Double-Strand Break Damage and Associated DNA Repair Genes Predispose Smokers to Gene Methylation

Shuguang Leng, Christine A. Stidley, Randy Willink, Amanda Bernauer, Kieu Do, Maria A. Picchi, Xin Sheng, Melissa A. Frasco, David Van Den Berg, Frank D. Gilliland, Christopher Zima, Richard E. Crowell, and Steven A. Belinsky

Abstract
Gene promoter hypermethylation in sputum is a promising biomarker for predicting lung cancer. Identifying factors that predispose smokers to methylation of multiple gene promoters in the lung could affect strategies for early detection and chemoprevention. This study evaluated the hypothesis that double-strand break (DSB) repair capacity and sequence variation in genes in this pathway are associated with a high methylation index in a cohort of current and former cancer-free smokers. A 50% reduction in the mean level of DSB repair capacity was seen in lymphocytes from smokers with a high methylation index, defined as three or more of eight genes methylated in sputum, compared with smokers with no genes methylated. The classification accuracy for predicting risk for methylation was 88%. Single nucleotide polymorphisms within the MRE11A, CHEK2, XRCC3, DNA-PKc, and NBN DNA repair genes were highly associated with the methylation index. A 14.5-fold increased odds for high methylation was seen for persons with seven or more risk alleles of these genes. Promoter activity of the MRE11A gene that plays a critical role in recognition of DNA damage and activation of ataxia-telangiectasia mutated was reduced in persons with the risk allele. Collectively, ours is the first population-based study to identify DSB DNA repair capacity and specific genes within this pathway as critical determinants for gene methylation in sputum, which is, in turn, associated with elevated risk for lung cancer. [Cancer Res 2008;68(8):3049–56]

Introduction
Lung cancer, the leading cause of cancer mortality in both men and women in the United States, now accounts for ~30% of all deaths from cancer (1). The lack of a validated screening approach for early detection and the resistance of advanced-stage tumors to therapy are largely responsible for the 5-year survival rate of 14% for lung cancer patients (2). The discovery of field cancerization in the respiratory tract of smokers prompted studies leading to the discovery that inactivation of genes, such as p16, by promoter hypermethylation occurs in precursor lesions to non–small cell lung cancer (3). This finding suggested that methylation, when detected in exfoliated cells within sputum, could serve as a biomarker for the early stages of lung carcinogenesis (4). To test this hypothesis, our group examined a large panel of genes for their ability to predict lung cancer in a nested case-control study. A combination of six genes was identified whose methylation in sputum predicted lung cancer before clinical diagnosis with both a sensitivity and specificity of 65% (5). A better understanding of the susceptibility factors that predispose smokers to the acquisition of multiple epigenetic alterations in key cellular regulatory genes within the respiratory epithelium could improve prediction of lung cancer risk and affect strategies for early detection and chemoprevention.

The precise mechanisms by which carcinogens disrupt the capacity of the cells to maintain the normal epigenetic code during DNA replication and repair are largely unknown. Smoking accounts for >90% of lung cancer. Carcinogens within tobacco induce single and double-strand breaks (DSB) in DNA, and reduced capacity for repair of DNA damage has been associated with lung cancer (6). Accumulating evidence suggests that extensive DNA damage, manifested through DSBs, could in part be responsible for the acquisition of aberrant gene promoter methylation during lung carcinogenesis. For example, the prevalence of promoter methylation of the p16 gene was significantly greater in adenocarcinomas from workers occupationally exposed to plutonium, an exposure that predominantly produces DSBs, than in cancer from unexposed smokers (7). The prevalence of p16 methylation increased with increasing plutonium exposure. In a second study, the prevalence of methylation of the estrogen receptor α gene promoter was greater in plutonium-induced adenocarcinomas in rodent lung tumors compared with tumors induced by 4-(methylnitrosamino)-1-(3-pyridyl)-1-butaneone, diesel exhaust, or carbon black exposures, which mainly induce single-strand breaks of DNA (8). These studies support the hypothesis that DSBs may play a key role in the development of aberrant gene promoter hypermethylation. The purpose of this study was to test the hypothesis that a high methylation index (defined as the methylation of three or more gene-specific promoters detected in sputum) is associated with a reduced capacity to repair DSBs. We also hypothesize that sequence variation in genes from the DSB repair pathway will predict for high methylation index.

