Published in final edited form as:
*Hum Mutat*. 2008 January ; 29(1): 113–122.

# A Variant of the Cockayne Syndrome B Gene *ERCC6* Confers Risk of Lung Cancer

**Zhongning Lin**[1,4], **Xuemei Zhang**[2], **Jingsheng Tuo**[3], **Yongli Guo**[2], **Bridgett Green**[1], **Chi-Chao Chan**[3], **Wen Tan**[2], **Ying Huang**[1], **Wenhua Ling**[4], **Fred F. Kadlubar**[5], **Dongxin Lin**[2,*], and **Baitang Ning**[1,*]

[1]*Division of Personalized Nutrition and Medicine, National Center for Toxicological Research, U.S. Food and Drug Administration, Jefferson, Arkansas* [2]*Department of Etiology and Carcinogenesis, Cancer Institute, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China* [3]*Laboratory of Immunology, National Eye Institute, National Institutes of Health, Bethesda, Maryland* [4]*School of Public Health, Sun Yat-sen University, Guangzhou, China* [5]*Department of Epidemiology, College of Public Health, University of Arkansas for Medical Sciences, Little Rock, Arkansas*

## Abstract

Cockayne syndrome B protein (ERCC6) plays an essential role in DNA repair. However, the Cockayne syndrome caused by the ERCC6 defect has not been linked to cancer predisposition; likely due to the fact that cells with severe disruption of the ERCC6 function are sensitive to lesion-induced apoptosis, thus reducing the chance of tumorigenesis. The biological function and cancer susceptibility of a common variant rs3793784:C>G (c.−6530C>G) in the ERCC6 was examined. We show that the c.−6530C allele has lower binding affinity of Sp1 by EMSA and displays a lower transcriptional activity in vitro and in vivo. We then examined the contribution of this polymorphism to the risk of lung cancer in a case–control study with 1,000 cases and 1,000 controls. The case–control analysis revealed a 1.76-fold (P = × $10^{-9}$) excess risk of developing lung cancer for the c.−6530CC carriers compared with noncarriers. The c.−6530CC interacts with smoking to intensify lung cancer risk, with the odds ratio (OR) = 9 for developing lung cancer among heavy smokers. Our data constituted strong evidence that ERCC6 rs3793784:C>G alters its transcriptional activity and may confer personalized susceptibility to lung cancer.

### Keywords

polymorphism; lung cancer; DNA repair; susceptibility; ERCC6

## Introduction

Cellular DNA is continuously under attack by both endogenous reactive species and exogenous mutagenic agents. Therefore, DNA repair systems are critical for the maintenance of genomic stability and integrity [Hoeijmakers, 2001; Wood et al., 2001]. A significant amount of evidence indicates that the decreased DNA repair capacity in humans is associated with

*Correspondence to: Dongxin Lin, M.D., Department of Etiology and Carcinogenesis, Cancer Hospital & Institute, Chinese Academy of Medical Sciences, Beijing 100021, China. Fax: (86)10-67722460. E-mail: dlin@public.bta.net.cn, Baitang Ning, Ph.D., 3900 NCTR Road, HFT-100, Jefferson, AR 72079. E-mail: baitang.ning@fda.hhs.gov.
Zhongning Lin and Xuemei Zhang contributed equally to this work.

Communicated by Daniel W. Nebert

increased rates of birth defects, cancer, aging, and reduced lifespan [Ronen and Glickman, 2001]. The normal expression and function of DNA repair proteins are therefore essential for cells to remove damage and thus prevent carcinogenesis.

ERCC6 (also known as Cockayne syndrome [CS] complementation group B [CSB]; MIM# 609413) is an important DNA repair protein, encoded by the *ERCC6* gene located on human chromosome 10q11 [Troelstra et al., 1992a]. ERCC6 was first reported as a member of a helicase-like protein family involved in Cockayne syndrome and DNA repair by Troelstra et al. [1992b]. Further studies revealed that ERCC6 participates in many cellular processes, including transcription-coupled DNA repair, transcription elongation, transcription regulation, global genome base excision repair (BER) and chromatin remodeling [Licht et al., 2003]. ERCC6 is involved in repairing various types of DNA damage, such as ultraviolet-induced dimers [Brosh et al., 1999; Orren et al., 1996; van der Horst et al., 1997], chemical carcinogen-induced bulky base adducts [Muftuoglu et al., 2002; van Oosterwijk et al., 1998; Wade and Chu, 1979]. It also functions as a BER protein involved in repairing some types of DNA oxidative damage in nuclei and mitochondria [Dianov et al., 1999; Osterod et al., 2002; Stevnsner et al., 2002; Tuo et al., 2003]. No predisposition to cancer is observed in Cockayne syndrome, which may be explained by the fact that the Cockayne syndrome cells are particularly sensitive to lesion-induced apoptosis, thereby preventing them from further tumorigenesis. However, it has been shown that ERCC6-disrupted mice had an increased susceptibility to skin cancer [van der Horst et al., 1997].

Individual differences in DNA repair capacity may affect an individual's susceptibility to DNA damage and, in turn, affect that individual's risk for developing cancer. Genetic polymorphism in DNA repair genes is one of the major reasons introducing individual differences in DNA repair capacity. Hundreds of polymorphisms in DNA repair genes have been identified and some of them have been consistently associated with cancer susceptibility [Hung et al., 2005; Spitz et al., 2003; Ribas et al., 2006]. For example, the *OGG1* p.S326C and the *XRCC1* p.R194W have been shown to be associated with an increased risk of various types of human cancer; while the *BRCA2* p.N372H was reported to be more specifically associated with an increased risk of breast cancer [Goode et al., 2002; Gu et al., 2005]. A number of SNPs were also identified in the *ERCC6* gene and a recent study reported that the *ERCC6* p.M1097V had a significant impact on bladder cancer recurrence [Gu et al., 2005].

Lung cancer is the leading cause of cancer-related death in the world and about 1.2 million deaths were reported annually [Parkin et al., 2005]. Tobacco smoke contains numerous carcinogens that can produce various types of DNA damage, which is believed to be the major mechanism underlying lung carcinogenesis [Hecht, 1999]. Because DNA repair pathways are vital in eliminating DNA damage and preventing carcinogenesis, genetic variation in these pathways has attracted much attention in researches on lung cancer susceptibility [Butkiewicz et al., 2001; Spitz et al., 2003, 2001; Zhang et al., 2005; Zhou et al., 2003]. ERCC6 is an important component of the DNA repair network; therefore, functional polymorphisms in the *ERCC6* gene resulting in altered expression and/or protein activity of ERCC6 could contribute to the altered susceptibility to lung cancer.

In the current study, we sought to determine whether the rs3793784:C>G polymorphism in the 5′-flanking region of the *ERCC6* gene has an effect on the expression of ERCC6 in vitro and in vivo. We also performed a case–control study in a Chinese population with a large sample size to examine the association between this genetic polymorphism and the risk of lung cancer.

## Materials and Methods

### Subjects for the Case–Control Study

This study recruited 1,000 lung cancer case patients and 1,000 control subjects. All subjects were unrelated ethnic Han Chinese. The characteristics of the study subjects have been described previously [Liang et al., 2003; Zhang et al., 2005]. Briefly, the case patients with newly diagnosed, histopathologically confirmed, and previously untreated (by radiotherapy or chemotherapy) primary lung cancer were recruited between January 1997 and June 2002 at Cancer Hospital, Chinese Academy of Medical Sciences (Beijing, China). There were no age, sex, stage, or histology restrictions; however, patients with previous malignancy or metastasized cancer from other organs were excluded. The control subjects were selected from a pool of cancer-free subjects recruited from a nutritional survey conducted in the same region during the same time period as the case patients were collected. The selection criteria include no prior history of malignancy, and control subjects were frequency-matched to the case patients by age (±5 years) and sex. At recruitment, informed consent was obtained from each subject, who was then interviewed for detailed information on demographic characteristics and lifetime history of tobacco use. The study was approved by the Institutional Review Board of the Cancer Institute and Hospital.

