Large-Scale Investigation of Base Excision Repair Genetic Polymorphisms and Lung Cancer Risk in a Multicenter Study

Rayjean J. Hung, Paul Brennan, Federico Canzian, Neonila Szeszenia-Dabrowska, David Zaridze, Jolanta Lissowska, Peter Rudnai, Eleonora Fabianova, Dana Mates, Lenka Foretova, Vladimir Janout, Vladimir Bencko, Amelie Chabrier, Stephane Borel, Janet Hall, Paolo Boffetta

Background: Base excision repair (BER) is a highly conserved essential mechanism for maintaining genome integrity. We examined associations among four well-characterized polymorphisms of BER genes (OGG1 Ser326Cys, XRCC1 Arg194Trp, XRCC1 Arg280His, and XRCC1 Arg399Gln) and lung cancer risk. Methods: A total of 2188 patients with lung cancer and 2198 control subjects without lung cancer recruited at 15 centers in six Eastern European countries from February 1998 to October 2002 provided DNA samples for genotype analysis. Genetic polymorphisms were analyzed by the fluorescence 5’ exonuclease and AmplifiKer assays. Unconditional multivariable logistic regression was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs). We estimated the false-positive reporting probability (FPRP) for our results by incorporating a range of prior probabilities that specific polymorphisms are associated with lung cancer risk. All statistical tests were two-sided. Results: The overall odds ratio for lung cancer among those with the OGG1 Cys/Cys genotype compared with those with the OGG1 Ser/Ser genotype was 1.34 (95% CI = 0.95 to 1.88); the association was most prominent for adenocarcinoma risk (OR = 1.66, 95% CI = 1.04 to 2.66). Overall, the XRCC1 polymorphisms were not associated with the risk of lung cancer. However, the XRCC1 Arg194Trp and Arg280His variants were each associated with a reduced risk of lung cancer among subjects in the highest quartile of pack-years of smoking compared with common allele homozygotes (ORs of 0.65 [95% CI = 0.46 to 0.93] and 0.56 [95% CI = 0.36 to 0.86], respectively). The associations between the OGG1 Cys/Cys genotype and adenocarcinoma risk and between XRCC1 Arg194Trp polymorphism and lung cancer risk among heavy smokers remained robust given prior probabilities of 25% (FPRP = 0.238) and 10% (FPRP = 0.276), respectively. Conclusions: Our results do not support a major independent role of BER gene polymorphisms in lung cancer risk. However, we cannot exclude the possibility that the OGG1 Ser326Cys and XRCC1 Arg194Trp polymorphisms play minor roles in lung carcinogenesis. [J Natl Cancer Inst 2005;97:567–76]

A affiliations of authors: International Agency for Research on Cancer, Lyon, France (RJH, PB, FC, AC, SB, JH, PB); Department of Epidemiology, Institute of Occupational Medicine, Lodz, Poland (NS-D); Institute of Carcinogenesis, Cancer Research Centre, Moscow, Russia (DZ); Department of Cancer Epidemiology and Prevention, Maria Sklodowska Curie and Institute of Oncology, Warsaw, Poland (JL); Johan National Institute of Public Health, Budapest, Hungary (PR); Specialized Institute of Hygiene and Epidemiology, Banska Bystrica, Slovakia (EF); Institute of Public Health, Bucharest, Romania (DM); Department of Cancer Epidemiology and Genetics, Masaryk Memorial Cancer Institute, Brno, Czech Republic (VF); Institute of Preventive Medicine, Palacky University Faculty of Medicine, Olomouc, Czech Republic (VJ); Charles University in Prague, First Faculty of Medicine, Institute of Hygiene and Epidemiology, Prague, Czech Republic (VB).

Correspondence to: Paul Brennan, IARC, 150 cours Albert-Thomais, 69008 Lyon, France (e-mail: brennan@iarc.fr).

See “Notes” following “References.”

DOI: 10.1093/jnci/dji101
Journal of the National Cancer Institute, Vol. 97, No. 8, © Oxford University Press 2005, all rights reserved.
Several studies have examined the association between OGG1 genotypes and enzyme activity, with inconsistent results [reviewed in (13)]. One study (14) found no association between different human OGG1 genotypes and OGG1 enzyme activity (15). However, Kohno et al. (12) found by using a complementation assay of an *Escherichia coli* mutant defective in the repair of 8-oxoG that 326Ser-containing OGG1 has a sevenfold higher activity for repairing 8-oxoG than the 326Cys-containing OGG1. Previous evidence indicated higher levels of 8-oxoG in lung tissue of lung cancer patients than in lung tissue of patients without cancer (16). We therefore hypothesized that the Ser326Cys polymorphism of OGG1 may be associated with an increased risk of lung cancer. This possibility is supported by results of two studies, from Japan and Hawaii, which reported that the Cys/Cys genotype is associated with an increased risk of lung cancer (17,18). However, results of two other studies, from Germany and Japan, did not show such an association (19,20). Each of the four studies had fewer than 300 lung cancer case patients, however, and were therefore of limited statistical power (Table 1).

The apurinic/apyrimidinic sites that remain after damaged bases are removed are cytotoxic and mutagenic and require further processing to ensure cell viability and genomic integrity. Apurinic/apyrimidinic endonuclease (also known as APEX1 or APE1) is involved in this phase of BER; it cleaves the DNA backbone at the 5' side of the apurinic/apyrimidinic site (6,21), after which the BER pathway proceeds to short-patch DNA repair (when the gap is only one nucleotide) or long-patch DNA repair (when the gap is two to eight nucleotides long) (2). The essential protein known as X-ray repair cross-complementing group 1 (XRCC1) has a central role in short-patch DNA repair in the BER pathway; it also senses single-strand breaks in DNA through its interactions with poly(ADP-ribose) polymerase (PARP) and proliferating cell nuclear antigen (PCNA) (22–24). Three genetic polymorphisms have been identified that produce variant forms of XRCC1: the Arg194Trp polymorphism in exon 6 (dbSNP no. rs1799782), the Arg280His polymorphism in exon 9 (dbSNP no. rs25489), and the Arg399Gln polymorphism in exon 10 (dbSNP no. rs25487). The Arg194Trp variant has been shown to be associated with decreased sensitivity to the mutagens bleomycin and benzo[a]pyrene-diol-epoxide (BPDE) (25). The functional significance of the Arg280His variant is not yet well understood; however, the 280His allele is located in the PCNA-binding region of XRCC1, and results of a small (80-subject) study suggested that this allele is associated with increased bleomycin sensitivity (23,26). The codon 399 (G→A) transversion is located in the carboxyl-terminal region of the PARP-interacting domain of XRCC1. The 399Gln allele has been shown to be associated with increased levels of aflatoxin B1-DNA adduct (a highly unstable form of DNA damage whose subsequent lesions can be removed by the BER pathway) and increased bleomycin sensitivity (25,27) in several studies (28); however, another study (29) found no such association.

