Original Contribution

Comprehensive Evaluation of the Impact of 14 Genetic Variants on Colorectal Cancer Phenotype and Risk

Steven J. Lubbe, Maria Chiara Di Bernardo, Peter Broderick, Ian Chandler, and Richard S. Houlston*

* Correspondence to Dr. Richard S. Houlston, Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey SM2 5NG, United Kingdom (e-mail: richard.houlston@icr.ac.uk).

Initially submitted April 1, 2011; accepted for publication July 21, 2011.

To comprehensively evaluate the impact of recently identified colorectal cancer (CRC) variants at 1q41, 3q26.2, 8q23.3, 8q24.21, 10p14, 11q23.13, 14q22.2, 15q13.3, 16q22.1, 18q21.1, 19q13.11, 20p12.3, and 20q13.3 on risk and CRC phenotype, the authors analyzed 8,878 cases and 6,051 controls from the United Kingdom ascertained in 1999–2007. The impact of variants on the familial CRC risk was enumerated from age-, sex-, and calendar-specific CRC rates in the 50,924 first-degree relatives of cases. Each of the 14 susceptibility loci independently influences CRC with the risk increasing with increasing number of risk alleles carried (per allele odds ratio $= 1.13; P = 2.99 \times 10^{-8}$) and, for those within the upper quintile, there is a 2.3-fold increased risk. In first-degree relatives of cases with $\leq 17$, 18–21, and $\geq 22$ risk alleles, standardized incidence ratios were 1.76, 2.08, and 2.25, respectively. Although the discriminatory attributes of the 14 CRC susceptibility loci for individual risk prediction are poor (area under the curve $= 0.58$), they may allow subgroups of the population at different CRC risks to be distinguished.

colorectal neoplasms; genetic loci; genotype; polymorphism, single nucleotide; risk

Abbreviations: CI, confidence interval; CRC, colorectal cancer; NSCCG, National Study of Colorectal Cancer Genetics; SNP, single nucleotide polymorphism; TGF-β, transforming growth factor-β.

Many colorectal cancers (CRCs) develop in genetically susceptible individuals, most of whom are not carriers of germline mismatch repair or adenomatous polyposis coli gene (APC) mutations, and much of the heritable risk of CRC is thought to be the consequence of the coinheritance of multiple low-risk variants (1–3). Genome-wide association studies have identified single nucleotide polymorphisms (SNPs) localizing to 14 chromosome regions—1q41 (4), 3q26.2 (4), 8q23.3 (5), 8q24.21 (6, 7), 10p14 (5), 11q23.1 (8), 12q13.13 (4), 14q22.2 (9), 15q13.3 (10), 16q22.1 (9), 18q21.1 (11), 19q13.11 (9), 20p12.3 (9), and 20q13.3 (4)—that influence CRC risk.

Although the risk of CRC associated with variation at each locus is modest (4–12), risk genotypes are common in the population and, hence, their contribution to the overall incidence of CRC is significant.

An important long-term outcome of genome-wide association studies of CRC is that the knowledge gained about the underlying molecular basis of CRC may lead to the development of innovative therapeutic and preventative measures. In this respect, the 14 risk loci identified so far have already provided important novel insights into CRC biology, with several suggesting the involvement of components of the transforming growth factor-β (TGF-β) superfamily-signaling pathway as an etiologic basis for CRC development (13).

A more immediate value of the associations identified from genome-wide association studies may be in assigning CRC risk. Individual SNPs at each of the 14 chromosomal regions confer only small CRC risks; hence, individually, each has limited use in the assessment of disease risk in an individual patient. However, by acting in concert, variants have the potential to confer a more profound impact on CRC risk. Characterization of putative genotype-phenotype correlations also has possible benefit in improving the efficacy of CRC management and prevention strategies for CRC.
To date, studies of the 14 CRC risk variants have been conducted largely through the analysis of small studies (14–18). To robustly evaluate the relation between variation at these 14 genetic loci and CRC, we have conducted a large study within the United Kingdom of 14,929 subjects, examining for cumulative effects and genotype-phenotype relations.

