Polymorphisms in the DNA Repair Genes *XRCC1* and *ERCC2*, Smoking, and Lung Cancer Risk[1]

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[2]

Occupational Health Program, Departments of Environmental Health [W. Z., G. L., D. P. M., L. L. X., L. S., D. C. C.] and Biostatistics [S. W. T.], Harvard School of Public Health, Boston, Massachusetts, 02115, and Hematology-Oncology [G. L., T. J. L.], Pulmonary and Critical Care Unit [D. C. C.], Department of Medicine, and Thoracic Surgery Unit, Department of Surgery [J. C. W.], Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114

Abstract

XRCC1 (X-ray cross-complementing group 1) and ERCC2 (excision repair cross-complementing group 2) are two major DNA repair proteins. Polymorphisms of these two genes have been associated with altered DNA repair capacity and cancer risk. We have described statistically significant interactions between the *ERCC2* polymorphisms (*Asp312Asn* and *Lys751Gln*) and smoking in lung cancer risk. In this case-control study of 1091 Caucasian lung cancer patients and 1240 controls, we explored the gene-environment interactions between the *XRCC1 Arg399Gln* polymorphism, alone or in combination with the two *ERCC2* polymorphisms, and cumulative smoking exposure in the development of lung cancer. The results were analyzed using logistic regression models, adjusting for relevant covariates. Overall, the adjusted odds ratio (OR) of *XRCC1 Arg399Gln* polymorphism (*Gln/Gln versus Arg/Arg*) was 1.3 [95% confidence interval (CI), 1.0–1.8]. Stratified analyses revealed that the ORs decreased as pack-years increased. For nonsmokers, the adjusted OR was 2.4 (95% CI, 1.2–5.0), whereas for heavy smokers (≥55 pack-years), the OR decreased to 0.5 (95% CI, 0.3–1.0). When the three polymorphisms were evaluated together, the adjusted ORs of the extreme genotype combinations of variant alleles (individuals with 5 or 6 variant alleles) *versus* wild genotype (individuals with 0 variant alleles) were 5.2 (95% CI, 1.7–16.6) for nonsmokers and 0.3 (95% CI, 0.1–0.8) for heavy smokers, respectively. Similar gene-smoking interaction associations were found when pack-years of smoking (or smoking duration and smoking intensity) was fitted as a continuous variable. In conclusion, cumulative cigarette smoking plays an important role in altering the direction and magnitude of the associations between the *XRCC1* and *ERCC2* polymorphisms and lung cancer risk.

Introduction

Tobacco smoke contains an array of potent chemical carcinogens and reactive oxygen species that may produce DNA bulky adducts, cross-links, oxidative or base DNA damage, and DNA strand breaks. Among the several major DNA repair pathways that operate on specific types of damaged DNA by cigarette smoking, base-excision repair is involved in repair of DNA base damage and single strand breaks, and NER[3] is involved in the repair of bulky monoadducts, cross-links, and oxidative damages (1–3).

XRCC1, 1 of >20 genes that participate in the base-excision repair pathway, has multiple roles in repairing DNA base damage and single-strand DNA breaks (4). Inconsistent results have been reported regarding the associations between the *Arg399Gln* (exon 10) polymorphism of *XRCC1* and either functional significance or the risk of tobacco associated cancers (5–16). The *Gln* allele of this polymorphism was associated with higher levels of DNA adducts (5, 11) and glycophorin A variants (5), increased sister chromatid exchange frequencies (6, 7), and higher sensitivity to ionizing radiation (8); however, two other studies found no association between this polymorphism and elevated DNA adduct levels (9, 10). *ERCC2* (also known as *XPD*), 1 of >20 genes that participate in the NER pathway, is involved in transcription-coupled NER and in the removal of a variety of structurally unrelated DNA lesions (17). Contradictory results are also reported on either the functional significance or lung cancer risk of the *Asp312Asn* (exon 10) and *Lys751Gln* (exon 23) polymorphisms of *ERCC2* (7, 10, 11, 14, 18–23).

Cigarette smoking may induce DNA damage, and individuals with a reduced DRC have a high level of carcinogen-DNA adducts in their tissues (24). Lung cancer patients may have lower DRC when compared with healthy subjects (19, 25). The profound role of cigarette smoking in both lung cancer development and DNA damage suggests that smoking may be more than a simple confounding variable. Recently, we described statistically significant gene-environment interactions between the *Asp312Asn* or *Lys751Gln* polymorphism of *ERCC2* and cumulative cigarette smoking in lung cancer risk, arguing for a possible biological interaction (26). Several other smaller studies found gene-smoking interactions for either *XRCC1* or

Received 8/14/02; revised 12/20/02; accepted 1/5/03.
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.
[1] Supported by NIH Grants CA74386, ES/CA 06409, and ES00002. W. Z. was supported by Fogarty International Institute Training Grant TW00828; G. L. by a Doris Duke Clinician Scientist Award; and D. P. M. by NIH Training Grant T3206709.
[2] To whom requests for reprints should be addressed, at Occupational Health Program, Harvard School of Public Health, 665 Huntington Avenue, Boston, MA 02115. Phone: (617) 432-3323; Fax: (617) 432-6981; E-mail: dchris@hohp.harvard.edu.

[3] The abbreviations used are: NER, nucleotide excision repair; XRCC1, X-ray cross-complementing group 1; ERCC2, excision repair cross-complementing group 2; DRC, DNA repair capacity; OR, odds ratio; CI, confidence interval.

ERCC2 polymorphisms in lung cancer risk (14–16, 23). Thus, the primary aim of this study was to determine whether cumulative smoking exposure altered the relationship between the *XRCC1 Arg399Gln* polymorphism and lung cancer risk in a gene-environment interaction analysis. If *XRCC1* genotypes were found to interact with cigarette smoking exposure, a secondary aim of this study was to determine the joint association between *XRCC1* (*Arg399Gln*) and *ERCC2* (*Asp312Asn* and *Lys751Gln*) polymorphisms and lung cancer risk, after accounting for potential interactions with cigarette smoking.

