Int. J. Cancer: **118,** 714–720 (2006)

Polymorphism of *xeroderma pigmentosum group G* and the risk of lung cancer and squamous cell carcinomas of the oropharynx, larynx and esophagus

Yan Cui[1], Hal Morgenstern[2], Sander Greenland[1,3], Donald P. Tashkin[4], Jenny Mao[4], Wei Cao[1], Wendy Cozen[5], Thomas M. Mack[5] and Zuo-Feng Zhang[1]*

[1]*Department of Epidemiology, University of California at Los Angeles, Los Angeles, CA, USA*
[2]*Department of Epidemiology, University of Michigan, Ann Arbor, MI, USA*
[3]*Department of Statistics, University of California at Los Angeles, Los Angeles, CA, USA*
[4]*Division of Pulmonary and Critical Care Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA*
[5]*Department of Preventive Medicine, Keck School of Medicine at University of Southern California, Los Angeles, CA, USA*

We investigated the effects of *XPG* His1104Asp polymorphism (rs17655) on the risk of lung cancer and squamous cell carcinomas of the oropharynx, larynx and esophagus (SCCOLE). This population-based case-control study involves 611 new cases of lung cancer, 601 new cases of oropharyngeal, laryngeal and esophageal cancers, and 1,040 cancer-free controls. The *XPG* polymorphism was assayed by PCR-RFLP method for 497 lung cancer cases, 443 cases of oropharyngeal, laryngeal and esophageal cancers and 912 controls. Binary and polytomous unconditional logistic regression models were fitted to assess the main effects and the effect modifications between the polymorphism and environmental exposures. With the adjustment for potential confounders, the *XPG* Asp1104Asp genotype was inversely associated with lung cancer (odds ratio [OR] = 0.62, 95% confidence limits [CL] = 0.38, 1.0) and SCCOLE (OR = 0.47, 95% CL = 0.27, 0.82), with the combined His1104His and His1104Asp genotypes as the referent. With subjects having genotype Asp1104Asp and no tobacco smoking exposure as the common referent, the ORs on lung cancer were 13 (95% CL = 4.4, 37) for heavy tobacco smoking (>20 pack-years), 1.9 (95% CL = 0.78, 4.5) for having at least one copy of 1104His, and 23 (95% CL = 9.5, 56) for the joint effect, respectively. Compared to non-smokers with the Asp1104Asp genotype, the adjusted OR on SCCOLE for heavy smokers (>20 pack-years) having at least one copy of 1104His was 8.0 (95% CL = 2.7, 24). Similarly, compared to non-drinkers with the Asp1104Asp genotype, the adjusted OR on SCCOLE for heavy drinkers (≥3 drinks/day) with at least one copy of 1104His was 10 (95% CL = 2.7, 38). In conclusion, our study suggests that the *XPG* Asp1104Asp genotype may be associated with decreased susceptibility to lung cancer and SCCOLE.

Key words: *XPG*; polymorphism; lung cancer; SCCOLE

Various DNA alterations can be caused by exposure to carcinogenic and cytotoxic compounds from environmental agents and endogenous metabolites.[1] Most of these alterations, if unrepaired, may result in genetic instability, mutagenesis, and cell death. DNA repair mechanisms are important for maintaining DNA integrity and preventing oncogenesis. Nucleotide excision repair (NER) can remove bulky adducts induced by environmental agents such as ultraviolet irradiation, tobacco smoking, and dietary factors. NER is one of the most important DNA repair pathways against tobacco-related carcinogenesis.[2] It has been postulated that the phenotype of DNA repair capacity varies in the general population, and that individuals with suboptimal DNA repair capacity are at an increased risk of developing various cancers.[3] Genetic polymorphisms in the coding and regulatory sequences of repair pathway components are among the major contributors to the variation of DNA repair capacity, thereby modifying cancer susceptibility.[4]

Xeroderma Pigmentosum Group G (*XPG*), one important component of the NER pathway, encodes a structure-specific endonuclease catalyzing 3′ incision and involves the subsequent 5′ incision by ERCC1-XPF heterodimer.[5,6] It is also involved in the stabilization of a pre-incision complex on the damaged DNA.[7] *XPG* gene maps to chromosome 13q33.[8] One common non-synonymous single nucleotide polymorphism (SNP) in its coding region has been identified, which causes the amino acid change of His1104Asp (rs17655). The His1104Asp polymorphism locates in the XPG C-terminus, which is required for the interactions between XPG and other components of the NER pathway such as TFIIH subunits.[9] The His1104Asp amino acid change may influence these protein-protein interactions and consequently predispose individuals with the suboptimal variants to various cancers. To our knowledge, there has been no study published on the potential effect of the *XPG* codon 1104 polymorphism on the risk of lung cancer or squamous cell carcinomas of the oropharynx, larynx, and esophagus (SCCOLE) in the United States population.

We investigated the effects of the *XPG* His1104Asp polymorphism on the risk of developing lung cancer and SCCOLE in a population-based case-control study. We also investigated potential modifications of tobacco- and alcohol-induced carcinogenesis of the above-mentioned cancers by the polymorphism.

Material and methods

Study design and study subjects

This population-based case-control study comprised 611 new cases of lung cancer, 601 new cases of oropharyngeal, laryngeal and esophageal cancers (303 oral cancers, 100 pharyngeal cancers, 90 laryngeal cancers and 108 esophageal cancers), and 1,040 cancer-free controls. Histologically confirmed new cases were obtained by the rapid ascertainment system of the Cancer Surveillance Program for Los Angeles County, which is administered by the Keck School of Medicine and Norris Comprehensive Cancer Center at the University of Southern California. The USC Cancer Surveillance Program is the population-based cancer registry for Los Angeles County, which has collected basic clinical and demographic information on all invasive cancers (except non-melanoma skin cancer) diagnosed among residents of Los Angeles County since 1972. Over 95% of cancer reports are histologically verified; the remainder verified by MRI, CT scan or other diagnostic methods. Cases of oropharyngeal, laryngeal, esophageal, and lung can-

Abbreviations: CL, confidence limit; MDA, malondialdehyde; NER, nucleotide excision repair; OR, odds ratio; PAH, polycyclic aromatic hydrocarbons; SCCOLE, squamous cell carcinomas of the oropharynx, larynx and esophagus; SNP, single nucleotide polymorphism; *XPG*, *xeroderma pigmentosum group G*.