Materials and Methods

Study population and sample collection. The Lovelace Smokers Cohort (n = 1,860) was established in 2001 to conduct longitudinal studies on molecular markers of respiratory carcinogenesis in biological fluids, such as sputum, from people at risk for lung cancer. The institutional review board of the Lovelace Respiratory Research Institute approved a protocol outlining the procedures for subject enrollment, biological specimen collection (sputum and blood), processing, and data management. Cohort...
subjects were recruited through advertisement in local newspaper and in radio and television programs and were mainly residents in the Albuquerque metropolitan area. Enrolled subjects were between 40 and 75 y of age with at least a 20 pack-year smoking history. All participants signed the consent form. At enrollment, individual information about medical, family, and smoking exposure history and quality of life was collected through a computer-based system. Induced sputum and blood were collected and pulmonary function testing was performed. Blood was processed within 2 h after blood draw to isolate lymphocytes and plasma.

Cytoscreening of lymphocytes began in 2005.

Cytologically adequate sputum samples from 824 cohort subjects were evaluated for gene promoter methylation of eight genes as described below. High methylation index was defined as the methylation of three or more gene-specific promoters in sputum. We selected persons from our cohort that exhibited a high (cases) or low [controls (zero of eight genes)] methylation index. To increase the stringency for case selection, GATA4, which was most commonly methylated in sputum, was excluded as one of the three methylated genes needed for case classification and 131 of 824 cohort subjects met this criteria. Cases were frequency matched by gender to controls. Cases (n = 131) and controls (n = 130) were selected for the genetic association study. Among the 131 cases, 77 had adequate number of cryopreserved lymphocytes for the mutagen sensitivity assay. Seventy-eight controls were selected from the 130 controls, with frequency matching by gender maintained, for the mutagen sensitivity assay.

**Sputum cytology and nested methylation-specific PCR.** Sputum samples were stored in Saxenmann’s fixative. Three slides were made for each sputum sample to check for adequacy defined as the presence of deep lung macrophages or CURschmann’s spiral (9). The methylation-specific PCR (MSP) assay was only performed on cytologically adequate sputum lung macrophages or CURschmann’s spiral (9). The methylation-specific PCR (MSP) assay was only performed on cytologically adequate sputum samples. Eight genes [p16, O-methylgauganine-DNA methyltransferase (MGMT), death-associated protein kinase (DAPK), ras effector homologue 1 (RASSF1A), PAX5a, PAX5b, GATA4, and GATA5] were selected for analysis of methylation in sputum based on our previous studies establishing their association with risk for lung cancer (5, 10–12). Nested MSP was used to detect methylated alleles in DNA recovered from the sputum samples as described (5, 10–12).

**Evaluation of DSB repair capacity in peripheral lymphocytes.** Phytohemagglutinin (PHA)-stimulated lymphocytes were treated with bleomycin to evaluate the generation of chromosome aberrations as an index of DSB repair capacity (DSBRC; ref. 13). Briefly, cryopreserved lymphocytes were thawed and cultured in RPMI 1640 supplemented with fetal bovine serum (20%) and PHA (1.5%) at a cell density of <0.5 × 10^6/mL. Sixty-seven hours after PHA stimulation, the culture was split into two T25 flasks and treated with bleomycin or vehicle for 5 h. The final concentration for bleomycin in culture medium was 3 units/L, a concentration defined as SNPs and case-control status, and the outcome variable, DSBRC, with the appropriate statistical method implemented in the program PHASE (version 2.1) was used to reconstruct the haplotypes from the SNPs in the MRE11A gene for the 261 subjects. Two subjects homozygous for the haplotype that contained the rs171042 SNP associated with high methylation index were selected. The other four people selected were each homozygous for one of the other four haplotypes. The MRE11A promoter fragment (−2,541 to −5 with +1 being the translational start site) was amplified from lymphocyte DNA from these six subjects. The promoter fragment was directionally subcloned into the pG2-basic luciferase reporter vector (Promega) upstream of the luciferase coding sequence. Five clones from each person were commercially sequenced to identify variants within the promoter region (Sequencher).

