Methyl-CpG Binding Domain 1 Gene Polymorphisms and Risk of Primary Lung Cancer

Jin-Sung Jang,[1] Su Jeong Lee,[1] Jin Eun Choi,[3] Sung Ick Cha,[4] Eung Bae Lee,[5] Tae In Park,[6] Chang Ho Kim,[4] Won Kee Lee,[2] Sin Kam,[2] Je-Yong Choi,[1] Young Mo Kang,[4] Rang-Woon Park,[1] In-San Kim,[1] Young Lae Cho,[7] Tae Hoon Jung,[4] Sung Beom Han,[8] and Jae Yong Park[1,3]

Departments of [1]Biochemistry and [2]Preventive Medicine, School of Medicine; [3]Cancer Research Institute; Departments of [4]Internal Medicine, [5]Thoracic Surgery, [6]Pathology, and [7]Obstetrics and Gynecology, Kyungpook National University Hospital; and [8]Department of Internal Medicine, School of Medicine, Keimyung University, Daegu, Korea

Abstract

The methyl-CpG binding domain 1 (MBD1) protein plays an important role for transcriptional regulation of gene expression. Polymorphisms and haplotypes of the *MBD1* gene may have an influence on MBD1 activity on gene expression profiles, thereby modulating an individual's susceptibility to lung cancer. To test this hypothesis, we investigated the association of *MBD1* −634G>A, −501delT (−501 T/T, T/−, −/−), and Pro^{401}Ala genotypes and their haplotypes with the risk of lung cancer in a Korean population. The *MBD1* genotype was determined in 432 lung cancer patients and in 432 healthy control subjects who were frequency matched for age and gender. The −634GG genotype was associated with a significantly increased risk of overall lung cancer compared with the −634AA genotype [adjusted odds ratio (OR), 3.10; 95% confidence interval (95% CI), 1.24-7.75; P = 0.016]. When analyses were stratified according to the tumor histology, the −634GG genotype was associated with a significantly increased risk of adenocarcinoma compared with the −634AA genotype (adjusted OR, 4.72; 95% CI, 1.61-13.82; P = 0.005). For the *MBD1* −501delT and Pro^{401}Ala polymorphisms, the −501 T/T genotype was associated with a marginal significantly increased risk of adenocarcinoma compared with the −501$^{-/-}$ genotype (adjusted OR, 2.07; 95% CI, 1.02-4.20; P = 0.045), and the Pro/Pro genotype was associated with a significantly increased risk of adenocarcinoma compared with the Ala/Ala genotype (adjusted OR, 3.41; 95% CI, 1.21-9.60; P = 0.02). Consistent with the genotyping analyses, the −634G/−501T/^{401}Pro haplotype was associated with a significantly increased risk of overall lung cancer and adenocarcinoma compared with the −634A/−501^{-}/^{401}Ala haplotype (adjusted OR, 1.44; 95% CI, 1.08-1.91; P = 0.012 and P_c = 0.048; adjusted OR, 1.75; 95% CI, 1.20-2.56; P = 0.004 and P_c = 0.016, respectively). On a promoter assay, the *−634A* allele had significantly higher promoter activity compared with the *−634G* allele in the Chinese hamster ovary cells and A549 cells (P < 0.05 and P < 0.001, respectively), but the −501delT polymorphism did not have an effect on the promoter activity. When comparing the promoter activity of the *MBD1* haplotypes, the −634A/−501^{-} haplotype had a significantly higher promoter activity than the −634G/−501T haplotype (P < 0.001). These results suggest that the *MBD1* −634G>A, −501delT, and Pro^{401}Ala polymorphisms and their haplotypes contribute to the genetic susceptibility for lung cancer and particularly for adenocarcinoma. (Cancer Epidemiol Biomarkers Prev 2005;14(11):2474–80)

Introduction

Although cigarette smoking is the major cause of lung cancer, only a fraction of smokers develop lung cancer during their lifetime, suggesting that genetic and epigenetic factors are of importance in determining an individual's susceptibility to lung cancer (1, 2).

DNA cytosine methylation in CpG dinucleotides is a major epigenetic mechanism that regulates chromosomal stability and gene expression (3, 4). Many human cancers, including lung cancer, have both global hypomethylation and regional hypermethylation of CpG islands (5-8). Such aberrant DNA methylation may contribute to carcinogenesis in several ways. Hypomethylation may lead to chromosomal instability, reactivation of transposable elements, and loss of imprinting (6, 9). Methylation of CpG sequences may facilitate C-to-T transition mutations in tumor suppressor genes and/or oncogenes through deamination of 5-methylcytosine to thymine (10). Methylated CpG sequences may also increase susceptibility to attack by some environmental carcinogens (11, 12). Finally, *de novo* hypermethylation of promoter CpG islands may lead to silencing of tumor suppressor genes and DNA repair genes (4, 6, 9).

Methylated CpG sites are recognized by a family of protein factors containing the methyl-CpG binding domain (MBD); to date, five family members have been identified in mammals: MeCP2, MBD1, MBD2, MBD3, and MBD4 (13-15). Four of these proteins (MeCP2, MBD1, MBD2, and MBD3) play important roles for methylation-mediated transcriptional silencing by recruiting chromatin-modifying factors, such as histone deacetylases, to the methylated promoters (14, 15). In contrast to the other family members, MBD4 protein has a thymine glycosylase activity and binds preferentially to 5mCpG-TpG mismatches, which are the primary products of deamination at methyl-CpG. Therefore, MBD4 protein is thought to function as a DNA repair enzyme to minimize mutation at 5-methylcytosine (16, 17).

MBD1 is known to act as a transcriptional repressor through the cooperation of MBD, cysteine-rich CXXC domains, and a COOH-terminal transcription repression domain (18, 19). Among the MBD family of proteins, MBD1 is characterized by two or three cysteine-rich CXXC domains that were originally found in DNA methyltransferase and human trithorax protein HRX (20). MBD1 has at least five isoforms that are the result of

Received 6/9/05; revised 8/16/05; accepted 9/1/05.