### *ERCC6* Genotyping

The genotyping of *ERCC6* rs3793784:C>G polymorphism (c.−6530C>G, located 6,530 bp upstream of the translation start site) was performed using PCR-based restriction fragment length polymorphism. ERCC6 genomic sequence with GenBank accession no. AY204752.1 was used. A 157-bp DNA fragment containing the *ERCC6* rs3793784:C>G site was amplified with the primer pairs of ERCC6F (5′-gggggaacagagaagcaggacagctAt-3′) and ERCC6R (5′-gccatgcgaatgtaaatcct-3′). A mismatch was introduced into the forward primer (capital A) to generate a *Nsi* restriction site. PCR was performed in a 10-μl reaction mixture containing 25 ng of DNA, 0.1 μM of each primer and 1 × JumpStart ReadyMix REDTaq (Sigma, Milwaukee, WI). The reaction was accomplished with a profile consisting of an initial melting step of 2 minutes at 94°C, followed by 34 cycles of 30 seconds at 94°C, 40 seconds at 60°C, 55 seconds at 72°C, and a final elongation step of 5 minutes at 72°C. Restriction enzyme *Nsi* was used to distinguish the genotype. The CC genotype had a single 157-bp band; the GG genotype had two bands, 126-bp and 31-bp, whereas the CG heterozygous genotype had all three bands, 157-bp, 126-bp, and 31-bp. The restriction products were separated on 2% agarose gels and visualized after ethidium bromide staining. Genotyping was performed without knowledge of the case/control status. A 10% random sample was tested in duplicate by different persons, and all results were 100% concordant. Eight samples of each genotype were confirmed by direct sequencing.

### Statistical Analysis

The $\chi^2$ test was used to compare the distribution of *ERCC6* genotypes between case patients and control subjects. The associations between genotype and risk of lung cancer were estimated by calculating the odds ratio (OR) and their 95% confidence interval (95% CI) with unconditional logistic regression models. The ORs were adjusted for age, sex, and pack-years smoked. Light, moderate, and heavy smokers were categorized by using the 25th and 75th percentile pack-year [(cigarettes per day/20) × (years smoked)] value of control subjects as the cutoff point (i.e., ≤ 16, 16–28, and>28 pack-years). We tested the null hypotheses of multiplicative gene-smoking interactions by evaluating departures from multiplicative joint effect models [Brennan, 2002]. A more than multiplicative joint effect was suggested when $OR_{11}>OR_{10} \times OR_{01}$. Departures from these multiplicative models were assessed by including main effect variables and their product terms in the logistic regression model. The homogeneity test was performed to compare the difference between smoking-related ORs among different

genotypes. We also analyzed the genotype-smoking interaction model, considering cumulative smoking exposure (square root of pack-years smoked) as a separate continuous variables using logistic regression model [Zhou et al., 2003]. All analyses were performed with the computer programs of the Statistical Analysis System (version 8.2; SAS Institute, Cary, NC).

## In Silico Analysis

The sequence flanking the SNP was screened for transcription factor binding sites. Web-based AliBaba2.1 (www.alibaba2.com) software was used for the analysis.

## Electrophoretic Mobility Shift Assay

Cells (H460, A549, and H226) were maintained in RPMI medium supplemented with 10% fetal bovine serum, 100 μg/ml penicillin, and 100 μg/ml streptomycin in a humidified, 5% $CO_2$ incubator at 37°C. Nuclear proteins were extracted from confluent cultures of H460 or A549 cells using a Nuclear Extraction Kit (Active Motif, Carlsbad, CA) according to the manufacturer's instructions. Infrared fluorescence dye IRD™ 800–labeled DNA probes were purchased from LI-COR (Lincoln, NE). The double-stranded oligonucleotides for the DNA probes corresponding to the−6530C or−6530G sequence from the *ERCC6* 5′-flanking region were 5′-gacagctcttccatccttcccg-3′ and 5′-gacagctcttgcatccttcccg-3′, respectively. The double-stranded oligonucleotide for the Sp1 probe was 5′-attcgatcggggcggggcgag-3′. Unlabeled oligonucleotides and the corresponding complementary oligonucleotides were ordered from Integrated DNA Technologies, Inc. (Coralville, IA). The binding reactions were carried out for 20 minutes at room temperature. Each 5-μl binding reaction mixture contained 50 mM Tris-HCl (pH 7.5), 25 mM KCl, 5 mM $MgCl_2$, 100 ng Poly(dI).(dC), 1 mM DTT, 0.1 pmol labeled probe, with or without 2 μg nuclear extracts. For the competition assays, 0- to 250-fold molar excess of unlabeled probes were added to the mixture before incubation. For supershift experiments, antibody against stimulatory protein 1 (Sp1) and rabbit IgG (Santa Cruz Biotechnology, Santa Cruz, CA) were used. The antibody or rabbit immunoglobulin G (IgG) (400 ng) was incubated with nuclear extracts at 4°C for 30 minutes, followed by an additional incubation for 20 minutes at room temperature with labeled Sp1 probe or c.−6530G probe. Aliquot of reaction mixtures were run on a nondenaturing 6% polyacrylamide/bis gel (18 cm) in a LI-COR $IR^2$ Global System (Lincoln, NE) with 0.5 × Tris-borate EDTA buffer, at 200–400 V (30 mA) for 3 hr.

## Construction of Reporter Plasmids immunoglobulin G

To evaluate promoter activity of different parts of the *ERCC6* 5′-flanking region, a serial deletion analysis of reporter constructs containing different lengths of the ERCC6 5′-flanking region was performed. Serial deletion fragments F1 (976 bp), F2 (2750 bp), F3 (7145 bp), and F4 (2124 bp) were cloned into the pGL3-Basic Luciferase Reporter Vector (Promega, Madison, WI). These fragments of the *ERCC6* 5′-flanking region were made by PCR amplification using primers flanked with restriction sites. The forward primer for amplifying F1, F2, and F3 was 5′-gaatacgcgttgggttggggccgctgacaggag-3′, in which the underlined bases represented the restriction enzyme *MluI* recognition site. The reverse primers for amplifying F1, F2, and F3 were 5′-gcctaagcttgccgccagccttggaaccc-3′, 5′-gctactcgaggaggctggggcaggag-3′, and 5′-gcctaaagcttcctccccatcaccaccactttcttgcttgattgcc-3′, respectively, in which the underlined bases were recognized by restriction enzyme HindIII (F1 and F3) or *Xho*I (F2). The F4 fragment was amplified using primer pairs of 5′-ccgggctagctgacctgccctgctctgttg-3′ and 5′-gcctaagcttcctccccatcaccaccactttcttgcttgattgcc-3′, which were flanked by a *Nhe*I site and a HindIII site, respectively. Two sets of human genomic DNA, CC homozygote and GG homozygote, which were confirmed by direct sequencing, were used as the templates for PCR. The cloned reporter constructs were confirmed by DNA sequencing.

### Transient Transfections, Luciferase Assay, and Ultraviolet Radiation

H460 cells and A549 cells were cultured as described above. Cell line 16HBE was cultured in MEM (GIBICO-Invitrogen, Carlsbad, CA) supplemented with 10% fetal bovine serum, 100 μg/ml penicillin, and 100 μg/ml streptomycin at 37 °C in a humidified, 5% $CO_2$ incubator. For transient transfection experiments, cells were plated on 24-well plates at a density of $2 \times 10^5$ per well with 500 μl of medium for 12 hr. Transfection was carried out using LipofectAMINE 2000 (Invitrogen, Carlsbad, CA) according to the manufacturer's instruction. Cells were cotransfected with 800 ng of reporter plasmid and 200 ng of pRL-TK (Promega, Madison, WI); the latter was used to standardize transfection efficiency.

H460 cells was used to analysis the luciferase activity induced by different lengths of the *ERCC6* 5′-flanking region and/or different ERCC6–6530 allele. A549 cells and 16HBE cells were transfected to verify the difference that the ERCC6 polymorphism influences the luciferase transcriptional activity after ultraviolet (UV) irradiation. After transfection for 4 hr, culture media were taken out and the A549 cells and 16HBE cells were treated with or without UV irradiation (10 $J/m^2$) for 5 minutes, then cells were cultured in correspondent media for 44 hr, luciferase activity was measured.