**Transient transfection and reporter gene assays.** The Calu6 lung tumor-derived cell line was used for transient transfections. Cells (1.5 × 10^5) were plated into six-well dishes and transfected the following day. Plasmid DNA (1 μg) and the pSV-β-galactosidase control vector (0.5 μg; Promega) were cotransfected into cells with Fugene 6 transfection reagent (Roche Diagnostics) at a Fugene 6 to DNA ratio of 3:1. A promoter-pG2-luciferase vector and the pG2-control vector that contains the SV40 promoter were used as negative and positive controls, respectively. Forty-eight hours after transfection, cells were harvested and lysed. Immediately after lysing, cell extracts were assayed in a luminometer for luciferase activity using the Luminoskan Ascent luminometer (Thermo Electron) for luciferase activity using the Luciferase Assay System (Promega). β-Galactosidase activity in cell lysates was measured using the Galacto-Star Reporter Gene Assay System (Tropix). Promoter activity was calculated as the ratio of activities of luciferase and β-galactosidase. Transfections were done in duplicate in four independent experiments.

**Statistical analysis.** The two-sample t test, Wilcoxon rank sum test, and χ^2 test were used to compare the mean or distribution of several demographic variables and DSBRC results between cases and controls as appropriate. Because the DSBRC data and the number of spontaneous breaks were not normally distributed, analysis was also performed on log-transformed data. The results based on log-transformed data were similar to those based on untransformed data, so only results based on untransformed data are shown. Analysis of covariance and logistic regression were used to assess the association between selected variables, such as SNPs and case-control status, and the outcome variable, DSBRC, with adjustment of covariates selected a priori (age at sputum collection, sex, race, current smoking status, and pack-years). DSBRC was dichotomized for logistic regression models using the upper quartile of DSBRC in control participants. The selection of the upper quartile of DSBRC in controls as the cutoff value was based on the distribution of DSBRC in cases and controls. Analysis of covariance and logistic regression models,
stratified by status, were also examined for different associations between SNPs and DSBRC by case-control status. A receiver operator characteristic (ROC) curve was also drawn to compare the sensitivity and specificity of DSBRC induced by bleomycin for classifying cases (16). Multivariate unconditional logistic regression assessed the association between SNPs and the outcome of case-control status, with the same covariates outlined above. Model results are presented as odds ratios (OR) with 95% confidence intervals (95% CI) for having three or more methylated genes. Logistic regression modeling was extended to generalized logit models to more precisely examine the high methylation index. ORs and 95% CIs for the risk of having three, four, or five or more methylated genes with zero methylated genes as the reference group were obtained with adjustment for the same covariates.

The call rate for each SNP was assessed before data analysis. For the 294 SNPs assayed, 42 were deemed unsuitable because they were monomorphic, had a mean allele frequency of <0.05, had low yield (<80%), or showed a highly significant distortion from Hardy-Weinberg equilibrium (P < 0.0001). These SNPs were removed from analysis. Four models were tested: codominant, dominant, additive, and recessive. Because of power limitations, only results for the additive model are presented for each SNP.

Results

Gene methylation in sputum. Gene promoter methylation was assessed in sputum from 824 members of the Lovelace Smokers Cohort, a cohort of current and former cancer-free smokers (Table 1). Methylation of an eight-gene panel that included p16, MGMT, DAPK, RASSF1A, GATA4, GATA5, PAX5a, and PAX5b was evaluated. Methylation of these genes has been associated with increased risk for lung cancer (5, 10–12). The prevalence of methylation ranged from 12% for RASSF1A to 31% for GATA4 and was not associated with family history for lung cancer (Supplementary Table S2). Nineteen percent of cohort members were methylated for three or more genes (Supplementary Table S2). Our previous nested case-control study within the Colorado Cohort revealed that methylation of three or more genes from a six-gene panel (excluding GATA4 and PAX5a) was associated with a 6.5-fold increased risk for lung cancer (5).