**MATERIALS AND METHODS**

**Study subjects**

Cases, aged <80 years at diagnosis, were ascertained through the National Study of Colorectal Cancer Genetics (NSCCG), an ongoing population-based study in the United Kingdom (19) of histologically proven CRC patients within 5 years of diagnosis with no exclusion for known genetic susceptibility. Family history of CRC in relatives is collected by using a previously validated questionnaire, but no confirmation of diagnoses is undertaken. The current study is based on 8,878 cases ascertained between March 2003 and April 2007. Controls were the spouses of cancer cases (n = 6,051) and were ascertained between June 1999 and April 2007 through the NSCCG (n = 1,933), the Genetic Lung Cancer Predisposition Study (n = 1,628) (20), the Colorectal Adenoma Gene-Environment Interactions Study (n = 707) (21), and the Royal Marsden Hospital National Health Services Trust family history DNA database (n = 1,783). None of the controls had a personal history of malignancy at ascertainment. Demographic data were collected through a self-administered questionnaire that was based on the NSCCG format but did not ask for information on a family history of cancer. All cases and controls were British residents with self-reported European ethnicity. There were no obvious differences in the geographic region of residency between cases and controls. A subset of the cases (n = 2,802) and controls (n = 2,708) had previously been used in the replication of genome-wide association signals for CRC risk (6). Informed, written consent was obtained from all subjects, and the study was carried out with ethical approval (MultiResearch Ethics Committee (MREC) reference numbers 02/0/097, 98/2/67, and 98/4/055; Royal Marsden Hospital National Health Services Trust (RMHNHST)-Centre for Clinical Research reference number 1552).

**Selection of SNPs for genotyping**

DNA was extracted from ethylenediaminetetraacetic acid (EDTA) venous blood samples by conventional methodologies and PicoGreen quantified (Invitrogen Corporation, Carlsbad, California; now Life Technologies). We selected 14 SNPs that have been reported to be associated with CRC from 14 chromosomal regions—rs6691170 (1q41) (4), rs10936599 (3q26.2) (4), rs16892766 (8q23.3) (5), rs6983267 (8q24.21) (6, 7), rs10795668 (10p14) (5), rs3802842 (11q23.1) (8), rs11169552 (12q13.13) (4), rs4444235 (14q22.2) (9), rs4779584 (15q13.3) (10), rs9929218 (16q21.1) (9), rs4939827 (18q21.1) (11), rs10411210 (19q13.11) (9), rs961253 (20p12.3) (9), and rs4925386 (20q13.33) (4). SNP genotyping was performed by allele-specific polymerase chain reaction (Kbioscience, Hertfordshire, United Kingdom (http://www.kbioscience.co.uk; now LGC Limited), with primer sequences and conditions available on request. To monitor quality control, we included a set of 136 duplicate samples in assays; genotype concordance was >99.9%. To confirm genotypes, we sequenced 192 samples chosen randomly from cases and controls; concordance between genotypes was 100%. Microsatellite instability in CRCs from 2,813 of the cases was determined, as previously described by using BAT25 and BAT26 (19), which are highly sensitive microsatellite instability markers (22). Samples that showed novel alleles at either or both markers were assigned microsatellite instability status (which corresponded to high microsatellite instability).

**Statistical analysis**

\[ P \leq 0.05 \] (2 sided) was considered statistically significant. A Bonferroni correction was imposed to adjust for multiple testing in associations between SNP and CRC risk and between SNP genotype and CRC phenotype. Statistical analyses were principally performed by use of STATA, version 10.0, software (StataCorp LP, College Station, Texas). To test for population stratification, for each variant, we used the chi-square test to assess the Hardy-Weinberg equilibrium among control subjects. The CRC risk associated with SNPs was calculated by odds ratios using unconditional logistic regression. Recessive, dominant, or multiplicative (i.e., where the risk is the product of the risk associated with carrying 2 risk alleles) models were fitted on the log scale using logistic regression. The best fitting model of inheritance for each variant was individually determined by comparing the −2 times the log-likelihood of each model with a base model via a chi-square test with 1 df. Interaction between SNP genotypes was evaluated by likelihood ratio tests. We used logistic regression on the general model (i.e., individual SNP genotypes as indicator variables) to construct receiver-operating characteristic curves.

