Carcinogenesis vol.27 no.7 pp.1475–1480, 2006
doi:10.1093/carcin/bgi350
Advance Access publication March 7, 2006

Polymorphisms in DNA damage binding protein 2 (DDB2) and susceptibility of primary lung cancer in the Chinese: a case–control study

Zhibin Hu[†], Minhua Shao[1,†], Jing Yuan[2], Liang Xu[3], Feng Wang[2], Yi Wang[1], Wentao Yuan[3], Ji Qian[1], Hongxia Ma, Ying Wang[3], Hongliang Liu[1], Weihong Chen[2], Lin Yang[3], Guangfu Jin, Xiang Huo, Feng Chen, Li Jin[1], Qingyi Wei[4], Wei Huang[3], Daru Lu[1], Tangchun Wu[2,*] and Hongbing Shen[*]

Department of Epidemiology and Biostatistics, Cancer Research Center of Nanjing Medical University, Nanjing 210029, China, [1]State Key Laboratory of Genetic Engineering, School of Life Sciences, Fudan University, Shanghai 200433, China, [2]Institute of Occupational Medicine, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, 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

*To whom correspondence and reprints should be addressed
Email: hbshen@njmu.edu.cn

DNA damage binding protein 2 (DDB2) is one of the major DNA repair proteins involved in the nucleotide excision repair (NER) pathway. Mutations in the *DDB2* gene can cause a repair-deficiency syndrome xeroderma pigmentosum group E. Because tobacco carcinogens can cause DNA damage that is repaired by NER and suboptimal NER capacity is reported to be associated with lung cancer risk, we hypothesized that common variants in the *DDB2* gene are associated with lung cancer risk. To test this hypothesis, we conducted a case–control study of 1010 patients with incident lung cancer and 1011 cancer-free controls and genotyped two *DDB2* single nucleotide polymorphisms (SNPs) (rs830083 and rs3781620) that are in linkage disequilibrium with other untyped SNPs. We found that compared with the rs830083CC, subjects carrying the heterozygous rs830083CG genotype had a significantly 1.31-fold increased risk of lung cancer [95% confidence interval (CI) 1.08–1.60] and those carrying the homozygous rs830083GG genotype had a non-significantly 1.22-fold elevated risk (95% CI 0.89–1.67). In addition, effects of the combined rs830083CG/GG variant genotypes were more evident in young subjects, heavy smokers and subjects with a positive family history of cancer. These findings indicate, for the first time, that the *DDB2* rs830083 polymorphism may contribute to the etiology of lung cancer. Further functional studies on this SNP and/or related variants are warranted to elucidate the underlying molecular mechanisms of the association.

Abbreviations: CI, confidence interval; DDB2, DNA damage binding protein 2; LD, linkage disequilibrium; MAF, minor allele frequency; NER, nucleotide excision repair; OR, odds ratio; SNP, single nucleotide polymorphism; XPE, xeroderma pigmentosum group E.

Correspondence may also be addressed to wut@mails.tjmu.edu.cn
[†]The first two authors contributed equally to this work.

Introduction

Lung cancer remains the leading cause of cancer-related deaths worldwide, with an estimated 1.35 million new cases and 1.18 million deaths in 2002 (1). Owing to continuous increase in the prevalence of smoking, the incidence rate of lung cancer in China has been significantly increasing in both urban and rural areas in the last two decades (2,3). It was estimated that ~90% of lung cancer cases are caused by tobacco smoking (4); however, only a small fraction of the smokers (usually <20%) developed lung cancer, suggesting an individual susceptibility to lung cancer.

DNA repair, a central defense system by which the cells cope with different DNA damage, maintains the integrity of human genome, and faulty DNA repair confers to the individual cancer susceptibility (5). Among several partly overlapped DNA repair pathways, nucleotide excision repair (NER) is the major repair pathway for removing DNA damage caused by tobacco smoking and deals with a wide class of helix-distorting lesions that interfere with base pairing and obstruct replication and transcription (6,7). Mutations in at least one of the eight NER core proteins can cause xeroderma pigmentosum (XP) with corresponding complementation groups characteristic of repair-deficiency phenotype (i.e. XP: A–G and V) (8,9), coupled with high UV-light sensitivity and increased cancer risk (10).

Interestingly, the XP group E (XPE) protein is not necessary for NER *in vitro* but is evidently required *in vivo*; XPE cells are biochemically heterogeneous, and some XPE cell lines lack the activity of binding to damaged DNA owing to mutations in the *DDB2* gene (11). The protein product of the *DDB2* gene, p48, interacts with the p127 protein and forms a UV-damaged DNA binding complex (termed DDB) (12,13). This DDB complex can bind to various forms of damaged DNA and participate in global NER (14–16). Fibroblasts isolated from the XPE patients harboring mutations in the *DDB2* gene could complement the deficiency by microinjecting of the wild-type protein (17,18). Recently, Itoh *et al*. (19) and Yoon *et al.* (20) generated, separately, a strain of $DDB2^{-/-}$ mice and demonstrated a role of DDB2 in preventing UV-induced skin carcinogenesis. Furthermore, the DDB2-deficient mice developed spontaneous tumors at a high rate, suggesting tumor suppression function of DDB2 (20).