Luciferase activity was determined according to the manufacturer's protocol using a luciferase assay system (Promega). Briefly, after transfection for 48 hr, cells were scraped into lysis reagent, transferred to microcentrifuge tubes and centrifuged for 30 seconds at 12,000g. Luciferase activity was measured using a luminometer (Turner Biosystems, Sunnyvale, CA; TD $20/20^n$) by mixing 100 μl of luciferase assay reagent with 20 μl of 1:10-diluted cell lysate, and the value for each sample was recorded three times at 10-second intervals. For each plasmid construct, three independent transfection experiments were carried out, and each was done in triplicate. The promoter activity was reported as a relative light unit, calculated by defining the firefly activity of empty pGL3-Basic vector as 1. Statistical analysis of differences was carried out with the *t*-test for comparison of means. P of <0.01 was considered significant.

### Real-Time Analysis of *ERCC6* mRNA

A total of 50 (38 males and 12 females) normal lung tissues adjacent to the tumors were obtained from surgically removed specimens of individual patients. The normal tissues sampled at least 5 cm away from the margin of the tumor were immediately placed in liquid nitrogen and then stored at −80°C before analysis. Total RNA was isolated from tissues using the Trizol Reagent (Molecular Research Center, Inc., Cincinnati, OH) and converted to cDNA using an $oligo(dT)_{15}$ primer and Superscript II (Invitrogen). Relative gene expression quantitation for *ERCC6*, with β-*actin* as an internal reference gene, was carried out using the ABI Prism 7000 sequence detection system (Applied-Biosystems, Foster City, CA) in triplicate, based on the SYBR-Green method. The primers used for *ERCC6* were 5′-ctggaacagggagtgcttcag-3′ and 5′-actccttctccacgtcaacga-3′; and for β-*actin* were 5′-gccctgaggcactct tcca-3′ and 5′-cggatgtccacgtcacacttc-3′. The PCR reaction mixture consisted of 0.1 μM of each primer, 1 × SYBR® Premix EX Taq™ (Perfect Real Time) premix reagent (TaKaRa, Dalian, China), and 50 ng cDNA to a final volume of 20 μl. Cycling conditions were 2 minutes at 50°C and 10 minutes at 95°C, followed by 40 cycles of 15 seconds at 95°C and 1 minute at 60°C, and followed by 15 seconds at 95°C, 20 seconds at 60°C, and 15 seconds at 95°C. PCR specificity was confirmed by dissociation curve analysis and gel electrophoresis. All analysis was performed in a blinded fashion with the laboratory persons unaware of genotyping data. The expression of individual *ERCC6* measurements was calculated relative to expression of β-*actin* using a modification of the method described previously [Lehmann and Kreipe, 2001].

## Results

### Promoter Activity of the c.−6530C/G Encompassing Region

Promoter activity of serial deletion constructs derived from the 5′-flanking region of *ERCC6* was analyzed using H460 cells (human non-small-cell lung cancer cells). As shown in Fig. 1, the highest promoter activity was obtained from construct F3, which contained the longest 5′-flanking sequence (−7025→+121). Construct F2 (−7025→−4275) reserved almost similar promoter activity to construct F3, indicating that critical components of transcriptional cis-elements may be located in this region (−7025→−4275). With only 15% promoter activity of construct F3, construct F4 (−2003→+121) had the lowest promoter activity, suggesting that important elements for *ERCC6* transcription are not within this part of sequence. Construct F1 (−7025→−6040) had>50% activity of construct F3, indicating the presence of regulatory element(s) in this region, where the rs3793784:C>G SNP is located. The *ERCC6* promoter constructs exhibited similar activity patterns when transfected into human lung squamous carcinoma cell line H226 and human lung adenocarcinoma cell line A549 (data not shown), suggesting a common regulatory mechanism for the expression of *ERCC6* in these cell types.

### Differential Promoter Activity and UV Sensitivity of c.−6530C/G Alleles

To explore the possible impact of rs3793784:C>G SNP on *ERCC6* transcriptional activity, we generated six luciferase reporter gene constructs, representing the C or G allele presented in different lengths of the 5′-flanking region of *ERCC6*, and transfected them into H460 cells. As shown in Fig. 2, all constructs containing the G allele consistently had higher expression level of luciferase compared with constructs containing the C allele. Specifically, construct F1 containing the G allele had a two-fold higher promoter activity than construct F1 containing the C allele ($P<0.01$). Likewise, constructs F2 and F3 containing the G allele had a 1.7-fold increase in promoter activity compared with the C allele counterparts ($P<0.01$). Taken together, these data suggested that the rs3793784:C>G SNP had a significant effect on the *ERCC6* promoter activity, with the G allele containing promoter having greater expression level.

To investigate if these two alleles exhibit different sensitivities to UV irradiation, we transfected the pGL3b-ERCC6P-F3 with the C or G allele into human bronchial epithelial 16HBE cells and human lung adenocarcinoma A549 cells. Data indicated that the G allele and C allele responded to UV radiation differently. As shown in Fig. 3, the G allele–containing construct had a higher luciferase activity with UV radiation compared with its luciferase activity without UV radiation ($P<0.01$). However, in the C allele plasmid transfected cell lines, UV irradiation did not show a significant alternation in luciferase activity.

### Alteration of Nuclear Protein Binding Behavior

In silico analysis revealed that the *ERCC6* c.−6530C→G transversion altered the nucleotide c.−6530 encompassing fragment from an Sp1 binding element (C allele) to a potential binding element for Sp1, Oct1, and GATA1 (G allele). To validate this finding, electrophoretic mobility shift assays were performed. Nuclear protein extracts from H460 or A549 cells were incubated with infrared dye IRD™-labeled oligonucleotide probes containing c.−6530C or c.−6530G allele. A DNA-protein complex (complex A) was detected in both incubations containing the c.−6530G or c.−6530C probe; however, the amount of the complex formed by the c.−6530G probe was much greater than that formed by the c.−6530C probe (Fig. 4A, Lanes 2 and 7), suggesting that the c.−6530G allele has a higher affinity to bind nuclear protein(s). The difference in binding affinity of these two probes was further confirmed by competition assays. The DNA-protein complex in the incubation containing the c.−6530C probe was able to be completely inhibited by the addition of 100-fold molar excess of unlabeled c.−6530C probe (Fig. 4A, Lane 4), but it was not completely competed out by the same amount of unlabeled c.−6530G probe in the incubation containing the c.−6530G probe (Fig. 4A, Lane 9). The

competition assays also indicated that the DNA-protein binding is in a sequence-specific manner. Supershift assays were performed to confirm Sp1 binding using the c.−6530G probe in the presence of antibody against Sp1. A supershift band was detected in the incubations with 200 or 400 ng anti-Sp1 antibody (Fig. 4B). In contrast, no such band was found in assays without anti-Sp1 antibody. Furthermore, competition assays using an Sp1 consensus showed that the interaction of nuclear protein with infrared dye IRD™-labeled C allele or G allele probe could be competed out by the addition of 100-fold molar excess of unlabeled Sp1 probe (Fig. 4C, Lanes 4 and 8). In addition, infrared dye IRD™-labeled Sp1 probe interacted with nuclear protein extracts and produced a migrated band that was identical to that seen in assays with IRD™-labeled c.−6530C or c.−6530G probe (Fig. 4C, Lanes 10–12).

### *ERCC6* mRNA Levels in Lung Tissues from Different Genotype Carriers

The levels of *ERCC6* mRNA in individual lung tissues were quantitated by real-time PCR (Fig. 5). Of the 50 lung tissue samples, six failed to produce PCR products, probably due to degradation of mRNA in the tissue samples. Among the 44 valid samples, 20 were *ERCC6* c.−6530CC genotype, 24 were c.−6530CG genotype, and two were c.−6530GG genotype, which was combined with the CG genotype for analysis. It was found that the *ERCC6* c.−6530CC genotype carriers had significantly lower *ERCC6* mRNA levels (mean±standard error [SE]) than the *ERCC6* c.−6530CG or c.−6530GG genotype carriers (0.537±0.212 vs. 4.203±1.566; $P<0.05$).