Grant sponsor: National Institute of Health; Grant numbers: CA90833, DA11386, CA77954, CA09142, CA96134, ES 011667; Grant sponsor: Alper Research Program for Environmental Genomics.

***Correspondence to:** Department of Epidemiology, University of California, Los Angeles, 71-225 CHS, Box 951772, 650 Charles E. Young Drive, South, Los Angeles, CA, 90095-1772. Fax: +1-310-206-6039.
E-mail: zfzhang@ucla.edu

Received 23 March 2005; Accepted after revision 6 June 2005
DOI 10.1002/ijc.21413
Published online 10 August 2005 in Wiley InterScience (www.interscience.wiley.com).

cers were excluded if they had a previous diagnosis; this information was determined from Cancer Surveillance Program records and verified from cases' self-reports in their interviews. To be eligible as a potential control, an individual must not have had any history of oropharyngeal, laryngeal, esophageal, or lung cancer. Controls were identified by using a formal algorithm to take a census of households in the neighborhood of the case, and matching potentially available controls to the case by age (within 10-year categories) and gender. All study subjects met the following criteria: (*i*) be residents of Los Angeles County at time of diagnosis for cases or at time of recruitment for controls; (*ii*) be 18–65 years of age during the observation period, 1999–2004; and (*iii*) speak either English or Spanish, or have translators available at home. The recruitment rates were 39% (611 of 1,556) for eligible lung cancer cases, 46% (601 of 1,301) for eligible oropharyngeal, laryngeal and esophageal cancer cases and 79% (1,040 of 1,321) for contacted eligible controls, respectively. We could not recruit some eligible cases into our study due to the following reasons: (*i*) the patients died before we contacted them (10% for oropharyngeal, laryngeal and esophageal cancers, 25% for lung cancer); (*ii*) we could not contact the patients due to incorrect addresses (18% for oropharyngeal, laryngeal and esophageal cancers, 14% for lung cancer); (*iii*) the patients were too ill to get interviewed (4% for oropharyngeal, laryngeal and esophageal cancers, 5% for lung cancer); (*iv*) the patients were not willing to participate the study (21% for oropharyngeal, laryngeal and esophageal cancers, 16% for lung cancer); and (*v*) the physicians refused our requests to contact their patients (1% for both oropharyngeal, laryngeal and esophageal cancers and lung cancer). In-person interviews were conducted with each case and control using standardized questionnaires. Each interview was administered by specially trained staff personnel and took about 45–60 min to complete. Both the case and his or her matched control were interviewed by the same staff member; and, to the extent possible, we matched the gender and ethnicity of the subjects with the gender and ethnicity of their interviewers. Self-reported information was collected on a variety of factors known to affect, or might affect, the risk of investigated cancers or that might otherwise be methodologically relevant to the analysis. These factors include socio-demographic characteristics, history of tobacco smoking, environmental tobacco smoking, drug and alcohol use, occupational and environmental exposures, selected clinical factors, diet and other behaviors. The lifetime accumulated history of tobacco smoking and alcohol drinking for study subjects was recorded in the questionnaires, which provided us the information on age of onset, duration of exposures, intensity of exposures, cumulative exposures and years of quitting before diagnosis for cases or interview for controls. After the interview, buccal swabs were collected for DNA extraction. A total of 1,986 buccal cell samples were collected from study subjects. The protocol was approved by the Institutional Review Boards of the University of California at Los Angeles and the University of Southern California. Informed consent was obtained from all study subjects.

Genotype analysis

Buccal cells were collected using a standard protocol by brushing the buccal mucosa, rinsing with mouthwash and expectorating into a plastic container.[10] Genomic DNA was subsequently extracted from the buccal cells using a modified phenol-chloroform protocol.[10] The *XPG* His1104Asp polymorphism was determined by a PCR-RFLP assay described by Jeon *et al.*[11] Briefly, 100 ng of the DNA sample was amplified using 0.2 μM primers (5′-GACCTGCCTCTCAGAATCATC-3′ and 5′-CCTCGCACGTCTTAGTTTCC-3′), 20 μM dNTP, 1 U Taq DNA polymerase (Promega, Madison, WI), and 1.5 mM $MgCl_2$ in a total volume of 20 μl. Thermal cycling was carried out under the following conditions: initial denaturation at 95°C for 5 min; subsequently 35 cycles of 94°C for 30 sec, 60°C for 45 sec and 72°C for 45 sec; and a final elongation step of 72°C for 5min. The PCR product (271 bp fragment) was then digested for 6 hr with 10 U of *Nla*III (New England Biolabs Inc., Beverly, MA) at 37°C and followed by resolution process on 4% agarose gel stained with ethidium bromide. The His/His genotype was determined by the presence of 2 bands at 227 and 44 bp, the Asp/Asp genotype by the presence of an uncut 271 bp band, and the heterozygous His/Asp genotype by the presence of three bands at 271, 227 and 44 bp. Standard laboratory procedure was followed in the study. The laboratory personnel were blinded on the disease status of study subjects. Negative control (without sample DNA in the reaction system) was applied in the PCR step to rule out potential contamination. PCR products were checked for validity according to their relative position to the markers (100 bp DNA ladder). Both positive (his1104 homozygote) and negative (Asp1104 homozygote) controls were applied in the digestion to ensure that the samples were fully digested by the enzyme and to rule out potential contamination. Independent repetition of genotyping in a randomly selected sample ($n = 120$) produced the same results. The *XPG* polymorphism was successfully genotyped for 497 lung cancer cases, 443 cases of oropharyngeal, laryngeal and esophageal cancers and 912 controls. The genotyping success rate was 93.2% (1,852 of 1,986 samples).