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 case-control population have been described previously (26–30). In brief, all of the histologically confirmed, newly diagnosed lung cancer patients at Massachusetts General Hospital were recruited between December 1992 and December 2000. Controls were recruited among friends and nonblood-related family members of the cases (usually spouses), with no specific matching characteristics. In some cases, if friends of lung cancer patients were not available, controls were recruited from friends and family of patients receiving thoracic surgery, chemotherapy, or radiation treatment for a condition other than lung cancer. Interviewer-administered questionnaires collected information on demographic and detailed smoking histories from each subject.

***XRCC1* and *ERCC2* Genotyping.** DNA was extracted from peripheral blood samples using the Puregene DNA Isolation kit (Gentra Systems, Minneapolis, MN). The *XRCC1 Arg399Gln* polymorphism was detected using modified PCR-restriction fragment length polymorphism methods, and published primer sequences (7). In brief, a 242-bp PCR product that included the *Arg/Gln* (A→C) allele in exon 10 (codon 399) was amplified, followed by *MspI* enzyme digestion (New England BioLabs, Beverly, MA).

The genotyping methods for the *ERCC2 Asp312Asp* (exon 10) and *Lys751Gln* (exon 23) polymorphisms have been described in detail (7, 26). Briefly, two separate PCR assays were used to detect the polymorphisms in exon 10 and exon 23 of *ERCC2* using published primer sequences. *DpnII* and *MspI* (for exon 10), and *Mbo II* enzyme (for exon 23) digestion (New England BioLabs) were used for restriction fragment analyses.

For quality control, a random 5% of the samples were repeated to assess the reproducibility of results. Two authors independently reviewed 100% of the agarose gels, with a third author arbitrating inconsistencies.

Statistical Analysis. Caucasian are 96% of the current samples. We analyzed all of the Caucasians with complete information on age, gender, smoking status (non-, ex-, and current smokers), pack-years of smoking, and years since smoking cessation (for ex-smokers). As described previously, generalized additive models suggested that analyses should incorporate square root of pack-years and log-transformed cigarettes per day in place of their untransformed values where appropriate (26–29, 31).

Analyses of all of the genotype associations with lung cancer risk were based on logistic regression models (26–29, 32). Logistic regression models were fit to examine the relationship between the log odds of lung cancer, and each covariate in the whole population and in different subgroups, after adjusting for the following factors: age, gender, smoking status, square root of pack-years of smoking, and years since smoking cessation (if ex-smoker). Years since smoking cessation were defined as zero for both current and nonsmokers.

In the gene-smoking interaction analysis, we used multiple approaches to evaluate consistency of results, including crude and adjusted analyses in specific categories of cumulative smoking exposure (*i.e.*, pack-years), and genotype-smoking interactions models that considered separate continuous variables for cumulative smoking exposure (*i.e.*, square root of pack-years, smoking duration, and log-transformed cigarettes per day). We fit the interactions between the *XRCC1* polymorphism and square root of pack-years, or various genotype combinations of *XRCC1* and *ERCC2* polymorphisms and square root of pack-years in separate gene-environment interaction models. The interaction between smoking status and square root of pack-years was also included in all of the gene-smoking interaction models when using the square root of pack-years as a continuous variable, because it was found to be statistically significant in previous analyses of this population (26, 28, 29). Where appropriate, OR and 95% CI for the risk of lung cancer were calculated from these models. Statistical analyses were all undertaken using the SAS statistical packages (SAS Institute, Cary, NC).

Results

Baseline 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 2574 (99.2%) of 2597 enrolled subjects were genotyped successfully for all of the *XRCC1* (*Arg399Gln*) and *ERCC2* (*Asp312Asn* and *Lys751Gln*) polymorphisms. We restricted our analysis to the 2331 Caucasians with complete data on age, gender, and smoking variables. Of these, there were 1091 lung cancer cases and 1240 controls. There was 100% concordance of randomly repeated samples and 99.7% agreement in independent gel interpretation between two individuals.

Detailed demographic characteristics including cigarette smoking history of cases and controls have been reported previously (26). Compared with control subjects, cases were older, more likely to be male, more likely to be current smokers, accrued significantly greater pack-years of smoking, and had fewer years since smoking cessation. The mean age (and SD) of all of the cases was 64.9 ± 10.8 years (range, 26–91), and of controls, 58.5 ± 12.4 years (range, 19–100). Females represented 46.4% of cases and 54.4% of controls. Among cases, the mean pack-years of ex-smokers was 55.3 ± 36.0, and of current smokers, 64.2 ± 35.8; among controls, the numbers were 29.3 ± 27.8 and 38.1 ± 25.2 respectively. The years since smoking cessation for ex-smokers were 14.2 ± 11.0 and 19.4 ± 11.9 for cases and controls, respectively.

The distribution of smoking variables in our controls was similar to the general Massachusetts population over age 45.[4] The proportions of non-, ex-, 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 per day (controls: 21.2 cigarettes; Massachusetts: 21.4 cigarettes) and earliest age of smoking (controls: 17.9 years; Massachusetts: 17.9) were similar. For ex-smokers, the proportions of those who

[4] Massachusetts Tobacco Survey, Massachusetts Department of Public Health Publication. Internet address: http://www.state.ma.us/dph/mtcp/report/mats.htm.

