Generalizability and Epidemiologic Characterization of Eleven Colorectal Cancer GWAS Hits in Multiple Populations

Jing He[1], **Lynne R. Wilkens**[2], **Daniel O. Stram**[1], **Laurence N. Kolonel**[2], **Brian E. Henderson**[1], **Anna H. Wu**[1], **Loic Le Marchand**[2], and **Christopher A. Haiman**[1]

[1] Department of Preventive Medicine, Keck School of Medicine, University of Southern California/ Norris Comprehensive Cancer Center, Los Angeles, California 90033, USA

[2] Epidemiology Program, Cancer Research Center of Hawaii, University of Hawaii, Honolulu, Hawaii 96813, USA

Abstract

Background—Genome-wide association studies (GWAS) in populations of European ancestry have identified several loci that confer an increased risk of colorectal cancer (CRC).

Methods—We studied the generalizability of the associations with 11 risk variants for CRC on 8q23 (rs16892766), 8q24 (rs6983267), 9p24 (rs719725), 10p14 (rs10795668), 11q23 (rs3802842), 14q22 (rs4444235), 15q13 (rs4779584), 16q22 (rs9929218), 18q21 (rs4939827), 19q13 (rs10411210) and 20p12 (rs961253) in a multiethnic sample of 2,472 CRC cases, 839 adenoma cases and 4,466 controls comprised of European American, African American, Native Hawaiian, Japanese American and Latino men and women. Because findings for CRC and adenoma were similar, we combined both groups in the analyses.

Results—We confirmed the associations with an increased risk of CRC/adenoma for the 8q24, 11q23 and 15q13 loci in European Americans, and observed significant associations between the 8q24 and 20p12 loci with CRC/adenoma risk in African Americans. Moreover, we found statistically significant cumulative effects of risk alleles on CRC/adenoma risk in all populations (odds ratio (OR) per allele = 1.07–1.09, p≤0.039) except in Japanese Americans (OR=1.01, p=0.52). We found heterogeneity in the associations by tumor subsite, age of CRC/adenoma onset, sex, body mass index (BMI) and smoking status for some of the variants.

Conclusions—These results provide evidence that the known variants are in aggregate significantly associated with CRC/adenoma risk in multiple populations except Japanese Americans, and the influences may differ across groups defined by clinicopathological characteristics for some variants.

Impact—These results underline the importance of studying the epidemiologic architecture of these genetic effects in large and diverse populations.

Keywords

Colorectal Cancer; Genetic Susceptibility; Cancer in minority; Underserved populations; Multiethnic Cohort

Correspondence to: Dr. Christopher Haiman, Associate Professor, Department of Preventive Medicine, Keck School of Medicine, University of Southern California, USC/Norris Comprehensive Cancer Center, Harlyne Norris Research Tower, 1450 Biggy Street, room 1504A, Los Angeles, California 90033 (haiman@usc.edu).

INTRODUCTION

Genome-wide association studies (GWAS) have identified several common single-nucleotide polymorphisms (SNPs) that confer risk for colorectal cancer (CRC) in populations of European ancestry. The variant rs6983267 within the chromosome 8q24 region was the first to be identified in multiple studies (1–3). Additional studies uncovered and validated other common risk variants, on 8q23 (rs16892766) (4), 9p24 (rs719725) (2,5), 10p14 (rs10795668) (4), 11q23 (rs3802842) (6), 15q13 (rs4779584) (7) and 18q21 (rs4939827) (6,8). A recent meta-analysis of two GWAS conducted in the UK revealed four new loci associated with CRC, on 14q22 (rs4444235), 16q22 (rs9929218), 19q13 (rs10411210) and 20p12 (rs961253) (9). Of the 11 variants, 5 are located in proximity of genes involved in TGF-β signaling (14q22, 15q13, 18q21, 19q13, 20p12) which highlights a previously implicated pathway important in the pathogenesis of CRC (6–9). Each of the identified CRC risk variants is common in European populations, with allele frequencies ranging from 7% to 49% (CEU Hapmap) and has modest effect sizes of 1.10 to 1.26 per allele (Supplemental Table 1). Together, these 11 loci explain <10% of the heritability of CRC in populations of European ancestry (9), which suggests that the vast majority of the inherited variations underlying risk of CRC have yet to be found.

A number of studies have characterized associations with subsets of these loci in populations of European ancestry (5,10–19); however, replication efforts in other racial/ethnic populations have been limited (3,20–22). The risk variant at 8q24 (rs6983267) is suggested to be a biologically functional allele (23,24) and we, and others, have reported rs6983267 to be a pan-ethnic marker of risk in multiple populations (3,20,22). Testing of all risk loci for CRC in non-European populations will be required to determine whether they may serve as markers of risk more broadly in the U.S. population and globally. While much emphasis has been placed on the low predictive value of common variants revealed through GWAS, modeling of their aggregate effects in multiple populations remains a crucial next step to characterize the risk conferred by these markers across racial/ethnic populations and identify subgroups of the population at greater risk.

In the present study, we evaluated the independent and aggregate effects of the 11 validated risk variants for CRC in four multiethnic studies which in total include 3,311 cases of CRC or adenoma and 4,466 controls comprised of European American, African American, Native Hawaiian, Japanese American and Latino men and women. Here we report on the associations of these risk variants with CRC risk by racial/ethnic group, disease subgroup (stage and anatomical subsite), and main CRC risk factors (age of onset, sex, BMI, smoking status, and first degree family history of colorectal cancer).