Grant support: National Cancer Control R&D Program 2003 and Ministry of Health and Welfare, Republic of Korea.

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.

Note: J-S. Jang and S.J. Lee contributed equally to this work.

Requests for reprints: Jae Yong Park, Department of Internal Medicine, School of Medicine, Kyungpook National University, Samduk 2Ga 50, Daegu, 700-412, Korea. Phone: 82-53-420-5536; Fax: 82-53-426-2046. E-mail: jaeyong@kyungpook.ac.kr

doi:10.1158/1055-9965.EPI-05-0423

alternative splicing within the regions of the CXXC domains and the COOH terminus. These MBD1 isoforms preferentially repress transcription from the methylated gene promoters, but the MBD1 isoforms containing three CXXC domains can also repress transcription from the unmethylated promoters, suggesting that MBD1 plays an important role for the establishment and maintenance of local chromatin states to regulate gene activities (18, 21). In addition to transcriptional gene regulation, a recent study suggests that MBD1 is also involved in DNA repair through its interaction with methylpurine DNA glycosylase, which removes the damaged purines produced by methylating or oxidative agents (22).

Single nucleotide polymorphisms are the most common form of human genetic variation and may contribute to the individual susceptibility for lung cancer. We previously showed that some variants in the DNA repair and DNA methyltransferase (DNMT) genes affect either the expression or the activities of enzymes and are therefore associated with the risk of lung cancer (23-27). Several candidate single nucleotide polymorphisms in the *MBD1* gene have been recently deposited in the public databases (http://www.ncbi.nlm.nih.gov/SNP). Although the functional effects of these polymorphisms have not been elucidated, we hypothesized that some of these variants, particularly their haplotypes, may influence *MBD1* activity on gene regulation and genome stability, thereby modulating the susceptibility to lung cancer. To test this hypothesis, a case-control study was conducted to evaluate the association between *MBD1* genotypes/haplotypes and the risk of lung cancer. Among the candidate single nucleotide polymorphisms in the *MBD1* gene, we have focused on the amino acid substitution variants ($Pro^{13}Leu$, 29295427G>A in exon 3 and $Pro^{401}Ala$, 29289281C>G in exon 12; Genbank accession no. NT_010966) and the variants in promoter regions (−870G>A, −750C>A, −634G>A, and −501delT; Genbank accession no. NT_010966), because these are the most likely to affect gene functioning. In the present study, we evaluated the association of −634G>A, −501delT (−501 T/T, T/−, −/−), and $Pro^{401}Ala$ genotypes and their haplotypes with lung cancer, because $Pro^{13}Leu$, −870G>A, and −750C>A polymorphisms were not detected in a preliminary study that consisted of 27 lung cancer patients and 27 healthy controls.

Materials and Methods

Study Population. This case-control study included 432 lung cancer patients and 432 healthy controls. The details of the study population have been described elsewhere (26-29). In brief, the eligible cases included all the patients who were newly diagnosed with primary lung cancer at Kyungpook National University Hospital, Daegu, Korea from January 2001 to February 2002. There were no age, gender, histologic, or stage restrictions, but those patients with a prior history of cancers were excluded from this study. The cases included 210 (48.6%) squamous cell carcinomas, 141 (32.6%) adenocarcinomas, 73 (16.9%) small cell carcinomas, and eight (1.9%) large cell carcinomas. The control subjects were randomly selected from a pool of healthy volunteers who had visited the general health check-up center at Kyungpook National University Hospital during the same period. The control subjects were frequency matched (1:1) to the cancer cases based on gender and age (±5 years). All the cases and the controls were ethnic Koreans and resided in Daegu City or in the surrounding regions. A detailed questionnaire was completed for each patient and each control by a trained interviewer. The questionnaire included information on the average number of cigarettes smoked daily and the number of years the subjects had been smoking. For the smoking status of the subjects, a person who had smoked at least once a day for >1 year during his or her lifetime was regarded as a smoker. A former smoker was defined as one who had stopped smoking at least 1 year before the lung cancer diagnosis (cases) or before the date signed on an informed consent (controls). The cumulative cigarette dose (pack-years) was calculated by using the following formula: pack-years = (packs per day) × (years smoked).

***MBD1* Genotyping.** Genomic DNA was extracted from peripheral blood lymphocytes by proteinase K digestion and phenol/chloroform extraction. The *MBD1* −634G>A, −501delT, and $Pro^{401}Ala$ genotypes were determined by using a PCR-RFLP assay. The PCR primers were designed based on the Genbank reference sequence (accession no. NT_010966). The PCR primers for the *MBD1* −634G>A, −501delT, and $Pro^{401}Ala$ polymorphisms were 5′-CCTCTGCCCGTGGGAGGCT-3′ (forward) and 5′-CCCTGCAGGAGGCGGAG(mutated T→G)CC-3′ (reverse); 5′-TTCCAGCCTCAACCTGAAGG-3′ (forward) and 5′-CTCAGTTTACCTGCGGTGTG-3′ (reverse); and 5′-AAGCAGATTCCGTGCTGGGA-3′ (forward) and 5′-CTTTCCACGACGGTAAGGTC(mutated G→C)-3′ (reverse), respectively. The PCR reactions were done in a total volume of 20 μL containing 100 ng genomic DNA, 25 pmol/L of each primer, 0.2 mmol/L deoxynucleotide triphosphates, 75 mmol/L Tris-HCl (pH 9.0), 15 mmol/L ammonium sulfate, 0.1 g/μL bovine serum albumin, 2.5 mmol/L $MgCl_2$, and 1 unit of Taq polymerase (Takara Shuzo Co., Otsu, Shiga, Japan). The PCR cycle conditions consisted of an initial denaturation step at 95°C for 5 minutes followed by 35 cycles of 30 seconds at 94°C; 30 seconds at 56°C for −634G>A, 59°C for −501delT, and 53°C for $Pro^{401}Ala$; 30 seconds at 72°C; and a final elongation at 72°C for 10 minutes. The PCR products were digested overnight at 37°C with the appropriate restriction enzymes (New England BioLabs, Beverly, MA). The restriction enzymes for the −634G>A, −501delT, and $Pro^{401}Ala$ genotypes were *Ban*II, *Ava*I, and *Ava*II, respectively. The digested PCR products were resolved on 6% acrylamide gels and stained with ethidium bromide for visualization under UV light. For quality control, the genotyping analysis was done ''blind'' with respect to the case/control status. About 10% of the samples were randomly selected to be genotyped again by a different author, and the results were 100% concordant. To confirm the genotyping results, selected PCR-amplified DNA samples (n = 2, respectively, for each genotype) were examined by DNA sequencing, and the results were also 100% concordant.