Accumulating evidence supports that polymorphisms in the NER genes may contribute to genetic susceptibility to lung cancer (21,22). To date, however, no studies investigate the association between common polymorphisms in the *DDB2* gene and cancer susceptibility. The *DDB2* gene is polymorphic, having 14 single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) >10% among those included in the Environmental Genome Project (EGP) database of resequencing data from 90 USA individuals (http://egp.gs.washington.edu/finished_genes.html). Because *DDB2* plays an important role in the NER pathway, tobacco

carcinogens can cause DNA damage that is repaired by NER and suboptimal NER capacity is reported to be associated with lung cancer risk (23), we hypothesized that genetic variants in the *DDB2* gene are associated with lung cancer risk. To test this hypothesis, we conducted a case–control study of 1010 incident lung cancer cases and 1011 age and sex frequency-matched cancer-free controls in a Chinese population.

Materials and methods

Study populations

All the subjects were genetically unrelated ethnic Han Chinese. The cases included all eligible patients newly diagnosed with incident lung cancer according to the National Diagnosis Standard for Lung Cancer, who were consecutively recruited between July 2002 and November 2004 from four hospitals of three metropolitan cities along the Yangzi River, including the Cancer Hospital of Jiangsu Province, the First Affiliated Hospital of Nanjing Medical University, the Shanghai Cancer Hospital and the Wuhan Zhongnan Hospital, without the restrictions of age, sex and histology. The exclusion criteria included reported previous cancer history and radiotherapy or chemotherapy for unknown conditions. A total of 1299 cases with histopathologically confirmed lung cancer were recruited, of whom 1010 patients consented to participate in the study and provided blood samples, resulting in a response rate of 77.8% (1010 out of 1299). The control subjects consisted of outpatients with diseases other than cancer in other departments of the same hospital during the same time period when the cases were recruited, who were seen at departments of general surgery, gynecology, internal medicine, orthopedics and otorhinolaryngology. All the control subjects were frequency-matched to the cases on age (±5 years), sex and residential area (urban or countryside), and the response rate was 81.3% (1011 out of 1244).

Each participant was scheduled for an interview after a 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 <1 cigarette per day and shorter than 1 year in their lifetime were defined as non-smokers; otherwise, they were considered as smokers. Those smokers who quit for >1 year were considered former smokers. Pack-years [(cigarettes per day/20) × years smoked] were calculated to indicate the cumulative smoking dose and the smokers being further dichotomized by the cumulative dose of 29 pack-years according to the distribution of controls. Family history of cancer was defined as any reported cancer in first-degree relatives (parents, siblings or children). After interview, ~5 ml venous blood sample was collected from each participant. The study was approved by the institutional review boards of Nanjing Medical University, Fudan University and Tongji Medical College of Huazhong University of Science and Technology.

Selection of SNPs in DDB2

We used the EGP database (http://egp.gs.washington.edu/finished_genes.html) to select SNPs from among the reported polymorphisms on the basis of their pairwise linkage disequilibrium (LD). As shown in Table I, a total of 14 SNPs with a MAF >10% were identified, of which rs830083 (nt18100C>G) is unique because it has a >0.4 LD parameters $r^2$with all other SNPs, close to the r^2 threshold 0.5 recommended by Carlson *et al.* (24) for the tagging SNP selection. Furthermore, because the 90 USA individuals in the EGP database were a mixed ethnic and Asian population (including Chinese) that was defined as lower haplotype diversity and higher pairwise LD compared with other populations (25), we thought the 0.4 r^2 threshold in our study was adequately stringent. To provide a test for our assumption, we also selected and genotyped another SNP, rs3781620 (nt23314C>G), to compare the pairwise r^2 between our Chinese population and the USA mixed population ($r^2 = 0.68$ in the database).