Several epidemiologic studies have examined the association between lung cancer risk and XRCC1 polymorphisms (Table 1). However, most of those studies were underpowered, and their results were inconclusive (24,30–33). Results of a large case-control study (1091 case patients and 1240 control subjects) suggested an interaction between cumulative tobacco consumption and the XRCC1 Gln/Gln genotype, with an increased risk of lung cancer in nonsmokers with the genotype and a decreased risk in heavy smokers with the genotype (34). The BER pathway is highly conserved and provides an essential mechanism for ensuring the integrity of the genome. It is estimated that 2000 to 10,000 apurinic/apyrimidinic sites arise per day in a mammalian cell grown under physiologic conditions (35). Reliable information about which BER gene polymorphisms are associated with lung cancer risk may elucidate the disease mechanism. We analyzed four well-characterized polymorphisms of BER genes in a large-scale case-control study with adequate statistical power to detect modest associations between sequence variants and lung cancer risk. We hypothesized that OGG1 326Cys, XRCC1 280His, and XRCC1 399Gln would be associated with increased lung cancer risk, whereas XRCC1 194Trp would be associated with decreased lung cancer risk. Furthermore, we hypothesized that the associations might be modulated by exposure to carcinogens, especially tobacco smoke. In addition, we estimated the false-positive report probability by incorporating the prior probability that these specific SNPs are associated with lung cancer risk (36).

**Subjects and Methods**

**Study Population**

This study was conducted in 15 centers in six Eastern European countries: the Czech Republic (Prague, Olomouc, and Brno), Hungary (Borsod, Heves, Szabolcs, Szolnok, and Budapest), Poland (Warsaw and Lodz), Romania (Bucharest), Russia (Moscow), and Slovakia (Banska Bystrica, Bratislava, and Nitra). Each center followed an identical protocol and was responsible for recruiting a consecutive group of patients who were newly diagnosed with lung cancer and a comparable group of hospital-based control subjects without lung cancer from February 1998 to October 2002.

All patients who were diagnosed with lung cancer from February 1998 to October 2002 at participating hospitals were eligible for this study. All cancer diagnoses were confirmed histologically or cytologically. Eligible subjects (case patients and control subjects) must have resided in the study area for at least 1 year before recruitment. Lung cancer case patients were identified through an active search of the records of clinical and pathology departments at the participating hospitals. All centers attempted to recruit all eligible patients as soon as possible after the patient had received an initial diagnosis of lung cancer; the maximum time interval between diagnosis and recruitment was 3 months. All study subjects (case patients and control subjects) and their physicians provided written informed consent. This study was approved by the institutions at all study centers, and ethical approval was obtained from the International Agency for Research on Cancer (Lyon, France), the coordinating center.

At all centers except the Warsaw center, control subjects were chosen from among inpatients and outpatients admitted to the same hospital as the case patients; case patients and control subjects from each hospital were frequency matched to each other by sex, age (±3 years), center, and referral (or residence) area. Control subjects were eligible for this study if they had been diagnosed with non–tobacco-related diseases or had undergone minor surgical procedures or had benign disorders, common infections, eye conditions (except cataract or diabetic retinopathy), or common orthopedic diseases (except osteoporosis). At the Warsaw center, control subjects were selected by random sampling of the general population using the Electronic List of Polish Residents.
Table 1. Summary of the previous studies on associations between base excision repair genetic polymorphisms and lung cancer risk

