Colorectal carcinoma (CRC) remains a frequent cause of cancer-associated mortality in the UK and still has a relatively poor outcome. Single gene defects account for up to 2–6% of cases, but twin studies suggest a hereditary component in 35%. CRC represents a paradigm for cancer genetics. Almost all the major-gene influences on CRC have been identified, and the identification of the remaining susceptibility alleles is proving troublesome. Only a few low-penetrance alleles, such as methylene tetrahydrofolate reductase C677T, appear convincingly to be associated with CRC risk. To identify the remaining CRC genes, parallel approaches, including strategies based on linkage and association and complementary analyses such as searches for modifier genes, must be employed. To gain sufficient evidence to prove that a gene is involved in CRC predisposition, it is probably necessary for multiple, adequately-powered studies to demonstrate an association with the disease, especially if the allelic variants have only a small differential effect on risk. It may also be possible to show how genes interact with each other and the environment, although this will be even more difficult. Accurate quantitation of the allele-specific risks in different populations will be necessary, but problematic, especially if those risks combine in a fashion which is not of a straightforward additive or multiplicative type. Without any good prior evidence of the nature of the remaining genetic influence on CRC, the possibility remains that this is a truly polygenic trait or that multiple, rare variants contribute to the increased risk; in these cases, identification of the genes involved will be very difficult. Despite these potential problems, the effectiveness of preventive measures for CRC, especially in high-risk individuals, means that the search for new predisposition genes is justified.
(reviewed in 5). Furthermore, hereditary factors may influence the severity of the disease in the Mendelian CRC syndromes, as exemplified by the number of colorectal polyps presenting within FAP families or in individuals with identical germ line APC mutations (6).

There are two factors that, in combination, make the search for additional CRC predisposition genes so important. First, identification and removal of colorectal adenomas (and possibly other types of bowel predisposition) almost certainly greatly reduce cancer incidence; identification of cancers before they present clinically may have similar benefits. Second, regular, frequent and effective screening for colorectal tumours is not currently available at the population level. The identification of controlling genetic factors therefore allows individual cancer risk assessment and the targeting of screening to those at highest risk; there may also be spin-offs in terms of effective chemoprevention and the development of new therapies. The challenge, therefore, is to identify all the relevant genes, and to specify their contribution in controlling CRC predisposition and severity in human populations. A major part of this challenge is to choose the most appropriate methods to locate and identify novel susceptibility genes.

In this short review, we shall detail some of the recent developments in the study of inherited susceptibility to CRC. Progress has been made in the understanding of the Mendelian syndromes and towards the identification of the small number of remaining high-penetrance CRC genes. We shall also present recent evidence for low-penetrance CRC alleles and for modifier genes of the Mendelian syndromes’ phenotypes. We complete our review by outlining future strategies for use in the continued search for genes predisposing to CRC.

**PROGRESS IN UNDERSTANDING THE HEREDITARY COLORECTAL TUMOUR SYNDROMES**

**MYH-associated polyposis**

MYH is a DNA glycosylase involved in the response to oxidative damage of DNA by base excision repair (BER). MYH was identified as a recessive CRC and adenoma predisposition gene by Al-Tassan et al. (7), who found an unusual somatic APC mutation spectrum (mostly G → T changes) in the polyps of a family that they were investigating for the pathogenic effects of the APC missense variant E1317Q. The MAP phenotype generally resembles attenuated (<100 adenomas) or mild, classical (100–1000 adenomas) FAP (8), although MAP tumours seem to be confined to the gastrointestinal tract. MAP individuals with early-onset CRC but no detected colorectal adenomas have been reported (9), although small adenomas can be missed using colonoscopy without dye-spray. MAP accounts for about 1% of all CRCs and one-third of patients with 15–100 colorectal adenomas (8,10). MYH screening should be considered in patients with multiple colorectal adenomas, although APC remains as a much stronger candidate in families with vertical transmission of disease. Testing MYH in patients with early-onset CRC and no adenomas would be justified if further data support those of Wang et al. (9). Mutation testing should be tailored to the ethnic group, with mutations Y165C and G382D more common in Caucasians, Y90X in Pakistanis, E466X in Indians and nt1395delGGA in those of Mediterranean origin (10–13). The site-specific nature of MAP tumours remains puzzling. BER deficiency does not specifically target APC and many MAP tumours harbour specific G → T changes in K-ras (14). The few MAP cancers studied so far are microsatellite-stable and near-diploid (14).

