylation pattern to *MGMT* (22). Thus, it is possible that rs16906252 is in perfect LD with other polymorphisms present in the euchromatin and heterochromatin boundaries surrounding the *MGMT* and/or *EBF3* which could affect gene regulation. Currently, testing this hypothesis will necessitate the completion of the 1000 Genome Project that will provide a dense map of genetic variation by sequencing the genomes of 1200 people (43).

The refractory nature of unresected lung cancer to chemotherapeutic combinations tested over three decades necessitates developing targeted therapies based on underlying genetic and epigenetic tumor profiles. Responses have been extremely positive for patients with mutations in the epidermal growth factor receptor receiving the tyrosine kinase inhibitors gefitinib or erlotinib (44). Alkylating agents such as 1,3-bis(2-chloroethyl)-1-nitrosourea were tested as chemotherapeutics for lung cancer decades ago with response rates limited by systemic toxicity (35,45,46). TMZ has been evaluated in a phase II study for efficacy in pretreated lung cancer, melanoma, and breast cancer patients with brain metastasis. Treatment was provided without knowledge of *MGMT* methylation status, and 26% of patients had a clinical response with median overall survival greatest for lung cancer patients (47). Our *in vitro* studies in lung cancer cell lines demonstrating sensitivity to TMZ that is highly correlated with levels of *MGMT* methylation suggest that a subset of lung cancer patients with relatively low *MGMT* expression could respond to this drug, a hypothesis that requires further preclinical studies prior to initiating clinical trials.

Translational relevance

Epigenetic silencing of *MGMT* by promoter methylation is a biomarker for identifying persons at high risk for lung cancer and for predicting therapeutic response to alkylating agents in glioblastoma. The A/G allele of rs16906252 in *MGMT* promoter-enhancer region predicts for gene methylation in sputum from cancer free smokers. This association was increased substantially in primary lung tumors supporting this variant as a key determinant for epigenetic silencing of *MGMT*. Importantly, the sensitivity of lung

cancer cell lines to temozolamide (TMZ) was strongly correlated with levels of *MGMT* methylation and expression, suggesting TMZ treatment may benefit a subset of lung cancer patients methylated for *MGMT*, a hypothesis that warrants further preclinical studies prior to initiating clinical trials.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Abbreviations

ASM	allele specific methylation
COBRA	combined bisulfite modification and restriction analysis
HWE	Hardy-Weinberg equilibrium
IC50	the half maximal inhibitory concentration
LD	linkage disequilibrium
LSC	Lovelace Smokers Cohort
MAF	minor allele frequency
MSP	methylation specific PCR
MGMT	<i>O⁶-methylguanine-DNA methyltransferase</i>
NHBEC	normal human bronchial epithelial cell
TMZ	temozolamide
VSC	Veteran Smokers Cohort

Acknowledgments

Grant Support

This work was supported by National Cancer Institute (U01 CA097356 and R01 ES008801 to S.A.B. and 1K01CA128823-01 to K.G.F.) and the State of New Mexico as a direct appropriation from the Tobacco Settlement Fund to S.A.B.

References

1. Belinsky SA. Silencing of genes by promoter hypermethylation: key event in rodent and human lung cancer. *Carcinogenesis*. 2005; 26:1481–1487. [PubMed: 15661809]
2. Belinsky SA, Klinge DM, Liechty KC, et al. Plutonium targets the p16 gene for inactivation by promoter hypermethylation in human lung adenocarcinoma. *Carcinogenesis*. 2004; 25:1063–1067. [PubMed: 14742312]
3. Cuzzo C, Porcellini A, Angrisano T, et al. DNA Damage, Homology-Directed Repair, and DNA Methylation. *PLoS Genet*. 2007; 3:e110. [PubMed: 17616978]
4. O'Hagan HM, Mohammad HP, Baylin SB. Double strand breaks can initiate gene silencing and SIRT1-dependent onset of DNA methylation in an exogenous promoter CpG island. *PLoS Genet*. 2008; 4:e1000155. [PubMed: 18704159]
5. Leng S, Stidley C, Willink R, et al. Double-strand break damage and associated DNA repair genes predispose smokers to gene methylation. *Cancer Res*. 2008; 68:3049–3056. [PubMed: 18413776]
6. Stidley CA, Picchi MA, Leng S, et al. Multivitamins, folate, and green vegetables protect against gene promoter methylation in the aerodigestive tract of smokers. *Cancer Res*. 2010; 70:568–574. [PubMed: 20068159]

