

Haplotype Analysis of Common Vitamin D Receptor Variants and Colon and Rectal Cancers

Carol Sweeney,¹ Karen Curtin,¹ Maureen A. Murtaugh,¹ Bette J. Caan,²
John D. Potter,³ and Martha L. Slattery¹

¹Health Research Center, Department of Family and Preventive Medicine, University of Utah, Salt Lake City, Utah; ²Kaiser Permanente Medical Research Program, Oakland, California; and ³Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, Washington

Abstract

Inherited variants of the vitamin D receptor (*VDR*) gene may influence cancer risk by altering the effect of vitamin D on cell growth and homeostasis. Studies have examined genotypes for common *VDR* polymorphisms, including a single nucleotide polymorphism (SNP) detected by *Bsm1*, a polyadenosine [poly(A)] repeat polymorphism, and a SNP detected by *Fok1*, as candidates for susceptibility to cancer, but most have not evaluated haplotypes for these markers. We investigated haplotypes for these polymorphisms in case-control studies of colon cancer (1,811 cases and 1,451 controls) and rectal cancer (905 cases and 679 controls). We used the expectation-maximization algorithm to estimate haplotypes for White, Hispanic, African-American, and Asian subjects, tested for differences in *VDR* haplotype distribution, and calculated odds ratios (OR) for association between haplotype and cancer. The distribution of hap-

lotypes differed by race or ethnic group, but four common haplotypes accounted for the majority of alleles in all groups. *VDR* haplotype distributions differed between colon cancer cases and controls ($P = 0.0004$). The common haplotype bLF, containing *Bsm1* b (*Bsm1* restriction site present), poly(A) long (18-22 repeats), and *Fok1* F (restriction site absent) was associated with increased risk of colon cancer, OR 1.15 (95% confidence interval, 1.03-1.28), as was the rare haplotype BLF, containing *Bsm1* B (restriction site absent), poly(A) long, and *Fok1* F (OR, 2.40; 95% confidence interval, 1.43-4.02). No case-control differences were detected for rectal cancer. In this analysis, haplotypes of the *VDR* influenced risk of colon cancer, but haplotype variables had only slightly better ability to explain case-control differences than genotype variables. (Cancer Epidemiol Biomarkers Prev 2006;15(4):744-9)

Introduction

The vitamin D receptor (*VDR*) mediates signals from circulating vitamin D to the cell nucleus, affecting transcription of a number of genes, including those encoding the insulin-like growth factor-binding proteins (1-3), which modulate insulin-like growth factor effects on cell growth. Thus, the antiproliferative effect of vitamin D on cell cultures, and its hypothesized protective effect against cancer, is thought to occur through binding to the *VDR* (4).

Several common polymorphisms of the *VDR* gene have been studied as possible genetic contributors to cancer susceptibility. A group of polymorphisms are found near the 3' end of the of the *VDR*, including a single nucleotide polymorphism (SNP) in an intron between exons 8 and 9 detected by the *Bsm1* restriction enzyme and a polyadenosine [poly(A)] length polymorphism in the 3' untranslated region (5). These polymorphisms are in extensive linkage disequilibrium (6). Although these specific DNA sequence changes do not affect *VDR* function (7), epidemiologic studies, reviewed by Thakkinstian (8), show that these *VDR* variants influence bone density, providing evidence that the 3' region polymorphisms are markers for functionally different alleles. A SNP detected by the *Fok1* restriction enzyme is recognized as functional, altering a translation initiation site (9) and affecting bone mineral density (10). The *Fok1* polymorphism has been reported to exhibit little linkage disequilibrium with other *VDR* polymorphisms (11).

The potential associations between risk of colorectal cancers or adenomas and the *VDR* 3' region polymorphisms and/or the *Fok1* polymorphism have been investigated in several studies (12-22). The implications of the evidence from these studies are not completely clear, as some studies detected associations with the 3' region polymorphisms and others report associations with the *Fok1* polymorphism. Some of the published studies have considered only one of the known common variants, whereas others evaluate genotype for two or more. With the exception of one small study (23), studies of these *VDR* variants in colorectal cancer risk have not applied haplotype-based analysis. The functional effect of genetic variation on gene expression, protein structure, or enzyme activity is expected to be better described by a summary of all genetic variants present on an allele (i.e., a haplotype) than by the presence or absence of individual polymorphisms (24, 25).

In the present study, we constructed *VDR* haplotypes from genotype data determined for three *VDR* polymorphisms: the intronic SNP detected by the *Bsm1* restriction enzyme, the 3' untranslated region poly(A) repeat length polymorphism, and the *Fok1* SNP. Using data from population-based, case-control studies of colon and rectal cancer, we estimated haplotype frequencies for four race or ethnic groups, and evaluated associations between haplotype and cancer risk. Further, we compared *VDR* haplotype variables to *VDR* genotype variables for their ability to explain case-control differences.

