

Genetic Variation in the Nucleotide Excision Repair Pathway and Bladder Cancer Risk

Montserrat García-Closas,¹ Núria Malats,² Francisco X. Real,^{2,3} Robert Welch,⁴ Manolis Kogevinas,² Nilanjan Chatterjee,¹ Ruth Pfeiffer,¹ Debra Silverman,¹ Mustafa Dosemeci,¹ Adonina Tardón,⁵ Consol Serra,⁶ Alfredo Carrato,⁷ Reina García-Closas,⁸ Gemma Castaño-Vinyals,² Stephen Chanock,^{1,4} Meredith Yeager,⁴ and Nathaniel Rothman¹

¹Division of Cancer Epidemiology and Genetics; ²Institut Municipal d'Investigació Mèdica and ³Universitat Pompeu Fabra, Barcelona, Spain; ⁴Core Genotype Facility at the Advanced Technology Center of the National Cancer Institute, Department of Health and Human Services, Bethesda, Maryland; ⁵Universidad de Oviedo, Oviedo, Spain; ⁶Consorci Hospitalari Parc Taulí, Sabadell, Spain; ⁷Hospital Universitario de Elche, Elche, Spain; and ⁸Unidad de Investigación, Hospital Universitario de Canarias, La Laguna, Spain

Abstract

Nucleotide excision repair (NER) is critical for protecting against damage from carcinogens in tobacco smoke. We evaluated the influence of common genetic variation in the NER pathway on bladder cancer risk by analyzing 22 single nucleotide polymorphisms (SNP) in seven NER genes (*XPC*, *RAD23B*, *ERCC1*, *ERCC2*, *ERCC4*, *ERCC5*, and *ERCC6*). Our study population included 1,150 patients with transitional cell carcinoma of the urinary bladder and 1,149 control subjects from Spain. Odds ratios (OR) and 95% confidence intervals (95% CI) were adjusted for age, gender, region, and smoking status. Subjects with the variant genotypes for SNPs in four of the seven genes evaluated had small increases in bladder cancer risk compared to subjects with the homozygous wild-type genotypes: *RAD23B* IVS5-15A>G (OR, 1.3; 95% CI, 1.1-1.5; $P = 0.01$), *ERCC2* R156R (OR, 1.3; 95% CI, 1.1-1.6; $P = 0.006$), *ERCC1* IVS5+33A>C (OR, 1.2; 95%

CI, 1.0-1.5; $P = 0.06$; $P_{\text{trend}} = 0.04$), and *ERCC5* M254V (OR, 1.4; 95% CI, 1.0-2.0; $P = 0.04$). A global test for pathway effects indicated that genetic variation in NER characterized by the 22 SNPs analyzed in this study significantly predicts bladder cancer risk ($P = 0.04$). Pairwise comparisons suggested that carrying variants in two genes could result in substantial increases in risk. Classification tree analyses suggested the presence of subgroups of individuals defined by smoking and NER genotypes that could have substantial increases in risk. In conclusion, these findings provide support for the influence of genetic variation in NER on bladder cancer risk. A detailed characterization of genetic variation in key NER genes is warranted and might ultimately help identify multiple susceptibility variants that could be responsible for substantial joint increases in risk. (Cancer Epidemiol Biomarkers Prev 2006;15(3):536-42)

Introduction

The importance of nucleotide excision repair (NER) in protection against cancer has been evident because of the discovery of cancer-prone syndromes, such as xeroderma pigmentosum, which result from rare germ line mutations in NER genes (1). NER is a complex and versatile mechanism that consists of the following critical steps (2): (a) damage recognition that differs depending on whether the damage is in transcriptionally silent (global genome repair involving XPC-RAD23B complex, XPA, and RPA) or transcriptionally active (transcription coupled repair involving a large protein complex, including CSA and ERCC6 proteins); (b) local unwinding of the DNA helix around the lesion by the transcription factor IIH complex that contains two DNA helicases (XPB and ERCC2); (c) dual incision of oligonucleotide containing the damage by 5' (ERCC1-ERCC4 complex) and 3' (ERCC5) endonucleases; and (d) repair of the nucleotide gap by DNA synthesis using the opposite normal DNA strand as a template which requires DNA polymerases (δ or ϵ) and the

accessory replication proteins: proliferating cell nuclear antigen, RPA, and RFC. This mechanism can repair a wide range of DNA lesions, including bulky DNA adducts caused by aromatic amines and other carcinogens in tobacco smoke (2). This suggests that common genetic variation in NER might influence the risk of smoking-related cancers, such as bladder cancer (3).

Functional studies in humans have shown that common variation in NER genes can affect the capacity to repair DNA (4-6), and epidemiologic studies have provided some evidence supporting their role in the pathogenesis of smoking-related cancers (3, 7). A few epidemiologic studies, including a range of 124 to 547 cases per study, have evaluated associations with bladder cancer risk (8-13). The gene that has been most studied is *ERCC2* (excision repair cross-complementary group 2), previously named *XPD*, which codes for a DNA helicase subunit of the core transcription factor IIH essential for NER and transcription (2). Specifically, five case-control studies of bladder cancer evaluated a nonsynonymous variant (*K751Q*) in *ERCC2* and found no significant associations with bladder cancer risk (8, 10-13). Other single nucleotide polymorphisms (SNP) that have been evaluated in relation to bladder cancer risk include *ERCC2* D312N (10); *XPC* K939Q, *PAT*, and *IVS11-6* (8, 9); and *ERCC5*, previously named *XPG* D1104H (8). The only statistically significant findings were from a Swedish study of 327 cases and same number of controls that found an increased risk for *ERCC5* K939Q homozygous variants and reduced risk for *ERCC5* D1104H homozygous variants (8).

