Abstract. High-penetrance mutations in a small group of genes have been identified as the causal agent of colorectal cancer (CRC) in high-risk families. Our understanding of the sporadic cases is, however, much more limited and only in the past two years have multicentric genome-wide association studies (GWAS) started to unravel the complex genetic architecture behind this common form. To date, ten loci have been associated with an increased risk of CRC. Environmental factors play a role as well as other genetic factors yet to be discovered. The search for common variants with a low penetrance has come to an end, at least in the European population, and the focus now moves to less common variants (with higher penetrance) and to unclassified variants of unknown significance. As yet, less than 10% of the 35% genetic contribution to CRC is known.

Colorectal cancer (CRC) is a common disease, increasing in the Western world and ranking second among the most common causes of cancer-related deaths. We can distinguish between familial and sporadic forms, the latter representing the majority. CRC has been postulated to start as a premalignant lesion which in association with several molecular events, gradually develops into cancer (1). Several well-known high-risk CRC syndromes exist and while the predisposing genes are known, they account for fewer than 5% of the cases, even though studies of twins indicate that 35% of cases have a genetic component (2). In families with a history of CRC, genetic counselling and pre-symptomatic testing are justified by an increased likelihood such individuals developing the disease. In this way, the premalignant lesions, or polyps, can be detected and removed (3, 4). Thus, early detection is of major importance in reducing the morbidity and mortality in CRC. Several hereditary forms of CRC are known [for a review see: (5)]. In this review we will give an overview of the already known familial syndromes, as well as explaining CRC as a complex disease.

Familial Adenomatous Polyposis (FAP)

Incidence. Adenomatous polyposis coli (APC) is a tumour suppressor gene and individuals carrying a mutation in APC will develop familial adenomatous polyposis (FAP). The incidence of FAP is usually reported as 1:10,000 (6) to 1:7,000 (7) and accounts for fewer than 1% of all CRC cases (5).

Individuals with FAP usually develop hundreds to thousands of adenomatous polyps in the colon and rectum in late childhood and adolescence that, if left untreated, will eventually develop into cancer. The lifetime risk of developing CRC is 100% before the early forties, and by the age of 35 years, 95% of individuals with FAP will have polyps. The mean age at colon cancer diagnosis in untreated individuals is 39 years (range 34-43 years). The penetrance of other intestinal and extraintestinal manifestations is less understood and may depend in part on where the mutation is located in the APC gene.

FAP variants. Mutations in different regions of the APC gene can lead to different phenotypes and they are usually referred to as APC-associated polyposis conditions. They include: i) Gardner syndrome (GS), characterized by the association of colonic adenomatous polyposis of classical FAP with osteomas and soft tissue tumors (epidermoid cysts, fibromas, desmoid tumors) (8). These extraintestinal growths are benign and can occur in up to 20% of individuals and families with FAP. ii) Turcot syndrome (TS) is the association between colonic polyposis (or CRC) and central nervous system (CNS) tumors. CNS tumors in TS patients with a mutation in the APC gene are typically medulloblastoma. iii) Attenuated FAP (AFAP) is
characterized by fewer polyps (average of 30) and a later clinical presentation. The exact lifetime risk of CRC in AFAP is unclear at present; the cumulative risk by age 80 is approximately 70% (9). The average age of onset is 50-55 years, which is 10-15 years earlier than in those with classic FAP, but still earlier than for sporadic cases (10, 11).