**AXIN2**

Apart from APC, studies have generally failed to find germline mutations in other components of the Wnt pathway (such as beta-catenin, APC2, axin, conductin (axin2) and glycogen synthase kinase 3-beta) in patients with multiple or early-onset colorectal tumours (15). Recently, however, a Finnish family with dominantly-transmitted oligodontia and colorectal polyps was shown to carry a germline R656X mutation in AXIN2 (16). The colorectal phenotype in the family was highly variable, ranging from 68 adenomas through 10 hyperplastic polyps to no disease. Loss of the wild-type AXIN2 allele was not tested in the polyps.

**Peutz–Jeghers syndrome**

Germline mutations of the serine/threonine kinase gene LKB1(STK11) cause PJS. PJS patients develop harmatomatous...
polyps of the gastrointestinal tract, and characteristic freckling, and are at increased risk of colorectal and other cancers. LKB1 has been reported to function in a variety of pathways, including those involving mTOR/AMP-activated protein kinase, Wnt, COX2, NF-kappaB, VEGF, cell polarity and p53 (17–25). It remains to be seen whether mutations in such genes play a role in PJS. From a strictly genetic perspective, there is ongoing controversy as to the existence of a second PJS locus, primarily because LKB1 mutations cannot be found in 30–80% of PJS patients. PJS kindreds unlinked to LKB1 (26) and some linked to 19q13 (27) have been reported, but no second locus has been identified. In this regard, it should be noted that, although PJS polyps are probably pathognomonic, several conditions have pigmented lesions that may mimic PJS (28).

Hereditary non-polyposis CRC

Hereditary non-polyposis colon cancer (HNPPC) is an autosomal dominant disorder characterized by early-onset CRC (often right-sided), few adenomas and specific extra-colonic cancers, including those of the ovary, stomach, small bowel and uroepithelial tract. HNPPC accounts for 1–5% of CRCs and results from germline mutations in MLH1 and MSH2 in about 90% of cases and more rarely in PMS2 and MSH6 (remaining 10% of cases). HNPPC cancers are characterized by microsatellite instability (MSI). The MMR genes are not mutated in sporadic CRCs, although MLH1 is silenced by promoter methylation in 10–15% of these tumors, most of which are right-sided and tend to occur in older females. In the light of these data, Suter et al. (29) hypothesized that methylation of the MLH1 promoter could occur in the germline and act as an alternative to mutation in predisposing to HNPPC. They found evidence of such a MLH1 'epimutation' in two patients with multiple, MSI+ primary tumours which showed mismatch repair deficiency. The epimutation occurred as a mosaic and was also present in spermatozoa of one of the individuals, indicating a germline defect and the potential for transmission to offspring. It was not clear whether an uncharacterized, primary genetic defect had led to the epimutation. Other workers have focussed on the role of additional MMR genes in the HNPPC families, with evidence that MLH3 (30–32) and EXO1 (33–36) might be involved in a small number of cases, although data are inconsistent. There is also accumulating evidence to show that the genetic pathways of sporadic, MSI+ CRCs and HNPPC CRCs are non-identical. Johnson et al. (37), for example, showed that beta-catenin mutations were specific to MSI+ carcinomas from HNPPC, but were rare in the HNPPC adenomas, suggesting that Wnt activation was not always an early event in HNPPC tumorigenesis. Two further issues the functions of MMR proteins outside simple DNA repair (38) and whether or not MMR loss initiates tumorigenesis in HNPPC, remain under close scrutiny.

