Novel single nucleotide polymorphism associations with colorectal cancer on chromosome 8q24 in African and European Americans

Sonia K.Kupfer, Jada Benn Torres, Stanley Hooker, Jeffrey R.Anderson, Andrew D.Skol, Nathan A.Ellis and Rick A.Kittles*†

Sections of Gastroenterology and Genetic Medicine, Department of Medicine, University of Chicago Medical Center, Chicago, IL 60637, USA

*To whom correspondence should be addressed. Section of Genetic Medicine, Department of Medicine, University of Chicago Medical Center, MC 6091, Chicago, IL 60637, USA. Tel: +1 773 834 2271; Fax: +1 773 702 2567. Email: rkittles@medicine.bsd.uchicago.edu

Regions on chromosome 8q24 harbor susceptibility alleles for multiple cancers including colorectal (region 3) and prostate cancer (regions 1–4). The objectives of the present study were (i) to test whether single-nucleotide polymorphisms (SNPs) in region 4 are associated with colorectal cancer (CRC) in European or African Americans; (ii) to test whether 8q24 SNPs previously shown to be associated with colorectal and prostate cancer also show association in our multiethnic series and (iii) to test for association between 100 ancestry informative markers (AIMs) and CRC in both the African American and European American cohorts. In total, we genotyped nine markers on 8q24 and 100 linked AIMs in 569 CRC cases and 439 controls (490 European Americans and 518 African Americans) obtained retrospectively from a hospital-based sample. We found rs7008482 in 8q24 region 4 to be significantly associated with CRC in European Americans (𝑃=0.03). Also in region 4, we found that a second SNP, rs16903050, trended toward association with CRC in African Americans. The rs983267 in region 3, previously implicated in CRC risk, trended toward association with disease in European Americans but not in African Americans. Finally, none of the 100 AIMs tested for association reached statistical significance after correction for multiple hypothesis testing. In summary, these results are evidence that 8q24 region 4 contains novel CRC-associated alleles in European and African Americans.

Introduction

Recent genome-wide association studies in both colorectal cancer (CRC) and prostate cancer have shown strong evidence for association at chromosome 8q24 (1–11). Risk loci for these cancers as well as breast and ovarian cancer (12) have been identified within four distinct regions of 8q24 (Figure 1). In CRC, studies have reported and replicated single nucleotide polymorphism (SNP) associations in region 3 (1,2,8–10), whereas in prostate cancer, associations have been reported in region 3 as well as regions 1 and 2 (11). Our group recently reported a significant prostate cancer association in African Americans in region 4 of 8q24 (13). While multiple studies in prostate cancer including our own have focused on African Americans (7), only Haiman et al. (6) has studied 8q24 associations in African American CRC cases. This study identified rs6983267, located in region 3 of 8q24, as a common risk variant for CRC and prostate cancer in both European and African Americans after controlling for population stratification (6).

Determining genetic risk factors for CRC in African Americans is especially important because this population has the highest CRC incidence and mortality rates of any USA population (14). Estimates suggest that higher mortality rates from CRC account for 25% of the disparity in overall cancer mortality rates between African and European American women and 11% of this disparity in men (15). Even when controlling for tumor stage, socioeconomic status and comorbidities, disparities persist between African and European Americans (16) suggesting that biological factors, including genetics and environmental exposures, play a role. However, genetic studies in CRC including sufficient numbers of African Americans are limited.

Given that 8q24 appears to harbor common risk loci for multiple cancers, we tested whether novel and previously validated 8q24 SNPs are associated with CRC in our multiethnic series while controlling for population admixture using ancestry informative markers (AIMs). Moreover, we tested whether any of these AIMs are associated with CRC in African American and European Americans in order to discover potential population-specific risk SNPs.

Materials and methods

Cases and controls

Individuals with CRC (𝑛=569) who underwent surgical resection at the University of Chicago Medical Center between 1994 and 2005 were retrospectively ascertained from the Cancer Center and Pathology Department databases. Individuals known to have hereditary syndromes (familial adenomatous polyposis and Lynch syndrome) or inflammatory bowel disease were excluded. Available baseline characteristics including age, gender, race, colorectal tumor location, histological grade, depth of invasion, nodal involvement and metastases were recorded.

Hospital-based control samples (𝑛=439) were ascertained through our Pathology Department database and included cancer-free individuals who had thyroidectomies and amputations performed at our institution. Controls were matched to cases by age at diagnosis, 10 year birth cohort, gender and race as recorded in the database. While we matched closely on clinical characteristics a priori, we could not match on ancestry estimates as these were obtained through genotyping. After determination of West African ancestry estimates through genotyping, we found no significant ancestry differences between cases and controls (see below).

