

Review

The Continuing Search For Predisposing Colorectal Cancer Variants

SIMONE PICELLI, SUSANNA VON HOLST and PETRA WESSENDORF

Department of Molecular Medicine and Surgery, Karolinska Institute, Stockholm, Sweden

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 forms. 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

pre-malignant 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

Correspondence to: Simone Picelli, Department of Molecular Medicine and Surgery L8:02, Karolinska Institute, Stockholm, Sweden. Tel: +46 0851773924, Fax: +46 0851773620, e-mail: simone.picelli@ki.se

Key Words: SNP, *APC*, linkage disequilibrium, risk, penetrance, polyposis, review.

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). However, the role of PMS2 is still under debate given that germline mutations are very rare and have been described only for a few individuals (19).

HNPCC variants. Some HNPCC syndromes have been described: i) Muir-Torre syndrome is defined by the combination of sebaceous neoplasms of the skin and other malignancies, commonly those seen in HNPCC (20, 21). ii) TS with mutations in the MMR genes comprises colorectal cancer commonly associated with glioblastoma.

Polyposis Associated with *mutY* Homolog

The colonic phenotype of *mutY*-homolog associated polyposis is similar to FAP but is inherited as an autosomal recessive trait. It is caused by bi-allelic germline mutations in the human homolog of the *Escherichia coli* base excision repair (BER) gene *mutY* (*MYH*). BER plays an important

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 and 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 ($p=6.40 \times 10^{-9}$). 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% CI=1.16-1.39; $p=8.80 \times 10^{-8}$) and 1.47 (95% CI=1.34-1.62; $p=4.44 \times 10^{-15}$) 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% CI=1.10-1.34; $p=6.9 \times 10^{-5}$). Tomlinson *et al.* found SNP rs10505477 to be strongly correlated with rs6983267 ($r^2=0.92$) 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

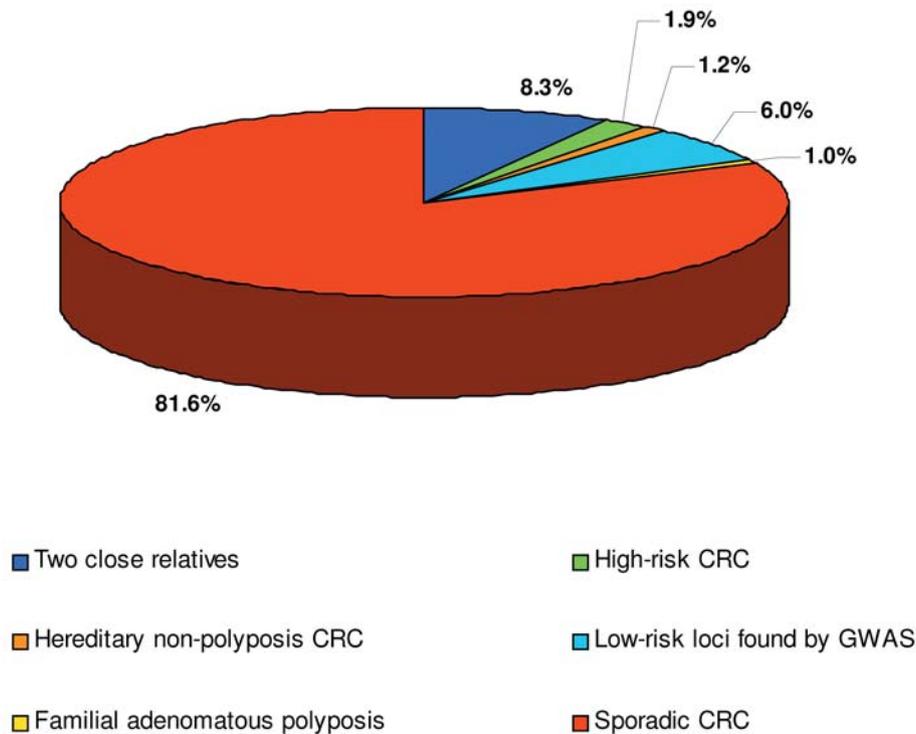


Figure 1. The frequencies of different types of CRC with respect to genetic background in Sweden (adapted and modified from (41)). GWAS, Genome-wide association studies.

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 ($OR_{het}=1.27$, 95% CI=1.20-1.34 and $OR_{hom}=1.43$, 95% CI=1.13-1.82), the homozygous state for rs10795668 was associated with a reduced risk of CRC ($OR_{het}=0.87$, 95% CI=0.83-0.91 and $OR_{hom}=0.80$, 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

Table I. CRC loci identified in the genome-wide association studies performed so far.

