

# Gene-Smoking Interaction Associations for the *ERCC1* Polymorphisms in the Risk of Lung Cancer

Wei Zhou,<sup>1</sup> Geoffrey Liu,<sup>1,3</sup> Sohee Park,<sup>2</sup> Zhaoxi Wang,<sup>1</sup> John C. Wain,<sup>4</sup> Thomas J. Lynch,<sup>3</sup> Li Su,<sup>1</sup> and David C. Christiani<sup>1,3</sup>

<sup>1</sup>Occupational Health Program, Departments of Environmental Health and <sup>2</sup>Biostatistics, Harvard School of Public Health; Departments of <sup>3</sup>Medicine and <sup>4</sup>Surgery, Massachusetts General Hospital, Boston, Massachusetts

## Abstract

Cigarette smoking may induce DNA damage. Lower DNA repair capacities have been associated with higher risk of lung cancer. Excision repair cross-complementing group 1 (*ERCC1*) is the lead enzyme in the nucleotide excision repair process, and low expression of *ERCC1* mRNA levels has been associated with higher risk of cancers. We examined the association between two polymorphisms of *ERCC1*, *8092C > A* (rs3212986) and *19007T > C* (codon 118, rs11615), which are associated with altered *ERCC1* mRNA stability and mRNA levels, in 1,752 Caucasian lung cancer patients and 1,358 controls. The results were analyzed using logistic regression models, adjusting for relevant covariates. The two polymorphisms were in Hardy-Weinberg disequilibrium and in linkage disequilibrium. There was no overall association between *ERCC1* polymorphisms and lung cancer risk, with the adjusted odds ratios (AOR) of 1.26 [95% confidence interval (95% CI), 0.81-1.96] for the

*8092C > A* polymorphism (*A/A* versus *C/C*) and 0.93 (95% CI, 0.67-1.30) for the *19007T > C* polymorphism (*C/C* versus *T/T*). Stratified analyses revealed that the AORs for the *8092C > A* polymorphism (*A/A* versus *C/C*) decreased significantly as pack-years increased, with the AOR of 2.11 (95% CI, 1.03-4.31) in never smokers and 0.50 (95% CI, 0.25-1.01) in heavy smokers ( $\geq 56$  pack-years), respectively. Consistent results were found when gene-smoking interaction was incorporated by joint effects and interactions models that considered both discrete and continuous variables for cumulative smoking exposure. The same direction for the gene-smoking interaction was found for the *19007T > C* polymorphism, although the interaction was not statistically significant. In conclusion, *ERCC1 8092C > A* polymorphism may modify the associations between cumulative cigarette smoking and lung cancer risk. (Cancer Epidemiol Biomarkers Prev 2005;14(2):491-6)

## Introduction

DNA repair and maintenance is essential in protecting the genome of the cell from environmental hazards such as tobacco smoke. Nucleotide excision repair (NER) is the major pathway in humans for repairing DNA damages induced by smoking-related carcinogens, such as benzo[*a*]pyrene diol epoxide adducts (1). This complex DNA repair process consists of ~30 proteins involved in sequential damage recognition, chromatin remodeling, incision of the damaged DNA strand on both sides of the lesion, excision of the oligonucleotide containing the damage, and gap-filling DNA synthesis followed by strand ligation (2). Excision repair cross-complementing group 1 (*ERCC1*), a highly conserved enzyme, is the lead component of the NER process and is absolutely required for the incision step of NER (3, 4). Among the proteins involved in NER, a defect in *ERCC1* seems to be associated with the most severe DNA repair deficiency (3). Low expression of *ERCC1* mRNA levels in peripheral blood lymphocytes have been associated with a statistically significant increased risk of squamous cell carcinoma of the head and neck and a nonsignificant higher risk of lung cancer (5, 6).

Two common polymorphisms of *ERCC1*, *8092C > A* (rs3212986) and *19007T > C* (codon 118, rs11615), have been reported. The *8092C > A* polymorphism is located in

the 3'-untranslated region of the gene and may affect *ERCC1* mRNA stability (7) and has been associated with significantly higher risk of adult-onset glioma (7) and nonsignificant higher risk of squamous cell carcinoma of the head and neck (8). The synonymous *19007T > C* polymorphism is associated with differential *ERCC1* mRNA levels (9, 10), and there is no report on the association between this polymorphism and the risk of cancers.

Impaired DNA repair capacity (DRC) may increase carcinogenesis and lead to the development of lung cancer; however, decreased DNA repair may contribute to the persistence of functional platinum-DNA adducts that confer antitumor activity. Therefore, DNA repair is regarded as a "double-edged sword" in cancer susceptibility and drug resistance (11). We and other groups have found statistically significant gene-smoking interactions between polymorphisms of DNA repair genes X-ray repair cross-complementing group 1 (*XRCC1*) and *ERCC2* and the risk of lung cancer (12-16). Associations between *ERCC1* polymorphisms and lung cancer risk have not been reported before. We hypothesized that the association between *ERCC1* polymorphisms and the risk of lung cancer is modified by cumulative smoking exposure. Drawing from a large sample, we tested this hypothesis using gene-smoking interaction analysis.