Statistical analysis

We used the goodness-of-fit χ^2 to test whether the *XPG* genotypes were in the Hardy-Weinberg equilibrium. To increase the precision of estimation and the power to test hypotheses, we used binary and polytomous unconditional logistic regression models including terms for the matching factors age and gender, which allowed us to include cases with no matched control.[12] Odds ratios (ORs) and 95% confidence limits (CLs) were calculated with and without adjustment for potential confounders, respectively. The potential confounders included age, gender (male/female), ethnic origin/ethnicity (Caucasian, Mexican-American, African-American, Asian-American, other), educational level, alcohol drinking and tobacco smoking. To minimize age confounding, age was controlled in fine categories (under 34, 35–36, 37–38, 39–40, 41–42, 43–44, 45–46, 47–48, 49–50, 51–52, 53–54, 55–56, 57–58, 59–62), and we excluded controls that were more than 3 years younger than the youngest case or 3 years older than the oldest case from the analysis. The age ranges were 32–59 for lung cancer cases and 20–59 for oropharyngeal, laryngeal and esophageal cancer cases, respectively. Overall 11 controls were excluded from the analysis of lung cancer and one control was excluded from the analysis of oropharyngeal, laryngeal and esophageal cancers. Educational level (years of schooling), tobacco smoking (pack-years) and alcohol consumption (drinks/day * years) were controlled as continuous variables.

We initially explored the main effect of the *XPG* polymorphism in two different ways by assuming either additive or dominant model. We primarily report the results on the *XPG* polymorphism with subjects having at least one allele of 1104His as the referent because our data supported that 1104His allele might dominate 1104Asp allele. Modification of tobacco smoking and alcohol effects by the polymorphism was also investigated. Heavy smoking was defined as smoking more than 20 pack-years, and heavy drinking was defined as consuming 3 drinks/day or more. All data analyses were conducted in SAS V8.2 (SAS Institute, Cary, NC). *p*-Values are two-sided.

Results

Among the lung cancer cases, there were 95 diagnosed as squamous cell carcinoma, 297 as adenocarcinoma, 115 as large-cell lung cancer, and 75 as small-cell lung cancer. Among oropharyngeal, laryngeal and esophageal cancer cases, 497 (300 oral cancers, 75 pharyngeal cancers, 90 laryngeal cancers and 32 esophageal cancers) were diagnosed as squamous cell carcinoma. The distributions of demographic characteristics and potential risk factors are summarized in Table I. The ORs for age and gender were not presented because the estimates from regression models are not valid for these two matching variables. Tobacco smoking was the

TABLE I – DISTRIBUTIONS OF DEMOGRAPHIC CHARACTERISTICS AND POTENTIAL RISK FACTORS

Key variable	Oropharyngeal, laryngeal and esophageal cancers			Lung cancer		
	Controls N (%)	Cases N (%)	Adjusted OR (95% CL)[1]	Controls N (%)	Cases N (%)	Adjusted OR (95% CL)[1]
Age[2,3]						
17–34	51 (4.9)	32 (5.3)	—	41 (4.0)	4 (0.6)	—
35–44	171 (16.5)	77 (12.8)	—	171 (16.6)	57 (9.3)	—
45–54	499 (48.0)	267 (44.4)	—	499 (48.5)	301 (49.3)	—
>54	318 (30.6)	225 (37.5)	—	318 (30.9)	249 (40.8)	—
Gender[2]						
Female	416 (40.0)	147 (24.5)	—	413 (40.1)	308 (50.4)	—
Male	623 (60.0)	454 (75.5)	—	616 (59.9)	303 (49.6)	—
Ethnic origin/ethnicity						
Caucasian	634 (61.1)	341 (56.9)	1.0	628 (61.1)	359 (58.9)	1.0
Mexican	149 (14.3)	70 (11.7)	0.53 (0.35, 0.81)	148 (14.4)	53 (8.7)	0.91 (0.57, 1.5)
African	102 (9.8)	69 (11.5)	1.1 (0.78, 1.6)	102 (9.9)	96 (15.7)	2.0 (1.4, 2.9)
Asian	62 (6.0)	64 (10.7)	2.7 (1.8, 4.1)	60 (5.8)	70 (11.5)	4.7 (3.1, 7.2)
Other	91 (8.8)	55 (9.2)	0.95 (0.63, 1.4)	90 (8.8)	32 (5.2)	0.78 (0.47, 1.3)
Missing	1	2		1	1	
Education (years of schooling)						
0–12	299 (28.8)	273 (45.4)	1.0	299 (29.1)	265 (43.4)	1.0
13–16	481 (46.3)	259 (43.1)	0.63 (0.48, 0.82)	471 (45.8)	275 (45.0)	0.79 (0.60, 1.0)
>16	258 (24.9)	69 (11.5)	0.36 (0.25, 0.52)	258 (25.1)	71 (11.6)	0.55 (0.37, 0.80)
Missing	1	0		1	0	
			$p_{trend} < 0.0001$			$p_{trend} = 0.0021$
Alcohol drinking (drinks/day)						
0–2	946 (91.4)	446 (74.5)	1.0	937 (91.3)	520 (85.3)	1.0
≥3	89 (8.6)	153 (25.5)	1.9 (1.4, 2.7)	89 (8.7)	90 (14.7)	0.89 (0.59, 1.3)
Missing	4	2		3	1	
Pack-years of tobacco smoking						
Never	491 (47.3)	182 (30.3)	1.0	484 (47.1)	110 (18.0)	1.0
1–20	353 (34.0)	147 (24.4)	0.97 (0.73, 1.3)	350 (34.1)	102 (16.7)	1.4 (1.0, 1.9)
21–40	136 (13.1)	146 (24.3)	2.0 (1.4, 2.8)	136 (13.2)	202 (33.1)	7.9 (5.6, 11)
>40	58 (5.6)	126 (21.0)	3.5 (2.3, 5.3)	58 (5.6)	197 (32.2)	22 (14, 34)
Missing	1	0		1	0	
			$p_{trend} < 0.0001$			$p_{trend} < 0.0001$