Materials and Methods

Study Subjects. We report on participants in collaborative population-based case-control studies of cancers of the colon and rectum. The colon cancer study was conducted in eight counties in Utah, the Kaiser Permanente Medical Care Program in northern California, and seven counties in Minnesota. Individuals with diagnoses of colon cancer

Received 10/18/05; revised 1/30/06; accepted 2/9/06.

Grant support: NIH grants CA48998 and CA85846.

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: Carol Sweeney, Health Research Center, Department of Family and Preventive Medicine, University of Utah, 375 Chipeta Way, Suite A, Salt Lake City, UT 84108. Phone: 801-581-5865; Fax: 801-581-3623. E-mail: Carol.Sweeney@hrc.utah.edu

Copyright © 2006 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-05-0814

(ICD02 18.0 and 18.2-18.9) between October 1991 and September 1994 were identified as cases for the study through the Utah Cancer Registry, the Northern California Cancer Registry, the Sacramento Cancer Registry, and the Minnesota Cancer Surveillance System. The rectal cancer study was conducted in the state of Utah and Kaiser Permanente Medical Care Program of northern California only; individuals with diagnoses of rectal cancer (ICD02 19, 20) between May 1997 and May 2001 were identified through the registries in those states. In northern California, only cases who were members of the Kaiser Permanente Medical Care Program were eligible for the studies. In Utah, control subjects ages ≤ 64 years were randomly selected from computerized drivers' license lists and those ages ≥ 65 years from Center for Medicare Studies (formerly Health Care Finance Administration) lists. Membership lists of the Kaiser Permanente Medical Care Program were used as a sampling frame for controls in northern California. In Minnesota, controls were selected from drivers' license lists. Control groups were frequency-matched to the age and sex distributions of incident cancer cases. Additional eligibility criteria for both cases and controls were as follows: being 30 to 79 years of age, being able to complete an interview in English, and having no previous diagnosis of colon or rectal cancer. Details of the procedures for recruitment of participants have been described (26-28). Recruitment and interviewing were conducted in accordance with human subject research protocols approved at each responsible institution. Participants took part in a structured interview. Each participant was asked to identify his or her race and ethnic background. Interview response rates among colon cancer study cases and controls were 67.7% and 52.4%, respectively, and in the rectal cancer study were 65.2% and 65.3% (27, 29).

Genotyping. Participants were asked to provide a blood sample for DNA extraction. An A>G SNP in the untranslated region (rs154410), intron 8 to 9, was detected by PCR and RFLP using the *Bsm1* enzyme, according to a published method (30) with modifications as described (22). Presence of the *Bsm1* restriction site is represented by "b" and its absence as "B." The length of the poly(A) repeat was determined by PCR amplification as described (22, 30); the size of PCR products was determined after separation by gel electrophoresis and categorized as short "S," 14 to 17 repeats, or long, "L," 18 to 22 repeats. The T>C SNP (rs10735810) that affects the first of two transcription start codons (ATG) was detected by PCR and *Fok1* RFLP (10); the presence of the *Fok1* restriction site (first ATG intact) is represented by "f" and its absence (first ATG start codon lost, protein product will be shorter by three amino acids) by "F."

Data Analysis. This analysis is restricted to subjects who provided a blood sample and for whom all three *VDR* polymorphisms were successfully genotyped: 1,452 colon cancer cases (73% of 1,993 interviewed cases), 1,811 colon cancer study controls (75% of 2,410 interviewed), 679 rectal cancer cases (75% of 952 interviewed), and 905 rectal cancer study controls (77% of 1,205 interviewed). Subjects were categorized as belonging to one of several mutually exclusive racial or ethnic groups: White (Caucasian or White; not Hispanic), Hispanic (Hispanic or Latino; not African-American or Black), African-American (African-American or Black), or Asian (Asian or Pacific Islander), based on self-report of race and ethnicity. A few individuals reported race as Native American, but they were excluded from this analysis because there were too few subjects to conduct haplotype analysis for this group.

We used genotype data from the control subjects, considering each racial or ethnic group separately, to infer chromosomal phase of linked loci and to estimate haplotype frequencies. We used the expectation-maximization algorithm (31), implemented in SAS/Genetics software (SAS/Genetics,

Cary, NC, 2002), to develop maximum likelihood estimates of population haplotype frequencies. The algorithm converges on the haplotype frequencies that have the highest probability of generating the observed genotypes. An omnibus test comparing haplotype distribution of two groups (32) was used to test for differences by race. Lewontin's D' statistic for linkage disequilibrium was calculated for each pair of polymorphisms. The values of Lewontin's D' can range from -1 to 1 ; we present the absolute value of D' , with zero indicating complete independence between two markers and one representing complete linkage disequilibrium. The expectation-maximization algorithm output also estimates each subject's probabilities of carrying particular pairs of haplotypes, based on his or her genotypes. These probabilities can be used in regression models to evaluate disease association (33). Haplotype probabilities for cases were assigned according to the estimated haplotype probabilities of a control from the same race or ethnic group with the same genotypes. Eight haplotypes were possible, based on the three polymorphisms considered. The haplotype probabilities were used to create haplotype dose variables (34). The values of the haplotype dose for an individual can range from zero, indicating that the haplotype is not possible based on the subject's genotypes, to two, indicating homozygosity for the haplotype; the sum of the eight haplotype probabilities equals to two.