Table 1 Genotype frequencies and ORs (95% CIs) of *XRCC1 Arg399Gln* polymorphism for different pack-years of smoking[a]

	n (case/control)	Percentage (case/control)			Crude ORs[b]		Adjusted ORs[b]	
		Arg/Arg	*Arg/Gln*	*Gln/Gln*	*Arg/Gln*	*Gln/Gln*	*Arg/Gln*	*Gln/Gln*
Total	1091/1240	42.8/44.4	42.9/44.0	14.3/11.5	1.0 (0.8–1.2)	1.3 (1.0–1.7)	1.0 (0.8–1.3)	1.3 (1.0–1.8)
Non-	73/434	35.6/44.0	43.8/45.4	20.6/10.6	1.2 (0.7–2.1)	2.4 (1.2–4.9)	1.2 (0.7–2.1)	2.4 (1.2–5.0)
Mild-	149/389	39.6/45.5	46.3/42.2	14.1/12.3	1.3 (0.8–1.9)	1.3 (0.7–2.4)	1.3 (0.8–1.9)	1.3 (0.7–2.4)
Moderate-	378/286	40.0/46.9	44.2/43.4	15.9/9.8	1.2 (0.9–1.7)	1.9 (1.1–3.1)	1.1 (0.8–1.6)	1.7 (1.0–3.0)
Heavy-	491/131	47.1/37.4	40.7/46.6	12.2/16.0	0.7 (0.5–1.1)	0.6 (0.3–1.1)	0.7 (0.4–1.1)	0.5 (0.3–1.0)

[a] Mild, moderate, and heavy smokers correspond to the three tertiles of pack-years in ever smokers for all participants. The tertiles were divided at 25 and 55 pack-years.
[b] Stratified analysis. In all logistic regression models, the *Arg/Arg* genotype was the reference group. For the adjusted analysis, the logistic regression model included the following covariates: age, gender, square root of pack-years, smoking status, time since smoking cessation (in years), and genotype groups.

have quit smoking for >5 years were 87.4% (controls) and 85.5% (Massachusetts).

Distribution of DNA Repair Gene Polymorphisms Among Cases and Controls. The *XRCC1* polymorphism in this control population was consistent with Hardy-Weinberg equilibrium ($P = 0.95$, χ^2 goodness of fit). The frequencies of *Arg/Arg*, *Arg/Gln*, and *Gln/Gln* were 42.8, 42.9, and 14.3%, respectively, in cases, and 44.5, 44.0, and 11.5% in controls. For older subjects (age ≥55 years), the genotype frequencies were very similar between the 886 cases and the 769 controls, with the frequencies of 42.2, 44.1, and 13.7% in cases, and 42.8, 44.7, and 12.5% in controls, respectively. For younger subjects (age <55 years), cases have higher frequency of the *Gln/Gln* genotype than controls; the frequencies of *Arg/Arg*, *Arg/Gln*, and *Gln/Gln* were 45.4, 37.6, and 17.1%, respectively, in the 205 cases, and 47.1, 42.9, and 10.0% in the 471 controls, respectively. In categories of higher pack-years of smoking, the frequency of *Arg/Arg* genotype showed a trend of increase in cases ($P_{trend} = 0.06$ for general linear models) but not in controls ($P_{trend} = 0.43$; Table 1; two-step test; Refs. 26, 29, 33).

The frequency distributions of the *ERCC2 Asp312Asn* and *Lys751Gln* polymorphisms have been described previously; the controls are also in Hardy-Weinberg equilibrium (26). Trends in different pack-year categories of smoking for these two polymorphisms were similar to the *XRCC1* polymorphism.

Associations between the *XRCC1 Arg399Gln* Polymorphism and Lung Cancer Risk. There was a borderline overall relationship between *XRCC1* polymorphism and lung cancer risk. Compared with the *Arg/Arg* genotype, the adjusted ORs of lung cancer were 1.0 (95% CI, 0.8–1.3) for the *Arg/Gln* genotype, and 1.3 (95% CI, 1.0–1.8) for the *Gln/Gln* genotype, which were similar to the crude results (Table 1). Similar associations between this polymorphism and lung cancer risk were found for males and females separately, and for different histological (adenocarcinoma and squamous cell carcinoma) and clinical stages (stage I/II and stage IIIA/IIIB/IV) subgroups.

We found a statistically significant interaction ($P = 0.01$) between *XRCC1* genotype (*Gln/Gln versus Arg/Arg*) and age subgroup (older *versus* younger) in the risk of lung cancer. The adjusted ORs of the *Gln/Gln* genotype *versus* the *Arg/Arg* genotype were 2.4 (95% CI, 1.3–4.2) for younger subjects (age <55 years; Ref. 30) and 1.0 (95% CI, 0.7–1.5) for older subjects (age ≥55 years). Similar trends were obtained when the age subgroups were dichotomized at 50, 60, or 65 years.

Relationship among Cumulative Cigarette Smoking, DNA Repair Gene Polymorphisms, and Lung Cancer Risk. For the *XRCC1* polymorphism, compared with the *Arg/Arg* genotype, the *Gln/Gln* genotype was a risk factor for lung cancer in nonsmokers and light smokers (<25 pack-years), but a protective factor in heavy smokers (≥55 pack-years), both in crude and adjusted analyses (Table 1). Similar associations were found when we stratified the population into corresponding levels by tertiles or quartiles of pack-years.

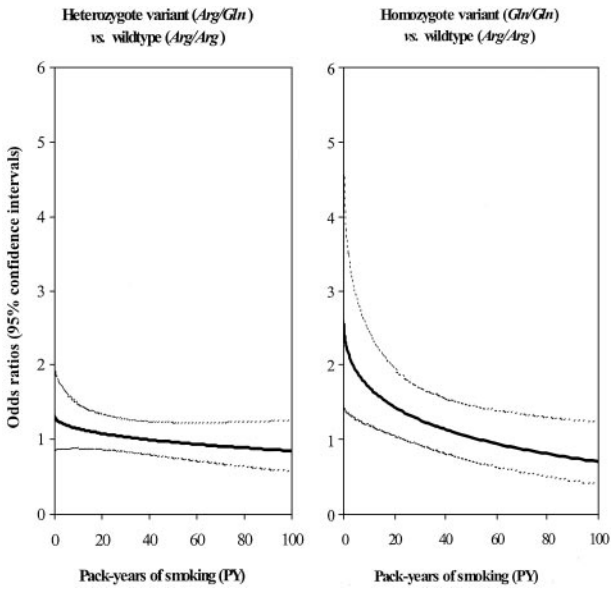

Fig. 1. Adjusted interaction model evaluating *XRCC1 Arg399Gln* polymorphism and smoking exposure (continuous variable). The fitted ORs (———) and 95% CIs (····) represent the lung cancer risk of carrying *Arg/Gln* genotype (*left*) or *Gln/Gln* genotype (*right*), when each is compared individually to the *Arg/Arg* genotype, at different numbers of pack-years (*PY*). The logistic regression model included the following covariates: age, gender, square root of pack-years, smoking status, time since smoking cessation (in years), genotypes, interaction terms between heterozygote and homozygote variant genotypes and square root of pack-years, and interaction between smoking status and square root of pack-years. The *P*s for interaction terms between genotypes and square root of pack-years were 0.22 for the heterozygote variant and <0.01 for the homozygote variant.