Laboratory assays

Genotyping was performed by the 5′-nuclease (TaqMan) assay, using the ABI PRISM 7900HT Sequence Detection System (Applied Biosystems, Foster City, CA), in 384-well format, in Chinese National Human Genome Center at Shanghai, China. The TaqMan primers and probes were designed using the Primer Express Oligo Design software v2.0 (ABI PRISM). Primer and probe sequences are as follows: rs830083F, CACCTCAGCCTCCCAAGTG; rs830083R, CAACGTGACAAAACCCCATCTTAAA; rs830083 probe1, CCAGCTAATTTTTCTATTTT-VIC; and rs830083 probe2, CAGCTAATTT-TTGTATTTT-FAM; rs3781620F, CTCAGCCTCCCAAAGTGTTGA; rs3781620R, ACATCAAAAGTATTCTTTGGGCTGAGT; rs3781620 probe1, CTCACAGCTGTAATC-VIC; and rs3781620 probe2, CTCACA-CCTGTAATC-FAM. PCR reactions were carried out in a reaction volume of 5 μl containing 5 ng DNA, 2.5 μl 2× TaqMan universal PCR Master MixNo AmpErase UNG (Applied Biosystems), 0.083 μl 40× Assay Mix. PCR reaction condition included 95°C for 10 min followed by 20 cycles of 15 s at 92°C and 1 min at 60°C followed by 30 cycles of 15 s at 89°C and 1.5 min at 60°C. Two blank control (water) and two duplicated samples in each 384-well format were used for quality-control procedure. The intensity of each SNP should meet the criteria of three clear clusters in two scales generated by SDS software (ABI). Genotyping failed in 49 (4.9%) cases and 51 (5.0%) controls in rs830083 locus and 43 (4.3%) cases and 33 (3.3%) controls in rs37816203 locus owing to DNA quantity or quality, and 921 (91.2%) cases and 931 (92.1%) controls were successfully genotyped in both loci.

Statistical analyses

Differences in select demographic variables, smoking status, pack-years smoked and frequencies of the *DDB2* genotypes, alleles and haplotypes between the cases and controls were evaluated by using the χ^2-test. The associations between the frequencies of *DDB2* variant genotypes and lung cancer risk were estimated by computing the odds ratios (ORs) and 95% confidence intervals (CIs) from both univariate and multivariate logistic regression analyses with adjustment for age, sex, family history of cancer and pack-years of smoking. We used the PHASE 2.0 program (26) to infer haplotype frequencies based on the observed *DDB2* genotypes. The association between the frequencies of *DDB2* rs830083 variant genotypes and lung cancer risk was also evaluated by stratification analyses by age, sex, pack-years of smoking, family history of cancer and histological types. The potential gene–environment interaction was also evaluated by logistic regression analysis and tested by comparing the changes in deviance (−2 log likelihood) between the models of main effects with or without the interaction term. All the statistical analyses were performed with Statistical Analysis System software (v.8.0e; SAS Institute, Cary, NC).

Table I. LD (r^2) between SNPs in *DDB2* gene region

	9434	9539	17293	18059	18100 (rs830083)	19646	20756	21388	22902	23314 (rs3781620)	23718	24043	24323	24370
9434	–	0.95	0.77	0.19	0.56	0.47	0.72	0.55	0.75	0.73	0.34	0.55	0.39	0.34
9539		–	0.73	0.19	0.54	0.47	0.68	0.52	0.71	0.68	0.32	0.54	0.37	0.32
17293			–	0.27	0.70	0.63	0.94	0.72	0.96	0.92	0.44	0.69	0.51	0.46
18059				–	0.43	0.32	0.29	0.20	0.28	0.29	0.15	0.23	0.14	0.14
18100 (rs830083)					–	0.68	0.69	0.50	0.66	0.68	0.42	0.55	0.44	0.41
19646						–	0.62	0.47	0.62	0.63	0.43	0.62	0.45	0.44
20756							–	0.67	0.90	0.95	0.45	0.71	0.52	0.44
21388								–	0.73	0.66	0.34	0.48	0.38	0.34
22902									–	0.91	0.41	0.65	0.48	0.43
23314 (rs3781620)										–	0.45	0.72	0.52	0.44
23718											–	0.53	0.79	0.77
24043												–	0.58	0.50
24323													–	0.86
24370														–

Results

The distributions of selected characteristics between lung cancer patients and controls are summarized in Table II. There was no significant difference in the distributions of age and sex between the cases and the controls, suggesting that our frequency-matching was adequate. However, the controls were more likely to be non-smokers (47.8%) than were the cases (31.2%), and more cases (44.5%) smoked >30 pack-years than did the controls (25.4%), and these differences were statistically significant ($P < 0.0001$). Furthermore, 17.1% of the lung cancer cases reported a family history of cancer in their first-degree relatives, which was significantly higher than that of the controls (12.8%), and this difference accounted for a significantly 41% increased lung cancer risk (OR = 1.41, 95% CI = 1.10–1.81). Of the 1010 cancer patients, 430 (42.6%) were adenocarcinoma, 335 (33.2%) squamous cell carcinoma, 65 (6.4%) small cell carcinoma and 180 (17.8%) large cell, mixed cell or undifferentiated carcinomas.

Allele frequencies between the cases and controls were significantly different for rs830083 ($P = 0.0481$) but not for rs3781620 ($P = 0.1609$). The LD analyses showed that the rs830083 locus was in nearly complete LD with the rs37816203 locus ($r^2 = 0.9891$ and $D' = 0.9951$). There were four haplotypes derived from these two *DDB2* genotypes, of which the haplotypes rs830083C-rs3781620C and rs830083G-rs3781620G accounted for 99.5% of the chromosomes for the cases and 99.4% of the chromosomes for the controls (Table II). Compared with the rs830083C-rs3781620C haplotype, the rs830083G-rs3781620G haplotype was associated with a borderline 1.07-fold increased risk of lung cancer (95% CI = 1.00–1.14).