[1]Adjusted for the variables listed in Table I.–[2]Odds ratios (ORs) are not presented because age and gender are matching variables and their OR estimated by the regression models are not valid.–[3]Age was controlled in fine categories although it was presented here in 4 broad categories.

strongest risk factor for lung cancer with an extremely increased risk among subjects who smoked >40 pack-years (OR = 22, 95% CL = 14, 34). Both alcohol drinking (≥3 drinks/day *vs.* <3 drinks/day: OR = 1.9, 95% CL = 1.4, 2.7) and tobacco smoking (>40 pack-years *vs.* nonsmoking: OR = 3.5, 95% CL = 2.3, 5.3) were strongly associated with oropharyngeal, laryngeal and esophageal cancers.

Genotyping data on the *XPG* His1104Asp polymorphism were available for 497 lung cancer cases, 443 oropharyngeal, laryngeal and esophageal cancer cases (202 oral, 74 pharyngeal, 73 laryngeal and 94 esophageal cancer cases), and 912 controls. Overall, the allele frequencies were 72% for G allele (His) and 28% for C allele (Asp) in the control group, which were very similar to those reported by the Cancer Genome Anatomy Project.[8] The ethnic origin/ethnicity-specific frequencies for C allele were 22% for Caucasians, 34% for Mexican-Americans, 48% for African-Americans, and 49% for Asian-Americans. No departures from Hardy-Weinberg equilibrium were detected for the overall allelic frequency distribution ($p = 0.27$) and the ethnic origin/ethnicity-specific allelic frequency distributions ($p = 0.37$ for Caucasians; $p = 0.058$ for Mexican-Americans; $p = 0.35$ for African-Americans; $p = 0.41$ for Asian-Americans). Overall, the Asp1104Asp genotype was slightly under-represented in both the oropharyngeal, laryngeal and esophageal cancer group (7.9%) and the lung cancer group (8.3%), compared to the controls (8.8%).

With the adjustment for age, gender, ethnic origin/ethnicity, educational level and pack-years of tobacco smoking, the Asp1104Asp genotype was associated with a slight decrease in the risk of lung cancer (OR = 0.62, 95% CL = 0.38, 1.0) with the combined His1104His and His1104Asp genotypes as the referent (Table II). Stratified analyses displayed consistent negative associations for Caucasians (OR = 0.47, 95% CL = 0.18, 1.2), Mexican-Americans (OR = 0.41, 95% CL = 0.11, 1.6), African-Americans (OR = 0.55, 95% CL = 0.18, 1.7), and Asian-Americans (OR = 0.69, 95% CL = 0.23, 2.0). Polytomous logistic regression exhibited inverse associations of Asp1104Asp genotype with adenocarcinoma (OR = 0.69, 95% CL = 0.39, 1.2), small-cell lung cancer (OR = 0.55, 95% CL = 0.18, 1.7), and large-cell lung cancer (OR = 0.31, 95% CL = 0.10, 0.92), but not with squamous cell carcinoma (OR = 1.0, 95% CL = 0.45, 2.3).

With the adjustment for age, gender, ethnic origin/ethnicity, educational level, alcohol drinking and pack-years of tobacco smoking, the Asp1104Asp genotype was associated with a decreased risk of oropharyngeal, laryngeal and esophageal cancers with the combined His1104His and His1104Asp genotypes as the reference group (OR = 0.67, 95% CL = 0.42, 1.1) (Table III). Genotyping data was available for 355 SCCOLE cases (200 for oral cavity, 56 for pharynx, 73 for larynx and 26 for esophagus). The inverse association between the Asp1104Asp genotype and SCCOLE was stronger (OR = 0.47, 95% CL = 0.27, 0.82). Polytomous logistic regression exhibited inverse associations with squamous cell carcinoma of the oral cavity (OR = 0.48, 95% CL = 0.24, 0.96), pharynx (OR = 0.44, 95% CL = 0.14, 1.4), larynx (OR = 0.46, 95% CL = 0.16, 1.3), and esophagus (OR = 0.35, 95% CL = 0.07, 1.7). Stratified analyses exhibited a tendency toward negative associations for Caucasians (OR = 0.55, 95% CL = 0.20, 1.5), African-Americans (OR = 0.51, 95% CL = 0.15, 1.8) and Asian-Americans (OR = 0.06, 95% CL = 0.01, 0.44). Association for Mexican-Americans could not be assessed due to limited numbers.

Table IV summarizes the effect modification of tobacco smoking on lung cancer and the consumption of tobacco and alcohol on SCCOLE by the *XPG* polymorphism. With subjects having genotype Asp1104Asp and no tobacco smoking exposure as the com-