The population attributable fraction ascribable to SNP genotypes was estimated by the formula: \[ \frac{\sum_{i=1}^{n} \frac{\hat{p}_i \times (\hat{q}_i - \hat{p}_i)}{\hat{q}_i}}{\sum_{i=1}^{n} \frac{\hat{p}_i \times (\hat{q}_i - \hat{p}_i)}{\hat{q}_i}} \times 100\% \] where \( x_i \) = \( (1 - p) + 2p(1 - p)\hat{p}_i + p^2\hat{q}_i \), p is the population allele frequency, and \( \hat{p}_i \) and \( \hat{q}_i \) are the odds ratios associated with hetero- and homozygosity, respectively (23).

Associations between SNPs and tumor site (colon and rectal cancer), International Classification of Diseases, Ninth Revision (ICD-9), codes 153 and 154, respectively), and family history of CRC in first-degree relatives, and microsatellite instability status were tested among case subjects by using logistic regression only, with and without correcting for known confounders, such as age and sex.

To enumerate the collective impact of SNPs on the familial CRC risk, we first estimated the cumulative probability of CRC in first-degree relatives of cases by the Kaplan-Meier product-limit method (24). The variance of the product-limit estimator was estimated according to the formulae of Greenwood (25) and was used to calculate 95% confidence intervals. First-degree relatives were censored at the date of pedigreed ascertainment, emigration, or last contact with the case. CRC incidence in first-degree relatives was truncated at age 80 years to avoid miscertification of registered cancer diagnosis and...
death. Comparison of CRC risk in first-degree relatives with the risk in the general population for England and Wales was based on 1998 rates (Office for National Statistics; http://www.statistics.gov.uk) using the formula \( 1 - e^{-\text{cumulative incidence}} \).

Expected numbers of CRCs in first-degree relatives were calculated by adjustment for age, sex, and calendar-period incidence rates for England and Wales (1971–2004) using the Person-Years program (26). Observed numbers were compared with expected numbers by means of the standardized incidence ratio assuming a Poisson distribution. Poisson regression was used to calculate the proportion of the phenotypic variance in familial risk associated with SNPs using age-, sex-, and calendar-adjusted expected numbers rates and observed numbers for families.

**RESULTS**

Complete clinical and demographic characteristics of the subjects studied are provided in Table 1. The median age at diagnosis of CRC in the 8,878 cases was 60.0 years (mean, 58.9 years; standard deviation, 8.2). The median age of controls was 60.0 years (mean, 59.4 years; standard deviation, 10.5). A total of 5,343 cases (60%) had colonic and 3,535 (40%) had rectal tumors; 330 of the 2,813 cases (12%) evaluated for microsatellite instability status were microsatellite instability positive. Genotypes were generated for >95% of cases and controls for each of the 14 SNPs, with no differential genotype calling between cases and controls. Genotype frequencies of all 14 SNPs in controls were in accordance with Hardy-Weinberg equilibrium after adjustment for multiple testing.

Table 2 lists the allele frequencies of the 14 SNPs among cases and controls and shows per allele odds ratios for CRC. Significantly different frequencies (\( P < 0.05 \)) between cases and controls were observed for all 14 SNPs. Although a third of the total series had previously been used as replication for genome-wide association signals, there was no statistical evidence for the “winner’s curse” (i.e., upward bias of the effect

