

# Genetic Variation in the Nucleotide Excision Repair Pathway and Colorectal Cancer Risk

Sonja I. Berndt,<sup>1</sup> Elizabeth A. Platz,<sup>1</sup> M. Daniele Fallin,<sup>1</sup> Lucy W. Thuita,<sup>1</sup> Sandra C. Hoffman,<sup>1</sup> and Kathy J. Helzlsouer<sup>1,2</sup>

<sup>1</sup>Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health and <sup>2</sup>Prevention and Research Center, The Weinberg Center for Women's Health and Medicine, Mercy Medical Center, Baltimore, Maryland

## Abstract

Nucleotide excision repair (NER) enzymes are critical for the removal of bulky DNA adducts caused by environmental carcinogens, such as heterocyclic amines and polycyclic aromatic hydrocarbons, which are found in two putative risk factors for colorectal cancer, tobacco smoke and meat cooked at high temperature. To examine the association between common genetic variants in NER genes and the risk of colorectal cancer, we conducted a case-cohort study within the CLUE II cohort. Twenty-two single nucleotide polymorphisms in 11 NER genes were genotyped in 250 colorectal cancer cases and a subcohort of 2,224 participants. Incidence rate ratios (RR) and 95% confidence intervals (95% CI) were estimated using a modified Cox regression model and robust variance estimate. The *ERCC6 1213G* variant,

which is thought to reduce NER capacity, was associated with an increased risk of colorectal cancer compared with the homozygous wild type (RR, 1.36; 95% CI, 1.00-1.86 and RR, 2.64; 95% CI, 1.53-4.58 for the *RG* and *GG* genotypes respectively with  $P_{\text{trend}} = 0.0006$ ). Having at least one *XPC 492H* allele was also associated with an increased risk of colorectal cancer (RR, 1.75; 95% CI, 1.20-2.57). When the combined effects of *ERCC6 R1213G* and *XPC R492H* were examined, the risk of colorectal cancer significantly increased with increasing number of variant alleles ( $P_{\text{trend}} = 0.00003$ ). Our study suggests that genetic polymorphisms in the NER genes, *ERCC6* and *XPC*, may be associated with an increased risk of colorectal cancer. (Cancer Epidemiol Biomarkers Prev 2006;15(11):2263-9)

## Introduction

Genetic factors are thought to play a role in susceptibility to colorectal cancer with heritable factors estimated to account for 35% of the risk (1). Individuals who have first-degree relative with colorectal cancer have a 2-fold increased risk of the colorectal cancer compared with those without a family history (2), suggesting that either shared environmental or genetic factors contribute to risk for colorectal carcinogenesis. Although known high-penetrance genes cause specific inherited colorectal cancer syndromes, such as familial adenomatous polyposis coli and hereditary nonpolyposis colorectal cancer (3), other genetic risk factors for colorectal cancer are not well characterized, and low-penetrance genes may contribute to colorectal cancer susceptibility by interacting with environmental risk factors, such as diet.

DNA repair mechanisms play a central role in maintaining the integrity of the genome and preventing the development of somatic mutations and alterations, which can lead to carcinogenesis. Defects in DNA repair have been linked to two familial types of colorectal cancer. Hereditary nonpolyposis colorectal cancer is thought to be caused by mutations in genes involved in mismatch repair (3), and mutations in the base excision repair gene *MUTYH* are believed to be responsible for a rare, autosomal recessive form of colorectal polyposis (4, 5). However, few studies have examined the role of common polymorphisms in DNA repair genes and the risk of colorectal cancer.

Nucleotide excision repair (NER) is one of the most versatile DNA repair pathways, and deficiencies in the pathway could increase cancer susceptibility. NER enzymes respond to a wide range of DNA damage but are particularly important for the removal of bulky adducts caused by environmental carcinogens, such as heterocyclic amines and polycyclic aromatic hydrocarbons, which are found in tobacco smoke and meats cooked at high temperature. Both tobacco smoke and meat consumption, particularly red meat consumption, are putative environmental risk factors for colorectal neoplasia (6, 7). Prospective studies have shown that high intake of red meat is consistently associated with an increased risk of colorectal cancer [reviewed in ref. 7]. Although the mechanism by which red meat increases the risk of colorectal neoplasia is not entirely clear, heterocyclic amines and polycyclic aromatic hydrocarbons that are formed in the meat when it is cooked at high temperatures are hypothesized to contribute to the increased risk observed (8, 9). Because NER enzymes may play an important role in repairing damage from these carcinogens, they may protect against development of colorectal cancer.

NER consists of two pathways, global genome repair and transcription-coupled repair, which share many of the same enzymes but differ in their recognition of DNA damage. Global genome repair removes DNA damage anywhere in the genome, whereas transcription-coupled repair preferentially removes damaging lesions from actively transcribed genes. In the global genome repair pathway, DNA damage is recognized by the xeroderma pigmentosum complementation group C (XPC)-hHR23B complex (10), which binds to the lesion and recruits other repair enzymes to the damaged site (11). In transcription-coupled repair, the stalled RNA polymerase II complex on the transcribed gene is thought to trigger the recruitment of Cockayne syndrome complementation group A (CSA) and group B (CSB) proteins, as well as other repair enzymes, to the damaged site (12). In both pathways, the DNA is locally unwound by two DNA helicases, XPD and XPB, and stabilized by RPA and XPA (13). XPG and the

Received 5/30/06; revised 8/10/06; accepted 9/13/06.

Grant support: National Institute on Aging grant IU01 AG18033.

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: Kathy J. Helzlsouer, Prevention and Research Center, Mercy Medical Center, 227 St. Paul Place, Baltimore, MD 21202. Phone: 410-951-7950; Fax: 410-951-7931.

E-mail: [khelzlsouer@mdmercy.com](mailto:khelzlsouer@mdmercy.com)

Copyright © 2006 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-06-0449

ERCC1/XPF complex cleave the DNA (14, 15), removing the damage, and DNA replication factors synthesize new nucleotides to fill the gap.