Hereditary Non-polyposis CRC

Incidence. Hereditary non-polyposis CRC (HNPCC) is caused by germline mutations in DNA mismatch-repair (MMR) genes. HNPCC is the most common form of hereditary CRC, accounting for 1-3% of all CRC cases (12). It is also called Lynch syndrome after the oncologist Henry Lynch who pioneered the study of this disease (13). HNPCC is an autosomal dominant disorder characterized not only by early onset CRC and microsatellite instability (MSI) but also by cancer at other sites such as the endometrium (20-60% lifetime risk and the second most common cancer in HNPCC patients), ovary, stomach, hepatobiliary tract, upper urinary tract, brain and skin. HNPCC patients present few polyps that rapidly (within 1-2 years) develop into cancer. Two thirds of the tumours are located in the proximal colon and tend to be poorly differentiated (14-16). The cumulative cancer risk in HNPCC is about 70% and previous studies indicated an average age of onset of 44 years. However, more recent population-based data have suggested a later age at diagnosis, 61 years of age (17). The lifetime risk of developing HNPCC has been reported to be sex dependent, being 69% for men and 52% for women (17).

The MMR pathway includes several proteins that act in concert to identify and remove single nucleotide mismatches or insertion and deletion loops. At least five different proteins take part in the process, four of which have been implicated in HNPCC: MLH1, MSH2, MSH6 and PMS2 (18). The role in the mutations induced by reactive oxygen species (ROS) that are generated during aerobic metabolism (22). 8-Oxo-guanine is the most stable product of oxidative DNA damage and can easily mispair with adenine residues during DNA replication leading to G:C→T:A transversion (23, 24). MYH functions to excise the mispaired adenine and allow the repair of the DNA strand (25, 26). Al-Tassan and collaborators were the first to observe a higher frequency of somatic G:C→T:A transversion in the APC gene in tumours from 3 siblings affected with multiple colorectal adenomas and cancer with homozygous mutations in MYH (27).

Differential Diagnosis

Peutz-Jeghers syndrome. Peutz-Jeghers syndrome is inherited as an autosomal dominant tract and is characterized by the association of gastrointestinal polyps and mucocutaneous pigmentation. Individuals with this syndrome are at increased risk for intestinal as well as extraintestinal tumours, such as colorectal, gastric, breast, gynecological, lung and pancreatic cancer (28). Molecular genetics analyses often reveal alterations in STK11.

Phosphate and tensin homologue (PTEN) hamartoma tumor syndrome. PTEN hamartoma tumor syndrome includes: Cowden syndrome and Bannayan-Riley-Ruvalcaba syndrome as well as the less characterized Proteus syndrome and Proteus-like syndrome. Approximately 80% of patients who meet the criteria for Cowden syndrome and 60% of patients with Bannayan-Riley-Ruvalcaba have a mutation in the PTEN gene.

Juvenile polyposis syndrome. Juvenile polyposis syndrome is characterized by predisposition for hamartomatous polyps in the gastrointestinal tract, especially in the stomach, small intestine and rectum. Their incidence in affected families range from 9% to 50%. Three genes are known to be associated with the disease: SMAD4, BMPRIA and ENG and the mode of inheritance is autosomal dominant.

Hereditary mixed polyposis syndrome. Hereditary mixed polyposis syndrome is associated with an increased risk of CRC tumors (juvenile polyps, adenomatous polyps, hyperplastic polyps and carcinomas). Recently, a locus associated with this syndrome has been mapped to 15q13.3-q14 in individuals of Ashkenazi Jewish descent (29).

The Importance of Common Variants in MMR Genes

Several studies have focused on MMR genes, where many variants of unknown significance, called unclassified variants, have been detected. While obvious pathogenic
mutations (such as frameshift mutations which lead to a premature stop codon) are responsible for the development of CRC, the consequence of unclassified variants is less clear. Being able to determine the risk associated with each variant could mean more efficient preventative and surveillance programs.

In a recent study, Koessler et al. reported the first comprehensive candidate gene study that systematically tagged all the known common variants in the MMR genes and tested these tags for association with CRC susceptibility (30). A total of 2,299 cases and 2,284 unrelated controls were genotyped for 68 tagging single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) higher than 5% in 7 genes of the MMR pathway (MLH1, MLH3, MSH2, MSH3, MSH6, PMS1, PMS2). The most plausible association was found for the variant A1045T in MSH3, also notably identified in another recent case cohort study (31).