Familial adenomatous polyposis and APC modifier genes

FAP is caused by germline APC mutations. Affected individuals develop hundreds to thousands of adenomatous colorectal polyps during the second and third decades, progressing to cancer in nearly 100% of cases without treatment. Specific extra-colonic features exist (39). We shall focus on three aspects of recent FAP research. First, it is clear, in both FAP and sporadic CRCs, that APC is not a simple tumour suppressor gene: the ‘two hits’ are strongly associated in terms of their location within the gene (39–42). This association probably provides an optimal level of Wnt signalling for tumour growth, mediated through beta-catenin levels. The mutation spectrum, and presumably optimal level of Wnt signalling, varies among tissues (40,43). The ‘first hit–second hit’ association at APC has also provided a new mechanism for determining disease severity (40), given that certain ‘second hits’ such as allelic loss probably occur spontaneously at a relatively high frequency and are only selected in certain FAP patients (with mutations close to codon 1300), who consequently have severe disease. Second, clues have emerged as to the cause of attenuated disease (AFAP, AAPC) in individuals with germline APC mutations in exons 1–4, 9 and the latter half of exon 15, owing to the discovery of ‘third hits’ in the tumours of some of these patients (44). In order to explain the paradox that a requirement for ‘third hits’ should in theory cause AFAP patients to have no more tumours than the general population, Su et al. (45) suggested that a broader spectrum of somatic mutations was selected in AFAP individuals than in those with no germline APC mutation (although their model was not placed in the context of the ‘first hit–second hit’ associations described earlier). Third, the site of the germline APC mutations cannot explain all the phenotypic variation which is seen in FAP and AFAP. It is likely that modifier genes distinct from APC influence features such as the number of tumours in each region of the gastrointestinal tract (46), the development of desmoids (47) and progression to cancer. In ApcMin mice, a modifier gene, Mom1/Pla2g2a, influences polyp number (48–50). FAP modifier genes in humans are likely to be common polymorphisms, and candidate gene approaches have been used to analyse these. One study has suggested effects of polymorphisms in the N-acetyltransferases (NAT1 and NAT2) on polyp numbers in FAP (51).

APC I1307K

The APC I1307K variant is present in about 6% of Ashkenazi Jews, but is much rarer in those of other ethnic groups. I1307K creates an As tract that appears to be somatically unstable, leading to frameshift mutations (52). Originally, 11307K was proposed to lead to an AFAP-like phenotype, although extra-colonic disease does not appear to feature in this group. Recent data have shown that APC I1307K only confers a small excess risk of CRC (53), inconsistent prima facie with an AFAP-like phenotype. Moreover, only a minority of colorectal tumours from I1307K carriers shows slippage of the As tract. It may be the case, therefore, that although the 1307K allele is somewhat unstable, this effect is subtle in most carriers; only some individuals, perhaps with an unidentified, additional germline factor, go on to develop an AFAP-like phenotype (54).
RECENT DEVELOPMENTS THROUGH LINKAGE ANALYSIS

Hereditary mixed polyposis syndrome

Hereditary mixed polyposis syndrome (HMPS) was described as a distinct syndrome of colorectal adenomas, hyperplastic polyps, atypical ‘mixed’ polyps and carcinoma by Whitelaw et al. (55) in a large Ashkenazi family. The extra-colonic features of FAP and the typical features of PJ were absent. The phenotype had limited similarity to JPS. Linkage of HMPS to chromosome 6q16–q21 was demonstrated by Thomas et al. (56). Tomlinson et al. (57) showed linkage to chromosome 15q14–q15 (CRAC1 locus) in another Ashkenazi family with multiple colorectal adenomas (including serrated lesions) and CRC. The development of multiple adenomas in a HMPS individual without the disease-associated haplotype led to the reassignment and stricter implementation of affection status in that family in order to minimize the phenotype rate. Subsequent linkage analysis showed HMPS and CRAC1 to be one-and-the-same (58). Affected individuals in these families and in several other Ashkenazi kindreds with a similar phenotype were shown to share an ancestral haplotype between D15S1030 and D15S118. The gene itself is, as yet, unidentified and its influence in other population groups, undetermined.