<table>
<thead>
<tr>
<th>Gene</th>
<th>SNP</th>
<th>First author (reference)</th>
<th>Year</th>
<th>Country</th>
<th>No. case patients/No. control subjects</th>
<th>Source of control subjects</th>
<th>Matching</th>
<th>Main reported results†</th>
</tr>
</thead>
<tbody>
<tr>
<td>OGG1</td>
<td>Ser326Cys</td>
<td>Sugimura (17)</td>
<td>1999</td>
<td>Japan</td>
<td>241/197</td>
<td>Hospital</td>
<td>None</td>
<td>SCC: 3.01 (1.33 to 6.83) for Cys/Cys vs. Ser/Ser or Ser/Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Wikman (19)</td>
<td>2000</td>
<td>Germany</td>
<td>105/105</td>
<td>Hospital</td>
<td>Frequency</td>
<td>Adenocarcinoma: null</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ito (20)</td>
<td>2002</td>
<td>Japan</td>
<td>138/241</td>
<td>Hospital</td>
<td>Null</td>
<td>Hawaiian ethnicity: 3.6 (1.0 to 11.9) for Cys/Cys vs. Ser/Ser</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Le Marchand (18)</td>
<td></td>
<td></td>
<td>298/405</td>
<td>Population</td>
<td>Frequency</td>
<td>SCC: 3.7 (1.7 to 8.3) for Cys/Cys vs. Ser/Ser or Ser/Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Adenocarcinoma: 2.1 (1.1 to 3.9) for Cys/Cys vs. Ser/Ser or Ser/Cys</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SCC: 3.4 (1.1 to 10.4) for Cys/Cys vs. Ser/Ser or Ser/Cys</td>
</tr>
<tr>
<td>XRCC1</td>
<td>Arg194Trp</td>
<td>Ratnasinghe (57)</td>
<td>2001</td>
<td>China</td>
<td>108/216</td>
<td>Population</td>
<td>Individual Null</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>David-Beabes (31) (AA)</td>
<td>2001</td>
<td>United States</td>
<td>180/461</td>
<td>Population</td>
<td>Frequency</td>
<td>Overall: 0.4 (0.2 to 0.9) for Arg/Trp or Trp/Trp vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;20 Cig/day: 0.2 (0.1 to 0.9) for Arg/Trp or Trp/Trp vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td>Arg280His</td>
<td>Chen (48)</td>
<td>2002</td>
<td>China</td>
<td>109/109</td>
<td>Population</td>
<td>Individual Null</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ratnasinghe (57)</td>
<td>2002</td>
<td>China</td>
<td>106/209</td>
<td>Population</td>
<td>Frequency</td>
<td>Overall: null</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: 1.8 (1.0 to 3.4) for Arg/His or His/His vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td>Arg399Gln</td>
<td>Misra (50)</td>
<td>2003</td>
<td>Finland</td>
<td>315/315</td>
<td>Population</td>
<td>Individual Null</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Ratnasinghe (57)</td>
<td>2001</td>
<td>China</td>
<td>107/208</td>
<td>Population</td>
<td>Individual Null</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>David-Beabes (31) (AA)</td>
<td>2001</td>
<td>United States</td>
<td>154/243</td>
<td>Population</td>
<td>Null</td>
<td>Overall: 1.8 (1.0 to 3.4) for Arg/His or His/His vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: 1.8 (1.0 to 3.4) for Arg/His or His/His vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>David-Beabes (31) (Cauc)</td>
<td>2001</td>
<td>United States</td>
<td>180/461</td>
<td>Population</td>
<td>Frequency</td>
<td>Overall: null</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;20 Cig/day: 0.6 (0.4 to 1.1) for Arg/Gln vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&gt;20 Cig/day: 0.4 (0.2 to 1.0) for Gln/Gln vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>for Gln/Gln vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Divine (32)</td>
<td>2001</td>
<td>United States</td>
<td>172/143</td>
<td>Hospital</td>
<td>None</td>
<td>Adenocarcinoma among non-Hispanics: 2.81 (1.2 to 7.9) for Gln/Gln vs. Arg/Arg or Arg/Gln</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: null</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: null</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: 1.3 (1.0 to 1.8) for Gln/Gln vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: 1.3 (1.0 to 1.8) for Gln/Gln vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: 1.3 (1.0 to 1.8) for Gln/Gln vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: 1.3 (1.0 to 1.8) for Gln/Gln vs. Arg/Arg</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Overall: 1.3 (1.0 to 1.8) for Gln/Gln vs. Arg/Arg</td>
</tr>
</tbody>
</table>

*SNP = single-nucleotide polymorphism; SCC = squamous cell carcinoma; SCLC = small-cell lung cancer; AA = African American; Cig/day = cigarettes smoked per day; Cauc = Caucasian; PY = pack-years; Hosp. healthy = healthy subjects recruited in the hospitals or health check-up center.†Odds ratios (95% confidence intervals).

Overall, the average participation rate was 91.0% among case patients and 91.2% among control subjects; the Warsaw center had the lowest participation rates for case patients (80.4%) and control subjects (79.8%). Case patients and control subjects underwent an identical in-person interview during which they completed a detailed questionnaire and provided blood samples. The questionnaire collected information about demographic variables, such as sex, date of birth, and education level; medical history; family history of cancer; history of tobacco consumption, including frequency, intensity, duration, and status; history of alcohol consumption; diet history (using a general food-frequency questionnaire); and occupational history. Blood samples were stored in liquid nitrogen. Of the 2633 case patients and 2884 control subjects who agreed to participate in this study, 2188 case patients (83%) and 2198 control subjects (76%) provided blood samples during the interview. Only subjects for whom DNA was available were included in the analyses.

### Laboratory Techniques

Genomic DNA was extracted from blood samples with the use of a QIAamp 96 DNA Blood Kit (Qiagen, Hilden, Germany); DNA concentrations were measured by using PicoGreen dsDNA quantification kits (Molecular Probes, Leiden, The Netherlands). All polymorphisms except for the XRCC1 Arg399Gln polymorphism were analyzed by a polymerase chain reaction (PCR)-based fluorescence 5′ exonuclease assay (i.e., the TaqMan assay) (37). Briefly, 10-ng aliquots of genomic DNA were placed into separate wells of a 96-well or 384-well PCR plate, along with a PCR cocktail that included both fluorescently labeled allele-specific probes (200 nM each), forward and reverse primers (1 μM each), and 5 μL of 2 × TaqMan Universal PCR Master Mix (including ROX passive dye as internal control, dNTPs, buffer, and Taq polymerase [Applied Biosystems, Foster City, CA]) in a final volume of 10 μL. PCR amplification was achieved by using
the following thermocycle conditions: 10 minutes at 95 °C, followed by 45 cycles of 30 seconds at 94 °C and 1 minute at 60 °C, followed by storage at 4 °C. The fluorescence of the PCR products was then plotted, and genotype was determined according to the signal of the two probes. The XRCC1 Arg399Gln polymorphism was genotyped with the use of the PCR-based AmpliFluor assay, as previously described (38). PCR amplification for XRCC1 Arg399Gln was achieved by using the following thermocycle conditions: 4 minutes at 96 °C, followed by 20 cycles of 10 seconds at 96 °C, 5 seconds at 55 °C, 10 seconds at 72 °C, followed by 22 cycles of 10 seconds at 96 °C, 20 seconds at 55 °C, 40 seconds at 72 °C, followed by storage at 4 °C. The sequences of PCR primers and probes are available online (http://jnccancerspectrum.oupjournals.org/jnci/content/vol97/issue8).