### Association Between *ERCC6* Genotype and Lung Cancer Risk

We then examined the contribution of this polymorphism to the risk of lung cancer in a case–control study. The detailed characteristics of these 1,000 case patients and 1,000 control subjects have been described elsewhere [Liang et al., 2003; Zhang et al., 2005]. Specifically, in this case–control panel, the control subjects and case patients were well matched in sex and age. It was found that more smokers were represented in the case patients compared with the control subjects (65.0% vs. 51.1%; OR = 1.80, 95% CI = 1.50–2.25; $P<1.0 \times 10^{-7}$). Furthermore, the case patients had higher values of pack-years smoked than the control subjects; 65.1% of smokers among the case patients smoked>28 pack-years, whereas the value was 33.5% among control subjects ($P<1.0 \times 10^{-7}$). Of the 1000 case patients, 448 (44.8%) were diagnosed as squamous cell carcinoma, 297 (29.7%) as adenocarcinoma, and 255 (25.5%) as other types, including undifferentiated carcinoma (n = 90), bronchioalveolar carcinoma (n = 92), and mixed cell carcinoma (n = 73).

The frequencies of the major allele c.−6530C and minor allele c.−6530G were 0.755 and 0.245 in control subjects compared with 0.830 and 0.170 in case patients, respectively (Table 1). The distributions of observed genotypes among control subjects and case patients were in agreement with the Hardy-Weinberg equilibrium (P = 0.971 and 0.062, respectively). The frequencies of the c.−6530GG,−GC, and −CC genotype in case patients differed significantly from those in control subjects ($\chi^2 = 49.94$, $P <1.0 \times 10^{-7}$, df = 2), with the CC homozygotes being higher in case patients than in control subjects (70.6% vs. 57.2%; $P<1.0 \times 10^{-7}$). Because both CG and GG genotypes were more prevalent in control subjects and the CG heterozygous genotype also showed functional significance, these genotypes were combined for risk estimation. Multivariate logistic regression analysis revealed that the *ERCC6* c.−6530CC genotype carriers were at higher risk for developing overall lung cancer compared with noncarriers (OR = 1.76, 95% CI = 1.46–2.14, $P<1.0 \times 10^{-7}$). Stratified analyses showed that the c.−6530CC genotype was associated with increased risk of all three histological types of lung cancer (Table 1). Sex and age had no effect on the risk of lung cancer related to this polymorphism (data not shown).

### Effect of Gene-Smoking Interaction on Lung Cancer Risk

We further investigated whether there was a gene–environment interaction between the rs3793784:C>G SNP and smoking on risk of lung cancer (Table 2). We found that among nonsmokers the c.−6530CC genotype was associated with a significant increase in risk of developing lung cancer, with the OR = 1.52 (95% CI = 1.13–2.04, P = 0.0052) compared with the associated GG and CG genotype. However, among smokers the OR for the c.−6530CC genotype rose to 3.70 (95% CI = 2.63–5.20), which was statistically significantly higher than that for the associated GG and CG genotype (OR = 1.87, 95% CI = 1.48–3.32; P<0.0001, test for homogeneity). The OR (3.70) for the CC genotype carriers in smoking status was greater than the multiple product (1.87 × 1.52 = 2.84) of OR for smoking (1.87) and the OR for GG/CG genotype (1.52). These results indicated a supermultiplicative interaction between smoking and the ERCC6 c.−6530CC genotype in intensifying risk for developing lung cancer based on the multiplicative joint effect models [Brennan, 2002]. However, logistic regression analysis did not demonstrated the interaction terms between ERCC6 genotypes and the square root of pack-years smoked were statistically significant (CC vs. GG/GC, P = 0.344 for the interaction term) [Zhou et al., 2003]. When the interaction between smoking and the ERCC6 c.−6530CC genotype was analyzed with cumulative smoking dose (i.e., pack-years), a supermultiplicative interaction was observed in all three smoking levels. Among the c.−6530GG or CG genotype carriers, the ORs of developing lung cancer for smoking ≤ 16, 16–28, or>28 pack-years were 0.79 (95% CI = 0.45–1.40), 1.34 (95% CI = 0.77–2.32), and 4.42 (95% CI = 2.76–7.09); however, the corresponding ORs among the CC genotype carriers were 1.88 (95% CI = 1.23–2.04), 2.41 (95% CI = 1.53–3.81), and 8.87 (95% CI = 5.74–13.71). A homogeneity test showed that the all these differences in ORs between the GG or CG and CC genotype were statistically significant (P<0.01 or P<0.05, respectively; Table 2).

## Discussion

The *ERCC6* participates in several DNA repair pathways and other cellular processes [Licht et al., 2003] and cells with severely disrupted ERCC6 are hypersensitive to DNA-damaging agents [Brosh et al., 1999; de Waard et al., 2003; Muftuoglu et al., 2002; Orren et al., 1996; Tuo et al., 2001; van der Horst et al., 1997; van Oosterwijk et al., 1998; Wade and Chu, 1979], indicating a potential and possible important role of ERCC6 in maintaining genomic stability and perhaps preventing carcinogenesis. In the current study, our functional analysis of the *ERCC6* rs3793784:C>G SNP suggested that the association between the c.−6530CC genotype and increased risk of developing lung cancer vs. the c.−6530GG or CG genotype might be attributed to low expression of the CC genotype. Furthermore, we found that the G allele containing reporter construct showed an increased promoter activity after UV irradiation in 16HBE cells and A549 cells, compared with non-UV-treated cells (P<0.01), but the C allele containing the reporter construct did not respond to UV radiation. It further demonstrated that the *ERCC6* c.−6530GG and c.−6530CC polymorphism might contribute *ERCC6* with different sensitivities to UV irradiation. Although the SNP is positioned 6,530 bp upstream of the translation start codon, it is only 492 bp away from the transcriptional start site. Bioinformatics analysis implied that this portion of 5′-flanking region of the *ERCC6* is likely to be involved in gene transcription. Through electrophoretic mobility shift assays, the Sp1-DNA complex was detected as preferentially binding to the c.−6530G allele but weakly to the c.−6530C allele. Because Sp1 is a well-known transcriptional factor that upregulates a diversity of genes and is ubiquitously expressed in human cells [Cook et al., 1999; Kadonaga et al., 1987; Li et al., 2004; Turner and Crossley, 1999], it would be logical that an allele that has low affinity to bind Sp1 might have low transcriptional activity and thus a low level of mRNA or protein expression. In accordance with these results, our luciferase reporter gene assays consistently showed that reporter gene constructs bearing the c.−6530G allele have a higher promoter activity. Moreover, real-time analysis of *ERCC6* RNA in lung tissues demonstrated a

significant difference in transcriptional activity between the c.−6530C and the c.−6530G alleles in vivo. Yet, we can not absolutely exclude the possibility that other SNPs might be in linkage disequilibrium with rs3793784:C>G polymorphism and might be causative factors for the altered expression of *ERCC6* and increased susceptibility to lung cancer; however, our experiments have demonstrated, in vitro and in vivo, that the SNP rs3793784:C>G has distinctive functional impacts on the expression of the gene.

ERCC6 plays an important role in DNA repair and DNA metabolism; reduced expression of this gene over a lifetime would be anticipated to impair genome stability and integrity and hence be associated with higher susceptibility to tumor formation. One of the notable functions of ERCC6 is its involvement in the DNA transcription-coupled repair pathway that is the major pathway in humans for repairing DNA adducts induced by smoking-related carcinogens. Previous study has shown a 12.5% decrease in the baseline expression level in lung cancer patients compared with control subjects [Cheng et al., 2000]. It has also been reported that reduced expression of ERCC6 is associated with a more than two-fold increased risk of squamous cell carcinoma of the head and neck [Cheng et al., 2002]. This data is in agreement with our observation.

