Gene-Environment Interaction for the ERCC2 Polymorphisms and Cumulative Cigarette Smoking Exposure in Lung Cancer

Wei Zhou, Geoffrey Liu, David P. Miller, Sally W. Thurston, Li Lian Xu, John C. Wain, Thomas J. Lynch, Li Su, and David C. Christiani

Abstract

Excision repair cross-complementing group 2 (ERCC2), a major DNA repair protein, is involved in nucleotide excision repair and basal transcription. The ERCC2 polymorphisms have been associated with altered DNA repair capacity. We investigated two ERCC2 polymorphisms, Asp312Asn and Lys751Gln, in 1092 Caucasian lung cancer patients and 1240 spouse and friend controls. The results were analyzed using generalized additive models and logistic regression, adjusting for relevant covariates. The overall adjusted odds ratios (ORs) and 95% confidence intervals (CIs) were 1.47 (95% CI, 1.1–2.0) for the Asp312Asn polymorphism (Asn/Asn versus Asp/Asp) and 1.06 (95% CI, 0.8–1.4) for the Lys751Gln polymorphism (Gln/Gln versus Lys/Lys). Gene-smoking interaction analyses revealed that the adjusted ORs for each of the two polymorphisms decreased significantly as pack-years increased. When comparing individuals with Asn/Asn + Gln/Gln versus individuals with Asp/Asp + Lys/Lys, the fitted ORs (95% CIs) were 2.56 (95% CI, 1.3–5.0) in nonsmokers and 0.69 (95% CI, 0.4–1.2) in heavy smokers (80 pack-years; P < 0.01 for the interaction term). Consistent and robust results were found when models incorporated different definitions of cumulative cigarette smoking. A stronger gene-smoking interaction was observed for the Asp312Asn polymorphism than for the Lys751Gln polymorphism. In conclusion, cumulative cigarette smoking modifies the associations between ERCC2 polymorphisms and lung cancer risk.

Introduction

DNA repair and maintenance is essential in protecting the genome of the cell from environmental hazards such as tobacco smoke. Reduced DRC can render a higher risk of developing many types of cancer, including lung cancer (1–4). ERCC2 (also known as XPD), a major DNA repair protein, is involved in transcription-coupled nucleotide excision repair and in the removal of a variety of structurally unrelated DNA lesions, including those induced by tobacco carcinogens (5, 6). Contradictory results are reported on the functional significance of the Asp312Asn (exon 10) and Lys751Gln (exon 23) polymorphisms of ERCC2 (2, 7–9). The Asp312Asn polymorphism (Asn allele) has been associated with higher risk of lung cancer in one study (10) and had a joint association with the Lys751Gln polymorphism on lung cancer risk in another (8). In both studies, the Asp312Asn and the Lys751Gln polymorphisms were in linkage disequilibrium (8, 10). The Lys751Gln polymorphism is reported to have no relationship with lung cancer risk in several studies (10–13).

Subsequently, one of these studies reported, in an expanded sample, that the Gln allele was associated with a higher trend for lung cancer (P > 0.05; Ref. 8).

Smoking can affect DNA repair capabilities (3, 14–16). We hypothesized the existence of an association between the ERCC2 polymorphisms, cumulative cigarette smoking, and lung cancer risk. There is precedence for these a priori hypotheses, because recently, we detected statistically significant gene-smoking interactions between microsomal epoxide hydrolase polymorphisms and lung cancer risk (17). Drawing from a large sample, we tested this hypothesis using gene-smoking interaction analysis and in analyses where the population was stratified by age, gender, clinical stage, and histological subgroups.

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. Details of this population have been described previously (17, 18). In brief, eligible cases (patients with histologically confirmed incident lung cancers) at Massachusetts General Hospital were recruited between December 1992 and December 2000. Controls were recruited first among the friends and non-blood-related family members of the lung cancer cases with no specific matching characteristics. In a small minority of cases, such individuals were not available, and controls were recruited among friends and non-blood-related family members of non-lung cancer patients admitted to the cardiothoracic wards. Interviewer-administered questionnaires collected information on demographic and detailed smoking histories from each subject.