TABLE II – ASSOCIATION BETWEEN THE XPG POLYMORPHISM AND THE RISK OF LUNG CANCER

XPG	Cases N (%)	Controls N (%)	OR (95% CL)	
			Crude	Adjusted[1]
All lung cancer cases				
His/His	244 (49.1)	468 (51.9)	1.0	1.0
His/Asp	212 (42.7)	356 (39.5)	1.1 (0.91, 1.4)	1.1 (0.80, 1.4)
Asp/Asp	41 (8.2)	78 (8.6)	1.0 (0.67, 1.5)	0.65 (0.39, 1.1)
			$p_{trend} = 0.51$	$p_{trend} = 0.36$
His/His + His/Asp	456 (91.8)	824 (91.4)	1.0	1.0
Asp/Asp	41 (8.2)	78 (8.6)	0.95 (0.64, 1.4)	0.62 (0.38, 1.0)
Ethnic origin/ethnicity-specific				
Caucasian				
His/His + His/Asp	276 (95.5)	532 (95.7)	1.0	1.0
Asp/Asp	13 (4.5)	24 (4.3)	1.0 (0.52, 2.1)	0.47 (0.18, 1.2)
Mexican-American				
His/His + His/Asp	41 (91.1)	116 (85.3)	1.0	1.0
Asp/Asp	4 (8.9)	20 (14.7)	0.57 (0.18, 1.8)	0.41 (0.11, 1.6)
African-American				
His/His + His/Asp	69 (85.2)	60 (79.0)	1.0	1.0
Asp/Asp	12 (14.8)	16 (21.0)	0.65 (0.29, 1.5)	0.55 (0.18, 1.7)
Asian-American				
His/His + His/Asp	40 (78.4)	37 (74.0)	1.0	1.0
Asp/Asp	11 (21.6)	13 (26.0)	0.78 (0.31, 2.0)	0.69 (0.23, 2.0)
Histology-specific				
Adenocarcinoma				
His/His + His/Asp	221 (91.3)	824 (91.4)	1.0	1.0
Asp/Asp	21 (8.7)	78 (8.6)	1.0 (0.61, 1.7)	0.69 (0.39, 1.2)
Squamous cell carcinoma				
His/His + His/Asp	68 (86.1)	824 (91.4)	1.0	1.0
Asp/Asp	11 (13.9)	78 (8.6)	1.7 (0.87, 3.4)	1.0 (0.45, 2.3)
Small-cell lung cancer				
His/His + His/Asp	60 (93.7)	824 (91.4)	1.0	1.0
Asp/Asp	4 (6.3)	78 (8.6)	0.70 (0.25, 2.0)	0.55 (0.18, 1.7)
Large-cell lung cancer				
His/His + His/Asp	85 (95.5)	824 (91.4)	1.0	1.0
Asp/Asp	4 (4.5)	78 (8.6)	0.50 (0.18, 1.4)	0.31 (0.10, 0.92)

[1]Adjusted for age, gender, ethnic origin/ethnicity, educational level and pack-years of tobacco smoking.

mon referent, the odds ratios on lung cancer risk were 13 (95% CL = 4.4, 37) for heavy tobacco smoking (>20 pack-years), 1.9 (95% CL = 0.78, 4.5) for having at least one copy of 1104His, and 23 (95% CL = 9.5, 56) for exposure to both conditions, respectively. The upward departure from additivity (23 − 13 − 1.9 + 1 = 9.1, 95% CL = −2.9, 21.7) suggests that the polymorphism might act synergistically with tobacco smoking on lung carcinogenesis among heavy smokers. Compared to non-smokers with the Asp1104Asp genotype, the adjusted odds ratio on SCCOLE for heavy smokers (>20 pack-years) having at least one copy of 1104His was 8.0 (95% CL = 2.7, 24). Similarly, compared to non-drinkers with the Asp1104Asp genotype, the adjusted OR on SCCOLE for heavy drinkers (≥3 drinks/day) with at least one copy of 1104His was 10 (95% CL = 2.7, 38). Modification of tobacco and alcohol effects on SCCOLE by the polymorphism among heavy consumers could not be precisely assessed due to limited numbers (smoking: 8.0 − 3.9 − 3.6 + 1 = 1.5, 95% CL = −2.3, 5.4; alcohol: 10− 3.9 − 4.4 + 1 = 2.7, 95% CL = −3.4, 8.7). When compared to non- or light-consumers of tobacco (<= 20 pack-years) and alcohol (<3 drinks/day) with the Asp1104Asp genotype, subjects with heavy exposures to tobacco (>20 pack-years) and alcohol (≥3 drinks/day) and with at least one copy of 1104His were at an extremely increased risk of SCCOLE (OR = 18, 95% CL = 7.5, 45) (Table V).

Discussion

The *XPG* gene encodes a 133 kDa acidic protein with an open reading frame of 1186 amino acids.[13] It consists of 2 non-synonymous SNPs (His1104Asp, rs17655; Ser529Cys, rs2227869), one synonymous SNP (His46His, rs1047768) and 2 SNPs in untranslated region (rs2296148; rs873601), and more than 35 intron SNPs with their heterozygosity ≥9.5%. Our study focused on the non-synonymous SNP His1104Asp (rs17655), which has been validated by National Center for Biotechnology Information SNP database, Cancer Genome Anatomy SNP database, and Hap-Map project. Another non-synonymous SNP, Ser529Cys (rs2227869), was not studied because it is uncommon in the general population (reported heterozygosity <10%) and our study does not have enough power to detect the association, if there is any. Decreased expression of XPG in lymphocytes has been associated with an increased risk of lung cancer and squamous cell carcinoma of head and neck.[14,15] We found that the *XPG* Asp1104Asp genotype was associated with a decreased risk of developing lung cancer and SCCOLE. The results for lung cancer were consistent with that reported by a Korean Study, which found the Asp1104Asp genotype was associated with a 46% decreased risk of lung cancer compared to the combination of His1104His and His1104Asp genotypes (OR = 0.54, 95%CL = 0.37, 0.80).[11] The His1104Asp polymorphism locates in the XPG C-terminus, which is required for the interactions between XPG and TFIIH subunits in the NER pathway.[9] The amino acid change from acidic Aspartic Acid to basic Histidine may change the XPG protein structure and therefore influence the protein-protein interactions and the stability of the pre-incision complex. The potential harmful effect incurred by 1104His is also supported by computational function analyses, as the SIFT (sorting intolerant from tolerant) program predicts the amino acid change from aspartic acid to histidine "intolerant" and PolyPhen predicts the amino acid change "possibly damaging".[16]

We found a negative association of the *XPG* Asp1104Asp genotype with lung adenocarcinoma, small-cell lung cancer, and large-cell lung cancer, suggestive of protective effects, but not with lung squamous cell carcinoma. This histological difference might be

