

From the Department of Molecular Medicine and Surgery  
Karolinska Institutet, Stockholm, Sweden

## **Investigation of genetic factors involved in colorectal cancer predisposition**

Tatjana Djureinovic



**Karolinska  
Institutet**

Stockholm 2006

All previously published papers were reproduced with permission from the publisher.

Published and printed by Karolinska University Press  
Box 200, SE-171 77 Stockholm, Sweden  
© Tatjana Djureinovic, 2006  
ISBN 91-7140-864-9

## ABSTRACT

Colorectal cancer (CRC) is one of the leading causes of cancer-related death in the western world. Several hereditary CRC predisposing syndromes have been described and their genetic component has been elucidated. Taken together, these syndromes account for at most 5% of all CRC cases. Twin studies have, however, suggested a much stronger contribution of genetic factors suggesting that a number of CRC predisposing genes remain to be identified.

The genome-wide linkage analysis in 18 non-FAP/non-HNPCC colorectal cancer families revealed regions of interest on three chromosomes, when the analysis was performed under the assumption of locus heterogeneity. The region on chromosome 22q was suggested in the parametric analysis, while the results of both parametric and nonparametric analysis provided support for the regions on chromosomes 11q and 14q. After finemapping of the regions on chromosomes 11q and 14q both the HLOD score and the NPL scores were reduced but still within the range of suggestive linkage. Families exhibiting linkage to chromosomes 11 and 14 were identified and overlapping regions were determined (11q13.2–13.4, 11q22.1–23.1 and 14q23.1–24.1). (Paper I)

The *SMARCA3* gene has been shown to be a common target for methylation in colon and gastric cancer. Germline mutation screening of the *SMARCA3* gene was performed in order to evaluate the role of this gene among Swedish colorectal cancer patients, some of whom also had a family history of gastric cancer. The lack of pathogenic germline mutations suggests that the gene has very little role, if any, in the predisposition to colorectal and gastric cancer. Several identified variants in the *SMARCA3* gene could act as modifying or low-risk alleles, however additional studies are needed to determine their role. (Paper II)

The *CHEK2* 1100delC, novel low-risk breast cancer predisposing allele, has been found at particularly high frequency in families with both colorectal and breast cancer. To investigate the possible role of this variant in CRC predisposition the variant frequency was determined in CRC cases and controls. No over-representation of the variant was detected in cases. However, due to the low frequency of the variant in the Swedish population, a very low penetrance effect of *CHEK2* 1100delC could not be excluded. (Paper III)

Bi-allelic germline mutations in the *MUTYH* gene are known to predispose to recessively inherited MAP syndrome, characterized by the occurrence of an increased number of colonic polyps. We found no evidence for the contribution of *MUTYH* in familial CRC with a low polyp number. The two most common *MUTYH* mutations, Y165C and G382D, have been identified in the Swedish population and evidence for a slightly increased CRC risk among heterozygote carriers of these mutations was seen. In addition, three novel variants affecting the same amino acid position, R423Q, R423P and R423R, have been detected among sporadic CRC cases, however their significance remains to be determined. (Paper IV)

Keywords: colorectal cancer (CRC), hereditary colorectal cancer (HCRC), two-close relatives (TCR), linkage analysis, case-control

# LIST OF PUBLICATIONS

This thesis is based on the following papers, which will be referred to in the text by their Roman Numerals:

- I. **Tatjana Djureinovic\***, Johanna Skoglund\*, Jana Vandrovцова, XiaoLei Zhou, Antonia Kalushkova, Lennart Iselius and Annika Lindblom. A genome-wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer. *Gut*, 2006 Mar;55(3):362-6
- II. **Tatjana Djureinovic**, Simone Picelli, Antonia Kalushkova and Annika Lindblom. Mutation screening of the *SMARCA3* gene in Swedish colorectal cancer patients. *Manuscript*
- III. **Tatjana Djureinovic**, Annika Lindblom, Johan Dalén, Stefan Dedorson, David Edler, Fredrik Hjern, Jörn Holm, Claes Lenander, Ulrik Lindfors, Nils Lundqvist, Hans Olivecrona, Louise Olsson, Lars Pahlman, Jörgen Rutegård, Kennet Smedh, Anders Törnqvist, Hans Eiberg and Marie Luise Bisgaard. The CHEK2 1100delC variant in Swedish colorectal cancer. *Anticancer research* – in press
- IV. Xiao-Lei Zhou, **Tatjana Djureinovic**, Barbro Werelius, Gudrun Lindmark, X-F Sun, Annika Lindblom and the Swedish Low-Risk Colorectal Cancer Group. Germline mutations in the MYH gene in Swedish familial and sporadic colorectal cancer. *Genetic testing*, 9(2):147-51

# CONTENTS

1	INTRODUCTION.....	1
1.1	GATEKEEPERS AND CARETAKERS .....	1
1.2	CANCER EPIGENETICS .....	3
1.3	HEREDITARY CANCER SYNDROMES.....	5
1.4	COLORECTAL CANCER.....	6
1.4.1	Genetic model of colorectal carcinogenesis .....	7
1.4.2	Genomic instability .....	9
1.4.3	Sporadic vs. familial CRC .....	12
1.4.4	Familial adenomatous polyposis (FAP) .....	13
1.4.5	Hereditary nonpolyposis colorectal cancer (HNPCC) .....	14
1.4.6	MUTYH-associated polyposis (MAP) .....	16
1.4.7	Harmatomatous syndromes.....	17
1.4.8	Non-HNPCC/non-FAP familial CRC .....	18
1.4.9	Novel CRC loci .....	21
1.5	STRATEGIES FOR FINDING NOVEL CANCER GENES .....	22
1.5.1	Linkage analysis .....	23
1.5.2	Allelic association .....	25
1.5.3	Somatic alterations in tumors.....	26
2	AIMS .....	27
3	MATERIAL AND METHODS.....	28
3.1	Patient material .....	28
3.2	Linkage analysis .....	29
3.3	Direct sequencing .....	29
3.4	Reverse-Transcriptase PCR (RT-PCR).....	30
3.5	Single nucleotide primer extension (SNuPE) analysis.....	31
3.6	Southern blot.....	32
3.7	Allele-specific PCR amplification .....	32
3.8	Denaturing high performance liquid chromatography (DHPLC) ...	33
3.9	Statistical analysis.....	34
4	RESULTS AND DISCUSSION.....	35
4.1	Paper I .....	35
4.2	Paper II.....	36
4.3	Paper III.....	37
4.4	Paper IV .....	38
5	CONCLUSIONS .....	40
6	Acknowledgements .....	42
7	References.....	44

## LIST OF ABBREVIATIONS

3'UTR	3' untranslated region
5'UTR	5' untranslated region
ACN	Acetonitril
AFAP	Attenuated familial adenomatous polyposis colo
APC	Adenomatous polyposis coli
BER	Base excision repair
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CRC	Colorectal cancer
ddNTP	Dideoxynucleoside triphosphate
DHPLC	Denaturing high performance liquid chromatography
DNA	Deoxiribonucleic acid
dNTP	Deoxynucleoside triphosphate
FAP	Familial adenomatous polyposis
HBCC	Hereditary breast and colon cancer
HCRC	Hereditary colorectal cancer
HLOD	Heterogeneity logarithm of the odds
hMLH1	Human MutL homolog 1 gene
hMLH3	Human MutL homolog 3 gene
HMPS	Hereditary mixed polyposis syndrome
hMSH2	Human MutS homolog 2 gene
hMSH3	Human MutS homolog 3 gene
hMSH6	Human MutS homolog 6 gene
HNPCC	Hereditary nonpolyposis colorectal cancer
hPMS1	Human post meiotic segregation homolog 1 gene
hPMS2	Human post meiotic segregation homolog 2 gene
IBD	Identical by descent
IBS	Identical by state
LOD	Logarithm of the odds
LOH	Loss of heterozygosity
LOI	Loss of imprinting
MAP	MUTYH-associated polyposis
MMR	Mismatch repair
MSI	Microsatellite instability
MUTYH	Mut Y homolog
NPL	Nonparametric linkage
P53	Tumor protein 53 gene
PCR	Polymerase chain reaction
Rb1	Retinoblastoma gene
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase polymerase chain reaction
SNP	Single nucleotide polymorphism
SNUPE	Single nucleotide primer extension
TCR	Two close relatives