The genotype distributions of *DDB2* rs830083 and rs3781620 in the cases and the controls are shown in Table III. The observed genotype frequencies for these two polymorphisms were all in Hardy–Weinberg equilibrium in the controls ($P = 0.9608$ and 0.9644, respectively). The *DDB2* rs830083 genotype frequencies were 38.7 (CC), 50.3 (CG) and 11.0% (GG) in the cases and 44.7 (CC), 44.4 (CG) and 10.9% (GG) in the controls, and the difference was statistically significant ($\chi^2 = 7.63$, $P = 0.0220$). Logistic regression analysis revealed that compared with the rs830083CC genotype, subjects carrying the heterozygous rs830083CG genotype had a significantly 1.31-fold increased risk of lung cancer (95% CI = 1.08–1.60) and those carrying the homozygous rs830083GG genotype had a non-significantly 1.22-fold elevated risk (95% CI = 0.89–1.67). When we

Table II. Distribution of select variables, *DDB2* alleles/haplotypes in lung cancer cases and cancer-free controls

Variable	Cases ($n = 1010$)		Controls ($n = 1011$)		*P*-value[a]
	No.	(%)	No.	(%)	
Age (years)					0.9824
≤60	500	49.5	500	49.5	
>60	510	50.5	511	50.5	
Sex					0.3037
Male	777	76.9	758	75.0	
Female	233	23.1	253	25.0	
Smoking status					<0.0001
Non-smokers	315	31.2	483	47.8	
Former smokers	301	29.8	161	15.9	
Current smokers	394	39.0	367	36.3	
Pack-years of smoking					<0.0001
0	315	31.2	483	47.8	
1–29	245	24.3	271	26.8	
30+	450	44.5	257	25.4	
Pack-years (mean ± SD)[b]	40.7 ± 27.4		31.7 ± 21.2		<0.0001
Family history of cancer					0.0059
No	837	82.9	882	87.2	
Yes	173	17.1	129	12.8	
rs830083[c]					0.0481
C allele	1227	63.8	1284	66.9	
G allele	695	36.2	636	33.1	
rs37816203[c]					0.1609
C allele	1239	64.1	1295	66.2	
G allele	695	35.9	661	33.8	
Haplotypes					0.0364
rs830083C-rs37816203C	1281	63.4	1341	66.3	
rs830083C-rs37816203G	4	0.2	10	0.5	
rs830083G-rs37816203C	6	0.3	2	0.1	
rs830083G-rs37816203G	729	36.1	669	33.1	
Histological types					
Adenocarcinomas	430	42.6			
Squamous cell	335	33.2			
Small cell	65	6.4			
Other carcinomas[d]	180	17.8			

[a]Two-sided χ^2-test.
[b]Average pack-years of smoking was calculated only for smokers.
[c]Genotyping failed in 49 cases and 51 controls in rs830083 locus and 43 cases and 33 controls in rs37816203 locus.
[d]Other carcinomas means the large cell, mixed cell carcinomas or undifferentiated carcinoma.

combined the rs830083CG and rs830083GG genotypes assuming a co-dominant allele effect, the combined rs830083CG/GG variant genotypes were associated with a 30% (adjusted OR = 1.30; 95% CI = 1.08–1.56) increased risk for lung cancer (Table III). Similarly, compared with rs3781620CC wild-type genotype, elevated risks were associated with the rs3781620CG, rs3781620GG, or their combined genotypes rs3781620CG/GG [ORs and 95% CIs being 1.28 (1.06–1.56), 1.11 (0.81–1.51) and 1.25 (1.04–1.50), respectively] (Table III).

Because the rs830083 polymorphism was in nearly complete LD with the rs3781620 variant and was the most representative locus in the *DDB2* gene, we further evaluated the associations of the rs830083CG/GG variant genotypes with lung cancer risk stratified by selected variables and histological types. As shown in Table IV, the effect of combined variant genotypes was more evident in young subjects (adjusted OR = 1.40; 95% CI = 1.07–1.83) and subjects with a positive family history of cancer (adjusted OR = 1.70; 95% CI = 1.03–2.80). However, although the elevated risk was more evident in heavy smokers (>29 pack-years, adjusted OR = 1.48; 95% CI = 1.07–2.04), we found that the main effect of *DDB2* variant genotypes was very similar between smokers and non-smokers. For the stratification of major histological types, risk of lung cancer associated with rs830083CG/GG was mainly limited in subjects with adenocarcinoma. Furthermore, compared with the non-smokers having the rs830083CC wild-type genotype, we observed a 5.04-fold increased risk of lung

Table III. Analysis of association between the *DDB2* polymorphisms and risk of lung cancer