Repair capacity associates with methylation index. The mutagen sensitivity assay was used to assess DSBRC (13). The
number of chromatid breaks induced in lymphocytes following exposure to bleomycin, a radiomimetic agent, was used to measure DSBRC. We selected persons from our cohort who exhibited a high [cases (three or more methylated genes)] or low [controls (zero of eight genes methylated)] methylation index because of the increased risk for lung cancer seen in our Colorado nested, case-control study when three or more genes were methylated in sputum. Cryopreserved lymphocytes were available for assessment of DSBRC for 77 cases and 78 controls. Demographics and smoking history for cases and controls are detailed in Table 1. A highly statistically significant difference was seen in DSBRC ($P < 0.001$) between cases and controls with a mean number of chromosome breaks per cell of $0.47 \pm 0.11$ and $0.32 \pm 0.10$, respectively (Fig. 1A). The mean number of bleomycin-induced chromatid breaks per cell was significantly higher in cases than in controls when subjects were stratified by age, sex, race, chronic airway obstruction, pack-years, and smoking status, indicating that none of these covariates was major confounders for the strong association seen between DNA repair capacity and methylation index (Supplementary Table S3). We further classified the cases into three groups based on the number of methylated genes (three, four, and five or more methylated genes) and found that the number of chromatid breaks per cell induced by bleomycin increased with the increasing number of methylated genes in sputum ($P < 0.001$; Fig. 1B). Age did not differ in cases with three, four, and five or more methylated genes. Finally, after adjusting for sex, race, current smoking status, cigarette pack-years, seeding number of lymphocytes, cryopreservation time, and log-transformed spontaneous chromatid breaks per cell, age was the only factor significantly associated with chromatid breaks induced by bleomycin in both cases and controls (Supplementary Table S3). The reduction of DNA repair capacity with age is well established and supports the accuracy of the mutagen sensitivity assay in this study (13).

A ROC curve was generated to determine how well DSBRC distinguished cases from controls. The ROC curve shows that DSBRC significantly ($P < 0.0001$) increased the classification accuracy from 66% to 88% for predicting risk for promoter methylation (Fig. 1C). With the sensitivity set at 80%, the false-positive rate was <20%.

The OR and 95% CI were calculated to further characterize the association between methylation index and DSBRC. The OR associated with an increment of 0.1 chromatid breaks per cell for having three or more methylated genes was 6.6 (95% CI, 3.7–13.2) for bleomycin treatment after adjustment for selected covariates. When the number of methylated genes was used as a multinomial response variable, the OR associated with an increment of 0.1 chromatid breaks per cell for methylation of three, four, and five or more genes was 5.3 (95% CI, 2.7–11.4), 7.1 (95% CI, 3.8–14.8), and

![Figure 1](https://example.com). DNA repair capacity is associated with gene promoter methylation in sputum. A, bleomycin treatment causes an increased number of chromatid breaks per cell in lymphocytes from cases (methylated group) compared with controls (unmethylated group; $P < 0.0001$). B, positive association between number of methylated genes and chromatid breaks per cell. Sample size for each group is indicated in parentheses. C, ROC curve comparing sensitivity and specificity of DNA repair capacity for classifying cases and controls. The covariates included in the ROC curve were age at sputum collection, sex, race, current smoking status, and pack-years. D, distribution of chromatid breaks per cell by case-control status.
8.5 (95% CI, 4.2–19.3), respectively. A histogram detailing the distribution of chromatid breaks per cell by case-control status revealed that 75% of controls compared with 18% of cases accumulated <0.38 breaks per cell (Fig. 1D). We chose to dichotomize the number of DSBs per cell at the 75th percentile for controls and this gave an adequate overlap in the distribution of breaks per cell in cases and controls. The results did not differ significantly with other cut points. When methylation index was compared with chromatid breaks per cell, the ORs for detecting methylation increased monotonically from 10 to 15 to 26 (Table 2). Overall, chromatid breaks ≥0.38 per cell were associated with a 14.5-fold increased risk of having three or more methylated genes in sputum (Table 2).