ERCC2 Genotyping. DNA was extracted from peripheral blood samples using the Puregene DNA Isolation kit (Genta Systems, Minneapolis, MN). The ERCC2 Asp312Asn polymorphism was detected using a modified PCR-RFLP method involving published primer sequences (7). In brief, a 751-bp PCR product that included the Asp312Asn allele in exon 10 (codon 312) was amplified, followed by DpnII and Styl enzyme digestion (New England Biolabs, Beverly, MA).

The Lys751Gln polymorphism was detected using a modified PCR-RFLP method involving published primer sequences (16). In brief, a 184-bp PCR product that included the Lys/Gln allele in exon 23 (codon 751) was amplified, followed by MboII enzyme digestion (New England Biolabs).

For quality control, a random 5% of the samples were repeated. Two authors independently reviewed 100% of the agarose gels, data entry, and statistical analyses.

Statistical Analysis. Although individuals of all races were recruited for this study, we restricted our analyses to Caucasians to minimize confounding because of allele frequency variation by ethnicity. We analyzed all Caucasians with complete information on age, gender, smoking status (nonsmokers, ex-smokers, and current smokers), PYs of smoking, and years since smoking cessation (for ex-smokers). We used multiple approaches to evaluate consistency of results, including crude analyses in specific categories of cumulative smoking exposure (i.e., PYs) and genotype-smoking joint effects and interactions models that considered both discrete and continuous variables for cumulative smoking exposure.

A GAM (18, 19) was used to examine the relationship between the odds of lung cancer and each continuous covariate. The GAM extends the generalized linearized models framework, such as logistic regression, by allowing the relationship between the outcome and each covariate to be an unspecified but smooth function. Plots of the log odds of lung cancer versus the smooth function of

Received 6/13/01; accepted 12/28/01.

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.

This study was supported by NIH Grants CA74386 (to D. C. C), ES/CA 06409 (to D. C. C.), and ES00002 (to D. C. C.); Fogarty International Institute Training Grant TW00828 (to W. Z.); and a Noah Herndon Fellowship (to G. L.).

The abbreviations used are: DRC, DNA repair capacity; ERCC2, excision repair cross-complementing group 2; GAM, generalized additive model; OR, odds ratio; CI, confidence interval; PY, pack-year.
each covariate, in a model that adjusts for the other covariates, were created in S-plus (20) using GAM. The plots were examined for departures from linearity. If such departures were found, the plots were examined further to see if the shape of the relationship suggested that a parametric transformation of the covariate would be linearly related to the odds of lung cancer. Using this technique, the square root of PYs, log-transformed cigarettes/day, and the untransformed continuous variables of age, time since smoking cessation, and smoking duration (years) were all approximately linearly associated with the log odds of lung cancer probability in the GAM models.

Analyses of all genotype associations with lung cancer risk were based on logistic regression models (21). 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, the square root of PYs of smoking, and years since smoking cessation (if ex-smoker). We fit the interactions between either the Asp312Asn or the Lys751Gln polymorphism and square root of PYs in separate gene-environment interaction models. The interaction between smoking status and square root of PYs was also included in all gene-smoking interaction models, because it was found to be statistically significant in previous analyses of this population (17). Where appropriate, ORs and 95% CIs for the risk of lung cancer were calculated from these models. Statistical analyses were all undertaken using the S-plus (MathSoft, Inc., Cambridge, MA) and 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 (>87% participation rate) and controls (>90% participation rate). A total of 2575 (99.2%) of 2597 enrolled subjects were genotyped successfully for both ERCC2 polymorphisms. The distributions of race, gender, age, and smoking characteristics for those with genotype data were similar to the corresponding distributions observed for the entire study population. Complete information on age, gender, and smoking variables was available for 2419 subjects (93.1%). We restricted our analysis to the 2332 Caucasians with complete data. Of these, there were 1092 lung cancer cases and 1240 controls. There was 100% concordance of randomly repeated samples and data entry and 99.7% agreement in independent gel interpretation between two individuals (resolved by a third individual).