MATERIALS AND METHODS

Study Populations

The Multiethnic Cohort (MEC)—The MEC includes 215,251 men and women in Hawaii and Los Angeles (with additional African-Americans from elsewhere in California) (25). The cohort is comprised predominantly of African Americans, Japanese Americans, Native Hawaiians, Latinos and European Americans who entered the study between 1993 and 1996 by completing a 26-page self-administered questionnaire that requested detailed information about dietary habits, demographic factors, personal behaviors, history of medical conditions, family history of cancer, and for women, reproductive history and exogenous hormone use. The participants were between the ages 45 and 75 at recruitment. Incident cancers in the MEC are identified by linkage to population-based Surveillance, Epidemiology and End Results (SEER) cancer registries covering Hawaii and Los Angeles County, and to the California State Cancer Registry covering all of California. From the registries, information

about stage of disease and site of tumor (colon versus rectum) is available. Beginning in 1994, blood samples were collected from incident colorectal cancer cases and a random sample of MEC participants to serve as a control pool for genetic analyses in the cohort. Starting in 2001, blood samples were collected prospectively from ~70,000 cohort participants. Eligible cases in this CRC case-control study consisted of men and women with incident CRC (n=1,549) diagnosed after enrollment in the MEC through December 2005. Controls (n=2,173) were participants without colorectal, breast or prostate cancer prior to entry into the cohort and without a diagnosis of CRC up to December 2005. This study was approved by the Institutional Review Boards at the University of Hawaii and the University of Southern California.

The Los Angeles County CRC Case-Control Study (LA CRC)—The LA CRC is a population-based case-control study of CRC. Eligible cases included English-speaking women with a histologically confirmed incident colorectal cancer, diagnosed at ages 55 to 74 years, from January 1998 through December 2002 and who were residents of Los Angeles County at the time of diagnosis. Cases were identified from the Los Angeles County Cancer Surveillance Program. Controls were selected from the neighborhoods where cancer cases resided at the time of diagnosis. A personal interview and a blood sample were obtained from each subject. A reference date was de ned as 1 year before the date of diagnosis of the case. This same reference date was used for each case's matched control subject. This study includes 497 cases and 507 controls that were available for genotyping (26). This study was approved by the Institutional Review Board at the University of Southern California.

The Hawaii Multiethnic CRC Case-Control Study (HI CRC)—The HI CRC is a population-based case-control study in Hawaii and has been described in detail previously (27). Cases were identified through the Hawaii SEER registry and consisted of Japanese American, European American and Native Hawaiian residents of Oahu, Hawaii, who were newly diagnosed with colon or rectal cancer between January 1994 and August 1998. Controls were selected from participants in an ongoing population-based health survey conducted by the Hawaii State Department of Health and from Health Care Financing Administration participants. In-person interviews were conducted at the subjects' homes by trained interviewers. The median interval between diagnosis and interview for cases was 4.5 months (25th–75th percentiles, 3.3–8.4 months). The questionnaire included detailed information on demographics, including the race of each grandparent; a quantitative food frequency questionnaire; a lifetime history of tobacco, alcohol, and aspirin use; a history of recreational sports activities since age18; a personal history of various relevant medical conditions; a family history of CRC in parents and siblings; information on height and weight at different ages; and for women, a history of reproductive events and hormone use. This study includes 426 cases and 616 controls with DNA for genetic analysis. This study was approved by the Institutional Review Board at the University of Hawaii.

The Adenoma Study—For the adenoma study (28), two flexible-sigmoidoscopy screening clinics were first used to recruit participants on Oahu, Hawaii. Adenoma cases were identified either from the baseline examination at the Hawaii site of the Prostate Lung Colorectal and Ovarian (PLCO) cancer screening trial (1996–2000) or at the Kaiser Permanente Hawaii (KPH)'s Gastroenterology Screening Clinic (1995–2007). In addition, from 2002 to 2007, we recruited all eligible colonoscopy patients in the KPH Gastroenterology Department. Cases were patients with histologically confirmed first-time colorectal adenoma(s), of Japanese, Caucasian or Hawaiian race/ethnicity. Controls were selected among patients with a normal colorectum, and were matched to the cases on age at exam, sex, race/ethnicity, screening date (±3 months), clinic and type of examination

(colonoscopy or flexible sigmoidoscopy). Exposure information was collected via interview-administered questionnaires designed to obtain demographic and lifestyle information, including lifetime histories of physical activity, tobacco smoking, and alcohol drinking; medical history; family cancer history; and, for females, reproductive and hormone use history. The interview also included a validated food frequency questionnaire with 268 food items, and a detailed history of vitamin and mineral supplement use was also taken. The median interval between endoscopy and interview was 11.3 months (25th–75th percentiles, 5.2–21.1 months). This study includes 839 cases and 1170 controls available for genetic analysis. This study was approved by the Institutional Review Board at the University of Hawaii.

Altogether, the present analysis included 3,311 CRC/adenoma cases and 4,466 controls (European American (1,171/1,534), African American (382/510), Native Hawaiian (323/472), Japanese American (1,042/1,426) and Latino (393/524)).

Genotyping

Genotyping was conducted by the TaqMan allelic discrimination assay (Applied Biosystems, Foster City, CA) (29). For all SNPs, genotype call rates were >95% among case and control groups in each population. Individuals who were missing information on 4 or more SNPs were excluded from the analysis (197 individuals were excluded). Blinded replicates (5–10%) were included within and across all 96-well DNA plates. For all the SNPs, the concordance rates for the duplicates QC samples were 99.4% in MEC, 99.7% in HI CRC and Adenoma study, and were 100% in the LA CRC study.

Statistical Analysis

Genotype frequencies in the controls were tested for departure from Hardy-Weinberg Equilibrium (HWE) in each racial/ethnic group in each study. Odds ratios (OR) and 95% confidence intervals (95% CI) were estimated for each variant using unconditional logistic regression. The "low-risk" allele based on previous GWAS reports was used as the reference allele (Supplemental Table 1). The effect of each SNP on disease risk was evaluated under a log-additive mode of inheritance, as well as using separate indicator variables to allow unrestricted effects of heterozygotes and homozygotes of the risk allele. For each SNP, we performed a 1-df likelihood ratio test (LRT) to compare the additive to the unrestricted model. For all SNPs, the unrestricted model which allows dominant and recessive effects was not significantly better than the additive model. The statistical power of capturing the ORs reported in previous GWAS were calculated using Quanto (version 1.2.4).