## Materials and Methods

**Study Population.** The study was approved by the Human Subjects Committees of Massachusetts General Hospital and the Harvard School of Public Health, Boston, MA. Details of this case-control population have been described previously (15, 17-19). In brief, all eligible cases (patients with histologically confirmed lung cancers) at Massachusetts General Hospital were recruited between December 1992 and April 2003. Before the year 1997, only early-stage (stages I and II)

Received 8/18/04; revised 9/23/04; accepted 9/28/04.

**Grant support:** NIH grants CA74386, CA90578, ES/CA 06409, and ES00002; Flight Attendants Medical Research Institute Young Clinical Scientist Award; Doris Duke Charitable Foundation; and Sue's Fund.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Note:** W. Zhou and G. Liu contributed equally to this work.

**Requests for reprints:** David C. Christiani, Occupational Health Program, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115. Phone: 617-432-3323; Fax: 617-432-6981. E-mail: dchristi@hsph.harvard.edu

Copyright © 2005 American Association for Cancer Research.

patients were recruited in this study; after 1997, all stages of lung cancer patients were recruited in this study. Controls were recruited among healthy friends and nonblood-related family members (usually spouses) of several groups of hospital patients: (a) patients with cancer, whether related or not related to a case or (b) patients with a cardiothoracic condition undergoing surgery (e.g., cardiac valvular disease, which typically afflicts patients with similar age and gender demographics as lung cancer patients). No matching was done. Importantly, none of the controls were themselves patients. Potential controls who carried a previous diagnosis of any cancer (other than nonmelanoma skin cancer) were excluded from participation. Over 85% eligible cases and over 90% controls participated in this study and provided blood samples. Interviewer-administered questionnaires collected information on demographic and detailed smoking histories from each subject.

**ERCC1 Genotyping.** DNA was extracted from peripheral blood samples using the Puregene DNA Isolation Kit (Gentra Systems, Minneapolis, MN). The *ERCC1 8092C > A* and *19007T > C* polymorphisms were genotyped by the 5' nuclease assay (Taqman) using the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Foster city, CA). The primers, probes, and reaction conditions were available upon request. Genotyping was done by laboratory personnel blinded to case-control status, and a random 5% of the samples were repeated to validate genotyping procedures. Two authors reviewed independently all genotyping results.

**Statistical Analysis.** Although individuals of all races were recruited for this study, we restricted our analyses to Caucasians (97%) to minimize confounding due to allele frequency variation by ethnicity. We analyzed all Caucasians with complete information on age, gender, smoking status (never smokers, ex-smokers, and current smokers), pack-years of smoking, and years since smoking cessation (for ex-smokers).

Detection of linkage disequilibrium between the two polymorphisms was based on Lewontin's  $D'$  in controls (20). Haplotypes of the two *ERCC1* polymorphisms were generated using the Partition Ligation-Expectation Maximization (PL-EM) version 1.0 (21, 22), which has been used in other association studies (23, 24). This software uses an efficient variant of the EM algorithm, to reconstruct individual probabilities for individual phasing accuracy based on unphased genotype data. At the same time, it also provides estimates on the overall haplotype frequencies as well as their SEs.

Analyses of all genotype and haplotype associations with lung cancer risk were based on logistic regression models (25). Logistic regression models were fit to examine the relationship between the log odds of lung cancer and each covariate, after adjusting for possible confounding factors such as age, gender, smoking status, pack-years of smoking, and years since smoking cessation (if ex-smoker). For gene-smoking interaction analyses, we used multiple approaches to evaluate consistency of results, including stratified analyses in specific categories of cumulative smoking exposure (i.e., pack-years), and genotype-smoking joint effects and interactions models that considered both discrete and continuous variables for cumulative smoking exposure. As suggested before (14, 15, 26, 27), square root of pack-years (SR-PY) and log transformed cigarettes per day were used in the analyses instead of the original untransformed variables. We fit the interactions between either the *8092C > A* or the *19007T > C* polymorphism or *ERCC1* haplotypes and SR-PY in separate gene-environment interaction models. A lack of fit test, as described in Hosmer and Lemeshow (25), was done to summarize the goodness-of-fit for each logistic regression model. Where

appropriate, odds ratio (OR) and 95% confidence interval (95% CI) for the risk of lung cancer were calculated from these models. Statistical analyses were all undertaken using SAS statistical packages (SAS Institute, Cary, NC).

## Results

**Population Characteristics.** There were no significant demographic differences (age and gender) between enrolled and unenrolled eligible cases and controls. A total of 3,345 of 3,371 (99%) enrolled subjects were genotyped successfully for both *ERCC1* polymorphisms. The distributions of race, gender, age, and smoking characteristics for those with genotype data was similar to the corresponding distributions observed for the entire study population. Complete information on age, gender, and detail smoking variables was available for 3,220 subjects (96%). We restricted our analysis to the 3,110 Caucasians with complete data (110 non-Caucasians were excluded). Of these, there were 1,752 lung cancer cases and 1,358 controls. There was 100% concordance of the 5% randomly repeated samples and 10% of data entry.