The notion that functional variation in *ERCC6* confers to the host predisposition to cancer formation does not conflict with the phenomenon that Cockayne syndrome caused by ERCC6 defect shows no cancer predisposition. Because in Cockayne syndrome, cells completely lacking the ERCC6 function are particularly sensitive to lesion-induced apoptosis [Lu et al., 2001] but do not affect chromosomal instability, which may reduce the chance of tumorigenesis. However, in cells where the expression of ERCC6 is subtly low but not null due to genetic polymorphism, DNA damage caused by endogenous reactive species or exogenous carcinogens would not be so severe as to induce cell apoptosis but would accumulate to finally initiate carcinogenesis. Low expression of ERCC6 could influence transcription-coupled repair, thus resulting in abnormal chromosomal instability, a crucial element in carcinogenesis. Further studies are needed to elucidate this important issue.

On the other hand, the phenomenon that CS patients do not get cancer is always confounded by the fact that CS mice do get cancer. We should consider that genomes between mice and humans are different. By comparisons of the cDNA database of humans and mice, Reed et al. [2003] revealed that humans have additional genes, not evident in mice, related to apoptosis, suggesting that there is a more complex and more precise surveillance system in humans. Furthermore, Murai et al. [2001] showed that, in rodents, global genome repair and transcription-coupled repair genes have important nonoverlapping functions, thus the insufficient repair capacity of ERCC6 may not be complemented by another NER, in turn to trigger tumorigenesis other than apoptosis.

In human cells, the balance between spontaneous/induced DNA damage and DNA repair is very well maintained by multiple DNA repair pathways. However, once the balance is uneven, such that DNA damage surpasses the DNA repair capacity, accumulation of multiple mutations may introduce carcinogenesis/tumorigenesis [Jackson et al., 2001]. Insufficient capacity of the DNA repair gene was evidently related to the increased susceptibility to cancers [Vaish, 2007; Gologan et al., 2005]. Therefore, decreased expression of ERCC6 by polymorphisms may cause a certain level of insufficient DNA repair that was not corrected by other DNA repair enzymes somehow, resulting in a higher susceptibility to lung cancer for these variant carriers.

The stratified analysis presented in this report suggests that the *ERCC6* rs3793784:C>G SNP interacts with tobacco smoking in intensifying the risk of lung cancer in a smoking dose–dependent manner. Strikingly, among subjects who smoked >28 pack-years, the *ERCC6* c.

−6530CC genotype rendered an OR of nearly 9 for developing lung cancer. However, different from the results of stratified analyses, the interaction between smoking the and ERCC6 variant did not statistically alter the risk of the lung cancer in the genotype–smoking interaction model. Further study should be targeted to the biological mechanism of that interaction. These results are consistent with many previous studies reporting gene–environment interactions between smoking and SNPs in other DNA repair genes [Brennan, 2002; Hu et al., 2005; Ito et al., 2004; Liang et al., 2003; Pfeifer et al., 2002; Zhang et al., 2005; Zhou et al., 2002, 2003] and further support our prior hypothesis that functional polymorphism in *ERCC6* may contribute to phenotypic variation in susceptibility to lung cancer. Tobacco smoke is an established major risk for lung cancer and DNA damage caused by tobacco carcinogens is believed to be an important mechanism underlying lung carcinogenesis [Hecht, 1999; Pfeifer et al., 2002; Spitz et al., 2003]. Therefore, it is biologically plausible to observe that subjects who carried the ERCC6 c.−6530C allele, which has a diminished function compared with the −6530G allele, and are exposed to extra carcinogens had an higher risk of lung cancer.

Although the study subjects recruited in this case–control study were from one hospital and might not be representative of the general population, the results of this study, which used incident case patients and had large sample size and significantly increased ORs with small P values, are unlikely to be due to selection bias or chances. The fact that genotype frequencies among control subjects and case patients are consistent with those derived from the Hardy-Weinberg equilibrium further supports the randomness of our subject selection. Moreover, the observed effect of the *ERCC6* rs3793784:C>G SNP on predisposition of lung cancer was not affected by other potential predictors of lung cancer risk such as age, sex, and smoking. Thus, it is unlikely that subject selection or unknown confounding factors could have biased our results in the case–control analysis.

In summary, we observed an association between the variant ERCC6 c.−6530C allele and the risk of lung cancer. Functional analyses showed that the variant c.−6530C allele contributes to significantly decreased expression of ERCC6 in vitro and in vivo in the target tissues. These data constitute strong evidence in support of the notion that the variant of *ERCC6* may confer an individual's susceptibility to lung cancer.

**Acknowledgements**

We acknowledge the Oak Ridge Institute for Science and Education (ORISE) and the Chinese State Key Basic Research and Development Program (grant number 2002CB512903) for supporting Zhongning Lin to participate in this project.

Grant sponsor: Oak Ridge Institute for Science and Education (ORISE); Grant sponsor: Chinese “State Key Basic Research Program”; Grant numbers: 2004CB518701; and 2002CB512903; Grant sponsor: U.S. Food and Drug Administration (FDA); Grant number: E0702101.

## References

Brennan P. Gene-environment interaction and aetiology of cancer: what does it mean and how can we measure it? Carcinogenesis 2002;23:381–387. [PubMed: 11895852]

Brosh RM Jr, Balajee AS, Selzer RR, Sunesen M, Proietti De Santis L, Bohr VA. The ATPase domain but not the acidic region of Cockayne syndrome group B gene product is essential for DNA repair. Mol Biol Cell 1999;10:3583–3594. [PubMed: 10564257]

Butkiewicz D, Rusin M, Enewold L, Shields PG, Chorazy M, Harris CC. Genetic polymorphisms in DNA repair genes and risk of lung cancer. Carcinogenesis 2001;22:593–597. [PubMed: 11285194]

Cheng L, Spitz MR, Hong WK, Wei Q. Reduced expression levels of nucleotide excision repair genes in lung cancer: a case–control analysis. Carcinogenesis 2000;21:1527–1530. [PubMed: 10910954]

Cheng L, Sturgis EM, Eicher SA, Spitz MR, Wei Q. Expression of nucleotide excision repair genes and the risk for squamous cell carcinoma of the head and neck. Cancer 2002;94:393–397. [PubMed: 11900225]

Cook T, Gebelein B, Urrutia R. Sp1 and its likes: biochemical and functional predictions for a growing family of zinc finger transcription factors. Ann NY Acad Sci 1999;880:94–102. [PubMed: 10415854]

de Waard H, de Wit J, Gorgels TG, van den Aardweg G, Andressoo JO, Vermeij M, van Steeg H, Hoeijmakers JH, van der Horst GT. Cell type–specific hypersensitivity to oxidative damage in CSB and XPA mice. DNA Repair (Amst) 2003;2:13–25. [PubMed: 12509265]

Dianov G, Bischoff C, Sunesen M, Bohr VA. Repair of 8-oxoguanine in DNA is deficient in Cockayne syndrome group B cells. Nucleic Acids Res 1999;27:1365–1368. [PubMed: 9973627]

Gologan A, Sepulveda AR. Microsatellite instability and DNA mismatch repair deficiency testing in hereditary and sporadic gastrointestinal cancers. Clin Lab Med 2005;25:179–196. [PubMed: 15749237]

Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. Cancer Epidemiol Biomarkers Prev 2002;11:1513–1530. [PubMed: 12496039]

Gu J, Zhao H, Dinney CP, Zhu Y, Leibovici D, Bermejo CE, Grossman HB, Wu X. Nucleotide excision repair gene polymorphisms and recurrence after treatment for superficial bladder cancer. Clin Cancer Res 2005;11:1408–1415. [PubMed: 15746040]

Hecht SS. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst 1999;91:1194–1210. [PubMed: 10413421]

Hoeijmakers JH. Genome maintenance mechanisms for preventing cancer. Nature 2001;411:366–374. [PubMed: 11357144]