DNA extraction

Germline DNA for cases and controls was prepared from archived formalin-fixed surgical specimens from the paraffin block tissue repository. For each case, a block of normal colorectal, thyroid or soft tissue from the surgical margins was pulled. Five sections, 10 μm, thick were cut from each block with a fresh blade and the sections were transferred to a microfuge tube. The paraffin was removed by extraction with xylene followed by extraction with octane to remove residual xylene. The DNA was extracted using the PureGene kit according to the manufacturer’s instructions.

Genotyping

A total of 109 SNPs (9 on 8q24 and 100 AIMs) were genotyped using the Sequenom MassARRAY platform. The nine 8q24 SNPs localized to four regions were as follows: (Figure 1) rs10090154 and rs1447295 in region 1; rs6983561 and rs16901979 in region 2; rs10505477, rs6983267 and rs7000448 in region 3 and rs16900305 and rs7008482 in region 4. SNPs in all four regions have been associated with prostate cancer, whereas only SNPs in region 3 have been previously associated with CRC. The AIMs used have been published previously (13).

iPLEX assays were designed using the Sequenom Assay Design software. Primer sequences for polymerase chain reaction (PCR) and single-base extension are available on request. Multiplex PCR was performed to amplify 5–10 ng of genomic DNA extracted from the surgical specimens. PCR reactions were treated with shrimp alkaline phosphatase to neutralize unincorporated deoxynucleotide triphosphates. A post-PCR single-base extension reaction was performed for each multiplex reaction using concentrations of 0.625 μM for low-mass primers and 1.25 μM for high-mass primers. Reactions were diluted with 16 μl of H2O and fragments were purified with resin, spotted onto Sequenom SpectroCHIP microarrays and scanned by MALDI-TOF mass spectrometry. Individual SNP genotype calls were generated using...
Sequenom Typer 4.0 software, which automatically calls allele-specific peaks according to their expected masses.

**Quality control**

We tested for departures from Hardy–Weinberg equilibrium in cases and controls independently. All nine SNPs had Hardy–Weinberg equilibrium \( P \text{-values} < 0.01 \); thus, none were excluded from the analysis. We excluded SNPs with minor allele frequencies \(<0.005\) (rs7000448 in both populations and rs16900305 in European Americans only). Hence, we included eight SNPs in African Americans and seven SNPs in European Americans in our final association analysis. Thirty-five African Americans (7%) and 29 European Americans (6%) were excluded from the 8q24 analysis for low genotyping call rates (\( >10\% \) missing). Genotyping quality control for all SNPs was accessed using blinded duplicate DNA for 24 samples. A genotype concordance rate of 100% was observed for all markers.

**Statistical analysis**

**Genetic ancestry estimation** Global individual ancestry was determined for each individual using AIMs for European and West African ancestry. A total of 100 highly informative genome-wide markers were genotyped by iPLEX assay. Individual ancestry estimates were obtained from the genotype results using the Bayesian Markov Chain Monte Carlo method implemented in the program STRUCTURE 2.1 (17). STRUCTURE 2.1 assumes an admixture model using prior population information and independent allele frequencies. The Markov Chain Monte Carlo model was run using \( K = 2 \) populations (60 Europeans and 131 West Africans) and a burn-in length of 30 000 iterations followed by 70 000 replications.

**Association analysis** We tested chromosome 8q24 SNPs for association with CRC for European and African Americans separately. We also calculated odds ratios (ORs) and 95% confidence intervals (CIs) using logistic regression assuming an additive effect (on the log scale) of allele dosage. We controlled for individual admixture by including West African ancestry estimates in the logistic regression model. We determined the significance of each SNP-disease association empirically using max(T) permutation procedure [number of permutations \((n) = 10\ 000\)] , which corrects for the number of SNPs tested while accounting for the linkage disequilibrium between them. These tests were done using the program PLINK (18).

We also tested for association between each of the 100 AIMs and disease status in both populations. AIMs have large allele frequency differences between the populations they are meant to discriminate. This is also a characteristic of a locus under selective pressure, which is viewed as a good a priori characteristic for loci with important functional consequences. We continued to include individual ancestry estimates in the logistic regression models; thus, any evidence of association with an AIM is unlikely to be attributed to percent ancestry. Empirical \( P \text{-values} \) were determined using max(T) permutation \((n = 10\ 000\) permutations).