In this report, we examined the association between colorectal cancer risk and 22 common polymorphisms in 11 genes [*ERCC6* (CSB), *XPC*, *RAD23B*, *ERCC2* (XPD), *ERCC3* (XPB), *XPA*, *RPA2*, *ERCC5* (XPG), *ERCC4* (XPF), *ERCC1*, and *LIG1*] involved in global genome and transcription-coupled NER using a case-cohort study within the CLUE II cohort. We hypothesized that variants that decreased the ability of NER to repair damage would increase the risk of colorectal cancer.

## Materials and Methods

**Study Population.** CLUE II is a community-based cohort that was established in 1989 to investigate risk factors for cancer and other diseases. A total of 22,887 adult residents of Washington County, Maryland consented to participate in the cohort, the majority of whom were white (98%), reflecting the racial composition of this county, and female (59%). At baseline, participants provided a blood specimen and completed a brief questionnaire that ascertained information on demographic factors, risk factors (e.g., smoking status, height, and weight), and medical history (e.g., previous cancer diagnoses). Blood specimens were collected by trained personnel using 20-mL heparinized Vacutainer tubes (Becton Dickinson, Rutherford, NJ) and immediately refrigerated until they were centrifuged (usually 2-6 hours later), divided into aliquots of plasma, buffy coat, and RBC, and subsequently frozen at  $-70^{\circ}\text{C}$  for future use. Information on diet was obtained at baseline with a 60-item Block food frequency questionnaire (16). Additional data on risk factors (e.g., family history of cancer), cancer screening, and medical history were obtained from follow-up questionnaires administered in 1996, 1998, and 2000. All participants provided written consent at study entry, and the study was approved by the Institutional Review Board at the Johns Hopkins Bloomberg School of Public Health.

**Identification of Subcohort.** For this study, we used a case-cohort approach in which person-years at risk were estimated from a random sample of the cohort. A subcohort was identified by taking a 10% age-stratified, random sample of CLUE II participants ( $n = 2,289$ ) who donated a blood specimen and were adult residents of Washington County, Maryland. Participants diagnosed with anal cancer ( $n = 1$ ) or colorectal cancer before 1989 ( $n = 21$ ) were excluded, and 43 participants with insufficient DNA for genotyping were eliminated, leaving 2,224 persons available for the analysis.

**Identification of Cases.** Incident cases of colorectal cancer were identified through linkage to the Washington County Cancer Registry and, since 1992, the Maryland Cancer Registry. The Washington County Cancer Registry identifies cancer cases from discharge records and pathology reports from Washington County Hospital, which is the only hospital in the county, and from death certificates. A comparison of the colorectal cancer cases ascertained through the Washington County Cancer Registry to those identified by the Maryland Cancer Registry in 1998 revealed that 98% of the colorectal cases from Washington County residents were captured by the Washington County Registry. Since 1993, all hospitals, cancer diagnostic laboratories, and radiation therapy centers in Maryland have been required by law to report incident cancer cases to the Maryland Cancer Registry, and currently, the Maryland Cancer Registry is certified as being >95% complete by the North American Association of Central Cancer Registries.

A total of 272 cases of colorectal cancer were identified among CLUE II participants in Washington County, Maryland

from the date of blood draw through December 2003. Twenty-two of these cases had insufficient DNA for genotyping and were excluded. Of the 250 remaining cases, 198 individuals were diagnosed with colon cancer [International Classification of Diseases, Ninth Revision (ICD-9) codes 153.0-153.9 and International Classification of Diseases, Tenth Revision (ICD-10) codes C18.0-C18.9] and 52 were diagnosed with rectal cancer (ICD-9 codes 154.0, 154.1, and 154.8 and ICD-10 codes C19 and C20). All cases were confirmed by pathology records, and 96% of cancers were adenocarcinoma. Tumors from the cecum through the splenic flexure were considered to be proximal colon cancers, whereas tumors from the descending and sigmoid colon were classified as distal colon cancers. Of the 198 colon cancer cases, 106 cases had proximal colon tumors, 78 cases had distal colon tumors, and 14 cases had tumors in an unspecified location of the colon or had tumors in both the proximal and distal colon. Information on cancer stage was available for 191 cases, including 148 colon cancers and 43 rectal cancers. Forty-one percent of the cases were diagnosed at an advanced stage (stage III-IV).

**Laboratory Assays.** DNA was extracted from the buffy coat fraction of the blood specimens using an alkaline lysis method (17), and aliquots of the extracted DNA were shipped overnight on dry ice to Applied Biosystems (Foster City, CA), where the genotyping was conducted. Twenty-two single nucleotide polymorphisms in 11 NER genes were selected from public databases for genotyping. Whenever possible, we chose polymorphisms that were suggested to alter function (18-21), encoded for a nonsynonymous amino acid change, or were located within the 5' or 3' untranslated region of the gene and thus could potentially alter mRNA stability. All of the polymorphisms [*ERCC1* *IVS3+74G>C* (rs3212948), *ERCC1* *G>T* 196 bp 3' of *STP* (a.k.a. *ASE-1* *Gln504Lys*; rs3212986), *ERCC2* *Asp312Asn* (rs1799793), *ERCC2* *IVS19-70G>A* (rs1799787), *ERCC2* *Gln751Lys* (rs13181), *ERCC3* *IVS6-108A>C* (rs4150416), *ERCC3* *G>A* 487 bp 3' of *STP* (rs2276583), *ERCC4* *Arg415Gln* (rs1800067), *ERCC4* *Ser662Pro* (rs2020955), *ERCC5* *Cys529Ser* (rs2227869), *ERCC5* *Asp1104His* (rs17655), *ERCC6* *Met1097Val* (rs2228526), *ERCC6* *Arg1213Gly* (rs2228527), *ERCC6* *Arg1230Pro* (rs4253211), *LIG1* *Ex2-24C>T* in 5' untranslated region (rs20579), *RAD23B* *Ala249Val* (rs1805329), *RPA2* *Ex9-51T>C* in 3' untranslated region (rs7356), *XPA* *Ex6-327C>G* in 3' untranslated region (rs3176751), *XPC* *Arg492His* (rs2227999), *XPC* *Ala499Val* (rs2228000), *XPC* *Arg687Arg* (rs3731151), and *XPC* *Lys939Gln* (rs2228001)] were genotyped using TaqMan assays. Laboratory personnel were blinded to case-control status and duplicate samples were added to the genotype assays for quality control evaluation. Of the study participants, genotyping was successfully completed for 93% to 98% of subjects, depending on the assay.