### SNPs within DNA repair genes and risk for methylation

The finding that DNA repair capacity strongly predicts for high methylation index, combined with its high heritability, suggests that variants in genes involved in repair should also be predictive (26). We selected 16 candidate genes from the DSBR and cell cycle control pathways for tag SNP–based genotyping (Supplementary Table S1). A total of 294 SNPs were evaluated for 131 cases and 130 controls that included the subset evaluated in the mutagen sensitivity assay.

### Table 2. Analysis of the relationship between chromatid breaks per cell categorized by upper quartile of control subjects and risk of methylation

<table>
<thead>
<tr>
<th>Variable</th>
<th>Chromatid breaks/cell</th>
<th>Crude OR (95% CI)</th>
<th>Adjusted OR (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>&lt;0.38 (n)</td>
<td>≥0.38 (n)</td>
<td></td>
</tr>
<tr>
<td><strong>Methylation index</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>58</td>
<td>20</td>
<td>1.0 (reference)</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>14</td>
<td>8.1 (2.6–25.4)</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>32</td>
<td>13.3 (5.1–34.7)</td>
</tr>
<tr>
<td>≥5</td>
<td>2</td>
<td>17</td>
<td>24.7 (5.2–116.2)</td>
</tr>
<tr>
<td><strong>Case control</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Controls</td>
<td>58</td>
<td>20</td>
<td>1.0 (reference)</td>
</tr>
<tr>
<td>Cases</td>
<td>14</td>
<td>63</td>
<td>13.1 (6.0–23.2)</td>
</tr>
</tbody>
</table>

*Obtained from models with adjustment for age at sputum collection, sex, race, smoking status, and pack-years.
† Upper quartile of breaks per cell induced by bleomycin in control subjects.
‡ Using generalized logit model.
§ Using logistic regression model.

### Table 3. Summary of associations between genes and promoter methylation using stepwise logistic regression

<table>
<thead>
<tr>
<th>Gene/SNPs*</th>
<th>OR† (95% CI)</th>
<th>P</th>
<th>Permuted P‡</th>
<th>Statistical power§</th>
<th>Prior probability of FPRPk</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRE11A/rs7117042</td>
<td>3.97 (1.77–8.89)</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.37</td>
<td>0.006</td>
</tr>
<tr>
<td>CHEK2/rs5762763</td>
<td>1.89 (1.20–2.97)</td>
<td>0.0064</td>
<td>0.0058</td>
<td>0.88</td>
<td>0.021</td>
</tr>
<tr>
<td>XRCC3/rs2295146</td>
<td>0.54 (0.35–0.83)</td>
<td>0.0051</td>
<td>0.0073</td>
<td>0.83</td>
<td>0.018</td>
</tr>
<tr>
<td>DNA-Kc/rs7830743</td>
<td>0.38 (0.18–0.80)</td>
<td>0.0117</td>
<td>0.0142</td>
<td>0.70</td>
<td>0.048</td>
</tr>
<tr>
<td>NBN/rs6998169</td>
<td>0.47 (0.23–0.93)</td>
<td>0.0452</td>
<td>0.0308</td>
<td>0.89</td>
<td>0.132</td>
</tr>
<tr>
<td>LIG4/rs1151402</td>
<td>0.68 (0.44–1.06)</td>
<td>0.0859</td>
<td>0.1078</td>
<td>0.98</td>
<td>0.208</td>
</tr>
<tr>
<td>XRCC2/rs3218400</td>
<td>0.55 (0.28–1.06)</td>
<td>0.0751</td>
<td>0.0823</td>
<td>0.92</td>
<td>0.197</td>
</tr>
<tr>
<td>Ku80/rs828911</td>
<td>1.55 (1.02–2.37)</td>
<td>0.0416</td>
<td>0.059</td>
<td>0.95</td>
<td>0.116</td>
</tr>
<tr>
<td>RAD50/rs2244012</td>
<td>1.64 (0.94–2.76)</td>
<td>0.0864</td>
<td>0.1132</td>
<td>0.98</td>
<td>0.209</td>
</tr>
<tr>
<td>CHEK1/rs537046</td>
<td>0.64 (0.37–1.12)</td>
<td>0.1176</td>
<td>0.1091</td>
<td>0.97</td>
<td>0.267</td>
</tr>
</tbody>
</table>