A CRC locus on chromosome 9q?

Wiesner et al. (59) undertook linkage analysis using the affected sibling pair method in 53 families with individuals affected at <65 years of age by CRC or adenomas (>1 cm or displaying high-grade dysplasia). Six potential loci were identified with excess allele sharing among concordantly affected sib-pairs (P < 0.016). Of these, a region on chromosome 9 gave a significant result when additionally examined for deficient allele sharing among discordant sib pairs (P = 0.014). The Haseman–Elston regression method provided an overall probability of linkage of P = 0.00045.

EVIDENCE FOR LOW-PENETRANCE ALLELES ASSOCIATED WITH AN INCREASED RISK OF CRC

Some high- or moderate-penetrance CRC predisposition genes remain to be detected and their pursuit is of great importance. It is generally believed, nevertheless, that low-penetrance variants, such as common alleles at SNPs, are responsible for much of the uncharacterized inherited influence on CRC. Many different SNPs and other common polymorphisms have been evaluated for their effects on CRC risk, and space precludes detailed descriptions or analysis here. Case–control (association) analysis has been used most frequently. Such studies have often been criticized, some of the reasons being the following:

- sample sizes produce insufficient statistical power;
- cases and controls are inappropriately chosen;
- statistical thresholds are chosen with insufficient rigour, so that most reported associations are inevitably false (60);
- the ‘multiple testing’ problem and tendency to inappropriate sub-group analyses are endemic;
- publication bias remains problematic, even though methods have been developed to assess this (61,62);
- studies rarely attempt to replicate their findings in the same population;
- findings which are not replicated in different populations may be inappropriately dismissed;
- the functional effects of variants have often not been clearly shown;
- some methods, such as PCR–RFLP analysis, are prone to systemic error unless control measures are carefully applied.

The recognition of these problems has arguably led to too pessimistic an attitude to studies of low-penetrance variants as CRC susceptibility alleles. Most recent studies have used larger, better defined populations of cases and controls, together with genotyping methods with greater accuracy and throughput. Formal assessments of the roles of several polymorphisms in CRC predisposition were made by Houlston and Tomlinson (63) and by de Jong et al. (64). In summary, there is already good evidence to show that a small number of polymorphisms is associated with a differential risk of CRC (Table 2). These include methylene tetrahydrofolate reductase MTHFR677T, H-ras VNTR rare alleles and N/172 alleles which cause a rapid acetylator phenotype. Other polymorphisms have been tested (Table 2) and many are unproven, although in a small number of cases, loci continue to be cited as being associated with CRC risk despite the evidence in their favour being minimal.

Recent data have suggested that loss of imprinting (LOI) of the insulin-like growth factor II (IGF2) in normal colonic mucosa and/or lymphocytes is indicative of increased CRC risk (89–92). It has been suggested that IGF2 LOI is a genetically determined trait. It is not clear whether the association between IGF2 LOI and CRC might be a direct effect of increased IGF2 levels, or a marker of a general tendency to, for example, hypomethylation or aberrant gene silencing.

Studies are increasingly testing risks associated with combinations of putative CRC genes, or of genes and specific environments, although inherent problems of statistical power mean that most of the reported associations of this type must currently be taken as no more than suggestive. Even for the MTHFR polymorphisms, advice on dietary modification so that folate intake is increased cannot yet be given reliably or with confidence. A further question is whether or not a true CRC risk factor should affect the risk of both rectal and colonic cancers, conditions that are usually treated as essentially the same disease (albeit with some different molecular and clinical features) by geneticists, but as different conditions by epidemiologists. Further studies are beginning to test the influence of germline variation on factors other than disease susceptibility, including the presenting features of CRCs, response to therapy and toxicity (including pharmacogenetic analyses), and recurrence/survival.