<table>
<thead>
<tr>
<th>SNP</th>
<th>Chromosome Position</th>
<th>Gene*</th>
<th>Major Allele</th>
<th>Minor Allele</th>
<th>Case MAF</th>
<th>Control MAF</th>
<th>OR</th>
<th>95% CI</th>
<th>P Value</th>
<th>Sample Size, no.</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6691170</td>
<td>1q41</td>
<td>G</td>
<td>T*</td>
<td>0.38</td>
<td>0.35</td>
<td>1.13</td>
<td>1.08, 1.19</td>
<td>5.16 × 10⁻⁷</td>
<td>14,309</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs10936599</td>
<td>3q26.2</td>
<td>MYNN</td>
<td>C*</td>
<td>0.23</td>
<td>0.24</td>
<td>1.07</td>
<td>1.02, 1.14</td>
<td>0.011</td>
<td>14,386</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs16892766</td>
<td>8q23.3</td>
<td>EIF3H</td>
<td>A</td>
<td>0.10</td>
<td>0.08</td>
<td>1.28</td>
<td>1.18, 1.39</td>
<td>8.92 × 10⁻⁹</td>
<td>14,764</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs6983267</td>
<td>8q24.21</td>
<td>c-MYC</td>
<td>G*</td>
<td>0.44</td>
<td>0.48</td>
<td>1.16</td>
<td>1.11, 1.22</td>
<td>2.62 × 10⁻¹⁰</td>
<td>14,671</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>rs10795668</td>
<td>10p14</td>
<td>Gβ1</td>
<td>A</td>
<td>0.30</td>
<td>0.32</td>
<td>1.14</td>
<td>1.08, 1.20</td>
<td>3.77 × 10⁻⁷</td>
<td>14,375</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs3802842</td>
<td>11q23.1</td>
<td>C11orf93</td>
<td>A</td>
<td>0.32</td>
<td>0.29</td>
<td>1.17</td>
<td>1.11, 1.23</td>
<td>3.86 × 10⁻⁹</td>
<td>14,657</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs11169552</td>
<td>12q13.13</td>
<td>Cβ</td>
<td>T</td>
<td>0.25</td>
<td>0.28</td>
<td>1.12</td>
<td>1.06, 1.18</td>
<td>5.82 × 10⁻⁵</td>
<td>14,327</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4444235</td>
<td>14q22.2</td>
<td>BMP4</td>
<td>T</td>
<td>0.49</td>
<td>0.47</td>
<td>1.09</td>
<td>1.05, 1.14</td>
<td>1.84 × 10⁻⁴</td>
<td>14,789</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>rs4779588</td>
<td>15q13.33</td>
<td>GREMI1/SCG5</td>
<td>C</td>
<td>0.22</td>
<td>0.19</td>
<td>1.20</td>
<td>1.14, 1.27</td>
<td>3.61 × 10⁻¹⁰</td>
<td>14,715</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>rs9929218</td>
<td>16q22.1</td>
<td>CDH1</td>
<td>A</td>
<td>0.28</td>
<td>0.30</td>
<td>1.10</td>
<td>1.04, 1.15</td>
<td>3.44 × 10⁻⁴</td>
<td>14,766</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>rs4939827</td>
<td>18q21.1</td>
<td>SMAD7</td>
<td>T*</td>
<td>0.44</td>
<td>0.48</td>
<td>1.16</td>
<td>1.10, 1.21</td>
<td>8.25 × 10⁻⁵</td>
<td>14,717</td>
<td>11, 28</td>
<td></td>
</tr>
<tr>
<td>rs10411210</td>
<td>19q13.11</td>
<td>RHPN2</td>
<td>Cβ</td>
<td>0.08</td>
<td>0.10</td>
<td>1.14</td>
<td>1.06, 1.24</td>
<td>1.08 × 10⁻³</td>
<td>14,722</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>rs961253</td>
<td>20p12.3</td>
<td>C</td>
<td>A*</td>
<td>0.38</td>
<td>0.35</td>
<td>1.14</td>
<td>1.08, 1.19</td>
<td>2.47 × 10⁻⁷</td>
<td>14,662</td>
<td></td>
<td></td>
</tr>
<tr>
<td>rs4925386</td>
<td>20q13.33</td>
<td>LAMA5</td>
<td>Cβ</td>
<td>0.29</td>
<td>0.32</td>
<td>1.11</td>
<td>1.06, 1.17</td>
<td>5.82 × 10⁻⁵</td>
<td>14,355</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; MAF, minor allele frequency; OR, odds ratio; SNP, single nucleotide polymorphism.

* Gene symbols: BMP4, bone morphogenetic protein 4 gene; C11orf93, chromosome 11 open reading frame 93 (hypothetical gene); CDH1, E-cadherin (epithelial) gene; c-MYC, v-myc avian myelocytomatosis viral oncogene homolog; EIF3H, eukaryotic translation initiation factor 3, subunit H, gene; GREM1, gremlin 1 gene; LAMA5, laminin, subunit alpha 5, gene; MYNN, myoneurin gene; RHPN2, Ras-homologous (Rho)-GTPase binding protein 2 gene; SCG5, secretogranin V gene; SMAD7, mothers against decapentaplegic homolog 7 gene.