# 1 INTRODUCTION

Cancer is a genetic disease, however there is a difference between cancer and classical genetic diseases, such as Cystic fibrosis or Duchenne muscular dystrophy, in which inherited mutations in corresponding genes inevitably cause the disease. Although a certain fraction of cancers are associated with clear genetic component, inherited mutations in cancer-associated genes are not sufficient for cancer development. Instead, cancer will only develop after the accumulation of additional somatic mutations in a number of genes (Vogelstein and Kinzler 1993). Genetic alterations commonly found in cancer cells include subtle alterations (such as single-base substitutions, small insertions or deletions), chromosome number changes, chromosomal translocations and amplifications. In some instances exogenously introduced genes of viral origin have been associated with development of cancers (as in the case of cervical cancer and human papilloma viral infection).

Several hundreds of different types of cancer affecting different organs have been described. Depending on the origin of cells they are derived from cancers are in general classified as carcinomas (epithelial cells), lymphomas and leukemias (blood and bone marrow cells), melanomas (melanocytes), sarcomas (mesenchymal cells), retinoblastomas (stem cells of retina), neuroblastomas (stem cells of neurons) and glioblastomas (glial stem cells). Although cancers represent a heterogeneous group of diseases that differ in the cause, symptoms and prognosis, common properties that are shared between different cancers have been found. Six essential physiological changes found in almost all cancer types include: self-sufficiency in growth signals, insensitivity to antigrowth signals, evasion of apoptosis, limitless replicative potential, sustained angiogenesis and tissue invasion and metastasis (Hanahan and Weinberg 2000).

## 1.1 GATEKEEPERS AND CARETAKERS

As mentioned above, cancer development is a multistep process in which mutations in a number of genes accumulate over the time. Genes commonly found mutated in cancers can broadly be classified as gatekeepers and caretakers. Gatekeepers are composed of oncogenes and tumor suppressor genes. Both groups of genes, when

mutated, directly regulate tumor growth by stimulating cell growth or inhibiting cell death (Kinzler and Vogelstein 1997).

Oncogenes were initially identified as viral elements that, when transferred to normal cells, initiate tumorigenesis (Huebner and Todaro 1969). Their cellular counterparts, proto-oncogenes, are highly conserved genes whose products play important roles in the regulation of cell cycle progression, cell division and differentiation (Vogelstein and Kinzler 2004). Based on their normal function proto-oncogenes can be classified into five groups: growth factors, growth factor receptors, transcription factors, signal transducers and regulators of programmed cell death. Proto-oncogenes are activated by gain-of-function mutations that can either change the structure of proteins making them constitutively active (like in the case of subtle changes within the coding region or chromosomal translocations generating fusion products) or that can increase the level of normal protein (which is achieved by gene amplification or translocation to a region of active transcription). These activating mutations act as a dominant event since a change in only one allele is usually enough to confer the growth advantage.

