Biomarkers, 2008; 13(6): 607–617

informa
healthcare

Methyl-CpG binding domain 1 gene polymorphisms and lung cancer risk in a Chinese population

HONGLIANG LIU[1]*, GUANGFU JIN[2]*, HAIFENG WANG[3], WENTING WU[1], YANHONG LIU[1], JI QIAN[1], WEIWEI FAN[1], HONGXIA MA[2], RUIFEN MIAO[2], ZHIBIN HU[2], WEIWEI SUN[3], YI WANG[3], LI JIN[1,3], QINGYI WEI[4], HONGBING SHEN[2], WEI HUANG[3], & DARU LU[1]

[1]*State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200433, China,* [2]*Department of Epidemiology and Biostatistics, Cancer Research Center of Nanjing Medical University, Nanjing 210029, China,* [3]*Department of Genetics, Chinese National Human Genome Center at Shanghai, Shanghai 201203, China and* [4]*Department of Epidemiology, the University of Texas M.D. Anderson Cancer Center, Houston, TX 77030, USA*

Abstract

Polymorphisms of the methyl-CpG binding domain 1 (*MBD1)* gene may influence MBD1 activity on gene expression profiles, thereby modulating individual susceptibility to lung cancer. To test this hypothesis, we investigated the associations of four *MBD1* polymorphisms and lung cancer risk in a Chinese population. Single locus analysis revealed significant associations between two polymorphisms (rs125555 and rs140689) and lung cancer risk ($p = 0.011$ and $p = 0.005$, respectively). Since the two polymorphisms were in linkage disequilibrium, further haplotype analyses were performed and revealed a significant association with lung cancer (global test p-value $= 0.0041$). Our results suggested that *MBD1* polymorphisms might be involved in the development of lung cancer. Validation of these findings in larger studies of other populations is needed.

Keywords: *Lung cancer, susceptibility,* MBD1*, polymorphism, haplotype*

(Received 1 September 2007; accepted 18 April 2008)

Introduction

Lung cancer is the leading cause of cancer-related death for men and women in the world, and there were an estimated 1.35 million new cases worldwide in 2002 (Parkin et al. 2005). Although overwhelming epidemiological evidence exists that smoking is the primary risk factor for lung cancer, <20% of lifetime smokers develop lung cancer, suggesting that genetic and epigenetic factors are of importance

Correspondence: Daru Lu, State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200433, China. Tel: +86 21 65642799. Fax: +86 21 65642799. E-mail: drlu@fudan.edu.cn or huangwei@chgc.sh.cn

*These authors contribute equally to this work.

ISSN 1354-750X print/ISSN 1366-5804 online
DOI: 10.1080/13547500802168031

in determining an individual's susceptibility to lung cancer (Moore et al. 2003, Spitz et al. 2003).

DNA methylation at position 5 of cytosine in CpG dinucleotides is the major epigenetic modification of mammalian genomes and is required for transcription, DNA replication, repair and genome stability (Robertson & Jones 2000). Hypermethylation of CpG islands may be implicated in tumorigenesis, acting as a mechanism to inactivate specific gene expression of a diverse array of genes (Rountree et al. 2000). Genes that have been reported to be regulated by *CpG* hypermethylation, include tumour suppressor genes, cell cycle related genes, DNA mismatch repair genes, hormone receptors and tissue or cell adhesion molecules (Yan et al. 2001). For example, tumor-specific deficiencies in the expression of the DNA repair genes MLH1(Yu et al., 2004) and MGMT (Danam et al., 2005), and the tumor suppressors p16, CDKN2 and MTS1, have been directly correlated to hypermethylation (Merlo et al., 1995; Yoon et al., 2003).

The effect of DNA methylation is due in part to structural alteration of DNA, which prevents some transcription factors from binding to their cognate sequences (Robertson 2002). In addition, DNA methylation affects chromatin structure due to recruitment of co-repressors and chromatin-modifying activities by proteins that bind specifically to methylated DNA. Most of the methyl-DNA binding proteins described so far belong to the methyl-CpG binding domain (MBD) family, defined by the methyl-CpG binding domain (Wade 2001). To date, five family members have been identified in mammals: MeCP2, MBD1, MBD2, MBD3 and MBD4.