SNP ID	Locus	Gene	Reference
rs6983267	8q24.21	-	(42), (44), (47), (51)
rs16892766	8q23.3	-	(46)
rs10795668	10p14	-	(46)
rs4939827	18q21.1	<i>SMAD7</i>	(47), (51)
rs3802842	11q23.1	<i>FLJ45803 & LOC120376</i>	(51)
rs4779584 & rs10318	15q13.3	<i>GREM1</i>	(29)
rs961253 & rs355527	20p12.3	-	(55)
rs4444235	14q22.2	<i>BMP4</i>	(55)
rs10411210 & rs7259371	19q13.1	<i>RHPN2</i>	(55)
rs9929218	16q22.1	<i>CDH1</i>	(55)

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 rs3802842, located close to a gene called *POU2AF1* and encoding a POU transcription factor, similarly to rs7014346 on 8q24. Furthermore, substantial population-specific differences in risk were observed, with different allelic effects between the Scottish population used in the discovery phase and a Japanese population, one of the eight cohorts used in the replication phase. The risk of rectal cancer was greater than that for colon cancer both for rs3802842 (11q23) and rs4939827 (18q21), and with a different magnitude for Japanese and Caucasian populations, explaining the aforementioned difference in risk (51).

Hereditary mixed polyposis syndrome, which is characterized by the development of different colorectal

tumours, had been previously mapped to the locus *CRAC1* on chromosome 15q13.3-q14 in individuals of Ashkenazi Jewish descent (52, 53). In the attempt to refine the location of *CRAC1*, eight more affected Ashkenazi individuals from the linked families were included in the study and the region was investigated using the Illumina Hap550 SNP array (29). Despite these efforts, it was not possible to pinpoint a specific causative variant common to all the affected individuals, only to narrow down the region, containing then three known genes (*SCG5*, *GREM1* and *FMN1*) and three hypothetical genes (29). Jaeger and collaborators hypothesized that the risk locus might not only harbor high-penetrance mutations that cause CRC in Ashkenazi Jews, but also low-penetrance variants that increase the risk in the general UK population. Using CRC cases selected for family history and/or early onset, as well as CRC cases unselected for family history, they were able to find two SNPs associated with the disease, rs4779584 and rs10318 (Table I). The overall p_{allele} was 4.44×10^{-14} for rs4779584 (OR=1.26, 95% CI=1.19-1.34) and 7.93×10^{-9} for rs10318 (OR=1.19, 95% CI=1.12-1.26) after four stages of replication (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 *TGFBRI* (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 7.65×10^{-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.* $p < 5.0 \times 10^{-7}$) 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; $p = 2.0 \times 10^{-10}$) and rs355527, also with a combined OR of 1.12 (95% CI=1.08-1.17; $p = 2.1 \times 10^{-10}$). 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; $p = 8.1 \times 10^{-10}$). 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 TGF β family and therefore could play an important role in CRC (56). More specifically, BMP signalling inhibits intestinal stem cell self-renewal through suppression of Wnt β -catenin signalling (57). It is worth noting that inactivating mutations in the BMP receptor subunit *BMPR1A* are a known cause of the rare juvenile polyposis syndrome, associated with a very high risk of CRC (58, 59). Furthermore, as discussed earlier, SNPs close to the BMP antagonist *GREM1* are associated with an increased risk of CRC (29). The third locus was given by two SNPs in moderate LD ($r^2 = 0.54$), rs10411210 and

rs7259371 on 19q13.1, with combined OR of 0.83 (95% CI=0.78-0.88; $p = 1.1 \times 10^{-9}$) and 0.89 (95% CI=0.85-0.93; $p = 2.2 \times 10^{-7}$), respectively. They encompass the Rho GTPase binding protein 2 gene (*RHPN2*), encoding a Rho GTPase involved in the regulation of the actin cytoskeleton and cell motility (60). RhoA proteins have been implicated in several types of cancer, including CRC, due to their capacity for promoting invasiveness (61). The fourth locus was rs9929218 on 16q22.1, which maps to intron 1 of the gene encoding cadherin 1 (*CDH1*). The combined OR was 0.90 (95% CI=0.87-0.94; $p = 1.2 \times 10^{-7}$) and there is evidence of association also for rs1862748, an SNP in strong LD ($r^2 = 0.91$) with rs9929218. *CDH1* is a known risk factor since somatic inactivation leads to an increased activity of the β -catenin TCF transcription factor pathway (62).

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).

***APC* I1307K.** 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 Sward/Ulrika Eklund Foundation. We are grateful to Professor Annika Lindblom for her valuable help with the manuscript.