For the gene-smoking interaction analysis between the *XRCC1* polymorphism and lung cancer risk, the magnitude and statistical significance of the interaction term between genotype group and square root of pack-years for the *Gln/Gln versus Arg/Arg* genotype comparison was stronger, respectively, than that for the *Arg/Gln versus Arg/Arg* comparison, suggesting a dose-response relationship for the number of *Gln* alleles (Fig. 1).

Consistency of the results was evaluated by considering smoking variables in different ways. We decomposed pack-years into its component parts of smoking intensity (mean number of cigarettes per day) and duration (in years). Similar gene-smoking interaction associations were found when either

Table 2 Genotype frequencies of the combined *XRCC1 Arg399Gln, ERCC2 Asp312Asn,* and *ERCC2 Lys751Gln* polymorphisms for different pack-years of smoking[a,b]

	n (case/control)	Percentage (case/control)						
Variant alleles		Group 0 (0)	Group 1 (1)	Group 2 (2)	Group 3 (3)	Group 4 (4)	Group 5 (5)	Group 6 (6)
Total	1091/1240	14.1/14.4	20.6/21.9	24.8/27.4	21.3/21.2	12.4/10.7	5.7/3.5	1.2/1.1
Non-	73/434	8.2/13.8	19.2/22.4	15.1/27.2	23.3/22.6	20.6/9.9	9.6/3.0	4.1/1.2
Mild-	149/389	16.1/17.2	22.2/20.3	19.5/27.8	21.5/17.7	10.7/12.1	9.4/3.6	0.7/1.3
Moderate-	378/286	13.5/14.3	19.8/25.9	25.4/28.7	23.0/20.6	12.7/7.7	4.5/2.5	1.1/0.4
Heavy-	491/131	14.9/7.6	21.0/16.0	27.3/24.4	19.6/28.2	11.4/15.3	4.9/6.9	1.0/1.5

[a] Mild, moderate, and heavy smokers correspond to the three tertiles of pack-years in ever smokers for all participants. The tertiles were divided at 25 and 55 pack-years.
[b] Combined genotype group 0 were the individuals with 0 variant alleles of the three polymorphisms (wild-type for all of the three polymorphisms); group 1 were the individuals with 1 variant allele; group 2 were the individuals with 2 variant alleles; group 3 were the individuals with 3 variant alleles; group 4 were the individuals with 4 variant alleles; group 5 were the individuals with 5 variant alleles; and group 6 were the individuals with 6 variant alleles.

Table 3 Adjusted ORs (95% CI) for the joint effect of *XRCC1* polymorphism (or combined *XRCC1* and *ERCC2* polymorphisms) and different smoking categories[a]

Genotype	Smoking				Marginal
	Non-	Mild-	Moderate-	Heavy-	
XRCC1 Arg399Gln					
Arg/Arg	1.0	6.0 (3.4–10.6)	12.8 (7.7–21.2)	42.9 (24.9–74.0)	1.0
Arg/Gln	1.2 (0.7–2.1)	7.9 (4.5–13.8)	14.8 (8.9–24.6)	30.4 (17.8–51.8)	1.1 (0.9–1.3)
Gln/Gln	2.6 (1.2–5.4)	8.5 (4.1–17.4)	22.1 (11.6–42.1)	24.6 (12.5–48.4)	1.3 (1.0–1.8)
Marginal	1.0	5.7 (3.8–8.4)	11.7 (8.4–16.3)	27.4 (19.2–39.1)	–
XRCC1 Arg399Gln + ERCC2 Asp312Asn + ERCC2 Lys751Gln[b]					
Group 0	1.0	9.2 (3.3–25.2)	18.8 (7.1–49.8)	91.5 (30.4–275.7)	1.0
Group 1	1.5 (0.5–4.3)	10.1 (3.7–27.1)	16.2 (6.4–41.3)	60.2 (22.2–163.1)	0.9 (0.7–1.3)
Group 2	0.9 (0.3–2.7)	7.2 (2.7–19.5)	18.5 (7.3–46.9)	54.9 (21.0–143.6)	0.9 (0.6–1.2)
Group 3	1.8 (0.6–4.8)	11.8 (4.3–31.8)	22.9 (8.9–58.9)	34.7 (13.3–90.5)	1.0 (0.7–1.4)
Group 4	3.8 (1.3–11.0)	8.7 (3.0–25.3)	30.1 (10.9–83.3)	34.6 (12.5–95.7)	1.1 (0.8–1.7)
Group 5/6	5.4 (1.7–17.4)	22.5 (7.1–70.7)	36.9 (11.0–123.9)	30.2 (9.8–93.1)	1.7 (1.0–2.7)
Marginal	1.0	5.6 (3.8–8.4)	11.9 (8.5–16.7)	27.7 (19.4–39.5)	–

[a] Interaction Models: separate logistic regression model for the *XRCC1 Arg399Gln* polymorphism alone or the combined three polymorphisms included the following covariates: age, gender, three dummy variables for pack-years, smoking status, time since smoking cessation (in years), genotype groups, 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.
[b] Combined genotype group 0 were the individuals with 0 variant alleles of the three polymorphisms (wild-type for all of the three polymorphisms); group 1 were the individuals with 1 variant allele; group 2 were the individuals with 2 variant alleles; group 3 were the individuals with 3 variant alleles; group 4 were the individuals with 4 variant alleles; and group 5/6 were the individuals with 5 or 6 variant alleles.

log-transformed cigarettes per day or smoking duration (both as continuous variables) were substituted for square root of pack-years in the regression models. The gene-smoking interaction term between *XRCC1* genotype (*Gln/Gln versus Arg/Arg*) and log-transformed cigarettes per day was statistically significant ($P = 0.03$), and the interaction between genotype and years of smoking (*Gln/Gln versus Arg/Arg*) was borderline significant ($P = 0.07$).