Even though the study was well powered, variants with a MAF lower than 5% (30) were not included and even common variants could have been missed due to sub-maximal linkage disequilibrium (LD) with tagging SNPs and the limited sample size of the cohorts. This indicates that MMR genes can confer a high risk of CRC in the presence of a pathogenic mutation (one which alters the reading frame of the protein leading to a premature stop codon, for example) as well as low-risk variants when the change is due to a SNP of unknown significance.

**Linkage Analysis In The Future**

Linkage analysis has been the tool of choice for finding genes for monogenic Mendelian diseases such as APC, MLH1 and MSH2 (32-34), but it requires the use of large families and a clearly defined phenotype. Besides its limited resolution, linkage analysis also has a low power in detecting weak effects and a high sensitivity to locus heterogeneity, two major issues in complex traits. While the resolution problem has been overcome by using specific linkage chips (Illumina 6K arrays and Affymetrix 10K arrays), locus heterogeneity remains a problem. One possible solution is to perform studies in specific kindreds where the phenotype has been accurately described (35) or to subgroup the families according to the degree they are affected.

Some genome-wide linkage studies have given direct evidence for the existence of further high- or moderate-penetrant CRC loci, even though the causal mutations have not been identified yet (36-40).

**Unknown High-risk and Low-risk Syndromes**

**Searching for high- or moderate-risk genes.** Families segregating known high-risk genes (APC and MMR genes) account for fewer than 5% of the CRC cases. There are families showing an aggregation of CRC cases for which the molecular component has not yet been identified. A Swedish study estimated the frequency of non-FAP non-HNPCC families having three or more first-degree relatives with CRC in at least two generations at 1.9% (Figure 1) (41). These families show a dominant pattern of inheritance and affected members could segregate a predisposing mutation in high-risk genes. The life-time risk of CRC is similar to that observed in HNPCC families, but with a later age of onset (4).

Another 8.3% of CRC cases come from families with two affected first- or second-degree relatives, where patients have a lower risk (10-20%) and are characterized by a higher incidence of colorectal adenomas compared to the high-risk families previously described (Figure 1) (41). In this group, several mildly or moderately penetrant alleles could explain the familial aggregation, but environmental factors cannot be excluded (3, 4, 41).

**Searching for low-risk variants.** In the past two years, the search for new susceptibility loci has been boosted by large genome-wide association studies (GWAS) performed with SNP chips. The first of them led to the identification of a susceptibility region on 8q24 (Table I) (42), where the most significant SNP was rs10505477 (tagged by rs6983267), which gave an odds ratio of 1.19 (p=6.40×10⁻⁹). It is worth noting that the region on 8q24 is also a known risk locus for prostate cancer, as found by Haiman and collaborators (43). The SNP rs10505477 is located within an uncharacterized gene called DQ515897 with multiple alternatively spliced products and close to a processed pseudogene of the transcription factor POU5F1 (which encodes the transcription factor OCT4) called POU5F1P1 and whose expression has been shown to occur in some tumour types. More interestingly, it is just 340,873 bp telomeric to the oncogene MYC, which is known for its role in colon cancer biology. Immunohistochemical analysis did not show a significant relation between the risk genotype and MYC expression (42). At the same time, a group in the United Kingdom found the same locus to be associated with an increased risk for CRC (44). They identified rs6983267 as the SNP with the strongest association, with an odds ratio (OR) of 1.27 (95% Cl=1.16-1.39; p=8.80×10⁻⁹) and 1.47 (95% Cl=1.34-1.62; p=4.44×10⁻¹⁵) for heterozygous and rare homozygous individuals, respectively. Moreover, a separate analysis in additional cancer-free adenoma cases provided evidence that rs6983267 is associated with an elevated risk of adenoma development, with an odds ratio of 1.21 (95% Cl=1.10-1.34; p=6.9×10⁻⁵). Tomlinson et al. found SNP rs10505477 to be strongly correlated with rs6983267 (r²=0.92) and with risk of colorectal neoplasia, with a p-value of 3.73×10⁻¹⁴ after replication (44). As described above, two transcripts lie in the vicinity of these two SNPs: DQ515897 and POU5F1P1. Tomlinson and collaborators failed to detect a concordance between the expression of the two genes in 49
CRC cell lines and the genotype at this locus (44). A recent study investigated the nature of the putative oncogenic change in the rs6983267 region and showed that the SNP affects the last nucleotide in a 9 bp-long transcription factor-binding site for transcription factor 4 (TCF4) (45). The possible existence of additional genetic variants around rs6983267 cannot be excluded but the results at least in part explain the increased risk for the carriers (45).