Common variation in individual genes in a complex pathway involving multiple genes, such as NER, is unlikely to have strong associations with cancer risk. Previous studies of NER and bladder cancer had limited statistical power to

Received 9/23/05; revised 12/7/05; accepted 1/5/06.

Support: Intramural Research Program of the NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics; and FIS/Spain grants 00/0745, G03/174, G03/160, C03/09, and C03/10.

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

Note: Supplementary data for this article are available at Cancer Epidemiology Biomarkers and Prevention Online (<http://cebp.aacrjournals.org/>).

Requests for reprints: Montserrat García-Closas, Hormonal and Reproductive Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, Room 7076, 6120 Executive Boulevard, MSC 7234, Rockville, MD 20852-7234. Phone: 301-435-3981; Fax: 301-402-0916; E-mail: montse@nih.gov

Copyright © 2006 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-05-0749

evaluate small to modest associations; thus, studies of larger sample sizes are required to further evaluate this critical pathway. We evaluated the influence of genetic variation in the NER pathway on bladder cancer risk among 1,150 cases and 1,149 controls participating in the Spanish Bladder Cancer Study. Specifically, we analyzed 22 genetic variants in seven NER genes [*XPC*, *RAD23B*, *ERCC6* (previously named *CSB*), *ERCC2*, *ERCC5*, *ERCC1*, and *ERCC4* (previously named *XPF*)].

Materials and Methods

Study Population. The study population has been previously described (14). Briefly, cases were patients participating in the Spanish Bladder Cancer Study diagnosed with histologically confirmed carcinoma of the urinary bladder in 1998 to 2001, ages 21 to 80 years (mean \pm SD = 66 \pm 10 years), of which 87% were males. Controls were selected from patients admitted to participating hospitals for diagnoses believed to be unrelated to the exposures of interest, individually matched to the cases on age at interview within 5-year categories, gender, ethnicity, and region. Demographic and risk factor information was collected at the hospitals using computer-assisted personal interviews. Dietary data were collected with a food frequency questionnaire and nutrient composition of foods, including folic acid, was obtained from a Spanish food composition table (15).

Eighty-four percent of eligible cases and 88% of eligible controls agreed to participate in the study and were interviewed. Of the 1,219 cases and 1,271 controls interviewed, 1,188 (97%) cases and 1,173 (92%) controls provided a blood or buccal cell sample for DNA extraction. Seven cases and 11 controls were excluded because of low amounts of DNA. To reduce heterogeneity, 16 cases with neoplasias of nontransitional histology and six non-Caucasian subjects (5 cases and 1 control) were excluded from the analyses. Fifteen subjects (7 cases and 8 controls) with missing smoking status information and seven subjects (3 cases and 4 controls) with DNA quality control problems were also excluded from the analyses.

Thus, the final study population available for analysis included 1,150 cases and 1,149 controls. We obtained informed consent from potential participants in accordance with the National Cancer Institute and local institutional review boards.

Subjects were categorized as never smokers (29% of controls) if they smoked <100 cigarettes in their lifetime, and ever smokers otherwise. Ever smokers were further classified as regular smokers (63% of controls) if they smoked one cigarette per day for 6 months or longer, and occasional smokers (8% of controls) otherwise. Of the regular smoker controls, 37% were current smokers (i.e., they smoked within a year of the reference date), and 63% were former smokers. Most (81%) regular smoker controls with information on whether they smoked black or blond tobacco (information available in 82% of controls) reported smoking black tobacco (48% smoked black tobacco only and 33% both tobacco types).

Genotyping. DNA for genotype assays was extracted from leukocytes using the Puregene DNA Isolation kit (Gentra Systems, Minneapolis, MN) for most cases ($n = 1,107$) and controls ($n = 1,032$) included in the analysis. DNA from an additional 43 cases and 117 controls was extracted from mouthwash samples using a phenol-chloroform extraction.

We selected 22 SNPs in seven NER genes (Table 1) with an expected rare allele frequency in Caucasians of >5% and assays available at Core Genotyping Facility of the Division of Cancer Epidemiology and Genetics, National Cancer Institute at the time of analysis. Selection favored nonsynonymous SNPs, those previously evaluated in relation to bladder cancer risk, or those with evidence for functional significance. Genotype assays were done at the Core Genotyping Facility using randomly sorted DNA samples from cases and controls, including duplicate samples for quality control. Description and methods for each genotype assay can be found at <http://snp500cancer.nci.nih.gov> (16).

