

Large-Scale Evaluation of Genetic Variants in Candidate Genes for Colorectal Cancer Risk in the Nurses' Health Study and the Health Professionals' Follow-up Study

Aditi Hazra,¹ Stephen Chanock,^{5,6} Edward Giovannucci,^{1,2,3} David G. Cox,^{1,3} Tianhua Niu,¹ Charles Fuchs,^{3,4} Walter C. Willett,^{1,2,3} and David J. Hunter^{1,2,3}

¹Program in Molecular and Genetic Epidemiology, Department of Epidemiology, and ²Department of Nutrition, Harvard School of Public Health; ³Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School; ⁴Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts; and ⁵Division of Cancer Epidemiology and Genetics and ⁶Core Genotype Facility at the Advanced Technology Centre, National Cancer Institute, Department of Health and Human Services, Bethesda, Maryland

Abstract

Advances in genomics offer new strategies for assessing the association of common genetic variations at multiple loci and risk of many diseases, including colorectal cancer. Low-penetrance alleles of genes in many biological pathways, such as DNA repair, metabolism, inflammation, cell cycle, apoptosis, and Wnt signaling, may influence the risk of nonfamilial colorectal cancer. To identify susceptibility genes for colorectal cancer, we designed a large-scale case-control association study nested within the Nurses' Health Study (190 cases and 190 controls) and the Health Professionals' Follow-up Study (168 cases and 168 controls). We used a custom GoldenGate (Illumina) oligonucleotide pool assay including 1,536 single nucleotide polymorphisms (SNP) selected in candidate genes from cancer-related pathways, which have been sequenced and genotyped in the SNP500Cancer project; 1,412 of the 1,536 (92%) of the SNPs were genotyped successfully within 388

genes. SNPs in high linkage disequilibrium ($r^2 \geq 0.90$) with another assayed SNP were excluded from further analyses. As expected by chance (and not significant compared with a corrected Bonferroni $P = 0.00004$), in the additive model, 11 of 1,253 (0.9%) SNPs had a $P_{\text{trend}} < 0.01$ and 38 of 1,253 (3.0%) SNPs had a $P_{\text{trend}} \geq 0.01$ and $P_{\text{trend}} < 0.05$. Of note, the *MGMT Lys¹⁷⁸Arg* (rs2308237) SNP, in linkage disequilibrium with the previously reported *MGMT Ile¹⁴³Val* SNP, had an inverse association with colorectal cancer risk (*MGMT Lys¹⁷⁸Arg*: odds ratio, 0.52; 95% confidence interval, 0.35-0.78; unadjusted $P_{\text{trend}} = 0.0003$ for the additive model; gene-based test global $P = 0.00003$). The SNP500Cancer database and the Illumina GoldenGate Assay allowed us to test a larger number of SNPs than previously possible. We identified several SNPs worthy of investigation in larger studies. (Cancer Epidemiol Biomarkers Prev 2008;17(2):311-9)

Introduction

Colorectal cancer, a complex disease arising from both genetic and environmental factors, is the third most common cancer and the second most common cause of death due to cancer in the United States (1).

So far, susceptibility to colorectal cancer has been characterized by the identification of rare inherited mutations in a small number of established genes and by diet and lifestyle factors, including intake of red meat and alcohol as well as smoking, physical activity, and obesity. High-penetrance mutations in the *APC/WNT*

pathway (in the *APC*, *AXIN*, and *CTNNB1* genes) and mismatch repair pathway (in the *MLH1*, *MSH1*, *MSH2*, *MSH3*, *PMS1*, and *PMS2* genes) are found in a proportion of cases of familial colorectal cancer, but these alterations account for only a small fraction of the risk of colorectal cancer in the general population. However, genetic variation in common, low-penetrance genes in multiple biological pathways, such as DNA repair, metabolism, inflammation, cell cycle, apoptosis, and WNT signaling, may also contribute to the etiology of inherited and sporadic cases of colorectal cancer. The spectrum of allelic differences in low-penetrance genes could account for the interindividual variation in response to diet and lifestyle factors.

The advent of highly annotated single nucleotide polymorphism (SNP) data from the International HapMap Project, coupled with the development of high-throughput genotyping platforms, has made SNPs attractive markers for large-scale association studies in candidate genes (2-5). The candidate gene approach may offer insight valuable for detecting genetic associations with colorectal cancer. However, most published studies have been limited to a single or a small number of SNPs in a small number of genes. Therefore, we conducted a prospective nested case-control study in the Nurses' Health Study (NHS) and the Health Professionals'

Received 3/2/07; revised 10/5/07; accepted 11/21/07.

Grant support: NIH research grants CA70817, CA87969, and CA55075, Entertainment Industry Foundation National Colorectal Cancer Research Alliance, and NIH training grant T-32 CA 09001-30 (A. Hazra).

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 & Prevention Online (<http://cebp.aacrjournals.org/>).

Requests for reprints: Aditi Hazra, Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA 02115. Phone: 617-525-2035; Fax: 617-525-2008.

E-mail: ahazra@hsph.harvard.edu

Copyright © 2008 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-07-0195

Table 1. Select characteristics of the study population

	HPFS		NHS		P*
	Cases (n = 168)	Controls (n = 168)	Cases (n = 197)	Controls (n = 197)	
Mean age at diagnosis or selection (y)	71.0	70.9	65.5	65.5	—
Family history of colorectal cancer (%)	22.6	16.2	22.6	16.8	0.04
Never smokers (%)	43.5	45.5	42.1	47.4	0.64
Mean pack-years of smoking [†]	25.6	28.0	29.7	25.8	0.09
Regular aspirin use (%)	19.1	25.2	34.2	42.6	0.03
Current use of postmenopausal hormone (%) [‡]	—	—	32.0	39.6	—
Multivitamin use (%)	54.2	50.9	45.3	51.6	0.63
Mean body mass index, kg/m ²	26.1	25.8	26.2	26.4	0.44
Mean physical activity, MET-h/wk	30.2	35.1	16.2	17.1	0.22
Red meat intake ≥1 serving/d (%)	36.3	34.1	18.9	16.3	0.55
Mean dietary folate, µg/d [§]	365.5	371.7	285.5	296.9	0.35
Mean alcohol intake, g/day [§]	10.9	11.2	6.5	5.9	0.85
Colon cancer by site					
Proximal (%)	44.9	—	43.3	—	—
Distal (%)	31.3	—	33.9	—	—
Rectal (%)	23.8	—	22.8	—	—

*P value for the combined NHS and HPFS cases and controls.

[†] Pack-years of smoking were calculated among past and current smokers only.

[‡] Percentage of women using hormone replacement therapy was calculated among postmenopausal women only.

[§] Mean values calculated at baseline.

Follow-up Study (HPFS) to evaluate common sequence variants at 1,536 loci in 388 genes to assess associations with susceptibility to colorectal cancer.