In a follow-up study, Tomlinson and collaborators further investigated 42,708 SNPs and, after three stages of replication, they identified two SNPs, rs16892766 (8q23.3) and rs10795668 (10p14) (Table I) with combined significance levels in replication phases of \( p = 8.7 \times 10^{-18} \) and \( p = 1.3 \times 10^{-13} \), respectively (46).

SNP rs16892766 lies in a 220-kb LD block that includes EIF3H (a gene which regulates cell growth and viability) and UTP23 but it is not associated with rs6983267 on 8q24, located 10 Mb more telomeric (46).

Elucidating the role of rs10795668 is even more difficult, since the SNP is located in an 82-kb LD block but there are no proven protein-coding transcripts in the region and the nearest genes are more than 0.4 Mb away from this SNP (0.4 Mb for the proximally located BC031880, 0.7 Mb for the distal LOC389936).

While for rs16892766 the homozygous state was associated with an increased risk of CRC (ORhet=1.27, 95% CI=1.20-1.34 and ORhom=1.43, 95% CI=1.13-1.82), the homozygous state for rs10795668 was associated with a reduced risk of CRC (ORhet=0.87, 95% CI=0.83-0.91 and ORhom=0.80, 95% CI=0.74-0.86), both in a dose-dependent manner. Furthermore, it was shown that the susceptibility allele of rs10795668 was more prevalent in rectal tumours, while the effect of rs16892766 was significantly stronger in younger patients (46).

In another GWAS study, Broderick and collaborators were able to replicate the association of CRC with rs6983267 on 8q24.21 and to find a new risk locus on 18q21.1 (Table I) (47). The significant SNP, rs4939827, maps to SMAD7, an intercellular antagonist of transforming growth factor (TGF) β signalling, which binds to the receptor complex and is able to block the activation of downstream signalling events (48). Furthermore, two other SNPs in SMAD7, rs12953717 and rs4464148, were among those with the most significant \( p \)-values (47).

Computational analysis for rs4939827, rs12953717 and rs4464148 showed four common haplotypes, of which three were associated with risk. The most significant was TTC \( (p=5.6 \times 10^{-13}) \) present in 29.3% of chromosomes from...
affected individuals and 26.3% of controls, while the second most significant was TTT ($p=9.9 \times 10^{-6}$) (47).

Loss of 18q is very common among individuals with CRC (49) and, for this reason, Broderick and co-workers focused on SMAD7 expression in lymphoblastoid cell lines. They detected a significant association between a lower mRNA expression and risk alleles at rs12953717 and rs4464148; their conclusion was that since even subtle changes in SMAD7 expression can affect β-catenin levels, these variants can lead to increased Wnt signalling and thus to carcinogenesis (49). More recently, an extensive resequencing of the genomic region on 18q21 surrounding these three SNPs was performed (50). One SNP, with a MAF of 0.47 and unlisted by dbSNP (therefore called Novel 1) was the strongest associated with CRC ($p=5.2 \times 10^{-6}$); as expected, Novel 1 is strongly correlated with the three previously reported SNPs, with $r^2>0.94$. A functional analysis of Novel 1 in a Xenopus model system was able to prove that the G allele in Novel 1 was associated with a reduced expression of SMAD7 in the colorectum (50) and therefore is the causative SNP.