Case-control comparisons were made for each of three race or ethnic groups, White, Hispanic, and African-American; small numbers of Asian subjects precluded meaningful case-control analysis for this racial group. An omnibus test for differences in the distribution of haplotypes between cases and controls was run, implementing a model-free statistical comparison (32). For our primary analysis, the haplotype dose variables were entered, as continuous variables, into logistic regression models to estimate odds ratios (OR) for the association between the haplotype and colon or rectal cancer. We first ran logistic regression models that included all haplotype variables. One model included all White, Hispanic, and African-American subjects and was adjusted for age, sex, study center, and race. Additional models were run separately for Whites, Hispanics, and African-Americans, adjusting for age, sex, and study center. We then used stepwise selection procedures to identify haplotype variables most likely to predict case-control status and to estimate ORs for haplotypes retained in the more parsimonious model. We also ran logistic models using haplotype variables coded 0, 1, 2, representing each subject's most likely pair of haplotypes, instead of haplotype dose. Finally, we compared *VDR* haplotype variables to *VDR* genotype variables, which have been reported on previously in this population (15, 22), on ability to explain colon cancer case-control differences. Using the logistic model, we evaluated improvement in $-2 \log$ -likelihood values; in other words, the improvement in model fit when a haplotype or genotype variable (or set of variables) was added to a restricted logistic model.

Results

The allele frequencies for the *VDR Bsm1*, poly(A), and *Fok1* polymorphisms among controls, by race and ethnicity, are shown in Table 1. The genotype distributions were in Hardy-Weinberg equilibrium within each group, with the exception that a smaller than expected number of heterozygotes for the *Bsm1* and poly(A) variants was observed among the small sample ($n = 26$) of Asian controls. The absolute values of the D' statistics indicated that the linkage disequilibrium between the *Bsm1* b and poly(A) L genotypes was almost complete among White and Hispanic subjects. Linkage disequilibrium was also seen between *Bsm1* and poly(A) among African-American and Asian subjects, but was not as strong. There was little evidence

Table 1. Allele frequencies and linkage disequilibrium coefficients for VDR Bsm1, poly(A), and Fok1 polymorphisms, by race, in population-based control subjects

Race or ethnicity	n	Allele frequency						Linkage disequilibrium coefficient (D')*		
		Bsm1 [†]		poly(A) [†]		Fok1 [†]		Bsm1/poly(A)	Bsm1/Fok1	poly(A)/Fok1
		b	B	L	S	F	f			
White	2,503	0.59	0.41	0.59	0.41	0.61	0.39	0.97	0.01	0.02
Hispanic	122	0.69	0.31	0.68	0.32	0.58	0.42	1.00	0.17	0.22
African-American	91	0.71	0.29	0.69	0.31	0.80	0.20	0.49	0.01	0.06
Asian	26	0.88	0.12	0.90	0.10	0.69	0.31	0.77	0.42	0.35

*Absolute value of Lewontin's D' statistic for linkage disequilibrium between this pair of markers.

[†]Genotype symbols are as follows: "b" presence and "B" absence of the intron 8 to 9 Bsm1 restriction site (rs154410); "S" short, 14 to 17 repeats, and "L" long, 18 to 22 repeats, of the poly(A) polymorphism; "F" presence and "f" absence of the ATG start codon Fok1 restriction site (rs10735810).

of linkage disequilibrium between the Bsm1 and Fok1 polymorphisms, and similarly little between the poly(A) polymorphism and Fok1. When data were analyzed separately by study center, genotype frequencies and linkage disequilibrium coefficients within each race or ethnic group were essentially identical for each center.

The haplotype frequencies predicted by the expectation-maximization algorithm are shown in Table 2. The haplotype bLF was the most frequent haplotype among all race and ethnic groups studied. The four haplotypes containing Bsm1 b with poly(A) S, or Bsm1 B with poly(A) L, were uncommon in all groups. The four uncommon haplotypes were particularly rare, <1% frequency for each haplotype, in the White and Hispanic populations, in keeping with the strong linkage disequilibrium between Bsm1 and poly(A) polymorphisms in these groups. The haplotype distributions differed significantly by race.