The distributions of demographic characteristics for cases and controls are summarized in Table 1. Compared with the controls, cases were older, had a higher proportion of males, more likely to be current smokers or heavy smokers, and had a shorter time since smoking cessation (if an ex-smoker) and larger pack-years of smoking. Adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and small-cell carcinoma represented 50%, 22%, 8%, and 9% of cases, respectively. Eleven percent were of mixed histologic subtype or had more than one primary tumor. Clinical American Joint Committee on Cancer stage data were available for 1,714 cases: 53% were early stage (I or II).

The distribution of smoking variables in our controls was similar to the general Massachusetts population over age 45 years (Massachusetts Tobacco Survey, Massachusetts Department of Public Health Publication, <http://www.state.ma.us/dph/mtcp/report/mats.htm>). The proportions of never smokers, ex-smokers, and current smokers were 35%, 46%, and 19% in our controls and 36%, 47%, and 17% in the general Massachusetts population over age 45 years, respectively. For current smokers, mean cigarettes per day (controls, 20 cigarettes; Massachusetts, 21 cigarettes) and earliest age of smoking (controls, 18 years; Massachusetts, 18) were similar. For ex-smokers, the proportions of those who have quit smoking for >5 years were 88% (controls) and 86% (Massachusetts).

**Distribution of ERCC1 Polymorphisms among Cases and Controls.** Both *ERCC1* polymorphisms in this control and case populations were consistent with Hardy-Weinberg equilibrium ( $P > 0.05$ ,  $\chi^2$  goodness-of-fit). The frequencies of C/C, C/A, and A/A genotypes for the *8092C > A* polymorphism were 56%, 37%, and 7% in cases, and 56%, 38%, and 6% in controls, respectively. The frequencies of T/T, C/T, and C/C genotypes for the *19007T > C* polymorphism were 37%, 47%, and 16% in both cases and controls. For the *19007T > C* polymorphism, although the T/T genotype generates the less commonly associated triplet codon sequence encoding the amino acid and has been termed the "variant" by convention, the T/T genotype indeed has higher frequencies in a number of studies (9, 10, 28, 29). Hence, the T/T genotype is used as reference group in all of our analyses. Genotype frequencies of the two polymorphisms were comparable to previous studies (7, 8, 29).

The two polymorphisms were in linkage disequilibrium ( $D' = 0.79$ ). Genotype concordances between the two polymorphisms were 71% in cases and 70% in controls (i.e., carriage of the A allele of *8092C > A* polymorphism correlated with carriage of the C allele of *19007T > C* polymorphism). *ERCC1*

**Table 1. Distribution of demographic characteristics among lung cancer cases and controls**

| Characteristics                                     | Case (n = 1,752) | Controls (n = 1,358) | P      |
|---|------------------|----------------------|--------|
|   | n (%)            | n (%)                |        |
| Age*, median (range)                                | 67 (26-91)       | 60 (27-100)          | <0.001 |
| Gender†, n (%)                                      |                  |                      |        |
| Male  | 922 (53%)        | 602 (44%)            | <0.001 |
| Female  | 830 (47%)        | 756 (56%)            |        |
| Smoking status†, n (%)                              |                  |                      |        |
| Never   | 135 (8%)         | 481 (35%)            | <0.001 |
| Ex-smoker   | 927 (53%)        | 618 (46%)            |        |
| Current smoker                                      | 690 (39%)        | 259 (19%)            |        |
| Pack-years*‡, median (range)                        | 51 (0.1-231)     | 25 (0.1-210)         | <0.001 |
| Cigarettes/d*‡, median (range)                      | 30 (0.1-120)     | 20 (0.1-100)         | <0.001 |
| Smoking duration*‡, median (range)                  | 40 (0.5-73)      | 26 (0.5-65)          | <0.001 |
| Years since quitting smoking*§, median (range)      | 12 (1-59)        | 18 (1-65)            | <0.001 |
| <i>ERCC1 8092C</i> > <i>A</i> polymorphism†, n (%)  |                  |                      |        |
| <i>A/A</i>  | 117 (7%)         | 86 (6%)              | 0.93   |
| <i>C/A</i>  | 658 (37%)        | 512 (38%)            |        |
| <i>C/C</i>  | 977 (56%)        | 760 (56%)            |        |
| <i>ERCC1 19007T</i> > <i>C</i> polymorphism†, n (%) |                  |                      |        |
| <i>C/C</i>  | 279 (16%)        | 213 (16%)            | 0.98   |
| <i>C/T</i>  | 819 (47%)        | 636 (47%)            |        |
| <i>T/T</i>  | 654 (37%)        | 509 (37%)            |        |

\*Tested by nonparametric Wilcoxon rank sum test.

†Cases and controls were compared using  $\chi^2$  tests.

‡Excludes individuals who have never smoked.

§Ex-smokers only.

haplotype frequencies are presented in Table 2. Applying Partition Ligation-Expectation Maximization algorithm, the posterior probabilities of individual haplotypes were all >99%; therefore, we assigned each individual haplotype with the highest posterior probability. There were no statistically significant differences in the distribution of genotype or assigned haplotype frequencies of cases and controls by different subgroups of gender, age, and histologic subtypes.