In general, acceptance of a polymorphism as a CRC risk factor can only come through repeated analysis by different studies and by consequent slow progress to a consensus judgement. Accurate quantitation of the associated risk will be just
Table 2. Summary of polymorphisms tested for association with differential CRC risk

<table>
<thead>
<tr>
<th>Gene</th>
<th>Name</th>
<th>Polymorphism</th>
<th>Rating</th>
<th>Notes</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADH3</td>
<td>Alcohol dehydrogenase</td>
<td>Codon 350</td>
<td>**</td>
<td>May depend on alcohol intake</td>
<td>(63,64)</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatous polyposis coli</td>
<td>E1317Q</td>
<td>**</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>APOE</td>
<td>Apolipoprotein E</td>
<td>42/3</td>
<td>**</td>
<td></td>
<td>(64)</td>
</tr>
<tr>
<td>BLM</td>
<td>Bloom syndrome</td>
<td>BLM&lt;sup&gt;inh&lt;/sup&gt;</td>
<td>**</td>
<td></td>
<td>(65,66)</td>
</tr>
<tr>
<td>CBS</td>
<td>Cystathionine beta-synthase</td>
<td>Ins68bp x8</td>
<td>*</td>
<td></td>
<td>(67)</td>
</tr>
<tr>
<td>CDDN1</td>
<td>Cyclin D1</td>
<td>A870G</td>
<td>**</td>
<td></td>
<td>(68,69)</td>
</tr>
<tr>
<td>CDH1</td>
<td>E-cadherin</td>
<td>del1100C</td>
<td>*</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>CHEK2</td>
<td>Checkpoint kinase 2</td>
<td>R8W</td>
<td>**</td>
<td>May depend on aspirin intake</td>
<td>(71)</td>
</tr>
<tr>
<td>COX1</td>
<td>Prostaglandin H synthase 1</td>
<td>L15Δ</td>
<td></td>
<td></td>
<td>(71)</td>
</tr>
<tr>
<td>COX2</td>
<td>Prostaglandin H synthase 2</td>
<td>P17L</td>
<td>*</td>
<td></td>
<td>(71)</td>
</tr>
<tr>
<td>CYP1A1</td>
<td>Cytochrome P450 1A1</td>
<td>H426Y</td>
<td></td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>EPHX</td>
<td>Epoxide hydrolase (microsomal)</td>
<td>Y113H</td>
<td></td>
<td></td>
<td>(64)</td>
</tr>
<tr>
<td>GCP2</td>
<td>Glutamate carboxypeptidase</td>
<td>H475Y</td>
<td></td>
<td></td>
<td>(65)</td>
</tr>
<tr>
<td>GSTM1</td>
<td>Glutathione S-transferase mu1</td>
<td>Null alleles</td>
<td></td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>GSTOP1</td>
<td>Glutathione S-transferase pi1</td>
<td>I101V</td>
<td></td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>GSTT1</td>
<td>Glutathione S-transferase theta1</td>
<td>Null alleles</td>
<td></td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>HRAS1</td>
<td>Harvey rat sarcoma virus 1</td>
<td>VNTR rare alleles</td>
<td>***</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>HFE</td>
<td>Haemochromatosis</td>
<td>H63D, C282Y</td>
<td>**</td>
<td>May depend on iron intake</td>
<td>(73)</td>
</tr>
<tr>
<td>IGF1</td>
<td>Insulin-like growth factor 1</td>
<td>C repeat</td>
<td>**</td>
<td></td>
<td>(74,75)</td>
</tr>
<tr>
<td>IGFBP3</td>
<td>IGF binding protein</td>
<td>–202 A &gt; C</td>
<td>*</td>
<td></td>
<td>(76)</td>
</tr>
<tr>
<td>IL6</td>
<td>Interleukin 6</td>
<td>–174 G &gt; C</td>
<td>**</td>
<td></td>