In case-control comparisons for colon cancer (Table 3), the omnibus test detected differences between cases and controls in the distribution of VDR haplotypes for all races combined ($P = 0.0004$) and for Whites ($P = 0.004$). The logistic regression model results from the full model, including haplotype dose variables for all eight haplotypes, indicated that the haplotypes bLF and BLF were associated with increased risk of colon cancer, and that the haplotype bLf was associated with reduced risk of colon cancer. The confidence intervals (CI) on the ORs for each of these three haplotypes excluded one in the model for all three races and in the model limited to Whites. ORs for bLF and colon cancer among African-Americans and Hispanics, and for BLF and colon cancer in African-Americans, were >1, suggestive of an increased risk of a similar or greater magnitude to the association observed for Whites, but CIs were wide in these smaller groups of subjects.

In the stepwise modeling analysis, with P values of 0.05 specified for entry into the model, the bLF variable, with an OR of 1.15 (95% CI, 1.03-1.28), and the BLF variable, with an OR of 2.40 (95% CI, 1.43-4.02), were selected in the model that included colon cancer cases and controls of all races. The same two

haplotypes were selected by the stepwise model in an analysis restricted to Whites. For Hispanics and African-Americans, no variables were entered by the stepwise model. The OR for the bLF haplotype dose variable represents a 15% increased risk of colon cancer for individuals with one allele for the haplotype bLF compared with those with no copies of this allele, and individuals who are homozygous for the bLF haplotype are at 30% increased risk. The haplotype bLF was common, with an estimated frequency >35% in all of the race and ethnic groups that we studied, so ~12% of the population is expected to be homozygous for this haplotype. The OR for the BLF haplotype dose variable indicates a >2-fold increased risk for individuals carrying one BLF allele. Because of the low frequency of the BLF haplotype, homozygosity for BLF would be extremely rare.

The omnibus test did not detect differences in haplotype distributions between rectal cancer cases and controls, and there were no haplotype variables that were significantly associated with rectal cancer in the full logistic model (Table 3). The stepwise selection procedure did not identify any haplotype variables as associated with rectal cancer at $P = 0.05$ nor at $P = 0.10$.

When we ran the logistic models using variables coded 0, 1, 2 for each haplotype, representing each subject's most likely pair of haplotypes, results were very similar to those shown in Table 3, although some OR estimates were very slightly closer to the null. For example, in the full model for colon cancer for all races, the ORs for bLF and BLF were 1.12 (95% CI, 1.01-1.24) and 2.14 (95% CI, 1.36-3.40), respectively, from the most likely pairs analysis, compared with 1.13 (95% CI, 1.01-1.26) and 2.26 (95% CI, 1.36-3.78) from the haplotype dose model.

Logistic models containing VDR genotype or VDR haplotype variables were compared (Table 4). The difference in the $-2 \log$ likelihoods associated with two haplotype-dose variables, bLF and BLF, was 16.58 units, whereas the corresponding value for adding dummy variables representing Fok1 Ff and ff genotypes was 7.32; both represent statistically significant improvement in model fit, and thus significant association with case-control

Table 2. VDR haplotype frequencies, by race or ethnicity, in population-based control subjects

Haplotype*	White (n = 2,503)	Hispanic (n = 122)	African-American (n = 91)	Asian (n = 26)
bLF	0.359	0.373	0.466	0.632
BSF	0.243	0.201	0.142	0.022
bLf	0.223	0.312	0.120	0.233
BSf	0.160	0.106	0.044	0.055
BLF	0.006	<0.001	0.088	0.019
bSf	0.005	0.008	0.101	0.020
bSf	0.003	<0.001	0.027	<0.001
BLf	0.001	<0.001	0.012	0.019

NOTE: Population haplotype frequencies were estimated from genotype data using the expectation-maximization algorithm. The distribution of haplotype frequencies in Whites differed from Hispanics ($P = 0.01$), African-Americans ($P < 0.001$), and Asians ($P < 0.001$).

*Genotype symbols are as follows: "b" presence and "B" absence, of intron 8 to 9 Bsm1 restriction site (rs154410); "S" short, 14 to 17 repeats, and "L" long, 18 to 22 repeats, of the poly(A) polymorphism; "F" presence and "f" absence of the ATG start codon Fok1 restriction site (rs10735810).

status, but the ability of the haplotype variables to explain variability in the model is better than that of genotype variables. The model with the two haplotype variables selected by the stepwise model was associated with more improvement in -2 log-likelihood values than any genotype variable or variables. However, for the model containing one haplotype dose variable for the common haplotype bLF, the model improvement, 4.91 units, was not greater than that associated with the *Fok1* genotype variable, 7.29 units.