Materials and Methods

Study Population. The NHS is an ongoing prospective study of 121,700 U.S. female registered nurses. Details of the design and follow-up of this cohort have been described previously (9, 10). Briefly, at enrollment in 1976, the participants, ages 30 to 55 years, completed a questionnaire providing information on risk factors for cancer and cardiovascular disease. Exposure and disease information are updated biennially. From 1989 to 1990, blood samples were collected from 32,826 of the NHS participants. After blood collection through June 2000, 197 incident cases of colorectal cancer were confirmed through medical records or death reports, of which 190 cases were successfully genotyped. Controls were randomly selected from women who were alive and free of cancer at the time of case ascertainment. One control was matched to each case on year of birth and month of blood draw.

The HPFS began in 1986 when 51,529 U.S. male dentists, optometrists, osteopaths, podiatrists, pharmacists, and veterinarians, ages 40 to 75 years, responded to a mailed questionnaire (9). These men provided baseline information on age, marital status, height, weight, ancestry, medications, smoking history, medical history, physical activity, and diet. Exposure and medical history information are updated every 2 years. Blood samples were collected from 18,225 of the HPFS participants between 1993 and 1995. Among these men, 168 incident cases of colorectal cancer were identified between the date of blood draw and January 2002. Men who were alive and free of diagnosed cancer at the time of case ascertainment were selected as controls and were matched to cases on year of birth and month of blood draw.

Sample Collection. Venous blood samples were separated into plasma, buffy coat, and RBCs and stored in liquid nitrogen. Genomic DNA was extracted from 50 µL buffy coat diluted with 150 µL PBS using the QIAmp (Qiagen) 96-spin blood protocol according to the manufacturer's instructions. Concentrations of genomic DNA were measured in 96-well format with PicoGreen technology (Molecular Probes).

Gene and SNP Selection. The SNP500Cancer database (<http://snp500cancer.nci.nih.gov>), a component of the National Cancer Institute's Cancer Genome Anatomy Project, provides sequence and genotype assay information for candidate SNPs in genes hypothesized to be related to cancer. National Cancer Institute's SNP500Cancer reports sequence analysis and allele prevalence information in anonymized control DNA samples (n = 102 Coriell samples representing four self-described ethnic groups: African/African American, Caucasian, Hispanic, and Pacific Rim; refs. 2, 4).

We designed an Illumina oligonucleotide pool assay (OPA) selecting candidate genes that were resequenced in the SNP500Cancer project (4). As of November 2004, there were a total of 5,800 SNPs with a minor allele frequency (MAF) of >3% in the combined four self-described populations (African/African American, Caucasian, Hispanic, and Pacific Rim) annotated in the SNP500 database (2). Overall, the SNP selection approach for this study was to examine 10 kb upstream and 10 kb downstream in accordance with design score validations based on Illumina in-house measurements and the 60-bp limitation (a SNP cannot be closer than 60 bp to another SNP on this OPA). After excluding SNPs with high r^2 (defined as $r^2 \geq 0.8$), we did a preliminary screen of the remaining 3,072 SNPs in the SNP500Cancer database (2-8, 11). From the 3,072 SNP panel, the final SNP selection of the 1,536 polymorphisms chosen for the OPA included SNPs with a MAF of >3% in the unrelated (that is, parents) HapMap CEPH Utah (CEU, with European

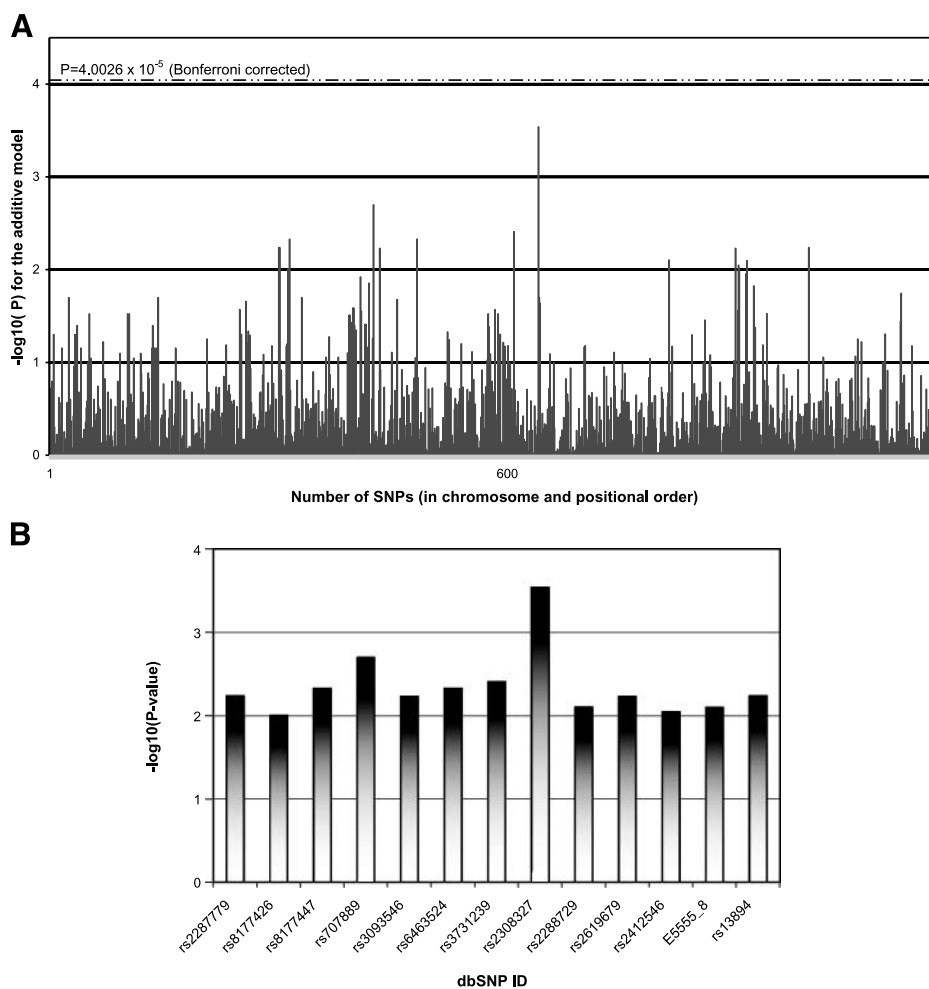


Figure 1. A. SNPs arranged in chromosomal and positional order. B. SNPs with a $\log_{10} P$ values < 2. SNP names: rs8177426 (*GPX3*), rs8177477 (*GPX3*), rs707889 (*HFE*), rs3093546 (*LTA*), rs6463524 (*PMS2*), rs3731239 (*CDKN2A*), rs2308327 (*MGMT*), rs2288729 (*APAF1*), rs2412546 (*RAD51*), rs28566535 (*CYP19A1*), and rs13894 (*SAT2*).

ancestry) samples favored nonsynonymous SNPs and SNPs with evidence for functional significance as well as those SNPs evaluated previously in relation to cancer risk (2, 4).