All genotypes under study were in Hardy-Weinberg equilibrium in the control population, except for a slightly higher than expected frequency of homozygous variants for *RAD23B* *IVS5-57A>T* (3% versus 2%, $P = 0.01$), *ERCC4* *R415Q*

Table 1. Genes and SNPs in NER evaluated in the Spanish Bladder Cancer Study

Gene symbol	Gene name	Chromosomal location	Nucleotide change	Amino acid change	dbSNP ID	Minor allele frequency in the control population
<i>XPC</i>	Xeroderma pigmentosum, complementation group C	3p25	<i>Ex9-398A>G</i>	<i>R493R</i>	rs2227999	0.07
			<i>Ex9-377C>T</i>	<i>A500V</i>	rs2228000	0.26
			<i>Ex16+211A>C</i>	<i>K939Q</i>	rs2228001	0.40
<i>RAD23B</i>	<i>RAD23</i> homologue B	9q31.2	<i>IVS5-66A>G</i>		rs1805332	0.26
			<i>IVS5-57A>T</i>		rs1805331	0.16
			<i>IVS5-15A>G</i>		rs1805335	0.42
			<i>Ex7+65C>T</i>	<i>A249V</i>	rs1805329	0.15
			<i>Ex18-90G>C</i>	<i>R1230P</i>	rs4253211	0.08
<i>ERCC6</i>	Excision repair cross-complementary group 6, formerly known as CBS	10q11.23	<i>Ex18-90G>C</i>			
<i>ERCC2</i>	Excision repair cross-complementary group 2, formerly known as <i>XPD</i>	19q13.3	<i>Ex6-10A>C</i>	<i>R156R</i>	rs238406	0.47
<i>ERCC1</i>	Excision repair cross-complementary group 1	19q13.2-q13.3	<i>Ex10-16G>A</i>	<i>D312N</i>	rs1799793	0.31
			<i>IVS19-70C>T</i>		rs1799787	0.27
			<i>Ex23+61A>C</i>	<i>K751Q</i>	rs13181	0.33
			<i>IVS5+33A>C</i>		rs3212961	0.12
<i>ERCC4</i>	Excision repair cross-complementary group 4, formerly known as <i>XPF</i>	16p13.3-p13.11	<i>196bp 3' of STP G>T</i>	<i>Q504K</i>	rs3212986	0.25
			<i>Ex8+31G>A</i>	<i>R415Q</i>	rs1800067	0.11
			<i>IVS9-35C>T</i>		rs1799799	0.36
<i>ERCC5</i>	Excision repair cross-complementary group 5, formerly known as <i>XPG</i>	13q22;13q33	<i>Ex10-34T>C</i>	<i>S662P</i>	rs2020955	0.01
			<i>Ex11-247T>C</i>	<i>S835S</i>	rs1799801	0.31
			<i>Ex2+50T>C</i>	<i>H46H</i>	rs1047768	0.43
			<i>Ex7+88A>G</i>	<i>M254V</i>	rs1047769	0.03
			<i>Ex8-369G>C</i>	<i>C529S</i>	rs2227869	0.04
			<i>Ex15-344G>C</i>	<i>D1104H</i>	rs17655	0.27

Table 2. Association between selected polymorphisms in NER genes and bladder cancer risk, adjusted for gender, age, region, and smoking status (1,150 cases and 1,149 controls; see Supplementary Table S1 for a full report of all SNPs evaluated)

Gene	SNP	Genotype	Cases, n (%)	Controls, n (%)	OR (95% CI)	P
XPC	K939Q	AA	374 (33)	411 (36)	1.0	
		AC	575 (51)	536 (47)	1.2 (1.0-1.4)	0.07
		CC	188 (17)	191 (17)	1.1 (0.8-1.4)	0.61
		AC or CC			1.2 (1.0-1.4)	$P_{\text{trend}} = 0.35$
RAD23B	IVS5-15A>G	GG	337 (30)	390 (35)	1.0	
		AG	571 (51)	524 (47)	1.3 (1.1-1.6)	0.01
		AA	216 (19)	205 (18)	1.2 (0.9-1.6)	0.12
				AG or AA		
ERCC2	R156R	AA	268 (24)	327 (29)	1.0	
		AC	578 (51)	528 (47)	1.3 (1.1-1.6)	0.01
		CC	287 (25)	270 (24)	1.3 (1.0-1.7)	0.04
		AC or CC			1.3 (1.1-1.6)	$P_{\text{trend}} = 0.03$
ERCC2	D312N	GG	517 (46)	538 (48)	1.0	
		AG	474 (42)	467 (42)	1.1 (0.9-1.3)	0.55
		AA	138 (12)	117 (10)	1.2 (0.9-1.6)	0.16
				AG or AA		
ERCC2	K751Q	AA	490 (43)	512 (46)	1.0	
		AC	500 (44)	487 (43)	1.1 (0.9-1.3)	0.45
		CC	146 (13)	126 (11)	1.2 (0.9-1.6)	0.22
				AC or CC		
ERCC1	IVS5+33A>C	CC	851 (75)	885 (78)	1.0	
		AC	255 (23)	225 (20)	1.2 (1.0-1.5)	0.11
		AA	24 (2)	18 (2)	1.6 (0.8-3.0)	0.19
				AC or AA		
ERCC5	H46H	CC	385 (35)	366 (33)	1.0	
		CT	530 (48)	506 (46)	1.0 (0.8-1.2)	0.99
		TT	188 (17)	222 (20)	0.8 (0.6-1.0)	0.06
				CT or TT		
ERCC5	M254V	AA	979 (91)	995 (93)	1.0	
		AG	93 (9)	72 (7)	1.4 (1.0-1.9)	0.08
		GG	5 (0)	1 (0)	—	
				AG or GG		
ERCC5	D1104H	GG	629 (55)	607 (53)	1.0	
		CG	434 (38)	445 (39)	0.9 (0.8-1.1)	0.55
		CC	78 (7)	84 (7)	0.9 (0.6-1.3)	0.54
				CG or CC		

(2% versus 1%, $P = 0.01$), and *ERCC4 IVS9-35C>T* (15% versus 13%, $P = 0.01$). Duplicate quality control DNA samples ($n = 93$ pairs) showed $\geq 98\%$ agreement for all assays except for a 96% agreement for *ERCC2 D312N*. *NAT2* and *GSTM1* genotypes were determined as previously described (14).