**Association between *ERCC1* Genotypes and Lung Cancer Risk.** There was no overall relationship between either *ERCC1* polymorphism and lung cancer risk. For the *8092C* > *A* polymorphism, the adjusted OR (AOR; 95% CI) that adjusted for age, gender, smoking status, SR-PY, and years since smoking cessation were 1.26 (95% CI, 0.81-1.96) for the *A/A* genotype and 1.00 (95% CI, 0.80-1.25) for the *C/A* genotype when each was compared with the *C/C* genotype. For the *19007T* > *C* polymorphism, the AORs were 0.93 (95% CI, 0.67-1.30) for the *C/C* genotype and 1.03 (95% CI, 0.83-1.29) for the *C/T* polymorphism, when each was compared with the *T/T* genotype. No association was found between either *ERCC1* polymorphism and lung cancer risk in different age or gender groups.

**Association between *ERCC1* Genotypes and Cumulative Cigarette Smoking in Lung Cancer Risk.** We first classified pack-years of smoking into discrete categories. Mild, moderate, and heavy smokers were stratified by tertiles of pack-years (divided at 30 and 56) in ever smokers of all participants. For the *ERCC1 8092C* > *A* polymorphism, the *A/A* genotype was a risk factor in nonsmokers, but a protective factor in heavy smokers (i.e., highest tertile of pack-years) when compared with the *C/C* genotype, with the AORs (95% CI) of 2.11 (1.03-4.31), 1.12 (0.61-2.05), 1.46 (0.73-2.93), and 0.50 (0.25-1.01) for never, mild, moderate, and heavy smokers, respectively.

Similar trends of the gene-smoking interactions were also suggested for the *ERCC1 19007T* > *C* polymorphism; however, none of the crude or AORs were statistically significant in each smoking subgroup. The AORs (95% CI) of *C/C* versus *T/T*

were 1.51 (0.86-2.65), 0.99 (0.63-1.54), 0.98 (0.61-1.55), and 0.76 (0.45-1.29) for never, mild, moderate, and heavy smokers, respectively.

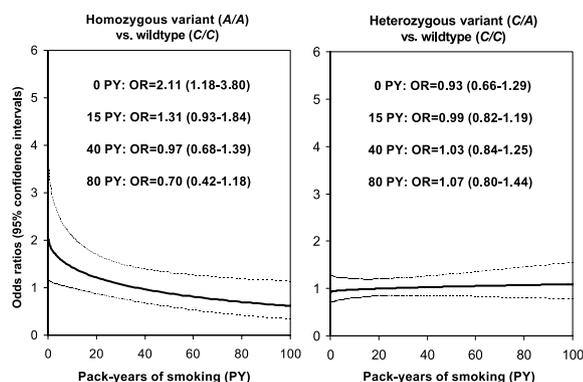
In addition to the stratified analysis by different smoking categories, we also investigated the joint effects of *ERCC1* polymorphisms and different levels of cigarette smoking. For the *8092C* > *A* polymorphism, the combined *C/C* genotype + heavy smoking has the highest AOR when compared with the reference group, the *C/C* genotype + never smoking (Table 3), which is consistent to the stratified analysis. Similar gene-smoking joint effects were also found for the *19007T* > *C* polymorphism.

We then evaluated a genotype-smoking interaction model that considered SR-PY as a continuous variable. The interaction terms between *ERCC1* genotypes and SR-PY were statistically significant for the *8092C* > *A* polymorphism (*A/A* versus *C/C*,  $P = 0.02$  for the interaction term; Fig. 1). Similar to the results of stratified analyses, the AORs of the *A/A* versus *C/C* genotypes decreased significantly as pack-years increased. No statistically significant interaction was found between the *19007T* > *C* polymorphism (*C/C* versus *T/T*) and SR-PY ( $P = 0.15$  for the interaction term).

**Table 2. Distribution of assumed *ERCC1* haplotypes among lung cancer cases and controls**

| <i>ERCC1</i> haplotypes*                        | Case (n = 3,504) |     | Controls (n = 2,716) |     |
|---|------------------|-----|----------------------|-----|
|   | Frequency (%)    | SE  | Frequency (%)        | SE  |
| Haplotype 1<br>( <i>8092C</i> + <i>19007T</i> ) | 60               | 0.8 | 60                   | 0.9 |
| Haplotype 2<br>( <i>8092C</i> + <i>19007C</i> ) | 15               | 0.6 | 15                   | 0.7 |
| Haplotype 3<br>( <i>8092A</i> + <i>19007T</i> ) | 1                | 0.1 | 1                    | 0.2 |
| Haplotype 4<br>( <i>8092A</i> + <i>19007C</i> ) | 24               | 0.7 | 24                   | 0.8 |

\*The sample size of haplotypes are based on number of chromosomes.