Genotyping Methods. Whole-genome amplified DNA plates were sent to Illumina for high-throughput genotyping by the highly multiplexed GoldenGate assay (5). The assay was conducted on DNA that was whole genome amplified using the multiple displacement amplification method relying on the Phi29 DNA polymerase protocol (12, 13) and the high concordance rates and minimal excess of significant departures from Hardy-Weinberg equilibrium (HWE) support that the high fidelity of the whole-genome amplified procedure. The GoldenGate assay queries genomic DNA with probes (two allele-specific oligonucleotides and one locus-specific oligonucleotide) and creates DNA fragments that can be amplified by standard PCR methods with three universal primers. Fluorescently labeled strands are hybridized to the appropriate probe on the microbead array, which allows identification of a particular SNP (5). Of the 1,536 SNPs, 1,421 loci were successfully genotyped with high completion rates (>80%) and MAFs >0.03 in controls, including 9 duplicate SNPs (to verify concordance); thus, there were 1,412 unique SNPs in the OPA.

SNP Characteristics. The SNPs included in the OPA had a MAF of >3% in 60 unrelated parents of CEU trios from HapMap (3). Of the 1,536 assays chosen for this study, 103 were dropped from the analysis because of low MAF (<3% in our study population) or assay problems (low genotyping success rate determined by Illumina using the GenCall score threshold at 50%). Thus, we obtained data on 1,421 SNPs (including 9 duplicate SNPs) in 388 genes. The characteristics of the 1,421 SNPs we genotyped were 51% intronic, 37% exonic, and 12% in the promoter region. Overall, in the 388 genes studied, the mean number of SNPs per gene was 3.5 and the maximum number was 37 (in the *GSK3B* gene). Of the 1,412 unique SNPs included in this study, 310 were coding SNPs, 145 of which were nonsynonymous SNPs.

In ~55 (14%) genes, there was an attempt to select of an adequate number of SNPs to tag common haplotypes. This category also included genes, *AXIN2*, *CDKN2A*, *CTNBN1*, *GSK3B*, *KRAS*, *MSH2*, and *PMS1*, which were submitted because of their high probability of association with colorectal cancer. These genes were then tagged with Tagger based on the CEU population with a MAF of >5% in dbSNP and HapMap data available in January 2005.

Statistical Analyses. Controls were matched to cases 1:1 according to age, month of blood draw, and fasting status at blood draw. Colorectal cancer risk was

Table 2. Polymorphisms associated with *P* values < 0.01 in the SNP500Cancer-Illumina OPA and risk of colorectal cancer in the NHS and HPFS

dbSNP ID	SNP500 Cancer ID	Nucleotide position*	Cases, <i>n</i> (%)	Controls, <i>n</i> [†] (%)	Conditional OR [‡] (95% CI)	Multivariate OR [‡] (95% CI)
Alleles [§]						
rs8177426	GPX3 28**	IVS1-1961A>G				
	GG		249 (71.8)	215 (62.9)	1.0	1.0
	GA		98 (28.2)	125 (36.5)	0.66 (0.48-0.91)	0.69 (0.48-0.98)
	AA		0 (0)	2 (0.6)	—	—
rs8177447	GA + AA GPX3 16	IVS4-14 T>C	98	127	0.66 (0.48-0.90)	0.68 (0.48-0.97)
	CC		260 (72.6)	224 (62.8)	1.0	1.0
	CT		89 (24.9)	116 (32.5)	0.69 (0.50-0.94)	0.73 (0.51-1.04)
	TT		9 (2.5)	17 (4.8)	0.47 (0.20-1.10)	0.43 (0.17-1.08)
rs707889	CT + TT HFE 08	IVS6 + 462G>A	98	133	0.65 (0.49-0.89)	0.69 (0.49-0.97)
	CC		231 (65.3)	195 (54.8)	1.0	1.0
	CT		116 (32.8)	138 (38.8)	0.69 (0.49-0.97)	0.60 (0.41-0.87)
	TT		7 (2.0)	23 (6.5)	0.28 (0.11-0.71)	0.22 (0.08-0.60)
rs3093546	CT + TT LTA 05	Ex1 + 50A>G	123	161	0.63 (0.46-0.87)	0.54 (0.38-0.78)
	CC		332 (95.4)	311 (89.4)	1.0	1.0
	CT		15 (4.3)	36 (10.34)	0.38 (0.20-0.72)	0.38 (0.19-0.76)
	TT		1 (0.29)	1 (0.29)	—	—
rs6463524	CT + TT PMS2 11	S260S; Ex7-24G>C	16	37	0.40 (0.22-0.74)	0.40 (0.21-0.79)
	GG		221 (61.9)	253 (70.9)	1.0	1.0
	GC		119 (33.3)	95 (26.6)	1.51 (1.07-2.13)	1.54 (1.06-2.24)
	CC		17 (4.76)	9 (2.52)	2.30 (1.00-5.27)	2.34 (0.95-5.79)
rs3731239	GC + CC CDKN2A 12	IVS2 + 185C>T	136	104	1.57 (1.12-2.20)	1.60 (1.11-2.30)
	AA		153 (42.7)	119 (33.3)	1.0	1.0
	AG		161 (45.0)	176 (49.3)	0.72 (0.52-0.99)	0.72 (0.51-1.01)

NOTE: The number of cases and controls do not add to the total number of study participants because individuals with missing exposure data were excluded from the analysis.

*For additional information on the nomenclature for description of the sequence variants, please see the SNP500Cancer database.

† Conditional logistic regression conditioned on the strata are defined by the matching variables.

‡ Conditional logistic regression conditioned on the strata are defined by the matching variables and adjusted for family history of colon cancer, pack-years of smoking, body mass index, postmenopausal hormone use, aspirin intake, physical activity, alcohol intake, dietary folate consumption, and red-meat consumption.

§ Allele names are based on Illumina assay.

|| *P* value for the Armitage trend test.

considered in relation to the 1,412 SNPs. Each polymorphic locus was tested for HWE in the control population. χ^2 statistics were used to test for differences in the distributions of the genotypes or other potential risk factors between cases and controls. Conditional logistic regression was used to compute odds ratios (OR) and 95% confidence intervals (95% CI) for the main effect of genotype and to control for potentially confounding variables. Polymorphisms were categorized into three genotypes, homozygous wild-type, heterozygous, and homozygous variant. Homozygosity for the more common allele was treated as the reference category. Tests for linear trend of log ORs (additive model) were calculated using an ordered categorical variable by assigning scores to the genotypes: 0 (no variant allele), 1 (carrying one variant allele), and 2 (carrying two variant alleles). For polymorphisms with less than five individuals with the homozygous variant genotype, homozygotes for the variant genotype were combined with heterozygotes. There were no significant

differences between men and women in the main effect of genotype for SNPs with additive $P_{\text{trend}} < 0.01$. Therefore, to maximize power only, the combined analyses are presented. Analyses were controlled for known risk factors for colorectal cancer, including family history of colorectal cancer, smoking history, body mass index, multivitamin use, aspirin use, postmenopausal hormone use (among women in the NHS), physical activity, and intake of red meat, folate, and alcohol.