Statistical Analysis. For each individual polymorphism, we estimated odds ratios (OR) and 95% confidence intervals (95% CI) using logistic regression models adjusting for gender, age at interview in 5-year categories, region, and smoking status (never, occasional, former, and current). These unconditional models provided estimates similar to conditional logistic regression models for individually matched pairs (data not shown). A global test for the association between genetic variation in NER pathway as a whole was performed based on the maximum of trend statistics of all the individual polymorphisms. The P s for the global test was computed by the permutation method (17).

Gene-gene and gene-smoking interactions were assessed using pairwise comparisons in logistic regression models, as well as classification trees (CART) implemented in the S-Plus “tree” function. CART is an exploratory technique that uses splitting rules to stratify data into groups with homogenous risk (18). Its advantage over logistic regression is the ability to

identify subgroups of individuals defined by environmental and/or genetic characteristics that are at high risk, suggesting the presence of gene-gene or gene-environment interactions. Indicator variables for smoking status (ever versus never) and genotypes (homozygous wild-type versus heterozygous or homozygous variants) were included in the CART models. Ten-fold cross-validation was used to reduce overfitted trees to their optimal size. Indicator variables for terminal nodes in the final tree were used in logistic regression models to estimate ORs and 95% CIs.

Unless otherwise specified, statistical analyses were done with STATA version 8.2, Special Edition (STATA Corp., College Station, TX).

Results

A description of the seven NER genes and 22 SNPs evaluated in this study, including minor allele frequencies in the control population, is shown in Table 1. Frequencies were similar to those previously reported in Caucasian populations (3). Exploration of associations for each individual SNP with bladder cancer risk revealed significant associations with SNPs in four of seven NER genes (Supplementary Table S1). Table 2

Table 3. Association between selected polymorphisms in five genes involved in NER and bladder cancer risk for never and ever cigarette smokers, adjusted for age, region, and gender

Gene	SNP	Smoking status	Homozygous wild type		Heterozygous/homozygous variant		OR (95% CI)	P	P _{interaction}
			Cases	Controls	Cases	Controls			
XPC	K939Q	Never	131	279	24	47	1.1 (0.6-1.8)	0.84	0.82
		Ever	825	696	131	96	1.1 (0.9-1.5)	0.38	
RAD23B	IVS5-15A>G	Never	48	109	107	221	1.1 (0.7-1.7)	0.64	0.55
		Ever	289	281	680	508	1.3 (1.0-1.6)	0.02	
ERCC2	R156R	Never	26	105	129	230	2.2 (1.3-3.5)	0.002	0.04
		Ever	242	222	736	568	1.2 (1.0-1.5)	0.05	
	D312N	Never	67	160	89	169	1.2 (0.8-1.8)	0.34	0.61
		Ever	450	378	523	415	1.1 (0.9-1.3)	0.42	
ERCC1	IVS5+33A>C	Never	66	158	93	173	1.3 (0.9-1.9)	0.19	0.40
		Ever	424	354	553	440	1.1 (0.9-1.3)	0.45	
ERCC5	H46H	Never	126	259	33	72	0.9 (0.6-1.5)	0.78	0.27
		Ever	725	626	246	171	1.3 (1.0-1.6)	0.04	
ERCC5	M254V	Never	49	101	103	217	1.0 (0.6-1.5)	0.86	1.00
		Ever	336	265	615	511	1.0 (0.8-1.2)	0.70	
	D1104H	Never	139	287	11	22	1.1 (0.5-2.3)	0.85	0.47
		Ever	840	708	87	51	1.5 (1.0-2.1)	0.03	
		Never	78	172	79	163	1.0 (0.7-1.5)	0.95	0.76
		Ever	551	435	433	366	0.9 (0.8-1.1)	0.57	

shows ORs (95% CI) for selected SNPs with significant or borderline significant associations in the current study, or SNPs evaluated in relation to bladder cancer risk in previous reports. Compared with homozygous wild-type individuals, those carrying genotypes with variant alleles for *RAD23B* IVS5-15A>G (OR, 1.3; 95% CI, 1.1-1.5; $P = 0.01$), *ERCC2* R156R (OR, 1.3; 95% CI, 1.1-1.6; $P = 0.006$), *ERCC1* IVS5+33A>C (OR, 1.2; 95% CI, 1.0-1.5, $P_{\text{trend}} = 0.04$), and *ERCC5* M254V (OR, 1.4; 95% CI, 1.0-2.0; $P = 0.04$) had a significant increase in risk. None of the SNPs evaluated in *XPC*, *ERCC6*, or *ERCC4* were significantly related to risk (Supplementary Table S1). A global test for pathway effects as determined by the 22 NER polymorphisms indicated that variation in this pathway significantly predicts bladder cancer risk ($P = 0.04$).