* Risk allele for each SNP.

* Gene annotated for nonintragenic markers.

size in the discovery compared with subsequently genotyped samples) and, in view of this, analyses were conducted by using pooled data (Table 1, the first of 4 Web tables and 1 Web figure posted on the Journal’s Web site [http://aje.oupjournals.org/]). Each SNP showed effect sizes similar to those previously reported (Table 2) (5, 6, 9, 11). The strongest association was attained for SNPs mapping to 8q24.21, those previously reported (Table 2) (5, 6, 9, 11). The strongest association was attained for SNPs mapping to 8q24.21, those previously reported (Table 2) (5, 6, 9, 11). The strongest association was attained for SNPs mapping to 8q24.21, those previously reported (Table 2) (5, 6, 9, 11).
of risk alleles among the controls where the reference group consisted of the quintile containing the control median. As shown in Figure 2B, a more stable risk estimate that is more relevant to the distribution of risk alleles in the population was generated. Using cumulative SNP data permits the risk of CRC in the population to be partitioned; for the 3% of the

Table 3. Selected Genotype-Phenotype Correlations, London, United Kingdom, 1999–2007

<table>
<thead>
<tr>
<th>Gender</th>
<th>OR_{crude}</th>
<th>95% CI</th>
<th>P Value</th>
<th>OR_{adjusted}</th>
<th>95% CI</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs3802842</td>
<td>Male</td>
<td>1.22</td>
<td>1.13, 1.32</td>
<td>2.47 × 10^{-7}</td>
<td>1.22</td>
<td>1.13, 1.32</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1.10</td>
<td>1.02, 1.18</td>
<td>0.012</td>
<td>1.10</td>
<td>1.02, 1.18</td>
</tr>
<tr>
<td>rs961253</td>
<td>Male</td>
<td>1.17</td>
<td>1.09, 1.26</td>
<td>2.79 × 10^{-5}</td>
<td>1.17</td>
<td>1.09, 1.26</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1.08</td>
<td>1.01, 1.16</td>
<td>0.022</td>
<td>1.08</td>
<td>1.01, 1.16</td>
</tr>
<tr>
<td>rs9929218</td>
<td>Male</td>
<td>1.11</td>
<td>1.03, 1.19</td>
<td>0.008</td>
<td>1.11</td>
<td>1.02, 1.19</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1.13</td>
<td>1.05, 1.22</td>
<td>0.001</td>
<td>1.13</td>
<td>1.05, 1.22</td>
</tr>
<tr>
<td>rs4925386</td>
<td>Male</td>
<td>1.06</td>
<td>0.98, 1.15</td>
<td>0.122</td>
<td>1.06</td>
<td>0.98, 1.14</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>1.19</td>
<td>1.10, 1.28</td>
<td>3.32 × 10^{-6}</td>
<td>1.19</td>
<td>1.10, 1.28</td>
</tr>
<tr>
<td>Age of diagnosis^b</td>
<td>rs11169552</td>
<td>\leq 49 years</td>
<td>1.25</td>
<td>1.12, 1.40</td>
<td>5.35 × 10^{-5}</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>\geq 50 years</td>
<td>1.10</td>
<td>1.04, 1.16</td>
<td>0.001</td>
<td>1.12</td>
<td>1.06, 1.18</td>
</tr>
<tr>
<td>Family history of CRC^c</td>
<td>rs6983267</td>
<td>FH positive</td>
<td>1.26</td>
<td>1.15, 1.37</td>
<td>1.16 × 10^{-7}</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>FH negative</td>
<td>1.09</td>
<td>1.01, 1.19</td>
<td>0.034</td>
<td>1.09</td>
<td>1.01, 1.19</td>
</tr>
<tr>
<td>Location of tumor^c</td>
<td>rs3802842</td>
<td>Colon</td>
<td>1.12</td>
<td>0.06, 1.18</td>
<td>1.45 × 10^{-4}</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>Rectum</td>
<td>1.24</td>
<td>1.16, 1.32</td>
<td>5.38 × 10^{-11}</td>
<td>1.24</td>
<td>1.16, 1.32</td>
</tr>
<tr>
<td>MSI status^c</td>
<td>rs4444235</td>
<td>MSS</td>
<td>1.12</td>
<td>1.04, 1.21</td>
<td>0.001</td>
<td>1.11</td>
</tr>
<tr>
<td></td>
<td>MSI</td>
<td>0.96</td>
<td>0.82, 1.11</td>
<td>0.559</td>
<td>0.95</td>
<td>0.82, 1.11</td>
</tr>
<tr>
<td></td>
<td>MSI</td>
<td>1.14</td>
<td>1.07, 1.22</td>
<td>3.33 × 10^{-5}</td>
<td>1.15</td>
<td>1.08, 1.23</td>
</tr>
<tr>
<td></td>
<td>MSI</td>
<td>0.98</td>
<td>0.84, 1.16</td>
<td>0.753</td>
<td>0.98</td>
<td>0.84, 1.14</td>
</tr>
<tr>
<td>Dukes' stage^c</td>
<td>rs16892766</td>
<td>A + B</td>
<td>1.43</td>
<td>1.26, 1.62</td>
<td>4.50 × 10^{-8}</td>
<td>1.43</td>
</tr>
<tr>
<td></td>
<td>C + D</td>
<td>1.23</td>
<td>1.08, 1.40</td>
<td>0.002</td>
<td>1.21</td>
<td>1.06, 1.38</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; CRC, colorectal cancer; FH, family history; MSI, microsatellite instability; MSS, microsatellite stable; OR, odds ratio.