In a recent publication, we described similar genotype-smoking interaction associations for the *ERCC2 Asp312Asn* and *Lys751Gln* polymorphisms where there was a dose-response relationship for the number of variant *Asn* or *Gln* alleles (26). The variant *Asn* or *Gln* alleles were risk factors in nonsmokers, but protective factors in heavy smokers when compared with the *Asp/Asp* or *Lys/Lys* wild-type.

Combined Effects of Two DNA Repair Genes on Lung Cancer Risk. We dichotomized the population into six genotype groups based on the number of variant alleles of the three *XRCC1* and *ERCC2* polymorphisms: group 0, the reference group, with no variant alleles of either gene, *i.e.,* individuals with wild genotype for all of the three polymorphisms; group 1, with 1 variant allele; group 2, with 2 variant alleles; and so forth. The frequency distributions of each genotype combination group by different pack-year categories are shown in Table 2.

In all of the logistic regression analyses, the genotype group 5 and group 6 were combined into one group (group 5/6) because of the small sample sizes. For subjects of different pack-year categories of smoking, when compared with the reference genotype group (group 0), the genotype combination groups with >3 variant alleles were risk factors for lung cancer in nonsmokers but protective factors in heavy smokers (Table 3). In stratified regression models, the adjusted ORs of the extreme genotype combinations of variant alleles (group 5/6) *versus* group 0 were 5.2 (95% CI, 1.7–16.6) for nonsmokers, 2.2 (95% CI, 0.9–5.2) for mild smokers, 2.0 (95% CI, 0.8–5.1) for moderate smokers, and 0.3 (95% CI, 0.1–0.8) for heavy smokers, respectively. Table 3 presented the results of the combined effects of cigarette smoking and different genotype groups (*XRCC1 Arg399Gln* polymorphism alone or the joint *XRCC1* and *ERCC2* polymorphisms) on the risk of lung cancer.

For the formal gene-smoking interaction analysis between the *XRCC1* and *ERCC2* combined genotypes and lung cancer risk, the magnitude and statistical significance of the interaction

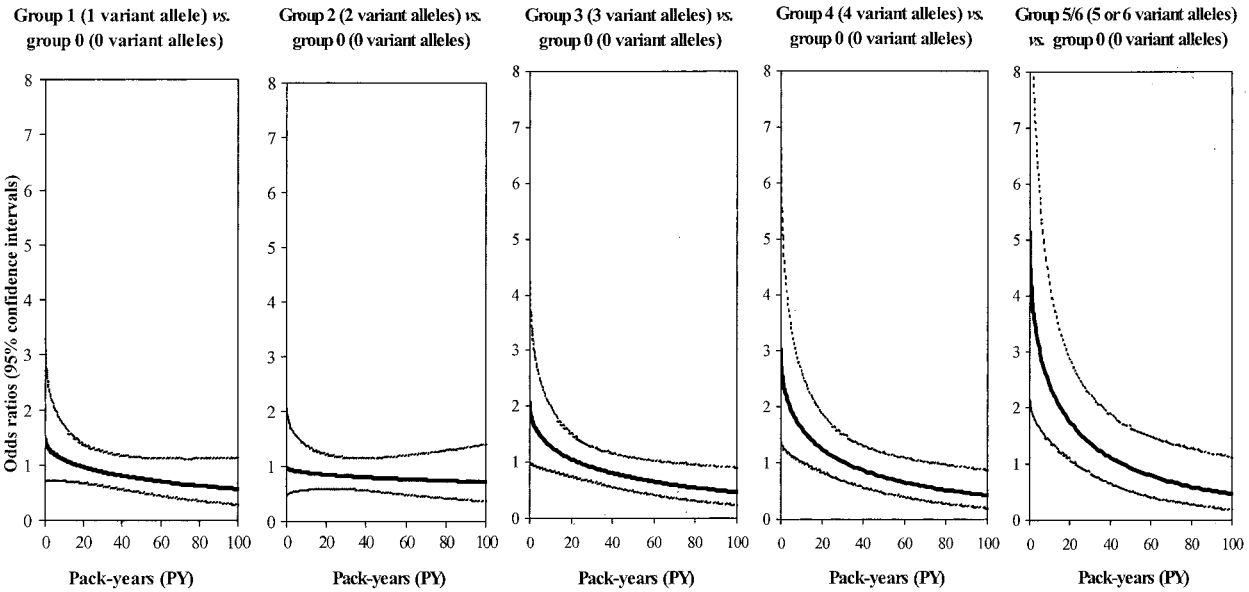

Fig. 2. Adjusted interaction model evaluating combined *XRCC1 Arg399Gln*, *ERCC2 Asp312Asn*, and *ERCC2 Lys751Gln* polymorphisms and smoking exposure (continuous variable). Combined genotype group 0 were the individuals with 0 variant alleles of the three polymorphisms; group 1 were the individuals with 1 variant allele; group 2 were the individuals with 2 variant alleles; group 3 were the individuals with 3 variant alleles; group 4 were the individuals with 4 variant alleles; group 5/6 were the individuals with 5 or 6 variant alleles. The fitted ORs (———) and 95% CIs (····) represent the lung cancer risk of combined genotype group 1 (*first on the left*), group 2 (*second on the left*), group 3 (*middle*), group 4 (*second on the right*), or group 5/6 (*first on the right*), when each is compared with the combined genotype group 0, at different numbers of pack-years (*PY*). The logistic regression model included the following covariates: age, gender, square root of pack-years, smoking status, time since smoking cessation (in years), combined genotype groups, interaction terms between different combined genotype groups and square root of pack-years, and interaction between smoking status and square root of pack-years. The *P*s for interaction terms between combined genotype groups and square root of pack-years were 0.13 for group 1, 0.61 for group 2, 0.02 for group 3, <0.01 for group 4, and <0.01 for group 5/6.

term between genotype group and square root of pack-years was strongest when comparing group 5/6 to group 0 (Fig. 2). The adjusted ORs of all five of the comparisons decreased as pack-years of smoking increased, and the fitted ORs were comparable with the results when pack-years was stratified into different categories. Similar gene-smoking interaction associations were found when either log-transformed cigarettes per day or smoking duration (both as continuous variables) were substituted for square root of pack-years in the regression models. The gene-smoking interaction terms between the extreme genotype combinations (group 5/6 *versus* group 0) and either years of smoking or log-transformed cigarettes per day were statistically significant ($P < 0.01$). Similar associations between the combined polymorphisms and lung cancer risk were found for males and females, younger and older subjects, and for different histological (adenocarcinoma and squamous cell carcinoma) and clinical stage (stage I/II and stage IIIA/IIIB/IV) subgroups.