Uniquely among MBD proteins, MBD1 has a methyl-CpG binding domain, two or three cysteine-rich CXXC motifs, and the C-terminal transcriptional repression domain (TRD). MBD1 has at least five isoforms that are the result of 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 (Nakao et al. 2001). Recent studies have demonstrated that MBD1 participated in the transcriptional silencing of *MGMT, MLH1* and *p53BP2* and so on (Yu et al. 2004, Danam et al. 2005, Lyst et al. 2006). In addition, previous study has suggested that MBD1 might also be involved in DNA repair through its interaction with methylpurine DNA glycosylase, which removes the damaged purines produced by methylating or oxidative agents (Shoda et al. 2003). Mutation analyses also revealed that a number of mutations in the *MBD1* gene exist in colon and lung cancer cell lines (Bader et al. 2003). All of these studies support the evidence that *MBD1* contributes to the risk of cancer.

Single nucleotide polymorphisms (SNPs) are the most common forms of human genetic variation and may contribute to individual susceptibility to lung cancer. A previous case–control study in a Korean population showed that the polymorphisms in *MBD1* may modulate an individual's susceptibility to lung cancer (Jang et al. 2005). However this has not been confirmed in any other population. Therefore we conducted a case–control study to evaluate the association between *MBD1* genotypes/haplotypes and the risk of lung cancer in a Chinese population.

Material and methods

Study populations

The study population and subject characteristics have been described previously (Hu et al. 2007). In brief, this was a hospital-based case–control study including 500 lung cancer patients and 517 cancer-free controls. All the subjects were genetically unrelated ethnic Han Chinese. Patients, who were newly diagnosed with incident lung cancer according to the National Diagnosis Standard for Lung Cancer, were consecutively recruited between July 2002 and November 2004 from the Cancer Hospital of Jiangsu Province, the First Affiliated Hospital of Nanjing Medical University, without restrictions of age, gender and histology. Those with a history of previous cancer, metastasized cancer, and previous radiotherapy or chemotherapy were excluded. The response rate for cases was 90.5%. Cancer-free controls were randomly selected from 10 500 individuals who participated in a community-based screening programme for non-infectious diseases conducted in Jiangsu Province during the same time period when the cases were recruited, with a response rate of 83.8%. These control subjects had no history of cancer and were frequency-matched to the cases by age ($\pm$5 years), gender and residential area (urban or rural areas). Each participant was scheduled for an interview after written informed consent was obtained, and a structured questionnaire was administered by interviewers to collect information on demographic data and environmental exposure history including tobacco smoking. Those who had smoked less than one cigarette per day and for less than 1 year in their lifetime were defined as non-smokers, otherwise, they were considered as ever smokers. Those smokers who quit for more than 1 year were considered as former smokers. Pack-years smoked ((cigarettes per day $\div$20) $\times$ years smoked) were calculated to indicate the cumulative smoking dose. Family history of cancer was defined as any self-reported cancer in first-degree relatives (parents, siblings or children). After the interview, approximately 5-ml venous blood sample was collected from each participant. The study was approved by the Institutional Review Board of Fudan University.

Genotyping assays

SNPs were selected from the NCBI SNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/) based on: (1) polymorphism density in genes, (2) predicted function and genomic context, (3) minor allele frequency, (4) quality of validation evidence and (5) compatibility with the genotyping platform. A greedy algorithm was used during the selection process. The more detailed description about the selection method can be found in our previous articles (Hu et al. 2006, 2007). A total of four SNPs in *MBD1* were selected, including one non-synonymous polymorphism (rs125555, Pro401Ala) and three intronic variants (rs2851716, rs140689, rs999199).

SNPs were genotyped using the Illumina Genotyping Facility which is a highly efficient genotyping assay combining the Illumina Golden GateTM assay, SentrixTM array matrices and SherlockTM scanner technology (Illumina Corp, Foster City, CA, USA), at the Chinese National Human Genome Center in Shanghai, China. More detailed descriptions of each step performed by the Illumina Genotyping facility are available in our previous paper (Hu et al. 2007) and the HapMap website (http://www.hapmap.org/downloads/protocols_overview.html).

Statistical analyses

The cases and controls were compared using the Student's *t*-test for continuous variables and the χ^2 test for categorical variables. The Hardy–Weinberg equilibrium test was done for each SNP among controls using Fisher's probability test statistic, as implemented in the software package SNPassoc (Gonzalez et al. 2007). The single locus association analysis was estimated by computing the odds ratios (OR) and 95% confidence intervals (CI) with adjustment for age, gender square root of pack-years of smoking and family cancer lung history (fmc) using SNPassoc under five genetic models (co-dominant, dominant, recessive, overdominant and log-additive) (Iniesta et al. 2005). We also evaluated the interactions of SNPs with gender (male, female), dichotomized age ($\leq$60 years, $>$60 years), fmc and trichotomized cumulative smoking doses (non-smokers, light smokers: $\leq$30 pack-years, heavy smokers: $>$30 pack-years) under the co-dominant and dominant genetic model. The issue of multiple-test was controlled by using 10 000-fold permutation test.