References

- 1 Fearon ER and Vogelstein B: A genetic model for colorectal tumorigenesis. *Cell* 61: 759-767, 1990.
- 2 Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A and Hemminki K: Environmental and heritable factors in the causation of cancer-analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med* 343: 78-85, 2000.
- 3 Liljegren A, Lindblom A, Rotstein S, Nilsson B, Rubio C and Jaramillo E: Prevalence and incidence of hyperplastic polyps and adenomas in familial colorectal cancer: correlation between the two types of colon polyps. *Gut* 52: 1140-1147, 2003.
- 4 Lindgren G, Liljegren A, Jaramillo E, Rubio C and Lindblom A: Adenoma prevalence and cancer risk in familial non-polyposis colorectal cancer. *Gut* 50: 228-234, 2002.
- 5 de la Chapelle A: Genetic predisposition to colorectal cancer. *Nat Rev Cancer* 4: 769-780, 2004.
- 6 Potter JD: Colorectal cancer: molecules and populations. *J Natl Cancer Inst* 91: 916-932, 1999.
- 7 Kinzler KW and Vogelstein B: Lessons from hereditary colorectal cancer. *Cell* 87: 159-170, 1996.
- 8 Gardner EJ and Richards RC: Multiple cutaneous and subcutaneous lesions occurring simultaneously with hereditary polyposis and osteomatosis. *Am J Hum Genet* 5: 139-147, 1953.
- 9 Neklason DW, Stevens J, Boucher KM, Kerber RA, Matsunami N, Barlow J, Mineau G, Leppert MF and Burt RW: American founder mutation for attenuated familial adenomatous polyposis. *Clin Gastroenterol Hepatol* 6: 46-52, 2008.
- 10 Spirio L, Olschwang S, Groden J, Robertson M, Samowitz W, Joslyn G, Gelbert L, Thliveris A, Carlson M, Otterud B, Lynch H, Watson P, Lynch P, Laurent-Puig p, Burt R, Hughes JP, Thomas G, Leppert M and White R: Alleles of the *APC* gene: an attenuated form of familial polyposis. *Cell* 75: 951-7, 1993.
- 11 Giardiello FM, Brensinger JD, Petersen GM, Luce MC, Hyland LM, Bacon JA, Booker SV, Parker RD and Hamilton SR: The use and interpretation of commercial *APC* gene testing for familial adenomatous polyposis. *N Engl J Med* 336: 823-827, 1997.
- 12 de la Chapelle A: The incidence of Lynch syndrome. *Fam Cancer* 4: 233-237, 2005.
- 13 Lynch HT and Krush AJ: Cancer family G revisited: 1895-1970. *Cancer* 27: 1505-1511, 1971.
- 14 Lynch HT, Smyrk TC, Watson P, Lanspa SJ, Lynch JF, Lynch PM, Cavalieri RJ and Boland CR: Genetics, natural history, tumor spectrum, and pathology of hereditary nonpolyposis colorectal cancer: an updated review. *Gastroenterology* 104: 1535-1549, 1993.
- 15 Lynch HT and Smyrk T: Hereditary nonpolyposis colorectal cancer (Lynch syndrome). An updated review. *Cancer* 78: 1149-1167, 1996.
- 16 Aarnio M, Sankila R, Pukkala E, Salovaara R, Aaltonen LA, de la Chapelle A, Peltomaki P, Mecklin JP and Jarvinen HJ: Cancer risk in mutation carriers of DNA mismatch-repair genes. *Int J Cancer* 81: 214-218, 1999.
- 17 Hampel H, Stephens JA, Pukkala E, Sankila R, Aaltonen LA, Mecklin JP and de la Chapelle A: Cancer risk in hereditary nonpolyposis colorectal cancer syndrome: later age of onset. *Gastroenterology* 129: 415-421, 2005.
- 18 Peltomaki P: Role of DNA mismatch repair defects in the pathogenesis of human cancer. *J Clin Oncol* 21: 1174-1179, 2003.
- 19 Hendriks YM, Jagmohan-Changur S, van der Klift HM, Morreau H, van Puijenbroek M, Tops C, van Os T, Wagner A, Ausems MG, Gomez E, Breuning MH, Brocker-Vriends AH, Vasen HF and Wijnen JT: Heterozygous mutations in *PMS2* cause hereditary nonpolyposis colorectal carcinoma (Lynch syndrome). *Gastroenterology* 130: 312-322, 2006.
- 20 Schwartz RA and Torre DP: The Muir-Torre syndrome: a 25-year retrospect. *J Am Acad Dermatol* 33: 90-104, 1995.
- 21 Misago N and Narisawa Y: Sebaceous neoplasms in Muir-Torre syndrome. *Am J Dermatopathol* 22: 155-161, 2000.
- 22 Sieber OM, Lipton L, Crabtree M, Heinemann K, Fidalgo P, Phillips RK, Bisgaard ML, Orntoft TF, Aaltonen LA, Hodgson SV, Thomas HJ and Tomlinson IP: Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in *MYH*. *N Engl J Med* 348: 791-799, 2003.
- 23 Cheng KC, Cahill DS, Kasai H, Nishimura S and Loeb LA: 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G→T and A→C substitutions. *J Biol Chem* 267: 166-172, 1992.
- 24 Shibutani S, Takeshita M and Grollman AP: Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. *Nature* 349: 431-434, 1991.
- 25 Slupska MM, Baikalov C, Luther WM, Chiang JH, Wei YF and Miller JH: Cloning and sequencing a human homolog (*hMYH*) of the *Escherichia coli mutY* gene whose function is required for the repair of oxidative DNA damage. *J Bacteriol* 178: 3885-3892, 1996.
- 26 Slupska MM, Luther WM, Chiang JH, Yang H and Miller JH: Functional expression of hMYH, a human homolog of the *Escherichia coli MutY* protein. *J Bacteriol* 181: 6210-6213, 1999.
- 27 Al-Tassan N, Chmiel NH, Maynard J, Fleming N, Livingston AL, Williams GT, Hodges AK, Davies DR, David SS, Sampson JR and Cheadle JP: Inherited variants of *MYH* associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* 30: 227-232, 2002.
- 28 Giardiello FM, Brensinger JD, Tersmette AC, Goodman SN, Petersen GM, Booker SV, Cruz-Correa M and Offerhaus JA: Very high risk of cancer in familial Peutz-Jeghers syndrome. *Gastroenterology* 119: 1447-1153, 2000.
- 29 Jaeger E, Webb E, Howarth K, Carvajal-Carmona L, Rowan A, Broderick P, Walther A, Spain S, Pittman A, Kemp Z, Sullivan K, Heinemann K, Lubbe S, Domingo E, Barclay E, Martin L, Gorman M, Chandler I, Vijaykrishnan J, Wood W, Papaemmanuil E, Penegar S, Qureshi M, Farrington S, Tenesa A, Cazier JB, Kerr D, Gray R, Peto J, Dunlop M, Campbell H, Thomas H, Houlston R and Tomlinson I: Common genetic variants at the *CRAC1* (HMPS) locus on chromosome 15q13.3 influence colorectal cancer risk. *Nat Genet* 40: 26-28, 2008.