Published genotype frequencies among Caucasians could not be found for *RPA2* *Ex9-51T>C* or *XPA* *Ex6-327C>G*. However, the genotype frequencies in the subcohort for the remaining polymorphisms evaluated in our study were similar to those reported for Caucasians on the National Cancer Institute SNP500 Cancer Project website.<sup>3</sup> The genotype frequencies in the subcohort by race were consistent with Hardy-Weinberg proportions ( $P > 0.05$ ) using a goodness-of-fit  $\chi^2$  test (or exact test if cell counts were small) for all SNPs, except for *ERCC2* *D312N* ( $P = 0.05$  for Caucasians) and *XPA* *Ex6-327C>G* ( $P = 0.02$  for Caucasians). Although statistically significant, a closer inspection of the data revealed that the observed genotype frequencies for the *XPA* polymorphism were not much different than the expected frequencies under Hardy-Weinberg equilibrium (CC, 99.40% versus 99.35%;

<sup>3</sup> <http://snp500cancer.nci.nih.gov>.

**Table 1. Baseline characteristics of the CLUE II cohort and subcohort**

	Full cohort ( <i>n</i> = 22,287)	Subcohort ( <i>n</i> = 2,224)	<i>P</i>
Age at blood draw (mean ± SD), y	48.4 ± 16.3	48.5 ± 16.3	0.90
Female, %	58.5	58.1	0.72
Caucasian, %	98.3	98.3	0.83
Education, %			
Less than 12 y	20.4	21.3	
High school graduate	45.3	45.6	
Beyond high school	34.3	33.1	0.42
Smoking status, %			
Never	50.0	48.3	
Former cigarette smoker	28.1	29.0	
Current cigarette smoker	19.7	20.5	
Ever cigar or pipe smoker but never cigarette smoker	2.1	2.2	0.49
Cholesterol level (mean ± SD), mg/dl	205.6 ± 40.6	206.2 ± 39.5	0.50
Body mass index (mean ± SD), kg/m <sup>2</sup>	26.2 ± 4.9	26.2 ± 5.0	0.54

CT, 0.55% versus 0.65%; TT, 0.05% versus 0%). These were included in subsequent analyses.

**Dietary Data.** The food frequency questionnaire was returned by ~80% of the participants included in this study. Participants who reported eating fewer than three foods per day, left more than half the questions blank, or had values for total energy intake outside the range of 500 to 3,500 kcal/d for women or 800 to 5,000 kcal/d for men were considered to have an invalid food frequency questionnaire and were excluded, leaving a total of 202 cases and 1,563 subcohort participants for dietary analyses.

Usual intake of meat was assessed with the information ascertained from the 13 questions in the food frequency questionnaire about meat intake over the past year. Consumption of each meat item in grams per day was estimated by multiplying the number of grams per medium serving size by the average frequency per day. Total meat and red meat intake was assessed by summing the relevant meat items for each group. For this analysis, red meat included hamburgers, cheeseburgers, meatloaf, beef, beef stew, pork, hot dogs, bacon, sausage, ham, bologna, salami, and other lunch meats. Thirds of meat consumption were estimated using the distribution of meat intake in the subcohort.

**Statistical Analysis.** For this study, we analyzed the data using a case-cohort approach. Person-years at risk for the entire cohort were estimated from the subcohort, and cases outside the subcohort only entered the analysis at the time of their diagnosis. Follow-up time for the subcohort participants without colorectal cancer was determined from the date of blood draw to the date of death or the end of follow-up (December 30, 2003), whichever came first. As standard practice for a case-cohort analysis, colorectal cancer cases within the subcohort were given two entries. Follow-up time for the first entry was estimated from the date of blood draw to the date of colorectal cancer diagnosis, and a negligible amount of follow-up time at the time of diagnosis was given for the second entry. All colorectal cases outside the subcohort were assigned a negligible amount of follow-time at the time of their diagnosis. Incidence rate ratios (RR) and 95% confidence intervals (95% CI) for colorectal cancer were estimated using a modified Cox regression model (22) with a robust variance estimate (23) to account for the fact that cases outside the subcohort appear only at the time of the event.

The RRs for colorectal cancer and individual polymorphisms were adjusted for age as a continuous variable and race (Caucasian or other). Adjustment for age as a categorical

variable (<60, 60-69, and 70+ years) and sex yielded comparable results. Additional adjustment for body mass index and education did not significantly alter the RRs, and the results were similar when the analysis was restricted to Caucasians or limited to individuals ≥35 years of age at baseline. In sensitivity analyses, exclusion of cases and person-years <50 years of age did not significantly alter the results. Test for trend was assessed by including a single variable for genotype coded as the number of variant alleles in the regression model.

Interactions between the polymorphisms and other covariates were assessed by including the cross-product terms as well as the main effect terms in a Cox regression model. For continuous exposures, categories were determined based on set cutpoints (e.g., thirds for red meat intake), and the categorical variables were used in evaluating possible interactions. The statistical significance of the interaction was evaluated by comparing nested models with and without the cross-product terms using a likelihood ratio test.

Measures of linkage disequilibrium and haplotype analyses were conducted among Caucasians. For polymorphisms located within the same gene or chromosomal region, pairwise linkage disequilibrium measures (*D'* and *r*<sup>2</sup>) were examined using Haploview.<sup>4</sup> Haplotype analyses were conducted using HaploStats.<sup>5</sup> Haplotypes were estimated using an expectation-maximization algorithm (24); overall differences between the cases and noncases were assessed using a global score test (25); and the risks for individual haplotypes were estimated using a generalized linear model (25, 26).