Promoter-Luciferase Constructs. To examine the potential effects of the −634G>A and −501delT polymorphisms on the *MBD1* transcription activity, the fragments, including the −634G>A (−872 to −576, transcription start site of exon 1 counted as +1) and the −501delT (−601 to −290), were synthesized, respectively, by PCR using the genomic DNA from donors that were carrying either the wild-type allele or the polymorphic allele of each polymorphism. The PCR primers used for the −634G>A and the −501delT promoter regions were 5′-GAAGCTGTC TCCACATTGCT-3′ (forward) and 5′-CACACCGCAGGTAAACTGA-3′ (reverse) and 5′-CGCGTGCCTCAGTTTACCT-3′ (forward) and 5′-CGCTTTCCAGCCTCAACCT-3′ (reverse), respectively. The PCR products were inserted upstream of the SV40 promoter in the pGL3-promoter plasmid (Promega, Madison, WI). For comparison of the haplotypes' promoter activities (haplotype −634G/−501T versus haplotype −634A/−501$^-$), the fragments of the *MBD1* promoter region (−994 to +113) were synthesized by PCR using the genomic DNA from donors carrying either the −634G/−501T haplotype or the −634A/−501$^-$ haplotype. The PCR primers for the *MBD1* promoter were 5′-GCCCATCTGCCTTTATCAGA-3′ (forward) and 5′-CAGCCCCAAGGCTGTCT-3′ (reverse). The PCR products were inserted upstream of the luciferase gene in the pGL3-basic plasmid (Promega). The correct sequence of all the clones was verified by DNA sequencing.

Transient Transfection and Luciferase Assay. The promoter activity was measured using the Dual-Luciferase Reporter Assay System (Promega) in Chinese hamster ovary cells and A549 lung cancer cells. The Chinese hamster ovary cells were cultured in minimal essential medium with 10% heat-inactivated fetal bovine serum, and the A549 cells were grown in RPMI 1640 that was supplemented with 10% heat-inactivated fetal bovine serum. Cells (1×10^5) were plated in a six-well plate the day before transfection so that the cells were ~60% confluent by the next day. The pRL-SV40 plasmid and the pGL3-basic plasmid with the synthesized fragments of the *MBD1* promoter region were cotransfected using Lipofectine reagent (Invitrogen, Carlsbad, CA). The pRL-SV40 vector that provided the constitutive expression of *Renilla* luciferase was used as an internal control to correct for the differences in transfection and harvesting efficiency. The cells were collected 48 hours after transfection, and the cell lysates were prepared according to Promega's instruction manual. Luciferase activity was measured using a Lumat LB953 luminometer (EG&G Berthhold, Bad Wildbad, Germany), and the results were normalized using the activity of *Renilla* luciferase. Independent triplicate experiments were done four times, and the results were reported as mean ± SD.

Statistical Analysis. The cases and controls were compared using Student's t test for continuous variables and the χ^2 test for categorical variables. Hardy-Weinberg equilibrium was tested for with a goodness-of-fit χ^2 test with one degree of freedom to compare the observed genotype frequencies among the subjects with the expected genotype frequencies. We examined the widely used measure of linkage disequilibrium between pairs of biallelic loci, Lewontin's D' ($|D'|$; ref. 30). The haplotypes and their frequencies were estimated based on a Bayesian algorithm using the Phase program (31), which is available at http://www.stat.washington.edu/stephens/phase.html. Unconditional logistic regression analysis was used to calculate the odds ratios (ORs) and 95% confidence intervals (95% CI), with adjustment for possible confounders (gender and family history of lung cancer as a nominal variable and age and pack-years as continuous variables). In addition to the overall association analysis, we did a stratified analysis by age, gender, smoking status, and tumor histology to further explore the association between *MBD1* genotypes/haplotypes and the risk of lung cancer in each stratum. To test which one of the three polymorphisms is more likely to be the main cause of the observed association, we compared seven different logistic regression models (each polymorphism alone, any two of the three polymorphisms, and then all the three polymorphisms together). When multiple comparisons are made, the corrected Ps (P_c) were also calculated for multiple testing by using Bonferroni's inequality method. All analyses were done using Statistical Analysis Software for Windows, version 8.12 (SAS Institute, Cary, NC).

Table 1. Characteristics of the study population

Variable	Cases (n = 432)	Controls (n = 432)
Age (y)	61.6 ± 9.0	60.9 ± 9.3
Sex		
Male	352 (81.5)*	352 (81.5)
Female	80 (18.5)	80 (18.5)
Smoking status†		
Current	317 (73.4)	229 (53.0)
Former	39 (9.0)	98 (22.7)
Never	76 (17.6)	105 (24.3)
Pack-years‡	39.9 ± 17.9	34.4 ± 17.6

*Column percentage in brackets.
†P = 0.001.
‡In current and former smokers (P < 0.001).

Results

The demographics of the cases and controls enrolled in this study are shown in Table 1. There were no significant differences between the cases and controls in mean age or gender distribution, suggesting that the matching based on these two variables was adequate. The case group had a higher prevalence of current smokers than the controls (P < 0.001), and the number of pack-years in smokers was significantly higher in the cases than in the controls (39.9 ± 17.9 versus 34.4 ± 17.6 pack-years; P < 0.001). These differences were controlled in the later multivariate analyses.