Tumor suppressor genes play a crucial role in growth inhibition, cell cycle arrest and promotion of apoptosis. The first identified tumor suppressor gene was the retinoblastoma (*Rb1*) gene (Friend *et al.* 1986). Based on statistical study of retinoblastoma in children Alfred Knudson proposed the “two-hit” hypothesis which was later found to be applicable for majority of tumor suppressor genes (Knudson 1971). He observed that children with bilateral tumors had a significantly earlier age of onset, compared to the children with a unilateral tumor. Based on that observation he proposed that two genetic hits were required to turn a normal cell into a tumor cell and that, in the case of bilateral tumors, the first hit was inherited. These hits were, in subsequent studies, found to be inactivating mutations in the *Rb1* tumor suppressor gene. Unlike dominantly acting gain-of-function mutations in oncogenes, recessive loss-of-function mutations that inactivate both alleles of the tumor suppressor genes occur during tumorigenesis. In accordance with the two hit hypothesis, most tumor suppressor genes are inactivated by intragenic mutations (point mutations, small insertions/deletions) of one allele in combination with chromosomal changes resulting in the deletion of the other allele. In addition to deletions, one of the alleles of most tumor suppressor genes can also be silenced by DNA hypermethylation (Jones and Baylin 2002). Not all tumor suppressor genes however follow the Knudson’s two hit

hypothesis. For some tumor suppressor genes, like *DMP1*, haploinsufficiency or loss of only one allele while the other remains active is found to be enough to confer the selective advantage (Santarosa and Ashworth 2004).

Caretakers or stability genes are involved in the maintenance of genome integrity. (Kinzler and Vogelstein 1997). Several repair pathways with partially overlapping functions are known to play a role in genome maintenance. Subtle lesions that occur during normal DNA replication or that are introduced after exposure to mutagens are repaired by nucleotide excision repair (NER), mismatch repair (MMR) and base excision repair (BER) system, while homologous recombination and non-homologous end joining are involved in the repair of lesions affecting larger parts of chromosomes such as double strand breaks. In contrast to gatekeepers, caretakers regulate tumor growth indirectly. Their inactivation leads to genetic instability resulting in an increased mutation rate. Mutations will affect all genes, however only when gatekeepers are affected there is a high probability for tumors to develop. Similar to tumor suppressor genes, two inactivating loss-of-function mutations affecting both alleles of caretaker genes are required for disease initiation.

The role of caretakers has particularly been highlighted after the identification of inherited cancer predisposing syndromes caused by germline mutations in these genes. Increased predisposition to skin cancer is seen in patients affected with Xeroderma pigmentosum (XP), an inherited NER disease characterized by extreme sun sensitivity (Cleaver 2005). Germ line mutations in the MMR genes, such as *MSH2* and *MLH1* are known to cause hereditary nonpolyposis colorectal cancer (HNPCC) (Lindblom *et al.* 1993; Peltomaki *et al.* 1993). Bi-allelic mutations in the *MUTYH* gene, a member of the BER pathway, are also known to predispose to colorectal cancer causing recessive *MUTYH*-associated polyposis (MAP) syndrome (Al-Tassan *et al.* 2002). Mutations in genes involved in the repair of double strand breaks, such are *ATM*, *BRCA1* and *BRCA2*, predispose to cancer-prone Ataxia telangiectasia and breast cancer, respectively (Hoeijmakers 2001).

## **1.2 CANCER EPIGENETICS**

Epigenetics refers to all modifications to genes, other than changes in the DNA sequence itself, which are heritable during cell division. It is becoming increasingly

apparent that cancer arises through a series of not only genetic but also epigenetic alterations. The first epigenetic abnormality detected in cancer cells was loss of DNA methylation at CpG nucleotides. In 1983 Fearon *et al.* reported global hypomethylation of CpGs in colorectal tumors compared with surrounding normal cells (Feinberg and Vogelstein 1983). Hypomethylation leads to the activation of genes through demethylation of normally methylated CpG islands located within their promoter regions. Many of the genes found to be over-expressed due to hypomethylation, like *HRAS* and *cyclin D2*, are known to play an important role in cancer development (Feinberg and Vogelstein 1983; Oshimo *et al.* 2003). The mechanisms leading to global hypomethylation in cancers are still unknown, however some data suggest that components of the SWI/SNF chromatin remodeling complex may play a role in this process. ATRX and Lsh, both members of the SNF2 family, have been found to be required for the maintenance of a normal methylation profile (Gibbons *et al.* 2000; Fan *et al.* 2003). In addition, mutations in some members of the SWI/SNF complex have been detected in a number of primary cancers and cell lines (Versteeg *et al.* 1998; Hiramoto *et al.* 1999; Wong *et al.* 2000; Schmitz *et al.* 2001). Whether these mutations are associated with global hypomethylation in the corresponding cancers remains to be determined.