We also performed analyses to model the cumulative effect of all 11 variants on disease risk. In these analyses, each individual was assigned a risk score, which was the total number of risk alleles carried for the 11 variants. For individuals missing data for <4 SNPs (n=970), the average number of risk alleles for a SNP in each population was used in the calculation of the risk score. ORs were estimated to evaluate the per allele effect of the risk score on CRC risk. We also categorized the risk score in quintiles based on the risk score distribution in controls in each racial/ethnic group. Results for the risk score analyses were not materially changed when we restricted it to individuals with complete genotype data for all 11 SNP (2909 of 3311 cases and 3898 of 4466 controls).

We tested whether dominant or recessive effects for each SNP contributed additional information on top of the risk score by including indicator variables for heterozygotes and homozygotes in the risk score model for each SNP one by one. This model was compared to the model with just the risk score using a 2-df LRT. As with the additive model, allowing dominant or recessive genetic effects for any SNP did not improve the model fit. Also, to

incorporate known genetic effect size into the assessment of cumulative effects, we created a weighted risk score, in which SNPs were weighted by their log (OR), based on ORs reported in the GWAS in European Americans (Supplemental Table 1). The two scores were highly correlated (Pearson correlation coefficient was 0.95, P<0.0001) and the unweighted risk score was used to estimate cumulative genetic effects in all analyses.

Case-only analyses were performed to test whether individual and aggregate effects of SNPs were different on CRC versus adenoma risk. For invasive CRC, we conducted case-only analyses for each SNP and the risk score by tumor stage at diagnosis (localized vs. regional/distant), for colon and rectal cancer, and by location in the colon (left vs. right). In addition to race/ethnicity, we examined whether the associations were modified by sex, age of onset (based on the median in the controls: <66 and ≥66 years), first-degree family history of colorectal cancer, BMI (<23, ≥23 ≥25, ≥25–30, ≥30 kg/m^2) and smoking status (never and ever smokers) using a LRT of heterogeneity by adding interaction term to the model.

Genetic effects were adjusted for age (categorized by quartiles of the age distribution in controls in the pooled sample: <59, ≥59–66, ≥66–72, ≥72 and unknown), study, and race/ethnicity in pooled analyses. For each covariate, an indicator variable was used for those individuals with missing data. To assess the potential confounding effects of population stratification, we conducted a principal component (PCA) analysis using genotype data for 1389 SNPs from candidate genes studies among 1098 CRC cases and 1489 controls from the MEC (30–32). In testing each CRC variant, we included the first 10 principal components (PCs) along with self-reported ethnicity in the logistic regression models to assess the impact of population stratification.

RESULTS

Descriptive characteristics of participants are presented in Table 1 by study. The ages of individuals ranged from 21 to 86 years, and in general, mean ages were similar between cases and controls within each study. The Native Hawaiians were the youngest group (mean age, 61 years) and African Americans the oldest (mean age, 67 years).

Ten of the 11 SNPs were polymorphic among controls in each racial/ethnic group (Figure 1). SNP rs16892766 was monomorphic in Japanese Americans and had a minor allele frequency (MAF) <10% in European Americans, Native Hawaiians and Latinos, while SNP rs10795668 had a MAF of 7% in African Americans. All other SNPs had a MAF >10% in each racial/ethnic group. Allele frequencies differed substantially across ethnic groups, with variations ranged from 12% for rs16892766 to 63% for rs4779584 (Table 2 and Figure 1). None of the SNPs displayed departure from HWE within study/ethnic group (P>0.01).

In ethnic-pooled case-only analysis comparing CRC (n=2,472) and adenoma cases (n=839) we observed no significant association between any SNP and case group (Supplemental Table 2). For each SNP the OR estimates were not materially different when combining in situ CRC (n=202 cases) with invasive CRC cases compared to invasive cases alone (results not presented). Based on these observations, we estimated the individual and aggregate genetic effects of the CRC variants among all cases combined. Within an ethnic group, the allele frequencies of the 11 SNPs were consistent across studies. We also observed no strong evidence of significant within-ethnic-group heterogeneity in the association of each SNP with CRC/adenoma risk between studies (only one SNP rs3802842 has P$_{het}$<0.01 among European Americans) (Supplemental Table 3).

In European Americans, the directions of association with the 11 variants were generally consistent with previous GWAS reports, positive associations were observed with the risk alleles of 9 of the 11 SNPs (8 with OR≥1.05) and the other 2 SNPs had ORs very close to

one (Table 2). Nominally significant associations were observed with rs6983267 (OR=1.12, 95% CI: 1.01–1.25; p=0.038), rs3802842 (OR=1.28, 95% CI: 1.14–1.44; p=4×10^{-5}) and rs4779584 (OR=1.20, 95% CI: 1.04–1.37; p=0.010), while rs16892766 also contributed to elevated disease risk, but the association was only borderline significant (OR=1.18, 95% CI: 0.97–1.43, p=0.095). The effect sizes for the other 5 SNPs were relatively small (OR<1.10) and did not reach statistical significance.

Positive associations (OR>1) with CRC/adenoma were also noted with 9 SNPs in Latinos (8 with OR≥1.05) and with 8 SNPs in each African Americans (5 with OR≥1.05) and Native Hawaiians (7 with OR≥1.05; Table 2). However, the associations were weaker in the four non-European populations, with only SNPs rs6983267 at 8q24 (OR=1.52, 95% CI: 1.11–2.07, p=0.0090 (3)) and rs961253 at 20p12 (OR=1.24, 95% CI: 1.01–1.53, p=0.042), found to be significantly associated with risk in African Americans (Table 2). Interestingly, in Japanese Americans (1,042 cases, 1,426 controls), the second largest group after European Americans, only 2 SNPs had effect estimates >1.05 (rs6983267 at 8q24 and rs4444235 at 14q22) with only the OR for rs6983267 approaching statistical significance (1.12, 95% CI: 0.99–1.26). Tests for heterogeneity suggested no difference in genetic effects across racial/ethnic groups for these 11 SNPs (Table 2). The association by racial/ethnic group and genotype are presented in Supplemental Table 4. In pooling all populations, the pattern of associations was similar to what was observed in European Americans, with significant positive associations observed for rs16892766, rs6983267, rs3802842, and the effect of rs4779584 became weaker than that observed in European Americans (Table 2). After removing European Americans from the pooled analysis, the associations were positive for 10 of the 11 SNPs, but only rs6983267 remained significantly associated with increased CRC/adenoma risk (data not shown; OR=1.16, 95% CI: 1.06–1.27, p=0.0011).