^a Adjusted for age.

^b Adjusted for gender.

^c Adjusted for age and gender.
population carrying ≥19 risk alleles, a 2.3-fold increased CRC risk was observed.

To estimate the ability of the cumulative effect of the 14 SNPs to distinguish cases from controls, we conducted specificity and sensitivity analyses, constructing receiver-operating characteristic curves and calculating statistics for the area under the curve. Incorporation of age and sex did not appreciably improve the discriminatory power of a genetic test based on the 14 SNPs with the area under the curve = 0.58 (95% CI: 0.57, 0.59) (Web Figure 1) relatively unaltered.

To enumerate the impact of variants on familial CRC risk, we calculated standardized incidence ratios in first-degree relatives stratified by the number of risk alleles carried by index cases. The 8,878 CRC cases provided data on a total of 50,924 first-degree relatives. Standardized incidence ratios were 1.76, 2.08, and 2.25 for first-degree relatives of index cases with ≤17, 18–21, and ≥22 risk alleles, respectively (Table 4). Increased familial CRC risk by age 70 was associated with an increased number of risk alleles, with first-degree relatives of index cases with ≥16 risk alleles having cumulative increased risks corresponding to a 4% (95% CI: 3, 5), 5% (95% CI: 4, 5), and 5% (95% CI: 5, 6), respectively (Figure 3). Collectively the 14 SNPs account for 4% (95% CI: 0.0, 5.4; P = 0.02). There was evidence that the collective contribution of the 14 variants to familial CRC risk is stronger for early onset CRC disease: 6.0% (95% CI: 0.0, 15.3) of the familial risk when cases were diagnosed at ≤45 years compared with 5.3% (95% CI: 0.0, 10.5) if 46–54 years and 1.3% (95% CI: 0.0, 4.3) if ≥55 years, consistent with genetic enrichment with early onset disease.

**DISCUSSION**

This analysis provides further support for a role of genetic variation at these 14 chromosome regions as risk factors for CRC. Although we found no strong evidence for deviation from additivity, some SNPs clearly have stronger effects than others on CRC risk. Although each of the SNPs was only moderately associated with CRC, they have a strong cumulative association with CRC risk.

To date, a functional basis for only the 8q23.3, 8q24.21, and 18q21.1 risk loci has been elucidated (27–29). rs6983267 mapping to 8q24.21 has been shown to directly interact with the MYC promoter (29); similarly, the 8q23.3 and 18q21.1 loci associations appear to be mediated through variants having cis-regulatory effects on the eukaryotic translation initiation factor 3 subunit H, gene (EIF3H) (27) and mothers against decapentaplegic homolog 7 gene (SMAD7) (28) expression, respectively. Although many of the 14 risk variants identified appear to have generic effects on CRC risk, it is entirely probable that variants may define different etiologic pathways of CRC development. Although our study was large, it may have been insufficiently powered to demonstrate some subtle genotype-phenotype relations. It is, however, noteworthy that both rs4444235 and rs4939827, which map to the bone morphogenetic protein 4 gene (BMP4) and SMAD7, appear to be preferentially associated with the development of microsatellite stable CRC. This is consistent with the observation that germ-line mutation in the TGF-β superfamily-signaling pathway genes is associated with microsatellite stable CRC (30–32).