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 more PYs. Adenocarcinoma, squamous cell carcinoma, large cell carcinoma, and small cell carcinoma represented 51, 24, 8, and 9% of cases, respectively. Eight % were of mixed histological subtype or had more than one primary tumor. Clinical American Joint Committee on Cancer stage data were available for 1041 cases: 58% were early stage (I or II).

The distribution of smoking variables in our controls was similar to the general Massachusetts population over age 45.4 The proportions of nonsmokers, ex-smokers, and current smokers were 35.0, 45.5, and 19.5% in our controls and 36.0, 47.0, and 17.0% in the general Massachusetts population over age 45 years, respectively. For current smokers, mean cigarettes/day (controls, 21.2 cigarettes; Massachusetts, 21.4 cigarettes) and earliest age of smoking (controls, 17.9 years; Massachussets, 17.9) were similar. For ex-smokers, the proportions of those who have quit smoking for >5 years were 87.4% (controls) and 85.5% (Massachusetts).

**Distribution of ERCC2 Polymorphisms among Cases and Controls.** Both ERCC2 polymorphisms in this control population were consistent with Hardy-Weinberg equilibrium ($P > 0.05$; $\chi^2$ goodness of fit; Table 2). Genotype frequencies were comparable with previous studies (5, 8, 22). The two polymorphisms were in linkage disequilibrium ($P = 4.24 \times 10^{-170}$ in cases and $P = 1.24 \times 10^{-175}$ in controls by Fisher’s exact test; Ref. 10). Genotype concordances between the two polymorphisms were 75.4% in cases and 75.7% in controls [i.e., the Asp allele (Asp312Asn) correlated with the Lys allele (Lys751Gln)]. There were no statistically significant differences in genotype frequencies for either cases or controls in specific subgroups of gender and age, histological subtypes, or clinical stage.

**Association between ERCC2 Genotypes and Lung Cancer Risk.** There was an overall relationship between ERCC2 polymorphism and lung cancer risk for the Asp312Asn polymorphism but not for the