- 30 Koessler T, Oestergaard MZ, Song H, Tyrer J, Perkins B, Dunning A, Easton D and Pharoah PP: Common variants in mismatch repair genes and risk of colorectal cancer. *Gut* 57: 1097-1101, 2008.
- 31 Berndt SI, Platz EA, Fallin MD, Thuita LW, Hoffman SC and Helzlsouer KJ: Mismatch repair polymorphisms and the risk of colorectal cancer. *Int J Cancer* 120: 1548-1554, 2007.
- 32 Lindblom A, Tannergard P, Werelius B and Nordenskjöld M: Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer. *Nat Genet* 5: 279-282, 1993.
- 33 Bodmer WF, Bailey CJ, Bodmer J, Bussey HJ, Ellis A, Gorman P, Lucibello FC, Murday VA, Rider SH, Scambler P, Solomon SE and Spurr NK: Localization of the gene for familial adenomatous polyposis on chromosome 5. *Nature* 328: 614-616, 1987.
- 34 Peltomäki P, Aaltonen LA, Sistonen P, Pylkkanen L, Mecklin JP, Jarvinen H, Green JS, Jass JR, Weber JL, Leach FS, Petersen GM, Hamilton SR, de la Chapelle A and Vogelstein B: Genetic mapping of a locus predisposing to human colorectal cancer. *Science* 260: 810-812, 1993.
- 35 Polychronakos C: Common and rare alleles as causes of complex phenotypes. *Curr Atheroscler Rep* 10: 194-200, 2008.
- 36 Djureinovic T, Skoglund J, Vandrovicova J, Zhou XL, Kalushkova A, Iselius L and Lindblom A: A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. *Gut* 55: 362-366, 2006.
- 37 Kemp Z, Carvajal-Carmona L, Spain S, Barclay E, Gorman M, Martin L, Jaeger E, Brooks N, Bishop DT, Thomas H, Tomlinson I, Papaemmanuil E, Webb E, Sellick GS, Wood W, Evans G, Lucassen A, Maher ER and Houlston RS: Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan. *Hum Mol Genet* 15: 2903-2910, 2006.
- 38 Kemp ZE, Carvajal-Carmona LG, Barclay E, Gorman M, Martin L, Wood W, Rowan A, Donohue C, Spain S, Jaeger E, Evans DG, Maher ER, Bishop T, Thomas H, Houlston R and Tomlinson I: Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom. *Cancer Res* 66: 5003-5006, 2006.
- 39 Picelli S, Vandrovicova J, Jones S, Djureinovic T, Skoglund J, Zhou XL, Velculescu VE, Vogelstein B and Lindblom A: Genome-wide linkage scan for colorectal cancer susceptibility genes supports linkage to chromosome 3q. *BMC Cancer* 8: 87, 2008.
- 40 Wiesner GL, Daley D, Lewis S, Ticknor C, Platzer P, Lutterbaugh J, MacMillen M, Baliner B, Willis J, Elston RC and Markowitz SD: A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2. *Proc Natl Acad Sci USA* 100: 12961-12965, 2003.
- 41 Olsson L and Lindblom A: Family history of colorectal cancer in a Sweden county. *Fam Cancer* 2: 87-93, 2003.
- 42 Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, Prendergast J, Olschwang S, Chiang T, Crowdy E, Ferretti V, Laflamme P, Sundararajan S, Roumy S, Olivier JF, Robidoux F, Sladek R, Montpetit A, Campbell P, Beziau S, O'Shea AM, Zogopoulos G, Cotterchio M, Newcomb P, McLaughlin J, Younghusband B, Green R, Green J, Porteous ME, Campbell H, Blanche H, Sahbatou M, Tubacher E, Bonaiti-Pellie C, Buecher B, Riboli E, Kury S, Chanock SJ, Potter J, Thomas G, Gallinger S, Hudson TJ and Dunlop MG: Genome-wide association scan identifies a colorectal cancer susceptibility locus on chromosome 8q24. *Nat Genet* 39: 989-994, 2007.