Discussion

We evaluated haplotypes based on three common variants of the *VDR* gene in a data set from large, population-based case-control studies of cancers of the colon and rectum. We observed tight linkage disequilibrium between the *Bsm1* and poly(A) polymorphisms, whereas there was almost no linkage disequi-

librium between the *Fok1* polymorphism and the other loci. The absence of linkage disequilibrium between the *Fok1* and other SNPs is consistent with what has previously been reported (11). There was variation in the extent of linkage disequilibrium among race or ethnic groups in this U.S. study population, as has been reported for *Bsm1* and poly(A) in another population (6).

The overall distribution of *VDR* haplotypes differed between colon cancer cases and controls, but not between rectal cancer cases and controls. The number of Hispanic and African-American subjects in this study population was too small to detect significant associations between haplotype and disease within these groups analyzed separately. However, there were qualitative similarities of the ORs between Whites, Hispanics, and African-Americans, indicating that, despite the differences in *VDR* haplotype distributions among these groups, it was appropriate to combine the groups in a race-adjusted model for case-control comparisons. For the

Table 3. ORs for associations between *VDR* haplotypes and colon and rectal cancers, among all subjects and by race

Haplotype*	Colon cancer		Rectal cancer	
	Full model† results	Stepwise selection results	Full model† results	Stepwise selection results
	OR‡ (95% CI)	OR‡ (95% CI)	OR‡ (95% CI)	OR‡ (95% CI)
All subjects				
<i>n</i> controls/cases	1,811/1,451		905/679	
bLF	1.13 (1.01-1.26)	1.15§ (1.03-1.28)	0.95 (0.81-1.12)	(No variables entered)
BSF	1.01 (0.89-1.15)		1.05 (0.88-1.27)	
bLf	0.87 (0.76-0.99)		1.09 (0.90-1.31)	
BSf	0.88 (0.75-1.03)		0.89 (0.70-1.13)	
BLF	2.26 (1.36-3.78)	2.40§ (1.43-4.02)	0.83 (0.38-1.82)	
bSF	0.76 (0.43-1.33)		0.92 (0.32-2.63)	
bSf	0.82 (0.31-2.13)		1.84 (0.42-7.96)	
BLf	2.45 (0.65-9.16)		3.20 (0.28-37.11)	
	<i>P</i> = 0.0004		<i>P</i> = 0.55	
Whites				
<i>n</i> controls/cases	1,696/1,333		807/604	
bLF	1.12 (1.00-1.26)	1.14§ (1.02-1.27)	0.91 (0.77-1.08)	(No variables entered)
BSF	1.03 (0.91-1.18)		1.10 (0.91-1.33)	
bLf	0.84 (0.74-0.97)		1.09 (0.89-1.33)	
BSf	0.89 (0.76-1.05)		0.90 (0.70-1.15)	
BLF	2.50 (1.37-4.57)	2.62§ (1.43-4.80)	0.77 (0.29-2.02)	
bSF	1.00 (0.51-1.95)		0.83 (0.21-3.27)	
bSf	0.72 (0.23-2.21)		2.05 (0.36-11.80)	
BLf	2.33 (0.60-9.15)		39.04 (0.35-999.9)	
	<i>P</i> = 0.004		<i>P</i> = 0.16	
Hispanics				
<i>n</i> controls/cases	65/54		57/48	
bLF	1.17 (0.62-2.20)	(No variables entered)	1.28 (0.69-2.39)	(No variables entered)
BSF	0.56 (0.26-1.23)		0.65 (0.27-1.57)	
bLf	1.97 (0.92-4.21)		0.94 (0.51-1.73)	
BSf	0.65 (0.22-1.88)		1.12 (0.41-3.07)	
BLF	NC [¶]		NC	
bSF	NC		NC	
bSf	NC		NC	
BLf	NC		NC	
	<i>P</i> = 0.51		<i>P</i> = 0.80	
African-Americans				
<i>n</i> controls/cases	50/64		41/27	
bLF	1.31 (0.71-2.43)	(No variables entered)	1.20 (0.57-2.53)	(No variables entered)
BSF	0.54 (0.21-1.38)		0.47 (0.11-2.10)	
bLf	1.09 (0.42-2.81)		1.72 (0.56-5.30)	
BSf	0.10 (0.004-2.18)		0.20 (0.01-3.12)	
BLF	1.94 (0.73-5.17)		0.95 (0.25-3.65)	
bSF	0.46 (0.15-1.39)		0.98 (0.19-5.17)	
bSf	1.38 (0.20-9.71)		1.17 (0.07-19.57)	
BLf	5.09 (0.04-745.6)		0.26 (0.002-31.92)	
	<i>P</i> = 0.53		<i>P</i> = 0.86	

*Genotype symbols are as follows: "b" presence and "B" absence, of the intron 8 to 9 *Bsm1* restriction site (rs154410); "S" short, 14 to 17 repeats, and "L" long, 18 to 22 repeats, of the poly-A polymorphism; "f" presence and "F" absence of the ATG start codon *Fok1* restriction site (rs10735810).

