

A Common 8q24 Variant and the Risk of Colon Cancer: A Population-Based Case-Control Study

Li Li,^{1,2,4} Sarah J. Plummer,⁵ Cheryl L. Thompson,^{1,2,4} Alona Merkulova,⁵ Louise S. Acheson,^{1,3} Thomas C. Tucker,⁶ and Graham Casey^{4,5}

Departments of ¹Family Medicine-Research Division, ²Epidemiology and Biostatistics, and ³Reproductive Biology, Case Western Reserve University/University Hospitals of Cleveland; ⁴Case Center for Transdisciplinary Research on Energetics and Cancer, Case Comprehensive Cancer Center; ⁵Department of Cancer Biology, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio; and ⁶Cancer Control Program, University of Kentucky, Lexington, Kentucky

Abstract

Three recent studies identified common variants on 8q24 that confer modestly increased susceptibility to colorectal cancer. Here, we replicate the association in a population-based case-control study of colon cancer, including 561 cases and 721 unrelated controls. The rs6983267 marker was significantly associated with colon cancer risk. Compared with those homozygous for the T allele, the heterozygous and homozygous carriers for the G allele had an age-adjusted odds ratio of 1.39 (95% confidence interval, 1.03-1.88) and 1.68 (95% confidence interval, 1.21-2.33), respectively. An additive model showed strong evidence for a gene-

dose response relationship ($P_{\text{trend}} = 0.0022$). The association remained statistically significant when restricted to Caucasians only (527 cases and 679 controls; $P_{\text{trend}} = 0.0056$). Further adjustment for other known risk factors did not alter the results. Stratified analysis revealed no evidence for effect modification by family history of colorectal cancer, age, or gender. These data replicate the association identified from recent studies, providing additional evidence supporting the rs6983267 genetic polymorphism as a marker predisposing to colon cancer. (Cancer Epidemiol Biomarkers Prev 2008;17(2):339-42)

Introduction

Recent studies have shown that several common genetic variants on chromosome 8q24 are associated with risk of prostate cancer (1-6). One of the identified variants (rs6983267) and a second tightly linked variant (rs10505477) have more recently been implicated in the risk of colorectal cancer and adenoma through genome-wide and focused association studies across several cohort and case-control study populations (4, 7, 8). To further confirm this association with colorectal cancer, we conducted a focused evaluation of the rs6983267 marker in a population-based case-control study of colon cancer.

Materials and Methods

Study Population. The study included 561 incident colon cancer cases and 721 population controls, recruited between July 2003 and March 2007. The population-based Surveillance, Epidemiology, and End Results Kentucky Cancer Registry identified all histopathologically confirmed primary colon cancer cases diagnosed in

Kentucky residents. We queried the cancer registry database every 3 months and identified all cases reported within 6 months of diagnosis preceding the recruitment. Rectal cancers were excluded.

We first sent an introductory letter explaining the study to potentially eligible cases. After 3 weeks, we phoned each subject for a screening interview to determine eligibility and their willingness to participate in the study. We then used a two-step approach to collect blood samples and lifestyle questionnaire data. First, we sent a prepacked phlebotomy kit with detailed written instructions for blood sample collection and written consent forms to each consenting case subject. The subjects were instructed to go to their physician offices or adjacent medical facilities for blood draw after overnight fasting. We then mailed each subject a self-administered lifestyle risk factor questionnaire developed by the National Cancer Institute Colon Family Cancer Registry⁷ to collect detailed information on lifestyle and behavioral risk factors, including demographics, comorbidities, family history of colorectal cancer, weight (2 years prior to recruitment), height, nonsteroidal anti-inflammatory medications (NSAID), smoking, alcohol consumption, and diet.

We used random digit dialing to recruit population controls following a protocol similar to that described above for cases. As a proxy for frequency matching the residential locations between the cases and the controls, the area codes and exchanges of the phone numbers of potential cases were used, along with randomly generated four-digit numbers, to produce the list of phone

Received 8/3/07; revised 10/10/07; accepted 11/21/07.