The genotype and polymorphic allele frequencies of the three *MBD1* polymorphisms (−634G>A, −501delT, and $Pro^{401}Ala$) among the controls and cases are shown in Table 2. The genotype distributions of the three polymorphisms among the controls were in Hardy-Weinberg equilibrium. The distribution of the *MBD1* −634G>A genotypes among the cases was significantly different from that among the controls (P = 0.04). When the cases were stratified by histologic type, the distribution of the −634G>A genotypes among the adenocarcinoma cases differed significantly from that among the controls (P = 0.001). The genotype distribution of the −501delT polymorphism among the cases was not significantly different from that among the controls. The distribution of the $Pro^{401}Ala$ genotypes among the cases was not significantly different from those among the controls, but the genotype distribution among the adenocarcinoma cases differed significantly from that among the controls (P = 0.02).

Table 3 shows the lung cancer risk related to the *MBD1* −634G>A, −501delT, and $Pro^{401}Ala$ genotypes, respectively. Adjusted ORs and their 95% CIs were calculated using the more common homozygous variant genotype as the reference group. The −634GG genotype was associated with a significantly increased risk of overall lung cancer compared with the −634AA genotype (adjusted OR, 3.10; 95% CI, 1.24-7.75; P = 0.016). When analyses were stratified according to the tumor histology, the risk associated with the −634G>A genotypes was more pronounced in patients with adenocarcinoma. The −634GG genotype was associated with a significantly increased risk of adenocarcinoma compared with the −634AA genotype (adjusted OR, 4.72; 95% CI, 1.61-13.82; P = 0.005). For the *MBD1* −501delT polymorphism, there was no significant association between this polymorphism and the risk of overall lung cancer. When the cases were categorized by tumor histology, however, the −501 T/T genotype was associated with a marginal significantly increased risk of adenocarcinoma compared with the $-501^{-/-}$ genotype (adjusted OR, 2.07; 95% CI, 1.02-4.20; P = 0.045). For the $Pro^{401}Ala$ polymorphism, compared with the Ala/Ala genotype, the Pro/Pro genotype was associated with a significantly increased risk of adenocarcinoma (adjusted OR, 3.41; 95% CI, 1.21-9.60; P = 0.02), and the Pro/Ala genotype was associated with a marginal significantly increased risk of adenocarcinoma (adjusted OR, 1.54; 95% CI, 1.00-2.34; P = 0.047).

The −634G>A and −501delT polymorphisms and the −501delT and $Pro^{401}Ala$ polymorphisms were in strong linkage disequilibrium ($|D'|$ = 0.927 and 0.929, respectively) in the study populations. The five common haplotypes accounted for 98.6% of the chromosome of the present study population. Three haplotypes that had a frequency of <1% were excluded [in the controls, 10 (1.2%) and in the cases, 15 (1.7%), respectively] from further analysis to avoid possible errors in either the genotyping or the estimation process. Table 4 shows the inferred haplotype distribution for the cases and controls, and the lung cancer risk related to the haplotypes. Consistent with the results of the genotyping analyses, the $-634G/-501T/^{401}Pro$ haplotype was associated with a significantly increased risk of overall lung cancer

Table 2. Genotype frequencies of *MBD1* polymorphisms in lung cancer cases and controls

Polymorphism	Variables	Genotype*			Polymorphic allele frequency
		1/1	1/2	2/2	
−634G>A	Controls	(1.6)†	118 (27.3)	307 (71.1)	0.847
	All cases	17 (3.9)	134 (31.0)	281 (65.0)‡	0.806
	Squamous cell carcinoma	6 (2.9)	62 (29.5)	142 (67.6)	0.824
	Adenocarcinoma	9 (6.4)	46 (32.6)	86 (61.0)§	0.773
	Large cell carcinoma	0 (0.0)	3 (37.5)	5 (62.5)	0.812
	Small cell carcinoma	2 (2.7)	22 (30.1)	49 (67.1)	0.822
−501delT	Controls	26 (6.0)	171 (39.6)	235 (54.4)	0.742
	All cases	33 (7.6)	186 (43.1)	213 (49.3)	0.708
	Squamous cell carcinoma	11 (5.2)	95 (45.2)	104 (49.5)	0.721
	Adenocarcinoma	16 (11.3)	55 (39.0)	70 (49.7)	0.691
	Large cell carcinoma	0 (0.0)	4 (50.0)	4 (50.0)	0.750
	Small cell carcinoma	6 (8.2)	32 (43.8)	35 (48.0)	0.699
$Pro^{401}Ala$ (C→G)	Controls	9 (2.1)	120 (27.8)	303 (70.1)	0.840
	All cases	15 (3.5)	138 (31.9)	279 (64.6)	0.806
	Squamous cell carcinoma	5 (2.4)	64 (30.5)	141 (67.1)	0.824
	Adenocarcinoma	8 (5.7)	49 (34.7)	84 (59.6)‖	0.770
	Large cell carcinoma	0 (0.0)	3 (37.5)	5 (62.5)	0.812
	Small cell carcinoma	2 (2.7)	22 (30.1)	49 (67.1)	0.822

*The wild-type allele (−634G, −501T, and ^{401}Pro) is denoted by 1 and the polymorphic allele is denoted by 2.
†Row percentage in brackets.
‡Difference from controls (P = 0.04).
§Difference from controls (P = 0.001).
‖Difference from controls (P = 0.02).

compared with the −634A/−501$^-$/^{401}Ala haplotype (adjusted OR, 1.44; 95% CI, 1.08-1.91; P = 0.012 and P_c = 0.048). When analyses were stratified by tumor histology, the −634G/−501T/^{401}Pro haplotype was associated with a significantly increased risk of adenocarcinoma compared with the −634A/−501$^-$/^{401}Ala haplotype (adjusted OR, 1.75; 95% CI, 1.20-2.56; P = 0.004 and P_c = 0.016).