**Figure 1.** Adjusted interaction model evaluating *ERCC1 8092C > A* polymorphism and smoking exposure (continuous variable). The fitted ORs of the lung cancer risk of carrying *A/A* genotype (left) or *C/A* genotype (right), when each is compared individually to the *C/C* genotype, at various of pack-years. The logistic regression model included the following covariates: age, gender, square root of pack-years, smoking status, time since smoking cessation (in years), genotype, and interaction terms between homozygote and heterozygote variant genotypes and square root of pack-years.  $P = 0.02$  for interaction terms between genotype and square root of pack-years for the homozygote variant;  $P = 0.83$  for interaction terms between genotype and square root of pack-years for the heterozygote variant. ORs (95% CI) for never smokers (0 pack-years) and for 15, 40, and 80 pack-years, which represent the approximate midpoints for the mild (1-30 pack-years), moderate (31-56 pack-years), and heavy (>56 pack-years) smokers.

**Association between *ERCC1* Haplotypes and Cumulative Cigarette Smoking in Lung Cancer Risk.** In the haplotype analysis, two indicator variables were generated to identify subjects with two copies or a single copy of haplotype 4 (*8092A + 19007C*), where each group was compared with subjects with zero copy of haplotype 4. The frequencies of subjects assumed to be carrying two, one, and zero copies of haplotype 4 were 6%, 37%, and 57% in both cases and controls. There was no significant overall association between *ERCC1* haplotypes and lung cancer risk, with the AORs (95% CI) of 1.02 (0.85-1.22) and 1.16 (0.81-1.66) for subjects with 1 and 2 copies of haplotype 4, respectively. In the haplotype-smoking interaction analyses where SR-PY was treated as a continuous variable, statistically significant interactions were found for subjects with two copies of haplotype 4 ( $P < 0.01$  for the interaction term between

haplotype groups and SR-PY), with the AORs decreased significantly as pack-years increased. The AORs (95% CI) of two copies of haplotype 4 versus zero copy of haplotype 4 was 2.18 (1.21-3.95) for never smokers and 0.67 (0.40-1.14) for heavy smokers (80 pack-years).

We repeated interaction analyses adjusting for smoking variables in different ways. Pack-years was decomposed into its component parts of smoking intensity (mean number of cigarettes per day) and duration (in years). When either component (as a continuous variable) was substituted for pack-years in the logistic regression models, the gene-smoking interaction term was statistically significant between *ERCC1* haplotype (two copies of haplotype 4 versus zero copy of haplotype 4) and log-transformed cigarettes per day ( $P < 0.01$ ) or between haplotypes and years of smoking ( $P = 0.03$ ). Similar results were obtained in haplotype-smoking interaction models that incorporated different stratifications of pack-years of smoking, in place of a continuous variable for pack-years (data not presented).

## Discussion

Cigarette smoking may induce DNA damage and also stimulate DNA repair (30). Lung cancer patients have lower DRC when compared with healthy subjects (30, 31). Therefore, smoking may modify both lung cancer risk and the function of DNA repair genes, and polymorphisms that can alter DRC may lead to synergistic effects with smoking on lung cancer development. We tested this biological hypothesis by evaluating the statistical relationships between two polymorphisms of the DNA repair gene *ERCC1*, cumulative cigarette smoking, and the risk of lung cancer. We found a statistically significant interaction between cumulative cigarette smoking and *ERCC1 8092C > A* polymorphism (or *ERCC1* haplotypes) in its association with lung cancer risk.

Tobacco carcinogens may induce various types of DNA damages including benzo[*a*]pyrene diol epoxide adduct, strand breaks, cross-links, and recombination, which are repaired through different DNA repair pathways. In addition to its critical role in NER pathway, *ERCC1* is also involved in other DNA metabolism pathways (32). The *ERCC1-ERCC4* complex participate in the cleavage of damaged DNA strand 5' to the DNA lesion (33), and *ERCC1*-defective cells or knockout mice are highly sensitive to DNA cross-linking agents (34, 35). Cells from *ERCC1*-deficient mice show increased genomic instability and a reduced frequency of S-phase-dependent illegitimate chromosome exchange and signs of premature aging in addition to a repair deficient phenotype (35).