Not only hypomethylation but also hypermethylation of CpG islands has frequently been detected in tumors. DNA hypermethylation plays a key role in gene silencing. The first tumor suppressor gene found to be silenced through promoter hypermethylation was *Rb1* (Sakai *et al.* 1991). Almost half of the tumor suppressor genes implicated in familial cancers are known to be inactivated by DNA hypermethylation and it seems that this is as frequent event as mutations occurring within the coding region of these genes.

Genomic imprinting refers to parent-of-origin-specific allele silencing which is maintained, at least in part, by differential methylation of regions where imprinted genes reside and which is reprogrammed in the germline. Imprinting abnormalities lead to aberrant gene expression and they are often manifested as developmental and neurological disorders, if changes occur early during development, and as cancers, if they occur later in life (Falls *et al.* 1999). In particular, loss of imprinting (LOI) of the insulin-like growth factor-II gene (*IGF2*) is detected in several common cancers including that of colorectum (Cui *et al.* 1998). LOI of the *IGF2* gene has also been

detected in normal colonic mucosa as well as in blood samples from colorectal cancer patients suggesting that it can identify a subset of population with colorectal cancer (CRC) or at risk of developing CRC (Cruz-Correa *et al.* 2004).

Recently, an epigenetic model of cancer has been proposed, suggesting that the first step in tumor formation is epigenetic alteration of progenitor cells (Feinberg *et al.* 2006). In addition, hypermethylation of the *hMLH1* gene and the *hMSH2* gene occurring in the germline have been observed (Suter *et al.* 2004; Chan *et al.* 2006). These epimutations are thought to act as an alternative to mutations in *hMLH1* and *hMSH2*, genes known to predispose to hereditary nonpolyposis colorectal cancer (HNPCC). Epimutations were shown to occur as a mosaic and, in the case of *hMSH2*, were also transmitted to offspring. Additional data are, however needed in order to elucidate the role of epimutations.

### **1.3 HEREDITARY CANCER SYNDROMES**

The cancer cells genome, compared to the genome of their progenitors, is altered at multiple sites. In most cases these changes represent somatic events, resulting in sporadic tumors. However, if one change is already present in the germline (predisposing event) this will cause cancer predisposition as seen in hereditary cancer syndromes. The presence of two or more family members affected with the same cancer or with related cancers (like colorectal and endometrial in the case of HNPCC) in combination with young age at onset (younger than in the case of sporadic cancers) and multiple primary cancers usually points to hereditary cancer syndromes. More than 200 hereditary cancer syndromes affecting different organs have been identified so far. Although these syndromes account for only a small fraction of all cancers, elucidation of their genetic basis has significantly contributed to our understanding of cancer biology.

A major fraction of known hereditary cancer syndromes exhibit an autosomal dominant mode of inheritance and are mostly caused by germline mutations in one of the tumor suppressor or caretaker genes. As both types of genes act recessively at the cellular level, an inherited mutation in one allele does not lead to the disease on its own. In addition to the mutation present in the germline, the second allele also needs to be inactivated and this is achieved somatically, in the target tissue. While most common

cancers, like breast and colorectal, are caused by inactivating mutations in tumor suppressor or repair genes, some less frequent cancer syndromes that are associated with activating oncogenic mutations also exist. The first proto-oncogene found to be implicated in inherited cancer susceptibility syndrome was the *RET* gene. Activating germline mutations in this gene are responsible for multiple endocrine neoplasia type 2 (MEN2) (Santoro *et al.* 2002). In addition to MEN2, gastrointestinal stromal tumors (GIST) and hereditary papillary renal carcinoma (HPRC) are also caused by activating germline mutations in the proto-oncogenes, *KIT* and *MET* respectively. Many of the known cancer susceptibility genes show high penetrance which can in some cases even reach 100%, like in the case of adenomatous polyposis coli (*APC*) gene.