---

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

<table>
<thead>
<tr>
<th></th>
<th>Case (n = 1092)</th>
<th>Controls (n = 1240)</th>
<th>Crude ORs (95% CI)*&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (yr)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;55</td>
<td>205 (18.8)</td>
<td>471 (38.0)</td>
<td>1.00</td>
</tr>
<tr>
<td>55–64</td>
<td>294 (26.9)</td>
<td>324 (26.1)</td>
<td>2.09 (1.7–2.6)</td>
</tr>
<tr>
<td>≥65</td>
<td>593 (54.3)</td>
<td>445 (35.9)</td>
<td>3.06 (2.5–3.8)</td>
</tr>
<tr>
<td><strong>Mean ± SD</strong></td>
<td>64.9 ± 10.8</td>
<td>58.5 ± 12.4</td>
<td></td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>507 (46.4)</td>
<td>674 (54.4)</td>
<td>1.00</td>
</tr>
<tr>
<td>Male</td>
<td>585 (53.6)</td>
<td>566 (45.7)</td>
<td>1.37 (1.2–1.6)</td>
</tr>
<tr>
<td><strong>Smoking status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non</td>
<td>73 (6.7)</td>
<td>434 (35.0)</td>
<td>1.00</td>
</tr>
<tr>
<td>Ex</td>
<td>571 (52.3)</td>
<td>564 (45.3)</td>
<td>6.02 (4.6–7.9)</td>
</tr>
<tr>
<td>Current</td>
<td>448 (41.0)</td>
<td>242 (19.5)</td>
<td>11.01 (8.2–14.8)</td>
</tr>
<tr>
<td><strong>PYs of smoking</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>73 (6.7)</td>
<td>434 (35.0)</td>
<td>1.00</td>
</tr>
<tr>
<td>1–19</td>
<td>104 (9.5)</td>
<td>326 (26.3)</td>
<td>1.90 (1.4–2.6)</td>
</tr>
<tr>
<td>20–39</td>
<td>222 (20.3)</td>
<td>236 (19.0)</td>
<td>5.59 (4.1–7.6)</td>
</tr>
<tr>
<td>40–59</td>
<td>268 (24.5)</td>
<td>131 (10.6)</td>
<td>12.16 (8.8–16.8)</td>
</tr>
<tr>
<td>60–79</td>
<td>175 (16.0)</td>
<td>55 (4.4)</td>
<td>18.92 (12.8–28.0)</td>
</tr>
<tr>
<td>≥80</td>
<td>250 (22.9)</td>
<td>58 (4.7)</td>
<td>25.63 (17.6–37.4)</td>
</tr>
<tr>
<td><strong>Years since smoking cessation (ex-smokers only)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1–5</td>
<td>148 (25.9)</td>
<td>71 (12.6)</td>
<td>1.00</td>
</tr>
<tr>
<td>6–15</td>
<td>214 (37.5)</td>
<td>176 (31.2)</td>
<td>0.58 (0.4–0.8)</td>
</tr>
<tr>
<td>16–30</td>
<td>158 (27.7)</td>
<td>215 (38.1)</td>
<td>0.35 (0.3–0.5)</td>
</tr>
<tr>
<td>&gt;30</td>
<td>51 (8.9)</td>
<td>102 (18.1)</td>
<td>0.24 (0.2–0.4)</td>
</tr>
</tbody>
</table>

*Crude OR of cases versus controls for each characteristic category (e.g., age ≥65 years), compared with reference category (e.g., age <55 years).

Non, Ex, and Current correspond to nonsmokers, ex-smokers, and current smokers.

---

<sup>a</sup> Massachusetts Tobacco Survey, Massachusetts Department of Public Health Publication. Internet address: http://www.state.ma.us/dph/mtcp/report/mats.htm.
Lys751Gln polymorphism. When adjusted for age, gender, square root of PYs, smoking status, and years since smoking cessation, the ORs (95% CIs) for lung cancer risk were 1.00 (95% CI, 0.8–1.2) for the Asp/Asn genotype and 1.47 (95% CI, 1.1–2.0) for the Asn/Asn genotype (when each was compared with the Asp/Asp genotype). In a separate model, adjusted ORs were 0.98 (95% CI, 0.8–1.2) for the Lys/Gln genotype and 1.06 (95% CI, 0.8–1.4) for the Gln/Gln genotype (when each was compared with the Lys/Lys genotype). Similar associations between each polymorphism and lung cancer risk were found between younger (<55 years) and older (≥55 years) subjects and in different histological and clinical stage subgroups. We found a borderline statistically significant interaction (P = 0.06) between the Asp312Asn polymorphism (Asn/Asn versus Asp/Asp) and gender in the risk of lung cancer. For males, the adjusted ORs were 1.03 (95% CI, 0.8–1.4) and 1.96 (95% CI, 1.3–3.1) for the Asp/Asn genotype and Asn/Asn genotype, respectively. For females, the adjusted ORs were 0.95 (95% CI, 0.7–1.3) and 1.06 (95% CI, 0.7–1.7), respectively. For the Lys/Gln and Gln/Gln genotypes of the Lys751Gln polymorphism, the respective adjusted ORs were 1.09 (95% CI, 0.8–1.4) and 1.31 (95% CI, 0.9–2.0) for males and 0.84 (95% CI, 0.6–1.1) and 0.84 (95% CI, 0.5–1.3) for females (gender-genotype interactions: P > 0.05).