The mean number of risk alleles and the distributions of the risk score were nearly identical across racial/ethnic groups (Table 3; Supplemental Figure 1 and 2). In all racial/ethnic groups the risk score was associated with an OR of 1.06 to 1.09 per allele (P's≤0.039) except in Japanese Americans (OR=1.01, 95% CI: 0.97–1.06, p=0.52; Table 3). A significant difference in the effect of the risk score on CRC/adenoma risk was detected between Japanese Americans and European Americans (1-df test, P_{het}=0.036), although there was no evidence of overall heterogeneity of the risk score effect across all racial/ethnic groups (4-df test, P_{het}=0.20; Table 3). In all populations except Japanese Americans (OR=1.10, p=0.43), CRC/adenoma risk for individuals in the highest quintile of the risk score was significantly elevated by a similar magnitude compared to those in the lowest quintile, with ORs ranging from 1.52 (p=0.036) in Latinos to 1.58 (p=9.38×10^{-5}) in European Americans (Table 3).

In the MEC, results were similar following adjustment for principal components versus adjusting only for self-reported ethnicity (Supplemental Table 5).

In ethnic-pooled case-only analyses, SNPs rs3802842 and rs4444235 were more associated with rectal cancer than with colon cancer (OR=1.15 for rectal versus colon cancer, 95% CI: 1.00–1.33, P=0.054 for rs3802842; OR=1.18, 95% CI: 1.03–1.35, P=0.020 for rs4444235; Supplemental Table 2). No significantly different patterns of association were observed for left versus right colon cancer, or by CRC stage (localized vs. distant/regional; Supplemental Table 2) for any of the variant. Test for interaction revealed significant effect modification by age, with the effects of rs3802842, rs4444235 and rs4779584 being stronger on CRC/adenoma risk at younger ages (before age 66) (P_{het}=0.041 for rs3802842; P_{het}=0.046 for rs4444235; P_{het}=0.0079 for rs4779584); while the effect of rs6983267 (P_{het}=0.013) and rs961253 (P_{het}=0.016) was stronger at older ages (after age 65) (Supplemental Table 6). The associations for rs10411210 (P_{het}=0.021) and rs9929218 (P_{het}=0.035) were stronger in men

than women. The association for rs4444235 was stronger in nonsmokers than smokers (P_{het}=0.033). Significant interaction with BMI was observed for rs4779584 (P_{het}=0.015), with stronger effects for group with BMI<23. No significantly different patterns of association were observed by family history of colorectal cancer (Supplemental Table 6) for any of the variant. Associations with risk score were similar by tumor site and stage (Supplemental Table 7), and no evidence of heterogeneity in associations of risk score was observed with age, sex, family history of colorectal cancer, BMI or smoking status overall (Supplemental Table 8) or in any population (results not shown).

DISCUSSION

In our study, we were able to detect statistically significant associations between rs6983267 (8q24), rs3802842 (11q23), rs4779584 (15q13) and risk of colorectal neoplasia in European Americans, which were consistent with the reports from GWAS of CRC (1,6,7) and various subsequent replication studies (3,5,10,12,15–19). The ORs estimated in European Americans in our study for rs6983267 and rs4779584 were reduced, and the OR for rs3802842 was increased compared to the ORs described in the original GWAS; however, the 95% CI overlapped substantially, suggesting that the differences were probably not significant. None of the three SNPs are located within known genes. SNP rs6983267 has been implicated in regulating MYC while rs4779584 is near *GREM1* which is a component of the TGF-β super-family signaling pathway. Further work will be required to better understand the clinical and biological implications of these associations.

We failed to replicate the association between rs10795668, rs719725, rs4939827, rs16892766, rs10411210, rs4444235, rs961253 and rs9929218 and CRC/adenoma risk in European Americans. Significant associations with variants rs9929218, rs10411210, rs961253 and rs4444235 were also not replicated in a study of 1786 cases and 1749 controls of Swedish origin (10), however, associations with rs10795668, rs4939827, rs6983267 and rs719715 have been confirmed in other studies in Europeans(5,10,13,18). Failure to replicate in our study may be explained by a lack of power. For the 8 SNPs that we failed to replicate in European Americans, the power of detecting the ORs reported in GWAS was 84% for rs4939827 and between 34% and 62% for the other 7 SNPs (Supplemental Table 9).

As we had found previously, rs6983267 was associated with elevated CRC/adenoma risk in multiple populations (3). However these observations are not independent as there is substantial overlap (~55%) between the cases and controls in our previous report and in the current study. The association with rs6983267 was statistically significant in African Americans, with the risk estimates in the other non-European ethnic groups of borderline significance. We also found a previously unreported significant association between rs961253 and CRC/adenoma risk in African Americans. No other statistically significant associations were found in non-European populations.

The failure to detect significant associations in non-European populations may be due to limited power. Except for rs4779584 and rs6983267 in Japanese Americans, the power to detect a nominally significant association in each non-European population was <80%, with the average power in each population for the variants tested ranging from 23% in Native Hawaiians to 47% in Japanese Americans (Supplemental Table 9). Lack of statistically significant associations in non-European populations may also be due to differences in linkage disequilibrium (LD) patterns across ethnic groups, with the risk variants being linked with functional variants in Europeans but not in all populations. Of the 11 SNPs, only rs6983267 has been implicated as being a functional variant (23, 24). However the specific biological mechanisms of all SNPs that have been associated with colorectal cancer risk remain unknown.