The observed genotype associations with bladder cancer risk were not substantially modified by gender, age, dietary folic acid intake, and *NAT2* or *GSTM1* genotypes (data not shown), or consistently found in any particular stratum defined by these variables. Associations for *RAD23B* IVS5-15A>G, *ERCC1* IVS5+33A>C, and *ERCC5* M254V seemed to be stronger for cigarette smokers than for never smokers; however, the number of never smokers was relatively small and differences were not statistically significant (Table 3). In contrast, we found a suggestion for a stronger association between bladder cancer and *ERCC2* R156R for never smokers compared with ever smokers (Table 3; $P_{\text{interaction}} = 0.04$). Genotype relative risks were similar for former and current smokers, and no significant differences were observed for

Table 4. Pairwise joint associations for four NER genes (*RAD23B* IVS5-15A>G, *ERCC2* R156R, *ERCC1* IVS5+33A>C, *ERCC5* M254V) and bladder cancer risk, adjusted for gender, age, and smoking status

Genotype 1	Genotype 2	Cases, n (%)	Controls, n (%)	OR (95% CI)	P	P _{interaction}	
<i>RAD23B</i> IVS5-15A>G	<i>ERCC2</i> R156R	GG	90 (8)	110 (10)	1.0 (reference)		
		GA or AA	175 (16)	211 (19)	0.9 (0.6-1.3)	0.60	
		GG	245 (22)	275 (25)	1.0 (0.7-1.4)	0.84	
		GA or AA	599 (54)	506 (46)	1.4 (1.0-1.9)	0.06	0.04
<i>RAD23B</i> IVS5-15A>G	<i>ERCC1</i> IVS5+33A>C	GG	244 (22)	294 (27)	1.0 (reference)		
		GA or AA	589 (53)	573 (52)	1.3 (1.0-1.6)	0.034	
		GG	86 (8)	93 (8)	1.1 (0.8-1.6)	0.49	
		GA or AA	186 (17)	146 (13)	1.6 (1.2-2.1)	0.002	0.63
<i>RAD23B</i> IVS5-15A>G	<i>ERCC5</i> M254V	GG	287 (27)	345 (33)	1.0 (reference)		
		GA or AA	676 (64)	633 (60)	1.3 (1.1-1.6)	0.01	
		GG	34 (3)	23 (2)	1.9 (1.1-3.4)	0.03	
		GA or AA	61 (6)	48 (5)	1.6 (1.0-2.5)	0.03	0.22
<i>ERCC2</i> R156R	<i>ERCC1</i> IVS5+33A>C	AA	232 (21)	280 (25)	1.0 (reference)		
		AC or CC	604 (54)	590 (53)	1.2 (1.0-1.5)	0.06	
		AA	32 (3)	43 (4)	1.0 (0.6-1.6)	0.881	
		AC or CC	246 (22)	197 (18)	1.5 (1.2-2.0)	0.002	0.37
<i>ERCC2</i> R156R	<i>ERCC5</i> M254V	AA	226 (21)	276 (26)	1.0 (reference)		
		AC or CC	740 (70)	704 (67)	1.3 (1.0-1.6)	0.03	
		AA	30 (3)	33 (3)	1.1 (0.7-1.9)	0.67	
		AC or CC	68 (6)	39 (4)	2.2 (1.4-3.4)	0.0007	0.24
<i>ERCC1</i> IVS5+33A>C	<i>ERCC5</i> M254V	CC	722 (68)	768 (73)	1.0 (reference)		
		AC or AA	240 (23)	217 (21)	1.2 (1.0-1.5)	0.11	
		CC	75 (7)	61 (6)	1.4 (0.9-2.0)	0.11	
		AC or AA	23 (2)	10 (1)	2.6 (1.2-5.7)	0.02	0.28

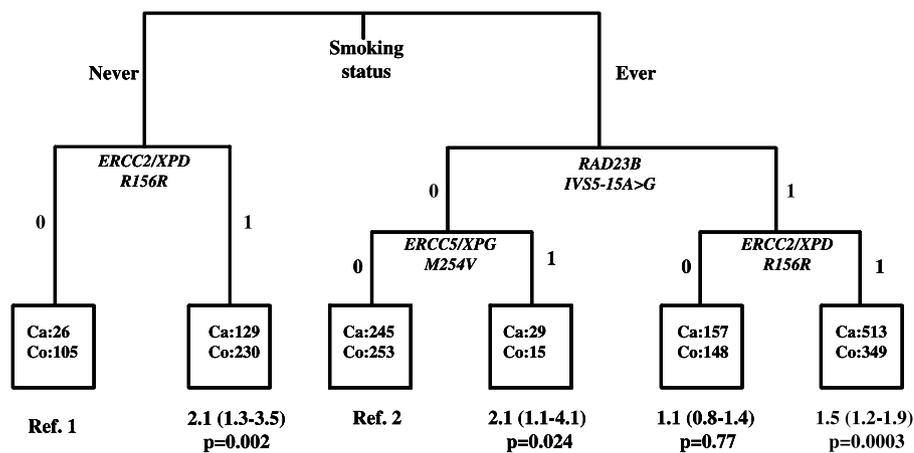


Figure 1. Classification tree model for cigarette smoking, NER polymorphisms and bladder cancer risk. *Ca*, cases; *Co*, controls. ORs (95% CIs) and *P*s under terminal nodes are for genotype-bladder cancer associations within smoking categories estimated from a logistic regression model (ref. 1 is the reference group among never smokers and ref. 2 is the reference group among ever smokers). Codes for genotypes are 0 for homozygous wild type and 1 for heterozygous or homozygous variants.

smokers of black or blond tobacco types nor for the effect of smoking intensity (average cigarettes per day) or duration on bladder cancer risk among smokers with and without the putative "at risk" alleles (data not shown). Evaluation of gene-gene pairwise interactions between the four SNPs associated with bladder cancer risk showed joint effects ranging from 1.4 to 2.6 (Table 4). Joint effects did not significantly depart from a multiplicative model, except for the combination of variant genotypes for *RAD23B IVS5-15A>G* and *ERCC2 R156R* (joint OR, 1.4; 95% CI, 1.0-1.9; $P_{\text{interaction}} = 0.04$).