7. McCabe MT, Lee EK, Vertino PM. A multifactorial signature of DNA sequence and polycomb binding predicts aberrant CpG island methylation. *Cancer Res.* 2009; 69:282–291. [PubMed: 19118013]
8. Feltus FA, Lee EK, Costello JF, et al. Predicting aberrant CpG island methylation. *Proc Natl Acad Sci U S A.* 2003; 100:12253–12258. [PubMed: 14519846]
9. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell.* 2006; 125:301–313. [PubMed: 16630818]
10. Widschwendter M, Fiegl H, Egle D, et al. Epigenetic stem cell signature in cancer. *Nat Genet.* 2007; 39:157–158. [PubMed: 17200673]
11. Bumber YA, Kondo Y, Chen X, et al. An Sp1/Sp3 binding polymorphism confers methylation protection. *PLoS Genet.* 2008; 4:e1000162. [PubMed: 18725933]
12. Kerkel K, Spadola A, Yuan E, et al. Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nat Genet.* 2008; 40:904–908. [PubMed: 18568024]
13. Zhang Y, Rohde C, Reinhardt R, et al. Non-imprinted allele-specific DNA methylation on human autosomes. *Genome Biol.* 2009; 10:R138. [PubMed: 19958531]
14. Hegi ME, Liu L, Herman JG, et al. Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity. *J Clin Oncol.* 2008; 26:4189–4199. [PubMed: 18757334]
15. Esteller M, Hamilton SR, Burger PC, et al. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Res.* 1999; 59:793–797. [PubMed: 10029064]
16. Esteller M, Toyota M, Sanchez-Cespedes M, et al. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is associated with G to A mutations in K-ras in colorectal tumorigenesis. *Cancer Res.* 2000; 60:2368–2371. [PubMed: 10811111]
17. Yin D, Xie D, Hofmann WK, et al. DNA repair gene O6-methylguanine-DNA methyltransferase: promoter hypermethylation associated with decreased expression and G:C to A:T mutations of p53 in brain tumors. *Mol Carcinog.* 2003; 36:23–31. [PubMed: 12503076]
18. Belinsky SA, Liechty KC, Gentry FD, et al. Promoter hypermethylation of multiple genes in sputum precedes lung cancer incidence in a high-risk cohort. *Cancer Res.* 2006; 66:3338–3344. [PubMed: 16540689]
19. Weller M, Stupp R, Reifenberger G, et al. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nat Rev Neurol.* 2010; 6:39–51. [PubMed: 19997073]
20. Harris LC, Remack JS, Brent TP. Identification of a 59 bp enhancer located at the first exon/intron boundary of the human O6-methylguanine DNA methyltransferase gene. *Nucleic Acids Res.* 1994; 22:4614–4619. [PubMed: 7984409]
21. Ogino S, Hazra A, Tranah GJ, et al. MGMT germline polymorphism is associated with somatic MGMT promoter methylation and gene silencing in colorectal cancer. *Carcinogenesis.* 2007; 28:1985–1990. [PubMed: 17621591]
22. Hawkins NJ, Lee JH, Wong JJ, et al. MGMT methylation is associated primarily with the germline C>T SNP (rs16906252) in colorectal cancer and normal colonic mucosa. *Mod Pathol.* 2009; 22:1588–1599. [PubMed: 19734844]
23. Tessema M, Willink R, Do K, et al. Promoter methylation of genes in and around the candidate lung cancer susceptibility locus 6q23–25. *Cancer Res.* 2008; 68:1707–1714. [PubMed: 18339850]
24. Chin LJ, Ratner E, Leng S, et al. A SNP in a let-7 microRNA complementary site in the KRAS 3' untranslated region increases non-small cell lung cancer risk. *Cancer Res.* 2008; 68:8535–8540. [PubMed: 18922928]
25. Tessema M, Yu YY, Stidley CA, et al. Concomitant promoter methylation of multiple genes in lung adenocarcinomas from current, former and never smokers. *Carcinogenesis.* 2009; 30:1132–1138. [PubMed: 19435948]
26. Nakagawachi T, Soejima H, Urano T, et al. Silencing effect of CpG island hypermethylation and histone modifications on O6-methylguanine-DNA methyltransferase (MGMT) gene expression in human cancer. *Oncogene.* 2003; 22:8835–8844. [PubMed: 14647440]

27. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet.* 2001; 68:978–989. [PubMed: 11254454]
28. Stephens M, Donnelly P. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet.* 2003; 73:1162–1169. [PubMed: 14574645]
29. Stram DO, Haiman CA, Hirschhorn JN, et al. Choosing haplotype-tagging SNPs based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the Multiethnic Cohort Study. *Hum Hered.* 2003; 55:27–36. [PubMed: 12890923]
30. Menashe I, Rosenberg PS, Chen BE. PGA: power calculator for case-control genetic association analyses. *BMC Genet.* 2008; 9:36. [PubMed: 18477402]
31. Egyházi S, Ma S, Smoczynski K, et al. Novel O6-methylguanine-DNA methyltransferase SNPs: a frequency comparison of patients with familial melanoma and healthy individuals in Sweden. *Hum Mutat.* 2002; 20:408–409.
32. Belinsky SA. Gene-promoter hypermethylation as a biomarker in lung cancer. *Nat Rev Cancer.* 2004; 4:707–717. [PubMed: 15343277]
33. Krześniak M, Butkiewicz D, Samoedny A, et al. Polymorphisms in TDG and MGMT genes - epidemiological and functional study in lung cancer patients from Poland. *Ann Hum Genet.* 2004; 68:300–312. [PubMed: 15225156]
34. Wu H, Zhang X, Ling G, et al. Mechanisms of differential expression of the CYP2A13 7520C and 7520G alleles in human lung: allelic expression analysis for CYP2A13 heterogeneous nuclear RNA, and evidence for the involvement of multiple Cis-regulatory single nucleotide polymorphisms. *Pharmacogenet Genomics.* 2009; 19:852–863. [PubMed: 20431511]
35. Cross SH, Bird AP. CpG islands and genes. *Curr Opin Genet Dev.* 1995; 5:309–314. [PubMed: 7549424]
36. Pulling LC, Divine KK, Klinge DM, et al. Promoter hypermethylation of the O6-methylguanine-DNA methyltransferase gene: more common in lung adenocarcinomas from never-smokers than smokers and associated with tumor progression. *Cancer Res.* 2003; 63:4842–4848. [PubMed: 12941804]
37. Zawlik I, Vaccarella S, Kita D, et al. Promoter methylation and polymorphisms of the MGMT gene in glioblastomas: a population-based study. *Neuroepidemiology.* 2009; 32:21–29. [PubMed: 18997474]
38. Rønneberg JA, Tost J, Solvang HK, et al. GSTP1 promoter haplotypes affect DNA methylation levels and promoter activity in breast carcinomas. *Cancer Res.* 2008; 68:5562–5571. [PubMed: 18632608]
39. Chan TL, Yuen ST, Kong CK, et al. Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer. *Nat Genet.* 2006; 38:1178–1183. [PubMed: 16951683]
40. Candiloro IL, Dobrovic A. Detection of MGMT promoter methylation in normal individuals is strongly associated with the T allele of the rs16906252 MGMT promoter single nucleotide polymorphism. *Cancer Prev Res (Phila Pa).* 2009; 2:862–867.
41. Song JZ, Stirzaker C, Harrison J, et al. Hypermethylation trigger of the glutathione-S-transferase gene (GSTP1) in prostate cancer cells. *Oncogene.* 2002; 21:1048–1061. [PubMed: 11850822]
42. Stirzaker C, Song JZ, Davidson B, et al. Transcriptional gene silencing promotes DNA hypermethylation through a sequential change in chromatin modifications in cancer cells. *Cancer Res.* 2004; 64:3871–3877. [PubMed: 15172996]
43. Via M, Gignoux C, Burchard EG. The 1000 Genomes Project: new opportunities for research and social challenges. *Genome Med.* 2010; 2:3. [PubMed: 20193048]
44. Rudin CM, Avila-Tang E, Harris CC, et al. Lung cancer in never smokers: molecular profiles and therapeutic implications. *Clin Cancer Res.* 2009; 15:5646–5661. [PubMed: 19755392]
45. Wolf J. Nitrosoureas as single agents in the treatment of pulmonary cancer. *Cancer Treat Rep.* 1976; 60:753–756. [PubMed: 782697]
46. Ahmann DL, Carr DT, Coles DT, et al. Evaluation of cyclophosphamide (NSC-26271) and 1,3-bis(2-chloroethyl)-1-nitrosourea (BCNU; NSC-409962) in the treatment of patients with inoperable or disseminated lung cancer. *Cancer Chemother Rep.* 1972; 56:401–403. [PubMed: 19051501]

47. Siena S, Crinò L, Danova M, et al. Dose-dense temozolomide regimen for the treatment of brain metastases from melanoma, breast cancer, or lung cancer not amenable to surgery or radiosurgery: a multicenter phase II study. *Ann Oncol.* 2010; 21:655–661. [PubMed: 19767314]

Figure 1. The LD map of nine SNPs identified in the 5' region of *MGMT* and the four common haplotypes is depicted. The number in each diamond represents the r^2 between any two corresponding SNPs. Three SNPs in grey rectangles (#1, 3, and 7) were selected as htSNPs. The base components of the four common haplotypes and their frequency in 49 subjects are displayed as well.