†Full models include all haplotype variables.

‡ORs for association between haplotype and colon or rectal cancer, from logistic regression models. Calculated using haplotype dose variables derived from the expectation-maximization model output. Adjusted for study center, age, and sex; "all subjects" model is also adjusted for race.

§ORs for haplotype variables selected at $P < 0.05$; where no OR is shown, the haplotype was not selected.

||Significance of omnibus test for case-control differences in haplotype distribution.

¶Not calculated: ORs could not be calculated for haplotypes BLF, bSF, bSf, and BLf among Hispanics due to low frequency.

Table 4. Comparison of contribution of VDR genotype versus VDR haplotype variables to colon cancer logistic regression models

Model	Model covariates	-2 log L	Models compared	df ^e	Likelihood ratio χ^2	P ^f
1	Age, sex, center, race	4,449.90				
2	Model 1 + bLF haplotype dose and BLF haplotype dose	4,433.32	2 vs 1	2	16.58	0.0003
3	Model 1 + <i>Fok1</i> Ff and ff genotypes	4,442.57	3 vs 1	2	7.32	0.03
4	Model 1 + <i>Fok1</i> genotype 0,1,2 ^g , and <i>Bsm1</i> /poly(A) genotype ^h 0,1,2 ^g	4,441.45	4 vs 1	2	8.44	0.01
5	Model 1 + bLF haplotype dose	4,444.99	5 vs 1	1	4.91	0.03
6	Model 1 + <i>Fok1</i> genotypes 0,1,2 ^g	4,442.60	6 vs 1	1	7.29	0.007
7	Model 1 + <i>Bsm1</i> /poly(A) genotype ^h 0,1,2 ^g	4,448.62	7 vs 1	1	1.27	0.26

^eDegrees of freedom for this comparison.

^fP value (likelihood ratio test) for significance of the genotype or haplotype variable(s) in this model.

^gVariable with three categories, representing number of copies of the minor allele, entered as a continuous variable.

^hRepresenting one or two alleles containing either *Bsm1* B or poly(A) S; for most subjects, *Bsm1* B and poly(A) S are concordant.

combined analysis, after stepwise selection of haplotype variables, ORs for colon cancer risk associated with haplotype dose were significantly greater than one for bLF, the most common haplotype, and for BLF, a relatively rare haplotype.

We considered haplotype variables as an alternative to variables representing genotypes for modeling colon cancer risk related to *VDR* variants. We compared logistic regression models that included the same study subjects and covariates, differing only in the variables used to represent *VDR* haplotype or genotype. The combination of the bLF and BLF haplotypes conferred the most improvement in model fit. However, a variable representing the *Fok1* genotype was also associated with a very significant improvement in model fit. Among the three polymorphisms examined, the *Bsm1* and poly(A) were in almost perfect linkage disequilibrium, but only very weak linkage disequilibrium existed between these two and the *Fok1* SNP. With this linkage disequilibrium pattern, genotype variables served almost as well as haplotypes to model disease risk; contrasts between explanatory power of haplotype versus genotype variables might be stronger for genes or polymorphisms with more complex linkage disequilibrium patterns.

The number of possible haplotypes in a data set can be large, so that when case-control comparisons of haplotypes are conducted, there is a possibility that a significant association for a single haplotype may represent a chance finding among multiple comparisons. It is important to consider results of a single test for differences in the distribution of haplotypes between cases and controls when making comparisons using haplotypes. In our analysis of *VDR* haplotypes, the detection of a highly significant difference in distribution of haplotypes between colon cancer cases and controls based on the omnibus test, a model-free comparison, improves our confidence in the conclusion that differences are not explained by chance. The omnibus test has certain limitations: It does not describe the magnitude of the association and it does not take into account covariates such as study center or race. In the present study, haplotype distributions were consistent across centers and case-control differences were similar by race. Logistic regression models were used to estimate ORs associated with haplotype dose, adjusted for covariates.

It has been suggested that only haplotypes with a population frequency of $\geq 5\%$ should be evaluated in case-control comparisons. However, associations between rare alleles and disease may be overlooked if this rule of thumb is applied. Infrequent haplotypes cannot be meaningfully analyzed in small data sets, but the present study evaluated a large number of colon and rectal cancer cases and controls so that we were able to retain rare haplotypes for analysis. We applied a stepwise modeling procedure to select haplotypes that best predicted disease risk; the BLF haplotype, which had a frequency of $< 1\%$ among Whites and Hispanics, proved to be statistically significantly

associated with colon cancer. A possible drawback of including all haplotypes in a model is that these highly correlated variables may produce output that is difficult to interpret. In this example, however, the ORs from the full and reduced models were very similar for the two haplotypes selected.