In another phased-design GWAS, Tenesa et al. replicated and refined the findings on 8q24 and 18q21 and identified a new risk locus on 11q23 (Table I) (51). The SNP on 11q23 associated with an increased risk of CRC was rs4779584, located close to a gene called GREM1 encoding a secreted bone morphogenetic protein (BMP) antagonist; this is not a pathogenic variant but it could play a role in CRC tumorigenesis since GREM1 acts in the TGF-β/BMP pathway. The other significant SNP, rs4779584, lies between GREM1 and SCG5. Thus, it is possible that SCG5 could also play a role in the tumorigenic process, even though it is a worse candidate than GREM1. It may, for example, influence cell proliferation in the large bowel, given its role in neuroendocrine signalling (29).

SNP rs10318 is located in the 3’-UTR of GREM1, which encodes a secreted bone morphogenetic protein (BMP) antagonist; this is not a pathogenic variant but it could play a role in CRC tumorigenesis since GREM1 acts in the TGF-β/BMP pathway. The other significant SNP, rs4779584, lies between GREM1 and SCG5. Thus, it is possible that SCG5 could also play a role in the tumorigenic process, even though it is a worse candidate than GREM1. It may, for example, influence cell proliferation in the large bowel, given its role in neuroendocrine signalling (29).

Since aberrations in the TGF-β pathway are strongly involved in CRC carcinogenesis, Valle et al. hypothesized that TGFBR1 (TGF-β type I receptor) could be an outstanding candidate gene which, when mutated, could cause predisposition to CRC or act as a modifier for other genes (54). They hypothesized that the putative change might be subtle, leading to a lower expression of the gene rather
than its complete silencing. Three SNPs (and another in the second set of samples) in the 3'-UTR of TGFBR1 were chosen and genotyped in 242 patients. The allele-specific expression (ASE) was tested in 96 individuals found to be heterozygous for the three SNPs and the p-value associated with the comparison between cases and controls was $7.65 \times 10^{-5}$ (54). To assess the effect of ASE on TGF-β signalling, cell lines from ASE and non-ASE healthy individuals were treated with TGF-β, which binds to TGFBR2 and lead to its dimerization with TGFBR1. They detected a difference in the level of phosphorylation in SMAD2 (pSMAD2) and SMAD3 (pSMAD3), important downstream effectors in the TGF-β signalling pathway, with lower levels in ASE patients (54). After trying to find the exact alteration responsible for CRC in the patient cohort, the authors concluded that these data are compatible with, but do not prove, the role of TGFBR1 in CRC; furthermore, they were unable to determine which mechanism causes ASE. Given the frequency of ASE in CRC patients and healthy controls, the population attributable risk would be 18.7% (frequency of 21% for cases and 3% for controls) with an OR of 9 (95% CI: 2.7-30.6).

A recently performed meta-analysis of GWAS identified four new CRC loci (Table I) (55). After replication, 23 SNPs associated with CRC risk at p-value<10^{-5} were identified, of which 14 map to regions previously identified through fast-tracking analyses. Of the remaining nine, seven reached strong levels of significance (i.e. p<5.0x10^{-7}) after the combined analysis, indicating four new predisposing loci for CRC. The strongest statistical evidence was provided for two SNPs in strong LD and mapping to a 38 kb region at 20p12.3: rs961253, with a combined OR of 1.12 (95% CI=1.08-1.16; p=2.0x10^{-10}) and rs355527, also with a combined OR of 1.12 (95% CI=1.08-1.17; p=2.1x10^{-10}). There are no genes or predicted protein-encoding transcripts in the vicinity of these SNPs; however, BMP2 maps 342 kb telomeric to the locus, and this may have some relevance due to a similar finding in the same study (see below). The second strongest evidence for an association was for rs4444235 on 14q22.2, with a combined OR of 1.11 (95% CI=1.08-1.15; p=8.1x10^{-10}). It maps 9.4 kb from the transcription start site of the gene encoding bone morphogenetic protein 4 preprotein (BMP4). Like BMP2, BMP4 belongs to the TGFβ family and therefore could play an important role in CRC (56). More specifically, BMP signalling inhibits intestinal stem cell self-renewal through suppression of Wntβ-catenin signalling (57). It has been estimated that the contribution to the familial risk of CRC for the ten loci identified so far would be less than 1% (55). In order to detect possible epistatic effects, Houlston and collaborators examined the pairwise interaction between the SNPs, without detecting any interacting effect. This suggests that each locus acts independently in CRC development (55). If an additive model is used, and considering that each individual could carry more than one risk variant, then the loci identified to date can collectively account for about 6% of the excess familial risk (Figure 1) (55).

