Investigation of the effects of DNA repair gene polymorphisms on the risk of colorectal cancer

Ian P. M. Tomlinson*, Richard S. Houlston1, Grant W. Montgomery2, Oliver M. Sieber3 and Malcolm G. Dunlop4

Molecular and Population Genetics Laboratory and Oxford NIHR Comprehensive Biomedical Research Centre, Nuffield Department of Clinical Medicine, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK, 1Section of Cancer Genetics, Institute of Cancer Research, Cotswold Road, Sutton SM2 5NG, UK, 2Genetic and Molecular Epidemiology Laboratories, Queensland Institute of Medical Research, 300 Herston Road, Herston Q4006, Australia, 3Ludwig Colon Cancer Initiative Laboratory, Ludwig Institute for Cancer Research, Royal Melbourne Hospital, Parkville, Victoria, Australia and 4Ludwig Colon Cancer Genetics Group, Institute of Genetics and Molecular Medicine, Western General Hospital, University of Edinburgh and MRC Human Genetics Unit, Edinburgh EH4 2XU, UK

*To whom correspondence should be addressed. Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Roosevelt Drive, Oxford OX3 7BN, UK, Tel: +44 (0)1865 287720; Fax: +44 (0)1865 287501; Email: iant@well.ox.ac.uk

Received on August 28, 2011; revised on September 13, 2011; accepted on September 15, 2011

Despite their prime candidate status, polymorphisms near genes involved in DNA repair or in other functions related to genome stability have been conspicuously under-represented in the significant associations reported from genome-wide association studies (GWAS) of cancer susceptibility. In this study, we assessed a set of single-nucleotide polymorphisms (SNPs) near 157 DNA repair genes in three colorectal cancer (CRC) GWAS. Although no individual SNP showed evidence of association, the set of SNPs as a whole was associated with colorectal cancer risk. When candidate SNPs were examined, our data did not support most of the previously reported associations with CRC susceptibility, an exception being an effect of the MLH1 promoter SNP −93G>A (rs1800734). Rare variants in CHEK2 (I157T and possibly del1100C) also appear to be associated with CRC risk. Overall, the absence to date of disease-associated DNA repair SNPs in cancer GWAS may be explained by a combination of the following: (i) many loci with individually very small effects on risk; (ii) rare alleles of moderate effect and (iii) subgroups of CRC, such as those with microsatellite instability, associated with specific variants. It will be particularly intriguing to determine whether any GWAS across cancer types identify DNA variants that predispose to cancers of more than one site.

Introduction

Colorectal cancer genetics

Colorectal carcinoma (CRC) affects 5–6% of the Western world population during their lifetime, with a small male preponderance. In Europe, CRC causes more deaths than any other cancer in the absence of a major avoidable risk factor such as smoking. Although environmental factors, such as diet, influence CRC risk, specific agents and mechanisms of action have been difficult to pin down. Inherited factors also play an important role in CRC predisposition, with a family history of the disease probably being the strongest, easily quantifiable risk factor (apart from a previous personal history of colorectal tumours). Most CRCs probably develop from benign lesions, mainly adenomatous polyps. Screening for adenomas and subsequent removal by colonoscopy or sigmoidoscopy can reduce CRC risk. Many Western countries now have population screening programmes based on these techniques or faecal occult blood testing. Thus, while it is expected that deaths from CRC will fall in Europe as population screening becomes more effective, CRC will continue to be a major killer and the public health burden of colorectal screening is likely to increase considerably.

The risks of many common diseases, including cancers, are known to depend on inherited factors. Recent progress has led to the identification of panels of common genetic variants that influence risk. For CRC, the known predisposition genes are of two types. Firstly, there patients who harbour high-penetrance mutations that cause rare Mendelian syndromes, often in the familial setting; these include CRC syndromes such as familial adenomatous polyposis (FAP), Lynch syndrome, MUTYH-associated polyposis (MAP) and juvenile polyposis (JPS). Secondly, individuals may carry common genetic variants associated with modest increases in CRC risk in the general population.