*Age, sex, ethnicity, smoking status, and pack-years were selected a priori and forced in the model. Stepwise selection was only used to select genetic susceptibility factors. The P values for both entry and inclusion of a variable in each round of variable selection were set at 0.1.
† ORs were calculated using an additive model where common homozygote, heterozygote, and rare homozygote are coded as 0, 1, and 2, respectively.
‡ Case and control status was permuted 10,000 times.
§ Statistical power is the power to detect an OR of 2.0 for individual tag SNPs under an additive model.
k FPRPs are calculated based on Wacholder et al. (9). Prior probabilities were set at 0.1 to 0.25 for rs5762763, rs2295146, rs7830743, rs6998169, rs1151402, and rs218400 and at 0.01 to 0.1 for rs828911, rs7117042, rs7906967, rs2244012, and rs537046. Findings with FPRP < 0.2 were considered noteworthy.
Forty-four SNPs were associated with risk for promoter methylation ($P < 0.15$) with adjustment for covariates (Supplementary Table S4). Because of the relatively high correlation between SNPs in these genes, we tested which SNP, or set of SNPs, was most significantly associated with risk for promoter methylation by using a stepwise logistic regression model. The underlined SNPs in Supplementary Table S4 were selected from each gene ($P < 0.15$) to represent the allelic status for those genes. These 16 SNPs were then included with the covariates in one model and with four or less alleles (Table 4). The inclusion of 10 SNPs did not greatly increase the ability to classify persons with a high methylation index. This outcome was not surprising because the added SNPs were only weakly associated with risk for promoter methylation.

Reduced activity of the MRE11A promoter. The genes included in the prediction model have biological plausibility. Two of the five genes whose sequence variation is associated with methylation, NBN and XRCC3, have shown association with lung cancer (18, 27). SNPs within the DNA-PKc and CHEK2 genes have been associated with breast and other cancers, whereas no studies have been conducted with MRE11A (19, 28). Assessment of the functional potential of the SNPs identified from our study for these genes revealed that DNA-PKc/rs7830743 is a nonsynonymous SNP changing amino acid residue 3434 from isoleucine to threonine in exon 73. This amino acid substitution is predicted to change the secondary structure and may influence the serine/threonine protein kinase activity of this protein (29). We have shown that reduced DNA-PK activity is associated with risk for lung cancer and sensitivity to cell killing by bleomycin (30), thus supporting an important role for this gene in lung cancer and aberrant gene promoter methylation. The SNPs from the other four genes are neither nonsynonymous nor in high linkage disequilibrium with any nonsynonymous SNP with known function. However, MRE11A/rs7117042 and NBN/rs6998169 are predicted to locate in the middle of the sequence, forming DNA triplexes that could inhibit DNA transcription (31). To begin addressing function of these SNPs, we tested whether MRE11A/rs7117042 is associated with a reduction in promoter activity.

Two subjects homozygous for the haplotype containing rs7117042 and four subjects, each homozygous for one of the other four common haplotypes, were selected for assessment of promoter activity (described in Materials and Methods; Supplementary Table S5). Sequencing of the 2,500-bp promoter construct revealed three haplotypes (ACGACTG, GCACAT, and AGGCTTG), with each haplotype present in two subjects. The most distinct sequence difference was the G to C change at −590 bp. We genotyped 100 subjects selected randomly from our study population for this SNP and found that the G allele was in complete linkage.

The calculated FPRP was <0.2 for four SNPs (rs7117042, rs5762763, rs2295146, and rs7830743) under the assigned prior probability range (Table 3). Findings with a FPRP of ≤0.2 are considered to be noteworthy.