Discussion

In this study, we used cumulative cigarette smoking as a surrogate for tobacco carcinogen exposure. We found statistically significant interactions between the three polymorphisms of the two DNA repair genes (*XRCC1* and *ERCC2*), both separately and combined, and cumulative cigarette smoking in lung cancer risk in this Caucasian sample. Invariably, the variant alleles of the three DNA repair genetic polymorphisms were risk factors in non- or light smokers, but tended to be protective in heavy smokers. This finding was consistent regardless of the analytical techniques and modeling for different subgroups by age, gender, histological cell types, and clinical stages.

In smaller studies, similar gene-smoking relationships have been described for the polymorphisms of both *XRCC1* (*Arg399Gln*) and *ERCC2* (*Asp312Asn* and *Lys751Gln;* Refs. 11, 14–16, 23). For *XRCC1*, one study with 308 healthy Italian individuals found that the *Gln/Gln* genotype had significantly higher DNA adduct levels in lymphocytes only in never-smokers but not in ever-smokers (11). Recent epidemiological studies involving Caucasian (180 cases), African-American (154 cases), or Korean (192 cases) suggested consistently that the variant *Gln* allele was associated with higher risk of lung cancer for lighter smokers and lower risk in heavier smokers (15, 16). For *ERCC2*, the *Asp/Asp* genotype of the *Asp312Asn* polymorphism was associated with higher lung cancer risk for light smokers only, and not for non- or heavier smokers in one study with 96 Caucasian cases (14). A recent study with 185 Swedish lung cancer cases suggested that the variant alleles of the two *ERCC2* polymorphisms were associated with increased risk for never-smokers only, and tended to be protective in ever-smokers (23). In addition, inconsistent results were demonstrated regarding the possible functional consequences of polymorphisms of both DNA repair genes, and between individual polymorphisms and lung cancer risk (5–10, 12, 13, 18–22). Discrepancies across these studies may be because of relatively small sample sizes analyzed in the various subgroups or because of associations that were specific to different cumulative smoking levels.

This is the first report showing the gene-smoking interactions between the joint effects of *XRCC1* and *ERCC2* genotypes and cumulative cigarette smoking exposure in lung cancer risk. Our results suggest that the magnitudes of these gene-smoking interactions are associated with the number of variant alleles of these three polymorphisms (Table 3; Fig. 2). Comparison of the

extreme joint genotype combinations of *XRCC1* and *ERCC2* is consistent with an additive effect of the two polymorphic genes at each level of smoking exposure (data not shown).

The exact mechanism of how cigarette smoking changes the DRC posed by each genotype of these DNA repair genetic polymorphisms is unknown. One possible explanation may be that different DRCs of different genotypes are overwhelmed by heavy smoking exposure. Alternately, cigarette smoking may stimulate DRC in response to the DNA damage caused by tobacco carcinogens, because heavy smokers among both lung cancer patients and controls may have more proficient DRC in lymphocytes than non- or light smokers (34), and DNA repair gene expression was increased in heavy smokers among both head and neck cancer patients and controls (35). Because heavy smokers tended to be older subjects (the frequencies of older subjects in non-, mild, moderate, and heavy smokers were 60.6, 60.4, 72.9, and 86.7%, respectively), the lower risk for the variant alleles in heavy smokers may reflect a prolonged survival for the lung cancer patients. A recent study has shown that effective host DRC may be associated with poorer survival in patients with non-small cell lung cancer who are treated with chemotherapy (36). We also found that the *Gln/Gln* genotype of the *XRCC1 Arg399Gln* polymorphism was associated with a markedly increased risk of lung cancer in younger individuals but not in older individuals. Additional analysis is needed to explore the effect of DNA repair genetic polymorphisms on the survival of lung cancer patients.

This study has a number of limitations. Firstly, our study was a hospital-based case-control study. Secondly, our controls were healthy spouses or friends of individuals with lung cancer or other cardiothoracic conditions. Because spouse or friend controls are more similar to cases than population controls because they are likely to share similar health behaviors, there may be some potential bias for our results. However, bias in the estimate of the stratum-specific OR because of a specific gene with strata defined by levels of measured confounders and effect modifiers (such as smoking, diet, and so forth) will only occur if, within these strata, spouse and friend controls are more likely than other types of controls to have a particular allele or alleles of the genotype under study. Although we did not match individually our controls to cases on age, gender, race, or smoking variables, we did adjust for these variables in the analyses and also performed stratified analyses by these variables, and each time found consistent gene-smoking associations. Thirdly, the exact biological mechanisms for the gene-smoking interaction associations are unclear, because there is no direct phenotype data for the function of these polymorphisms. Nonetheless the current study adds weight to existing epidemiological studies that have found a similar gene-smoking interaction association between these two DNA repair genes and cumulative cigarette smoking (11, 14–16, 23). Fourthly, we only evaluated the *Arg399Gln* polymorphism of *XRCC1*, and *Asp312Asn* and *Lys751Gln* polymorphisms of *ERCC2* in this study, and did not evaluate the other polymorphisms of these two genes, including the *XRCC1 Arg194Trp* polymorphism, because of scarce functional data and relatively low allele frequencies for these polymorphisms (5, 6, 14, 15, 35, 36), which may result in some misclassification in the functions of these two DNA repair genes. Besides, XRCC1 and ERCC2 only contribute partially to DNA repair capacity in their respective pathways, and it is possible that polymorphisms of other genes not evaluated in this study could play a role in determining lung cancer risk. However, evaluation of a greater number of polymorphisms will require large sample sizes. Fifthly, the data of environmental tobacco smoking and alcohol assumption are not adjusted in our analysis. In the case of environmental tobacco smoking, the use of spouse and friend controls may actually mitigate such differences, given that the environment of cases and controls may be generally similar. Alcohol assumption may modify the risk of head and neck, and esophageal cancers, whereas a recent review suggests that the evidence is far from conclusive in the risk of lung cancer (37). The above limitations may be one reason for the inconsistent trend in risk of lung cancer by *XRCC1* alone or the combined *XRCC1* and *ERCC2* genotypes across different smoking levels; *i.e.*, the OR of *XRCC1 Gln/Gln versus Arg/Arg* in moderate smoking group is bigger than that in the light smoking group (Table 1).