<td>(76)</td>
</tr>
<tr>
<td>IL8</td>
<td>Interleukin 8</td>
<td>–251 T &gt; A</td>
<td>*</td>
<td></td>
<td>(76)</td>
</tr>
<tr>
<td>IRS1</td>
<td>Insulin receptor substrate 1</td>
<td>G972R</td>
<td>**</td>
<td></td>
<td>(74,75)</td>
</tr>
<tr>
<td>IRS2</td>
<td>Insulin receptor substrate 2</td>
<td>G1057D</td>
<td>*</td>
<td></td>
<td>(74,75)</td>
</tr>
<tr>
<td>MLH1</td>
<td>MutL homologue (MMR)</td>
<td>D132H</td>
<td>**</td>
<td></td>
<td>(77)</td>
</tr>
<tr>
<td>MLH3</td>
<td>MutL homologue (MMR)</td>
<td>P844L</td>
<td>*</td>
<td></td>
<td>(78)</td>
</tr>
<tr>
<td>MMP1</td>
<td>Matrix metalloproteinase 1</td>
<td>2G</td>
<td>**</td>
<td></td>
<td>(79)</td>
</tr>
<tr>
<td>MMP3</td>
<td>Matrix metalloproteinase 3</td>
<td>6A</td>
<td>**</td>
<td></td>
<td>(79)</td>
</tr>
<tr>
<td>MSN2</td>
<td>MutS homologue 2 (MMR)</td>
<td>nt 2006 C &gt; T</td>
<td>*</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>MTHFD1</td>
<td>MTHF dehydrogenase</td>
<td>R653Q</td>
<td>*</td>
<td></td>
<td>(71)</td>
</tr>
<tr>
<td>MTHFR</td>
<td>MTHF reductase</td>
<td>C677T</td>
<td>***</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>MTR</td>
<td>Methionine synthase</td>
<td>A1298C</td>
<td></td>
<td>May be in LD with C677T</td>
<td>(63,64)</td>
</tr>
<tr>
<td>MTRR</td>
<td>Methionine synthase reductase</td>
<td>A66G</td>
<td>**</td>
<td></td>
<td>(80)</td>
</tr>
<tr>
<td>NAT1</td>
<td>N-Acetyltransferase 1</td>
<td>Multiple alleles, e.g. *10</td>
<td>*</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>NAT2</td>
<td>N-Acetyltransferase 2</td>
<td>Rapid acetylator alleles</td>
<td>***</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>PAI1</td>
<td>Plasminogen activator inhibitor 1</td>
<td>Ins G promoter</td>
<td>*</td>
<td></td>
<td>(82)</td>
</tr>
<tr>
<td>PLAT2A</td>
<td>Secretory phospholipase A2</td>
<td>nt 964 C &gt; G</td>
<td>*</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>PPARG</td>
<td>Peroxisome proliferator activated receptor</td>
<td>P10A</td>
<td>**</td>
<td></td>
<td>(76)</td>
</tr>
<tr>
<td>SHMT</td>
<td>Serine hydroxymethyltransferase</td>
<td>L474F</td>
<td>*</td>
<td></td>
<td>(71)</td>
</tr>
<tr>
<td>TGBF1</td>
<td>Transforming growth factor beta</td>
<td>L10P</td>
<td>*</td>
<td></td>
<td>(83)</td>
</tr>
<tr>
<td>TGFBR1</td>
<td>TGF beta receptor 1</td>
<td>del(Ala1)</td>
<td>**</td>
<td></td>
<td>(84)</td>
</tr>
<tr>
<td>TNFA</td>
<td>Tumour necrosis factor alpha</td>
<td>–308 G &gt; A</td>
<td>**</td>
<td></td>
<td>(76)</td>
</tr>
<tr>
<td>TNFB</td>
<td>Tumour necrosis factor beta</td>
<td>–238 G &gt; A</td>
<td>*</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>TP53</td>
<td>p53</td>
<td>R72P</td>
<td>*</td>
<td></td>
<td>(63,64)</td>
</tr>
<tr>
<td>TS</td>
<td>Thymidylate synthase</td>
<td>2R/3R promoter</td>
<td>**</td>
<td>May depend on folate intake</td>
<td>(85,86)</td>
</tr>
<tr>
<td>VDR</td>
<td>Vitamin D receptor</td>
<td>M1T</td>
<td>*</td>
<td></td>
<td>(87)</td>
</tr>
</tbody>
</table>