The traditional approach to investigating candidate genes in molecular epidemiology has been to select one or a few individual polymorphisms hypothesized to confer disease susceptibility through an effect on protein function or on the expression of the gene and to examine genotypes for these polymorphisms in relation to disease risk. More recently, investigators have been encouraged to take into account variability throughout the candidate gene by identifying haplotype-tagging SNPs and analyzing these markers in study populations (25, 35, 36). The second approach will be more resource-intensive, requiring that genotypes be determined for a larger number of markers of unknown functional significance, and only a limited number of examples have appeared in the cancer epidemiology literature (e.g., ref. 34). Analysis based on tag SNPs will become more feasible as our understanding of the human genome expands, the efficiency of genotyping methods improves, and statistical methods for haplotype analysis are further developed. Still, the traditional approach will remain valuable when there is knowledge of biological function associated with a polymorphism. A third, hybrid approach is possible: to focus on polymorphisms believed to be functional, but to assay multiple polymorphisms in the gene and to construct haplotypes based on the combination of these markers. The present study is an example of this hybrid approach. The *VDR Fok1* polymorphism is recognized to be functional, with the C allele altering a translation start codon so that the protein would differ in length by three amino acids. The *Bsm1* and poly(A) polymorphisms are among a group of variants that are associated with functional differences in bone density, although the polymorphisms are thought to be markers rather than functional variants. In our analysis, the comparison of likelihoods from logistic models showed that, using data from the same polymorphisms, the *VDR* haplotype variables were only somewhat better predictors of disease risk than *VDR* genotype variables.

Acknowledgments

We thank Michael Hoffman and Sandie Edwards for technical assistance with this study.

References

1. Welsh J, Wietzke JA, Zinser GM, et al. Impact of the vitamin D3 receptor on growth-regulatory pathways in mammary gland and breast cancer. *J Steroid Biochem Mol Biol* 2002;83:85-92.
2. Colston KW, Perks CM, Xie SP, Holly JM. Growth inhibition of both MCF-7