In conclusion, cumulative cigarette smoking exposure appears to play an important role in altering the direction and magnitude of the association between *XRCC1* and *ERCC2* polymorphisms, and the risk of lung cancer. Lung cancer risk by the polymorphisms of both *XRCC1* and *ERCC2* is dependent on cumulative smoking exposure. Additional studies are needed to detect the function of these polymorphisms, and its associations with cigarette smoking and clinical prognosis.

Acknowledgments

We thank the following staff members of the Lung Cancer Susceptibility Group: Barbara Bean, Jessica Shin, Andrea Solomon, Andrea Shaffer, So-Aye Shaneyfeet, Lucy Ann Principe, Salvatore Mucci, Richard Rivera-Massa, Lisa I. Wang, Stephanie Shih, and Maria Fragoso; and Dr. Panos Fidias, and the physicians and surgeons of the Massachusetts General Hospital Cancer Center for generous support.

References

1. Berwick, M., and Vineis, P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. J. Natl. Cancer Inst., *92:* 874–897, 2000.

2. Sancar, A., and Tang, M. S. Necleotide excision repair. Photochem. Photoibol., *57:* 905–921, 1993.

3. Wilson, D. M., III., and Thompson, L. H. Life without DNA repair. Proc. Natl. Acad. Sci. USA, *94:* 12754–7, 1997.

4. Thompson, L. H., and West, M. G. XRCC1 keeps DNA from getting stranded. Mutat. Res., *459:* 1–18, 2000.

5. Lunn, R. M., Langlois, R. G., Hsieh, L. L., Thompson, C. L., and Bell, D. A. XRCC1 polymorphisms: effects on aflatoxin B1-DNA adducts and glycophorin A variant frequency. Cancer Res., *59:* 2557–2261, 1999.

6. Abdel-Rahman, S. Z., and El-Zein, R. A. The *399Gln* polymorphism in the DNA repair gene *XRCC1* modulates the genotoxic response induced in human lymphocytes by the tobacco-specific nitrosamine NNK. Cancer Lett., *159:* 63–71, 2000.

7. Duell, E. J., Wiencke, J. K., Cheng, T. J., Varkonyi, A., Zuo, Z. F., Ashok, T. D., Mark, E. J., Wain, J. C., Christiani, D. C., and Kelsey, K. T. Polymorphisms in the DNA repair genes *XRCC1* and *ERCC2* and biomarkers of DNA damage in human blood mononuclear cells. Carcinogenesis (Lond.), *21:* 965–971, 2000.

8. Hu, J. J., Smith, T. R., Miller, M. S., Mohrenweiser, H. W., Golden, A., and Case, L. D. Amino acid substitution variants of *APE1* and *XRCC1* genes associated with ionizing radiation sensitivity. Carcinogenesis (Lond.), *22:* 917–922, 2001.

9. Matullo, G., Guarrera, S., Carturan, S., Peluso, M., Malaveille, C., Davico, L., Piazza, A., and Vineis, P. DNA repair gene polymorphisms, bulky DNA adducts in white blood cells and bladder cancer in a case-control study. Int. J. Cancer, *92:* 562–567, 2001.

10. Palli, D., Russo, A., Masala, G., Saieva, C., Guarrera, S., Carturan, S., Munnia, A., Matullo, G., and Peluso, M. DNA adduct levels and DNA repair polymorphisms in traffic-exposed workers and a general population sample. Int. J. Cancer, *94:* 121–127, 2001.

11. Matullo, G., Palli, D., Peluso, M., Guarrera, S., Carturan, S., Celentano, E., Krogh, V., Munnia, A., Tumino, R., Polidoro, S., Piazza, A., and Vineis, P. XRCC1, XRCC3, XPD gene polymorphisms, smoking and ^{32}P-DNA adducts in a sample of healthy subjects. Carcinogenesis (Lond.), *22:* 1437–1445, 2001.

12. Divine, K. K., Gilliland, F. D., Crowell, R. E., Stidley, C. A., Bocklage, T. J., Cook, D. L., and Belinsky, S. A. The *XRCC1* 399 glutamine allele is a risk factor for adenocarcinoma of the lung. Mutat. Res., *461:* 273–278, 2001.

13. Ratnasinghe, D., Yao, S. X., Tangrea, J. A., Qiao, Y. L., Andersen, M. R., Barrett, M. J., Giffen, C. A., Erozan, Y., Tockman, M. S., and Taylor, P. R. Polymorphisms of the DNA repair gene *XRCC1* and lung cancer risk. Cancer Epidemiol. Biomark. Prev., *10:* 119–123, 2001.

14. Butkiewicz, D., Rusin, M., Enewold, L., Shields, P. G., Chorazy, M., and Harris, C. C. Genetic polymorphisms in DNA repair genes and risk of lung cancer. Carcinogenesis (Lond.), *22:* 593–597, 2001.

15. David-Beabes, G. L., and London, S. J. Genetic polymorphism of *XRCC1* and lung cancer risk among African-Americans and Caucasians. Lung Cancer, *34:* 333–339, 2001.