Grant support: Damon Runyon Cancer Research Foundation, Clinical Investigator Award (CI-8), Case Center for Transdisciplinary Research on Energetics and Cancer (U54 CA116867), and National Cancer Institute K22 Award (K22 CA120545).

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.

Requests for reprints: Li Li, Department of Family Medicine-Research Division, Case Western Reserve University, 11001 Cedar Avenue, Suite 306, Cleveland, OH 44106-7136. Phone: 216-368-5437. E-mail: lxl62@cwru.edu

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-0713

⁷ http://epi.grants.cancer.gov/CFR/about_questionnaires.html

numbers for control recruitment. Recruitment was not conducted when a business was reached. Respondents who were 40 years or older and free of personal history of cancer other than skin cancer were included as controls. We excluded those with known inflammatory bowel diseases, family history of familial adenomatous polyposis, and hereditary nonpolyposis colorectal cancer. As with case subjects, each control first donated a sample of blood after fasting overnight and then was asked to complete the self-administered risk factor questionnaire.

The participation rate was 72.2% for case subjects. Due to the nature of random digit dialing, the response rate for controls cannot be precisely assessed as we could not determine the eligibility of those subjects that were reached but refused to disclose personal information. Nevertheless, among those individuals that allowed us to explain the study and were determined to be eligible, the participation rate was 62.5%. The overall participation rate among potential controls reached by random digit dialing, regardless of their eligibility, was 33.3%. Eighty-six percent of enrolled cases and 88.8% of controls returned blood samples and completed the risk factor questionnaire. All participants provided written informed consent. The study was approved by the Institutional Review Boards of the University of Kentucky, Lexington, and Case Western Reserve University/University Hospitals of Cleveland.

DNA Extraction and Genotyping. Fasting blood samples were collected locally and shipped overnight on frozen ice packs. Upon receipt, aliquots of plasma and buffy coat were prepared and frozen at -80°C . One aliquot of buffy coat was thawed and genomic DNA extraction was done on a Biorobot EZ1 (Qiagen). The genomic DNA was quantitated using the Quant-it picogreen kit (Invitrogen). After normalizing, the samples were arrayed in 96-well plates at a final concentration of 0.63 ng/ μL .

The rs6983267 variant in the 8q24 chromosomal region was genotyped using the TaqMan allelic discrimination assay with a custom primer/probe set (Applied Biosystems). The genotyping failure rate was $<0.1\%$. Two percent of the samples were repeated for quality control, and had a concordance rate of 100%. Laboratory personnel were blinded to case-control status.

Statistical Methods. We evaluated the association of the rs6983267 genotype with colon cancer in unconditional logistic regression models. The G allele of this single nucleotide polymorphism (SNP) was treated as the risk variant. Four possible genetic models for risk were evaluated by adding dummy variables (Z) into the logistic regressions: unconstrained genotype effect ($Z_1 = 0$ and $Z_2 = 0$ if TT, $Z_1 = 1$ and $Z_2 = 0$ if GT, $Z_1 = 0$ and $Z_2 = 1$ if GG), dominant ($Z = 1$ if carrying at least one G allele, 0 otherwise), additive ($Z = 2$ if carrying two copies of the G allele, 1 if carrying one G allele, 0 otherwise), or recessive ($Z = 1$ if carrying two copies of the G allele, 0 otherwise). We first estimated the crude and age-adjusted odds ratios (OR). Then we further controlled for race, sex, family history of colorectal cancer, body mass index (BMI), and nonsteroidal anti-inflammatory drug (NSAID) use. We further explored potential effect modification by stratifying analyses by age (≤ 65 or >65 years), sex, and family history of colorectal cancer. We also addressed potential population substructures by limiting our analyses to Caucasians (self-reported),

which includes 527 cases and 679 controls. All P values were two-sided, and all analyses were undertaken with SAS software (version 9.1; SAS Institute, Inc.).