Genotype	Cases		Controls		Crude OR (95% CI)	Adjusted OR (95% CI)[a]
	No.	(%)	No.	(%)		
rs830083	961		960			
CC	372	(38.7)	429	(44.7)	1.00	1.00
CG	483	(50.3)	426	(44.4)	1.31 (1.08–1.58)	1.31 (1.08–1.60)
GG	106	(11.0)	105	(10.9)	1.16 (0.86–1.58)	1.22 (0.89–1.67)
CG/GG	589	(61.3)	531	(55.3)	1.28 (1.07–1.53)	1.30 (1.08–1.56)
rs37816203	967		978			
CC	374	(38.7)	429	(43.9)	1.00	1.00
CG	491	(50.8)	437	(44.7)	1.29 (1.07–1.56)	1.28 (1.06–1.56)
GG	102	(10.6)	112	(11.5)	1.04 (0.77–1.41)	1.11 (0.81–1.51)
CG/GG	593	(61.4)	549	(56.2)	1.24 (1.03–1.48)	1.25 (1.04–1.50)

[a]Adjusted for age, sex, family history of cancer and pack-years of smoking.

Table IV. Stratified analyses between the combined *DDB2* rs830083 genotypes and lung cancer risk

	*DDB2*rs830083					
	Cases (n= 961)		Controls (n= 960)		Adjusted OR (95% CI)[a]	
	CC *N* (%)	CG/GG *N* (%)	CC *N* (%)	CG/GG *N* (%)	CC	CG/GG
Age (years)						
≤60	175 (37.2)	296 (62.9)	206 (44.4)	258 (55.6)	1.00	1.40 (1.07–1.83)
>60	197 (40.2)	293 (59.8)	223 (45.0)	273 (55.0)	1.00	1.20 (0.92–1.56)
Sex						
Male	293 (39.5)	448 (60.5)	319 (44.9)	392 (55.1)	1.00	1.26 (1.02–1.57)
Female	79 (35.9)	141 (64.1)	110 (44.2)	139 (55.8)	1.00	1.39 (0.96–2.03)
Smoking status						
Non-smokers	112 (38.0)	183 (62.0)	207 (45.1)	252 (54.9)	1.00	1.29 (0.96–1.75)
Smokers	260 (39.0)	406 (61.0)	222 (44.3)	279 (55.7)	1.00	1.26 (0.99–1.60)
Pack-years of smoking						
0	112 (38.0)	183 (62.0)	207 (45.1)	252 (54.9)	1.00	1.29 (0.96–1.75)
1–29	95 (39.8)	144 (60.3)	105 (40.4)	155 (59.6)	1.00	1.01 (0.70–1.45)
>29	165 (38.6)	262 (61.4)	117 (48.5)	124 (51.5)	1.00	1.48 (1.07–2.04)
Family history of cancer						
No	313 (39.5)	480 (60.5)	371 (44.3)	467 (55.7)	1.00	1.24 (1.01–1.51)
Yes	59 (35.1)	109 (64.9)	58 (47.5)	64 (52.5)	1.00	1.70 (1.03–2.80)
Histological types						
Adenocarcinomas	149 (36.5)	259 (63.5)	429 (44.7)	531 (55.3)	1.00	1.40 (1.10–1.79)
Squamous cell	133 (41.6)	187 (58.4)			1.00	1.21 (0.92–1.58)
Small cell	32 (49.2)	33 (50.8)			1.00	0.86 (0.52–1.43)
Other carcinomas[b]	58 (34.5)	110 (65.5)			1.00	1.56 (1.09–2.22)

[a]Adjusted for age, sex, family history of cancer and pack-years of smoking.
[b]Other carcinomas means the large cell, mixed cell carcinomas or undifferentiated carcinoma.

cancer in heavy smokers (>29 pack-years) carrying the rs830083CG/GG genotypes (95% CI = 3.56–7.14). Similarly, compared with subjects with a negative family cancer history having the rs830083CC genotype, a 1.94-fold risk of lung cancer in subjects with a positive family cancer history carrying the rs830083CG/GG genotypes was found (95% CI = 1.36–2.76). However, we did not find any statistical evidence for interaction between rs830083 variant genotypes and pack-years of smoking or family cancer history in the multivariate logistic regression model ($P = 0.6304$ and 0.2467 for the interaction term, respectively; data not shown).

Discussion

In this hospital-based case–control study, we investigated the associations of the rs830083 polymorphism of the *DDB2* gene with risk of lung cancer in a Chinese population. We found that the rs830083CG/GG genotypes were associated with a significantly increased risk of lung cancer compared with the rs830083CC genotype. The elevated risk was more pronounced among younger subjects (<60 years old), heavy smokers (30+ pack-years) and subjects with a positive family cancer history. Because the rs830083 locus is in LD with most of the *DDB2* SNPs, the pairwise LD between rs830083 and rs37816203 in our study population is >0.68 in the mixed USA population, which supports our assumption that this variant is representative of other untyped SNPs among Asian population. Our SNP selection was based on the r^2 LD statistic, because r^2 is directly related to statistical power to detect disease associations with unassayed loci (27). With the advantage of the relatively large sample size, our study, for the first time, provided the evidence that SNPs in the *DDB2* gene may be associated with lung cancer susceptibility.