For all covariates, we used updated data from the most recent questionnaire completed before diagnosis of cancer. Individuals with missing genotype data or covariate values were excluded from specific analyses. All *P* values are based on two-sided tests. All statistical analyses were done with SAS (version 9.1; SAS Institute; ref. 14).

SNP spectral decomposition was used to calculate the M_{eff} value to correct for multiple testing (<http://genepi.qimr.edu.au/general/daleN/SNPSPD/>; ref. 15). This correction strategy accounts for the linkage

Table 2. Polymorphisms associated with P values < 0.01 in the SNP500Cancer-Illumina OPA and risk of colorectal cancer in the NHS and HPFS (Cont'd)

dbSNP ID	SNP500 Cancer ID	Nucleotide position*	Cases, n (%)	Controls, n [†] (%)	Conditional OR [‡] (95% CI)	Multivariate OR [‡] (95% CI)
	Alleles [§]					
rs2308327	GG	Lys178Arg; Ex4 + 119A>G	44 (12.3)	62 (17.4)	0.55 (0.35-0.87)	0.54 (0.33-0.90)
	AG + GG		205	238	0.68 (0.50-0.92)	0.68 (0.49-0.94)
	MGMT 03		303 (85.1)	266 (74.7)	1.0	1.0
	AA		52 (14.6)	84 (23.6)	0.54 (0.37-0.79)	0.55 (0.36-0.82)
	AG		1 (0.28)	6 (1.7)	—	—
rs2288729	AG + GG	IVS12 + 2093G>A	53	90	0.51 (0.35-0.75)	0.52 (0.35-0.78)
	APAF1 04		193 (54.1)	164 (45.9)	1.0	1.0
	AA		144 (40.3)	158 (44.3)	0.75 (0.55-1.04)	0.76 (0.54-1.08)
	AG		20 (5.6)	35 (9.8)	0.46 (0.25-0.85)	0.46 (0.24-0.91)
	GG		164	193	0.70 (0.51-0.96)	0.71 (0.51-1.00)
rs2412546	AG + GG	IVS5-4480G>A	164	193	0.70 (0.51-0.96)	0.71 (0.51-1.00)
	RAD51 16		75 (21.0)	102 (28.6)	1.0	1.0
	AA		187 (52.2)	179 (50.1)	1.41 (0.98-2.02)	1.49 (1.00-2.21)
	AG		96 (26.8)	76 (21.3)	1.73 (1.12-2.66)	2.01 (1.24-3.25)
	GG		283	255	1.50 (1.06-2.12)	1.62 (1.11-2.37)
rs28566535	AG + GG	IVS2 + 14872T>G	283	255	1.50 (1.06-2.12)	1.62 (1.11-2.37)
	CYP19A1 30		307 (85.8)	328 (91.9)	1.0	1.0
	AA		47 (13.1)	28 (7.8)	1.73 (1.07-2.81)	1.64 (0.96-2.82)
	AC		4 (1.12)	1 (0.28)	—	—
	CC		51	29	1.82 (1.13-2.93)	1.84 (1.09-3.09)
rs13894	AA + CC	R126C; Ex6 + 31C>T	51	29	1.82 (1.13-2.93)	1.84 (1.09-3.09)
	SAT2 01		297 (83.2)	320 (89.9)	1.0	1.0
	GG		58 (16.3)	36 (10.1)	1.89 (1.19-3.01)	2.47 (1.46-4.17)
	GA		2 (0.56)	0 (0)	—	—
	AA		60 (16.9)	36 (10.1)	1.93 (1.21-3.07)	2.54 (1.50-4.29)
	GA + AA					

disequilibrium (LD) between polymorphic sites. Failure to account for the nonindependence of SNPs would make the Bonferroni correction overconservative. The ratio of observed eigenvalue variance, $\text{Var}(\lambda_{\text{obs}})$ to its maximum (M) gives the proportional reduction in the number of variables in a set. The effective number of variables (M_{eff}) is calculated as $M_{\text{eff}} = 1 + (M - 1) [1 - \text{Var}(\lambda_{\text{obs}}) / M]$ (15).

In addition to the SNP-based tests, we also evaluated two gene-based tests for association using the P value obtained from the likelihood ratio test comparing models with and without (a) terms for heterozygous and homozygous variant genotypes for each SNP in a gene ($df = 2 \times \text{number of SNPs per gene}$) and (b) terms for each SNP (genotypes assigned as 0, 1, 2) in a gene ($df = \text{number of SNPs per gene}$).

Results

In this study, 358 cases of colorectal cancer and 358 matched controls from the prospective NHS and HPFS cohorts were genotyped for 1,412 unique SNPs in cancer pathways. After excluding SNPs in high LD ($r^2 > 0.90$), we examined the association of 1,299 SNPs and colorectal cancer risk and present a summary of the OPA performance and the associations for the SNPs with a $P_{\text{trend}} < 0.01$ in the additive model.

OPA Characteristics. The OPA contained 1,421 successfully genotyped SNPs derived from 388 genes. The loci success rate was 92.5% (1,421 of 1,536 SNPs successfully genotyped). The genotyping success rate for the DNA samples was 98.4% [874 of 888 total samples, including quality-control (QC) samples]. Nine duplicate SNPs were included in the analysis for additional QC measures and were 99.9% concordant (data not shown). After duplicate SNPs were filtered, 1,412 unique SNPs remained for further analysis (see Supplementary Table S1 for a comprehensive list of the 1,299 SNPs analyzed after excluding SNPs in LD with an $r^2 > 0.90$).

QC Analysis. Fourteen percent blinded QC samples were included. Analysis of the 35 QC replicates, involving 130 samples, resulted a 99.95% overall concordance rate (184,640 of 184,730 QC genotype pairs were concordant). Of the 35 QC replicate sample sets, 20 (57.14%) had a 100% concordance rate. The other 15 QC sets had concordance rates ranging from 99.62% to 99.96%. The 90 discordant QC genotype samples were exclusive to 30 loci (2.11% of total loci), of which 6 (*rs3765459*, *rs1052576*, *rs3774268*, *rs3736228*, *rs7260*, and *E3359_310*) had concordance rates from 89.23% to 96.92%.