TABLE III – ASSOCIATION BETWEEN THE *XPG* POLYMORPHISM AND THE RISK OF OROPHARYNGEAL, LARYNGEAL AND ESOPHAGEAL CANCERS

XPG	Cases N (%)	Controls N (%)	OR (95% CL)	
			Crude	Adjusted[1]
All oropharyngeal, laryngeal and esophageal cancer cases				
His/His	214 (48.3)	474 (52.0)	1.0	1.0
His/Asp	194 (43.8)	357 (39.2)	1.2 (0.95, 1.5)	1.1 (0.85, 1.5)
Asp/Asp	35 (7.9)	80 (8.8)	0.97 (0.63, 1.5)	0.71 (0.43, 1.2)
			$p_{trend} = 0.45$	$p_{trend} = 0.61$
His/His + His/Asp	408 (92.1)	831 (91.2)	1.0	1.0
Asp/Asp	35 (7.9)	80 (8.8)	0.89 (0.59, 1.3)	0.67 (0.42, 1.1)
Squamous cell carcinomas of the oropharynx, larynx and esophagus (SCCOLE)				
His/His + His/Asp	333 (93.8)	831 (91.2)	1.0	1.0
Asp/Asp	22 (6.2)	80 (8.8)	0.69 (0.42, 1.1)	0.47 (0.27, 0.82)
Ethnic origin/ethnicity-specific (SCCOLE)				
Caucasian				
His/His + His/Asp	200 (97.1)	537 (95.7)	1.0	1.0
Asp/Asp	6 (2.9)	24 (4.3)	0.67 (0.27, 1.7)	0.55 (0.20, 1.5)
Mexican-American[2]				
His/His + His/Asp	47 (100.0)	116 (84.7)	1.0	1.0
Asp/Asp	0 (0.0)	21 (15.3)	—	—
African-American				
His/His + His/Asp	35 (81.4)	60 (79.0)	1.0	1.0
Asp/Asp	8 (18.6)	16 (21.0)	0.86 (0.33, 2.2)	0.51 (0.15, 1.8)
Asian-American				
His/His + His/Asp	27 (84.4)	38 (73.1)	1.0	1.0
Asp/Asp	5 (15.6)	14 (26.9)	0.50 (0.16, 1.6)	0.06 (0.01, 0.44)
Site-specific (SCCOLE)				
Oral cavity				
His/His + His/Asp	189 (94.5)	831 (91.2)	1.0	1.0
Asp/Asp	11 (5.5)	80 (8.8)	0.61 (0.32, 1.2)	0.48 (0.24, 0.96)
Pharynx				
His/His + His/Asp	52 (92.9)	831 (91.2)	1.0	1.0
Asp/Asp	4 (7.1)	80 (8.8)	0.80 (0.28, 2.3)	0.44 (0.14, 1.4)
Larynx				
His/His + His/Asp	68 (93.1)	831 (91.2)	1.0	1.0
Asp/Asp	5 (6.9)	80 (8.8)	0.76 (0.30, 1.9)	0.46 (0.16, 1.3)
Esophagus				
His/His + His/Asp	24 (92.3)	831 (91.2)	1.0	1.0
Asp/Asp	2 (7.7)	80 (8.8)	0.87 (0.20, 3.7)	0.35 (0.07, 1.7)

[1]Adjusted for age, gender, ethnic origin/ethnicity, educational level, pack-years of tobacco smoking and alcohol drinking.–[2]The association could not be assessed because no cases had Asp1104Asp genotype.

TABLE IV – EFFECT MODIFICATION OF TOBACCO SMOKING AND ALCOHOL DRINKING BY THE *XPG* POLYMORPHISM

Variable 1	Variable 2	Cases N (%)	Controls N (%)	Adjusted OR (95% CL)
Lung cancer[1]				
Tobacco	*XPG*			
Never	Asp/Asp	8 (1.6)	42 (4.7)	1.0
1-20 Pack-years	Asp/Asp	8 (1.6)	19 (2.1)	2.4 (0.70, 8.1)
>20 Pack-years	Asp/Asp	25 (5.0)	17 (1.9)	13 (4.4, 37)
Never	His/His + His/Asp	79 (15.9)	386 (42.8)	1.9 (0.78, 4.5)
1-20 Pack-years	His/His + His/Asp	71 (14.3)	288 (31.9)	2.2 (0.91, 5.3)
>20 Pack-years	His/His + His/Asp	306 (61.6)	150 (16.6)	23 (9.5, 56)
SCCOLE[2]				
Tobacco	*XPG*			
Never	Asp/Asp	4 (1.1)	43 (4.7)	1.0
1-20 Pack-years	Asp/Asp	5 (1.4)	20 (2.2)	2.2 (0.51, 9.6)
>20 Pack-years	Asp/Asp	13 (3.7)	17 (1.9)	3.8 (1.0, 14)
Never	His/His + His/Asp	100 (28.2)	391 (42.9)	3.6 (1.2, 11)
1-20 Pack-years	His/His + His/Asp	80 (22.5)	290 (31.8)	3.2 (1.1, 9.5)
>20 Pack-years	His/His + His/Asp	153 (43.1)	150 (16.5)	8.0 (2.7, 24)
SCCOLE[3]				
Alcohol	*XPG*			
Never	Asp/Asp	3 (0.9)	31 (3.4)	1.0
1-2 Drinks/day	Asp/Asp	12 (3.4)	39 (4.3)	2.9 (0.70, 12)
≥3 Drinks/day	Asp/Asp	7 (2.0)	10 (1.1)	3.9 (0.76, 21)
Never	His/His + His/Asp	55 (15.5)	192 (21.1)	4.4 (1.2, 16)
1-2 Drinks/day	His/His + His/Asp	184 (52.0)	568 (62.5)	4.4 (1.2, 16)
≥3 Drinks/day	His/His + His/Asp	93 (26.2)	69 (7.6)	10 (2.7, 38)

[1]Adjusted for age, gender, ethnic origin/ethnicity and educational level.–[2]Adjusted for age, gender, ethnic origin/ethnicity, educational level and alcohol drinking.–[3]Adjusted for age, gender, ethnic origin/ethnicity, educational level and tobacco smoking.