- 43 Haiman CA, Patterson N, Freedman ML, Myers SR, Pike MC, Waliszewska A, Neubauer J, Tandon A, Schirmer C, McDonald GJ, Greenway SC, Stram DO, Le Marchand L, Kolonel LN, Frasco M, Wong D, Pooler LC, Ardlie K, Oakley-Girvan I, Whittemore AS, Cooney KA, John EM, Ingles SA, Altshuler D, Henderson BE and Reich D: Multiple regions within 8q24 independently affect risk for prostate cancer. *Nat Genet* 39: 638-644, 2007.
- 44 Tomlinson I, Webb E, Carvajal-Carmona L, Broderick P, Kemp Z, Spain S, Penegar S, Chandler I, Gorman M, Wood W, Barclay E, Lubbe S, Martin L, Sellick G, Jaeger E, Hubner R, Wild R, Rowan A, Fielding S, Howarth K, Silver A, Atkin W, Muir K, Logan R, Kerr D, Johnstone E, Sieber O, Gray R, Thomas H, Peto J, Cazier JB and Houlston R: A genome-wide association scan of tag SNPs identifies a susceptibility variant for colorectal cancer at 8q24.21. *Nat Genet* 39: 984-988, 2007.
- 45 Tuupanen S, Turunen M, Lehtonen R, Hallikas O, Vanharanta S, Kivioja T, Bjorklund M, Wei G, Yan J, Niittymäki I, Mecklin JP, Jarvinen H, Ristimäki A, Di-Bernardo M, East P, Carvajal-Carmona L, Houlston RS, Tomlinson I, Palin K, Ukkonen E, Karhu A, Taipale J and Aaltonen LA: The common colorectal cancer predisposition SNP rs6983267 at chromosome 8q24 confers potential to enhanced Wnt signaling. *Nat Genet* 41: 868-869, 2009.
- 46 Tomlinson IP, Webb E, Carvajal-Carmona L, Broderick P, Howarth K, Pittman AM, Spain S, Lubbe S, Walther A, Sullivan K, Jaeger E, Fielding S, Rowan A, Vijayakrishnan J, Domingo E, Chandler I, Kemp Z, Qureshi M, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Penegar S, Barclay E, Wood W, Martin L, Gorman M, Thomas H, Peto J, Bishop DT, Gray R, Maher ER, Lucassen A, Kerr D, Evans DG, Schafmayer C, Buch S, Volzke H, Hampe J, Schreiber S, John U, Koessler T, Pharoah P, van Wezel T, Morreau H, Wijnen JT, Hopper JL, Southey MC, Giles GG, Severi G, Castellvi-Bel S, Ruiz-Ponte C, Carracedo A, Castells A, Forsti A, Hemminki K, Vodicka P, Naccarati A, Lipton L, Ho JW, Cheng KK, Sham PC, Luk J, Agundez JA, Ladero JM, de la Hoya M, Caldes T, Niittymäki I, Tuupanen S, Karhu A, Aaltonen L, Cazier JB, Campbell H, Dunlop MG and Houlston RS: A genome-wide association study identifies colorectal cancer susceptibility loci on chromosomes 10p14 and 8q23.3. *Nat Genet* 40: 623-630, 2008.
- 47 Broderick P, Carvajal-Carmona L, Pittman AM, Webb E, Howarth K, Rowan A, Lubbe S, Spain S, Sullivan K, Fielding S, Jaeger E, Vijayakrishnan J, Kemp Z, Gorman M, Chandler I, Papaemmanuil E, Penegar S, Wood W, Sellick G, Qureshi M, Teixeira A, Domingo E, Barclay E, Martin L, Sieber O, Kerr D, Gray R, Peto J, Cazier JB, Tomlinson I and Houlston RS: A genome-wide association study shows that common alleles of *SMAD7* influence colorectal cancer risk. *Nat Genet* 39: 1315-1317, 2007.
- 48 ten Dijke P and Hill CS: New insights into TGF-beta-Smad signalling. *Trends Biochem Sci* 29: 265-273, 2004.
- 49 Gaasenbeek M, Howarth K, Rowan AJ, Gorman PA, Jones A, Chaplin T, Liu Y, Bicknell D, Davison EJ, Fiegler H, Carter NP, Roylance RR and Tomlinson IP: Combined array-comparative genomic hybridization and single-nucleotide polymorphism-loss of heterozygosity analysis reveals complex changes and multiple forms of chromosomal instability in colorectal cancers. *Cancer Res* 66: 3471-3479, 2006.