The identification of common polymorphisms associated with CRC risk has occurred in two phases. Initially, such studies were confined to the analysis of candidate polymorphisms, but studies were small and, in retrospect, probably under-powered. It is debatable whether any of the reported associations were demonstrated to sufficiently stringent levels of evidence to be regarded as ‘proven’ (1, 2). However, more recently, a genome-wide hypothesis-free approach has been employed in which haplotype tagging single-nucleotide polymorphisms (tagSNPs) are used to screen for associations based on linkage disequilibrium (LD) mapping. For CRC, such genome-wide association studies (GWAS) have identified 17 tagSNPs associated with CRC risk modulation at stringent levels of significance ($P < 5 \times 10^{-8}$) (3–6). The studies have involved Discovery Phases based on hundreds of thousands of tagSNPs typed in a total of ~10 000 CRC cases and controls, and multiple Validation Phases involving tens of thousands of cases and controls derived from the UK and COGENT collaborators throughout Europe (5, 7).

Functions of CRC-associated single-nucleotide polymorphisms

The nature of the functional variation that is correlated with these single-nucleotide polymorphisms (SNPs) is largely unknown, and studies to identify such variation can be difficult
and time-consuming. For one of the CRC tagSNPs, rs6983267, there is evidence that it itself is functional (8). It lies a few hundred kilobases upstream of the c-Myc oncoprotein and changes a TCF4-binding site. The region around the SNP is capable of driving reporter gene expression in the same locations as endogenous Myc. A further CRC tagSNP, rs4779584, near Gremlin1 is an example of a synthetic association. Its association with CRC risk actually results from two independent functional variants tagged by rs16969681 and rs11632715. rs16969681 may in fact be functional (4).

Overall, GWAS in CRC have provided excellent evidence of the important molecular pathways that influence disease risk in the general population. Of the 17 SNPs above, 7 are adjacent to genes (Gremlin1, BMP2/4, SMAD7) that act in the bone morphogenetic protein (BMP) pathway. Although in theory the tagSNPs identified may act on other genes in cis or in trans, the probability of so many being near BPM genes is $10^{-9}$. Moreover, two other SNPs are near genes (Laminin A5 and DIP2B) for which there is evidence for interaction with the BMP pathway.

A further CRC tagSNP, rs10936599, lies within an intron of the myoeurin gene (9). However, the region is gene-rich and SNPs in very strong LD with rs10936599 lie very close to three or four other genes. The strongest candidate gene near rs10936599 is the telomerase RNA component (TERC), which encodes the RNA template on which telomerase reverse transcriptase acts to maintain chromosome ends. Detailed analyses have shown that the risk allele at rs10936599 is in perfect LD with an SNP within the TERC transcript. This allele is associated not only with CRC risk but also with longer telomeres in both peripheral blood and normal colorectal mucosa. Longer chromosome telomeres are, other things being equal, predicted to reduce cellular senescence and apoptosis. As is the case for the BMP pathway SNPs, TERC variation increases the risk of colorectal adenomas as well as CRCs, indicating an effect at early stages of tumorigenesis. Epidemiological evidence regarding associations between telomere length and CRC risk has been mixed, but the genetic findings have provided excellent evidence in favour of long telomeres being associated with disease.

DNA repair and CRC risk
Carcinogenesis generally depends on the acquisition of somatic mutations by a susceptible cell. On this basis, it is to be expected that individuals with profound defects in DNA repair will be at greatly increased risk of cancer and other malignancies. Many examples of patients with rare, but highly penetrant, germ line mutations in the major DNA repair pathways do indeed exist and almost all these patients have an increased risk of cancer. Mutations in genes involved in maintaining genome integrity, through processes such as normal chromosome segregation, also increase the risk of cancer. For reasons that remain partly obscure, different types of DNA repair defect predispose to different cancer spectra. Colorectal cancer, for example, can be caused by two forms of high-penetration DNA repair defect: recessive mutations in the DNA glycosylase MUTYH which cause defective base excision repair; and dominant mutations in any of the four DNA mismatch repair genes (MLH1, MSH2, MSH6 and PMS2). Some variants in genes such as MUTYH, MSH2 and MLH1 have been proposed to act as low- or moderate-penetrance susceptibility alleles for CRC, although evidence in their favour is generally mixed (see below).