Haploview (Atanassov et al. 2005) and JLIN (Carter et al. 2006) were used to calculate the two indicators of linkage disequilibrium: the D' index of Lewontin (Lewontin 1988) and the r^2 (square of correlation coefficient) index of Hill and Robertson (Bellacosa et al. 1999). The haplotypes/diplotypes and their frequencies were estimated based on a Bayesian algorithm using the Phase program (Stephens & Donnelly, 2003), which is available at http://www.stat.washington.edu/stephens/phase.html.

All the statistical analyses were done with R 2.4.1 package (http://CRAN.R-project.org/) if not mentioned specially.

Results

The distributions of selected characteristics between lung cancer patients and controls have been described elsewhere (Hu et al. 2007). Overall, cases and controls were matched well for age and gender ($p=0.661$ for age, and $p=1.000$ for gender). Smoking and self-reported family history of cancer in first-degree relatives were significant risk factors for lung cancer. Of the 500 cancer patients, 229 had adenocarcinoma, 141 had squamous cell carcinoma, 34 had small cell carcinoma and 96 had large cell, mixed cell, or undifferentiated carcinomas.

The genotype distributions of the four polymorphisms among the controls were in Hardy–Weinberg equilibrium ($p>0.05$). The critical values of multiple-test were 0.0164 for the dominant model and 0.0176 for the log-additive model, individually, which were calculated using the 10 000-fold permutation test. Single locus analysis revealed two SNPs significantly associated with lung cancer (Table I). A decreased risk of overall lung cancer was found for the variant carriers versus the homozygous wild-type of rs125555 (OR 0.71, 95% CI 0.54–0.93, $p=0.011$). Similar protective effects were also observed among patients with non-small cell lung cancer (NSCLC) but not small cell lung cancer (SCLC). As for the polymorphism rs140689, an increased risk was observed for genotype T/A (OR 1.41, 95% CI 1.02–1.96) and T/T (OR 2.99, 95% CI 1.03–8.68) compared with the wild-type A/A ($p_{\text{trend}}=0.005$ for log-additive model) in patients with NSCLC. Due to complete linkage disequilibrium ($r^2=1.0$) with rs125555, SNP rs2851716 showed similar protective effects on lung cancer risk.

Stratified analyses were performed according to gender, age, fmc and pack-years. Only the significant results are shown in Table II. A significant interaction was found

Table I. Adjusted OR (95% CI) for lung cancer associated with *MBD1* polymorphisms.

			All case		NSCLC		SCLC	
Polymorphism	Genotype	Control (%)	n (%)	OR[a] (95% CI)	n (%)	OR[a] (95% CI)	n (%)	OR[a] (95% CI)
Rs2851716	TT	311 (60.2)	332 (66.4)	1.00 (Ref)	308 (66.1)	1.00 (Ref)	24 (70.6)	1.00 (Ref)
	TC	185 (35.8)	148 (29.6)	0.72 (0.55–1.96)	138 (29.6)	0.74 (0.56–0.98)	10 (29.4)	0.73 (0.34–1.60)
	CC	21 (4.1)	20 (4.0)	0.77 (0.40–1.48)	20 (4.3)	0.84 (0.43–1.61)	–	–
	TC+CC	206 (39.8)	168 (33.6)	0.73 (0.56–0.95)[b]	158 (33.9)	0.75 (0.57–0.99)	10 (29.4)	0.64 (0.30–1.39)
rs125555	G/G	311 (60.2)	333 (66.9)	1.00 (Ref)	309 (66.6)	1.00 (Ref)	24 (70.6)	1.00 (Ref)
	C/G	185 (35.8)	147 (29.5)	0.71 (0.54–0.94)	137 (29.5)	0.73 (0.55–0.97)	10 (29.4)	0.73 (0.34–1.60)
	C/C	21 (4.1)	18 (3.6)	0.67 (0.34–1.33)	18 (3.9)	0.73 (0.37–1.43)	–	–
	C/G+C/C	206 (39.8)	165 (33.1)	0.71 (0.54–0.93)[c]	155 (33.4)	0.73 (0.56–0.96)	10 (29.4)	0.64 (0.30–1.39)
rs140689	A/A	423 (81.8)	375 (75.0)	1.00 (Ref)	345 (74.0)	1.00 (Ref)	30 (88.2)	1.00 (Ref)
	T/A	89 (17.2)	112 (22.4)	1.37 (0.99–1.89)	108 (23.2)	1.41 (1.02–1.96)	4 (11.8)	0.57 (0.19–1.68)
	T/T	5 (1.0)	13 (2.6)	2.70 (0.93–7.89)[d]	13 (2.8)	2.99 (1.03–8.68)[e]	–	–
	T/A+T/T	94 (18.2)	125 (25.0)	1.44 (1.05–1.98)	121 (26.0)	1.50 (1.09–2.05)	4 (11.8)	0.54 (0.18–1.60)
rs999199	CC	493 (95.4)	482 (96.4)	1.00 (Ref)	449 (96.4)	1.00 (Ref)	33 (97.1)	1.00 (Ref)
	AC	24 (4.6)	18 (3.6)	0.79 (0.41–1.51)	17 (3.6)	0.84 (0.44–1.63)	1 (2.9)	0.68 (0.09–5.3)
	AA	–	–	–	–	–	–	–