The *P* value from the robust sandwich estimate was used to assess statistical significance. However, to address the issue of multiple testing, *P* values for the associations between individual polymorphisms and colorectal cancer risk were also adjusted for the false discovery rate using the Benjamini and Hochberg method (27) with the software package R 2.0.1. Only the *P* values for the main effects of the polymorphisms and colorectal cancer risk were included in the calculation. Any *P* values adjusted using the Benjamini and Hochberg method are indicated as such. Unless otherwise specified, statistical analyses were done using SAS version 8.1 (SAS Institute, Inc., Cary, NC).

## Results

Baseline characteristics of the subcohort participants were similar to those of the full cohort (Table 1). Slightly more women than men participated in the cohort and subcohort, and the majority of participants were Caucasian. The mean follow-up time for the subcohort was 13.5 ± 2.7 years, which was similar to the full cohort. A total of 250 participants with DNA were diagnosed with colorectal cancer, and the average age at diagnosis was 70.6 ± 10.7 years. As expected, older age was associated with an increased risk of colorectal cancer (RR, 1.07 per year increase; 95% CI, 1.06-1.08), but no significant-difference was observed by sex (RR, 0.86; 95% CI, 0.66-1.11). Although not statistically significant, a slightly higher risk of colorectal cancer was observed among participants who were overweight or obese (body mass index ≥25 kg/m<sup>2</sup>: 1.18; 95% CI, 0.88-1.56) and ever smokers (RR, 1.23; 95% CI, 0.91-1.66) after adjusting for age and sex. Individuals with more than a high school education had a borderline significant lower risk of colorectal cancer than participants with less than a high school education (RR, 0.70; 95% CI, 0.49-1.01).

In examining individual polymorphisms, genetic variation in *ERCC6* was found to be associated with an increased

<sup>4</sup> <http://www.broad.mit.edu/personal/jcbarret/haploview/>.

<sup>5</sup> <http://www.mayo.edu/hsr/people/schaid.html>.

**Table 2. Risk of colorectal cancer associated with ERCC6 genotype stratified by cancer location**

Polymorphism	No. PY	Colorectal cancer		Colon cancer		Proximal colon cancer		Distal colon cancer		Rectal cancer	
		No. cases	RR (95% CI)	No. cases	RR (95% CI)	No. cases	RR (95% CI)	No. cases	RR (95% CI)	No. cases	RR (95% CI)
<i>M1097V</i>											
MM	18,426	137	1.0	108	1.0	54	1.0	48	1.0	29	1.0
MV	8,680	80	1.37 (1.00-1.87)	63	1.37 (0.97-1.94)	34	1.47 (0.94-2.32)	23	1.13 (0.67-1.90)	17	1.35 (0.73-2.50)
VV	1,212	21	2.49 (1.42-4.36)	17	2.56 (1.39-4.69)	11	3.28 (1.59-6.77)	5	1.70 (0.64-4.53)	4	2.24 (0.76-6.63)
<i>P<sub>trend</sub></i>			0.001		0.002		0.002		0.35		0.13
<i>R1213G</i>											
RR	18,493	135	1.0	107	1.0	52	1.0	48	1.0	28	1.0
RG	8,815	80	1.36 (1.00-1.86)	63	1.36 (0.96-1.91)	37	1.63 (1.04-2.55)	20	0.96 (0.56-1.65)	17	1.37 (0.74-2.55)
GG	1,211	22	2.64 (1.53-4.58)	19	2.88 (1.60-5.17)	12	3.72 (1.84-7.53)	6	2.03 (0.82-5.04)	3	1.73 (0.51-5.87)
<i>P<sub>trend</sub></i>			0.0006		0.0009		0.0003		0.42		0.22
<i>R1230P</i>											
R	23,372	205	1.0	164	1.0	89	1.0	64	1.0	41	1.0
RP	5,632	44	0.84 (0.58-1.21)	33	0.78 (0.52-1.18)	17	0.74 (0.43-1.28)	13	0.80 (0.43-1.49)	11	1.07 (0.54-2.10)
PP	330	0	—	0	—	0	—	0	—	0	—
<i>P<sub>trend</sub></i>			0.13		0.10		0.15		0.30		0.89

colorectal cancer risk (Table 2). Both the *ERCC6 1097V* and *ERCC6 1213G* alleles were associated with a dose-dependent increased risk of colorectal cancer with homozygotes having >2-fold increased risk of colorectal cancer compared with the wild type (RR, 2.49; 95% CI, 1.42-4.36;  $P = 0.001$  and RR, 2.64; 95% CI, 1.53-4.58;  $P = 0.0005$ , respectively). The association for both *ERCC6* variants was slightly stronger for cancers in the proximal colon compared with the distal colon. The risks were similar when the analysis was restricted to colorectal adenocarcinoma cases. Because the *ERCC6 M1097V* and *R1213G* polymorphisms were in strong linkage disequilibrium ( $D' = 1.0$ ) and almost completely correlated ( $r^2 = 0.99$ ), it was impossible to statistically differentiate the effects of one polymorphism from the other. Both polymorphisms were also in strong linkage disequilibrium ( $D' = 1.0$ ) but weakly correlated ( $r^2 = 0.03$ ) with the *ERCC6 R1230P* polymorphism, and adjustment for the *R1230P* polymorphism did not significantly affect the associations between *ERCC6 1097V* and *1213G* and colorectal cancer risk. In addition, the association between the *ERCC6 M1097V* and *R1213G* polymorphisms and colorectal cancer remained statistically signif-

icant after adjustment for multiple testing ( $P_{\text{trend}} = 0.01$  for both).