Figure 1 shows the final CART model for smoking status and NER genotypes. Ten-fold cross-validation indicated an optimal tree with six terminal nodes. The number of cases and controls are shown for each terminal node. Not surprisingly, the first split was for smoking status. Ever smokers were further split according to combinations of three genotypes, suggesting a possible three-way interaction between *RAD23B IVS5-15A>G*, *ERCC5 M254V*, and *ERCC2 R156R*. On the other hand, only *ERCC2 R156R* seemed to be important in determining risk among never smokers. This CART model suggested the presence of subgroups of individuals defined by smoking and NER genotypes that could have substantial increases in risk.

Discussion

In this report, we showed that genetic variation in the NER pathway, which is responsible for protecting against DNA damage from chemical carcinogens in tobacco smoke (2), significantly contributes to bladder cancer risk (global test $P = 0.04$). Of the 22 SNPs in seven NER genes evaluated, four were significantly related to a small increase in bladder cancer risk (*RAD23B IVS5-15A>G*, *ERCC2 R156R*, *ERCC1 IVS5+33A>C*, and *ERCC5 M254V*).

Previous epidemiologic studies have evaluated a limited number of variants in NER genes in relation to bladder cancer risk (refs. 8-13; Table 5). The *XPC* gene codes for a protein involved in the recognition of the DNA damage to be repaired by NER (2). The homozygous variant genotype for *XPC K939Q* significantly increased risk of bladder cancer in a previous study in Sweden (8) but not in a study in the United Kingdom (9) or in the current study population in Spain; however, a small increase in risk for homozygous variants cannot be excluded. Additional epidemiologic evidence and a better understanding of the functional significance of this amino acid change (5) would be needed to establish or rule out a potential small effect. This variant is in strong linkage disequilibrium

Table 5. Summary of previous studies on associations between polymorphisms in selected NER genes and bladder cancer risk

Gene	SNP	First author	Year	Country	Cases	Controls	Minor allele frequency (controls)	Heterozygous vs homozygous wild type, OR (95% CI)	Homozygous variants vs homozygous wild type, OR (95% CI)
<i>XPC</i>	<i>K939Q</i>	Sanyal S.	2003	Sweden	327	327	0.34	1.1 (0.8-1.6)	2.0 (1.1-3.6)
		Sak S.C.	2005	United Kingdom	547	579	0.40	0.8 (0.6-1.0)	1.0 (0.7-1.4)
	<i>PolyAT</i>	Current study		Spain	1,150	1,149	0.40	1.2 (1.0-1.4)	1.1 (0.8-1.4)
		Sak S.C.	2005	United Kingdom	547	579	0.40	0.8 (0.6-1.0)	1.0 (0.7-1.4)
<i>IVS11-6</i>		Sak S.C.	2005	United Kingdom	547	579	0.40	0.8 (0.6-1.0)	1.0 (0.7-1.4)
<i>ERCC2</i>	<i>K751Q</i>	Matullo G.*	2001	Italy	124	85	0.42	0.8 (0.4-1.6)	1.1 (0.4-2.9)
		Shen M.	2003	Italy	201	214	0.40	0.9 (0.6-1.4)	1.0 (0.6-1.8)
		Stern, M.C.	2002	United States	228	210	0.38	1.0 (0.7-1.5)	0.8 (0.4-1.3)
	<i>D312N</i>	Sanyal S.	2003	Sweden	327	327	0.38	1.1 (0.7-1.6)	1.3 (0.8-2.2)
		Schabath M.	2005	United States	505	486	0.34	1.0 (0.8-1.4)	1.2 (0.9-1.4)
	Current study	Schabath M.	2005	Spain	1,150	1,149	0.33	1.1 (0.9-1.3)	1.2 (0.9-1.6)
<i>ERCC5</i>	<i>D1104H</i>	Current study		Spain	1,150	1,149	0.31	1.1 (0.9-1.3)	1.2 (0.9-1.6)
		Sanyal S.	2003	Sweden	327	327	0.23	1.1 (0.8-1.6)	0.4 (0.2-0.9)
		Current study		Spain	1,150	1,149	0.27	0.9 (0.8-1.1)	0.9 (0.6-1.3)

*OR (95% CI) were calculated from published data.

with two other variants that could affect the function of the gene: *XPC* polyAT (*XPC-PAT*) has been linked to reduced repair capacity (19), and *XPC IVS11-6* alters protein function (20). However, none of these variants were associated with elevated bladder cancer risk in a study of 547 cases in the United Kingdom (9).

The protein coded by *RAD23B* forms a complex with *XPC* during damage recognition (2). In the current study, a variant in intron 5 of *RAD23B* (*IVS5-15A>G*) was associated with a significant increased risk. This variant has not been previously evaluated in relation to bladder cancer risk, and its functional significance is unknown.