[a]OR (95% CI) were adjusted for age, gender, family history of cancer (fmc), smoking status and pack-years; [b]$p=0.020$ for dominant model; [c]$p=0.012$ for dominant model; [d]$p=0.011$ for log-additive model; [e]$p=0.005$ for log-additive model.
NSCLC, non-small cell lung cancer; SCLC, small cell lung cancer.

only between the non-synonymous polymorphism (rs125555, Pro401Ala) and pack-years ($p_{trend} = 0.03996$), while the other two SNPs (rs2851716 and rs140689) showed a marginally significant interaction with pack-years ($p_{trend} = 0.051796$ and $p_{trend} = 0.053358$, respectively). The protective or risk effects mainly exist among non-smokers and light smokers.

Figure 1 shows that a strong linkage disequilibrium exists among rs125555, rs140689 and rs999199 ($D' > 0.8$), and SNP rs285176 showed complete linkage disequilibrium ($r^2 = 1.0$) with rs125555. Considering the minor allele frequency of rs999199 was 0.02, further haplotype and diplotype analyses were only carried out between rs125555 and rs140689. Table III shows the haplotype analysis results, which revealed that distribution of haplotype among the NSCLC patients was significantly different from that among the controls (global test p-value $= 0.0041$). Compared with haplotypes '*GA*', haplotype '*GT*' significantly increased the risk of NSCLC (OR 1.45, 95% CI 1.08–1.93, $p = 0.012$), while haplotype '*CA*' did not show significant association with lung cancer risk (OR 0.82, 95% CI 0.65–1.04, $p = 0.10$).

We also assessed the association of trichotomized diplotype (0, 1 or 2 copies of the haplotype) with lung cancer risk. As shown in Table IV, the distributions of two diplotypes built by haplotypes '*CA*' and '*GT*' showed a significant difference between NSCLC patients and controls ($p = 0.0386$ and $p = 0.0204$, individually). The trend test revealed a positive association between lung cancer risk and the haplotype copy number of '*GT*' ('1 copy' OR 1.42, 95% CI 1.03–1.47; '2 copies' OR 2.88, 95% CI 0.98–8.39; $p_{trend} = 0.0041$).

Discussion

The study was performed to investigate the association between MBD1 polymorphisms and lung cancer risk. Two polymorphisms were found to be significantly associated with lung cancer risk: the non-synonymous polymorphism (rs125555, Pro401Ala) decreases the lung cancer risk in the dominant genetic model ($p = 0.011$) and the intronic polymorphism (rs140689) increases the risk in the additive model ($p_{trend} = 0.011$). The latter effect was even more pronounced among patients with NSCLC ($p_{trend} = 0.005$). As the two polymorphisms were in strong linkage disequilibrium ($D' > 0.8$), haplotype and diplotype analyses of the two variants were performed and revealed a significant association with NSCLC. Stratified analysis also indicated an interaction relationship exist between the nonsynonymous variant rs125555 and trichotomized pack-years ($p_{trend} = 0.040$). The lack of association between SNPs and SCLC could be due to the relative small sample size or the different carcinogenic aetiology for NSCLC and SCLC.