Except for Japanese Americans, cases tend to carry more risk alleles than controls, and moreover, there was a gradual increase in the CRC/adenoma OR with an increased number of risk alleles in all ethnic groups. Similar cumulative effects of low-risk variants have been described in other studies (10,14,18,22).

We noted that both the individual and aggregate genetic effects of the CRC risk variants were weaker or absent in Japanese Americans compared with other racial/ethnic groups. Despite the relatively large sample size, only two SNPs, rs6983267 and rs444235, were found to confer more than a 5% increase in CRC/adenoma risk in Japanese Americans. Moreover, no enrichment in the number of risk alleles was found in Japanese American CRC/adenoma cases. Although this could be again due to lack of power and different LD patterns, the weak genetic effects were consistent with the fact that Japanese were traditionally a low risk population for colorectal cancer. However, their risk increased markedly upon migration to the U.S. and during recent decades in Japan, possibly reflecting a particularly strong susceptibility to the environmental exposures associated with a western lifestyle (27). Since GWAS hits were identified based on the reproducibility of their associations with CRC, they are unlikely to be modifiedby environmental risk factors, the frequencies of which do vary across populations.

Heterogeneity in the associations of the SNPs with CRC by clinicopathological characteristics has been described in a number of previous studies, including stronger relationship between rectal cancer and rs10795668(4, 22), rs3802842(6, 14) and rs4939827(19); between age of disease onset and rs16892766(4), rs6983267, rs10795668, rs10411210 (10); between sex and rs4939827(13) and rs9929218(9); between family history of CRC and rs6983267(1), rs10795668 (10); and between CRC stage and rs3802842 (22). However, past results have been inconsistent. In our pooled population, we noted heterogeneity of association by tumor site (rectum and colon), age of disease onset, sex, BMI and smoking status for some of the SNPs. We were unable to detect any significant difference in association patterns by tumor stage or by family history of colorectal cancer. Middeldorp et al (18) reported enrichment in the number of risk alleles in patients with a family history of colorectal cancer compared with solitary CRC patients, and in patients with early-onset disease (age ≤50) compared with patients with late-onset of CRC (age >50). We did not find any relationship between number of CRC risk alleles and these risk factors or any other clinicopathological characteristics. Based on the number of comparisons these observations will need to be confirmed in other studies.

In this large multiethnic study, we confirmed the associations between SNPs on 8q24, 11q23 and 15q13 and CRC/adenoma risk in European Americans, and observed associations for SNPs on 8q24 and 20p12 in African Americans. Despite many SNPs failing to reach statistical significance in individual analyses, we found significant cumulative effects of risk alleles on CRC/adenoma across populations except for Japanese Americans. The genetic effects differed by tumor site, age of onset, sex, BMI and smoking status for some of the variants. Since many comparisons were tested caution is needed in interpreting these results. Our results indicate that the risk alleles identified in many European studies are in aggregate associated with CRC/adenoma risk in most populations with the exception of Japanese Americans. This underscores both the role that common variants play in CRC/adenoma and the need to extend study to diverse populations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. 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. [PubMed: 17618284]

2. Zanke BW, Greenwood CM, Rangrej J, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. Nat Genet. 2007; 39:989–94. [PubMed: 17618283]

3. Haiman CA, Le Marchand L, Yamamato J, et al. A common genetic risk factor for colorectal and prostate cancer. Nat Genet. 2007; 39:954–6. [PubMed: 17618282]

4. Tomlinson IP, Webb E, Carvajal-Carmona L, et al. A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. Nat Genet. 2008; 40:623–30. [PubMed: 18372905]

5. Poynter JN, Figueiredo JC, Conti DV, et al. Variants on 9p24 and 8q24 are associated with risk of colorectal cancer: results from the Colon Cancer Family Registry. Cancer Res. 2007; 67:11128–32. [PubMed: 18056436]

6. Tenesa A, Farrington SM, Prendergast JG, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. Nat Genet. 2008; 40:631–7. [PubMed: 18372901]

7. Jaeger E, Webb E, Howarth K, et al. Common genetic variants at the CRAC1 (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. Nat Genet. 2008; 40:26–8. [PubMed: 18084292]

8. Broderick P, Carvajal-Carmona L, Pittman AM, et al. A genome-wide association study shows that common alleles of SMAD7 influence colorectal cancer risk. Nat Genet. 2007; 39:1315–7. [PubMed: 17934461]

9. Houlston RS, Webb E, Broderick P, et al. Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. Nat Genet. 2008; 40:1426–35. [PubMed: 19011631]

10. von Holst S, Picelli S, Edler D, et al. Association studies on 11 published colorectal cancer risk loci. Br J Cancer. 2010; 103:575–80. [PubMed: 20648012]

11. Li L, Plummer SJ, Thompson CL, et al. A common 8q24 variant and the risk of colon cancer: a population-based case-control study. Cancer Epidemiol Biomarkers Prev. 2008; 17:339–42. [PubMed: 18268117]

12. Pittman AM, Broderick P, Sullivan K, et al. CASP8 variants D302H and -652 6N ins/del do not influence the risk of colorectal cancer in the United Kingdom population. Br J Cancer. 2008; 98:1434–6. [PubMed: 18362937]

13. Thompson CL, Plummer SJ, Acheson LS, Tucker TC, Casey G, Li L. Association of common genetic variants in SMAD7 and risk of colon cancer. Carcinogenesis. 2009; 30:982–6. [PubMed: 19357349]

14. Pittman AM, Webb E, Carvajal-Carmona L, et al. Refinement of the basis and impact of common 11q23.1 variation to the risk of developing colorectal cancer. Hum Mol Genet. 2008; 17:3720–7. [PubMed: 18753146]