*No evidence of association; **some reports of association, but early data or unconfirmed evidence; ***good evidence of association. These classifications represent the authors’ personal opinions only. We apologize to some authors for citing reviews or meta-analyses rather than the original studies for reasons of space.
as important as its identification in the first place, and this will require multiple studies in individual populations which may differ in genetic background, environment and availability of screening and medical care. Quantitation of risks associated with variants at multiple predisposition loci will be even more difficult, especially if risks do not combine in a simple additive or multiplicative manner, yet if testing for low-penetrance alleles is to come into routine clinical practice, such risks must be determined.

Without any good prior evidence of the nature of the remaining genetic influence on CRC, the possibility remains that it is a truly polygenic trait in which risks associated with any single polymorphism are so low that identification of the genes involved (and their testing in clinical practice) will be very difficult. This is, perhaps, the worst case scenario. An alternative is that multiple, rare variants may contribute to the increased familial risk of CRC: in this case, identification of the variants will be possible through mutation screening methods, although evaluation of their functional importance will be time-consuming. The case of the sub-polymorphic variant APC E1317Q provides a case in point (93–97); it currently seems likely that this variant, which was discovered by mutation screening, has little functional effect, although ‘closing the door’ on such a variant is not easy. Nevertheless, the modestly increased risk of breast cancer associated with the CHEK2 del1100C sub-polymorphic variant (98) shows that important alleles of this type do exist. (Any association between CHEK2 del1100C and CRC remains controversial.) Should many of the remaining CRC alleles prove to be of the low-frequency/low-risk type, genetic testing in the clinical setting may not be worthwhile. Despite these potential problems, however, the effectiveness of preventive measures for CRC, especially in high-risk individuals, means that the search for new predisposition genes is justified.

**FUTURE STRATEGIES FOR IDENTIFYING CRC PREDISPOSITION GENES**

The successful detection of predisposition genes for cancer has largely been limited to high-penetrance genes such as APC, BRCA1/2 and CDKN2A. Although the search for low-penetrance genes has been informative in some other complex diseases, for example, APOE-4 in Alzheimer’s disease, those underlying cancer remain largely elusive. The characteristics of cancer create obstacles. The usual late onset of CRC means parents may be deceased and the disease is often fatal, making sample collection difficult. In addition, the requirement of random somatic mutations for the progression to carcinoma may cloud the picture. However, some of these problems can be circumvented by the selection of early-onset cases to increase the genetic component and to reduce the environmental component, and by the study of benign, precursor lesions.

The question therefore arises of which methods to use to detect novel cancer genes. In our view, given that there are undoubtedly still high and moderate penetrance CRC genes yet to be discovered, linkage analysis should certainly not be abandoned. The latter would require the use of non-parametric methodology in view of genetic heterogeneity and the existence of phenocopies. However, even for moderate penetrance genes, the numbers of multi-case families required to yield results are large. For example, under a dominant model with a gene frequency of 0.1 and a risk ratio of 6, about 500 affected sib pairs or 20 affected sib trios would be needed to give a LOD score of 3.3 (99).