Study Population. Select characteristics of the study population are shown in Table 1. Briefly, the mean age at diagnosis was 71.0 years (range, 50-86) in the HPFS and 65.5 years (range, 46-78) in the NHS. Cases were more

Table 3. Polymorphisms associated with P values <0.05 in the SNP500Cancer-Illumina OPA and risk of colorectal cancer, using the dominant genetic model, in the NHS and HPFS

SNP ID	SNP500Cancer ID	Nucleotide position	Cases (n)	Controls (n)	Conditional OR* [†]
rs3774937	<i>IL10-17</i>	-3584A>T	159	132	1.0
rs2236771	<i>PLA2G2A-03</i>	T32T; Ex4 + 56G>C	159 + 43*	175 + 54*	1.35 (1.01-1.81)
rs1537234	<i>GSTM3-06</i>	IVS7-30G>T	297	319	1.0
rs2066470	<i>MTHFR-03</i>	P39P Ex2-120C>T	61 + 1*	40 + 0*	0.63 (0.42-0.95)
rs879332	<i>HSD3B2-25</i>	-17124G>A	136	108	1.0
rs4149371	<i>TP5313-18</i>	-578A>G	172 + 53*	190 + 62*	1.42 (1.04-1.94)
rs2289019	<i>HADHA-10</i>	IVS13-163C>G	276	302	1.0
rs256567	<i>PMS1-54</i>	IVS9-938C>T	81 + 2*	56 + 3*	0.66 (0.45-0.95)
rs4673222	<i>FSD7-20</i>	Ex1-1251G>A	87	93	1.0
rs1049216	<i>CASP3-08</i>	Ex8 + 567A>G	161 + 113*	190 + 78*	0.92 (0.66-1.28)
rs1405938	<i>CASP3-07</i>	IVS3-46A>G	321	337	1.0
rs3087455	<i>CASP3-02</i>	IVS2-1555A>C	38 + 1*	23 + 0*	0.58 (0.34-0.98)
rs3774937	<i>NFKB1-02</i>	IVS1 + 11306C>T	236	209	1.0
rs4648059	<i>NFKB1-21</i>	IVS12-452C>G	113 + 12*	133 + 19*	1.36 (1.01-1.83)
Poly-0009047	<i>HSD17B4-08</i>	Ex24-76A>G	273	289	1.0
Poly-0009064	<i>GHR-21</i>	IVS2 + 29065C>T	80 + 7*	70 + 0*	0.75 (0.52-1.08)
rs6179	<i>GHR-01</i>	G186G; Ex6-61A>G	323	297	1.0
rs6413428	<i>SEPP1-02</i>	Ex5 + 710A>G	35 + 3*	62 + 1*	1.83 (1.18-2.86)
rs2235718	<i>FOXO1-02</i>	-3077C>T	213	177	1.0
			122 + 24*	159 + 24*	1.53 (1.13-2.07)
			222	246	1.0
			119 + 18*	102 + 12*	0.75 (0.55-1.02)
			166	144	1.0
			164 + 30*	173 + 43*	1.28 (0.95-1.73)
			187	153	1.0
			136 + 38*	161 + 46*	1.48 (1.09-2.00)
			323	338	1.0
			35 + 2*	23 + 0*	0.60 (0.35-1.03)
			308	326	1.0
			49 + 2*	32 + 0*	0.62 (0.39-1.00)
			229	250	1.0
			114 + 18*	101 + 8*	0.75 (0.55-1.02)
			188	211	1.0
			142 + 31*	132 + 17*	0.79 (0.59-1.05)
			211	231	1.0
			125 + 25*	117 + 13*	0.81 (0.61-1.08)
			306	278	1.0
			49 + 5*	75 + 7*	1.64 (1.12-2.38)

NOTE: Heterozygotes + variant homozygous compared with the common homozygous variant.

*Conditional logistic regression conditioned on the strata defined by the matching variables.

† Wild-type reference compared with variant carriers.

likely than controls to have a family history of colorectal cancer and to consume ≥ 1 serving per day of red meat, less likely to be regular users of aspirin, and tended to be less physically active and have lower intakes of dietary folate.

Hardy-Weinberg Analysis. The genotype distributions for the 1,421 polymorphisms were examined and tested for agreement with HWE. Among the controls, 72 loci (5.49%) had HWE χ^2 P values ≤ 0.05 and 6 loci (0.42%) had HWE χ^2 test P values ≤ 0.001 .

Correction for Multiple Comparisons. SNPs in high LD were excluded from the analysis. First, we evaluated which SNPs had a pairwise r^2 values < 0.9 on an individual chromosome basis. If there were multiple SNPs in one gene at the $r^2 = 0.9$ level, we included nonsynonymous SNPs and SNPs with MAF of $> 1\%$ (total number of SNPs: excluded 113 SNPs). The M_{eff} value (15) calculated by SNP spectral decomposition (see Supplementary Table S2) for the remaining 1,299 SNPs is 1,253 and the resultant Bonferroni-corrected genome-wide significance threshold is 3.99×10^{-5} . Thus, the interpretation of the M_{eff} correction is that the single-

locus tests for 1,299 SNPs distributed on the 22 human autosomes are effectively equivalent to 1,253 statistically independent single-locus tests.

SNPs with P_{trend} Values < 0.01 for the Additive Model. Although not significant after this Bonferroni correction, among the 1,299 SNPs genotyped (Fig. 1A), 11 SNPs had a $P_{\text{trend}} < 0.01$ (Fig. 1B) in the additive model. These SNPs were in *glutathione peroxidase 3 (GPX3; rs8177426 and rs8177477)*, *hemochromatosis (HFE; rs707889)*, *lymphotoxin α (LTA; rs3093546)*, *postmeiotic segregation increased 2 (PMS2; rs6463524)*, *cyclin-dependent kinase inhibitor 2A (CDKN2A; rs3731239)*, *O⁶-methylguanine-DNA methyltransferase (MGMT; rs2308327)*, *apoptotic peptidase activating factor 1 (APAF1; rs2288729)*, *RAD51 homologue (RAD51; rs2412546)*, *cytochrome P450, family19, subfamily A, polypeptide 1 (CYP19A1; rs28566535)*, and *spermidine/spermine N1-acetyltransferase (SAT2/SSAT; rs13894)* and were analyzed further in multivariate conditional models (Table 2) using a SNP-based approach and a gene-based approach. The distributions of all these SNPs, except the *GPX3* SNPs, were in agreement with HWE in cases and controls from both

Table 3. Polymorphisms associated with P values < 0.05 in the SNP500Cancer-Illumina OPA and risk of colorectal cancer, using the dominant genetic model, in the NHS and HPFS (Cont'd)