The *ERCC2* gene encodes a DNA helicase subunit of the core transcription factor IIIH that is essential for NER and transcription (2). A nonsynonymous variant (*K751Q*) in *ERCC2* that has been linked to deficiencies in NER repair in some functional studies (4-6) was not associated with a significant increase in bladder cancer risk in the current Spanish population, or in five previous studies conducted in the United States, Sweden, and Italy (refs. 8, 10-13; Table 5). The *ERCC2 D312N* polymorphism was also not associated with an overall increase in bladder cancer risk in our population nor in a previous study of 505 cases in the United States (ref. 10; Table 5).

The proteins coded by *ERCC4* and *ERCC1* form a heterodimeric protein with endonuclease activity that cuts the DNA strand at the 5' side of the damage (2). The *ERCC5* gene codes for an endonuclease that cuts the DNA strand at the 3' side of the damage. Our data suggested an increased risk of bladder cancer associated with variant alleles in *ERCC1 IVS5+33A>C* and *ERCC5 M254V*. These associations have not been previously reported and the functional significance of the variants is unknown. Thus, they need to be confirmed in future studies. Our data were consistent with a small reduction in risk associated with the variant allele for *ERCC5 D1104H*, as previously indicated in a bladder cancer study in Sweden (8) and a study of lung cancer (21). However, this protection was not significant in our study population.

Evaluation of pairwise joint associations between putative susceptibility variants suggested that individuals carrying two variants might have substantial increases in risk. CART, a technique to explore high-order interactions (18), suggested the presence of subgroups of individuals defined by smoking and NER genotypes that could have substantial increases in risk. We did internal cross-validation to determine the optimal tree model. However, exploratory techniques are prone to overfitting the data, and the ORs for specific genotype combinations indicated by these models need to be interpreted with caution. External validation in independent data is needed to confirm these findings.

The strengths of our study population include high participation rates and large sample size. Our study had adequate statistical power to detect relatively small genotype associations; however, the power to detect interactions was limited. Rather than carrying out a detailed characterization of the genetic variation in any particular NER gene, we selected a few SNPs in key NER genes to attempt to capture common variation in this pathway as a whole. Common variation in individual genes in a complex pathway involving multiple genes, such as NER, is unlikely to have strong associations with cancer risk. This is especially true for genetic markers of unknown functional significance that are used as potential surrogates for "causative" variants. When multiple genes in one pathway have weak associations with risk, a global test for pathway effects, such as the one used in this report, can be more powerful than individual tests to detect an association (22). In addition, because all SNPs are considered simultaneously, a global test also addresses the problem of multiple comparisons. Because we did not include a dense survey of SNPs in genes of interest intended to

capture haplotype diversity, it is possible that additional genetic variants are related to bladder cancer risk. Three genotypes that were not significantly associated with bladder cancer risk showed small but significant departures from Hardy-Weinberg equilibrium in the control population (i.e., *RAD23B IVS5-57*, *ERCC4 R415Q*, and *ERCC4 IVS9-35C>T*). Duplicate quality control samples showed $\geq 98\%$ genotype agreement for all three assays, indicating that departures were unlikely to be due to genotyping error. Furthermore, a sensitivity analysis where ORs and 95% CIs were reestimated using the expected genotype frequencies under Hardy-Weinberg equilibrium in the control population showed no substantial changes in estimated OR's.

In conclusion, our results provide support for an overall association between genetic variation in the NER pathway and bladder cancer risk and suggest the presence of gene-gene and gene-smoking interactions. However, it is unclear what the causative variants are, and a more detailed characterization of the genetic variation in key NER genes is warranted. Pooling comparable data from current and ongoing studies will be required to confirm small associations and to evaluate complex interrelationships between genetic variants and cigarette smoking suggested by this report. These efforts might ultimately help identify multiple susceptibility variants that jointly could be responsible for substantial increases in bladder cancer risk.

Participating Study Centers in Spain

Institut Municipal d'Investigació Mèdica, Universitat Pompeu Fabra, Barcelona: Coordinating Center (M. Kogevinas, N. Malats, F.X. Real, M. Sala, G. Castaño, M. Torà, D. Puente, C. Villanueva, C. Murta, J. Fortuny, E. López, S. Hernández, R. Jaramillo); Hospital del Mar, Universitat Autònoma de Barcelona, Barcelona (J. Lloreta, S. Serrano, L. Ferrer, A. Gelabert, J. Carles, O. Bielsa, K. Villadiego); Hospital Germans Tries i Pujol, Badalona, Barcelona (L. Cecchini, J.M. Saladié, L. Ibarz); Hospital de Sant Boi, Sant Boi, Barcelona (M. Céspedes); Centre Hospitalari Parc Taulí, Sabadell, Barcelona (C. Serra, D. García, J. Pujadas, R. Hernando, A. Cabezuelo, C. Abad, A. Prera, J. Prat); ALTHAIA, Manresa, Barcelona (M. Domènech, J. Badal, J. Malet); Hospital Universitario, La Laguna, Tenerife (R. García-Closas, J. Rodríguez de Vera, A.I. Martín); Hospital La Candelaria, Santa Cruz, Tenerife (J. Taño, F. Cáceres); Hospital General Universitario de Elche, Universidad Miguel Hernández, Elche, Alicante (A. Carrato, F. García-López, M. Ull, A. Teruel, E. Andrada, A. Bustos, A. Castillejo, J.L. Soto); Universidad de Oviedo, Oviedo, Asturias (A. Tardón); Hospital San Agustín, Avilés, Asturias (J.L. Guate, J.M. Lanzas, J. Velasco); Hospital Central Covadonga, Oviedo, Asturias (J.M. Fernández, J.J. Rodríguez, A. Herrero); Hospital Central General, Oviedo, Asturias (R. Abascal, C. Manzano); Hospital de Cabueñes, Gijón, Asturias (M. Rivas, M. Arguelles); Hospital de Jove, Gijón, Asturias (M. Díaz, J. Sánchez, O. González); Hospital de Cruz Roja, Gijón, Asturias (A. Mateos, V. Frade); Hospital Alvarez-Buylla, Mieres, Asturias (P. Muntañola, C. Pravia); Hospital Jarrio, Coaña, Asturias (A.M. Huescar, F. Huergo); Hospital Carmen y Severo Ochoa, Cangas, Asturias (J. Mosquera).