Clinical characteristics were compared between cases and controls as well as within cases and controls as a function of ancestry. Two-sided \( t \)-tests were used to compare continuous variables including age and ancestry estimates. Pearson chi-square tests of independence were used to compare categorical variables including gender and tumor site.

**Results**

The clinical characteristics of our CRC cases and controls are shown in Table I. Cases and controls were closely matched for age at diagnosis, gender and race as assigned in the Cancer Center database. African American cases and controls did not show significantly different West African ancestry estimates and had mean estimates similar to those previously estimated among individuals residing in Chicago, IL (R. Kittles, unpublished data).

Considering only CRC cases, we found differences between African and European Americans in several demographic and clinical characteristics. African American CRC cases were older, more likely to be female and had more proximal cancers compared with
European Americans. Regarding our hospital-based control subjects, since we included normal DNA from individuals undergoing two different surgical procedures, we tested whether there were differences between thyroidectomy and amputation patients. In both European and African Americans, we found no significant differences in minor allele frequencies of 8q24 SNPs between thyroid and amputation controls (P-values > 0.05 for all 8q24 SNPs; data not shown).

In reviewing our West African ancestry estimates using the 100 AIMS, we noted that 89 individuals had discrepancies between race assigned in the Cancer Center database and ancestry estimates. By reviewing electronic medical records, we were able to revise the Cancer Center database in accord with individual ancestry estimates in 30 subjects (22 cases; 8 controls). These were included in our association analyses. For the remaining 59 subjects, 28 subjects (22 cases; 6 controls) contained no clinical documentation of race, 16 subjects (seven cases; nine controls) had clinical documentation that contradicted the ancestry estimates and 15 European American individuals (eight cases; seven controls) were found to have >15% West African ancestry. Due to these discrepancies in race designation and ancestry estimates, all 59 individuals were excluded from our association analyses.

Allele frequencies for each 8q24 SNP are shown in Table II for both European and African Americans. After correcting for ancestry, we observed a statistically significant association between rs7008482 in region 4 and the risk of CRC in European Americans with an OR of 1.51 (95% CI 1.14–1.99). This association remained significant after correcting for multiple testing (P_{emp} = 0.03). This SNP was not significantly associated with CRC in African Americans. In addition to the allele frequencies differing between African and European Americans at this locus, the risk allele in the two populations differed (risk allele is T and G in African and European Americans, respectively). In African Americans, rs16900305, also located in region 4, was nominally associated with CRC with an OR of 1.60 (95% CI 1.05–2.45); however, the SNP did not remain significant after accounting for multiple testing (P_{emp} = 0.17).

In African Americans, we observed a trend toward an association for rs6983267 in region 3. This SNP previously has been associated with CRC (Figure 1). In contrast, in African Americans, we found no significant associations for region 3 SNPs. It is important to note that the presumptive risk-associated alleles of rs10505477 and rs6983267 have frequencies of 0.83 (A allele) and 0.90 (G allele), respectively, in African Americans. These high frequencies potentially limit our

---

**Table I. Clinical characteristics of European and African American CRC cases and controls**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>European American</th>
<th></th>
<th>African American</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cases</td>
<td>Controls</td>
<td>Cases</td>
<td>Controls</td>
</tr>
<tr>
<td>Number of participants</td>
<td>288</td>
<td>202</td>
<td>281</td>
<td>237</td>
</tr>
<tr>
<td>Mean age in years (SD)</td>
<td>65.4 (12.2)b</td>
<td>64.5 (11.9)</td>
<td>0.39</td>
<td>67.8 (12.1)b</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female (%)</td>
<td>128 (44.4)c</td>
<td>97 (48.0)</td>
<td>0.61</td>
<td>163 (58.0)c</td>
</tr>
<tr>
<td>Male (%)</td>
<td>160 (55.6)c</td>
<td>105 (52.0)</td>
<td></td>
<td>118 (42.0)c</td>
</tr>
<tr>
<td>Tumor site⁵</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Right (%)</td>
<td>88 (31.0)c</td>
<td>137 (49.8)c</td>
<td>0.15</td>
<td>0.84 (0.14)</td>
</tr>
<tr>
<td>Left (%)</td>
<td>196 (69.0)c</td>
<td>138 (50.2)c</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>P-value for heterogeneity.  <br><sup>b</sup>By two-sided t-test, age was statistically different in European versus African American cases (P-value = 0.02) but not in controls (P-value = 0.21).  <br><sup>c</sup>By Pearson chi-square test, gender was statistically different in European versus African American cases (P-value = 0.001) but not in controls (P-value = 0.11).  <br><sup>d</sup>Right, splenic flexure to cecum; left, rectum to splenic flexure.  <br><sup>e</sup>By Pearson chi-square test, tumor site was statistically different in European versus African American (P-value < 0.001).