**Table 3. AORs (95% CI) for the joint effect of *ERCC1* polymorphisms and cumulative smoking exposure**

| Genotype                                | Smoking (case/control)* |                  |                    |                     |                  |
|---|-------------------------|------------------|--------------------|---------------------|------------------|
|   | Never (135/481)         | Mild (338/504)   | Moderate (560/240) | Heavy (719/138)     | Marginal         |
| <i>ERCC1 8092C &gt; A</i> polymorphism  |                         |                  |                    |                     |                  |
| <i>C/C</i>                              | 1.00                    | 4.24 (2.81-6.38) | 8.60 (5.84-12.66)  | 19.89 (13.16-30.07) | 1.0              |
| <i>C/A</i>                              | 0.91 (0.59-1.38)        | 4.39 (2.85-6.74) | 9.72 (6.42-14.72)  | 17.37 (11.12-27.13) | 1.00 (0.84-1.20) |
| <i>A/A</i>                              | 2.24 (1.08-4.62)        | 4.75 (2.42-9.32) | 12.92 (11.6-42.1)  | 9.94 (4.95-19.94)   | 1.17 (0.83-1.66) |
| Marginal                                | 1.00                    | 4.20 (2.99-5.88) | 8.97 (6.60-12.19)  | 17.46 (12.67-24.06) | —                |
| <i>ERCC1 19007T &gt; C</i> polymorphism |                         |                  |                    |                     |                  |
| <i>T/T</i>                              | 1.00                    | 4.72 (2.94-7.58) | 9.27 (5.86-14.68)  | 19.07 (11.75-30.95) | 1.0              |
| <i>C/T</i>                              | 1.00 (0.64-1.57)        | 4.45 (2.81-7.06) | 10.25 (6.56-16.01) | 20.95 (13.00-33.77) | 1.03 (0.85-1.23) |
| <i>C/C</i>                              | 1.53 (0.86-2.71)        | 4.38 (2.52-7.60) | 9.23 (5.41-15.74)  | 14.34 (8.14-25.27)  | 0.99 (0.77-1.27) |
| Marginal                                | 1.00                    | 4.21 (3.01-5.90) | 9.00 (6.62-12.24)  | 17.52 (12.71-24.15) | —                |

NOTE: Interaction models: Separate logistic regression model for the *ERCC1 8092C > A* and *19007T > C* polymorphisms included the following covariates: age, gender, three dummy variables for pack-years, smoking status, time since smoking cessation (in years), genotype groups, and interaction terms between genotype groups and the three dummy variables of smoking. Logistic regression models for marginal calculations included the above covariates, except for the interaction terms. The ORs in different smoking categories are for current smokers or nonsmokers only. For ex-smokers, the ORs for each comparison are ~25% less than for current smokers.

\*Mild, moderate, and heavy smokers correspond to the three tertiles of pack-years in ever smokers for all participants. The tertiles were divided at 30 and 56 pack-years.

Although no functional difference has been described for the 8092C > A polymorphism, this polymorphism has been suggested to be associated with *ERCC1* mRNA stability (7) and may therefore, be associated with lower DRC. In a case-control study of squamous cell carcinoma of the head and neck where most of the cases were heavy smokers, the C/C genotype of this polymorphism was also associated with a nonsignificant higher risk of squamous cell carcinoma of the head and neck (36), consistent with our results in heavy smokers. Inconsistent results have been reported for the function of 19007T > C polymorphism: one study found that ovarian cell lines containing the T/T genotype have decreased *ERCC1* mRNA induction after cisplatin exposure when compared with the C/C genotype cell line (9). However, another study found a trend toward higher intratumoral *ERCC1* mRNA levels with an increasing number of T alleles in a study of 31 advanced colorectal cancer patients (10). In our study, we did not find a statistically significant association between the 19007T > C polymorphism and the risk of lung cancer. In the haplotype analysis, although we found statistically significant interactions between *ERCC1* haplotypes and cumulative cigarette smoking, the major effect of the haplotypes is driven by the 8092C > A polymorphisms, given the high linkage disequilibrium between the two polymorphisms (Table 2).

The exact mechanism of the gene-smoking interaction association for the *ERCC1* 8092C > A polymorphism warrants further investigation. One possible explanation is that in lifetime never smokers, the A/A genotype that is associated with lower DRC may not repair DNA damage from environmental exposures effectively and is associated with higher DNA damage levels and in turn, a higher risk of lung cancer. However, cigarette smoking may stimulate DRC (30). In heavy smokers, high exposure to tobacco carcinogens may induce serious DNA damage and generate preneoplastic cells. It is possible that effective DNA repair may decrease the apoptosis rate of some of the preneoplastic cells. Therefore, in heavy smokers lower DRC may also be associated with a higher apoptosis rate of these preneoplastic cells, which may be one reason that we observed lower risk of lung cancer for the A/A genotype in heavy smokers.

Gene-smoking interactions for the associations between DNA repair gene polymorphisms and the risk of lung cancer have been observed consistently. Homozygous variant genotypes of the DNA repair genes *XRCC1* (*Gln/Gln* genotype of the *Arg399Gln* polymorphism) and *ERCC2* (*Asn/Asn* genotype of *Asp312Asn* polymorphism and *Gln/Gln* genotype of *Lys751Gln* polymorphism) that are associated with lower DRC have been shown to be consistently associated with higher risk of lung cancer in never smokers, and lower risk in heavy smokers, when compared with the corresponding wild genotypes (12-16). Similar gene-environment interactions have also been suggested between the *XRCC1 Arg399Gln* polymorphism and sunburn exposure in the risk of skin cancer (37). In addition, the *Gln/Gln* genotype of the *XRCC1 Arg399Gln* polymorphism and the *Gln/Gln* genotype of the *ERCC2 Lys751Gln* polymorphism are associated with higher DNA adduct levels in WBC in never smokers only, and not in smokers (38).