A Few Common Predisposing Variants or Many Rare Variants?

The aim of genetic epidemiological studies is to determine the number and penetrance of alleles affecting disease risk, also known as the genetic architecture of a disease. This affects the strategy for identifying polymorphisms underlying disease susceptibility. The number and the distribution of disease alleles in the population depends on several factors, such as mutation rate, genetic selection, drift and population demography. Since several parameters have to be estimated, different assumptions give rise to different models and two main hypotheses have emerged: the common disease common variant and the common disease rare variant hypothesis (63-65).

Another factor affecting the genetic architecture of a complex disease is the number of genes determining the disease susceptibility. CRC is a multistage process involving several pathways and many genes important for cell-cycle control, apoptosis and angiogenesis (66). Therefore, it seems plausible to assume that a fair number of polymorphic variants rather than a few should affect cancer susceptibility. 

An analysis of ENCODE data (67) has shown that up to 60% of the SNPs have a MAF lower than 5% (68). SNPs with such a low frequency are poorly represented in the HapMap Database (http://hapmap.ncbi.nlm.nih.gov/) and in all the SNP platforms currently available, and this could be...
why the risk alleles found so far have a relatively high frequency in the population.

The risk associated with these variants is generally small (OR<2); had it been higher, the selection would have acted more strongly to remove them from the population. A summary of the OR recently found by GWAS has shown that most of the common variants have an OR of only between 1.2 and 1.5 (mean OR=1.36), while the rare variants have an OR of 2 or more (mean OR=3.74) (69).

Targeting rare SNPs in large case-control association studies could thus have more power to detect causal variants than targeting common SNPs. Not only should the non-synonymous SNPs (nsSNPs) be investigated, but so should the SNPs in the promoter and eventually also silent SNPs, which have been suggested to be pathogenic (70-73).

It is possible that the majority of common variants for CRC have already been found, at least in the European population. It will now be interesting to determine whether the genetic architecture is the same in the different human populations or, more likely, whether each population has its own pattern of susceptibility. Due to their low frequency and small contribution to the overall susceptibility for CRC, rare variants will not be detected even by very large GWAS. Instead, another approach should be taken, such as that used in searching for variants predisposing for colorectal adenomas (74, 75). Candidate genes, for example those already known to be involved in CRC development, are first sequenced in a carefully selected group of individuals. All the rare variants, provided that they are not obviously pathogenic, are then analysed in a control population. At this stage, it is also possible to perform biochemical and functional studies, as well as a bioinformatics analysis, to predict the possible effect of the nucleotide change.

A good candidate is an SNP that shows a statistically significant difference in frequency between cases and controls either if considered alone or (most often) in combination with other SNPs in the same or in closely related genes. By considering the cumulative frequency of mildly deleterious polymorphisms rather than their individual frequency, this method is the only one available if, as has been emerging in the past few years, there is a very heterogeneous spectrum of predisposing alleles (76). Fearnhead and collaborators (75) followed this approach when they analysed rare variants in the APC gene, other genes involved in WNT signalling (AXIN1 and CTNNB1), MLH1 and MSH2. Even if not significant when considered alone (probably due to a small sample size), there was a clear difference in the frequency of the rare variants between cases and controls, with an OR of 2.2. The lack of double variants (two SNPs in the same individual) is consistent with their low frequency and implies that these susceptibility factors could act independently in a non-additive way (75). The major disadvantage still limiting the widespread use of this approach, besides the analytical challenge of selecting the truly deleterious SNPs from much neutral genetic variation, has been the high cost of sequencing many genes in thousands of patients (77).