ROC curves were generated to evaluate the classification accuracy of this panel of SNPs to distinguish cases from controls. The area under the curve increased from 57% (covariates only) to 72% (covariates with the 5 most significant SNPs) and to 75% (covariates with all 10 SNPs; Fig. 2A). The difference between the area under the curve with only covariates and the two models that included both covariates and multiple SNPs is highly significant ($P < 0.001$). Restricting this analysis to include only cases and controls in which DSBRC was determined resulted in an area of 82% that increased to 93% when repair capacity was included in the model. To test the hypothesis that the identified SNPs in different genes would work additively to influence risk for promoter methylation, the joint effect of each SNP, inclusive of both putative susceptibility alleles, was evaluated. When the five SNPs with the strongest association with risk for promoter methylation were included, persons with five, six, or seven or more alleles were found to have a 2.5-, 2.8-, and 14.4-fold increased risk, respectively, for three or more methylated genes in sputum compared with those with four or less alleles (Table 4). The inclusion of 10 SNPs did not greatly increase the ability to classify persons with a high methylation index. This outcome was not surprising because the added SNPs were only weakly associated with risk for promoter methylation.
disorder with the T allele of the risk SNP rs7117042, identified to be most strongly associated with high methylation index. The highest promoter activity was seen in constructs containing the ACGACTG haplotype. With this haplotype as the reference, a 23% and 38% reduction in promoter activity was seen for the GCACAT and AGGCTTG haplotypes, respectively (Fig. 2B). These results show that the risk tag SNP is associated with a marked reduction in transcription of the MRE11A gene. MRE11A has a critical role in recognition of DSB damage. It complexes with Rad50 and Nbs1 to directly sense the DSBs, binds to the DNA, modifies the ends via 5′ to 5′ exonuclease activity, recruits ATM to the damaged DNA template, and dissociates the ATM dimer (32). Therefore, a reduction in level of the MRE11A protein could have a major effect on DSBRC.

Discussion

These results indicate a strong link between reduced DSBRC and risk for methylation in sputum. Our studies transcend from a functional assay for DNA repair to specific genotypes and finally show an activity deficit of the MRE11A gene that plays a critical role in recognition of DSB DNA damage and activation of the ATM gene (32). The mechanism underlying this association could, in part, be mediated by the genes that are recruited to sites of DSBs and the resultant modification of chromatin to facilitate repair. One of the earliest responses to DSB damage is phosphorylation by ATM kinase of the histone H2AX, which then facilitates accumulation of repair/signaling proteins and also SWI/SNF complexes that have been implicated in transcriptional silencing to chromatin regions distal to a DSB (33, 34). A recent study showed activation of H2AX by ATM in the A549 lung tumor-derived cell line by tobacco smoke (35).

Another key contributing factor for aberrant de novo methylation during DNA damage is the rapid recruitment of DNMT1 to sites of DNA damage (36). Le Gac et al. (37) found that in cells treated with doxorubicin, which induces DSBs, DNMT1 is recruited by activated p53 and binds to functional Sp1 sites within the damaged DNA template, and dissociates the ATM dimer (32). Therefore, a reduction in level of the MRE11A protein could have a major effect on DSBRC.

Table 4. Association between number of risk alleles and promoter methylation in the Lovelace Smokers Cohort

<table>
<thead>
<tr>
<th>No. high-risk alleles</th>
<th>Cases (n = 128), n (%)</th>
<th>Controls (n = 130), n (%)</th>
<th>OR (95% CI)*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Top 5 SNPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥4</td>
<td>9 (7.0)</td>
<td>29 (22.3)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>5</td>
<td>36 (28.1)</td>
<td>45 (34.6)</td>
<td>2.54 (1.06–6.53)</td>
</tr>
<tr>
<td>6</td>
<td>37 (29.8)</td>
<td>44 (33.9)</td>
<td>2.84 (1.18–7.33)</td>
</tr>
<tr>
<td>≥7</td>
<td>46 (35.9)</td>
<td>12 (9.2)</td>
<td>14.39 (5.37–42.45)</td>
</tr>
<tr>
<td>All 10 SNPs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤10</td>
<td>13 (10.3)</td>
<td>50 (39.1)</td>
<td>1.00 (reference)</td>
</tr>
<tr>
<td>12</td>
<td>31 (24.6)</td>
<td>34 (26.6)</td>
<td>3.68 (1.69–8.40)</td>
</tr>
<tr>
<td>≥13</td>
<td>56 (44.4)</td>
<td>18 (14.1)</td>
<td>13.73 (6.08–33.21)</td>
</tr>
</tbody>
</table>

*Unconditional logistic regression with adjustment for age, sex, ethnicity, smoking status, and pack-years.