This study has several limitations. First, this is a hospital-based case-control study, where a subset of the controls included healthy spouses and friends of lung cancer cases. This subset of controls has a tendency to be more similar to cases than population controls because they may share similar health behaviors; thus, a potential bias may exist. However, bias in the estimate of the stratum specific OR due to a specific gene with strata defined by levels of measured confounders and effect modifiers (such as smoking, etc.) will only occur if, within these strata, this subset of controls is more likely than other types of controls to have a particular allele or alleles of the genotype under study; an unlikely scenario. Although we

did not match individually our controls to cases on age, gender, or smoking variables, we did adjust for these variables in the analyses. We found consistent gene-smoking interactions for the 8092C > A polymorphism, regardless of how we defined or categorized the smoking variables, or which model was used. The crude and AORs and adjusted joint effects ORs for different pack-years categories of smoking were similar in magnitude and direction to the point estimates obtained from fitted ORs of the interaction models (Fig 1). Similar associations were also found between different histologic cell types (adenocarcinoma and squamous cell carcinoma) and clinical stages of patients (early and late), when each group of patients were compared with controls (data not shown). Second, we only evaluated two polymorphisms of *ERCC1*, which have been suggested to be associated with *ERCC1* mRNA levels and DRC. It is possible that other polymorphisms of this gene or polymorphisms of other DNA repair genes including *ERCC4* may affect the association between these two polymorphisms and lung cancer risk, which may help to explain why we did not observe the significant association between the C/A genotype of the *ERCC1* 8092C > A polymorphism and the risk of lung cancer. However, adding other polymorphisms into the analysis (gene-gene interaction or gene-gene joint effects) will introduce three-way interactions and require a larger sample size. Third, although we adjusted for various smoking variables in our analysis, second hand smoke exposure, alcohol consumption, diet, environmental, and occupational exposure data were not adjusted in our logistic regression models because of incomplete and missing information. Given the strong associations between cigarette smoking and DNA damage and between smoking and lung cancer risks, these confounders will only likely had mild effects on the results of gene-smoking interactions, if any. Fourth, although we and other epidemiologic studies suggested statistically significant interactions between DNA repair polymorphisms and lung cancer risk, more biological background data are needed to explain our results, especially for the protective effect in heavy smokers of a polymorphism entailing a lower DRC. Last, we used Partition Ligation-Expectation Maximization program to infer the haplotype probabilities of each individual. Although the posterior probabilities of individual haplotypes were all >99%, inherit error still existed.

In conclusion, our results suggested statistically significant interactions between *ERCC1* genotypes or haplotypes and cumulative cigarette smoking in lung cancer risk. The A/A genotype of the 8092C > A polymorphism and specific haplotypes of the two polymorphisms (8092A + 19007C) are associated with higher lung cancer risks in never smokers, and lower lung cancer risks in heavy smokers. The results need to be confirmed in other independent studies, and further studies are needed to detect the function of these polymorphisms and its associations with cigarette smoking exposure.

## Acknowledgments

We thank the following staff members of the Lung Cancer Susceptibility Group: Barbara Bean, Jessica Shinn, Andrea Solomon, Thomas Van Geel, Lucy Ann Principe, Salvatore Mucci, Richard Rivera-Massa, David P. Miller; and the generous support of Dr. Panos Fidiias and the physicians and surgeons of the Massachusetts General Hospital Cancer Center.

## References

1. Cleaver JE. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* 1968;218:652-6.
2. Wood RD. DNA repair in eukaryotes. *Annu Rev Biochem* 1996;65:135-67.
3. Wilson MD, Ruttan CC, Koop BF, Glickman BW. *ERCC1*: a comparative genomic perspective. *Environ Mol Mutagen* 2001;38:209-15.
4. Sijbers AM, de Laat WL, Ariza RR, et al. Xeroderma pigmentosum group F caused by a defect in a structure-specific DNA repair endonuclease. *Cell* 1996;86:811-22.