To ensure quality control, DNA samples from case patients and control subjects were randomly distributed on each PCR plate, and all genotyping was conducted by personnel who were blinded to the case–control status of the DNA sample. DNA samples from individuals of known base excision repair genotypes were added to each PCR plate as internal controls to ensure the validity of the genotyping. We randomly selected 10% of the study subjects (i.e., both case patients and control subjects) and re-analyzed their DNA samples for each polymorphism to examine the reliability of the genotyping assays.

**Statistical Analysis**

The frequency distributions of demographic variables and putative risk factors for lung cancer, including country of residence, age at recruitment (which, for case patients, was a proxy for age at diagnosis), sex, highest education level, and smoking status, were examined for case patients and control subjects. Former smokers were defined as smokers who stopped smoking at least 2 years before the interview. Tobacco consumption included smoking of cigarettes, pipes, and/or cigars. Cumulative tobacco consumption was calculated by multiplying smoking duration (in years) by smoking intensity (in the equivalent of cigarette packs) and expressed as pack-years. We categorized the subjects as light (≤14 pack-years), moderate (>14–38.26 pack-years), or heavy (>38.26 pack-years) smokers based on the quartiles of cumulative tobacco consumption among the control group.

We tested the Hardy–Weinberg equilibrium of allele distributions separately among case patients and control subjects. The minimum detectable odds ratio was calculated for each sequence variant based on its genotype frequency, our study sample size, and a statistical power of 80%, as previously described (39). Our study had an 80% power to detect a minimum OR of 1.4 for rare polymorphisms such as OGG1 Ser326Cys and a minimum OR of 1.2 for common polymorphisms such as XRCC1 Arg399Gln. We used unconditional multivariable logistic regression analysis to examine associations between genetic polymorphisms and lung cancer risk by estimating odds ratios (ORs) and 95% confidence intervals (CIs). Genotypes were categorized into three groups (major allele homozygous, heterozygous, and homozygous variant) when the allele frequencies allowed. In the stratified analyses and interaction analyses, genotypes were dichotomized into two categories (risk versus nonrisk genotypes) on the basis of prior knowledge of the functional significance of the variants, existing epidemiologic evidence, and the frequency distributions of the variants. The prior risk genotypes (i.e., genotypes that were hypothesized to be associated with increased risk) are as follows: OGG1 Cys/Cys; XRCC1 Arg/Arg (for the Arg194Trp polymorphism); Arg/His and His/His (for the Arg280His polymorphism); and Arg/Gln and Gln/Gln (for the Arg399Gln polymorphism).

We estimated the linkage disequilibrium by Lewontin’s D′ (40) and used expectation-maximization algorithms to estimate the haplotype frequency for the three polymorphisms of XRCC1 as previously described (41,42).

We conducted stratified analyses by histology to investigate associations between genetic polymorphisms and histologic subtypes of lung cancer. We also evaluated the modulating effects of cumulative tobacco exposure and age at diagnosis by comparing the stratum-specific risk estimates. Our rationale for stratifying by age at diagnosis is that one would expect a genetic factor to have a strong effect on cancer development among subjects with a young age of diagnosis than among subjects with an old age at diagnosis. The cut point of 50 years for age at diagnosis was chosen based on the age distribution at lung cancer diagnosis in Europe (43).

To evaluate the joint effect of the two designated risk genotypes (i.e., all combinations of every pair of risk genotypes among the four polymorphisms), we used subjects with two nonrisk genotypes as the reference group and compared the effects of having either one or both risk genotypes. Matching variables and potential confounders, including country of residence, age (continuous), sex, and pack-years of smoking (continuous), were included in all multivariable logistic regression models. For all interaction analyses, we tested departure from multiplicativity by estimating the interaction odds ratio, which was derived from the product term in a saturated model along with the covariates in the logistic model.

We conducted multigenic analyses, in which the independent variable was the number of prior risk genotypes a subject carried and was treated as a categorical variable or as a continuous variable. All statistical analyses were conducted using STATA software (version 8.0, StataCorp LP, College Station, TX). All statistical tests were two-sided.

We estimated the false-positive report probability for statistically significant observations using the methods described by Wacholder et al. (36). The prior probability that the association between a genetic variant and a disease is real is likely to be influenced by knowledge of the biologic function of a gene, the functional significance of variants, and the available epidemiologic evidence. However, prior probability is a subjective measure that can vary among investigators depending on the importance that is assigned to different pieces of evidence. For this reason, we calculated the false-positive report probabilities for prior probabilities ranging from 50% to 0.1%. We considered that a prior probability of 50% may be appropriate when there is very strong biologic plausibility that the association is real and consistent epidemiologic evidence for such an association exists and that a prior probability of 0.1% may be appropriate when both biologic knowledge and epidemiologic data are inadequate. For example, OGG1—326Cys protein has been shown to be less capable of repairing 8-oxoG lesions than the OGG1—326Ser protein (12), and several epidemiologic studies have reported an association between the OGG1 326Cys allele and lung cancer risk (Table 1). The combination of this evidence suggested a likely high prior probability of 10%–25%. However, because limited data are available on the function of the XRCC1 280His allele, we considered the most likely prior probability to be low (i.e., 0.1%–1.0%).
RESULTS

Table 2 shows the frequency distribution of demographic characteristics and the putative risk factors for lung cancer among the study participants. Control subjects were slightly more educated than case patients. Smoking was much more prevalent among case patients than among control subjects. The percentages of case patients and control subjects from the six countries ranged from 7% to 32%.

The frequencies of the OGG1 Ser326Cys, XRCC1 Arg194Trp, XRCC1 Arg280His, and XRCC1 Arg399Gln alleles among the control group were 20%, 7.4%, 4.8%, and 35%, respectively, and their distributions in the control group were in Hardy–Weinberg equilibrium, with \( P \) values of .22, .93, .59, and .11, respectively. The results of the initial and the repeat genotyping analyses were at least 99.4% concordant for the four SNPs. Due to the nature of PCR amplification, so their genotypes could not be determined. Data for OGG1 Ser326Cys, XRCC1 Arg194Trp, XRCC1 Arg280His, and XRCC1 Arg399Gln were missing from at least 99.4% concordant for the four SNPs. Due to the nature of the high-throughput fluorescence 5' exonuclease assay, a small fraction of the subjects could not be classified into one of the three genotype clusters (i.e., wild type/wild type, wild type/variant, variant/variant) during the genotyping process because of poor DNA quality or insufficient PCR amplification, so their genotypes could not be determined. Data for OGG1 Ser326Cys, XRCC1 Arg194Trp, XRCC1 Arg280His, and XRCC1 Arg399Gln were missing from 1.6%, 2.4%, 4.7%, and 7.3% of the subjects, respectively.