Figure 2. The A allele confers reduced transcription and expression of *MGMT*. (A) Haplotype 4 (HAP4) carrying the A allele of rs16906252 shows a 20 – 40% reduction in promoter activity in three lung cell lines compared to HAP1. Values are mean $\pm$ SD from three experiments. (B) A 153 bp fragment surrounding the rs16906252 was amplified from cDNA isolated from these two NHBEs and then digested by FspI. Densitometry analysis showed that the expression of the A allele was reduced 31–39% compared with the G allele, while rs16906252 dependent methylation of *MGMT* leads to selective loss of A allele expression in Calu-6 and H1993. As expected, expression of only the G allele is seen in cell lines (HCCC4006, H23, and H2009) with G/G genotype that are unmethylated for *MGMT*.

Figure 3.

ASM of *MGMT*. (A) Bisulfite sequencing of COBRA product from Calu-6 encompassing CpGs with locations from -5 to $+148$ relative to the translation start site shows methylation of only the A allele. (B) Bisulfite sequencing of the MSP product in a primary lung tumor sample encompassing CpGs with locations from $+27$ to $+181$ relative to the translation start site heterozygous for rs16906252 shows selective methylation of the A allele.

Figure 4. MGMT expression and TMZ IC50s in 5 lung tumor cell lines (Δ) is highly correlated ($r = 0.99$).

Table 1

Association between SNPs and risk for *MGMT* methylation in lung adenocarcinoma and in sputum from smokers

Samples/Cohorts (n)	SNPs	Genotypes	Methylated/unmethylated	OR (95%CI)
Lung Adenocarcinoma				
(n=169)	rs1711646 [*]	G/G	39/66	1.0
		G/T	15/35	0.7 (0.3 – 1.4)
		T/T	5/9	0.9 (0.2 – 3.2)
	rs1625649 [*]	C/C	17/56	1.0
		C/A	32/42	2.4 (1.1 – 4.9)
		A/A	10/12	2.7 (1.0 – 7.6)
	rs16906252 ^{*†}	G/G	36/109	1.0
		G/A	23/1	97.0 (12.1 – 778.4)
		A/A	Not detected	—
Sputum				
Lovelace Smokers cohort (n=1256)	rs16906252 [‡]	G/G	207/853	1.0
		G/A	113/75	6.3 (4.5 – 8.7)
		A/A	6/2	12.4 (2.5 – 62.8)
Veteran Smokers cohort (n=475)	rs16906252 [§]	G/G	82/332	1.0
		G/A	39/21	8.0 (4.4 – 14.5)
		A/A	1/0	Not calculated ^{//}

* Age, gender, race (white vs. nonwhite), and smoking status (ever vs. never) were used as adjustment factors in logistic regression.

† Due to small sample size in one cell, exact logistic regression was used.

‡ Age, gender, race, current smoking status, and pack years were used as adjustment factors in logistic regression in the Lovelace Smokers cohort.

§ Age, race, current smoking status, and pack years were used as adjustment factors in logistic regression in VA Smokers cohort.

¶ OR for A/A genotype in the VA cohort was not calculated because no A/A genotype was identified in the unmethylated group.

Table 2*MGMT* methylation status, mRNA expression, rs16906252 genotype and TMZ IC50s

Cell line	<i>MGMT</i> Methylation *	Genotype	% Expression [†]	TMZ IC50 [‡]
Lung				
HCC4006	Unmethylated	G/G	162	1655
H2085	Unmethylated	G/G	232	2500
Calu-6	Hemi-methylated	G/A	50	680
Calu3	Hemi-methylated	G/G	64	693
A549	Hemi-methylated	G/G	70	500
Brain				
U251	Fully methylated	G/G	0	270
T98G	Sparsely methylated	G/G	100	1590

* *MGMT* methylation status was measured by COBRA.

[†] % expression was the average from 3 cDNAs independently made from one mRNA sample and was relative to the average expression seen in the NHBEC lines.

[‡] TMZ IC50 was the average from 3 independent experiments.