Association between ERCC2 Genotypes and Cumulative Cigarette Smoking in Lung Cancer Risk. We first classified smoking into discrete categories to avoid the issue of potential misclassification of subjects with respect to smoking exposure. The variant Asn (Asp312Asn) and Gln (Lys751Gln) alleles were individual risk factors in nonsmokers but protective factors in heavy smokers (i.e., highest tertile of PYs) when compared with their respective wild types (Asp/Asp or Lys/Lys) in both crude analyses (Table 2) and adjusted joint-effects analyses (Table 3). We also separately evaluated cases and controls to ensure that our results were being driven by associations in the cases instead of controls (23). In controls, the genotype frequencies of the two polymorphisms were similar in different discrete categories of PYs. As expected in the cases, the frequencies of the Asp/Asp genotype (Asp312Asn polymorphism) and the Lys/Lys genotype (Lys751Gln polymorphism, males only) were higher for individuals in heavier smoking exposure categories compared with light/non-smokers.

We then evaluated a genotype-smoking interaction model that considered square root PYs as a continuous variable. The interaction terms between ERCC2 genotypes and square root of PYs were statistically significant for the Asp312Asn polymorphism. The magnitude and statistical significance of the interaction term between genotype and square root of PY for the Asn/Asn versus Asp/Asp genotype were larger and stronger, respectively, than that for the Asp/Asp versus Asp/Asp comparison, suggesting a gene dose-response relationship for the number of Asn alleles (Fig. 1). Similar to the results of Tables 2 and 3, the adjusted ORs of the Asp/Asn versus Asp/Asp and Asn/Asn versus Asp/Asp genotypes decreased significantly as PYs increased. The magnitude and statistical significance of the Lys751Gln polymorphism-smoking interaction were smaller and weaker, respectively, when compared with the Asp312Asn polymorphism (Fig. 2).

We also assessed the combined effects of both polymorphisms in a
Asp312Asn polymorphism and smoking exposure (continuous variable). The fitted ORs represent the lung cancer risk of carrying Asn/Asn genotype (left) or Asp/Asn genotype (right), when each is compared individually with the Asp/Asp genotype at different numbers of PYs. The logistic regression model included the following covariates: age, gender, square root of PYs, smoking status, time since smoking cessation (in years), genotype, interaction terms between homozygote and heterozygote variant genotypes and square root of PYs, and interaction between smoking status and square root of PYs. The P for interaction terms between genotype and square root of PYs was <0.01 for the homozygote variant and 0.04 for the heterozygote variant. ORs and 95% CIs are presented for nonsmokers (0 PYs) and for 15, 40, and 80 PYs, which represent the approximate midpoints for the mild (1–25 PYs), moderate (26–55 PYs), and heavy (>55 PYs) smokers.

We investigated the effects of patient age, clinical stage, histological subtype, and gender on the gene-smoking relationship. For both polymorphisms, similar associations between ERCC2 genotypes and lung cancer risk were found among younger (<55 years of age) and older (≥55 years of age) subjects and when cases of different histological subtypes or disease stages were compared with all controls. For the Asp312Asn polymorphism, a gene-smoking interaction was observed in both males (P = 0.03 for Asp/Asn versus Asp/Asp comparison and P < 0.01 for Asp/Asn versus Asp/Asp comparison) and females (P = 0.92 and P < 0.01, respectively). For the Lys751Gln polymorphism, the statistically significant interaction between genotypes and square root of PY was observed only in males (P = 0.02 for Lys/Gln versus Lys/Lys comparison and P < 0.01 for Gln/Gln versus Lys/Lys comparison) but not in females (P > 0.05 for both genotype comparisons).