Hu Z, Ma H, Lu D, Zhou J, Chen Y, Xu L, Zhu J, Huo X, Qian J, Wei Q, Shen H. A promoter polymorphism (−77T>C) of DNA repair gene XRCC1 is associated with risk of lung cancer in relation to tobacco smoking. Pharmacogenet Genomics 2005;15:457–463. [PubMed: 15970793]

Hung RJ, Hall J, Brennan P, Boffetta P. Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. Am J Epidemiol 2005;162:925–942. [PubMed: 16221808]

Ito H, Matsuo K, Hamajima N, Mitsudomi T, Sugiura T, Saito T, Yasue T, Lee KM, Kang D, Yoo KY, Sato S, Ueda R, Tajima K. Gene-environment interactions between the smoking habit and polymorphisms in the DNA repair genes, APE1 Asp148Glu and XRCC1 Arg399Gln, in Japanese lung cancer risk. Carcinogenesis 2004;25:1395–1401. [PubMed: 15044328]

Jackson AL, Loeb LA. The contribution of endogenous sources of DNA damage to the multiple mutations in cancer. Mutat Res 2001;477:7–21. [PubMed: 11376682]

Kadonaga JT, Carner KR, Masiarz FR, Tjian R. Isolation of cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. Cell 1987;51:1079–1090. [PubMed: 3319186]

Lehmann U, Kreipe H. Real-time PCR analysis of DNA and RNA extracted from formalin-fixed and paraffin-embedded biopsies. Methods 2001;25:409–418. [PubMed: 11846610]

Li L, He S, Sun JM, Davie JR. Gene regulation by Sp1 and Sp3. Biochem Cell Biol 2004;82:460–471. [PubMed: 15284899]

Liang G, Xing D, Miao X, Tan W, Yu C, Lu W, Lin D. Sequence variations in the DNA repair gene XPD and risk of lung cancer in a Chinese population. Int J Cancer 2003;105:669–673. [PubMed: 12740916]

Licht CL, Stevnsner T, Bohr VA. Cockayne syndrome group B cellular and biochemical functions. Am J Hum Genet 2003;73:1217–1239. [PubMed: 14639525]

Lu Y, Lian H, Sharma P, Schreiber-Agus N, Russell RG, Chin L, van der Horst GT, Bregman DB. Disruption of the Cockayne syndrome B gene impairs spontaneous tumorigenesis in cancer-predisposed Ink4a/ARF knockout mice. Mol Cell Biol 2001;21:1810–1818. [PubMed: 11238917]

Muftuoglu M, Selzer R, Tuo J, Brosh RM Jr, Bohr VA. Phenotypic consequences of mutations in the conserved motifs of the putative helicase domain of the human Cockayne syndrome group B gene. Gene 2002;283:27–40. [PubMed: 11867210]

Murai M, Enokido Y, Inamura N, Yoshino M, Nakatsu Y, van der Horst GT, Hoeijmakers JH, Tanaka K, Hatanaka H. Early postnatal ataxia and abnormal cerebellar development in mice lacking Xeroderma pigmentosum Group A and Cockayne syndrome Group B DNA repair genes. Proc Natl Acad Sci USA 2001;98:13379–13384. [PubMed: 11687625]

Orren DK, Dianov GL, Bohr VA. The human CSB (ERCC6) gene corrects the transcription-coupled repair defect in the CHO cell mutant UV61. Nucleic Acids Res 1996;24:3317–3322. [PubMed: 8811084]

Osterod M, Larsen E, Le Page F, Hengstler JG, Van Der Horst GT, Boiteux S, Klungland A, Epe B. A global DNA repair mechanism involving the Cockayne syndrome B (CSB) gene product can prevent the in vivo accumulation of endogenous oxidative DNA base damage. Oncogene 2002;21:8232–8239. [PubMed: 12447686]

Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA Cancer J Clin 2005;55:74–108. [PubMed: 15761078]

Pfeifer GP, Denissenko MF, Olivier M, Tretyakova N, Hecht SS, Hainaut P. Tobacco smoke carcinogens, DNA damage and p53 mutations in smoking-associated cancers. Oncogene 2002;21:7435–7451. [PubMed: 12379884]

Reed JC, Doctor K, Rojas A, Zapata JM, Stehlik C, Fiorentino L, Damiano J, Roth W, Matsuzawa S, Newman R, Takayama S, Marusawa H, Xu F, Salvesen G, Godzik A. RIKEN GER Group; GSL Members. Comparative analysis of apoptosis and inflammation genes of mice and humans. Genome Res 2003;13:1376–1388. [PubMed: 12819136]

Ribas G, González-Neira A, Salas A, Milne RL, Vega A, Carracedo B, González E, Barroso E, Fernández LP, Yankilevich P, Robledo M, Carracedo A, Benítez J. Evaluating HapMap SNP data transferability in a large-scale genotyping project involving 175 cancer-associated genes. Hum Genet 2006;118:669–679. [PubMed: 16323010]

Ronen A, Glickman BW. Human DNA repair genes. Environ Mol Mutagen 2001;37:241–283. [PubMed: 11317342]

Spitz MR, Wu X, Wang Y, Wang LE, Shete S, Amos CI, Guo Z, Lei L, Mohrenweiser H, Wei Q. Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. Cancer Res 2001;61:1354–1357. [PubMed: 11245433]

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;12:689–698. [PubMed: 12917198]

Stevnsner T, Nyaga S, de Souza-Pinto NC, van der Horst GT, Gorgels TG, Hogue BA, Thorslund T, Bohr VA. Mitochondrial repair of 8-oxoguanine is deficient in Cockayne syndrome group B. Oncogene 2002;21:8675–8682. [PubMed: 12483520]

Troelstra C, Landsvater RM, Wiegant J, van der Ploeg M, Viel G, Buys CH, Hoeijmakers JH. Localization of the nucleotide excision repair gene ERCC6 to human chromosome 10q11–q21. Genomics 1992a; 12:745–749. [PubMed: 1349298]

Troelstra C, van Gool A, de Wit J, Vermeulen W, Bootsma D, Hoeijmakers JH. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell 1992b;71:939–953. [PubMed: 1339317]

Tuo J, Muftuoglu M, Chen C, Jaruga P, Selzer RR, Brosh RM Jr, Rodriguez H, Dizdaroglu M, Bohr VA. The Cockayne Syndrome group B gene product is involved in general genome base excision repair of 8-hydroxyguanine in DNA. J Biol Chem 2001;276:45772–45779. [PubMed: 11581270]

Tuo J, Jaruga P, Rodriguez H, Bohr VA, Dizdaroglu M. Primary fibroblasts of Cockayne syndrome patients are defective in cellular repair of 8-hydroxyguanine and 8-hydroxyadenine resulting from oxidative stress. FASEB J 2003;17:668–674. [PubMed: 12665480]

Turner J, Crossley M. Mammalian Kruppel-like transcription factors: more than just a pretty finger. Trends Biochem Sci 1999;24:236–240. [PubMed: 10366853]

Vaish M. Mismatch repair deficiencies transforming stem cells into cancer stem cells and therapeutic implications. Mol Cancer 2007;26:26–34. [PubMed: 17407576]

van der Horst GT, van Steeg H, Berg RJ, van Gool AJ, de Wit J, Weeda G, Morreau H, Beems RB, van Kreijl CF, de Gruijl FR, Bootsma D, Hoeijmakers JH. Defective transcription-coupled repair in Cockayne syndrome B mice is associated with skin cancer predisposition. Cell 1997;89:425–435. [PubMed: 9150142]

van Oosterwijk MF, Filon R, de Groot AJ, van Zeeland AA, Mullenders LH. Lack of transcription-coupled repair of acetylaminofluorene DNA adducts in human fibroblasts contrasts their efficient inhibition of transcription. J Biol Chem 1998;273:13599–13604. [PubMed: 9593697]

Wade MH, Chu EH. Effects of DNA damaging agents on cultured fibroblasts derived from patients with Cockayne syndrome. Mutat Res 1979;59:49–60. [PubMed: 431551]

Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. Science 2001;291:1284–1289. [PubMed: 11181991]

Zhang X, Miao X, Liang G, Hao B, Wang Y, Tan W, Li Y, Guo Y, He F, Wei Q, Lin D. Polymorphisms in DNA base excision repair genes ADPRT and XRCC1 and risk of lung cancer. Cancer Res 2005;65:722–726. [PubMed: 15705867]

Zhou W, Liu G, Miller DP, Thurston SW, Xu LL, Wain JC, Lynch TJ, Su L, Christiani DC. Gene-environment interaction for the ERCC2 polymorphisms and cumulative cigarette smoking exposure in lung cancer. Cancer Res 2002;62:1377–1381. [PubMed: 11888908]

Zhou W, Liu G, Miller DP, Thurston SW, Xu LL, Wain JC, Lynch TJ, Su L, Christiani DC. Polymorphisms in the DNA repair genes XRCC1 and ERCC2, smoking, and lung cancer risk. Cancer Epidemiol Biomarkers Prev 2003;12:359–365. [PubMed: 12692111]

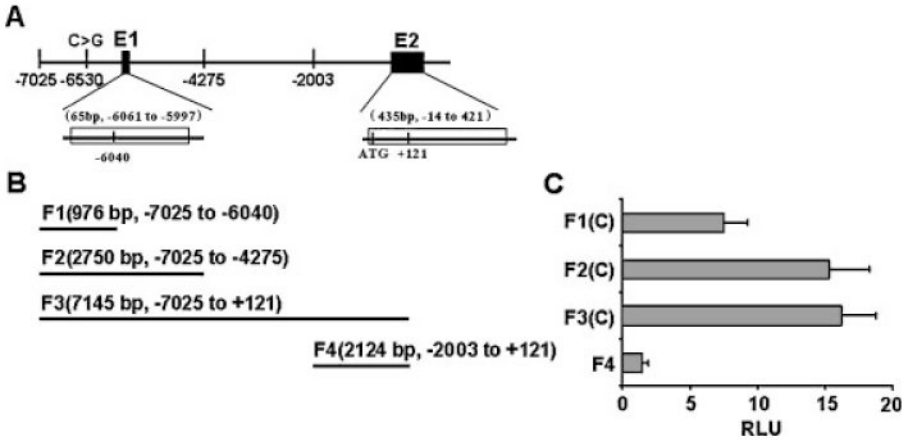


**Figure 1.**

Scheme of the 5′-flanking region of human *ERCC6* and promoter activity of fragments flanking the rs3793784:C>G SNP in *ERCC6*. **A:***ERCC6* 5′-untranslated region identified in the NCBI database. Boxes E1 and E2 represent exon 1 and exon 2. The SNP is located at c. −6530, which is 471 bp upstream the transcription start site located at c. −6061. The ERCC6 genomic reference sequence with GenBank accession no. AY204752.1 was used with the A of the translation initiation codon as nucleotide+1. Other numbers represent primer positions for cloning reporter constructs. **B:** Fragments F1–F4 were PCR-amplified to make reporter constructs, and their lengths and positions are shown in parentheses. **C:** Fragments F1–F4 with the C allele were cloned into the pGL3-Basic vector and transfected into H460 cells. Their activities were measured by dual luciferase assays. Results are expressed as fold increases in luciferase activity relative to the empty pGL3-Basic vector.

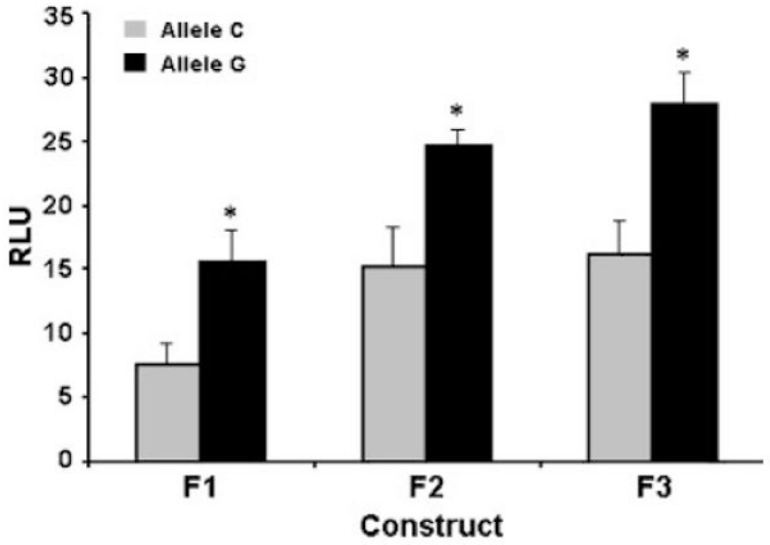


**Figure 2.**

Comparison of promoter activities between constructs with different allele. Fragments F1–F3 with the C or G allele were cloned into the pGL3-Basic vector and transfected into H460 cells. The lengths and positions of F1–F3 were the same as those in Fig.1. Their activities were measured by dual luciferase assays and the results are expressed as fold increases in luciferase activity relative to the empty pGL3-Basic vector. *Constructs with the G allele had higher activities compared with those with the C allele ($P<0.01$).

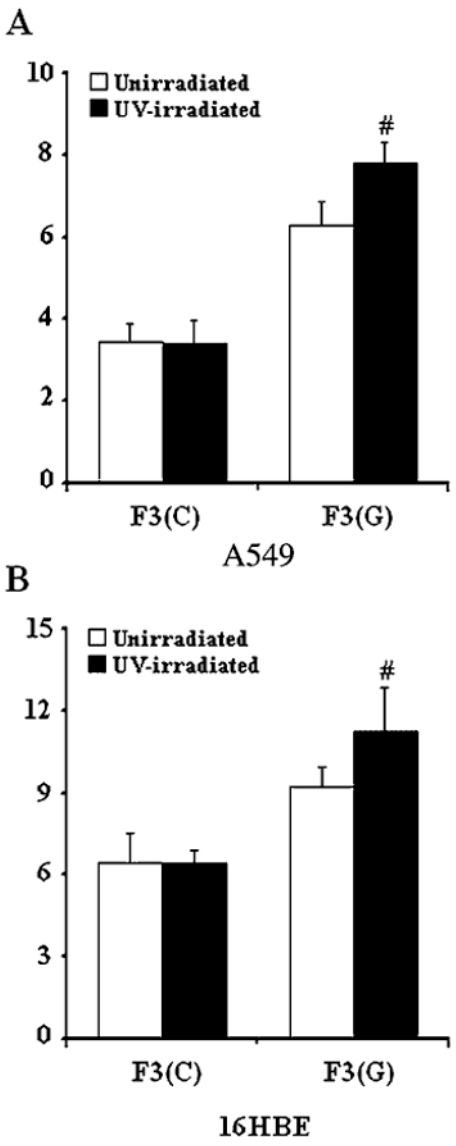


**Figure 3.**

UV irradiation impacts on allele-specific promoter activity. **A:** Comparison of promoter activities between C allele-containing construct and G allele-containing construct in A549 cells with or without UV irradiation. **B:** Comparison of promoter activities between C allele-containing construct and G allele-containing construct in 16HBE cells with or without UV irradiation. #After UV irradiation, constructs containing the −6530G allele induced higher luciferase expressions ($P<0.01$) in two cell lines.