15. Berndt SI, Potter JD, Hazra A, et al. Pooled analysis of genetic variation at chromosome 8q24 and colorectal neoplasia risk. Hum Mol Genet. 2008; 17:2665–72. [PubMed: 18535017]

16. Tuupanen S, Niittymaki I, Nousiainen K, et al. Allelic imbalance at rs6983267 suggests selection of the risk allele in somatic colorectal tumor evolution. Cancer Res. 2008; 68:14–7. [PubMed: 18172290]

17. Schafmayer C, Buch S, Volzke H, et al. Investigation of the colorectal cancer susceptibility region on chromosome 8q24.21 in a large German case-control sample. Int J Cancer. 2009; 124:75–80. [PubMed: 18839428]

18. Middeldorp A, Jagmohan-Changur S, van Eijk R, et al. Enrichment of low penetrance susceptibility loci in a Dutch familial colorectal cancer cohort. Cancer Epidemiol Biomarkers Prev. 2009; 18:3062–7. [PubMed: 19843678]

19. Curtin K, Lin WY, George R, et al. Cancer Epidemiol Biomarkers Prev. 2009; 18:616–21. [PubMed: 19155440]

20. Matsuo K, Suzuki T, Ito H, et al. Association between an 8q24 locus and the risk of colorectal cancer in Japanese. BMC Cancer. 2009; 9:379. [PubMed: 19857256]

21. Kupfer SS, Torres JB, Hooker S, et al. Novel single nucleotide polymorphism associations with colorectal cancer on chromosome 8q24 in African and European Americans. Carcinogenesis. 2009; 30:1353–7. [PubMed: 19520795]

22. Xiong F, Wu C, Bi X, et al. Risk of genome-wide association study-identified genetic variants for colorectal cancer in a chinese population. Cancer Epidemiol Biomarkers Prev. 2010; 19:1855–61. [PubMed: 20530476]

23. Pomerantz MM, Ahmadiyeh N, Jia L, et al. The 8q24 cancer risk variant rs6983267 shows long-range interaction with MYC in colorectal cancer. Nat Genet. 2009; 41:882–4. [PubMed: 19561607]

24. Tuupanen S, Turunen M, Lehtonen R, et al. The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. Nat Genet. 2009; 41:885–90. [PubMed: 19561604]

25. Kolonel LN, Henderson BE, Hankin JH, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol. 2000; 151:346–57. [PubMed: 10695593]

26. Wu AH, Siegmund KD, Long TI, et al. Hormone therapy, DNA methylation and colon cancer. Carcinogenesis. 2010; 31:1060–7. [PubMed: 20064828]

27. Le Marchand L, Hankin JH, Wilkens LR, et al. Combined effects of well-done red meat, smoking, and rapid N-acetyltransferase 2 and CYP1A2 phenotypes in increasing colorectal cancer risk. Cancer Epidemiol Biomarkers Prev. 2001; 10:1259–66. [PubMed: 11751443]

28. Hankin JH, Wilkens LR, Kolonel LN, Yoshizawa CN. Validation of a quantitative diet history method in Hawaii. Am J Epidemiol. 1991; 133:616–28. [PubMed: 2006649]

29. Lee LG, Connell CR, Bloch W. Allelic discrimination by nick-translation PCR with fluorogenic probes. Nucleic Acids Res. 1993; 21:3761–6. [PubMed: 8367293]

30. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9. [PubMed: 16862161]

31. Reich D, Price AL, Patterson N. Principal component analysis of genetic data. Nat Genet. 2008; 40:491–2. [PubMed: 18443580]

32. Serre D, Montpetit A, Pare G, et al. Correction of population stratification in large multi-ethnic association studies. PLoS One. 2008; 3:e1382. [PubMed: 18196181]
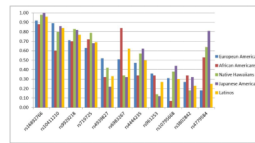
Figure 1.
Risk allele frequencies in controls across racial/ethnic groups, in descending ordered based on frequency in European Americans.

Table 1

Descriptive characteristics of each study[*]

	Adenoma		LA CRC		HI CRC		MEC		All	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
Total subjects	839	1170	497	507	426	616	1549	2173	3311	4466
Sex (%)										
Men	515(0.61)	733(0.63)	0	0	255(0.60)	355(0.58)	853(0.55)	1132(0.52)	1623(0.49)	2220(0.50)
Women	324(0.39)	437(0.37)	497(1.00)	507(1.00)	171(0.40)	261(0.42)	696(0.45)	1041(0.48)	1688(0.51)	2246(0.50)
Race (%)										
African Americans	0	0	67(0.13)	52(0.10)	0	0	315(0.20)	458(0.21)	382(0.12)	510(0.11)
Native Hawaiians	181(0.21)	233(0.20)	0	0	65(0.15)	83(0.13)	77(0.05)	156(0.07)	323(0.10)	472(0.11)
Japanese Americans	267(0.32)	380(0.32)	0	0	233(0.55)	373(0.61)	542(0.35)	673(0.31)	1042(0.31)	1426(0.32)
Latinos	0	0	59(0.12)	43(0.09)	0	0	334(0.22)	481(0.22)	393(0.12)	524(0.12)
European Americans	391(0.47)	557(0.48)	371(0.75)	412(0.81)	128(0.30)	160(0.26)	281(0.18)	405(0.19)	1171(0.35)	1534(0.34)
Age										
Mean(sd)	59.9(8.6)	61.9(8.4)	65.6(5.7)	64.6(6.8)	64.3(11.8)	65.0(11.6)	68.4(8.2)	66.8(8.1)	65.3(9.2)	65.1(8.9)
No. unknown	0	0	0	0	0	0	28	0	28	0
Site										
Unknown	--	--	0	--	2	--	28	--	30	--
Rectum	--	--	59	--	120	--	413	--	592	--
Colon	--	--	438	--	304	--	1108	--	1850	--
Left colon	--	--	187	--	169	--	509	--	865	--
Right colon	--	--	248	--	134	--	587	--	969	--
Other colon	--	--	3	--	1	--	12	--	16	--
Stage										
Unknown	--	--	1	--	2	--	28	--	31	--
In situ	--	--	30	--	42	--	130	--	202	--
Invasive	--	--	466	--	382	--	1391	--	2239	--
Localized	--	--	201	--	194	--	694	--	1089	--
Distant/regional	--	--	264	--	182	--	594	--	1040	--
Unknown	--	--	1	--	6	--	103	--	110	--