Two polymorphisms in *XPC* were associated with an altered risk of colorectal cancer (Table 3). Having at least one *XPC 492H* allele was associated with an increased risk of colorectal cancer compared with the homozygous wild type (RR, 1.75; 95% CI, 1.20-2.57,  $P = 0.004$ ), which seemed to be slightly stronger for the distal colon (RR, 2.57; 95% CI, 1.45-4.56) compared with the proximal colon (RR, 1.67; 95% CI, 0.96-2.91). Heterozygotes at *XPC R687R* were associated with a decreased risk of colorectal cancer compared with the wild type, but no association was observed with the homozygous variant. All four *XPC* polymorphisms were in strong linkage disequilibrium ( $D' = 0.88-1.0$ ) but weakly correlated ( $r^2 = 0.02-0.22$ ), and mutual adjustment had little effect on the observed risks for the *XPC 492H* allele with RRs ranging from 1.73 (95% CI, 1.18-2.55) to 1.98 (95% CI, 1.25-3.12) for the dominant model. When haplotype analyses were conducted, only the haplotype containing the *492H* allele was significantly associated with colorectal cancer risk (data not shown). No significant association was observed with the haplotype

**Table 3. Risk of colorectal cancer associated with XPC genotype stratified by cancer location**

Polymorphism	No. PY	Colorectal cancer		Colon cancer		Proximal colon cancer		Distal colon cancer		Rectal cancer	
		No. cases	RR (95% CI)	No. cases	RR (95% CI)	No. cases	RR (95% CI)	No. cases	RR (95% CI)	No. cases	RR (95% CI)
<i>R492H</i>											
RR	25,718	192	1.0	151	1.0	82	1.0	56	1.0	41	1.0
RH	3,455	40	1.71 (1.16-2.52)	33	1.80 (1.18-2.75)	16	1.60 (0.91-2.83)	17	2.48 (1.38-4.45)	7	1.37 (0.61-3.09)
HH	63	2	3.90 (0.61-25.1)*	2	5.01 (0.78-32.4)*	1	4.54 (0.44-46.9)*	1	6.77 (0.66-69.1)*	0	—
<i>P<sub>trend</sub></i>			0.003		0.002		0.06		0.0008		0.51
<i>A499V</i>											
AA	16,983	123	1.0	94	1.0	49	1.0	35	1.0	29	1.0
AV	10,515	85	1.13 (0.84-1.54)	71	1.24 (0.88-1.73)	42	1.41 (0.91-2.18)	26	1.21 (0.72-2.06)	14	0.79 (0.41-1.52)
VV	1,399	11	0.93 (0.46-1.86)	8	0.89 (0.40-1.96)	1	0.22 (0.03-1.61)	7	2.09 (0.87-5.06)	3	1.07 (0.31-3.67)
<i>P<sub>trend</sub></i>			0.68		0.50		0.86		0.14		0.71
<i>R687R</i>											
GG	15,784	154	1.0	127	1.0	72	1.0	49	1.0	27	1.0
AG	11,327	71	0.63 (0.47-0.86)	51	0.55 (0.39-0.78)	25	0.47 (0.30-0.76)	20	0.56 (0.33-0.96)	20	1.03 (0.57-1.85)
AA	2,096	21	1.06 (0.63-1.77)	16	0.98 (0.55-1.74)	7	0.75 (0.33-1.68)	7	1.12 (0.49-2.57)	5	1.43 (0.54-3.81)
<i>P<sub>trend</sub></i>			0.13		0.05		0.02		0.31		0.60
<i>K939Q</i>											
KK	9,791	81	1.0	62	1.0	27	1.0	30	1.0	19	1.0
KQ	14,277	110	0.94 (0.68-1.29)	88	0.98 (0.69-1.40)	51	1.30 (0.80-2.12)	30	0.69 (0.41-1.17)	22	0.80 (0.43-1.49)
KQ	5,168	53	1.19 (0.80-1.76)	43	1.25 (0.81-1.93)	26	1.74 (0.98-3.08)	15	0.91 (0.48-1.75)	10	0.97 (0.44-2.12)
<i>P<sub>trend</sub></i>			0.50		0.38		0.06		0.58		0.82

\*Because the cell counts are small, the normal approximation cannot be used. The true confidence intervals are slightly wider than the confidence intervals presented above.

**Table 4. Risk of colorectal cancer associated with thirds of red meat intake and ERCC6 genotype**

	Red meat intake (g/d)			$P_{\text{interaction}}^*$
	<44.0	44.0-86.3	≥86.3	
All participants	1.0	1.16 (0.80-1.70)	1.32 (0.86-2.02)	—
<i>M1097V</i>				
<i>MM</i>	1.0	1.05 (0.64-1.72)	1.18 (0.69-2.03)	
<i>MV or VV</i>	1.36 (0.79-2.34)	1.67 (0.96-2.87)	2.27 (1.29-3.99)	0.33
<i>R1213G</i>				
<i>RR</i>	1.0	1.01 (0.61-1.65)	1.10 (0.64-1.89)	
<i>RG or GG</i>	1.13 (0.65-1.97)	1.68 (0.99-2.86)	2.04 (1.16-3.60)	0.20

NOTE: RRs are based on the 202 cases and 1,583 subcohort participants with information on diet and are adjusted for age, race, and energy.

\* $P$  for interaction estimated using a likelihood ratio test and assuming linearity for thirds of red meat intake.

containing the *XPC R687R* variant. After adjustment for multiple testing, the association between the *XPC 492H* variant and colorectal cancer remained statistically significant ( $P_{\text{trend}} = 0.02$ ), but the association with *XPC R687R* was only borderline statistically significant ( $P = 0.05$ ).