Table 3 shows associations between OGG1 Ser326Cys polymorphisms and lung cancer risk, stratified by histology, smoking status, and age at diagnosis. The overall OR for the OGG1 Cys/Cys genotype was 1.34 (95% CI = 0.95 to 1.88). When we stratified by histology, the association between the Cys/Cys genotype and lung cancer risk was most prominent for adenocarcinoma risk (OR = 1.66, 95% CI = 1.04 to 2.66). When we stratified by age at diagnosis, there was no evidence of effect modification: among subjects who were recruited or diagnosed with lung cancer at age 50 years or younger, those with the Cys/Cys genotype had an OR of 2.10 (95% CI = 0.78 to 5.62), and among those older than 50 years, those with the Cys/Cys genotype had an OR of 1.28 (95% CI = 0.89 to 1.84) (\( P = .36; \) test for heterogeneity).

Table 4 shows associations between the XRCC1 Arg194Trp, Arg280His, and Arg399Gln polymorphisms and lung cancer risk. None of the XRCC1 polymorphisms was associated with the risk of lung cancer in the whole study population. Because of the rarity of the 194Trp and 280His alleles, the probability of two or more variants being present on the same chromosome is close to null. Therefore, no additional information was obtained from the haplotype analyses. Lewontin’s \( D' \) values of 194Trp-280His, 194Trp-399Gln, and 280His-399Gln were −0.99, −0.95, and −0.96, respectively. When we stratified by pack-years of cumulative tobacco consumption based on quartiles of consumption among the control subjects, the XRCC1 Arg194Trp (OR = 0.65, 95% CI = 0.46 to 0.93) and Arg280His (OR = 0.56, 95% CI = 0.36 to 0.86) polymorphisms were each associated with a decreased risk of lung cancer among heavy smokers (i.e., those with >38.26 pack-years of consumption) compared with common allele homozygotes.

Table 2. Frequency distribution of demographic variables and putative risk factors for lung cancer among study participants*

<table>
<thead>
<tr>
<th>Variable or risk factor</th>
<th>Case patients (N = 2188), N (%)</th>
<th>Control subjects (N = 2198), N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Country</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Romania</td>
<td>161 (7)</td>
<td>213 (10)</td>
</tr>
<tr>
<td>Hungary</td>
<td>343 (16)</td>
<td>296 (13)</td>
</tr>
<tr>
<td>Poland</td>
<td>686 (31)</td>
<td>705 (32)</td>
</tr>
<tr>
<td>Russia</td>
<td>406 (19)</td>
<td>323 (15)</td>
</tr>
<tr>
<td>Slovakia</td>
<td>293 (13)</td>
<td>215 (10)</td>
</tr>
<tr>
<td>Czech Republic</td>
<td>299 (14)</td>
<td>446 (20)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>483 (22)</td>
<td>581 (26)</td>
</tr>
<tr>
<td>Male</td>
<td>1705 (78)</td>
<td>1617 (74)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤40</td>
<td>26 (1)</td>
<td>60 (3)</td>
</tr>
<tr>
<td>41–50</td>
<td>313 (14)</td>
<td>320 (15)</td>
</tr>
<tr>
<td>51–60</td>
<td>691 (32)</td>
<td>676 (31)</td>
</tr>
<tr>
<td>61–70</td>
<td>829 (38)</td>
<td>796 (36)</td>
</tr>
<tr>
<td>&gt;70</td>
<td>329 (15)</td>
<td>346 (16)</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤Elementary (obligatory)</td>
<td>227 (10)</td>
<td>179 (8)</td>
</tr>
<tr>
<td>Technical school</td>
<td>765 (35)</td>
<td>692 (32)</td>
</tr>
<tr>
<td>High school</td>
<td>529 (24)</td>
<td>562 (26)</td>
</tr>
<tr>
<td>Beyond high school</td>
<td>423 (19)</td>
<td>435 (20)</td>
</tr>
<tr>
<td>≥University</td>
<td>239 (11)</td>
<td>325 (15)</td>
</tr>
<tr>
<td>Missing</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>167 (8)</td>
<td>736 (34)</td>
</tr>
<tr>
<td>Former</td>
<td>413 (19)</td>
<td>630 (29)</td>
</tr>
<tr>
<td>Current</td>
<td>1604 (73)</td>
<td>819 (37)</td>
</tr>
<tr>
<td>Missing</td>
<td>4</td>
<td>13</td>
</tr>
</tbody>
</table>

*Some percentages do not total 100% because of rounding.
Table 4. Association between XRCC1 polymorphisms and lung cancer risk, stratified by histology, cumulative tobacco consumption, and age at diagnosis *