Since MBD1 has been involved in mediating methylation-associated gene silencing in various human cancers (Fujita et al., 2000; Merlo et al., 1995; Yu et al., 2004), polymorphisms in *MBD1* may have an influence on MBD1 activity on gene expression profiles, thereby modulating an individual's susceptibility to lung cancer. To date, only one previous study has reported the relationship between *MBD1* polymorphisms and lung cancer risk, in a Korean population (Jang et al. 2005). However this was not confirmed in other studies with different populations. Our results were consistent with the previous study of the nonsynonymous polymorphism (rs125555, G→C, Pro401Ala) on lung cancer risk. FastSNP prediction proved that the polymorphism might lead to a non-conservative change to the protein structure (Hu et al. 2006). The

Table II. Stratified analysis of *MBD1* polymorphisms with NSCLC by cumulative smoking dose (pack-years).

		Never smoker		Light smoker		Heavy smoker	
Polymorphism	Genotype	Case/control	OR (95% CI)[a]	Case/control	OR (95% CI)[a]	Case/control	OR (95% CI)[a]
rs2851716[b]	TT	108/146	1.00 (Ref)	101/101	1.00 (Ref)	99/64	1.00 (Ref)
	TC	42/95	0.59 (0.38–0.93)	39/64	0.57 (0.35–0.94)	57/26	1.42 (0.81–2.51)
	CC	6/8	0.93 (0.31–2.80)	7/6	1.09 (0.35–3.41)	7/7	0.58 (0.19–1.79)
	TC+CC	48/103	0.62 (0.40–0.95)	46/70	0.62 (0.38–0.99)	64/33	1.24 (0.73–2.11)
rs125555[c]	G/G	109/146	1.00 (Ref)	101/101	1.00 (Ref)	99/64	1.00 (Ref)
	C/G	41/95	0.57 (0.36–0.89)	39/64	0.57 (0.35–0.94)	57/26	1.42 (0.81–2.51)
	C/C	6/8	0.92 (0.30–2.77)	5/6	0.77 (0.22–2.65)	7/7	0.58 (0.19–1.79)
	C/G+C/C	47/103	0.60 (0.39–0.92)	44/70	0.59 (0.36–0.95)	64/33	1.24 (0.73–2.11)
rs140689[d]	A/A	114/208	1.00 (Ref)	109/142	1.00 (Ref)	122/73	1.00 (Ref)
	T/A	38/37	1.81 (1.08–3.04)	33/28	1.59 (0.89–2.81)	37/24	0.92 (0.51–1.67)
	T/T	4/4	1.52 (0.36–6.34)	5/1	5.40 (0.60–48.53)	4/–	–
	T/A+T/T	42/41	1.78 (1.08–2.93)	38/29	1.73 (0.99–3.01)	41/24	1.02 (0.57–1.83)
rs999199	CC	150/233	1.00 (Ref)	140/166	1.00 (Ref)	159/94	1.00 (Ref)
	AC	6/16	0.62 (0.23–1.64)	7/5	1.56 (0.47–5.14)	4/3	0.63 (0.14–2.94)
	AA	–/–	–	–/–	–	–/–	–

[a]OR (95%CI) were adjusted for age, gender, family history of cancer (fmc) and smoking status; [b]$p_{trend} = 0.051796$ for interaction analysis under co-dominant model; [c]$p_{trend} = 0.03996$ for interaction analysis under co-dominant model; [d]$p_{trend} = 0.053358$ for interaction analysis under co-dominant model.