As a consequence of the strong linkage disequilibrium among the −634G>A, −501delT, and $Pro^{401}Ala$ polymorphisms, it is not easy to determine which polymorphism is more likely to be the cause of the observed association. In an attempt to resolve this, we fitted logistic regression models in which we allowed for the effects of the three polymorphisms, individually and jointly. The models incorporating −501delT, $Pro^{401}Ala$, or −501delT and $Pro^{401}Ala$ into −634G>A did not fit significantly better than the model with −634G>A alone (all comparisons, P > 0.05). However, the model with −501delT alone or $Pro^{401}Ala$ alone fitted less well than any joint models with −634G>A (all comparisons, P< 0.05).

The association between the *MBD1* genotypes/haplotypes and the risk of adenocarcinoma was further examined after stratifying for age, gender, smoking status, and histologic subtypes of adenocarcinoma [adenocarcinoma with/without bronchioloalveolar histology (BAC)]. There was no clear evidence that age, gender, or smoking status modified the effect of the MBD1 genotypes/haplotypes on the risk of adenocarcinoma in the stratified analysis. Moreover, the stratified analysis on the histologic subtypes of adenocarcinoma did not show any difference between adenocarcinomas with and without BAC histology (data not shown).

We investigated the effects of the −634G>A and −501delT polymorphisms on the promoter activity of *MBD1* by luciferase assay. The promoter activity of the *−634A* allele was significantly higher (1.5-fold) compared with the *−634G* allele in the Chinese hamster ovary cells (P < 0.001; Fig. 1A), but the −501delT polymorphism did not have an effect on the promoter activity (Fig. 1B). Because the −634G>A and −501delT polymorphisms were in linkage disequilibrium, we also compared the transcription activity of the haplotypes (haplotype −634G/−501T versus haplotype −634A/−501$^-$). In the Chinese hamster ovary cells, the −634A/−501$^-$ haplotype increased transcription activity by 2.1-fold compared with the −634G/−501T haplotype (P < 0.001). Similar result was observed in the A549 cells (P < 0.01; Fig. 1C).

Table 3. Adjusted ORs (95% CIs) for lung cancer associated MBD1 genotypes

Polymorphism	Genotype	All cases	Squamous cell carcinoma	Adenocarcinoma	Small cell carcinoma
−634G>A	G/G	3.10 (1.24-7.75)*	2.13 (0.68-6.60)	4.72 (1.61-13.82)†	2.26 (0.44-11.55)
	G/A	1.26 (0.93-1.71)	1.13 (0.77-1.65)	1.46 (0.95-2.24)	1.28 (0.74-2.21)
	A/A	1.0	1.0	1.0	1.0
−501delT	T/T	1.47 (0.84-2.57)	0.92 (0.43-1.98)	2.07 (1.02-4.20)‡	1.59 (0.60-4.21)
	T/−	1.17 (0.86-1.56)	1.25 (0.88-1.78)	1.08 (0.71-1.63)	1.24 (0.73-2.11)
	−/−	1.0	1.0	1.0	1.0
$Pro^{401}Ala$ (C→G)	Pro/Pro	2.12 (0.89-5.05)	1.49 (0.47-4.71)	3.41 (1.21-9.60)§	1.89 (0.38-9.39)
	Pro/Ala	1.24 (0.92-1.68)	1.11 (0.76-1.62)	1.54 (1.00-2.34)‖	1.14 (0.65-1.99)
	Ala/Ala	1.0	1.0	1.0	1.0

NOTE: Adjusted for age, sex, smoking status, pack-years of smoking, and family history of lung cancer.
*P = 0.016.
†P = 0.005.
‡P = 0.045.
§P = 0.02.
‖P = 0.047.

Discussion

We investigated the potential association of three *MBD1* polymorphisms (−634G>A, −501delT, and Pro^{401}Ala) with the risk of lung cancer. In addition, we estimated the *MBD1* haplotypes of the three polymorphisms and compared their frequency distributions in the lung cancer cases and controls. Compared with the −634A/−501^{-}/^{401}Ala haplotype, the −634G/−501T/^{401}Pro haplotype was associated with an increased risk of lung cancer and especially adenocarcinoma. This finding suggests that the *MBD1* −634G>A, −501delT, and Pro^{401}Ala polymorphisms and their haplotypes could be used as a marker for genetic susceptibility to adenocarcinoma. Of three major histologic types of lung cancer, the proportion of adenocarcinoma is increasing worldwide. Thus, identification of genetic factors that are responsible for the susceptibility to adenocarcinoma is indispensable for establishing novel and efficient ways of preventing the disease. This is the first case-control study of *MBD1* polymorphisms and haplotypes to examine their relation to lung cancer.

In the present study, carriers with the −634G/−501T/^{401}Pro haplotype were at increased risk of lung cancer compared with individuals having the −634A/−501^{-}/^{401}Ala haplotype. To determine whether the association between the *MBD1* polymorphisms and the risk of lung cancer is due to differences in the transcriptional activity of *MBD1* promoter, we compared the promoter activity of the wild-type allele or the polymorphic allele of these two polymorphisms by luciferase assay. *In vitro* promoter assay revealed that the *−634A* allele had a significantly higher transcriptional activity than the *−634G* allele, and the −501delT polymorphism did not have an effect on the transcriptional activity of the *MBD1* promoter. When we compared the promoter activity of the *MBD1* haplotypes, the −634A/−501^{-} haplotype had significantly higher transcriptional activity than the −634G/−501T haplotype. These results suggest that the genetic effects of *MBD1* polymorphisms on the risk of lung cancer can be mainly attributed to the −634G>A polymorphism, and these findings also suggest that the −634G>A change influences MBD1 expression, thus contributing to the genetic susceptibility to lung cancer. The mechanism by which the *MBD1* −634A/−501^{-} haplotype leads to higher promoter activity is currently unknown. Analysis of the potential transcription factor binding sites by the Alibaba2 program (32) showed that the −634G>A change leads to the creation of an cytoplasmic polyadenylation element binding site and eliminates an activator protein 2 α binding site and thyroid hormone receptor-α binding site, whereas the −501delT polymorphism has no effects on the transcription binding site of the *MBD1* promoter. Therefore, it is possible that the predicted changes in the putative transcription factor binding sites due to the −634G>A change may lead to enhanced promoter activity. However, this hypothesis has to be verified in future studies.