No biological conclusions can be made based on observed data, a limitation of our epidemiological study approach. However, we hypothesize that polymorphic variants of ERCC2 repair tobacco-associated DNA damage with differing efficiency in light and nonsmokers, whereas the DNA damage in heavy smokers is so great that differences in efficiencies by repair gene polymorphisms may not be as apparent. In both the crude (Table 2) and the adjusted models (Figs. 1 and 2), the ORs in nonsmokers and/or mild smokers most clearly excluded 1.0.

Similar to results in our moderate and heavy male smoker sub-
groups, the Asp/Asp genotype of the Asp312Asn polymorphism was associated with higher lung cancer risk when compared with the Asn allele in a study of 96 lung cancer cases, all smokers and predominantly male (10). As in our study, Spitz et al. (8) reported a joint effect between the Asn allele (Asp312Asn) and the Gln allele (Lys751Gln) polymorphism in lung cancer risk, with a stronger association involving the Asp312Asn polymorphism. The variant Asn allele was associated with lower levels of DRC in some studies (8) but not others (7). Contradictory results have also been observed for the Lys751Gln polymorphism. The Lys allele was associated with disease states and decreased DRC in several reports (2, 7, 22), whereas the Gln allele was associated with disease states and reduced DRC in other studies (5, 8, 24). Yet other studies report no overall relationship between the Lys751Gln polymorphism and lung cancer risk (10–13). Discrepant results may be explained, in part, by a gene-smoking interaction.

Males accounted for the majority of the increased lung cancer risk conferred by the homozygous variants of both ERCC2 polymorphisms in our study. Because males in our sample had a wider range of smoking exposures, perhaps differences in primary lung cancer risk were easier to detect. Yet even in the face of a stronger main lung cancer association, the interaction between the Lys751Gln polymorphism and cumulative cigarette smoking in lung cancer risk was observed mainly in males. Evidence indicates that males may have higher oxidative DNA damage, higher DRC, and altered DNA adduct levels in the lung as compared with females (3, 14–16) but may be less sensitive to the influence that age-related alterations in chromatin exert on DNA double-strand breaks (25). These differences may be clinically important because females appear to have higher lung cancer risk than males, given the same smoking exposure level, although this finding has been disputed (26–28). One can speculate that gender differences may be related partly to estrogen and other endogenous hormones or to lifestyle features that are gender specific (27, 28).

In conclusion, this is the first report of identified interactions between two ERCC2 polymorphisms (Asp312Asn and Lys751Gln) and cumulative cigarette smoking in lung cancer risk. We report a higher overall risk of lung cancer in individuals carrying the Asn/Asn genotype that is attributable primarily to non- and mild smokers. In heavy smoking individuals carrying the Asn/Asn genotype, there is a decreased lung cancer risk compared with the wild type, although the cancer risk conferred by primary smoking exposure was of a substantially greater magnitude than the cancer risk from variant ERCC2 genotypes. Nonetheless, a statistically significant gene-smoking interaction was still identified. More functional data on the ERCC2 polymorphic variants are needed to better understand the role of these variables in determining lung cancer risk.

ACKNOWLEDGMENTS

We thank the following staff members of the Lung Cancer Susceptibility Group: Barbara Bean, Jessica Shinn, Andrea Solomon, Linda Lineback, Lucy Ann Principe, Salvatore Mucci, Richard Rivera-Massa, Lisa I. Wang, Rong Fan, Yang Sai, Stephanie Shih, Maria Fragoso; and the generous support of Dr. Panos Fidas and the physicians and surgeons of the Massachusetts General Hospital Cancer Center.

REFERENCES

Gene-Environment Interaction for the ERCC2 Polymorphisms and Cumulative Cigarette Smoking Exposure in Lung Cancer

Wei Zhou, Geoffrey Liu, David P. Miller, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/62/5/1377

Cited articles
This article cites 24 articles, 8 of which you can access for free at:
http://cancerres.aacrjournals.org/content/62/5/1377.full#ref-list-1

Citing articles
This article has been cited by 19 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/62/5/1377.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.