DDB2 is an important DNA damage binding protein that participates in global NER. In addition to its role in DNA repair, DDB2 also interacts with the cell-cycle transcription factor E2F1 (28,29), binds to cullin 4A, an E3 ligase in the ubiquitin–proteasome pathway (29), and is associated with the CBP/p300 family of histone acetyl transferase (30). Moreover, the human *DDB2* gene is a downstream target of the tumor suppressor p53 and BRCA1 (18,31). Recent studies provided further evidence for a feedback loop in which DDB2 plays an important role in the regulation of p53 and controlling p53-mediated apoptosis (19,20,32). In the DDB2-deficient animal model, a high rate of spontaneous tumors in mice lacking NER is interesting (33,34). Further animal studies revealed that enhanced expression of DDB2 both delayed the onset of squamous cell carcinoma and decreased the number of tumors per mouse in chronically UV-B light-exposed hairless mice (35). Inoki *et al*. (36) found that two DDB2 variants (D1 and D2) generated by alternative splicing were dominantly negative inhibitors of DNA repair. These further suggested that DDB2 might play a multifunctional role in cancer development.

Because no common non-synonymous SNPs in the *DDB2* coding region and no known regulatory variants were identified in the *DDB2* gene, LD-based analysis provided a way to avoid a priori consideration of SNPs function; rather, it assumes that genotypes at unassayed, risk-related SNPs may be linked with the assayed SNPs. The power of the study to detect the disease-associated polymorphisms indirectly in N samples is equivalent to the power to detect it directly in $N*r^2$ samples (27). Therefore, genotyping the rs830083 locus with a $r^2 \geq 0.4$ is at least equivalent to genotyping >400 pairs of cases and controls at the potentially disease-associated locus. In addition, there is accumulating evidence that supports the importance of intronic polymorphisms as markers for cancer susceptibility. For example, intronic SNPs in *p53* (37) and ataxia telangiectasia mutated (ATM) (38) were associated with increased risk for lung cancer and breast cancer, respectively. In contrast to exons, the association of intronic polymorphisms with cancer risk could be due to their tight linkage with other susceptibility loci or genes, although their potential influence on mRNA instability has been postulated (22). Interestingly, we found that only heterozygotes of *DDB2*, but not homozygotes, had a higher risk for lung cancer. Although the reason for a higher risk associated with the variant heterozygotes remains unknown, it is possible that this heterozygous genotype may be in LD with other susceptibility loci, that the heterozygotes may have had impaired function because of the potential imbalance of the protein structure, or that some of homozygotes in combination with other adverse genotypes may have been embyogenically lethal. However, these hypotheses need to be tested in future studies.

In the present study, we found that the risk of lung cancer associated with *DDB2* SNPs was more pronounced among relatively younger subjects (<60 years old) and those with a family cancer history, suggesting that genetic susceptibility, often associated with an early age of onset, may play a more important role in the development of lung cancer among younger patients than among relatively older subjects. It is also plausible that *DDB2* rs830083 locus may be a potential genetic marker of lung cancer and is associated with familial aggregation. Although the elevated risk was more evident in heavy smokers (>29 pack-years), we did not find a significant interaction between *DDB2* genotypes and smoking. In addition, the main effect of *DDB2* variant genotypes was very similar between smokers and non-smokers, indicating that the role of the genotypes was independent of smoking status. Furthermore, we found that the risk associated with the *DDB2* variant genotypes was higher for lung cancer patients with adenocarcinoma, suggesting that different histopathological types may have different etiologies, not only in relation to environmental risk factors but also in genetic susceptibility. However, because of the small sample size in the subgroups, there findings were preliminary and need to be validated in further studies with larger sample sizes in these subpopulations.

Like all other case–control studies, inherited biases in the present study may have led to spurious findings. First of all, the response rate of the cases and controls were 77.8 and 81.3%, respectively, and ~5% of the DNA samples failed for genotyping at each locus, which may have caused selection bias. However, the general demographic characteristics and smoking habits of the subjects who were and were not included in the final analysis were similar, which should not lead to substantial bias, if any. In addition, potential confounding factors were minimized by frequency-matching on age, sex, and residential area between the cases and controls. Second, although we included >1000 cases and a comparable number of frequency-matched controls in the final analysis, this study was limited owing to exclusion of SNPs with low frequencies, which may be potentially important. Finally, no biological and functional relevance can be made on the basis of this tagging SNP approach, and additional studies may be needed to identify the disease-causing SNPs.

In conclusion, our study provides evidence for the first time that the *DDB2* polymorphisms may contribute to the etiology of lung cancer. The findings need to be substantiated by further studies on functional evaluations of *DDB2* polymorphisms as well as the determination of related lung cancer risk in larger molecular epidemiological studies with diverse ethnic populations and multiple *DDB2* SNPs.