- and Hs578T human breast cancer cell lines by vitamin D analogues is associated with increased expression of insulin-like growth factor binding protein-3. *J Mol Endocrinol* 1998;20:157–62.
3. Rozen F, Yang XF, Huynh H, Pollak M. Antiproliferative action of vitamin D-related compounds and insulin-like growth factor-binding protein 5 accumulation. *J Natl Cancer Inst* 1997;89:652–6.
 4. Uitterlinden AG, Fang Y, van Meurs JB, van Leeuwen H, Pols HA. Vitamin D receptor gene polymorphisms in relation to Vitamin D related disease states. *J Steroid Biochem Mol Biol* 2004;89–90:187–93.
 5. Ingles SA, Ross RK, Yu MC, et al. Association of prostate cancer risk with genetic polymorphisms in vitamin D receptor and androgen receptor. *J Natl Cancer Inst* 1997;89:166–70.
 6. Ingles SA, Haile RW, Henderson BE, et al. Strength of linkage disequilibrium between two vitamin D receptor markers in five ethnic groups: implications for association studies. *Cancer Epidemiol Biomarkers Prev* 1997;6:93–8.
 7. Durrin LK, Haile RW, Ingles SA, Coetzee GA. Vitamin D receptor 3'-untranslated region polymorphisms: lack of effect on mRNA stability. *Biochim Biophys Acta* 1999;1453:311–20.
 8. Thakkestian A, D'Este C, Eisman J, Nguyen T, Attia J. Meta-analysis of molecular association studies: vitamin D receptor gene polymorphisms and BMD as a case study. *J Bone Miner Res* 2004;19:419–28.
 9. Uitterlinden AG, Fang Y, Van Meurs JB, Pols HA, Van Leeuwen JP. Genetics and biology of vitamin D receptor polymorphisms. *Gene* 2004;338:143–56.
 10. Harris SS, Eccleshall TR, Gross C, Dawson-Hughes B, Feldman D. The vitamin D receptor start codon polymorphism (*FokI*) and bone mineral density in premenopausal American Black and White women. *J Bone Miner Res* 1997;12:1043–8.
 11. Nejentsev S, Godfrey L, Snook H, et al. Comparative high-resolution analysis of linkage disequilibrium and tag single nucleotide polymorphisms between populations in the vitamin D receptor gene. *Hum Mol Genet* 2004;13:1633–9.
 12. Curran JE, Vaughan T, Lea RA, Weinstein SR, Morrison NA, Griffiths LR. Association of A vitamin D receptor polymorphism with sporadic breast cancer development. *Int J Cancer* 1999;83:723–6.
 13. Ingles SA, Garcia DG, Wang W, et al. Vitamin D receptor genotype and breast cancer in Latinas (United States). *Cancer Causes Control* 2000;11:25–30.
 14. Slattery ML, Samowitz W, Hoffman M, Ma KN, Levin TR, Neuhausen S. Aspirin, NSAIDs, and colorectal cancer: possible involvement in an insulin-related pathway. *Cancer Epidemiol Biomarkers Prev* 2004;13:538–45.
 15. Slattery ML, Neuhausen SL, Hoffman M, et al. Dietary calcium, vitamin D, VDR genotypes and colorectal cancer. *Int J Cancer* 2004;111:750–6.
 16. Slattery ML, Murtaugh M, Caan B, Ma KN, Wolff R, Samowitz W. Associations between BMI, energy intake, energy expenditure, VDR genotype and colon and rectal cancers (United States). *Cancer Causes Control* 2004;15:863–72.
 17. Wong HL, Seow A, Arakawa K, Lee HP, Yu MC, Ingles SA. Vitamin D receptor start codon polymorphism and colorectal cancer risk: effect modification by dietary calcium and fat in Singapore Chinese. *Carcinogenesis* 2003;24:1091–5.
 18. Boyapati SM, Bostick RM, McGlynn KA, et al. Calcium, vitamin D, and risk for colorectal adenoma: dependency on vitamin D receptor *BsmI* polymorphism and nonsteroidal anti-inflammatory drug use? *Cancer Epidemiol Biomarkers Prev* 2003;12:631–7.
 19. Kim HS, Newcomb PA, Ulrich CM, et al. Vitamin D receptor polymorphism and the risk of colorectal adenomas: evidence of interaction with dietary vitamin D and calcium. *Cancer Epidemiol Biomarkers Prev* 2001;10:869–74.
 20. Peters U, Hayes RB, Chatterjee N, et al. Circulating vitamin D metabolites, polymorphism in vitamin D receptor, and colorectal adenoma risk. *Cancer Epidemiol Biomarkers Prev* 2004;13:546–52.
 21. Peters U, McGlynn KA, Chatterjee N, et al. Vitamin D, calcium, and vitamin D receptor polymorphism in colorectal adenomas. *Cancer Epidemiol Biomarkers Prev* 2001;10:1267–74.
 22. Slattery ML, Yakumo K, Hoffman M, Neuhausen S. Variants of the VDR gene and risk of colon cancer (United States). *Cancer Causes Control* 2001;12:359–64.
 23. Park K, Woo M, Nam J, Kim JC. Start codon polymorphisms in the vitamin D receptor and colorectal cancer risk. *Cancer Lett* 2005.
 24. Cox DG. AACR special conference: SNPs, haplotypes, and cancer—applications in molecular epidemiology, Key Biscayne, Florida, USA, 13–17 September 2003. *Breast Cancer Res* 2004;6:E9.
 25. Rebbeck TR, Ambrosone CB, Bell DA, et al. SNPs, haplotypes, and cancer: applications in molecular epidemiology. *Cancer Epidemiol Biomarkers Prev* 2004;13:681–7.
 26. Slattery ML, Potter J, Caan B, et al. Energy balance and colon cancer—beyond physical activity. *Cancer Res* 1997;57:75–80.
 27. Slattery ML, Caan BJ, Benson J, Murtaugh M. Energy balance and rectal cancer: an evaluation of energy intake, energy expenditure, and body mass index. *Nutr Cancer* 2003;46:166–71.
 28. Rogers A, Murtaugh MA, Edwards S, Slattery ML. Contacting controls: are we working harder for similar response rates, and does it make a difference? *Am J Epidemiol* 2004;160:85–90.
 29. Slattery ML, Potter JD, Duncan DM, Berry TD. Dietary fats and colon cancer: assessment of risk associated with specific fatty acids. *Int J Cancer* 1997;73:670–7.
 30. McClure L, Eccleshall TR, Gross C, et al. Vitamin D receptor polymorphisms, bone mineral density, and bone metabolism in postmenopausal Mexican-American women. *J Bone Miner Res* 1997;12:234–40.
 31. Excoffier L, Slatkin M. Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population. *Mol Biol Evol* 1995;12:921–7.
 32. Zhao JH, Curtis D, Sham PC. Model-free analysis and permutation tests for allelic associations. *Hum Hered* 2000;50:133–9.
 33. Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. *Hum Hered* 2002;53:79–91.
 34. Haiman CA, Stram DO, Pike MC, et al. A comprehensive haplotype analysis of CYP19 and breast cancer risk: the multiethnic cohort. *Hum Mol Genet* 2003;12:2679–92.
 35. Stram DO. Tag SNP selection for association studies. *Genet Epidemiol* 2004;27:365–74.
 36. Thompson D, Stram D, Goldgar D, Witte JS. Haplotype tagging single nucleotide polymorphisms and association studies. *Hum Hered* 2003;56:48–55.