16. Park, J. Y., Lee, S. Y., Jeon, H. S., Bae, N. C., Chae, S. C., Joo, S., Kim, C. H., Park, J. H., Kam, S., Kim, I. S., and Jung, T. H. Polymorphism of the DNA repair gene *XRCC1* and risk of primary lung cancer. Cancer Epidemiol. Biomark. Prev., *11:* 23–27, 2002.

17. Leadon, S. A., and Cooper, P. K. Preferential repair of ionizing radiation-induced damage in the transcribed strand of an active human gene is defective in Cockayne syndrome. Proc. Natl. Acad. Sci. USA, *90:* 10499–10503, 1993.

18. Lunn, R. M., Helzlsouer, K. J., Parshad, R., Umbach, D. M., Harris, E. L., Sanford, K. K., and Bell, D. A. *XPD* polymorphisms: effects on DNA repair proficiency. Carcinogenesis (Lond.), *21:* 551–555, 2000.

19. Spitz, M. R., Wu, X., Wang, Y., Wang, L. E., Shete, S., Amos, C. I., Guo, Z., Lei, L., Mohrenweiser, H., and Wei, Q. Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. Cancer Res., *61:* 1354–1357, 2001.

20. David-Beabes, G. L., Lunn, R. M., and London, S. J. No association between the *XPD (Lys751G1n*) polymorphism or the *XRCC3 (Thr241Met*) polymorphism and lung cancer risk. Cancer Epidemiol. Biomark. Prev., *10:* 911–912, 2001.

21. Seker, H., Butkiewicz, D., Bowman, E. D., Rusin, M., Hedayati, M., Grossman, L., and Harris, C. C. Functional significance of *XPD* polymorphic variants: attenuated apoptosis in human lymphoblastoid cells with the *XPD* 312 Asp/Asp genotype. Cancer Res., *61:* 430–434, 2001.

22. Hemminki, K., Xu, G., Angelini, S., Snellman, E., Jansen, C. T., Lambert, B., and Hou, S. M. *XPD* exon 10 and 23 polymorphisms and DNA repair in human skin *in situ*. Carcinogenesis (Lond.), *22:* 1185–1188, 2001.

23. Hou, S. M., Falt, S., Angelini, S., Yang, K., Nyberg, F., Lambert, B., and Hemminki, K. The *XPD* variant alleles are associated with increased aromatic DNA adduct level and lung cancer risk. Carcinogenesis (Lond.), *23:* 599–603, 2002.

24. Wei, Q., and Spitz, M. R. The role of DNA repair capacity in susceptibility to lung cancer: a review. Cancer Metastasis Rev., *16:* 295–307, 1997.

25. Wei, Q., Cheng, L., Hong, W. K., and Spitz, M. R. Reduced DNA repair capacity in lung cancer patients. Cancer Res., *56:* 4103–4107, 1996.

26. Zhou, W., Liu, G., Miller, D. P., Thurston, S. W., Xu, L. L., Wain, J. C., Lynch, T. J., Su, L., and Christiani, D. C. Gene-environment interaction for the *ERCC2* polymorphisms and cumulative cigarette smoking exposure in lung cancer. Cancer Res., *62:* 1377–1381, 2002.

27. Xu, L. L., Wain, J. C., Miller, D. P., Thurston, S. W., Su, L., Lynch, T. J., and Christiani, D. C. The NAD(P)H:quinone oxidoreductase 1 gene polymorphism and lung cancer: differential susceptibility based on smoking behavior. Cancer Epidemiol. Biomark. Prev., *10:* 303–309, 2001.

28. Zhou, W., Thurston, S. W., Liu, G., Xu, L. L., Miller, D. P., Wain, J. C., Lynch, T. J., Su, L., and Christiani, D. C. The interaction between microsomal epoxide hydrolase polymorphisms and cumulative cigarette smoking in different histological subtypes of lung cancer. Cancer Epidemiol. Biomark. Prev., *10:* 461–466, 2001.

29. Zhou, W., Liu, G., Thurston, S. W., Xu, L. L., Miller, D. P., Wain, J. C., Lynch, T. J., Su, L., and Christiani, D. C. Genetic polymorphisms in *N*-acetyltransferase-2 and microsomal epoxide hydrolase, cumulative cigarette smoking and lung cancer. Cancer Epidemiol. Biomark. Prev., *11:* 15–21, 2002.

30. Wang, L. I., Miller, D. P., Sai, Y., Liu, G., Su, L., Wain, J. C., Lynch, T. J., and Christiani, D. C. Manganese superoxide dismutase alanine-to-valine polymorphism at codon 16 and lung cancer risk. J. Natl. Cancer Inst., *93:* 1818–1821, 2001.

31. Hastie, T. J., and Tibshirani, R. J. Generalized Additive Models. New York: Chapman and Hall, 1990.

32. Hosmer, D. W., and Lemeshow, S. Applied Logistic Regression. New York: John Wiley and Sons, 1989.

33. Garcia-Closas, M., Thompson, W. D., and Robins, J. M. Differential misclassification and the assessment of gene-environment interactions in case-control studies. Am. J. Epidemiol., *147:* 426–433, 1998.

34. Wei, Q., Cheng, L., Amos, C. I., Wang, L. E., Guo, Z., Hong, W. K., and Spitz, M. R. Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. J. Natl. Cancer Inst., *92:* 1764–1772, 2000.

35. Wei, Q., Eicher, S. A., Guan, Y., Cheng, L., Xu, J., Young, L. N., Saunders, K. C., Jiang, H., Hong, W. K., Spitz, M. R., and Strom, S. S. Reduced expression of hMLH1 and hGTBP/hMSH6: a risk factor for head and neck cancer. Cancer Epidemiol. Biomark. Prev., *7:* 309–314, 1998.

36. Bosken, C. H., Wei, Q., Amos, C. I., and Spitz, M. R. An analysis of DNA repair as a determinant of survival in patients with non-small-cell lung cancer. J. Natl. Cancer Inst., *94:* 1091–1099, 2002.

37. Bandera, E. V., Freudenheim, J. L., and Vena, J. E. Alcohol consumption and lung cancer: a review of the epidemiologic evidence. Cancer Epidemiol. Biomark. Prev., *10:* 813–821, 2001.