† Top 5 SNPs include rs7117042, rs5762763, rs2295146, rs7830743, and rs6998169.

‡ All 10 SNPs include rs7117042, rs5762763, rs2295146, rs7830743, rs6998169, rs1131402, rs3218400, rs828911, and rs537046.

DSB DNA damage induced by doxorubicin, DNMT1 complexed with p53 was recruited to the survivin gene promoter followed by de novo methylation and gene silencing (39). Cuozzo et al. (40) provides even stronger support for a mechanistic link between DNA damage and methylation. In that study, a recombinant plasmid containing a 1-SCE1 restriction site within one copy of two inactivated tandem repeated green fluorescent protein (GFP) genes was introduced into HeLa or mouse embryonic stem cells. The restriction endonuclease 1-SCE1 was added to the cell to induce a DSB in the GFP gene at this site. Rapid gene silencing associated with homologous recombination and DNA methylation of the recombinant gene was seen and could be blocked by treatment with the demethylating agent 5-aza-deoxycytidine. Chromatin immunoprecipitation revealed that DNMT1 was bound specifically to the homologous GFP DNA. Together, these in vitro studies strongly support a direct mechanistic link between DNA damage and induction of de novo methylation by DNMT1. Our population-based studies now provide for the first time an in vivo association between DNA repair capacity and gene promoter methylation, both through a functional assay and genetic variants in genes within the DSB repair pathway. Thus, in the absence of efficient repair, the recruitment of p53, DNMT1, and transcriptional repressors to many genes, such as p16, which also contain Sp1 sites within its promoter, could lead to de novo methylation and gene silencing.

The identification of DSBRC and specific genes within this pathway as a critical determinant for gene promoter hypermethylation has important implications for basic and translational science. Our study substantiates that DNA damage that has long been recognized as an initiating event for mutagenesis is also likely a major factor in initiating aberrant promoter hypermethylation. Other DNA damage response pathways, such as apoptosis, nucleotide, and base excision repair, may also contribute to the induction of aberrant promoter hypermethylation. A major priority for our research is to replicate the provocative findings in this study along with our emerging methylation gene panel in a prospective population-based study. Genetic variants associated with promoter hypermethylation could be used to identify young smokers who would be most susceptible to induction of preneoplasia and, thus, should receive chemoprevention. In addition, the integration of these genetic variants with detection of gene promoter hypermethylation in sputum in long-term heavy smokers...
could lead to the first diagnostic test for incident lung cancer and affect long-term survival from this fatal disease.

Acknowledgments

Received 11/21/2007; revised 1/24/2008; accepted 2/15/2008.

Conflict of interest statement: S.A. Belinsky is a consultant to Oncomethylome Sciences. Under a licensing agreement between Lovelace Respiratory Research 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.

Grant support: U01 CA97956 and the State of New Mexico as a direct appropriation from the Tobacco Settlement Fund.

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.

We thank Tristin Maroney for her technical assistance.

References


Double-Strand Break Damage and Associated DNA Repair Genes Predispose Smokers to Gene Methylation

Shuguang Leng, Christine A. Stidley, Randy Willink, et al.


Updated version  Access the most recent version of this article at: http://cancerres.aacrjournals.org/content/68/8/3049

Cited articles  This article cites 38 articles, 25 of which you can access for free at: http://cancerres.aacrjournals.org/content/68/8/3049.full.html#ref-list-1

Citing articles  This article has been cited by 21 HighWire-hosted articles. Access the articles at: /content/68/8/3049.full.html#related-urls

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

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

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