5. 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;94:393-7.
6. 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;21:1527-30.
7. Chen P, Wiencke J, Aldape K, et al. Association of an ERCC1 polymorphism with adult-onset glioma. *Cancer Epidemiol Biomarkers Prev* 2000;9:843-7.
8. Sturgis EM, Dahlstrom KR, Spitz MR, Wei Q. DNA repair gene ERCC1 and ERCC2/XPD polymorphisms and risk of squamous cell carcinoma of the head and neck. *Arch Otolaryngol Head Neck Surg* 2002;128:1084-8.
9. Yu JJ, Lee KB, Mu C, et al. Comparison of two human ovarian carcinoma cell lines (A2780/CP70 and MCAS) that are equally resistant to platinum, but differ at 19007T>C of the ERCC1 gene. *Int J Oncol* 2000;16:555-60.
10. Park DJ, Stoehlmacher J, Zhang W, et al. ERCC1 polymorphism is associated with differential ERCC1 gene expression. *Proc AACR* 2002;93:1591.
11. Wei Q, Frazier ML, Levin B. DNA repair: a double-edged sword. *J Natl Cancer Inst* 2000;92:440-1.
12. David-Beabes GL, London SJ. Genetic polymorphism of XRCC1 and lung cancer risk among African-Americans and Caucasians. *Lung Cancer* 2001;34:333-9.
13. Park JY, Lee SY, Jeon HS, et al. Polymorphism of the DNA repair gene XRCC1 and risk of primary lung cancer. *Cancer Epidemiol Biomarkers Prev* 2002;11:23-7.
14. Zhou W, Liu G, Miller DP, et al. Gene-environment interaction for the ERCC2 polymorphisms and cumulative cigarette smoking exposure in lung cancer. *Cancer Res* 2002;62:1377-81.
15. Zhou W, Liu G, Miller DP, et al. Polymorphisms in the DNA repair genes XRCC1 and ERCC2, smoking, and lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 2003;12:359-65.
16. Misra RR, Ratnasinghe D, Tangrea JA, et al. Polymorphisms in the DNA repair genes XPD, XRCC1, XRCC3, and APE/ref-1, and the risk of lung cancer among male smokers in Finland. *Cancer Lett* 2003;191:171-8.
17. Xu LL, Wain JC, Miller DP, et al. The NAD(P)H:quinone oxidoreductase 1 gene polymorphism and lung cancer: differential susceptibility based on smoking behavior. *Cancer Epidemiol Biomarkers Prev* 2001;10:303-9.
18. Miller DP, Liu G, De Vivo I, et al. Combinations of the variant genotypes of GSTP1, GSTM1, and p53 are associated with an increased lung cancer risk. *Cancer Res* 2002;62:2819-23.
19. Miller DP, Neuberger D, De Vivo I, et al. Smoking and the risk of lung cancer: susceptibility with GSTP1 polymorphisms. *Epidemiology* 2003;14:545-51.
20. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* 1995;29:311-22.
21. Qin ZS, Niu T, Liu JS, Xu X. Partition-ligation-expectation-maximization algorithm for haplotype inference with single-nucleotide polymorphisms. *Am J Hum Genet* 2002;71:1242-7.
22. Niu T, Qin ZS, Xu X, Liu JS. Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms. *Am J Hum Genet* 2002;70:157-69.
23. Han J, Hankinson SE, De Vivo I, et al. A prospective study of XRCC1 haplotypes and their interaction with plasma carotenoids on breast cancer risk. *Cancer Res* 2003;63:8536-41.
24. Setiawan VW, Hankinson SE, Colditz GA, Hunter DJ, De Vivo I. HSD17B1 gene polymorphisms and risk of endometrial and breast cancer. *Cancer Epidemiol Biomarkers Prev* 2004;13:213-9.
25. Hosmer DW, Lemeshow S. *Applied Logistic Regression*. New York: John Wiley and Sons; 1989.
26. Zhou W, Thurston SW, Liu G, et al. The interaction between microsomal epoxide hydrolase polymorphisms and cumulative cigarette smoking in different histological subtypes of lung cancer. *Cancer Epidemiol Biomarkers Prev* 2001;10:461-6.
27. Zhou W, Liu G, Thurston SW, et al. Genetic polymorphisms in N-acetyltransferase-2 and microsomal epoxide hydrolase, cumulative cigarette smoking, and lung cancer. *Cancer Epidemiol Biomarkers Prev* 2002;11:15-21.
28. Park DJ, Zhang W, Stoehlmacher JT, et al. ERCC1 gene polymorphism as a predictor for clinical outcome in advanced colorectal cancer patients treated with platinum-based chemotherapy. *Clin Adv Hematol Oncol* 2003;1:162-6.
29. Winsey SL, Haldar NA, Marsh HP, et al. A variant within the DNA repair gene XRCC3 is associated with the development of melanoma skin cancer. *Cancer Res* 2000;60:5612-6.
30. Wei Q, Cheng L, Amos CI, et al. Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. *J Natl Cancer Inst* 2000;92:1764-72.
31. Wei Q, Cheng L, Hong WK, Spitz MR. Reduced DNA repair capacity in lung cancer patients. *Cancer Res* 1996;56:4103-7.
32. Sijbers AM, van der Spek PJ, Odijk H, et al. Mutational analysis of the human nucleotide excision repair gene ERCC1. *Nucleic Acids Res* 1996;24:3370-80.
33. Gaillard PH, Wood RD. Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair. *Nucleic Acids Res* 2001;29:872-9.
34. Westerveld A, Hoeijmakers JH, van Duin M, et al. Molecular cloning of a human DNA repair gene. *Nature* 1984;310:425-9.
35. Melton DW, Ketchen AM, Nunez F, et al. Cells from ERCC1-deficient mice show increased genome instability and a reduced frequency of S-phase-dependent illegitimate chromosome exchange but a normal frequency of homologous recombination. *J Cell Sci* 1998;111:395-404.
36. 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;94:393-7.
37. Han J, Hankinson SE, Colditz GA, Hunter DJ. Genetic variation in XRCC1, sun exposure, and risk of skin cancer. *Br J Cancer* 2004;91:1604-9.
38. Matullo G, Palli D, Peluso M, et al. XRCC1, XRCC3, XPD gene polymorphisms, smoking and (32)P-DNA adducts in a sample of healthy subjects. *Carcinogenesis* 2001;22:1437-45.