We hypothesized that the effects of the *ERCC6* and *XPC* variants on cancer risk may differ between persons with and without a family history of colorectal cancer. Information about family history of cancer was not collected until 1996 and available only for ~65% of the participants. The risk of colorectal cancer associated the *ERCC6 1097V* and *1213G* variants was slightly greater among persons who had a first-degree relative with colorectal cancer (RR, 2.63; 95% CI, 0.94-7.39 and RR, 2.35; 95% CI, 0.87-6.37 for at least one *1097V* or *1213G* variant allele, respectively) compared with those without a family history (RR, 1.49; 95% CI, 1.02-2.19 and RR, 1.48; 95% CI, 1.01-2.17 for *1097V* and *1213G*, respectively), but the interactions were not statistically significant ( $P = 0.23$  and  $P = 0.30$ , respectively). For the association between the *XPC 492H* variant and colorectal cancer, little difference in risk was observed between participants with or without a family history of colorectal cancer (RR, 1.83; 95% CI, 0.49-6.78 and RR, 1.63; 95% CI, 0.99-2.69, respectively).

Because NER enzymes correct damage caused by exposure to tobacco smoke or cooked meat, we hypothesized that variants in *ERCC6* or *XPC* may modify the risk of colorectal cancer associated with smoking or meat consumption. No significant differences in the risk of colorectal cancer were observed by smoking status for the *ERCC6 1097V* allele (RR, 1.74; 95% CI, 1.12-2.71 and RR, 1.35; 95% CI, 0.91-2.00 for never and ever smokers, respectively), the *ERCC6 1213G* allele (RR, 1.85; 95% CI, 1.19-2.88 and RR, 1.29; 95% CI, 0.88-1.91, respectively), or *XPC 492H* variant (RR, 1.97; 95% CI, 1.10-3.51 and RR, 1.54; 95% CI, 0.92-2.58, respectively). Although the interaction was not statistically significant, the risk of colorectal cancer associated with the *ERCC6 1097V* and *1213G* alleles was greater among those in the upper third for red meat intake (Table 4), and there was a slight trend toward increasing risk with greater red meat intake among *ERCC6* variant carriers ( $P_{\text{trend}} = 0.22$  for *ERCC6 1097V* carriers and  $P_{\text{trend}} = 0.19$  for *ERCC6 1213G* carriers after adjustment for energy).

**Table 5. Risk of colorectal cancer associated with the total number of risk alleles for ERCC6 R1213G and XPC R492H**

No. risk alleles*	No. cases	No. PY	RR <sup>†</sup> (95% CI)	$P$
0	111	16,155	1.0	
1	77	10,014	1.21 (0.88-1.68)	0.24
2	34	2,123	2.57 (1.64-4.03)	0.00004
3+	5	150	8.48 (2.67-26.98)	0.0003
$P_{\text{trend}}$				0.00003

\*Risk alleles are *ERCC6 1213G* and *XPC 492H*.

† Adjusted for age (continuous) and race.

The risk of colorectal cancer for the *XPC 492H* variant was not modified by red meat intake.

We speculated that the combined effects of the *ERCC6* and *XPC* variants may be associated with a substantial elevated risk of colorectal cancer because the enzymes encoded by these genes are specific to two distinct NER subpathways, which repair DNA damage from different parts of the genome. When the combined effects of the *ERCC6 1213G* and *XPC 492H* variants were examined, the risk of colorectal cancer significantly increased with increasing number of variant alleles ( $P_{\text{trend}} = 0.00003$ , Table 5). Similar results were observed for the joint effects of *ERCC6 M1097V* and *XPC R492H*.

No associations were observed between the other polymorphisms examined in this study and colorectal cancer risk (Supplementary Table A). In addition, no significant differences were observed between *ERCC6 R1213G*, *ERCC6 M1097V*, or *XPC R492H* and colorectal cancer risk by age at diagnosis, sex, folate intake, or body mass index.

## Discussion

In this prospective study, we found that genetic variants in two NER genes, *ERCC6* and *XPC*, were associated with an altered risk of colorectal cancer. These findings suggest that common genetic polymorphisms in DNA repair genes may influence susceptibility to colorectal cancer and that individuals with *ERCC6 1213G* or *XPC 492H* variants may have an increased risk of colorectal cancer.

*ERCC6* encodes the CSB protein, which is necessary for transcription-coupled repair. The precise role of CSB in transcription-coupled NER is still unclear; however, it may help initiate the repair process. The stalled RNA polymerase II is believed to act as a steric hindrance to the repair of damage on the actively transcribed gene, and CSB may be needed to displace or remove the stalled RNA polymerase II from the damage site to allow repair to occur (28). Recently, CSB has been shown to possess chromatin remodeling activities (29), suggesting that it may alter the chromatin structure at the damaged site to facilitate DNA repair. CSB may also recruit NER repair factors to the DNA damage site, stimulating the repair process (28). *CSB*<sup>-/-</sup> mice that have been exposed to UV radiation display an elevated risk of skin cancer (30), suggesting CSB deficiencies may increase the risk of carcinogenesis.

In our study, we found that individuals with the *ERCC6 1097V* or *1213G* variant had an increased risk of colorectal cancer. Because the *ERCC6 M1097V* and *R1213G* polymorphisms were in strong linkage disequilibrium and highly correlated in our study population, we could not differentiate the effects of the two polymorphisms on colorectal cancer risk in our study. However, the increased risk observed with *ERCC6* was likely due to the *1213G* variant. Unlike the *M1097V* polymorphism, the *1213G* variant encodes a nonconservative amino acid change, and the *R1213* allele is conserved

in the mouse genome, suggesting the amino acid may be important for the protein. In addition, *in vitro* evidence shows that the 1213G variant alters the function of the protein (21). After transfection with the *ERCC6* 1213G variant and exposure to UV irradiation, UV-sensitive hamster cells display decreased survival compared with the wild type (21), indicating that the 1213G variant leads to a reduction in the ability of cells to repair DNA damage. The decreased ability to repair DNA damage may increase the risk of colorectal cancer by allowing more somatic DNA mutations or alterations to occur.