	Adenoma		LA CRC		HI CRC		MEC		All	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
First-degree family history of colorectal cancer (%)										
Yes	151(0.18)	160(0.14)	65(0.13)	57(0.11)	74(0.17)	62(0.10)	182(0.12)	190(0.09)	472(0.14)	469(0.11)
No	688(0.82)	1010(0.86)	432(0.87)	450(0.89)	352(0.83)	554(0.90)	1367(0.88)	1983(0.91)	2839(0.86)	3997(0.89)
Smoking status (%)										
Never	364(0.43)	620(0.53)	221(0.45)	255(0.5)	174(0.41)	289(0.47)	532(0.34)	933(0.43)	1291(0.39)	2097(0.47)
Past	370(0.44)	473(0.4)	215(0.43)	197(0.39)	182(0.43)	260(0.42)	683(0.44)	912(0.42)	1450(0.44)	1842(0.41)
Current	105(0.13)	77(0.07)	61(0.12)	55(0.11)	67(0.16)	62(0.1)	234(0.15)	299(0.14)	467(0.14)	493(0.11)
Unknown	0	0	0	0	3(0.01)	5(0.01)	100(0.06)	29(0.01)	103(0.03)	34(0.01)
BMI (kg/m²) (%)										
<23	125(0.15)	256(0.22)	108(0.22)	122(0.24)	149(0.35)	195(0.32)	327(0.21)	536(0.25)	709(0.21)	1109(0.25)
≥23—25	150(0.18)	213(0.18)	86(0.17)	95(0.19)	86(0.20)	120(0.19)	268(0.17)	445(0.20)	590(0.18)	873(0.20)
≥25—30	324(0.39)	447(0.38)	160(0.32)	173(0.34)	128(0.30)	219(0.36)	577(0.37)	816(0.38)	1189(0.36)	1655(0.37)
≥30	240(0.29)	254(0.22)	143(0.29)	117(0.23)	63(0.15)	82(0.13)	277(0.18)	356(0.16)	723(0.22)	809(0.18)
Unknown	0	0	0	0	0	0	100(0.06)	20(0.01)	100(0.03)	20(0.004)

*
Adenoma: The Adenoma Study; LA CRC: The Los Angeles County CRC Case-Control Study; HI CRC: The Hawaii Multiethnic CRC Case-Control Study; MEC: The Multiethnic Cohort.

Table 2

The association of known CRC variants with CRC/adenoma risk by race/ethnicity[*]

SNP/Allele Tested[†]	Chr./Position[†]	Location/Nearest Gene	EA 1171 cases 1534 controls	AA 382 cases 510 controls	NH 323 cases 472 controls	JA 1042 cases 1426 controls	LA 393 cases 524 controls	Pooled 3311 cases 4466 controls
rs16892766	8q23.3	Intergenic	1.18 (0.97–1.43)	1.23 (0.92–1.63)	1.14 (0.59–2.21)	--	1.29 (0.82–2.05)	1.20 (1.04–1.39)
RAF for C	117699995	EIF3H	0.077	0.12	0.020	--	0.040	0.047
P_{het}[‡]								0.99
rs6983267	8q24.21	Intergenic	1.12 (1.01–1.25)	1.52 (1.11–2.07)	1.18 (0.95–1.46)	1.12 (0.99–1.26)	1.13 (0.92–1.39)	1.15 (1.07–1.23)
RAF for G	128482487	MYC	0.51	0.84	0.33	0.32	0.62	0.47
P_{het}								0.54
rs719725	9p24	Intergenic	0.99 (0.88–1.11)	1.06 (0.84–1.33)	1.21 (0.93–1.57)	0.98 (0.87–1.11)	0.96 (0.78–1.19)	1.00 (0.93–1.08)
RAF for A	6355683	TPD52L3	0.63	0.72	0.80	0.68	0.70	0.68
P_{het}								0.81
rs10795668	10p14	Intergenic	1.05 (0.93–1.18)	0.93 (0.64–1.33)	1.20 (0.97–1.47)	1.01 (0.90–1.13)	1.02 (0.83–1.26)	1.05 (0.97–1.12)
RAF for G	8741225	BC031880	0.69	0.93	0.62	0.56	0.70	0.67
P_{het}								0.55
rs3802842	11q23	Intergenic	1.28 (1.14–1.44)	1.11 (0.90–1.36)	1.27 (0.98–1.65)	1.02 (0.91–1.15)	1.07 (0.85–1.34)	1.15 (1.07–1.23)
RAF for C	110676919	FLJ45803	0.27	0.34	0.18	0.32	0.23	0.28
P_{het}								0.11
rs4444235	14q22.2	Intergenic	1.05 (0.94–1.17)	1.02 (0.83–1.25)	0.97 (0.79–1.18)	1.08 (0.96–1.21)	0.98 (0.82–1.19)	1.04 (0.97–1.11)
RAF for C	53480669	BMP4	0.47	0.34	0.57	0.62	0.50	0.52
P_{het}								0.84
rs4779584	15q13.3	Intergenic	1.20 (1.04–1.37)	1.03 (0.85–1.25)	1.02 (0.82–1.27)	0.94 (0.81–1.08)	1.13 (0.92–1.39)	1.07 (0.99–1.15)
RAF for T	30782048	GREM1, SCG5	0.18	0.54	0.64	0.81	0.25	0.48
P_{het}								0.18
rs9929218	16q22.1	Intron	0.99 (0.88–1.11)	1.01 (0.82–1.26)	0.92 (0.71–1.19)	0.98 (0.84–1.13)	1.09 (0.87–1.36)	0.99 (0.92–1.07)