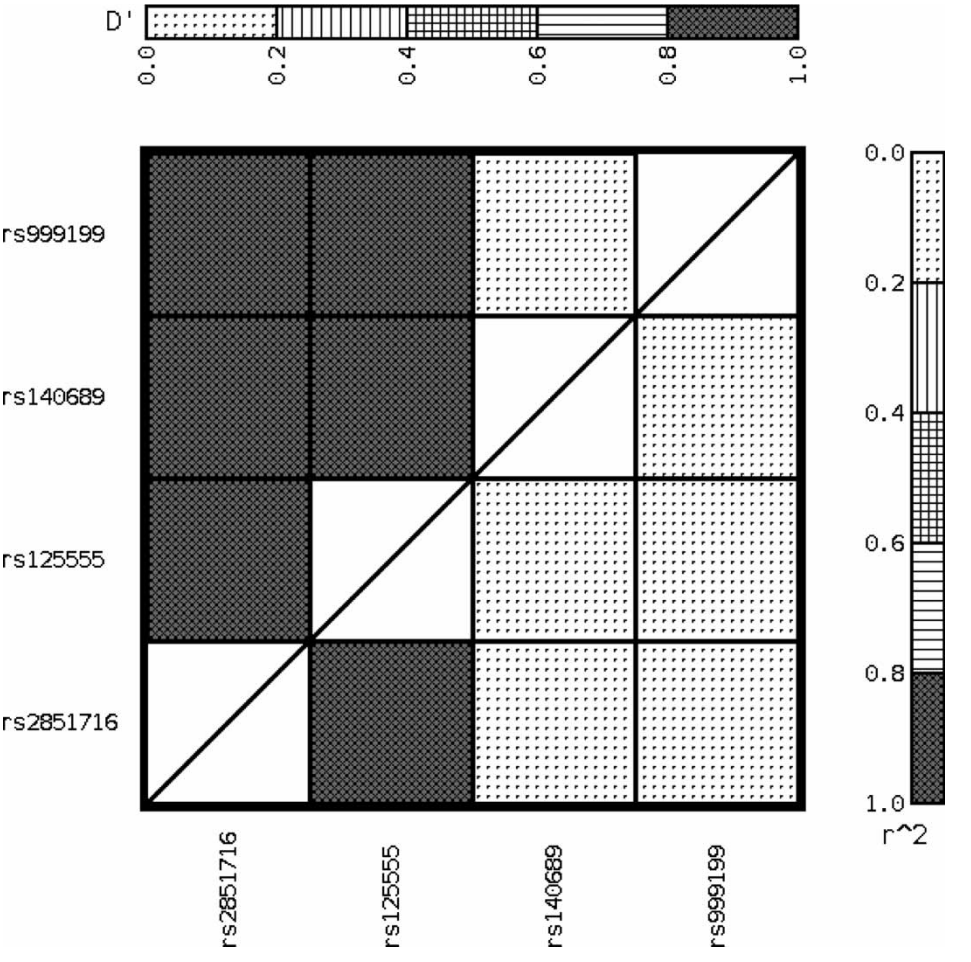

Figure 1. Linkage analysis of *MBD1* polymorphisms in controls by JLIN. Strong linkage disequilibrium exists among rs125555, rs140689 and rs999199 ($D' > 0.8$), and SNP rs285176 shows complete linkage disequilibrium with rs125555 ($r^2 = 1.0$).

exact nature of the functional alterations associated with the polymorphism will require further exploration with *in vitro* studies. Although another reported positive SNP (G-634A, rs732066) was not included in our study, it showed a strong linkage disequilibrium with Pro401Ala according to the HapMap data of Chinese population ($r^2 = 0.829$), and also the FastSNP prediction revealed that the latter had a higher

Table III. Associations between *MBD1* common haplotypes and NSCLC risk.

Haplotype	rs125555	rs140689	Case freq	Control freq	Total freq	OR (95% CI)[a]	*p*-Value[b]
1	G	A	0.6697	0.6847	0.6776	1.00 (Ref)	–
2	C	A	0.1866	0.2195	0.2039	0.82 (0.65–1.04)	0.100
3	G	T	0.1438	0.0956	0.1185	1.45 (1.08 –1.93)	0.012

[a]OR (95% CI) was adjusted for age, gender, family history of cancer, smoking status and pack-years; [b]global test *p*-value = 0.0041.

Table IV. Diplotype analysis of *MBD1* polymorphisms with NSCLC risk.

	0 copy		1 copy		2 copies			Global test
Haplotype	Case/ control	OR (95% CI)[a]	Case/ control	OR (95% CI)[a]	Case/ control	OR (95% CI)[a]	p_{trend}	p-value
1	57/46	1.00 (Ref)	193/234	0.72 (0.46–1.13)	216/237	0.85 (0.54–1.33)	0.5249	0.2713
2	311/311	1.00 (Ref)	137/185	0.70 (0.53–0.93)	18/21	0.73 (0.37–1.44)	0.0979	0.0386
3	345/423	1.00 (Ref)	108/89	1.42 (1.03–1.47)	13/5	2.88 (0.98–8.39)	0.0041	0.0204

[a]OR (95% CI) was adjusted for age, gender, family history of cancer, smoking status and pack-years.

potential risk than the former (http://fastsnp.ibms.sinica.edu.tw/pages/PrioritizeResult.jsp?taskid=TK8379&Submit=ENST00000269468).