Understanding the disease aetiology implies finding the true variant responsible for the increased susceptibility and there is a distinct difference between common and rare variants. For rare variants, the effect is due to the variant itself because of the method used for their discovery (resequencing) (69). Based on the GWAS published to date, the picture is quite different for common variants. In this case, the variant is unlikely to be functionally relevant, but rather in close LD with the true predisposing factor. Unfortunately, given the low associated OR, it is very difficult, if possible at all, to establish which of the closely linked variants is functionally important. Furthermore, if the OR is small, as it is for most of the common variants, the penetrance can be very small, even if the contribution of a particular variant may be large in terms of population attributable risk (69). Thus, since the penetrance determines whether a carrier will develop the disease, rare variants are likely to be more interesting than common variants.

Another issue is that the detection of common variants requires large cohorts of samples in order to achieve statistical significance, and the lower the risk associated with them, the higher the background due to environmental factors and genetic heterogeneity (69). However, common and rare variants are not mutually exclusive but should be seen as distinctive contributing factors. One study has shown how common variants can act as modifiers of the effect of rare variants, modulating the severity of the phenotype. Felix et al. (78) investigated the relationship between individuals carrying the same nonsense mutation in hMLH1 and variations in GSTM1 and GSTT1, genes known to be involved in the detoxification of xenobiotics. They showed that males who were null for GSTT1 were almost three times more at risk of developing CRC at any age when compared with males who had both copies of the gene (age of onset in GSTT1 null individuals of 39 years vs. 54 years in GSTT1 non-null) (78). Moreover, 21% of females who developed CRC had a GSTT1 null mutation compared to 44% of males. This also highlights the importance of including sex-linked or sex-limited genes in future studies (78).

**Low-penetrance Genetic Predisposition**

A distinctive feature shared by common and rare variants is the lack of familial aggregation, due to a reduced penetrance. It has been shown that assuming a penetrance of 10% of the heterozygote for a disease susceptibility allele, only 1.4% of families with four offspring will include more than one affected child (69). It is thus impossible to identify these susceptibility factors by linkage, although is feasible by
association studies. Before the rise in GWAS using SNP chips, this approach was successful in finding some low-risk alleles (for a review see (5)).

A large meta-analysis of 50 published studies investigating the association between common alleles in 13 genes and CRC concluded that significant results were found only for three polymorphisms: APC I1307K, HRAS1 VNTR and MTHFR A677V, the latter being recently replicated (79, 80).

APCI1307K. Codon 1307 (ATA) in exon 15 of the APC gene encodes isoleucine. The APC I1307K mutation causes a transversion T→A, so that the new codon (AAA) encodes lysine instead. This has no detectable effect per se, but it is nonetheless believed to increase the risk for CRC. In fact, it creates an A8 tract instead of the normal A3TA4 and this is thought to increase the errors in the replication process through somatic single nucleotide insertions or deletions due to slippage of DNA polymerase. This mutation was first analysed in the Ashkenazi Jewish population and its frequency has been estimated in 6.1% among healthy Ashkenazim compared to 10.4% in Ashkenazim with CRC (81). Furthermore, when the individuals are stratified according to their family history of CRC, around 28% of all probands are carriers of the variant (81). Other studies have confirmed the association between APC I1307K and CRC (65, 82, 83) but only in the Ashkenazi Jewish, while the importance in other human populations seems to be negligible. A pooled analysis of all the published studies gave an OR of 1.58 (95% CI=1.21-2.07) for carriers of the APC I1307K allele (80).

HRAS1 VNTR. The proto-oncogene HRAS1 is a member of the RAS family and encodes a protein involved in mitogenic signal transduction and differentiation (84). Activating point mutations in HRAS1 have been found in tumour cells from bladder, lung, colon and melanoma.