<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. case patients/ no. control subjects</th>
<th>OR (95% CI)</th>
<th>Genotype</th>
<th>No. case patients/ no. control subjects</th>
<th>OR (95% CI)</th>
<th>Genotype</th>
<th>No. case patients/ no. control subjects</th>
<th>OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
<td>Total</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>1878/1828</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>1901/1896</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>844/874</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>R/W</td>
<td>259/292</td>
<td>0.89 (0.73 to 1.08)</td>
<td>R/H</td>
<td>181/190</td>
<td>0.93 (0.74 to 1.18)</td>
<td>R/Q</td>
<td>951/881</td>
<td>1.06 (0.92 to 1.22)</td>
</tr>
<tr>
<td>W/W</td>
<td>10/12</td>
<td>1.16 (0.46 to 2.92)</td>
<td>H/H</td>
<td>6/6</td>
<td>1.25 (0.38 to 4.08)</td>
<td>Q/Q</td>
<td>254/260</td>
<td>0.95 (0.77 to 1.18)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>269/304</td>
<td>0.90 (0.74 to 1.09)</td>
<td>R/H or H/H</td>
<td>187/196</td>
<td>0.94 (0.75 to 1.19)</td>
<td>R/Q or Q/Q</td>
<td>1205/1141</td>
<td>1.03 (0.90 to 1.18)</td>
</tr>
<tr>
<td>Histology</td>
<td></td>
<td></td>
<td>Histology</td>
<td></td>
<td></td>
<td>Histology</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC</td>
<td></td>
<td></td>
<td>SCC</td>
<td></td>
<td></td>
<td>SCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>796/1828</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>800/1896</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>347/874</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>111/304</td>
<td>0.87 (0.67 to 1.13)</td>
<td>R/H or H/H</td>
<td>76/196</td>
<td>0.88 (0.65 to 1.20)</td>
<td>R/Q</td>
<td>401/881</td>
<td>1.10 (0.91 to 1.33)</td>
</tr>
<tr>
<td>Adenocarcinoma</td>
<td></td>
<td></td>
<td>Adenocarcinoma</td>
<td></td>
<td></td>
<td>Adenocarcinoma</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>428/1828</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>436/1896</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>206/874</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>62/304</td>
<td>0.89 (0.65 to 1.20)</td>
<td>R/H or H/H</td>
<td>44/196</td>
<td>0.88 (0.61 to 1.27)</td>
<td>Q/Q</td>
<td>208/881</td>
<td>0.97 (0.77 to 1.21)</td>
</tr>
<tr>
<td>Cumulative tobacco consumption†</td>
<td></td>
<td></td>
<td>Cumulative tobacco consumption†</td>
<td></td>
<td></td>
<td>Cumulative tobacco consumption†</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td></td>
<td></td>
<td>Never</td>
<td></td>
<td></td>
<td>Never</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>135/614</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>142/637</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>71/296</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>26/104</td>
<td>1.47 (0.88 to 2.45)</td>
<td>R/H or H/H</td>
<td>15/62</td>
<td>1.15 (0.60 to 2.18)</td>
<td>R/Q</td>
<td>66/286</td>
<td>0.98 (0.66 to 1.46)</td>
</tr>
<tr>
<td>Light‡</td>
<td></td>
<td></td>
<td>Light‡</td>
<td></td>
<td></td>
<td>Light‡</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>124/303</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>127/325</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>20/92</td>
<td>0.83 (0.46 to 1.48)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>21/54</td>
<td>0.88 (0.49 to 1.56)</td>
<td>R/H or H/H</td>
<td>14/34</td>
<td>0.88 (0.44 to 1.76)</td>
<td>R/Q</td>
<td>63/152</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>Moderate$</td>
<td></td>
<td></td>
<td>Moderate$</td>
<td></td>
<td></td>
<td>Moderate$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>836/607</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>840/622</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>384/287</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>120/87</td>
<td>1.01 (0.75 to 1.37)</td>
<td>R/H or H/H</td>
<td>101/61</td>
<td>1.17 (0.84 to 1.65)</td>
<td>R/Q</td>
<td>423/290</td>
<td>1.08 (0.87 to 1.35)</td>
</tr>
<tr>
<td>Heavy</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heavy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>777/297</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>785/306</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>120/87</td>
<td>1.02 (0.74 to 1.40)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>101/58</td>
<td>0.65 (0.46 to 0.93)</td>
<td>R/H or H/H</td>
<td>57/38</td>
<td>0.56 (0.36 to 0.86)</td>
<td>Q/Q</td>
<td>395/157</td>
<td>1.05 (0.80 to 1.39)</td>
</tr>
<tr>
<td>Age at diagnosis, y</td>
<td></td>
<td></td>
<td>Age at diagnosis, y</td>
<td></td>
<td></td>
<td>Age at diagnosis, y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤50</td>
<td></td>
<td></td>
<td>≤50</td>
<td></td>
<td></td>
<td>≤50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>294/311</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>295/333</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>128/135</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>38/60</td>
<td>0.71 (0.44 to 1.14)</td>
<td>R/H or H/H</td>
<td>37/33</td>
<td>1.24 (0.72 to 2.11)</td>
<td>R/Q</td>
<td>153/164</td>
<td>0.97 (0.68 to 1.39)</td>
</tr>
<tr>
<td>&gt;50</td>
<td></td>
<td></td>
<td>&gt;50</td>
<td></td>
<td></td>
<td>&gt;50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R/R</td>
<td>1584/1517</td>
<td>1.00 (referent)</td>
<td>R/R</td>
<td>1606/1563</td>
<td>0.90 (0.70 to 1.17)</td>
<td>R/R</td>
<td>716/739</td>
<td>1.00 (referent)</td>
</tr>
<tr>
<td>R/W or W/W</td>
<td>231/244</td>
<td>0.95 (0.77 to 1.18)</td>
<td>R/H or H/H</td>
<td>150/163</td>
<td>0.90 (0.70 to 1.17)</td>
<td>Q/Q</td>
<td>798/717</td>
<td>1.08 (0.92 to 1.26)</td>
</tr>
</tbody>
</table>

* Odds ratios (ORs) adjusted for country, age at diagnosis, sex, and cumulative tobacco consumption (in pack-years). CI = confidence interval; R = Arg; W = Trp; H = His; Q = Gln; SCC = squamous cell carcinoma.
† Odds ratios adjusted for country, age at diagnosis, and sex.
‡ First quartile (≤14 pack-years).
§ Second and third quartiles (>14 – 38.26 pack-years).
|| Fourth quartile (>38.26 pack-years).
OGG1 and XRCC1 polymorphisms do not play a major role in
respectively. Overall, however, our results suggest that the
ported by results of our false-positive report probability analy-
likelihood that these results represent true associations was sup-
with heavy smokers who carried the Arg/Arg genotype. The
Ser genotype and that heavy smokers who carried the XRCC1
underpowered studies. We found that subjects who carried the
association.