MBD1 has been thought to be involved in mediating methylation-associated gene silencing in various human cancers (18, 19). Therefore, it is possible that up-regulation of MBD1 expression may lead to a predisposition towards silencing of the methylated genes, thereby increasing the susceptibility to lung cancer. In contrast to this hypothesis, however, we observed in the present study that the *MBD1* genotype/haplotype with lower transcriptional activity was associated with an increased risk of lung cancer. In fact, this finding is supported by several previous reports (33-36) showing that MBD proteins were down-regulated in a variety of human cancers. Although the mechanism by which down-regulation of MBD1 expression leads to the enhanced susceptibility to lung cancer remains to be elucidated, this may happen because a decrease of MBD1 expression leads to a disturbance in the normal cellular homeostasis of the gene expression profiles; thus, it allows tumorigenesis via epigenetic instability (33–36). Another possible explanation is that MBD1 may function as a DNA repair system that is associated with methyl-CpG dinucleotides (22), as is the case of MBD4, thereby influencing the susceptibility to lung cancer.

Another interesting finding of the present study is that the *MBD1* polymorphisms, and their haplotypes had a more pronounced association with adenocarcinoma. Although the reason for the observed histology-specific difference in the risk conferred by the *MBD1* polymorphisms remains to be elucidated, this difference may be attributable to the differences in the carcinogenic pathways among the histologic types of lung cancer. Therefore, certain genotypes could confer a greater susceptibility to a particular histologic type of lung cancer (37-39). In our previous study (26), we have observed a similar finding that *DNMT3B* polymorphisms and their haplotypes were significantly associated with only the risk of adenocarcinoma. These findings of our previous and present studies suggest that alterations of the DNA methylation machinery, including DNMT3B and MBD1, may have a pronounced association with development of adenocarcinoma.

Several recent studies have shown that mutations in the kinase domain of *epidermal growth factor receptor* gene, like the *K-ras* mutations, frequently target adenocarcinoma but are more frequent in never smokers, females, adenocarcinomas

Table 4. Distribution of MBD1 haplotypes, predicted by Bayesian algorithm, in controls and cases

Haplotype*	Controls (n = 854)† n (%)	All cases (n = 849)†		Histologic type of lung cancer‡					
				Squamous cell carcinoma (n = 414)†		Adenocarcinoma (n = 275)†		Small cell carcinoma (n = 145)†	
		n (%)	OR§ (95% CI)	n (%)	OR§ (95% CI)	n (%)	OR§ (95% CI)	n (%)	OR§ (95% CI)
222	631 (73.9)	597 (70.3)	1.0	297 (71.7)	1.0	188 (68.4)	1.0	101 (69.7)	1.0
111	104 (12.2)	137 (16.1)	1.44 (1.08-1.91)‖	60 (14.5)	1.25 (0.87-1.79)	53 (19.3)	1.75 (1.20-2.56)¶	22 (15.2)	1.36 (0.81-2.28)
212	70 (8.2)	68 (8.0)	1.00 (0.70-1.43)	35 (8.5)	1.04 (0.67-1.62)	18 (6.5)	0.82 (0.47-1.44)	14 (9.7)	1.18 (0.63-2.20)
211	25 (2.9)	26 (3.1)	1.03 (0.58-1.83)	13 (3.1)	1.03 (0.50-2.13)	9 (3.3)	1.19 (0.54-2.66)	3 (2.1)	0.74 (0.21-2.60)
112	24 (2.8)	21 (2.5)	0.98 (0.54-1.81)	9 (2.2)	0.83 (0.37-1.86)	7 (2.5)	1.03 (0.43-2.50)	5 (3.4)	1.43 (0.53-3.89)

The order of polymorphisms for the haplotypes is as follows: −634G>A, −501delT, and Pro^{401}Ala. The wild-type allele (−634G*, *−501T*, and *^{401}Pro*) is denoted by 1 and the polymorphic allele is denoted by 2.

†Three haplotypes that had a frequency of <1 % were excluded from analysis; controls 10 and cases 15 (squamous cell carcinoma, n = 6; adenocarcinoma, n = 7; large cell carcinoma, n = 1; and small cell carcinoma, n = 1), respectively.

‡Eight large cell carcinoma cases were excluded from analysis.

§Adjusted for age, sex, pack-years of smoking, and family history of lung cancer.

‖P = 0.012 and P_c = 0.048 (Bonferroni corrected P).

¶P = 0.004 and P_c = 0.016.