Although we did not find a statistically significant interaction between *ERCC6* genotype and smoking or total meat intake, the risk of colorectal cancer associated with the 1213G allele was slightly greater among participants in the upper third of red meat intake compared with the lower third. Red meat cooked at high temperatures contains carcinogens, such as heterocyclic amines and polycyclic aromatic hydrocarbons, which create bulky DNA adducts. These bulky adducts are primarily repaired by NER, and deficiencies in repair could contribute to their carcinogenic potential. Our results suggest that persons with the *ERCC6* 1213G variant may have a greater risk of colorectal cancer associated with high red meat intake. However, we did not have information on cooking practices, which could have given us a better estimate of the carcinogenic exposure from red meat, and our power to examine the interaction between *ERCC6* genotype and red meat intake was limited.

Only one other study has examined the association between genetic variation in *ERCC6* and the risk of colorectal neoplasia. In a case-control study within the Prostate, Lung, Colorectal, and Ovarian Cancer Screening Trial, Huang et al. (31) did not observe a significant association between colorectal adenoma and the *ERCC6* 1097V variant (odds ratio, 1.2; 95% CI, 0.7-2.2 for *VV* versus *MM*). However, this study only examined the risk for distal colon and rectal adenomas, and the association that we observed for *ERCC6* 1097V was more pronounced for proximal colon cancers. It is also possible that genetic variation in *ERCC6* plays a larger role in the progression of colorectal adenoma to cancer than the formation of colorectal adenomas.

We also observed an association between the *XPC* 492H variant and colorectal cancer risk. *XPC* is required for global genome NER and plays an important role in damage recognition and the initiation of repair. *XPC* binds to hHR23B, forming a complex that attaches to the DNA lesion and recruits other repair factors, such as TFIIH and XPG, to the damaged site (32, 33). *XPC* is thought to be the limiting subunit of the *XPC*-hHR23B complex, as hHR23B can be replaced by hHR23A *in vitro* (34) and both hHR23B and hHR23A are more abundant than *XPC* *in vivo* (35). Inherited mutations in *XPC* are associated with xeroderma pigmentosum, a rare autosomal recessive disease characterized by a very high predisposition to skin cancer, and cultured cells from these patients are highly sensitive to exposure from chemical carcinogens (36).

In our study, we found that the *XPC* 492H allele was associated with an increased risk of colorectal cancer. The functional effects of the *XPC* 492H variant have not been studied. The polymorphism encodes a conservative amino acid change and is not well conserved across species, suggesting that the association that we observed with the *XPC* 492H variant and colorectal cancer may be due to linkage disequilibrium with another polymorphism in *XPC*. Several other polymorphisms within *XPC* have been suggested to have functional significance. The A allele at the splice acceptor site of *XPC* intron 11 has been shown to lead to a greater proportion of alternatively spliced mRNA with a deletion of exon 12 and decreased DNA repair activity (37). A biallelic poly(AT) insertion/deletion polymorphism (PAT) in intron 9 of *XPC* has also been associated with reduced DNA repair (19) and increased anti-B[a]PDE-DNA adduct levels (38). However,

because this polymorphism is in strong linkage disequilibrium with the splice acceptor site polymorphism in intron 11 ( $D' = 1$  and  $r^2 = 0.96$  for Caucasians; ref. 37), the decreased DNA repair observed with the *PAT*<sup>+/+</sup> variant may simply be due to the correlation with the splice acceptor site variant. Several polymorphisms have also been reported in the putative transcription factor binding sites for *XPC* and hypothesized to alter transcription activity (39), but the functional effects of these variants have not been rigorously studied.

In a case-control study, Huang et al. (31) found that the *XPC* R492H polymorphism significantly modified the risk of colorectal adenoma associated with cigarette smoking, with nonsmoking 492H carriers displaying an increased risk of adenoma. We did not observe a significant interaction between *XPC* R492H and smoking in our study. However, smoking is a weaker risk factor for colorectal cancer than adenoma (6), and this could have contributed to the different results seen between the two studies. In addition, there is some evidence to suggest that smoking may only be associated with colorectal cancer if a sufficiently long induction period is considered (6). However, we were unable to differentiate long-term smokers from recent smokers in this study. Consistent with our study, no significant associations with colorectal neoplasia were observed for polymorphisms in *ERCC1*, *ERCC2*, *ERCC4*, and *ERCC5* in other studies (31, 40, 41).

Our prospective study suggests that common genetic polymorphisms in DNA repair genes influence susceptibility to colorectal cancer. Our study was community based and had the advantage of recruiting participants from a fairly homogeneous population. However, because our study consisted primarily of Caucasians, the results of the study may not be generalizable to other populations. Because our study was nested in a large cohort, information on several environmental factors, including diet, was obtained before cancer diagnosis and thus not subjected to recall bias. However, we did not have detailed information on meat cooking practices and had limited power to detect gene-environment interactions. Thus, the effect modification of *ERCC6* genotype on red meat consumption that we observed requires replication. As with all genetic association studies, it is possible that the relationships we observed between the *ERCC6* and *XPC* variants and colorectal cancer risk are due to chance. However, for both *ERCC6* R1213G and *XPC* R492H, there was a significant dose-dependent increase in risk for increasing number of variant alleles and the associations remained statistically significant even after adjustment for multiple testing, suggesting that a chance association is less likely.

In conclusion, we found that genetic variants in *ERCC6* and *XPC* were associated with an increased risk of colorectal cancer. This finding indicates that deficiencies in NER could play a role in the development of colorectal cancer. However, additional studies are needed to confirm the results of this study and to further characterize the effects of genetic variation in *XPC* and *ERCC6* on the ability of NER to repair DNA damage.

## Acknowledgments

We thank Judy Hoffman-Bolton and Alyce E. Burke for managing the data, Kenneth W. Kinzler for his expertise, and the participants in the CLUE II cohort for making this study possible.

## References

1. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 2000;343:78–85.
2. Johns LE, Houlston RS. A systematic review and meta-analysis of familial colorectal cancer risk. *Am J Gastroenterol* 2001;96:2992–3003.
3. Kinzler KW, Vogelstein B. Colorectal tumors. In: Vogelstein B, Kinzler KW, editors. *The genetic basis of human cancer*. New York: McGraw-Hill Companies, Inc.; 2002. p. 583–612.