**Table II. SNPs, position, risk allele frequencies and P-values for association with CRC on chromosome 8q24 in European and African Americans**

<table>
<thead>
<tr>
<th>SNP</th>
<th>Position</th>
<th>Risk allele</th>
<th>Cases</th>
<th>Controls</th>
<th>OR (95% CI)</th>
<th>P*</th>
<th>P_{emp}</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Americans</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs16900305</td>
<td>126142853</td>
<td>A</td>
<td>0.003</td>
<td>0.000</td>
<td>NT</td>
<td>NT</td>
<td>NT</td>
</tr>
<tr>
<td>rs7008482</td>
<td>126336812</td>
<td>G</td>
<td>0.36</td>
<td>0.28</td>
<td>1.51 (1.14–1.99)</td>
<td>0.007</td>
<td>0.03</td>
</tr>
<tr>
<td>rs6983561</td>
<td>128176062</td>
<td>C</td>
<td>0.03</td>
<td>0.04</td>
<td>0.79 (0.40–1.57)</td>
<td>0.53</td>
<td>0.97</td>
</tr>
<tr>
<td>rs16901979</td>
<td>128194098</td>
<td>A</td>
<td>0.03</td>
<td>0.04</td>
<td>0.79 (0.40–1.57)</td>
<td>0.53</td>
<td>0.97</td>
</tr>
<tr>
<td>rs10505477</td>
<td>128476625</td>
<td>A</td>
<td>0.50</td>
<td>0.45</td>
<td>1.26 (0.97–1.63)</td>
<td>0.10</td>
<td>0.38</td>
</tr>
<tr>
<td>rs6983267</td>
<td>128482487</td>
<td>G</td>
<td>0.53</td>
<td>0.46</td>
<td>1.31 (1.01–1.69)</td>
<td>0.06</td>
<td>0.23</td>
</tr>
<tr>
<td>rs1447295</td>
<td>128554220</td>
<td>A</td>
<td>0.11</td>
<td>0.09</td>
<td>1.30 (0.84–1.99)</td>
<td>0.23</td>
<td>0.67</td>
</tr>
<tr>
<td>rs10090154</td>
<td>128601319</td>
<td>T</td>
<td>0.11</td>
<td>0.09</td>
<td>1.31 (0.84–2.03)</td>
<td>0.23</td>
<td>0.68</td>
</tr>
</tbody>
</table>

| African Americans |       |             |       |          |             |      |        |
| rs16900305 | 126142853 | A           | 0.13  | 0.09     | 1.60 (1.05–2.45) | 0.03  | 0.17   |
| rs7008482  | 126336812 | G           | 0.82  | 0.87     | 0.72 (0.50–1.03) | 0.15  | 0.68   |
| rs6983561  | 128176062 | C           | 0.43  | 0.46     | 0.88 (0.68–1.14) | 0.44  | 0.98   |
| rs16901979 | 128194098 | A           | 0.42  | 0.45     | 0.86 (0.67–1.21) | 0.36  | 0.95   |
| rs10505477 | 128476625 | A           | 0.83  | 0.82     | 1.09 (0.78–1.54) | 0.49  | 0.99   |
| rs6983267  | 128482487 | G           | 0.90  | 0.90     | 1.04 (0.68–1.59) | 0.60  | 0.99   |
| rs1447295  | 128554220 | A           | 0.29  | 0.33     | 0.83 (0.63–1.10) | 0.27  | 0.89   |
| rs10090154 | 128601319 | T           | 0.17  | 0.20     | 0.84 (0.60–1.28) | 0.37  | 0.96   |

Nucleotides shown are based on HapMap reference sequence. P_{emp}, empirical P-value based on 10 000 max(T) permutations of trait values in the sample controlling for West African ancestry. NT, not tested.

<sup>1</sup>P-value corrected for West African ancestry.
power to detect significant SNP associations in the African American population. In summary, we found evidence of association with region 4 8q24 SNPs and CRC in European Americans and nominal significance with region 3 and 4 SNPs in both European and African Americans.