Acknowledgments

We thank Robert C. Saal (Westat, Rockville, MD) and Leslie Carroll and Jane Wang (IMS, Silver Spring, MD) for their support in study and data management, Dr. Zeynep Kalaylioglu (IMS) for her support in data analysis, Dr. Maria Sala (Institut Municipal d'Investigació Mèdica, Barcelona, Spain) for her work in data collection, Francisco Fernandez for his work on data management, Dr. Montserrat Torà for her work in the coordination of sample collection and blood processing, and physicians, nurses, interviewers, and study participants for their efforts during field work.

References

- Hoeijmakers JH. Human nucleotide excision repair syndromes: molecular clues to unexpected intricacies. *Eur J Cancer* 1994;30A:1912–21.
- Friedberg EC. How nucleotide excision repair protects against cancer. *Nat Rev Cancer* 2001;1:22–33.
- Neumann AS, Sturgis EM, Wei Q. Nucleotide excision repair as a marker for susceptibility to tobacco-related cancers: a review of molecular epidemiological studies. *Mol Carcinog* 2005;42:65–92.
- Au WW, Navasumrit P, Ruchirawat M. Use of biomarkers to characterize functions of polymorphic DNA repair genotypes. *Int J Hyg Environ Health* 2004;207:301–13.
- Vodicka P, Kumar R, Stetina R, et al. Genetic polymorphisms in DNA repair genes and possible links with DNA repair rates, chromosomal aberrations and single-strand breaks in DNA. *Carcinogenesis* 2004;25:757–63.
- Pavanello S, Pulliero A, Siwinska E, Mielzynska D, Clonfero E. Reduced nucleotide excision repair and GSTM1-null genotypes influence anti-B[a]PDE-DNA adduct levels in mononuclear white blood cells of highly PAH-exposed coke oven workers. *Carcinogenesis* 2005;26:169–75.
- Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002;11:1513–30.
- Sanyal S, Festa F, Sakano S, et al. Polymorphisms in DNA repair and metabolic genes in bladder cancer. *Carcinogenesis* 2004;25:729–34.
- Sak SC, Barrett JH, Paul AB, Bishop DT, Kiltie AE. The polyAT, intronic IVS11–6 and Lys939Gln XPC polymorphisms are not associated with transitional cell carcinoma of the bladder. *Br J Cancer* 2005;92:2262–5.
- Schabath MB, Delcloas GL, Grossman B, et al. Polymorphism in XPD exons 10 and 23 and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 2005;14:878–84.
- Stern MC, Johnson LR, Bell DA, Taylor JA. XPD codon 751 polymorphism, metabolism genes, smoking, and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002;11:1004–11.
- Shen M, Hung RJ, Brennan P, et al. Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy. *Cancer Epidemiol Biomarkers Prev* 2003;12:1234–40.
- Matullo G, Guarrera S, Carturan S, et al. DNA repair gene polymorphisms, bulky DNA adducts in white blood cells and bladder cancer in a case-control study. *Int J Cancer* 2001;92:562–7.
- Garcia-Closas M, Malats N, Silverman D, et al. NAT2 slow acetylation, GSTM1 null genotype, and risk of bladder cancer: results from the Spanish Bladder Cancer Study and meta-analyses. *Lancet* 2005;366:649–59.
- Slimani N, Farran A, Torrent M. European Prospective Investigation into Cancer and Nutrition (EPIC): food composition tables-Spain. Lyon (France): International Agency for Research on Cancer (IARC); 1996.
- Packer BR, Yeager M, Staats B, et al. SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. *Nucleic Acids Res* 2004;32:D528–32.
- Westfall PH, Young SS. Resampling based multiple testing. New York (NY): John Wiley & Sons, Inc.; 1993.
- Zhang H, Bonney G. Use of classification trees for association studies. *Genet Epidemiol* 2000;19:323–32.
- Qiao Y, Spitz MR, Shen H, et al. Modulation of repair of ultraviolet damage in the host-cell reactivation assay by polymorphic XPC and XPD/ERCC2 genotypes. *Carcinogenesis* 2002;23:295–9.
- Khan SG, Muniz-Medina V, Shahlavi T, et al. The human XPC DNA repair gene: arrangement, splice site information content and influence of a single nucleotide polymorphism in a splice acceptor site on alternative splicing and function. *Nucleic Acids Res* 2002;30:3624–31.
- Jeon HS, Kim KM, Park SH, et al. Relationship between XPG codon 1104 polymorphism and risk of primary lung cancer. *Carcinogenesis* 2003;24:1677–81.
- Schaid DJ, McDonnell SK, Hebring SJ, Cunningham JM, Thibodeau SN. Nonparametric tests of association of multiple genes with human disease. *Am J Hum Genet* 2005;76:780–93.