In this study, one intronic variant (rs140689) also showed a risk effect for lung cancer which has not been reported in any other paper. There is no clear biological explanation for this result, and we cannot rule out the possibility that our observations for rs140689 were due to chance. However, according to the bioinformatics prediction by FastSNP (Yuan et al. 2006), this polymorphism may influence the binding of transcription factors. Because of functional evidence lacking, it can not be excluded that the intronic variant maybe just one tagging SNP of other functional SNPs (Johnson et al., 2001). Replication in a future study is needed.

The present study is a case–control study, with the major weakness that the disease occurred before patients were included in the study. Thus, the disease might have influenced some of the results. However, we believe that our results are unlikely to be attributable to selection bias because we matched the controls to the cases according to age, gender and residential area. Like most other susceptibility studies, another weakness is that the selection of SNPs has no biological basis, and the results just suggest association relationships exist between MBD1 SNPs and lung cancer risk. Further functional evidence should be provided as for the biological consequences. Moreover, due to the moderate sample size in the present study, the results also should be confirmed byother large population-based studies.

In conclusion, we observed significant associations between *MBD1* rs125555 and rs140689 polymorphisms and lung cancer risk especially for patients with NSCLC. Moreover an interaction between rs125555 and pack-years was also suggested. Further haplotype and diplotype analyses also supported the hypothesis that the MBD1 polymorphisms may be involved in the development of lung cancer. Replication in large epidemiological studies among diverse ethnic populations is recommended.

Acknowledgements

This work was funded by: China National Key Basic Research Program 2002CB512902, 2002BA711A10 and 2004CB518605; National Outstanding Youth Science Foundation of China 30425001 and Shanghai Science and Technology Program 06DZ19501.

Declaration of interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Atanassov B. S., Barrett J., Davis B. J. 2005. Homozygous germ line mutation in exon 27 of murine Brca2 disrupts the Fancd2-Brca2 pathway in the homologous recombination-mediated DNA interstrand cross-links' repair but does not affect meiosis. Genes Chromosomes & Cancer 44:429–437.

Bader S., Walker M., McQueen H. A., Sellar R., Oei E., Wopereis S., Zhu Y., Peter A., Bird A. P., Harrison D. J. 2003. MBD1, MBD2 and CGBP genes at chromosome 18q21 are infrequently mutated in human colon and lung cancers. Oncogene 22:3506–3510.

Bellacosa A., Cicchillitti L., Schepis F., Riccio A., Yeung A. T., Matsumoto Y., Golemis E. A., Genuardi M., Neri G. 1999. MED1, a novel human methyl-CpG-binding endonuclease, interacts with DNA mismatch repair protein MLH1. Proceedings of the National Academy of Sciences U S A 96:3969–3974.

Carter K. W., McCaskie P. A., Palmer L. J. 2006. JLIN: a java based linkage disequilibrium plotter. BMC Bioinformatics 7:60.

Danam R. P., Howell S. R., Brent T. P., Harris L. C. 2005. Epigenetic regulation of O6-methylguanine-DNA methyltransferase gene expression by histone acetylation and methyl-CpG binding proteins. Molecular Cancer Therapeutics 4:61–69.

Fujita N., Shimotake N., Ohki I., Chiba T., Saya H., Shirakawa M., Nakao M. 2000. Mechanism of transcriptional regulation by methyl-CpG binding protein MBD1. Molecular & Cellular Biochemistry 20:5107–5018.

Gonzalez J. R., Armengol L., Sole X., Guino E., Mercader J. M., Estivill X., Moreno V. 2007. SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23:644–645.

Hu Z., Jin G., Liu H., Sun W., Wang H., Shao M., Ma H., Wang Y., Wang Y. 2007. Genetic variants in MGMT and risk of lung cancer in Southeastern Chinese: a haplotype-based analysis. Human Mutation 28:431–440.

Hu Z., Xu L., Shao M., Yuan J., Wang Y., Wang F., Yuan W., Qian J., Ma H., Wang Y., Liu H., Chen W., Yang L., Jing G., Huo X., Chen F., Jin L., Wei Q., Wu T., Lu D., Huang W., Shen H. 2006. Polymorphisms in the two helicases ERCC2/XPD and ERCC3/XPB of the transcription factor IIH complex and risk of lung cancer: a case-control analysis in a Chinese population. Cancer Epidemiology, Biomarkers & Prevention 15:1336–1340.

Iniesta R., Guino E., Moreno V. 2005. Statistical analysis of genetic polymorphisms in epidemiological studies. Gacita Sanitaria 19:333–341.