Of cut point. Because our statistically signi-
fi
significance of both the OGG1 326Cys and XRCC1 194Trp al-
stances, especially
given that the distribution of the OGG1 Ser326Cys polymor-
study in a Hawaiian population that comprised Caucasian
Japanese individuals found that the OGG1 Cys/Cys geno-
type was associated with both adenocarcinoma and squamous
cell carcinoma risks, with the association being stronger for the
latter (18). By contrast, we found that the OGG1 Cys/Cys
genotype was associated mainly with adenocarcinoma risk.
This discrepancy might reflect the fact that our study population
included subjects of different ethnicities and environmental ex-
possibilities that may have modified the associations, especially
given that the distribution of the OGG1 Ser326Cys polymor-
phism differs among Chinese/Japanese and Caucasian popula-
tions (17–20). The frequency of the Cys allele among the control
subjects in our study (20%) was consistent with the range of
frequencies of this allele previously reported for Caucasians
[i.e., 13% to 22% (18,19,46)] but not with that previously re-
ported for Japanese and Chinese populations [i.e., 40% to 47%
(17,18,20,47)]. Another possible reason for the discrepancy in
the OGG1 risk associations is that the previous studies did not
have sufficient statistical power to detect an association with
adenocarcinoma risk.

We observed no interactions between any two risk genotypes.
Having an increased number of prior risk genotypes was not as-
associated with an increased risk of lung cancer. Subjects who car-
ried at least three prior risk genotypes had an OR of 1.34 (95% CI = 0.91 to 1.97) compared with subjects who carried none of
the prior risk genotypes (data not shown).

Table 5 shows the false-positive report probability for the
three statistically significant associations we observed. In
general, at a false-positive report probability cut point of 0.5,
the observed odds ratio for an association is likely to reflect a
true association with a prior probability greater than 10%.
However, an association with prior probability of 1% or lower
is likely to be a false-positive finding. On the other hand, us-
ing a more stringent false-positive report probability cut point
of 0.2, interpretations of the observed odds ratios with prior
probability of 10% or lower would be influenced by the change
of cut point. Because our statistically significant findings were
restricted to subgroups, a more stringent cut point for false-
positive report probability may be appropriate. Given the
available epidemiologic data and the known functional sig-
nificance of both the OGG1 326Cys and XRCC1 194Trp al-
leles, a prior probability of at least 10% is not unreasonable
for these two SNPs. However, given the limited data on the
function of the XRCC1 280His allele, a lower prior probability
may be appropriate (e.g., 1%–0.1%), and thus, the observed
odds ratio among heavy smokers may reflect a false-positive
association.

**Discussion**

We conducted a large-scale association study to reduce the
likelihood of false-positive and false-negative results typical of
underpowered studies. We found that subjects who carried the
OGG1 Cys/Cys genotype had a small increased risk of lung
adenocarcinoma compared with subjects who carried the Ser/
Ser genotype and that heavy smokers who carried the XRCC1
194Trp allele had a decreased risk of lung cancer compared
with heavy smokers who carried the Arg/Arg genotype. The
likelihood that these results represent true associations was sup-
ported by results of our false-positive report probability analy-
sis, which was based on prior probabilities of 25% and 10%,
respectively. Overall, however, our results suggest that the
OGG1 and XRCC1 polymorphisms do not play a major role in
lung carcinogenesis. Nonetheless, the association we observed
between the OGG1 Cys/Cys genotype and adenocarcinoma risk
is consistent with experimental evidence showing that the
OGG1 isoform encoded by the Cys allele exhibits decreased
base excision repair activity (12,44). Likewise, the association
we observed between the XRCC1 194Trp allele and lung can-
cer risk in heavy smokers is consistent with the evidence show-
ing that this allele is less sensitive to mutagens than the XRCC1
194Arg allele (25).

Results of a previous study in a Japanese population revealed
an association between squamous cell lung cancer and the
OGG1 Cys/Cys genotype (17). Paz-Elizur et al. (45) showed that
reduced OGG1 activity is associated with the occurrence of
non–small-cell lung cancer in an Israeli population (45). Results
of a study in a Hawaiian population that comprised Caucasian
and Japanese individuals found that the OGG1 Cys/Cys geno-
type was associated with both adenocarcinoma and squamous
cell carcinoma risks, with the association being stronger for the
latter (18). By contrast, we found that the OGG1 Cys/Cys
genotype was associated mainly with adenocarcinoma risk.
This discrepancy might reflect the fact that our study population
included subjects of different ethnicities and environmental ex-
positions that may have modified the associations, especially
given that the distribution of the OGG1 Ser326Cys polymor-
phism differs among Chinese/Japanese and Caucasian popula-
tions (17–20). The frequency of the Cys allele among the control
subjects in our study (20%) was consistent with the range of
frequencies of this allele previously reported for Caucasians
[i.e., 13% to 22% (18,19,46)] but not with that previously re-
ported for Japanese and Chinese populations [i.e., 40% to 47%
(17,18,20,47)]. Another possible reason for the discrepancy in
the OGG1 risk associations is that the previous studies did not
have sufficient statistical power to detect an association with
adenocarcinoma risk.

We found that the XRCC1 Arg399Gln polymorphism was not
associated with lung cancer risk and that the XRCC1
Arg194Trp and Arg280His polymorphisms were each associ-
ated with a decreased risk of lung cancer among heavy smokers.
Several epidemiologic studies have examined XRCC1 polymor-
phisms and lung cancer risk, but only limited data have been
available for the XRCC1 Arg280His polymorphism. Results of
previous studies that examined the association between the
XRCC1 Arg399Gln polymorphism and lung cancer risk were
inconsistent, possibly owing to the large random error in those
small studies (31–34,48–31). Our finding that XRCC1 Arg194Trp
variants are associated with decreased lung cancer risk among
heavy smokers is consistent with the results of a functional study
of the 194Trp allele (25) and with results of a previous study
reporting that the 194Trp allele was associated with a decreased
risk of lung cancer among subjects who smoked more than 20
cigarettes/day (31).
The molecular basis of the association between the 194Trp allele and decreased lung cancer risk in heavy smokers remains to be established. It is possible that the variant protein is associated with increased repair activity and that this increase is influenced by gene–environment interactions. In support of such a model, it has been shown that individuals homozygous for the common allele (Arg/Arg) have substantially more chromosome breaks per cell in mutagen sensitivity assays using both bleomycin and BPDE as challenge mutagens than individuals who carry one or two of the variant (Trp) alleles (25). However, in that study, although healthy heavy smokers (>42 pack-years) had fewer chromosomal breaks than never smokers after treatment with bleomycin or BPDE, the association with genotype was not examined (25). Lower levels of 8-oxoG in lymphocytes of smokers compared with nonsmokers, which could be explained by the presence of efficient DNA repair processes for the oxidative damage induced by smoking, have also been reported in some studies (52,53).