It is intriguing that the 11q23.1 risk variant is more common in rectal rather than colonic disease, as there are differences in the biology of these tumors in terms of both mutational spectra and environmental risk factors (33, 34). There is little evidence to date that any of the 14 variants influence disease progression per se. Specifically, although an association between the risk allele at 8q23.3 and tumor stage has been reported (14), we found no evidence to support this observation. Finally, the sex-specific associations observed between some SNPs and CRC risk raise the possibility that such variation may underscore some of the differences in CRC incidence between genders. It is notable in this respect that rs961253, rs992918, and rs4925386 map to genes encoding components of the extended TGF-β signaling pathway, raising the possibility of sex-mediated differential effects.

The major strengths of this study are its size and its population-based design. The familial CRC risk estimated from N스크G data is identical to that obtained from meta-analyses of epidemiologic case-control and cohort studies (35, 36), and this makes preferential selection of familial CRC unlikely. Similarly, survivorship is unlikely to have influenced study findings, even though case selection in N스크G is biased to Dukes’ stages A and B disease. Furthermore, this analysis is unlikely to be confounded by significant population stratification, as the analysis has been limited to individuals with self-reported European ethnicity.

One limitation of our study is that it was not a prospective study, and therefore the predictive power of the 14 risk variants we enumerated is an estimate. Although the results apply only to a single population, it is likely that our prediction estimates are reasonably accurate and that the effect sizes observed are representative of those that would be obtained in other European populations. They may not, however, apply to populations with substantially different ethnic origin or those having different environmental exposures. Although associations at 8q24.21, 10p14, 11q23.1, 12q13.13, 15q13.3, and 18q21.1 have been validated in Chinese populations (37, 38),...
little support for the same SNPs influencing CRC in African Americans has yet been provided (39). It is likely that the impact of risk variants will be contingent upon interaction with nongenetic risk factors. Thus, epidemiologic risk factor data should ideally be taken into consideration to allow the examination of interactions between known etiologic factors (e.g., dietary risk factors) and genetic risk variants. Unfortunately, the data available within the current study did not allow such an analysis.

Recent success in identifying SNPs associated with cancer risk through genome-wide association studies has led to the suggestions that such SNPs may be useful in predicting an individual’s cancer risk. In this study, we evaluated the ability of the 14 confirmed risk variants as risk predictors for CRC. The 14 currently identified risk variants have inadequate discriminatory ability for use in clinical practice for individualizing risk, and this is reflected in the contribution they make to the familial risk of CRC. As the area under the curve for the 14 risk variants so far identified is 0.58, genetic tests offered by many commercial companies to individuals have very limited predictive value. The absence of family history data in controls precluded comparison of the discriminatory power of the 14 SNPs in conjunction with family history data. Even so, on the basis of our findings, the currently identified 14 variants have potential to allow subgroups of the population at distinctly different CRC risks to be distinguished, which may have direct utility in the public health arena by facilitating stratifying population programs aimed at reducing CRC incidence. Even though most of the SNPs are themselves not directly functional and that it is likely that improved definition

**Figure 2.** Plots showing the increasing odds ratios for colorectal cancer with the increasing number of risk alleles, London, United Kingdom, 1999–2007. The odds ratios are given relative to the median number of 14 risk alleles (A) and the quintile containing the median number of 14 risk alleles (B). The vertical bars represent 95% confidence intervals. The horizontal line denotes the null value (odds ratio = 1.0).
of risk will be forthcoming with the enumeration of the causal variants at each of the loci within the European population, testing for the current set of 14 SNPs has potential health-care benefits. For example, in many Western countries, fecal occult blood testing or endoscopic surveillance is offered or advocated for individuals aged 50–55 years in the population. On the basis of absolute risk, our findings indicate that the population carrying the number of risk alleles within the upper quintile could profitably be offered such surveillance 5 years earlier (Figure 4). In a similar fashion, polygenic susceptibility defined by SNP genotype has recently been proposed to have implications for population breast and prostate screening (40). However, as previously commented, it will be important to test whether new markers provide improvements in the discriminative or predictive accuracy beyond existing predictive models for cancer risk (41), which, in the case of CRC, can be relatively inexpensive to formulate.