- 50 Pittman AM, Naranjo S, Webb E, Broderick P, Lips EH, van Wezel T, Morreau H, Sullivan K, Fielding S, Twiss P, Vijayakrishnan J, Casares F, Qureshi M, Gomez-Skarmeta JL and Houlston RS: The colorectal cancer risk at 18q21 is caused by a novel variant altering *SMAD7* expression. *Genome Res* 19: 987-993, 2009.
- 51 Tenesa A, Farrington SM, Prendergast JG, Porteous ME, Walker M, Haq N, Barnetson RA, Theodoratou E, Cetnarskyj R, Cartwright N, Semple C, Clark AJ, Reid FJ, Smith LA, Kavoussanakis K, Koessler T, Pharoah PD, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Schmidt CO, Hampe J, Chang-Claude J, Hoffmeister M, Brenner H, Wilkening S, Canzian F, Capella G, Moreno V, Deary IJ, Starr JM, Tomlinson IP, Kemp Z, Howarth K, Carvajal-Carmona L, Webb E, Broderick P, Vijayakrishnan J, Houlston RS, Rennert G, Ballinger D, Rozek L, Gruber SB, Matsuda K, Kidokoro T, Nakamura Y, Zanke BW, Greenwood CM, Rangrej J, Kustra R, Montpetit A, Hudson TJ, Gallinger S, Campbell H and Dunlop MG: Genome-wide association scan identifies a colorectal cancer susceptibility locus on 11q23 and replicates risk loci at 8q24 and 18q21. *Nat Genet* 40: 631-637, 2008.
- 52 Jaeger EE, Woodford-Richens KL, Lockett M, Rowan AJ, Sawyer EJ, Heinimann K, Rozen P, Murday VA, Whitelaw SC, Ginsberg A, Atkin WS, Lynch HT, Southey MC, Debinski H, Eng C, Bodmer WF, Talbot IC, Hodgson SV, Thomas HJ and Tomlinson IP: An ancestral Ashkenazi haplotype at the HMP5/CRAC1 locus on 15q13-q14 is associated with hereditary mixed polyposis syndrome. *Am J Hum Genet* 72: 1261-1267, 2003.
- 53 Tomlinson I, Rahman N, Frayling I, Mangion J, Barfoot R, Hamoudi R, Seal S, Northover J, Thomas HJ, Neale K, Hodgson S, Talbot I, Houlston R and Stratton MR: Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22. *Gastroenterology* 116: 789-795, 1999.
- 54 Valle L, Serena-Acedo T, Liyanarachchi S, Hampel H, Comeras I, Li Z, Zeng Q, Zhang HT, Pennison MJ, Sadim M, Pasche B, Tanner SM and de la Chapelle A: Germline allele-specific expression of *TGFBR1* confers an increased risk of colorectal cancer. *Science* 321: 1361-1365, 2008.
- 55 Houlston RS, Webb E, Broderick P, Pittman AM, Di Bernardo MC, Lubbe S, Chandler I, Vijayakrishnan J, Sullivan K, Penegar S, Carvajal-Carmona L, Howarth K, Jaeger E, Spain SL, Walther A, Barclay E, Martin L, Gorman M, Domingo E, Teixeira AS, Kerr D, Cazier JB, Niittymaki I, Tuupanen S, Karhu A, Aaltonen LA, Tomlinson IP, Farrington SM, Tenesa A, Prendergast JG, Barnetson RA, Cetnarskyj R, Porteous ME, Pharoah PD, Koessler T, Hampe J, Buch S, Schafmayer C, Tepel J, Schreiber S, Volzke H, Chang-Claude J, Hoffmeister M, Brenner H, Zanke BW, Montpetit A, Hudson TJ, Gallinger S, Campbell H and Dunlop MG: Meta-analysis of genome-wide association data identifies four new susceptibility loci for colorectal cancer. *Nat Genet* 40: 1426-1435, 2008.
- 56 Kim JS, Crooks H, Dracheva T, Nishanian TG, Singh B, Jen J and Waldman T: Oncogenic beta-catenin is required for bone morphogenetic protein 4 expression in human cancer cells. *Cancer Res* 62: 2744-2748, 2002.
- 57 He XC, Zhang J, Tong WG, Tawfik O, Ross J, Scoville DH, Tian Q, Zeng X, He X, Wiedemann LM, Mishina Y and Li L: BMP signaling inhibits intestinal stem cell self-renewal through suppression of Wnt-beta-catenin signaling. *Nat Genet* 36: 1117-1121, 2004.