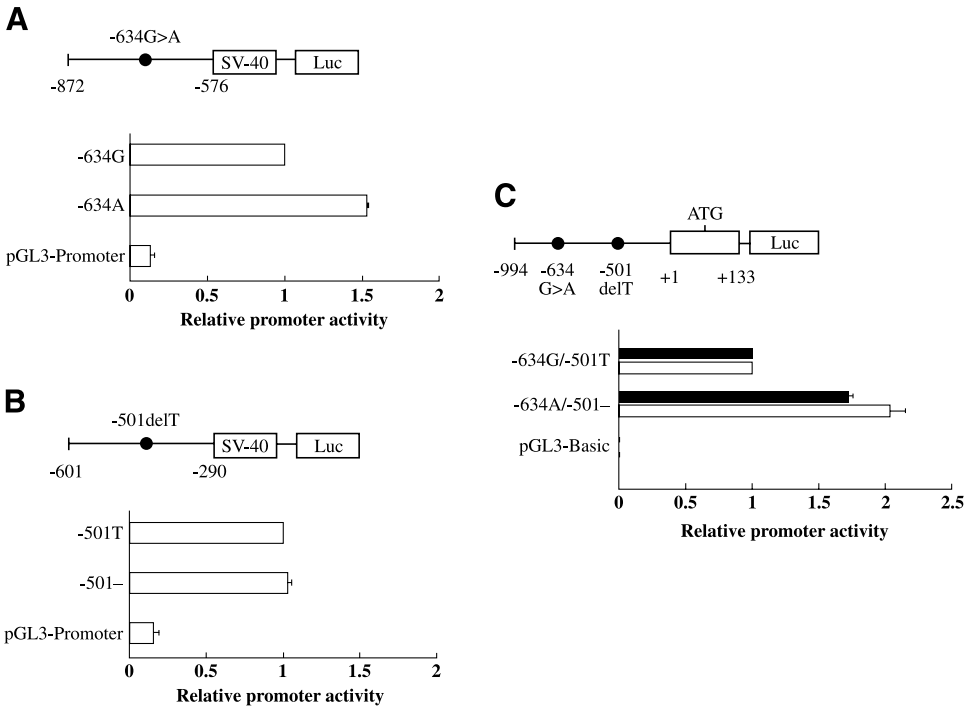

Figure 1. Transcription activity analyses of *MBD1* −634G>A and −501delT polymorphisms (*A* and *B*) and their haplotypes (*C*, haplotype −634G/−501T versus haplotype −634A/−501$^-$). The transcription activity was measured using the Dual-Luciferase Reporter Assay System in Chinese hamster ovary cells (*white columns*) and A549 cells (*black columns*). *Columns,* means from four independent experiments done in triplicate; *bars,* SD. The −634A construct increased the promoter activity by 1.5-fold compared with the −634G construct ($P < 0.001$). The −634A/−501$^-$ haplotype generated a significantly higher promoter activity compared with the −634G/−501T haplotype in Chinese hamster ovary cells and A549 cells ($P < 0.001$ and $P < 0.01$, respectively). *S,* SV40 promoter; *Luc,* luciferase.

with nonmucinous BAC histology, and the East Asian populations, whereas the *K-ras* mutations are more frequent in smokers, males, adenocarcinomas with mucinous BAC histology, and the Western populations (40, 41). These observations suggest that adenocarcinomas arising in the never smokers and smokers may be caused by different etiologies, not only in relation to environmental risk factors but also in relation to genetic susceptibility factors (40-42). Therefore, we did a stratification analysis to examine if the *MBD1* genotypes/haplotypes may have differential effects on the risk of adenocarcinoma according to age, gender, and smoking status and the histologic subtypes of adenocarcinoma (adenocarcinoma with/without BAC histology). In the current study, no risk modification was found with regard to age, gender, and smoking status. Moreover, the stratified analysis on the histologic subtypes of adenocarcinoma did not show any difference between adenocarcinomas with and without BAC histology (data not shown). However, because the number of subjects in the subgroups was small, our findings from the stratified analyses should be interpreted with caution before being confirmed in larger studies.

In the current study, the *MBD1* Pro^{13}Leu, −870G>A, and −750C>A polymorphisms were not detected in the preliminary study that included 27 healthy controls. These samples included 54 chromosomes, which provides at least a 95% confident level to detect alleles with frequencies >5%. Thus, it is very likely that if these polymorphisms exist, they may not play a major role in the genetic susceptibility to lung cancer in the Korean population (43, 44).

In conclusion, we found that the *MBD1* −634G>A, −501delT, and Pro^{401}Ala polymorphisms and their haplotypes were significantly associated with the risk of lung cancer and particularly adenocarcinoma. These results suggest that the *MBD1* gene may be involved in the development of lung cancer, although additional studies having larger sample sizes are required to confirm our findings. Future studies on the other *MBD1* sequence variants and their biological function are also needed to understand the role of the *MBD1* polymorphisms in determining the risk of lung cancer. Moreover, because genetic polymorphisms often vary between different ethnic groups, further studies are needed to clarify the association of the *MBD1* polymorphism with lung cancer in diverse ethnic populations.

References

1. Spitz MR, Wei Q, Dong Q, Amos CI, Wu X. Genetic susceptibility to lung cancer: the role of DNA damage and repair. Cancer Epidemiol Biomarkers Prev 2003;12:689–98.
2. Moore LE, Huang W-Y, Chung J, Hayes RB. Epidemiologic consideration to assess altered DNA methylation from environmental exposure to cancer. Ann N Y Acad Sci 2003;983:181–96.
3. Jones PA, Laird PW. Cancer epigenetics comes of age. Nat Genet 1999;21: 163–7.
4. Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet 2000;16:168–74.
5. Gama-Sosa MA, Slagel VA, Trewyn RW, et al. The 5-methylcytosine content of DNA from human tumors. Nucleic Acids Res 1983;11:6883–94.
6. Robertson KD. DNA methylation, methyltransferases, and cancer. Oncogene 2001;20:3139–55.
7. Woodson K, Mason J, Choi S-W, et al. Hypomethylation of p53 in peripheral blood DNA is associated with the development of lung cancer. Cancer Epidemiol Biomarkers Prev 2001;10:69–74.
8. Zochbauer-Muller S, Fong KM, Virmani AK, Geradts J, Gazdar AF, Minna JD. Aberrant promoter methylation of multiple genes in non-small cell lung cancer. Cancer Res 2001;61:249–55.
9. Esteller M, Herman JG. Cancer as an epigenetic disease: DNA methylation and chromatic alterations in human cancers. J Pathol 2002; 196:1–7.
10. Gonzalgo ML, Jones PA. Mutagenic and epigenetic effects of DNA methylation. Mutat Res 1997;386:107–18.
11. Denissenko MF, Chen JX, Tang M-S, Pfeifer GP. Cytosine methylation determines hot spots of DNA damage in the human *P53* gene. Proc Natl Acad Sci U S A 1997;94:3893–8.
12. Yoon JH, Smith LE, Feng Z, Tang M-S, Lee C-S, Pfeifer GP. Methylated CpG dinucleotides are the preferential targets for G-to-T transversion mutations induced by benzo[a]pyrene diol epoxide in mammalian cells: similarities with the p53 mutation spectrum in smoking-associated lung cancers. Cancer Res 2001;61:7110–7.
13. Hendrich B, Bird A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. Mol Cell Biol 1998;18:6538–49.
14. Wade PA. Methyl CpG-binding proteins: coupling chromatin architecture to gene regulation. Oncogene 2001;20:3166–73.
15. Ballestar E, Wolffe AP. Methyl CpG binding proteins: targeting specific gene repression. Eur J Biochem 2001;268:1–6.
16. Hendrich B, Hardeland U, Ng HH, Jiricny J, Bird A. The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites. Nature 1999;401:301–4.
17. Petronzelli F, Riccio A, Markham GD, et al. Biphasic kinetics of the human DNA repair protein MED1 (MBD4), a mismatch-specific DNA *N*-glycosylase. J Biol Chem 2000;275:32422–9.
18. Fujita N, Shimotake N, Ohki I, et al. Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. Mol Cell Biol 2000;20: 5107–18.
19. Fujita N, Watanabe S, Ichimura T, et al. MCAF mediates MBD1-dependent transcriptional repression. Mol Cell Biol 2003;23:2834–43.