In fact, DNA repair and genome integrity SNPs—henceforth called ‘DNA repair SNPs’—are striking by their absence in the list of common CRC predisposition variants no known DNA repair gene is tagged by the known predisposition SNPs for CRC. This has several potential causes. One highly plausible explanation is that DNA repair and genome integrity are so important to the cell and organism that any alleles with more than very small functional effects are strongly selected against. Such alleles are very unlikely to drift up to polymorphic levels. Nevertheless, common DNA repair alleles with very small effects on cancer risk may still exist, yet their effects are too small—with odds ratios (OR) <1.1 per allele—to be detected individually even in the several thousand samples employed for GWAS. En masse (or as a class), however, an effect of DNA repair SNPs on cancer susceptibility might be detectable.

We have therefore analysed our data from three GWAS from the UK to search specifically for effects of DNA repair SNPs on CRC risk.

Methods
Patients were recruited from three GWAS, the ColoRectal tumour Gene Identification (CORGI) study, the Scottish Colorectal Cancer Study (COOS) and the VICTOR/QUASAR2/1958 Birth Cohort data sets (VQ58). These cohorts comprised a total of 3334 CRC (or severe adenoma) cases and 4628 controls; further details are provided in (4). All studies received appropriate local research ethics committee approval.

Methods for genotyping using Illumina Hap550 and Hap300 SNP arrays have been described previously (4). All data had previously been subjected to standard, but rigorous quality control assessment (4) and converted into PLINK binary file format for analysis (http://pngu.mgh.harvard.edu/purcell/plink/).

One hundred and fifty-seven DNA repair genes and other genes involved in maintaining genome integrity were identified from a published list (10) (supplementary Table 1 is available at Mutagenesis Online). Each gene from this list had a set of SNPs retrieved using dbSNP126 and HapMap Release 24 Phase II by inputting genomic co-ordinates flanking the coding region of each gene by ±100 kb to cover control regions. SNPs within these regions flanking each gene were identified from the content of the Hap300 and/or Hap550 arrays.

Basic association analysis was performed for each autosomal SNP using the PLINK package. QQ-plots of the chi-squared association statistic were constructed for each cohort and we found $\lambda_{pc} < 1.05$ in all three cases (details not shown). Association tests for individual SNPs were carried out using allelic, Cochran-Armitage and recessive models. Meta-analysis was performed using the Mantel–Haenszel method on all SNPS present in two or all data sets. We reported fixed effects statistics unless there was evidence of heterogeneity among sample sets ($P_{het} < 0.05$ or $I^2 > 75\%$), in which case random effects statistics were reported. For the whole set of SNPs, a simple consistency test between the three data sets was performed based on direction of effect and an expectation that one quarter of SNPs would show the same direction by chance. Set-based tests were also performed on all SNPs and those most strongly associated with CRC in PLINK (http://pngu.mgh.harvard.edu/~purcell/plink/anal.shtml#set), which calculates a statistic for each set as the mean of these single-SNP statistics, and determines significance using permutation of case–control status. SNPs in LD ($r^2 > 0.2$) were removed from this part of this analysis.

Results
We identified 6216 SNPs within 100 kb of the 157 DNA repair loci that were present on the Illumina Hap550 arrays; ~60% of these SNPs were also present on the Hap370 arrays. Following meta-analysis in the three GWAS series, only one SNP (Table 1) showed evidence of association with CRC risk under the allelic, additive or recessive models using a threshold of $I_{meta} = 10^{-4}$ (approximating to a Bonferroni correction at $P = 0.05$). However, this SNP was rs10411210, a known CRC predisposition SNP that lies within an intron of rhophilin2 (RHPHN2) but is also within 100 kb of the DNA repair gene FAAP24. We have previously presented evidence that favours
Table I. Individual DNA repair tagSNP association statistics