SNP ID	SNP500Cancer ID	Nucleotide position	Cases (n)	Controls (n)	Conditional OR* [†]
rs9405496	<i>FOXC1-13</i>	-2049A>C	296	271	1.0
rs998075	<i>IGF2R-02</i>	T713T; Ex16 + 88A>G	59 + 6*	81 + 9*	1.50 (1.05-2.15)
rs5369	<i>EDN1-01</i>	E106E; Ex3-72A>G	84	110	1.0
rs2282151	<i>NFKB1E-03</i>	*760bp 3' of STP G>A	189 + 88*	176 + 75*	0.67 (0.47-0.95)
rs1549760	<i>CDK5-16</i>	-903G>A	292	262	1.0
rs2979895	<i>POLB-16</i>	IVS2-2264G>A	61 + 6*	92 + 6*	1.65 (1.15-2.38)
E3663_301	<i>TERF1-02</i>	IVS7 + 82C>T	236	210	1.0
rs3891248	<i>MYC-02</i>	IVS1-355A>T	108 + 17*	122 + 29*	1.34 (1.00-1.80)
rs10082466	<i>MBL2-27</i>	Ex4-1483T>C	228	209	1.0
Poly-0014684	<i>GPX2-07</i>	*2680bp 3' of STP T>A	120 + 13*	124 + 27*	1.22 (0.92-1.62)
rs4924496	<i>RAD51-17</i>	IVS3 + 1932T>C	318	302	1.0
rs730154	<i>CYP19A1-16</i>	IVS2 + 24809A>G	41 + 1*	53 + 6*	1.53 (0.98-2.39)
rs1869145	<i>LIPC-25</i>	IVS1-33033C>T	208	228	1.0
rs1968689	<i>LIPC-37</i>	IVS1-7747C>T	132 + 21*	123 + 9*	0.77 (0.56-1.05)
rs2159116	<i>FANCA-34</i>	IVS27-36G>T	279	254	1.0
Poly-0014870	<i>STK6-08</i>	*800bp 3' of STP G>C	74 + 7*	96 + 11*	1.44 (1.03-2.01)
rs10485805	<i>STK6-16</i>	IVS9-68T>C	197	225	1.0
rs4646312	<i>COMT-16</i>	IVS1-385C>T	141 + 23*	122 + 14*	0.73 (0.54-0.98)
rs4680	<i>COMT-01</i>	V158M; Ex4-12G>A	324	308	1.0
			34 + 0*	51 + 2*	1.61 (1.03-2.52)
			113	135	1.0
			173 + 75*	170 + 56*	0.76 (0.55-1.04)
			233	266	1.0
			108 + 19*	81 + 13*	0.65 (0.48-0.89)
			204	233	1.0
			132 + 20*	113 + 12*	0.71 (0.52-0.95)
			320	300	1.0
			39 + 2*	59 + 2*	1.59 (1.03-2.44)
			257	237	1.0
			78 + 12*	96 + 18*	1.42 (1.02-1.97)
			270	247	1.0
			86 + 5*	105 + 9*	1.43 (1.01-2.02)
			238	211	1.0
			109 + 14*	128 + 22*	1.40 (1.03-1.92)
			108	129	1.0
			178 + 74*	181 + 50*	0.80 (0.59-1.07)
			106	84	1.0
			180 + 66*	183 + 86*	0.70 (0.48-1.01)

the NHS and the HPFS cohorts. The nonsynonymous *MGMT* *Lys*¹⁷⁸*Arg* SNP had a $P_{\text{trend}} = 0.0003$ in the additive model (Fig. 1B). We found a significant inverse association among individuals with one copy (multivariate OR, 0.54; 95% CI, 0.37-0.79) and one or more copies of the *MGMT* *Lys*¹⁷⁸*Arg* variant allele (multivariate OR, 0.51; 95% CI, 0.35-0.75). The relationship remained after adjusting for other known risk factors for colorectal cancer (one or more copies of the variant allele: OR, 0.52; 95% CI, 0.35-0.78). The *MGMT* gene-based tests for association (Bonferroni corrected $P = 0.0001$), evaluating four SNPs, had a notable global P value for genotype associations ($P = 0.0003$) and a global P value based on trend test for individual SNPs ($P = 0.00003$).

Compared with the *HFE* *IVS6* + 462 CC genotype, the CT variant (multivariate OR, 0.60; 95% CI, 0.41-0.87) and the TT variant (OR, 0.22; 95% CI, 0.08-0.60) were associated with a reduced risk of colorectal cancer. Compared with the *LTA* 50 CC genotype, the CT + TT variants (OR, 0.40; 95% CI, 0.21-0.79) were inversely associated with colorectal cancer. The *APAF1* 2093 AG + GG variants were associated with a suggestive reduction in risk of colorectal cancer (OR, 0.71; 95% CI, 0.51-1.00).

The *PMS2* *S260S*, *RAD51* *A4480G*, and *SAT2* *Arg*¹²⁶*Cys* SNPs were associated with an increased risk of colorectal

cancer. The risk of colorectal cancer was increased in individuals carrying the *PMS2* 260 CC genotype (OR, 2.34; 95% CI, 0.95-5.79) relative to the risk in those carrying the GG genotype. The OR was slightly attenuated in the variant carrier model (OR, 1.60; 95% CI, 1.11-2.30). Differences were observed in the genotype distribution between cases and controls for *rs2412546* in intron 5 of *RAD51* ($P_{\text{trend}} = 0.009$). The minor allele of the *RAD51* SNP was associated with an increased risk of colorectal cancer (OR, 1.62; 95% CI, 1.11-2.34). The nonsynonymous SNP at *Arg*¹²⁶*Cys* (*rs13894*) in exon 6 of *SAT2* ($P_{\text{trend}} = 0.006$) was associated with an increased risk of colorectal cancer (variant carrier genotypes: OR, 2.54; 95% CI, 1.50-4.29).

SNPs with $P_{\text{trend}} > 0.01$ and $P_{\text{trend}} < 0.05$. Table 3 presents the main effects for 43 SNPs with a $P_{\text{trend}} < 0.05$ but $P_{\text{trend}} > 0.01$. Of interest is the functional nonsynonymous polymorphism *catechol-O-methyltransferase* (*COMT*) *Val*¹⁵⁸*Met* involved in the dopaminergic pathway and associated with several psychiatric conditions (16), which in this study was associated with a reduced risk of colorectal cancer (OR, 0.75 and 0.58 for the heterozygote and minor allele variant, respectively).

Nonsynonymous SNPs. A total of 145 nonsynonymous SNPs were successfully genotyped (see Supplementary

Table S3), including nonsynonymous SNPs in genes, such as *APC*, *AXIN*, *MLH1*, and *MHS3*, associated previously with colorectal cancer risk. All had *P* values >0.05, except for the previously noted *MGMT* R178K, *SAT2* R126C, and *COMT* Val¹⁵⁸Met polymorphisms (Bonferroni corrected *P* value for the 145 nonsynonymous SNPs = 0.0003).