The increasing attention being given to association studies to identify CRC genes is therefore justified. For CRC, these have previously been limited to examining the frequency of alleles at candidate loci in cases and controls. However, the advent of SNP maps has created the potential for genome-wide association (100). The power of these studies, especially for rare alleles can be increased by using cases with a family history of CRC (101) and by selecting in other ways for ‘genetic’ disease (for example, multiple tumours, cancers of more than one site or early age of presentation). The detection of a dominant gene with a frequency of 0.1 and a relative risk of 4 would require 200 cases and 200 controls in an association study, but over 1000 affected sib pairs in a traditional linkage analysis (99).

A more controversial area in the optimal design of association studies revolves around the question of how many SNPs to use and whether the approach should be haplotype-map or sequence based. Public and private SNP discovery projects report the identification of 1.4 and 2.1 million SNPs, respectively, although the total is estimated (102) to be about 15 million in reality. Obviously only a small proportion of these could currently be genotyped in any association study. A map-based approach depends on linkage disequilibrium occurring between blocks of haplotypes within which only a few common haplotypes are observed and there is little evidence of historical recombination. This allows the genome to be represented by a carefully chosen group of haplotype tag SNPs (htSNPs) (103). The number of SNPs required will increase with the genetic diversity of a population because of shorter haplotype blocks. For example, the International HapMap Project has recently decided to add 2.25 million SNPs to the original 600 000 for each of three population maps, owing to the longer evolutionary history of the Yuroban (Nigerian) population (88). This approach will mainly detect alleles with a frequency of ≥5%. Advocates of a sequence-based approach argue that SNPs should be chosen on the basis of their probable contribution to the phenotype. The focus of this approach is therefore on coding regions and SNPs resulting in amino acid changes and/or SNPs found in evolutionarily conserved areas. This would require far fewer SNPs (50 000–100 000) for a genome-wide screen and would potentially detect lower frequency alleles (1–20%). Botstein and Risch (102) have suggested that, at least initially, a ‘sequence-based’ approach on a selected 5% of the genome might provide the power to detect some of the CRC-associated variants.

For identifying sub-polymorphic variants associated with disease, a mutation screening strategy based on candidate genes (and followed if possible by a suitable association study) is probably the only practical method for the short term. It remains to be seen whether or not new technologies such as re-sequencing microarrays prove cost-effective for research and diagnosis here. The requirement for functional studies will be particularly important for sub-polymorphic variants, especially those which lead to conserved amino acid changes, or affect splice sites, promoters or control regions.

Although it is hoped that linkage, association and mutation screening studies will yield answers, the concurrent use of
more indirect methods is prudent and may contribute to the fine mapping of an identified region of interest. As with SMAD4 and more recently, MYH, the search for somatic mutations may also identify the site of germline mutations, as their effects in the cell are functionally similar. Other approaches include the study of cancer-associated phenotypes including normal variation (for example in hormone levels). Inflammatory bowel disease, which we have not discussed in detail earlier, is associated with an increase risk of CRC yet has a higher sibling relative risk (104), and some of the predisposition genes might be the same for the two conditions. The search for modifier genes is also potentially important, as a gene which influences the severity of a Mendelian syndrome should also, in principle, act as a low-penetrance predisposition gene.

Intensive, effective population-based screening for CRC and its precursor lesions is likely in the foreseeable future to be expensive and/or to have a relatively low take-up rate. Even if this screening for CRC becomes more acceptable, it will optimally be targeted to those at highest risk. The identification of CRC risk genes will permit targeted screening, and will open up possibilities for other preventive measures and, perhaps, therapies. It is highly likely that at least some CRC genes can be identified and will provide useful targets for routine genetic testing and preventive intervention.

ACKNOWLEDGEMENTS

We are grateful to colleagues, particularly Richard Houlston and Walter Bodmer, for discussion and help.

REFERENCES