- 58 Howe JR, Bair JL, Sayed MG, Anderson ME, Mitros FA, Petersen GM, Velculescu VE, Traverso G and Vogelstein B: Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nat Genet* 28: 184-187, 2001.
- 59 Zhou XP, Woodford-Richens K, Lehtonen R, Kurose K, Aldred M, Hampel H, Launonen V, Virta S, Pilarski R, Salovaara R, Bodmer WF, Conrad BA, Dunlop M, Hodgson SV, Iwama T, Jarvinen H, Kellokumpu I, Kim JC, Leggett B, Markie D, Mecklin JP, Neale K, Phillips R, Piris J, Rozen P, Houlston RS, Aaltonen LA, Tomlinson IP and Eng C: Germline mutations in *BMPRIA/ALK3* cause a subset of cases of juvenile polyposis syndrome and of Cowden and Bannayan-Riley-Ruvalcaba syndromes. *Am J Hum Genet* 69: 704-711, 2001.
- 60 Peck JW, Oberst M, Bouker KB, Bowden E and Burbelo PD: The RhoA-binding protein, rhopilin-2, regulates actin cytoskeleton organization. *J Biol Chem* 277: 43924-43932, 2002.
- 61 Bellovin DI, Simpson KJ, Danilov T, Maynard E, Rimm DL, Oettgen P and Mercurio AM: Reciprocal regulation of RhoA and RhoC characterizes the EMT and identifies RhoC as a prognostic marker of colon carcinoma. *Oncogene* 25: 6959-6967, 2006.
- 62 Wheeler JM, Kim HC, Efstathiou JA, Ilyas M, Mortensen NJ and Bodmer WF: Hypermethylation of the promoter region of the E-cadherin gene (*CDH1*) in sporadic and ulcerative colitis associated colorectal cancer. *Gut* 48: 367-371, 2001.
- 63 Risch N and Merikangas K: The future of genetic studies of complex human diseases. *Science* 273: 1516-1517, 1996.
- 64 Bodmer W: Familial adenomatous polyposis (FAP) and its gene, *APC*. *Cytogenet Cell Genet* 86: 99-104, 1999.
- 65 Frayling IM, Beck NE, Ilyas M, Dove-Edwin I, Goodman P, Pack K, Bell JA, Williams CB, Hodgson SV, Thomas HJ, Talbot IC, Bodmer WF and Tomlinson IP: The *APC* variants I1307K and E1317Q are associated with colorectal tumors, but not always with a family history. *Proc Natl Acad Sci USA* 95: 10722-10727, 1998.
- 66 Hanahan D and Weinberg RA: The hallmarks of cancer. *Cell* 100: 57-70, 2000.
- 67 The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306: 636-640, 2004.
- 68 Gorlov IP, Gorlova OY, Sunyaev SR, Spitz MR and Amos CI: Shifting paradigm of association studies: value of rare single-nucleotide polymorphisms. *Am J Hum Genet* 82: 100-112, 2008.
- 69 Bodmer W and Bonilla C: Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40: 695-701, 2008.
- 70 Denecke J, Kranz C, von Kleist-Retzow J, Bosse K, Herkenrath P, Debus O, Harms E and Marquardt T: Congenital disorder of glycosylation type Id: clinical phenotype, molecular analysis, prenatal diagnosis, and glycosylation of fetal proteins. *Pediatr Res* 58: 248-253, 2005.
- 71 Pfarr N, Prawitt D, Kirschfink M, Schroff C, Knuf M, Habermehl P, Mannhardt W, Zepp F, Fairbrother W, Loos M, Burge CB and Pohlentz J: Linking C5 deficiency to an exonic splicing enhancer mutation. *J Immunol* 174: 4172-4177, 2005.
- 72 Wicklow BA, Ivanovich JL, Plews MM, Salo TJ, Noetzel MJ, Lueder GT, Cartegni L, Kaback MM, Sandhoff K, Steiner RD and Triggs-Raine BL: Severe subacute GM2 gangliosidosis caused by an apparently silent *HEXA* mutation (V324V) that results in aberrant splicing and reduced *HEXA* mRNA. *Am J Med Genet A* 127A: 158-166, 2004.