4. Al Tassan N, Chmiel NH, Maynard J, et al. Inherited variants of MYH associated with somatic G:C->T:A mutations in colorectal tumors. *Nat Genet* 2002;30:227-32.
5. Jones S, Emmerson P, Maynard J, et al. Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G:C->T:A mutations. *Hum Mol Genet* 2002;11:2961-7.
6. Giovannucci E. An updated review of the epidemiological evidence that cigarette smoking increases risk of colorectal cancer. *Cancer Epidemiol Biomarkers Prev* 2001;10:725-31.
7. Sandhu MS, White IR, McPherson K. Systematic review of the prospective cohort studies on meat consumption and colorectal cancer risk: a meta-analytical approach. *Cancer Epidemiol Biomarkers Prev* 2001;10:439-46.
8. Sinha R, Chow WH, Kulldorff M, et al. Well-done, grilled red meat increases the risk of colorectal adenomas. *Cancer Res* 1999;59:4320-4.
9. Augustsson K, Skog K, Jagerstad M, Dickman PW, Steineck G. Dietary heterocyclic amines and cancer of the colon, rectum, bladder, and kidney: a population-based study. *Lancet* 1999;353:703-7.
10. Sugawara K, Ng JM, Masutani C, et al. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol Cell* 1998;2:223-32.
11. Volker M, Mone MJ, Karmakar P, et al. Sequential assembly of the nucleotide excision repair factors *in vivo*. *Mol Cell* 2001;8:213-24.
12. Mu D, Sancar A. Model for XPC-independent transcription-coupled repair of pyrimidine dimers in humans. *J Biol Chem* 1997;272:7570-3.
13. Evans E, Moggs JG, Hwang JR, Egly JM, Wood RD. Mechanism of open complex and dual incision formation by human nucleotide excision repair factors. *EMBO J* 1997;16:6559-73.
14. 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.
15. Sijbers AM, de Laat WL, Ariza RR, et al. Xeroderma pigmentosum group F caused by a defect in a structure specific DNA repair endonuclease. *Cell* 1996;86:811-22.
16. Block G, Hartman AM, Naughton D. A reduced dietary questionnaire: development and validation. *Epidemiology* 1990;1:58-64.
17. Klintschar M, Neuhuber F. Evaluation of an alkaline lysis method for the extraction of DNA from whole blood and forensic stains for STR analysis. *J Forensic Sci* 2000;45:669-73.
18. 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.
19. 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.
20. Seker H, Butkiewicz D, Bowman ED, et al. Functional significance of XPD polymorphic variants: attenuated apoptosis in human lymphoblastoid cells with the XPD 312 Asp/Asn genotype. *Cancer Res* 2001;61:7430-4.
21. Mallery DL, Tanganelli B, Colella S, et al. Molecular analysis of mutations in the CSB (ERCC6) gene in patients with Cockayne syndrome. *Am J Hum Genet* 1998;62:77-85.
22. Self SG, Prentice RL. Asymptotic distribution theory and efficiency results for case-cohort studies. *Ann Stat* 1988;16:64-81.
23. Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *J Am Stat Assoc* 1989;84:1074-8.
24. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;12:921-7.
25. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425-34.
26. Lake SL, Lyon H, Tantisira K, et al. Estimation and tests of haplotype-environment interaction when linkage phase is ambiguous. *Hum Hered* 2003;55:56-65.
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 1995;57:289-300.
28. Troelstra C, van Gool A, de Wit J, Vermeulen W, Bootsma D, Hoeijmakers JH. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. *Cell* 1992;71:939-53.
29. Citterio E, van den Boom V, Schnitzler G, et al. ATP-dependent chromatin remodeling by the Cockayne syndrome B DNA repair-transcription-coupling factor. *Mol Cell Biol* 2000;20:7643-53.
30. van der Horst GT, van Steeg H, Berg RJ, et al. Defective transcription-coupled repair in Cockayne syndrome B mice is associated with skin cancer predisposition. *Cell* 1997;89:425-35.
31. Huang WY, Berndt SI, Kang D, et al. Nucleotide excision repair gene polymorphisms and risk of advanced colorectal adenoma: XPC polymorphisms modify smoking-related risk. *Cancer Epidemiol Biomarkers Prev* 2006;15:306-11.
32. Yokoi M, Masutani C, Maekawa T, Sugawara K, Ohkuma Y, Hanaoka F. The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIIH to damaged DNA. *J Biol Chem* 2000;275:9870-5.
33. Araujo SJ, Nigg EA, Wood RD. Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome. *Mol Cell Biol* 2001;21:2281-91.
34. Sugawara K, Ng JM, Masutani C, et al. Two human homologs of Rad23 are functionally interchangeable in complex formation and stimulation of XPC repair activity. *Mol Cell Biol* 1997;17:6924-31.
35. van der Spek PJ, Eker A, Rademakers S, et al. XPC and human homologs of RAD23: intracellular localization and relationship to other nucleotide excision repair complexes. *Nucleic Acids Res* 1996;24:2551-9.
36. Bootsma D, Kraemer KH, Cleaver JE, Hoeijmakers JH. Nucleotide excision repair syndromes: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. In: Vogelstein B, Kinzler KW, editors. *The genetic basis of human cancer*. New York: McGraw-Hill Companies, Inc.; 2002. p. 211-37.
37. 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.
38. 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.
39. Lee GY, Jang JS, Lee SY, et al. XPC polymorphisms and lung cancer risk. *Int J Cancer* 2005;115:807-13.
40. Mort R, Mo L, McEwan C, Melton DW. Lack of involvement of nucleotide excision repair gene polymorphisms in colorectal cancer. *Br J Cancer* 2003;89:333-7.
41. Yeh CC, Sung FC, Tang R, Chang-Chieh CR, Hsieh LL. Polymorphisms of the XRCC1, XRCC3, & XPD genes, and colorectal cancer risk: a case-control study in Taiwan. *BMC Cancer* 2005;5:12.