Our results for the XRCC1 Arg399Gln polymorphism did not replicate the findings of Zhou et al. (34), who reported that the XRCC1 399Gln/399Gln genotype was associated with a decreased risk of lung cancer among heavy smokers. This discrepancy might be due, in part, to differences in the levels of tobacco consumption between the two populations: 491 case patients and 131 control subjects (11% of study population) in the study by Zhou et al. (34) had at least 55 pack-years of tobacco consumption compared with 272 case patients and 107 control subjects (5%) in our study population. The smaller percentage of heavy smokers in our study population might have limited our ability to detect an association, especially if the association between the 399Gln allele and a decreased risk of lung cancer risk is limited to very heavy smokers. Differences between two study populations in the prevalence of other potential effect modifiers (e.g., occupational exposures to lung carcinogens) that were not accounted for might have also contributed to the discrepancy.

Our study has several limitations. First, potential selection bias can occur when subjects who agree to participate in the study have characteristics that differ from those of subjects who are eligible for the study. The major reason for nonparticipation in this study was the failure of some subjects to provide written informed consent (6.7% of case patients and 7.3% of control subjects), which might represent a form of self-selection. Bias from self-selection may affect estimates of exposure to environmental factors; however, it is unlikely that self-selection would be related to a subject’s genotype. Results of a simulation study (54) suggested that selection factors that are related only to environmental factors might still lead to a biased estimate of genetic main effects if the environmental factors modify the genetic effects. Nevertheless, assuming that genotype does not influence subject participation conditional on exposures and disease, the genotype–environment interaction odds ratio should not be influenced (54). Furthermore, bias in the estimate of the genetic main effect due to the selection factor of environmental exposures can be adjusted for in the analyses by treating the environmental factors as potential confounders (54).

Second, only 83% of the case patients and 76% of the control subjects provided DNA samples. A bias may have occurred if the availability of DNA was related to both genotype and disease status. For example, if subjects who did not provide blood samples had more advanced stages of illness than subjects who provided blood samples and if the stages of illness are related to the genotype, then estimates of the genetic associations would be biased. However, the general demographic characteristics and smoking habits of subjects who provided DNA samples and the entire study population were similar, suggesting that our study was not limited by this potential source of bias.

Third, although recall bias is not relevant for the estimation of genetic associations, it can affect estimates of environmental exposures. Self-reports of lifestyle history are subject to differential recall between case patients and control subjects, which may bias the estimates toward or away from the null. Misclassification on the environmental exposures may subsequently bias estimates of gene–environment interactions (55). The bias for a multiplicative interaction term will be toward the null value if 1) both environmental and genetic factors are binary and independent, 2) the exposure measurement is better than random (the sum of sensitivity and specificity is greater than or equal to 1), and 3) misclassification is independent of the genetic factor (55). In other scenarios, the direction of bias cannot be easily predicted.

Another limitation is the missing data generated in the process of cluster identification of genotyping. These missing data were present in a random manner regarding demographic variables and smoking status, and the consequence of these missing data was to reduce the statistical power of our study. The missing data were associated with country: the quality of the DNA samples varied from one country to another. Nevertheless, the missing data were not likely to bias the estimates to a meaningful extent because their proportions were low and country was controlled for in the multivariable models.

Yet another potential limitation of our study was that the case patients and control subjects were not matched by ethnicity. However, more than 90% of our study population was of European ancestry. In addition, even in situations in which case patients and control subjects have very different disease prevalences and allele frequencies (e.g., when case patients and control subjects have different ethnic distributions), population stratification is unlikely to lead to substantial bias in well-designed case–control studies (56). Therefore, it is unlikely to be a concern in our study.

Our study has a number of strengths. First, the large sample size gives high precision and therefore is less susceptible to fluctuating results. In addition, this is the first study to use the false-positive report probability to interpret the results. Estimates of the false-positive report probability can be used to decide whether a statistically significant finding is noteworthy based on prior probabilities; however, prior probability varies among individuals, and investigators who use different prior probabilities may reach different conclusions. Publication bias may also influence the choice of prior probability because the published data may not be representative of the true association. However, we tested a range of prior probabilities, which allowed us to identify how sensitive our positive findings were to changing prior probabilities.

Given that our study had an 80% power to detect a minimum OR of 1.4 for rare polymorphisms such as OGG1 Ser326Cys, and a minimum OR of 1.2 for common polymorphisms such as XRCC1 Arg399Gln, our results do not support a major independent role for any of the polymorphisms investigated in this study in lung cancer risk. However, we cannot exclude the possibility that OGG1 Ser326Cys has a minor role in lung carcinogenesis and that the XRCC1 Arg194Trp polymorphism has a minor role in lung carcinogenesis among extremely heavy smokers.
REFERENCES


Downloaded from https://academic.oup.com/jnci/article-abstract/97/8/567/2544192 by guest on 05 March 2019


Notes

R. Hung and S. Borel worked on this study under the tenure of Special Training Awards from the International Agency for Research on Cancer. The study was funded by a National Cancer Institute R01 grant (contract CA 092039-01A2), and an Association for International Cancer Research grant (contract 03-281).

Manuscript received July 22, 2004; revised February 10, 2005; accepted February 25, 2005.