Results

The characteristics of the participants are shown in Table 1. The rs6983267 SNP was statistically significantly associated with increased colon cancer risk (Table 2). Compared with subjects homozygous for the T allele, heterozygotes and carriers homozygous for the G allele had an age-adjusted OR of 1.39 [95% confidence interval (95% CI), 1.03-1.88] and 1.68 (95% CI, 1.21-2.33), respectively. An additive model showed strong evidence for a gene-dose response relationship ($P_{\text{trend}} = 0.0022$). The association remained statistically significant when restricted to Caucasians only ($P_{\text{trend}} = 0.0056$). Results from the dominant and recessive models had similar patterns (data not shown). Adjusting for other known risk factors in the full logistic regression model did not materially change the association of rs6983267 with colon cancer: the OR for heterozygotes was 1.45 (95% CI, 1.05-2.00), and 1.69 (95% CI, 1.19-2.40) for those carrying two copies of the risk allele ($P_{\text{trend}} = 0.0095$). Stratified analyses revealed no evidence for effect modification by age, gender, or family history of colorectal cancer (Table 3).

Discussion

This population-based case-control study replicates the association identified from the genome-wide association

Table 1. Kentucky Colon Cancer Study participant characteristics

	Controls (<i>n</i> = 721)	Cases (<i>n</i> = 561)
Mean age (SD)	59.0 (10.9)	64.1 (10.8)
Sex (%)		
Female	442 (61.3)	287 (51.2)
Male	279 (38.7)	274 (48.8)
Race (%)		
African American	23 (3.2)	25 (4.5)
Caucasian	679 (94.2)	527 (94.1)
Other	19 (2.6)	8 (1.4)
Regular NSAID use (%)*		
Regular	445 (67.4)	309 (62.3)
Never	215 (29.8)	188 (33.5)
Missing	61 (8.5)	64 (11.4)
Mean body mass index (kg/m ² ; SD) [†]	28.1 (6.2)	29.3 (6.1)
Family history of colorectal cancer (%) [‡]		
Yes	166 (23.0)	177 (31.6)
No	496 (68.8)	330 (58.8)
Missing	59 (8.2)	54 (9.6)
rs6983267 (%)		
TT	175 (24.3)	103 (18.4)
GT	348 (48.3)	272 (48.5)
GG	198 (27.5)	186 (33.2)

*Regular NSAID use was defined as reporting ever use of aspirin or NSAID at least twice a week for at least 1 mo.

[†]Lifestyle variables were obtained from the RFQ. Body mass index was calculated based on self-reported current weight and height (kg/m²). Seventy cases and sixty-nine controls were missing on weight or height.

[‡]A positive family history of colorectal cancer was defined as self-report of colorectal cancer in ≥ 1 first-degree relative.

Table 2. Logistic regression analysis of rs6983267 and colon cancer

	Genotype	Cases (%)	Controls (%)	Crude		Age-adjusted*		Full model	
				OR (95% CI) [†]	P [‡]	OR (95% CI)*	P [‡]	OR (95% CI) [§]	P [‡]
Entire sample	GG	186 (33.2)	198 (27.5)	1.60 (1.17-2.19)	4.0 × 10 ⁻³	1.68 (1.21-2.33)	2.2 × 10 ⁻³	1.69 (1.19-2.40)	9.5 × 10 ⁻³
	GT	272 (48.5)	348 (48.3)	1.33 (0.99-1.78)		1.39 (1.03-1.88)		1.45 (1.05-2.00)	
	TT	103 (18.4)	175 (24.3)	1.0 (referent)		1.0 (referent)		1.0 (referent)	
Caucasians	GG	165 (31.3)	177 (26.1)	1.55 (1.12-2.14)	8.6 × 10 ⁻³	1.62 (1.16-2.26)	5.6 × 10 ⁻³	1.61 (1.36-2.30)	0.013
	GT	259 (49.2)	331 (48.8)	1.30 (0.97-1.74)		1.35 (1.00-1.83)		1.41 (1.02-1.94)	
	TT	103 (19.5)	171 (25.2)	1.0 (referent)		1.0 (referent)		1.0 (referent)	

*ORs and 95% CI estimates for unconstrained genotype effect adjusted for age.