- 73 Xie J, Pabon D, Jayo A, Butta N and Gonzalez-Manchon C: Type I Glanzmann thrombasthenia caused by an apparently silent beta3 mutation that results in aberrant splicing and reduced beta3 mRNA. *Thromb Haemost* 93: 897-903, 2005.
- 74 Fearnhead NS, Wilding JL, Winney B, Tonks S, Bartlett S, Bicknell DC, Tomlinson IP, Mortensen NJ and Bodmer WF: Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc Natl Acad Sci USA* 101: 15992-15997, 2004.
- 75 Fearnhead NS, Winney B and Bodmer WF: Rare variant hypothesis for multifactorial inheritance: susceptibility to colorectal adenomas as a model. *Cell Cycle* 4: 521-525, 2005.
- 76 Kryukov GV, Pennacchio LA and Sunyaev SR: Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet* 80: 727-739, 2007.
- 77 Hirschhorn JN and Altshuler D: Once and again-issues surrounding replication in genetic association studies. *J Clin Endocrinol Metab* 87: 4438-4441, 2002.
- 78 Felix R, Bodmer W, Fearnhead NS, van der Merwe L, Goldberg P and Ramesar RS: *GSTM1* and *GSTT1* polymorphisms as modifiers of age at diagnosis of hereditary nonpolyposis colorectal cancer (HNPCC) in a homogeneous cohort of individuals carrying a single predisposing mutation. *Mutat Res* 602: 175-181, 2006.
- 79 Webb EL, Rudd MF, Sellick GS, El Galta R, Bethke L, Wood W, Fletcher O, Penegar S, Withey L, Qureshi M, Johnson N, Tomlinson I, Gray R, Peto J and Houlston RS: Search for low penetrance alleles for colorectal cancer through a scan of 1467 non-synonymous SNPs in 2575 cases and 2707 controls with validation by kin-cohort analysis of 14704 first-degree relatives. *Hum Mol Genet* 15: 3263-3271, 2006.
- 80 Houlston RS and Tomlinson IP: Polymorphisms and colorectal tumor risk. *Gastroenterology* 121: 282-301, 2001.
- 81 Laken SJ, Petersen GM, Gruber SB, Oddoux C, Ostrer H, Giardiello FM, Hamilton SR, Hampel H, Markowitz A, Klimstra D, Jhanwar S, Winawer S, Offit K, Luce MC, Kinzler KW and Vogelstein B: Familial colorectal cancer in Ashkenazim due to a hypermutable tract in *APC*. *Nat Genet* 17: 79-83, 1997.
- 82 Woodage T, King SM, Wacholder S, Hartge P, Struwing JP, McAdams M, Laken SJ, Tucker MA and Brody LC: The *APC* I1307K allele and cancer risk in a community-based study of Ashkenazi Jews. *Nat Genet* 20: 62-65, 1998.
- 83 Gryfe R, Di Nicola N, Lal G, Gallinger S and Redston M: Inherited colorectal polyposis and cancer risk of the *APC* I1307K polymorphism. *Am J Hum Genet* 64: 378-384, 1999.
- 84 Krontiris TG, Devlin B, Karp DD, Robert NJ and Risch N: An association between the risk of cancer and mutations in the *HRAS1* minisatellite locus. *N Engl J Med* 329: 517-523, 1993.
- 85 Krontiris TG, DiMartino NA, Colb M, Mitcheson HD and Parkinson DR: Human restriction fragment length polymorphisms and cancer risk assessment. *J Cell Biochem* 30: 319-329, 1986.
- 86 Wyllie FS, Wynford-Thomas V, Lemoine NR, Williams GT, Williams ED and Wynford-Thomas D: *Ha-ras* restriction fragment length polymorphisms in colorectal cancer. *Br J Cancer* 57: 135-138, 1988.
- 87 Klingel R, Mittelstaedt P, Dippold WG and Meyer zum Buschenfelde KH: Distribution of *Ha-ras* alleles in patients with colorectal cancer and Crohn's disease. *Gut* 32: 1508-1513, 1991.
- 88 Gosse-Brun S, Sauvaigo S, Daver A, Larra F, Kwiatkowski F, Bignon YJ and Bernard-Gallon D: Association between *H-ras* minisatellite and colorectal cancer risk. *Anticancer Res* 18: 2611-2616, 1998.
- 89 Ceccherini-Nelli L, De Re V, Viel A, Molaro G, Zilli L, Clemente C and Boiocchi M: *Ha-ras-1* restriction fragment length polymorphism and susceptibility to colon adenocarcinoma. *Br J Cancer* 56: 1-5, 1987.
- 90 Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB and Issa JP: CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci USA* 96: 8681-8686, 1999.
- 91 Breivik J and Gaudernack G: Genomic instability, DNA methylation, and natural selection in colorectal carcinogenesis. *Semin Cancer Biol* 9: 245-254, 1999.
- 92 Kim YI, Pogribny IP, Salomon RN, Choi SW, Smith DE, James SJ and Mason JB: Exon-specific DNA hypomethylation of the *p53* gene of rat colon induced by dimethylhydrazine. Modulation by dietary folate. *Am J Pathol* 149: 1129-1137, 1996.
- 93 Jacques PF, Bostom AG, Williams RR, Ellison RC, Eckfeldt JH, Rosenberg IH, Selhub J and Rozen R: Relation between folate status, a common mutation in methylenetetrahydrofolate reductase, and plasma homocysteine concentrations. *Circulation* 93: 7-9, 1996.
- 94 Slattery ML, Potter JD, Samowitz W, Schaffer D and Leppert M: Methylenetetrahydrofolate reductase, diet, and risk of colon cancer. *Cancer Epidemiol Biomarkers Prev* 8: 513-518, 1999.
- 95 Park KS, Mok JW and Kim JC: The 677C>T mutation in 5,10-methylenetetrahydrofolate reductase and colorectal cancer risk. *Genet Test* 3: 233-236, 1999.
- 96 Ma J, Stampfer MJ, Giovannucci E, Artigas C, Hunter DJ, Fuchs C, Willett WC, Selhub J, Hennekens CH and Rozen R: Methylenetetrahydrofolate reductase polymorphism, dietary interactions, and risk of colorectal cancer. *Cancer Res* 57: 1098-1102, 1997.
- 97 Chen J, Giovannucci E, Kelsey K, Rimm EB, Stampfer MJ, Colditz GA, Spiegelman D, Willett WC and Hunter DJ: A methylenetetrahydrofolate reductase polymorphism and the risk of colorectal cancer. *Cancer Res* 56: 4862-4864, 1996.
- 98 Chen J, Giovannucci E, Hankinson SE, Ma J, Willett WC, Spiegelman D, Kelsey KT and Hunter DJ: A prospective study of methylenetetrahydrofolate reductase and methionine synthase gene polymorphisms, and risk of colorectal adenoma. *Carcinogenesis* 19: 2129-2132, 1998.
- 99 Scott RJ: Modifier genes and HNPCC: variable phenotypic expression in HNPCC and the search for modifier genes. *Eur J Hum Genet* 16: 531-532, 2008.

Received September 17, 2009

Revised October 30, 2009

Accepted October 30, 2009