Discussion

New technologies for large-scale SNP analysis enable much larger sets of SNPs to be tested for association simultaneously and may help in deciphering the complex nature of colorectal and other cancers. In this set of *a priori* candidate genes in cancer pathways, almost all the SNP associations were null, although our sample size and thus the statistical power to detect a modest association is limited. A limitation of genetic association studies is the lack of functional information for many SNPs. Nonsynonymous SNPs are common genetic variants that alter encoded amino acids in proteins; thus, nonsynonymous SNPs may modify the structure or function of expressed proteins. Our analysis included 145 nonsynonymous SNPs that had a higher probability of being associated with colorectal cancer risk; 3 of the nonsynonymous SNPs had nominally significant *P* values <0.05. An additional benefit of large-scale analysis, which yields an abundance of genotypic data, is that many SNPs that are not significantly associated with disease risk can be reported in a single study, thus reducing the probability of publication bias.

Quality Control. As this is one of the first case-control analyses using the Illumina GoldenGate assay to be reported, we did a detailed evaluation of the data quality. The high concordance of the QC data confirms that large-scale association studies using the GoldenGate technology provide highly reliable results at a cost of about five cents per genotype.

SNPs with P_{trend} Values <0.01. There were 11 SNPs in 10 genes in various biological pathways associated with risk of colorectal cancer at the P_{trend} 0.01 = value with the additive model. SNPs in *HFE*, *LTA*, *PMS2*, *CDKN2A*, *MGMT*, *APAF1*, *RAD51*, and *SAT2* genes were analyzed further in multivariate conditional models. Main-effects analysis identified the association of two nonsynonymous SNPs, *MGMT* Lys¹⁷⁸Arg ("benign" by polymorphism phenotyping, which combines a conservation score with additional properties to predict the functional importance of an amino acid alteration) and *SAT2* Arg¹²⁶Cys ("possibly damaging" by polymorphism phenotyping), with colorectal cancer risk (17). In a recent evaluation of 1,041 nonsynonymous SNPs in 2,575 colorectal cancer cases and 2,707 controls colorectal cancer, these nonsynonymous SNPs were not significantly associated with colorectal cancer risk (18).

MGMT eliminates mutagenic DNA adducts from the O⁶ position of the guanine nucleotide in the DNA direct reversal repair pathway (19, 20). The *MGMT* Lys¹⁷⁸Arg variant allele was strongly associated with a reduced risk of colorectal cancer. In addition, gene-based tests for *MGMT* also suggested associations with colorectal cancer. The *MGMT* Lys¹⁷⁸Arg SNP is in strong LD with *MGMT* Ile¹⁴³Val. The current findings overlay our previously reported associations for the *MGMT* Ile¹⁴³Val

among NHS women (10); however, the *MGMT* Lys¹⁷⁸Arg polymorphism among men in the HPFS have not been evaluated before. The Ile¹⁴³Val SNP, in exon 5, lies in close proximity to the 145Cys alkyl-receptor residue (21) and to the conserved estrogen-receptor interacting helix (10) and may alter the function of this DNA repair gene.

SAT2/SSAT catalyzes the transfer of the acetyl group from acetyl-CoA to the N1 position of spermidine or spermine and has a critical role in governing intracellular polyamine concentrations, compounds involved in cell proliferation. *SAT2* is overexpressed in colorectal cancer cells (22) in the presence of a *K-Ras* mutation. The *SAT2* Arg¹²⁶Cys SNP, associated with increased risk in our study, may modulate susceptibility to colorectal cancer via alterations in polyamine metabolism especially in association with aspirin/NSAID usage (23).

In this study, the *RAD51* 4480 AG + GG, the *CYP19A1* 14872 AC + CC and the *PMS2* 260 GC + CC variants were associated with a significantly increased risk of colorectal cancer. *RAD51* is involved in homologous recombination of double-strand break repair in the same pathway (24). *PMS2* forms a MutL α heterodimer with *MLH1* in the mismatch repair complex. Mutations in mismatch repair have been associated with hereditary nonpolyposis colorectal cancer as well as with a subset of colon tumors. Aberrations in mismatch repair related to hereditary nonpolyposis colorectal cancer are typically characterized by mutations in the *MLH1* and *MSH2* genes and occasional mutations in *MSH6* (25). Our data suggest that the *PMS2* S260S CC genotype is associated with an increased risk of colorectal cancer, although it does not alter the protein sequence; LD with another gene variant may be responsible for this association. The *HFE* 462 TT, *LTA* C50T, *GPX3* G1961A (intron 1) GA + AA, and *GPX3* C14T (intron 4) genotype were associated with a reduced risk of colorectal cancer. *HFE* SNPs involved in iron metabolism have been associated with colorectal adenoma and cancer risk (26-28). The *CDKN2A* A185G variants and the *APAF* A2093G variants involved in the cell cycle and apoptosis, respectively, were also associated with colorectal cancer risk.

SNPs with P_{trend} Values <0.05. The nonsynonymous *COMT* Val¹⁵⁸Met polymorphism regulates *COMT* activity with the Met/Met variant associated with a 3- to 4-fold difference in function (intermediate phenotype observed in heterozygotes; ref. 29). This *COMT* Met/Met genotype may be inversely associated with colorectal cancer risk via the estrogen metabolism pathway (30, 31). Previous studies have found inverse associations with the Met/Met genotype for postmenopausal breast cancer but an increased risk for premenopausal breast cancer (32).

This study has several limitations. Large-scale association studies entail the performance of numerous statistical tests. Evaluating each test by an uncorrected threshold would yield a surplus of loci deemed significant due to chance. Therefore, we corrected for multiple testing using the Bonferroni and SNP spectral decomposition (for tests that are not independent because of high LD between the markers) approaches to reduce spurious associations. However, given our modest sample size, we had limited power to detect significant differences at these extreme *P* values.

With 358 cases of colorectal cancer, we were underpowered to evaluate gene-environment and gene-gene

interactions. Although the SNPs were selected among candidate genes, we did not use a haplotype-tagging approach. Therefore, the association with colorectal cancer risk is not definitive for many genes, for which only one to two SNPs were included in this study. Nevertheless, there were 145 nonsynonymous SNPs in this analysis, of which two loci were strongly associated with colorectal cancer risk. In addition, the low MAF of several SNPs may make accurate detection of a modest association with colorectal cancer risk difficult. Further, we recognize that susceptibility to colorectal cancer is influenced by genetic epistasis and determined by synergistic interactions between environmental carcinogens and allelic variants of multiple genes in numerous pathways. Last, the functional relevance of many of the polymorphisms examined in this study is unknown.