Jang J. S., Lee S. J., Choi J. E., Cha S. I., Lee E. B., Park T. I., Kim C. H., Lee W. K., Kam S., Choi J. Y., Kang Y. M., Park R. W., Kim I. S., Cho Y. L., Jung T. H., Han S. B., Park J. Y. 2005. Methyl-CpG binding domain 1 gene polymorphisms and risk of primary lung cancer. Cancer Epidemiology, Biomarkers & Prevention 14:2474–2480.

Johnson G. C., Esposito L., Barratt B. J., Smith A. N., Heward J., Di Genova G., Ueda H., Cordell H. J., Eaves I. A., Dudbridge F., Twells R. C., Payne F., Hughes W., Nutland S., Stevens H., Carr P., Tuomilehto-Wolf E., Tuomilehto J., Gough S. C., Clayton D. G., Todd J. A. 2001. Haplotype tagging for the identification of common disease genes. Nature Genetics 29:233–237.

Lewontin R. C. 1988. On measures of gametic disequilibrium. Genetics 120:849–852.

Lyst M. J., Nan X., Stancheva I. 2006. Regulation of MBD1-mediated transcriptional repression by SUMO and PIAS proteins. EMBO Journal 25:5317–5328.

Merlo A., Herman J. G., Mao L., Lee D. J., Gabrielson E., Burger P. C., Baylin S. B., Sidransky D. 1995. 5′ CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers. Nature Medicine 1:686–692.

Moore L. E., Huang W. Y., Chung J., Hayes R. B. 2003. Epidemiologic considerations to assess altered DNA methylation from environmental exposures in cancer. Annals of the New York Academy of Sciences 983:181–196.

Nakao M., Matsui S., Yamamoto S., Okumura K., Shirakawa M., Fujita N. 2001. Regulation of transcription and chromatin by methyl-CpG binding protein MBD1. Brain Development 23(Suppl. 1):S174–S176.

Parkin D. M., Bray F., Ferlay J., Pisani P. 2005. Global cancer statistics, 2002. CA: A Cancer Journal for Clinicians 55:74–108.

Robertson K. D. 2002. DNA methylation and chromatin – unraveling the tangled web. Oncogene 21:5361–5379.

Robertson K. D., Jones P. A. 2000. DNA methylation: past, present and future directions. Carcinogenesis 21:461–467.

Rountree M. R., Bachman K. E., Baylin S. B. 2000. DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. Nature Genetics 25:269–277.

Shoda M., Ishida T., Yamaguchi K., Kamihira S., Utsunomiya A., Watanabe T. 2003. Microsatellite instability analysis of ATL cells. AIDS Research and Human Retroviruses 19:S–74.

Spitz M. R., Wei Q., Dong Q., Amos C. I., Wu X. 2003. Genetic susceptibility to lung cancer: the role of DNA damage and repair. Cancer Epidemiology, Biomarkers & Prevention 12:689–698.

Stephens M., Donnelly P. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. American Journal of Human Genetics 73:1162–1169.

Wade P. A. 2001. Methyl CpG binding proteins: coupling chromatin architecture to gene regulation. Oncogene 20:3166–3173.

Yan P. S., Chen C. M., Shi H., Rahmatpanah F., Wei S. H., Caldwell C. W., Huang T. H. 2001. Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. Cancer Research 61:8375–8380.

Yoon B. I., Li G. X., Kitada K., Kawasaki Y., Igarashi K., Kodama Y., Inoue T., Kobayashi K., Kanno J., Kim D. Y., Inoue T., Hirabayashi Y. 2003. Mechanisms of benzene-induced hematotoxicity and leukemogenicity: cDNA microarray analyses using mouse bone marrow tissue. Environmental Health Perspectives 111:1411–1420.

Yu H. P., Wang X. L., Sun X., Su Y. H., Wang Y. J., Lu B., Shi L. Y., Xiong C. l., Li Y. Y., Li F., Xu S. Q. 2004. Polymorphisms in the DNA repair gene XPD and susceptibility to esophageal squamous cell carcinoma. Cancer Genetics and Cytogenetics 154:10–15.

Yuan HY, Chiou JJ, Tseng WH, Liu CH, Liu CK, Lin YJ, Wang HH, Yao A, Chen YT, Hsu CN. 2006. FASTSNP: an always up-to-date and extendable service for SNP function analysis and prioritization. Nucleic Acids Research 34:635–641.