[†]Crude OR and 95% CI estimates for unconstrained genotype effect.

[‡]P for trend from additive model.

[§]OR and 95% CI estimates for unconstrained genotype effect adjusted for age, race, gender, family history, body mass index, and NSAID use (479 cases and 640 controls with data available).

studies, providing additional evidence that the rs6983267 genetic polymorphism is a marker for predisposing to colon cancer. The age-adjusted additive OR of 1.29 (95% CI, 1.10-1.51) for each copy of the risk allele, observed among predominantly Caucasian Kentucky residents in this study, was similar to the ORs derived from additive models in the three genome-wide association study populations [1.22, 1.27, and 1.17; refs. 4, 7, 8, respectively]. The study by Zanke et al. used a different SNP (rs10505477), which is in tight linkage disequilibrium with the rs6983267 marker (8). We pooled the crude results for the rs6983267 SNP from our study with the results from the previous studies of colorectal cancer in a meta-analysis to get a combined OR estimate of 1.28 (95% CI, 1.21-1.36; refs. 4, 7). Because 75% of control

participants carried at least one copy of the risk allele, the population-attributable risk of this polymorphism is likely to be substantial, and was estimated at 15.3% in our sample, whereas the predictive value of genotype for an individual is low.

The rs6983267 SNP marks a region of 8q24 with few known or predicted genes (6), although this region is frequently amplified in many tumor types, including colorectal cancer (9, 10). This variant is located close to the putative pseudogene *POU5F1P1*, a retrotransposed transcription factor expressed in colon cancer cell lines (7). It is also located >300 kb upstream of the *MYC* oncogene, which is overexpressed in many colon tumors. However, to date, no relationship has been found between *MYC* protein expression, *POU5F1P1/MYC* amplification, or

Table 3. Stratified logistic regression analysis of rs6983267 and colon cancer

	Genotype	Cases (%)	Controls (%)	Crude		Age-adjusted		Full model	
				OR (95% CI)*	P	OR (95% CI) [†]	P	OR (95% CI)	P
Age ≤65 y	GG	96 (32.1)	144 (27.5)	1.56 (1.03-2.36)	0.040	1.63 (1.07-2.49)	0.028	1.70 (1.05-2.73)	0.034
	GT	150 (50.2)	256 (48.9)	1.37 (0.94-2.00)		1.40 (0.94-2.06)		1.45 (0.94-2.24)	
	TT	53 (17.7)	124 (23.7)	1.0 (referent)		1.0 (referent)		1.0 (referent)	
Age >65 y	GG	90 (34.4)	54 (27.3)	1.70 (1.02-2.85)	0.044	1.77 (1.37-2.30)	0.031	1.68 (0.95-2.98)	0.10
	GT	122 (46.6)	93 (47.0)	1.34 (0.83-2.15)		1.33 (1.03-1.73)		1.58 (0.93-2.68)	
	TT	50 (19.1)	51 (25.8)	1.0 (referent)		1.0 (referent)		1.0 (referent)	
<i>P</i> _{interaction} Males	GG	92 (33.6)	76 (27.2)	1.47 (0.91-2.36)	0.10	1.43 (0.87-2.35)	0.14	1.48 (0.86-2.57) [§]	0.16
	GT	130 (47.5)	140 (50.2)	1.13 (0.73-1.74)		1.15 (0.73-1.82)		1.23 (0.74-2.02)	
	TT	52 (19.0)	63 (22.6)	1.0 (referent)		1.0 (referent)		1.0 (referent)	
Females	GG	94 (32.8)	122 (27.6)	1.69 (1.10-2.59)	0.020	1.89 (1.21-2.95)	6.5 × 10 ⁻³	1.80 (1.10-2.95)	0.028
	GT	142 (49.5)	208 (47.1)	1.50 (1.01-2.22)		1.60 (1.06-2.41)		1.72 (1.10-2.70)	
	TT	51 (17.8)	112 (25.3)	1.0 (referent)		1.0 (referent)		1.0 (referent)	
<i>P</i> _{interaction} +Family history	GG	57 (32.2)	44 (26.5)	1.51 (0.83-2.74)	0.17	1.53 (0.83-2.80)	0.17	1.69 (0.89-3.21)	0.11
	GT	84 (47.5)	80 (48.2)	1.23 (0.71-2.10)		1.24 (0.71-2.15)		1.34 (0.76-2.41)	
	TT	36 (20.3)	42 (25.3)	1.0 (referent)		1.0 (referent)		1.0 (referent)	
-Family history	GG	114 (34.7)	137 (27.6)	1.85 (1.23-2.79)	3.6 × 10 ⁻³	1.88 (1.23-2.89)	4.7 × 10 ⁻³	1.64 (1.05-2.56)	0.038
	GT	162 (49.2)	241 (48.6)	1.50 (1.02-2.19)		1.56 (1.05-2.32)		1.50 (1.00-2.25)	
	TT	53 (16.1)	118 (23.8)	1.0 (referent)		1.0 (referent)		1.0 (referent)	
<i>P</i> _{interaction}						0.81 [‡]		0.96 [‡]	