The Harvey ras-1 variable number of tandem repeats (HRAS1-VNTR) is located 1 kb downstream of HRAS1 and is composed of 30-100 units of a 28 bp consensus sequence. Even though more than 30 alleles of 1,000-3,000 bp have been described (85), the four most common account for 94% of the variability (84). Rare alleles have been proposed as risk factors for different types of cancer (84), but the underlying mechanism is still unclear. Originally it was hypothesised that the association may be the result of LD with a (functional) unknown variant (84). Alternatively, these repeats could modulate the expression of nearby genes by interacting with transcriptional regulatory elements, such as the rel/nuclear factor kappa B (NF-κB) family of regulatory factors (80). Five studies examined the risk of CRC associated with rare alleles of the HRAS1 VNTR (84, 86-89); all of them reported an OR higher than 1 but only in two studies were the results statistically significant (87, 88). A pooled analysis of all the five studies gave an OR of 2.5 (95% CI: 1.54-4.05) for CRC (80).

MTHFR C677T. Global and gene-specific changes in DNA methylation pattern contribute to loss of proto-oncogene and tumour suppressor expression. In CRC, this occurs during progression from adenoma to carcinoma (90, 91). There is some evidence that DNA methylation depends on the availability of methyl group donors, such as folate (92). The C677T polymorphism in the MTHFR gene causes the replacement of an alanine with a valine at position 222, producing a protein with a lower activity. Homozygotes for the variant, Val/Val, have about 30% of the normal enzyme activity and lower level of methylenetetrahydrofolate (methyl-THF) (93). Decreased levels of methyl-THF may negatively affect DNA methylation contributing to carcinogenesis. Furthermore, depletion of methyl-THF impairs thymidylate biosynthesis causing deoxynucleotide pool imbalances, making the DNA prone to strand breaks. Several studies have investigated this association and found that there is an inverse relationship with CRC (94-98). Pooled meta-analysis gave an OR of 0.76 (95% CI=0.62-0.92) for the Val/Val genotype compared to Ala/Ala and Ala/Val genotypes (80). A recent study performed in 2,575 cases and 2,707 controls (with validation by kin-cohort of 14704 first-degree relatives) confirmed this association, reporting an OR of 0.82 (95% CI=0.75-0.91) (79).

CRC Prevention

Linkage studies will hopefully result in the identification of new high- or moderate risk predisposing genes. This new knowledge could easily be implemented into clinical praxis where already there is a format for genetic counselling, testing and prevention programs.

New knowledge of low-risk genes will be more difficult to use clinically since any relative risk associated with a single variant is expected to be quite low (if considered alone). However, association studies aiming at finding the incidence of these alleles in the general population still have clinical and practical importance. This will allow the development of more reliable risk models able to guide screening programs and preventative strategies. The goal is to set up more effective preventative population-based surveillance programs in order to significantly reduce the incidence of CRC.

Low-penetrance alleles account for the majority of the sporadic cases but even if of undisputed utility, their value in diagnostics and prevention is limited at the present. Modifying alleles could also play a role, influencing the penetrance and the expressivity, although the effect is difficult to measure in humans and it has so far been studied only in animal models. It is reasonable to expect more than
one modifier exists and that some may cancel each other out. For this information to be implemented in clinical practice all modifier genes have to be identified and their interactions clarified (99).

Acknowledgements

This work was supported by the Swedish Cancer Society, the Stockholm Cancer Foundation, the Nilsson-Ehle Foundation (grant IDs 23267 & 24506) and the Anders Otto Swärd/Ulrika Eklund Foundation. We are grateful to Professor Annika Lindblom for her valuable help with the manuscript.

References

7 Kinzler KW and Vogelstein B: Lessons from hereditary colorectal cancer. Cell 87: 159-170, 1996.


Picelli et al: Predisposing Colorectal Cancer Variants (Review)

Received September 17, 2009
Revised October 30, 2009
Accepted October 30, 2009