In conclusion, although the combined information from the currently known susceptibility variants has poor predictive value for individual CRC risk discrimination, these variants may allow the identification of subgroups of the population at a substantially increased risk of CRC. Although the calculated multilocus population attributable risk derived here from the 14 SNPs is high, a large population attributable risk does not preclude the possibility that other unidentified risk loci and environmental factors substantially contribute to disease risk (42). Large-scale, ongoing meta-analyses of CRC genome-wide association studies are likely to lead to the identification of additional risk variants. As more risk variants are identified and the functional variants at each locus are unmasked, tests with better predictive performance are likely to become available (43), potentially providing a valuable addition to prevention strategies for CRC and other malignancies (42).

**Figure 3.** Colorectal cancer risk in first-degree relatives of cases stratified by number of risk alleles carried by cases, London, United Kingdom, 1999–2007. Risk at ages 40, 60, and 70 years: for ≤12 risk alleles, 0.6% (95% confidence interval (CI): 0.3, 1.2), 1.7% (95% CI: 1.0, 2.7), and 3.8% (95% CI: 3.1, 4.7); for 13–15 risk alleles, 0.2% (95% CI: 0.1, 0.3), 2.1% (95% CI: 1.9, 2.3), and 4.5% (95% CI: 4.0, 5.0); and for ≥16 risk alleles, 0.3% (95% CI: 0.2, 0.5), 2.0% (95% CI: 1.5, 2.6), and 4.9% (95% CI: 4.5, 5.5). Population risk at ages 40, 60, and 70 years, 0.1%, 0.8%, and 2.2%, respectively (for 1998).

**Figure 4.** Colorectal cancer risk stratified by number of risk alleles carried, London, United Kingdom, 1999–2007. Colorectal cancer screening via fecal occult blood testing or endoscopic surveillance is offered or advocated for individuals aged 50–55 years in many Western countries. On the basis of absolute risk findings, members of the population carrying the number of risk alleles within the upper quintile (dashed lines) could profitably be offered such surveillance 5 years earlier than members carrying the median number of risk alleles (solid lines).

### Table 4. Relation Between Familial Colorectal Cancer Risk and Number of Risk Alleles, London, United Kingdom, 1999–2007

<table>
<thead>
<tr>
<th>Age of Relative, years</th>
<th>≤17</th>
<th>18–21</th>
<th>≥22</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Observed CRC Cases, no.</td>
<td>SIR</td>
<td>95% CI</td>
<td>Observed CRC Cases, no.</td>
</tr>
<tr>
<td>&lt;55</td>
<td>42</td>
<td>2.82*</td>
<td>2.03, 3.81</td>
<td>195</td>
</tr>
<tr>
<td>≥55</td>
<td>86</td>
<td>1.49*</td>
<td>1.19, 1.84</td>
<td>459</td>
</tr>
<tr>
<td>All</td>
<td>128</td>
<td>1.76*</td>
<td>1.47, 2.09</td>
<td>654</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; CRC, colorectal cancer; SIR, standardized incidence ratio.

* P < 0.001.
ACKNOWLEDGMENTS

Author affiliations: Section of Cancer Genetics, Institute of Cancer Research, Sutton, Surrey, United Kingdom (Steven J. Lubbe, Maria Chiara Di Bernardo, Peter Broderick, Ian Chandler, Richard S. Houlston); and Department of Histopathology, Royal Devon and Exeter Hospital, Exeter, Devon, United Kingdom (Ian Chandler).

This work was supported by Cancer Research UK (C1298/A8362—Bobby Moore Fund for Cancer Research UK), CORE, and the National Cancer Research Network. S. L. is in receipt of a PhD studentship from Cancer Research UK. The authors thank Steven Penegar for his contributions toward this study. Finally, the authors acknowledge National Health Service funding to the National Institute for Health Research Biomedical Research Centre.

Conflict of interest: none declared.

REFERENCES