In summary, these data extend the current knowledge of genetic variation associated with colorectal cancer risk. SNP-based and gene-based approaches suggest that genetic variants in *MGMT* may be associated with colorectal cancer risk. Our study lends further support to the previously reported association of the *MGMT* *Ile*¹⁴³*Val*, in linkage with the *MGMT* *Lys*¹⁷⁸*Arg* SNP genotyped in this OPA, located near the 145Cys residue, with risk of colorectal cancer. In addition, we identified novel polymorphisms, including nonsynonymous SNPs *SAT* *Arg*¹²⁶*Cys* and *COMT* *Val*¹⁵⁸*Met*, associated with colorectal cancer risk. The *PMS2* *S260S* SNP is in a gene in the mismatch repair pathway, a major pathway with an established relation to colorectal cancer. In addition to replication of these findings in other populations, further investigation to establish the functional relationship of these SNPs with colorectal cancer is warranted. With advances in affordable genotyping technology and annotation of common human genetic variation, large-scale analyses such as this study have the potential to substantially clarify the inherited component of colorectal cancer risk.

Acknowledgments

We thank Pati Soule, Hardeep Ranu, and Craig Labadie for laboratory assistance and the dedicated participants of the NHS and the HPFS for ongoing commitment.

References

- American Cancer Society. Cancer facts and figures. 2006.
- Packer BR, Yeager M, Staats B, et al. SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. *Nucleic Acids Res* 2004;32:D528–32.
- A haplotype map of the human genome. *Nature* 2005;437:1299–320.
- Packer BR, Yeager M, Burdett L, et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res* 2006;34:D617–21.
- Steemers FJ, Gunderson KL. Illumina, Inc. *Pharmacogenomics* 2005;6:777–82.
- Foster CB, Aswath K, Chanock SJ, McKay HF, Peters U. Polymorphism analysis of six selenoprotein genes: support for a selective sweep at the glutathione peroxidase 1 locus (3p21) in Asian populations. *BMC Genet* 2006;7:56.
- Hughes AL, Packer B, Welch R, Chanock SJ, Yeager M. High level of functional polymorphism indicates a unique role of natural selection at human immune system loci. *Immunogenetics* 2005;57:821–7.
- Hunter DJ, Riboli E, Haiman CA, et al. A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. *Nat Rev Cancer* 2005;5:977–85.
- Tranah GJ, Giovannucci E, Ma J, Fuchs C, Hunter DJ. APC Asp1822Val and Gly2502Ser polymorphisms and risk of colorectal cancer and adenoma. *Cancer Epidemiol Biomarkers Prev* 2005;14:863–70.
- Tranah GJ, Bugni J, Giovannucci E, et al. O⁶-methylguanine-DNA methyltransferase Leu84Phe and Ile143Val polymorphisms and risk of colorectal cancer in the Nurses' Health Study and Physicians' Health Study (United States). *Cancer Causes Control* 2006;17:721–31.
- Garcia-Closas M, Malats N, Real FX, et al. Large-scale evaluation of candidate genes identifies associations between VEGF polymorphisms and bladder cancer risk. *PLoS Genet* 2007;3:e29.
- Tranah GJ, Lescault PJ, Hunter DJ, De Vivo I. Multiple displacement amplification prior to single nucleotide polymorphism genotyping in epidemiologic studies. *Biotechnol Lett* 2003;25:1031–6.
- Paynter RA, Skibola DR, Skibola CF, Buffler PA, Wiemels JL, Smith MT. Accuracy of multiplexed Illumina platform-based single-nucleotide polymorphism genotyping compared between genomic and whole genome amplified DNA collected from multiple sources. *Cancer Epidemiol Biomarkers Prev* 2006;15:2533–6.
- SAS. Genetics user's guide. Cary: SAS Institute, Inc.; 2002.
- Nyholt DR. A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. *Am J Hum Genet* 2004;74:765–9.
- Rothe C, Koszycki D, Bradwejn J, et al. Association of the Val158Met catechol O-methyltransferase genetic polymorphism with panic disorder. *Neuropsychopharmacology* 2006;31:2237–42.
- Sunyaev S, Ramensky V, Bork P. Towards a structural basis of human non-synonymous single nucleotide polymorphisms. *Trends Genet* 2000;16:198–200.
- Webb EL, Rudd MF, Sellick GS, et al. Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14 704 first-degree relatives. *Hum Mol Genet* 2006;15:3263–71.
- Lindahl T, Demple B, Robins P. Suicide inactivation of the *E. coli* O⁶-methylguanine-DNA methyltransferase. *EMBO J* 1982;1:1359–63.
- Pegg AE. Regulation of ornithine decarboxylase. *J Biol Chem* 2006;281:14529–32.
- Chueh LL, Nakamura T, Nakatsu Y, Sakumi K, Hayakawa H, Sekiguchi M. Specific amino acid sequences required for O⁶-methylguanine-DNA methyltransferase activity: analyses of three residues at or near the methyl acceptor site. *Carcinogenesis* 1992;13:837–43.
- Linsalata M, Giannini R, Notarnicola M, Cavallini A. Peroxisome proliferator-activated receptor γ and spermidine/spermine N1-acetyltransferase gene expressions are significantly correlated in human colorectal cancer. *BMC Cancer* 2006;6:191.
- Babbar N, Gerner EW, Casero RA, Jr. Induction of spermidine/spermine N1-acetyltransferase (SSAT) by aspirin in Caco-2 colon cancer cells. *Biochem J* 2006;394:317–24.
- Thacker J. The RAD51 gene family, genetic instability and cancer. *Cancer Lett* 2005;219:125–35.
- Modrich P. Mechanisms and biological effects of mismatch repair. *Annu Rev Genet* 1991;25:229–53.
- Stevens RG, Morris JE, Cordis GA, Anderson LE, Rosenberg DW, Sasser LB. Oxidative damage in colon and mammary tissue of the HFE-knockout mouse. *Free Radic Biol Med* 2003;34:1212–6.
- Shaheen NJ, Silverman LM, Keku T, et al. Association between hemochromatosis (HFE) gene mutation carrier status and the risk of colon cancer. *J Natl Cancer Inst* 2003;95:154–9.
- Chan AT, Ma J, Tranah GJ, et al. Hemochromatosis gene mutations, body iron stores, dietary iron, and risk of colorectal adenoma in women. *J Natl Cancer Inst* 2005;97:917–26.
- Mannisto PT, Kaakkola S. Catechol-O-methyltransferase (COMT): biochemistry, molecular biology, pharmacology, and clinical efficacy of the new selective COMT inhibitors. *Pharmacol Rev* 1999;51:593–628.
- Lotta T, Vidgren J, Tilgmann C, et al. Kinetics of human soluble and membrane-bound catechol O-methyltransferase: a revised mechanism and description of the thermolabile variant of the enzyme. *Biochemistry* 1995;34:4202–10.
- Dawling S, Roodi N, Mermaugh RL, Wang X, Parl FF. Catechol-O-methyltransferase (COMT)-mediated metabolism of catechol estrogens: comparison of wild-type and variant COMT isoforms. *Cancer Res* 2001;61:6716–22.
- Thompson PA, Shields PG, Freudenheim JL, et al. Genetic polymorphisms in catechol-O-methyltransferase, menopausal status, and breast cancer risk. *Cancer Res* 1998;58:2107–10.