<table>
<thead>
<tr>
<th>Chr</th>
<th>BP</th>
<th>SNP</th>
<th>P</th>
<th>Q</th>
<th>$I^2$</th>
<th>Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>38224410</td>
<td>rs10411210</td>
<td>1.27 x 10^{-5}</td>
<td>0.6882</td>
<td>0.3106</td>
<td>2.75</td>
</tr>
<tr>
<td>11</td>
<td>118569768</td>
<td>rs12417928</td>
<td>1.03 x 10^{-4}</td>
<td>1.3787</td>
<td>0.6793</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>20094459</td>
<td>rs1243647</td>
<td>4.00 x 10^{-4}</td>
<td>0.8168</td>
<td>0.6255</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>38237541</td>
<td>rs6510337</td>
<td>5.83 x 10^{-4}</td>
<td>0.8456</td>
<td>0.7768</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>48854794</td>
<td>rs2239372</td>
<td>7.78 x 10^{-4}</td>
<td>1.104</td>
<td>0.4561</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>32893868</td>
<td>rs1470217</td>
<td>9.28 x 10^{-5}</td>
<td>1.1212</td>
<td>0.466</td>
<td>0</td>
</tr>
</tbody>
</table>

Chr, chromosome; BP, position in base pairs (dbSNP Build 36); P, meta-analysis P value; OR, odds ratio per allele relative to the minor allele; $Q$, $P_{het}$, $I^2$, % heterogeneity; Gene, DNA repair gene in region of SNP.

Table II. Associations between CRC susceptibility and Candidate SNPs chosen in previously published studies

<table>
<thead>
<tr>
<th>Candidate SNP</th>
<th>Annotation</th>
<th>Proxy SNP</th>
<th>LD of proxy</th>
<th>OR</th>
<th>$P_{meta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1800734</td>
<td>MLH1-93G &gt; A</td>
<td>Present on Hap300/Hap550 arrays</td>
<td>1.07</td>
<td>0.0726</td>
<td></td>
</tr>
<tr>
<td>rs1799977</td>
<td>MLH1 Ile219Val</td>
<td>Present on Hap300/Hap550 arrays</td>
<td>0.99</td>
<td>0.800</td>
<td></td>
</tr>
<tr>
<td>rs2303425</td>
<td>MSH2 -118T&gt;C</td>
<td>rs10495944</td>
<td>$r^2 = 0.95, D’ = 1.00$</td>
<td>1.02</td>
<td>0.817</td>
</tr>
<tr>
<td>rs1981928</td>
<td>MSH2 intronic</td>
<td>rs3732183</td>
<td>$r^2 = 1.00, D’ = 1.00$</td>
<td>1.00</td>
<td>0.999</td>
</tr>
<tr>
<td>rs1042821</td>
<td>MSH6 Gly391Glu</td>
<td>rs6713506</td>
<td>$r^2 = 0.30, D’ = 0.76$</td>
<td>1.06</td>
<td>0.831</td>
</tr>
<tr>
<td>rs26279</td>
<td>MSH3 Ala1045Thr</td>
<td>Present on Hap300/Hap550 arrays</td>
<td>1.06</td>
<td>0.0814</td>
<td></td>
</tr>
<tr>
<td>rs1979005</td>
<td>MSH3 intronic</td>
<td>rs3776969</td>
<td>$r^2 = 0.87, D’ = 1.00$</td>
<td>0.94</td>
<td>0.370</td>
</tr>
<tr>
<td>rs1042522</td>
<td>TP53 Arg272Pro</td>
<td>rs7141</td>
<td>$r^2 = 0.45, D’ = 0.88$</td>
<td>1.01</td>
<td>0.837</td>
</tr>
<tr>
<td>rs1800056</td>
<td>ATM Phe858Leu</td>
<td>rs4987876</td>
<td>$r^2 = 0.21, D’ = 1.00$</td>
<td>0.97</td>
<td>0.636</td>
</tr>
<tr>
<td>rs2308321</td>
<td>MGMT Ile174Val</td>
<td>Present on Hap300/Hap550 arrays</td>
<td>1.00</td>
<td>0.989</td>
<td></td>
</tr>
<tr>
<td>rs25487</td>
<td>XRC1C Glu399Arg</td>
<td>rs1799778</td>
<td>$r^2 = 0.97, D’ = 1.00$</td>
<td>0.98</td>
<td>0.488</td>
</tr>
<tr>
<td>rs3218536</td>
<td>XRC2C R188H</td>
<td>Present on Hap300/Hap550 arrays</td>
<td>1.04</td>
<td>0.599</td>
<td></td>
</tr>
<tr>
<td>rs3218499</td>
<td>XRC2C intronic</td>
<td>rs3218408</td>
<td>$r^2 = 0.95, D’ = 1.00$</td>
<td>0.99</td>
<td>0.868</td>
</tr>
</tbody>
</table>

Results shown are from the meta-analysis of the three GWAS data sets. Note that for rs26279, the direction of effect is opposite to that found in other studies.
greater pairwise LD ($r^2 > 0.2$), the overall test statistic was significant ($P = 0.00799$). Thus, there may be an effect of some DNA repair SNPs on CRC risk, yet the effect of any individual SNP is too small to readily be detected in individual SNP analyses in the sample sets used.