NOTE: OR and 95% CI estimates and *P* value for unconstrained genotype effect adjusted for age, race, gender, family history of colorectal cancer, body mass index, and NSAID use (479 cases and 640 controls with data available).

*Crude OR and 95% CI estimates for unconstrained genotype effect.

[†]OR and 95% CI estimates for unconstrained genotype effect adjusted for age.

[‡]*P* for interaction between variable and genotype via a likelihood ratio test comparing the models with and without the interaction terms.

[§]OR and 95% CI estimates and *P* value for unconstrained genotype effect adjusted for age, race, family history of colorectal cancer, body mass index, and NSAID use (479 cases and 640 controls with data available).

^{||}OR and 95% CI estimates and *P* value for unconstrained genotype effect adjusted for age, race, gender, body mass index, and NSAID use (479 cases and 640 controls with data available).

POU5F1P1 RNA expression and genotypes of rs6983267, or the linked marker rs10505477 (7, 8).

Although the overall risk effect of this SNP is modest, Tomlinson et al. found a similar association of rs6983267 with colon adenomas, suggesting an effect early in the development of colon neoplasia (7). The discovery of a variant associated with two common but distinct types of cancer, possibly functioning at the early stages of cancer development, along with the high frequency of risk allele carriers, indicates the importance of further research into the exact mechanism of this association.

References

1. Amundadottir LT, Sulem P, Gudmundsson J, et al. A common variant associated with prostate cancer in European and African population. *Nat Genet* 2006;38:652–58.
2. Freedman ML, Haiman CA, Patterson N, et al. Admixture mapping identified 8q24 as a prostate cancer risk locus in African-American men. *Proc Natl Acad Sci U S A* 2006;103:14068–73.
3. Gudmundsson J, Sulem P, Manolescu A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 2007;39:631–7.
4. Haiman CA, Le Marchand L, Yanamoto J, et al. A common genetic risk factor for colorectal and prostate cancer. *Nat Genet* 2007;39:954–6.
5. Yeager M, Orr N, Hayes RB, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. *Nat Genet* 2007;39:645–9.
6. Schumacher FR, Feigelson HS, Cox DG, et al. A common 8q24 variant in prostate and breast cancer from a large nested case-control study. *Cancer Res* 2007;67:2951–5.
7. Tomlinson I, Webb E, Carvajal-Carmona L, et al. A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 2007;39:984–8.
8. Zanke BW, Greenwood CMT, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 2007;39:989–94.
9. Ried T, Knutzen R, Steinbeck R, et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. *Genes Chromosomes Cancer* 1996;15:234–45.
10. Douglas EJ, Fiegler H, Rowan A, et al. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res* 2004;64:4817–25.