Besides dominant inheritance, a small percentage of cancer susceptibility is attributed to recessively inherited cancer predisposing syndromes. Most of these syndromes, like Ataxia telangiectasia, Bloom syndrome, Fanconi anemia, Xeroderma pigmentosum, represent childhood diseases where increased cancer predisposition is only a part of a broad spectrum of phenotypic characteristics. *MUTYH*-associated polyposis (MAP) is the only known recessively inherited syndrome with disease developing in adulthood. Moreover, it is the only recessive syndrome with increased cancer susceptibility, to cancers of colon and rectum in this case, as its main feature.

#### **1.4 COLORECTAL CANCER**

Colorectal cancer (CRC) is the second most prevalent cancer worldwide and it affects men and women almost equally (Parkin *et al.* 2005). In 2002 CRC accounted for close to one million cancer diagnoses and about half million cancer deaths (Parkin *et al.* 2005). In Sweden, according to the data of cancer incidence from 2004, CRC ranked second in females, after breast cancer, and fourth in males, after prostate, skin and lung cancer (The National Board of Health and Welfare, 2004). The incidence of CRC differs at least 25-fold worldwide, with the highest rate seen in North America, Australia, New Zealand and Western Europe and a low rate seen in Africa and Asia (Parkin *et al.* 2005). Studies of migrants moving from low-risk to high-risk areas show that the incidence of CRC increases to the level of their adoptive country even within the first generation. These data clearly suggest a contribution of environmental and dietary factors in CRC development. A diet rich in animal fat, alcohol use and cigarette smoking are some of the factors positively related to the CRC risk, while increased

intake of dietary fiber, physical activity and hormone replacement therapy seems to lower risk of disease, although some data are still controversial (Boyle and Langman 2000). The majority of CRC patients develop disease during the sixth decade of life or later suggesting the age as one of the important risk factors. In addition, an increased predisposition to CRC has also been observed in conditions such as Crohn's disease and ulcerative colitis (Vagefi and Longo 2005).

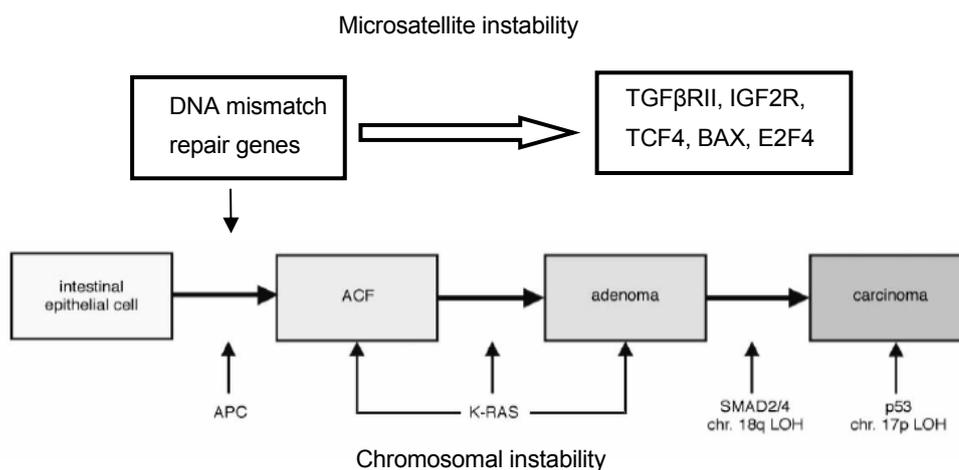
By far the single most important risk factor for CRC development is a family history of the disease. First degree relatives of CRC patients have a 2-3 fold increased risk for the disease (Johns and Houlston 2001). The familial risk increases with multiple affected relatives and it is also related to the age at diagnosis. Several CRC predisposing syndromes with highly penetrant single gene mutations have been identified. Familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC) are dominantly inherited syndromes that are caused by mutations in the adenomatous polyposis coli gene (*APC*) and mismatch repair genes (MMR), respectively (de la Chapelle 2004). Inactivating mutations in the *MUTYH* gene, member of the BER pathway, are the cause of recessively inherited *MUTYH*-associated polyposis (MAP) (Sampson *et al.* 2005). In the analysis of cancer incidence among twins, CRC was suggested to be one of the most heritable cancers with the contribution of genetic factors estimated to be ~ 35% (Lichtenstein *et al.* 2000). However, known CRC hereditary syndromes are found in at most 5% of all cases, indicating that a number of CRC predisposing genes remain to be identified (de la Chapelle 2004).