Discussion

Candidate gene analyses have failed to provide convincing evidence of DNA repair SNPs that are directly involved in predisposition to colorectal cancer. Here, we have analysed a panel of tagSNPs flanking DNA repair genes and failed to find good evidence that any individual SNP influences the risk of CRC in the general UK population. In general, GWAS of other cancers have failed to find similar associations. In fact, the only cancer susceptibility GWAS that has found variation likely to affect DNA repair is one on breast cancer (SNP rs999737 near RAD51L1) (39). We have, however, provided evidence using en masse (SNP set) analysis that variation around DNA repair loci does contribute to CRC risk. Clearly, set-based tests must be interpreted cautiously, but the examples of Lynch syndrome and MAP show that DNA repair variants can influence CRC risk. The possibility clearly exists that several DNA repair SNPs have small effects on CRC risk—probably too small to find easily using GWAS—and furthermore that an individual’s overall DNA repair capacity (40) might be associated with CRC susceptibility.

DNA repair capacity, measured by assays such as micronucleus formation and radiosensitivity, has been reported as highly heritable (41–43). Even if these results accurately reflect more general variation in DNA repair, they are not necessarily incompatible with the findings that common cancer susceptibility alleles have only rarely been found near DNA repair genes. Potential non-exclusive explanations for this apparent inconsistency include the following:

(i) heritability results from low variation in intrinsic DNA repair capacity, caused by a small genetic contribution, but even smaller environmental and random contributions;

(ii) polygenic model: many loci with individually very small effects influence DNA repair capacity;

(iii) multiple rare alleles of moderate effect influence DNA repair capacity, these alleles not drifting to high frequencies owing to strong selective constraints;

(iv) sub-groups of CRC, such as those with microsatellite instability, are associated with specific DNA variants, but most studies are not empowered to detect these effects; and

(v) the relevant DNA repair variants are not present on or tagged by commercial SNP array content.

The set test results reported above support (ii) and the CHECK2 data support (iii). The data in support of an association between MLH1-93G > A and microsatellite-unstable CRC provide evidence for (iv).

GWAS in CRC have not generally discovered variation in pathways that could be described as strong candidates based, for example, on previous studies of individual genes. For example, none of the nine main Mendelian CRC genes has been shown to harbour low-penetrance susceptibility variants to date. CRC GWAS have, however, thrown up a new list of candidate genes, and studies based on those have already been successful. In other cases, unexplained and unexpected associations have been found. For example, while bearing in mind that the genes influenced by functional variation may not be those nearest to the GWAS tagSNP, it seems likely that variation in EIF3H (44), an apparently ubiquitously expressed transcription elongation factor, has specific effects on CRC risk; the reasons for this are entirely unclear. Despite these examples, examination of variation in DNA repair genes remains a priority in forthcoming studies to detect and examine rarer cancer predisposition variants through screens based on genome-wide or focussed large-scale sequencing. Analysis of CRC risk in carriers of heterozygous MUTYH mutations, where ~1% of the population are at hypothetically increased risk, has shown both the promise and the inherent problems of such rare variant studies. It will also be intriguing to determine whether any large GWAS across cancer types can identify DNA repair variants that predispose to cancers of more than one site.

Supplementary data

Supplementary Table 1 is available at Mutagenesis Online.

Funding

This work was supported by Cancer Research UK.

Acknowledgements

Collaborators in the CORGI, NSCCG and COGS studies provided invaluable help with patient recruitment and sample collection. Members of the contributing laboratories undertook invaluable work in patient recruitment and sample preparation. Andrew Beggs kindly supplied the list of DNA repair genes.

Conflict of interest statement: None declared.

References