Acknowledgements

The authors would like to thank Dr Yijiang Chen (The First Affiliated Hospital of Nanjing Medical University), Drs Jiannong Zhou and Lin Xu (Jiangsu Cancer Hospital) for their assistance in recruiting the patients; Ms Yanhong Liu (Fudan University), Ms Juying Niu (Nanjing Medical University) and personnel from Tongji Medical College for their help in sample management; and the associates from Department of Genetics, Chinese National Human Genome Center in Shanghai, for their technical support. This work was supported in part by the China National Key Basic Research Program Grants 2002CB512902 (to D.L. and H.S.), 2002CB512905 (to T.W.), 2002BA711A10 and 2004CB518605 (to W.H.); National Outstanding Youth Science Foundation of China 30425001 (to H.S.) and National '211' Environmental Genomics Grant (to D.L.).

Conflict of Interest Statement: None declared.

References

1. Parkin,D.M., Bray,F., Ferlay,J. and Pisani,P. (2005) Global cancer statistics, 2002. *CA. Cancer J. Clin.*, **55**, 74–108.
2. Zhang,H. and Cai,B. (2003) The impact of tobacco on lung health in China. *Respirology*, **8**, 17–21.
3. Yang,L., Parkin,D.M., Li,L. and Chen,Y. (2003) Time trends in cancer mortality in China: 1987–1999. *Int. J. Cancer*, **106**, 771–783.
4. Parkin,D.M., Pisani,P., Lopez,A.D. and Masuyer,E. (1994) At least one in seven cases of cancer is caused by smoking. Global estimates for 1985. *Int. J. Cancer*, **59**, 494–504.
5. Friedberg,E.C. (2003) DNA damage and repair. *Nature*, **421**, 436–440.
6. Hoeijmakers,J.H. (2001) Genome maintenance mechanisms for preventing cancer. *Nature*, **411**, 366–374.
7. Friedberg,E.C. (2001) How nucleotide excision repair protects against cancer. *Nat. Rev. Cancer*, **1**, 22–33.
8. Sancar,A. (1996) DNA excision repair. *Annu. Rev. Biochem.*, **65**, 43-81.
9. Cleaver,J.E., Thompson,L.H., Richardson,A.S. and States,J.C. (1999) A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Hum. Mutat.*, **14**, 9–22.
10. de Boer,J. and Hoeijmakers,J.H. (2000) Nucleotide excision repair and human syndromes. *Carcinogenesis*, **21**, 453–460.
11. Itoh,T., Linn,S., Ono,T. and Yamaizumi,M. (2000) Reinvestigation of the classification of five cell strains of xeroderma pigmentosum group E with reclassification of three of them. *J. Invest. Dermatol.*, **114**, 1022–1029.
12. Chu,G. and Chang,E. (1988) Xeroderma pigmentosum group E cells lack a nuclear factor that binds to damaged DNA. *Science*, **242**, 564–567.
13. Keeney,S., Chang,G.J. and Linn,S. (1993) Characterization of a human DNA damage binding protein implicated in xeroderma pigmentosum E. *J. Biol. Chem.*, **268**, 21293–21300.
14. Hwang,B.J., Toering,S., Francke,U. and Chu,G. (1998) p48 activates a UV-damaged-DNA binding factor and is defective in xeroderma pigmentosum group E cells that lack binding activity. *Mol. Cell. Biol.*, **18**, 4391–4399.
15. Tang,J.Y., Hwang,B.J., Ford,J.M., Hanawalt,P.C. and Chu,G. (2000) Xeroderma pigmentosum *p48* gene enhances global genomic repair and suppresses UV-induced mutagenesis. *Mol. Cell*, **5**, 737–744.
16. Wakasugi,M., Kawashima,A., Morioka,H., Linn,S., Sancar,A., Mori,T., Nikaido,O. and Matsunaga,T. (2002) DDB accumulates at DNA damage sites immediately after UV irradiation and directly stimulates nucleotide excision repair. *J. Biol. Chem.*, **277**, 1637–1640.
17. Otrin,V.R., Kuraoka,I., Nardo,T., McLenigan,M., Eker,A.P.M., Stefanini,M., Levine,A.S. and Wood,R.D. (1998) Relationship of the xeroderma pigmentosum group E DNA repair defect to the chromatin and DNA binding proteins UV-DDB and replication protein A. *Mol. Cell. Biol.*, **18**, 3182–3190.
18. Hwang,B.J., Ford,J.M., Hanawalt,P.C. and Chu,G. (1998) Expression of *p48* xeroderma pigmentosum gene is p53-dependent and is involved in global genomic repair. *Proc. Natl Acad. Sci. USA*, **96**, 424–428.