We also performed association tests using 100 genome-spanning AIMs, comparing CRC cases and controls in both the African American and European American cohorts. Table III lists the association results where ancestry-adjusted $P$-values were noted to be <0.05. None of these associations remained significant after adjusting for multiple hypothesis testing. In African Americans, we observed an association signal for rs10059859 on chromosome 5 with an adjusted nominal $P$-value of 0.002 ($P = 0.14$ after adjusting for multiple testing). In both European Americans, we found three SNP associations with adjusted nominal $P$-values <0.05. These SNPs, rs6601288, rs1540979 and rs2021782, are located on chromosomes 8, 13 and 7, respectively.

**Discussion**

Several distinct regions across a gene desert on chromosome 8q24 have been associated with a number of common malignancies including colorectal, prostate, ovarian and breast cancer. Previously, SNPs associated with CRC have been localized within region 3, with one of these SNPs (rs6983267) also implicated in prostate cancer risk (6). Our group previously identified rs7008482 in region 4 as a prostate cancer risk factor in African Americans (13). This SNP is in region 4 of 8q24 located ~2.2 Mb proximal to previously identified association signals on 8q24. The rs7008482 is located in an intronic region of the NSMCE2 (also called Mms21) gene that has been implicated in DNA replication, recombination and repair (19). In region 4, we have also identified a second SNP rs16900305 as associated with prostate cancer (R. Kittles, unpublished data). This SNP is located in an intronic region of hypothetical gene KIAA0196 and 30 kb upstream of NSMCE2.

In our present study, we found that rs7008482 was significantly associated with CRC in European Americans with an OR of 1.51 for the G allele. This result is similar to that found in prostate cancer in African Americans (13). We did not find a significant association with rs7008482 in our African American CRC cohort. In fact, we noted that the T-allele frequency was slightly higher in African American cases than controls, though this was not significant in the association analysis. We may not have observed a similar association in the present study due to our smaller African American sample size. Genotyping rs7008482 in additional African American CRC cases and controls is currently underway. In the African American cohort, we obtained evidence for an association with rs16900305 also located in region 4 on 8q24. Considering the rs7008482 and rs16900305 data together, these results suggest that region 4 is a novel CRC-associated region on 8q24 in both African and European Americans.

Furthermore, we showed a trend toward association of rs6983267 in region 3 with CRC in our European American cohort, a finding that replicates previous studies in CRC (1,2,8–10). However, we did not find significant associations with this SNP in our African American cohort as had been reported by Haiman et al. (6) using 217 subjects from the multiethnic cohort. This study found a trend toward an association with rs6983267 in African American CRC cases and 1049 controls (OR 1.37 (95% CI 0.98–1.91)). It is possible that we were not adequately powered to detect a significant difference given that the risk allele frequencies for these two SNPs in African Americans are both >80%. Taken together, these results suggest that the previously tested region 3 SNPs may not be strongly associated with CRC in African Americans. Other variants in the 8q24 region may show stronger associations with CRC in African Americans.

Finally, we report CRC associations by testing AIMs for associations in African and European Americans. In African Americans, the C allele of rs10059859 on chromosome 5q was associated with disease. In European Americans, rs10059859 was not associated with disease, but we did find associations with rs6601288, rs1540979 and rs2021782 on chromosomes 8, 13 and 7, respectively. Additional genotyping in a larger sample size is required to replicate these preliminary results. There may be population-specific SNP associations in CRC and future admixture mapping may be warranted.

In summary, regions on chromosome 8q24 harbor susceptibility alleles for a number of common cancers including colorectal and prostate cancer. In this study and our previous work, we have broadened the region of interest to include SNPs in region 4. This novel CRC-associated region includes associated SNPs in both European and African Americans. A DNA repair gene NSMCE2 is an excellent candidate gene in region 4. Additional work is required to determine functional variants in this area and to elucidate their role in carcinogenesis.

**Funding**

The Cancer Research Foundation, Department of Medicine at the University of Chicago, Digestive Disease Research Core Center (P30 DK42086); National Institutes of Health, National Service Research Award (1F32CA132493 to S.S.K.); Department of Defense (DAMD W81XWH-07-1-0203 to R.A.K.); The University of Chicago Cancer Center.

**Acknowledgements**

We thank Nancy Cox for helpful discussions and The University of Chicago Genotyping Core for assistance with genotyping. We thank Heather Gold, Polly Cline and Meghan Daly for help collecting surgical blocks and DNA preparation.

**Conflict of Interest Statement:** None declared.
References


Received December 10, 2008; revised May 12, 2009; accepted May 14, 2009