SNP/Allele Tested[†]	Chr./Position[†]	Location/Nearest Gene	EA 1171 cases 1534 controls	AA 382 cases 510 controls	NH 323 cases 472 controls	JA 1042 cases 1426 controls	LA 393 cases 524 controls	Pooled 3311 cases 4466 controls
RAF for G	67378447	CDH1	0.71	0.70	0.83	0.82	0.77	0.76
P_{het}								0.91
rs4939827	18q21.1	Intron	1.05 (0.94–1.17)	0.98 (0.79–1.21)	1.14 (0.92–1.40)	0.99 (0.86–1.14)	1.13 (0.92–1.40)	1.05 (0.98–1.12)
RAF for T	44707461	SMAD7	0.52	0.32	0.42	0.22	0.33	0.37
P_{het}								0.67
rs10411210	19q13.1	Intron	1.06 (0.89–1.26)	0.99 (0.81–1.20)	0.92 (0.72–1.18)	1.00 (0.86–1.17)	1.15 (0.87–1.52)	1.02 (0.93–1.11)
RAF for C	38224140	RHPN2	0.89	0.60	0.81	0.86	0.85	0.83
P_{het}								0.82
rs961253	20p12.3	Intergenic	1.03 (0.92–1.15)	1.24 (1.01–1.53)	1.12 (0.83–1.50)	0.93 (0.78–1.12)	1.15 (0.93–1.42)	1.05 (0.97–1.14)
RAF for A	6352281	BMP2	0.36	0.33	0.14	0.12	0.27	0.25
P_{het}								0.35

*
Each cell gives odds ratios (and 95% confidence intervals) for allele dosage effects along with the risk allele frequency in controls. Odds ratios for CRC/adenoma are adjusted for age (quartiles), ethnicity (in pooled) and study. EA: European Americans; AA: African Americans; NH: Native Hawaiians; JA: Japanese Americans; LA: Latinos.

†
NBCI build 36.1, forward strand

‡
P value for heterogeneity of allele dosage effects across ethnic groups (4-df test)

Table 3

The risk of CRC/Adenoma associated with an aggregate risk score * by race/ethnicity

Mean risk score (range)	cases controls	EA	AA	NH	JA	LA	Pooled
		10.9(5–18)	11.8(7–18)	11.1(6–17)	10.7(4–17)	10.8(5–15)	11.0(4–18)
		10.6(4–18)	11.6 (6–18)	10.7(5.87–16)	10.7(4–17)	10.9(5–17.26)	10.7(4–18)
Quintiles[†]							
Q1	n (cases/ctrls)	271/423	92/139	64/122	260/361	97/159	784/1204
	OR(95% CI)[‡]	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)	1.00 (Ref)
Q2	n (cases/ctrls)	205/272	74/98	54/105	209/295	73/98	615/868
	OR(95% CI)	1.18(0.93–1.50)	1.29(0.85–1.95)	1.01(0.64–1.58)	0.98(0.77–1.24)	1.31(0.88–1.96)	1.10(0.96–1.26)
	P value	0.16	0.23	0.98	0.86	0.19	0.17
Q3	n (cases/ctrls)	222/305	69/100	70/80	195/288	69/93	625/866
	OR(95% CI)	1.15(0.91–1.45)	1.19(0.79–1.82)	1.76(1.12–2.75)	0.94(0.74–1.20)	1.26(0.84–1.91)	1.12(0.98–1.28)
	P value	0.24	0.41	0.014	0.60	0.27	0.11
Q4	n (cases/ctrls)	202/265	70/91	58/71	178/233	66/77	574/737
	OR(95% CI)	1.21(0.95–1.53)	1.32(0.86–2.01)	1.56(0.98–2.49)	1.06(0.82–1.37)	1.50(0.98–2.29)	1.22(1.06–1.41)
	P value	0.12	0.20	0.059	0.65	0.061	0.0059
Q5	n (cases/ctrls)	271/269	77/82	77/94	200/249	88/97	713/791
	OR(95% CI)	1.58(1.25–1.98)	1.53(1.00–2.32)	1.57(1.02–2.42)	1.10(0.86–1.41)	1.52(1.03–2.26)	1.39(1.22–1.60)
	P value	9.38×10^{-5}	0.048	0.042	0.43	0.036	1.83×10^{-6}
Per allele	n (cases/ctrls)	1171/1534	382/510	323/472	1042/1426	393/524	3311/4466
	OR(95% CI)	1.08(1.04–1.12)	1.07(1.00–1.15)	1.09(1.01–1.18)	1.01(0.97–1.06)	1.07(1.01–1.15)	1.06(1.03–1.08)
	P value	7.60×10^{-5}	0.039	0.023	0.52	0.034	7.81×10^{-7}
	$P_{hetvs.EA}$[§]	Ref.	0.83	0.72	0.036	0.89	$0.20^{\|}$

*
risk score is the total number of risk alleles

[†]Cut off values for risk score quintiles were based on the distribution of risk score in controls in each ethnic group, which are 9, 10, 11, 12 in EA, JA and LA; 10, 11, 12 and 13 in AA; and 9, 10, 11, 12.16 in NH.

[‡]Odds ratios (and 95% confidence intervals) adjusted for age (quartiles), ethnicity (pooled) and study. EA: European Americans; AA: African Americans; NH: Native Hawaiians; JA: Japanese Americans; LA: Latinos.

[§]P value for heterogeneity of risk score effects between European Americans and each of the other racial/ethnic groups (1-df test).

[‖]P value for heterogeneity of risk score effects across ethnic groups (4-df test).