TABLE V – JOINT EFFECT OF TOBACCO SMOKING, ALCOHOL DRINKING AND *XPG* POLYMORPHISM ON THE RISK OF SCCOLE

Smoking (pack-years)	Alcohol (drinks/day)	*XPG*	Cases N (%)	Controls N (%)	Adjusted OR (95% CL)[1]
<=20	<3	Asp/Asp	8 (2.2)	60 (6.6)	1.0
<=20	<3	His/His + His/Asp	158 (44.6)	641 (70.5)	2.6 (1.2, 5.8)
<=20	≥3	Asp/Asp	1 (0.30)	3 (0.3)	3.4 (0.29, 40)
<=20	≥3	His/His + His/Asp	23 (6.5)	39 (4.3)	4.6 (1.8, 12)
>20	<3	Asp/Asp	7 (2.0)	10 (1.1)	4.2 (1.2, 15)
>20	<3	His/His + His/Asp	81 (22.9)	119 (13.1)	5.7 (2.5, 13)
>20	≥3	Asp/Asp	6 (1.7)	7 (0.80)	5.3 (1.3, 21)
>20	≥3	His/His + His/Asp	70 (19.8)	30 (3.3)	18 (7.5, 45)

[1]Adjusted for age, gencer, ethnic origin/ethnicity and educational level.

due to chance or might be attributable to the etiologic difference among these four subtypes of lung cancer. Experimental data has suggested that tobacco polycyclic aromatic hydrocarbons (PAH) predominantly induce lung squamous cell carcinoma; whereas tobacco-specific nitrosamines predominantly induce lung adenocarcinoma.[17,18] It has been proposed that the decreasing incidence of lung squamous cell carcinoma over time might have been due to the reduced exposure of the central bronchi to tobacco PAH, and that the marked rise in lung adenocarcinoma might have partially resulted from an increased exposure of peripheral lung to tobacco-specific nitrosamines.[19] Unfortunately, tobacco-specific chemicals involved in the large-cell and small-cell lung cancers have not been very well documented. Nevertheless, it is possible that these four subtypes of lung cancer may be caused by different initiating or promoting agents found in tobacco and other environmental factors. The *XPG* polymorphism may have differential repair efficiency on different chemical-induced DNA adducts and consequently have differential protective effects against different histological types of lung cancer.

Interestingly, we found the *XPG* Asp1104Asp genotype negatively associated with SCCOLE, although we detected no association between the genotype and lung squamous cell carcinoma. This seeming contradiction might be due to chance or might be due to the fact that lung squamous cell carcinoma is different from SCCOLE etiologically and that the *XPG* polymorphism has differential modifying effects on different chemical-induced carcinogenesis. Alcohol drinking is a well-established risk factor for SCCOLE, but not a risk factor for lung cancer.[20] There are at least two lines of evidence supporting the possibility that the NER pathway is involved in alcohol-related carcinogenesis. First, alcohol consumption can generate lipid peroxidation products such as malondialdehyde (MDA) that is highly mutagenic and carcinogenic.[21] MDA can interact with DNA to form MDA-DNA adducts, which can be repaired via the NER pathway.[22] Second, the NER pathway can remove certain types of small oxidative DNA lesions induced by chronic alcohol consumption.[21,23] The *XPG* polymorphism may influence the capacity of the NER pathway to repair alcohol-induced DNA lesions. Furthermore, recent studies have showed that XPG contributes to base excision repair as it can stimulate binding of HsNth1 to damaged DNA and increases glycosylase/AP lyase activity.[24,25] Therefore, the *XPG* polymorphism may influence the efficiency of the base excision repair pathway to repair the damage due to reactive oxygen species generated by alcohol consumption. In summary, the different associations of the *XPG* polymorphism with lung squamous cell carcinoma and SCCOLE might be attributable to the etiologic difference between these two diseases and the differential repair capacity of the *XPG* polymorphism on alcohol and tobacco-induced carcinogenesis.

Our study is however subject to several limitations. Non-participation of certain cases and controls and unwillingness of some recruited cases and controls to provide buccal cells might lead to potential selection bias. Only 42% (1,212 of 2,857) of eligible cancer patients identified by the tumor registry system were recruited into the study. By contrast, a much higher fraction (79%) of contacted controls participated in the study. A large proportion but not all of this difference was due to death of cases before interview. Eighty-eight percent (88%) of the participants donated buccal cells. In order for bias to occur, participation among eligible cases would have to be associated with exposures such as *XPG* genotypes, and the associations would have to differ between potential cases and controls. Both conditions occurring to the extent needed to produce appreciable bias seem unlikely because genotypes are neither known to study subjects nor associated with uncontrolled factors that would differentially affect participation. In addition, we found similar distributions on age, gender, alcohol consumption and tobacco smoking between the subjects who donated buccal cells and those who did not. Ten percent of eligible oropharyngeal, laryngeal and esophageal cancer cases and 25% of eligible lung cancer cases (20% for adenocarcinoma, 26% for squamous cell carcinoma, 24% for small-cell lung cancer and 29% for large-cell lung cancer) died before they could be recruited. Selective-survival bias could have occurred if the *XPG* polymorphism were a prognostic factor for the above-mentioned cancers. Unfortunately, we could not assess the direction and magnitude of the potential selective-survival bias due to the lack of studies investigating the effect of the polymorphism on the prognosis of the above-mentioned cancers. Exposure misclassification might exist due to the retrospective nature of the measurements. Recall bias may exist because cases and controls may differentially recall certain exposures such as smoking history. Using the USC Cancer Surveillance Program as the sources of cases, empirical evidence from a previous case-control study on lung cancer suggested the effect of tobacco use is estimated with little or no bias.[26] Furthermore, there should be only minimal bias in misclassification of *XPG* genotypes because they were assayed by standard laboratory protocols and all the laboratory researchers were blinded on the disease status of study subjects. Last, residual confounding might exist when evaluating the main effects and interactions, although to ensure against this we adjusted the well-documented risk factors.