20. Cross SH, Meehan RR, Nan X, Bird A. A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. Nat Genet 1997;16:256–9.
21. Fujita N, Takebayashi SI, Okumura K, et al. Methylation-mediated transcriptional silencing in euchromatin by methyl-CpG binding protein MBD1 isoforms. Mol Cell Biol 1999;19:6415–26.
22. Watanabe S, Ichimura T, Fujita N, et al. Methylated DNA-binding domain 1 and methylpurine-DNA glycosylase link transcriptional repression and DNA repair in chromatin. Proc Natl Acad Sci U S A 2003;100:12859–64.
23. Park JY, Lee SY, Jeon H-S, et al. Polymorphism of the DNA repair gene *XRCC1* and risk of primary lung cancer. Cancer Epidemiol Biomarkers Prev 2002;11:23–7.
24. Park JY, Park SH, Choi JE, et al. Polymorphism of the DNA repair gene *XPA* and risk of primary lung cancer. Cancer Epidemiol Biomarkers Prev 2002;11:993–7.
25. Park SH, Lee GY, Jeon H-S, et al. −93G>A polymorphism of *hMLH1* and risk of primary lung cancer. Int J Cancer 2004;112:678–82.
26. Lee SJ, Jeon H-S, Jang J-S, et al. *DNMT3B* polymorphisms and risk of primary lung cancer. Carcinogenesis 2005;26:403–9.
27. Lee GY, Jang J-S, Lee SY, et al. *XPC* polymorphisms and lung cancer risk. Int J Cancer 2005;115:807–13.
28. Lee SJ, Lee SY, Jeon H-S, et al. *Vascular endothelial growth factor* gene polymorphisms and risk of primary lung cancer. Cancer Epidemiol Biomarkers Prev 2005;14:571–5.
29. Park JM, Lee GY, Choi JE, et al. No association between polymorphisms in the histone deacetylase genes and the risk of lung cancer. Cancer Epidemiol Biomarkers Prev 2005;14:1841–3.
30. Lewontin RC. On measures of gametic disequilibrium. Genetics 1998;120: 849–52.
31. Stephens M, Smith MJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 2001;68:978–89.
32. Grabe N. Alibaba2: context specific identification of transcription factor binding sites. In Silico Biol 2002;2:S1–1.
33. Sato M, Horio Y, Sekido Y, Minna JD, Shimokata K, Hasegawa Y. The expression of DNA methyltransferases and methyl-CpG-binding proteins is not associated with the methylation status of p14, p16 and RASSF1A in human lung cancer cell lines. Oncogene 2002;21:4822–9.
34. Muller-Tidow C, Kugler K, Diederichs S, et al. Loss of expression of HDAC-recruiting methyl-CpG-binding domain proteins in human cancers. Br J Cancer 2001;85:1168–74.
35. Prokhortchouk E, Hendrich B. Methyl-CpG binding proteins and cancer: are MeCpGs more important than MBDs? Oncogene 2002;21:5394–9.
36. Patra SK, Patra A, Zhao H, Carroll P, Dahiya R. Methyl-CpG-DNA binding proteins in human prostate cancer: expression of CXXC sequence containing MBD1 and repression of MBD2 and MeCP2. Biochem Biophys Res Commun 2003;302:759–66.
37. Gu J, Spitz MR, Yang F, Wu X. Ethnic differences in poly(ADP-ribose) polymerase pseudogene genotype distribution and association with lung cancer risk. Carcinogenesis 1999;20:1465–9.
38. Sugimura H, Kohno T, Wakai K, et al. *hOGG1* Ser^{326}Cys polymorphism and lung cancer susceptibility. Cancer Epidemiol Biomarkers Prev 1999;8: 669–74.
39. Liu G, Miller DP, Zhou W, et al. Differential association of the codon 72 p53 and GSTM1 polymorphisms on histological subtype of non-small cell lung cancer. Cancer Res 2001;61:8718–22.
40. Marchetti A, Martella C, Felicioni L, et al. *EGFR* mutations in non-small cell lung cancer: analysis of a large series of cases and development of a rapid and sensitive method for diagnostic screening with potential implications on pharmacologic treatment. J Clin Oncol 2005;23:857–65.
41. Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 2005;97:339–46.
42. Gazdar AF, Shigematsu H, Herz J, Minna JD. Mutations and addiction to *EGFR*: the Achilles 'heal' of lung cancers? Trends Mol Med 2004;10:481–6.
43. Risch NJ. Searching for genetic determinants in the new millennium. Nature 2000;405:847–56.
44. Rebbeck TR, Ambrosone CB, Bell DA, et al. SNPs, haplotypes, and cancer: applications in molecular epidemiology. Cancer Epidemiol Biomarkers Prev 2004;13:681–7.