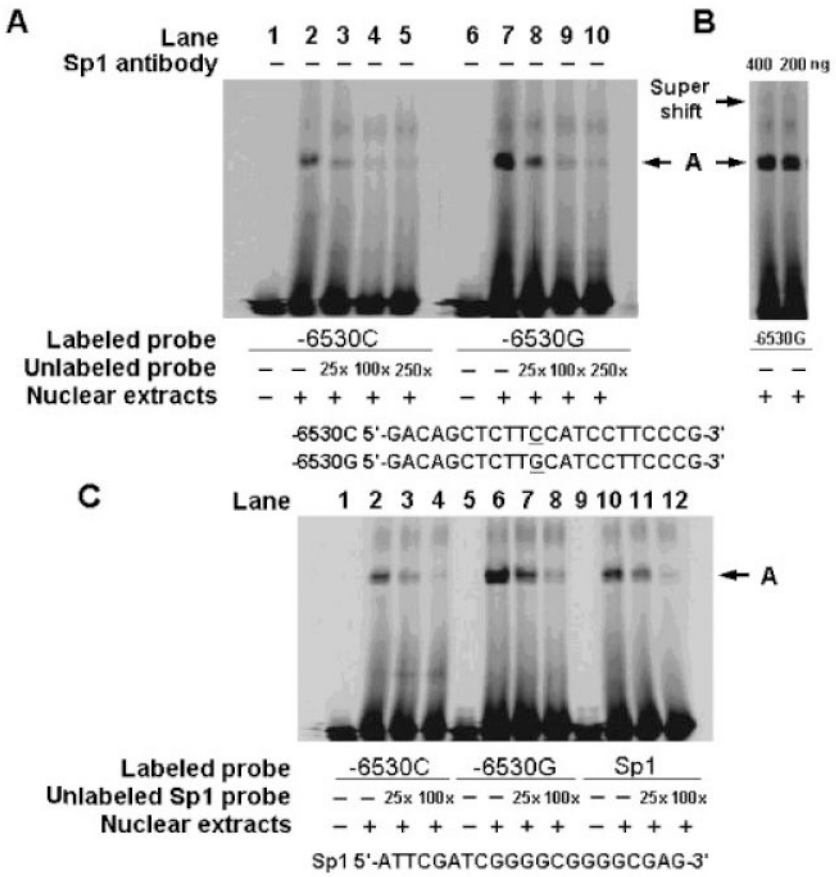


**Figure 4.**

Electrophoretic mobility shift assays using infrared dye IRD™-labeled probes and nuclear extracts from H460 cells. **A:** Nuclear extracts were incubated with infrared dye IRD™-labeled C allele probe or G allele probe. The DNA-protein complex (indicated by arrow A) showed density deference between the C and G allele. The competition assay was done by the addition of 25-, 100-, or 250-fold molar excess of unlabeled probes to the incubation mixtures. **B:** A faint supershift band was detected when 200 and 400 ng anti-Sp1 antibody was incubated with the nuclear extracts prior to the addition of the infrared dye IRD™-labeled G probe at room temperature. **C:** DNA-protein complex formed by incubation of IRD™-labeled C or G allele probe with nuclear extracts had the same pattern as that formed by incubation of IRD™-labeled Sp1 consensus probe with nuclear extracts. The formation of DNA-protein complex formed by the G or C probe was inhibited by 100-fold molar excess of unlabeled Sp1 probe.

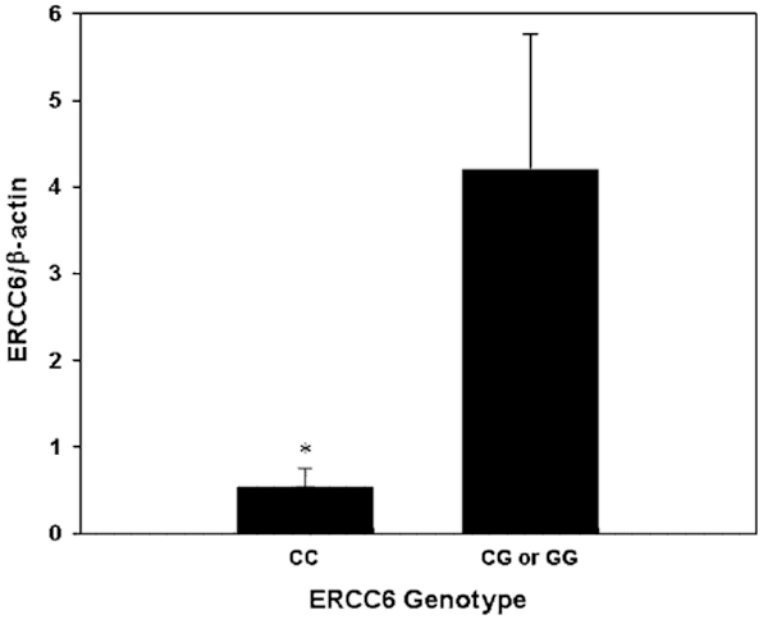


**Figure 5.**
Levels of *ERCC6* mRNA expression in lung tissues as a function of *ERCC6* genotype. Columns, mean; bars, ±SE normalized to β-actin. Expression level among the CC genotype was significantly lower than that among the CG or GG genotype; *P<0.05.

**Table 1**

Allele and Genotype Frequencies for *ERCC6* in 1,000 Case Patients and 1,000 Control Subjects and Their Association With the Risk of Lung Cancers*

| | Control subjects (n = 1,000, allele = 2,000) | | Total patients (n = 1,000, allele = 2,000) | | Patients with SCC (n = 448, allele = 896) | | Patients with AC (n = 297, allele = 594) | | Patients with other (n = 255, allele = 510) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | n | (%) | n | (%) | n | (%) | n | (%) | n | (%) |
| Allele frequency | | | | | | | | | | |
| C allele | 1,510 | (75.5) | 1,661 | (83.0) | 765 | (85.4) | 478 | (80.5) | 418 | (82.0) |
| G allele | 490 | (24.5) | 339 | (17.0) | 131 | (14.6) | 116 | (19.5) | 92 | (18.0) |
| Genotype frequency | | | | | | | | | | |
| GG | 62 | (6.2) | 45 | (4.5) | 17 | (3.8) | 16 | (5.4) | 12 | (4.7) |
| GC | 366 | (36.6) | 249 | (24.9) | 97 | (21.7) | 84 | (28.3) | 68 | (26.7) |
| CC | 572 | (57.2) | 706 | (70.6) | 334 | (74.5) | 197 | (66.3) | 175 | (68.6) |
| OR (95% CI)[a] | | | 1.76 (1.46–2.14) | | 2.26 (1.73–2.97) | | 1.40 (1.07–1.85) | | 1.59 (1.18–2.15) | |

*Data were calculated by unconditional logistic regression with the *ERCC6*-variant genotypes (GG or GC) as the reference group and adjusting for sex, age, and smoking status.

[a]Data were calculated by unconditional logistic regression and adjusted for sex, age, and smoking status.

SCC, Squamous cell carcinoma AC, Adenocarcinoma.

**Table 2**

Risk of Lung Cancer Associated With *ERCC6* Genotypes by Smoking Status Among 1,000 Case Patients and 1,000 Control Subjects

| Smoking status | *ERCC6* genotype | | | | | |
|---|---|---|---|---|---|---|
| | GG+GC[a] | OR(95%CI)[b] | P-value | CC[a] | OR (95% CI)[b] | P-value |
| Nonsmoker | 109/198 | 1.00 (reference) | | 241/291 | 1.52 (1.13–2.04) | 0.0052 |
| Smoker | 185/230 | 1.87 (1.48–3.32) | 0.0001 | 465/281 | 3.70 (2.63–5.20)[c] | $<1.0 \times 10^{-7}$ |
| ≤16 pack-years | 27/85 | 0.79 (0.45–1.40) | 0.4269 | 79/94 | 1.88 (1.23–2.86)[c] | 0.0034 |
| 16-28 pack-years | 35/69 | 1.34 (0.77–2.32) | 0.3026 | 86/92 | 2.41 (1.53–3.81)[d] | 0.0002 |
| >28 pack-years | 123/76 | 4.42 (2.76–7.09) | $<1.0 \times 10^{-7}$ | 300/95 | 8.87 (5.74–13.71)[c] | $<1.0 \times 10^{-7}$ |

[a] Number of patients/number of controls.

[b] Data were calculated by unconditional logistic regression and adjusted for sex and age.

[c] P < 0.01, test for homogeneity between smoking-related ORs among GG+CG and CC genotypes.

[d] P = 0.05, test for homogeneity between smoking-related ORs among GG+CG and CC genotypes.