Strengths of our study include a relatively large sample size, a population-based study design, and extensive relevant questionnaire data. In conclusion, we found that the *XPG* Asp1104Asp genotype was negatively associated with lung cancer and SCCOLE, suggesting protective effects on these cancers. We also found some suggestion that the genotype synergistically modifies the effect of tobacco use on lung cancer risk.

Acknowledgments

This work is supported by the Alper Research Program for Environmental Genomics of the UCLA Jonsson Comprehensive Cancer Center. We thank Mr. J. Dermand for supervising the administration of questionnaires and the collection of buccal smears and Dr. M. Hashibe, Ms. Y.A. Lee, Ms. N.-C. Yuko You, and Ms. T. McAfee for extracting DNA from buccal cells. We are also indebted to the study participants for their dedication and commitment.

References

1. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. Science 2001;291:1284–9.
2. van Hoffen A, Balajee AS, van Zeeland AA, Mullenders LH. Nucleotide excision repair and its interplay with transcription. Toxicology 2003;193:79–90.
3. Berwick M, Vineis P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. J Natl Cancer Inst 2000; 92:874–97.
4. Mohrenweiser HW, Jones IM. Variation in DNA repair is a factor in cancer susceptibility: a paradigm for the promises and perils of individual and population risk estimation? Mutat Res 1998;400:15–24.
5. O'Donovan A, Davies AA, Moggs JG, West SC, Wood RD. XPG endonuclease makes the 3′ incision in human DNA nucleotide excision repair. Nature 1994;371:432–5.
6. Wakasugi M, Reardon JT, Sancar A. The non-catalytic function of XPG protein during dual incision in human nucleotide excision repair. J Biol Chem 1997;272:16030–4.
7. Constantinou A, Gunz D, Evans E, Lalle P, Bates PA, Wood RD, Clarkson SG. Conserved residues of human XPG protein important for nuclease activity and function in nucleotide excision repair. J Biol Chem 1999;274:5637–48.
8. Cancer Genome Anatomy Project SNP500Cancer Database. National Cancer Institute 2004 (http://snp500cancer.nci.nih.gov/snp.cfm?both_snp_id=ERCC5-02).
9. Wakasugi M, Sancar A. Order of assembly of human DNA repair excision nuclease. J Biol Chem 1999;274:18759–68.
10. Cao W, Hashibe M, Rao JY, Morgenstern H, Zhang ZF. Comparison of methods for DNA extraction from paraffin-embedded tissues and buccal cells. Cancer Detect Prev 2003;27:397–404.
11. Jeon HS, Kim KM, Park SH, Lee SY, Choi JE, Lee GY, Kam S, Park RW, Kim IS, Kim CH, Jung TH, Park JY. Relationship between XPG codon 1104 polymorphism and risk of primary lung cancer. Carcinogenesis 2003;24:1677–81.
12. Greenland S. Introduction to Regression Models. In: Rothman KJ, Greenland S, eds. Modern epidemiology. Philadelphia: Lippincott Williams & Wilkins, 1998. p 395–6.
13. Clarkson SG. The XPG story. Biochimie 2003;85:1113–21.
14. Cheng L, Spitz MR, Hong WK, Wei Q. Reduced expression levels of nucleotide excision repair genes in lung cancer: a case-control analysis. Carcinogenesis 2000;21:1527–30.
15. Cheng L, Sturgis EM, Eicher SA, Spitz MR, Wei Q. Expression of nucleotide excision repair genes and the risk for squamous cell carcinoma of the head and neck. Cancer 2002;94:393–7.
16. NIEHS SNPS SIFT/PolyPhen Data. 2004 (http://egp.gs.washington.edu/pph-sift-stuff/pphSiftForEgpNonsynSnps.out).
17. Deutsch-Wenzel RP, Brune H, Grimmer G, Dettbarn G, Misfeld J. Experimental studies in rat lungs on the carcinogenicity and dose-response relationships of eight frequently occurring environmental polycyclic aromatic hydrocarbons. J Natl Cancer Inst 1983;71:539–44.
18. Hoffmann D, Djordjevic MV, Rivenson A, Zang E, Desai D, Amin S. A study of tobacco carcinogenesis. LI. Relative potencies of tobacco-specific N-nitrosamines as inducers of lung tumours in A/J mice. Cancer Lett 1993;71:25–30.
19. Le Marchand L, Sivaraman L, Pierce L, Seifried A, Lum A, Wilkens LR, Lau AF. Associations of CYP1A1, GSTM1, and CYP2E1 polymorphisms with lung cancer suggest cell type specificities to tobacco carcinogens. Cancer Res 1998;58:4858–63.
20. Poschl G, Seitz HK. Alcohol and cancer. Alcohol Alcohol 2004;39: 155–65.
21. Brooks PJ. DNA damage, DNA repair, and alcohol toxicity—a review. Alcohol Clin Exp Res 1997;21:1073–82.
22. Marnett LJ. Oxy radicals, lipid peroxidation and DNA damage. Toxicology 2002;181–2:219–22.
23. Moller P, Wallin H. Adduct formation, mutagenesis and nucleotide excision repair of DNA damage produced by reactive oxygen species and lipid peroxidation product. Mutat Res 1998;410:271–90.
24. Klungland A, Hoss M, Gunz D, Constantinou A, Clarkson SG, Doetsch PW, Bolton PH, Wood RD, Lindahl T. Base excision repair of oxidative DNA damage activated by XPG protein. Mol Cell 1999; 3:33–42.
25. Bessho T. Nucleotide excision repair 3′ endonuclease XPG stimulates the activity of base excision repair enzyme thymine glycol DNA glycosylase. Nucleic Acids Res 1999;27:979–83.
26. Wu AH, Fontham ET, Reynolds P, Greenberg RS, Buffler P, Liff J, Boyd P, Correa P. Family history of cancer and risk of lung cancer among lifetime nonsmoking women in the United States. Am J Epidemiol 1996;143:535–42.