19. Itoh,T., Cado,D., Kamide,R. and Linn,S. (2004) *DDB2* gene disruption leads to skin tumors and resistance to apoptosis after exposure to ultraviolet light but not a chemical carcinogen. *Proc. Natl Acad. Sci. USA*, **101**, 2052–2057.
20. Yoon,T., Chakrabortty,A., Franks,R., Valli,T., Kiyokawa,H. and Raychaudhuri,P. (2005) Tumor-prone phenotype of the DDB2-deficient mice. *Oncogene*, **24**, 469–478.
21. Kiyohara,C., Otsu,A., Shirakawa,T., Fukuda,S. and Hopkin,J.M. (2002) Genetic polymorphisms and lung cancer susceptibility: a review. *Lung Cancer*, **37**, 241–256.
22. Goode,E.L., Ulrich,C.M. and Potter,J.D. (2002) Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **11**, 1513–1530.
23. Wei,Q., Cheng,L., Amos,C.I., Wang,L.E., Guo,Z., Hong,W.K. and Spitz,M.R. (2000) Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. *J. Natl. Cancer Inst.*, **92**, 1764–1772.
24. Carlson,C.S., Eberle,M.A., Rieder,M.J., Yi,Q., Kruglyak,L. and Nickerson,D.A. (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am. J. Hum. Genet.*, **74**, 106–120.
25. Gabriel,S.B., Schaffner,S.F., Nguyen,H. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science*, **296**, 2225–2259.
26. Stephens,M. and Donnelly,P. (2003) Comparison of Bayesian methods for haplotype reconstruction from population genotype data. *Am. J. Hum. Genet.*, **73**, 1162–1169.
27. Pritchard,J.K. and Przeworski,M. (2001) Linkage disequilibrium in humans: models and data. *Am. J. Hum. Genet.*, **69**, 1–14.
28. Hayes,S., Shiyanov,P., Chen,X. and Raychaudhuri,P. (1998) DDB, a putative DNA repair protein, can function as a transcriptional partner of E2F1. *Mol. Cell. Biol.*, **18**, 240–249.
29. Shiyanov,P., Hayes,S.A., Donepudi,M., Nichols,A.F., Linn,S., Slagle,B.L. and Raychaudhuri,P. (1999) The naturally occurring mutants of DDB are impaired in stimulating nuclear import of the p125 subunit and E2F1-activated transcription. *Mol. Cell. Biol.*, **19**, 4935–4943.
30. Datta,A., Bagchi,S., Nag,A., Shiyanov,P., Adami,G.R., Yoon,T. and Raychaudhuri,P. (2001) The p48 subunit of the damaged-DNA binding protein DDB associates with the CBP/p300 family of histone acetyltransferase. *Mutat. Res.*, **486**, 89–97.
31. Hartman,A.R. and Ford,J.M. (2002) BRCA1 induces DNA damage recognition factors and enhances nucleotide excision repair. *Nat. Genet.*, **32**, 180–184.
32. Itoh,T., O'Shea,C. and Linn,S. (2003) Impaired regulation of tumor suppressor p53 caused by mutations in the xeroderma pigmentosum DDB2 gene: mutual regulatory interactions between p48(DDB2) and p53. *Mol. Cell. Biol.*, **23**, 7540–7553.
33. de Vries,A., van Oostrom,C.T., Hofhuis,F.M. *et al.* (1995) Increased susceptibility to ultraviolet-B and carcinogens of mice lacking the DNA excision repair gene *XPA*. *Nature*, **377**, 169–173.
34. Wijnhoven,S.W., Kool,H.J., Mullenders,L.H., van Zeeland,A.A., Friedberg,E.C., van der HorstG,T., van Steeg,H. and Vrieling,H. (2000) Age-dependent spontaneous mutagenesis in *Xpc* mice defective in nucleotide excision repair. *Oncogene*, **19**, 5034-5037.
35. Alekseev,S., Kool,H., Rebel,H., Fousteri,M., Moser,J., Backendorf,C., de Gruijl,F.R., Vrieling,H. and Mullenders,L.H. (2005) Enhanced DDB2 expression protects mice from carcinogenic effects of chronic UV-B irradiation. *Cancer Res.*, **65**, 10298–10306.
36. Inoki,T., Yamagami,S., Inoki,Y., Tsuru,T., Hamamoto,T., Kagawa,Y., Mori,T. and Endo,H. (2004) Human DDB2 splicing variants are dominant negative inhibitors of UV-damaged DNA repair. *Biochem. Biophys. Res. Commun.*, **314**, 1036–1043.
37. Wu,X., Zhao,H., Amos,C.I., Shete,S., Makan,N., Hong,W.K., Kadlubar,F.F. and Spitz,M.R. (2002) p53 genotypes and haplotypes associated with lung cancer susceptibility and ethnicity. *J. Natl. Cancer Inst.*, **94**, 681–690.
38. Lee,K.M., Choi,J.Y., Park,S.K. *et al.* (2005) Genetic polymorphisms of ataxia telangiectasia mutated and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.*, **14**, 821–825.

Received November 8, 2005; revised January 3, 2006; accepted January 16, 2006