#### **1.4.1 Genetic model of colorectal carcinogenesis**

The polyp-cancer sequence of CRC development recognized by Muto *et al.* suggested a slow evolution of tumors from normal colonic epithelium via adenomatous dysplasia, a process in which cells evolve from the ones showing mild dysplasia to the ones showing severe dysplasia, resulting in malignant transformation and eventually metastasis (Muto *et al.* 1975). Due to the relative availability of biopsies from different histopathological steps of tumor development, the molecular pathology of CRC has been studied more extensively than any other cancer type resulting in the discovery of cancer related genetic alterations. Based on these results a model that can also be applied to other common epithelial tumors has been developed.

The smallest lesions that can be observed in the colorectum are aberrant crypt foci (ACF), small areas of epithelium with irregular architecture that can exhibit either normal, hyperplastic or dysplastic morphology (Jen *et al.* 1994; Molatore and Ranzani 2004). Polyps, a mass of cells protruding from the bowel wall, are classified into two groups, according to their histology. The most common lesions found in adults are hyperplastic polyps (HP). They consist of a large number of cells lined up in a single row that exhibit normal morphology. Adenomatous polyps are on the other hand dysplastic. They show abnormal intra and extracellular organization, with several layers of epithelial cells lying on the basement membrane and with enlarged nuclei exhibiting a different position than those in normal cells. While adenomas are traditionally regarded as the precursor lesions for the majority of CRCs and HPs considered as non-neoplastic, recent data suggest the role of not just adenomas but also HPs in CRC development (Higuchi and Jass 2004).

The “adenoma-carcinoma sequence” proposed by Fearon and Vogelstein describes a stepwise evolution of CRC resulting from the accumulation of specific genetic changes, each of them associated with recognizable histopathological changes (Fearon and Vogelstein 1990). At least four genetic events are needed for CRC development and they include inactivation of three tumor suppressor genes and activation of one oncogene (Figure 1).



**Figure 1.** Schematic view over genetic events involved in CRC development (modified from Fearon *et al.* 1990 and Fodde *et al.* 2002).

The earliest and most frequent genetic alteration detected in colorectal tumors is inactivation of the *APC* gene (Powell *et al.* 1992). Germline mutations in *APC* are the cause of dominantly inherited familial adenomatous polyposis (FAP) syndrome whereas somatic inactivation of this gene has been detected in the majority of the sporadic CRCs (Kinzler and Vogelstein 1996). The protein encoded by the *APC* gene has several important functions, however it is believed that its tumor suppressing activity is based on the regulation of the intracellular  $\beta$ -catenin level within the WNT pathway (Fodde *et al.* 2001). Activating *KRAS* mutations are found in 50% of adenomas and carcinomas and seem to be required for adenoma growth and progression (Bos *et al.* 1987; Fodde *et al.* 2001). A high frequency of loss of heterozygosity (LOH) involving chromosomes 18q and 17p has been detected in a large proportion of adenomas and carcinomas. (Fearon *et al.* 1987; Fearon *et al.* 1990). The most probable tumor suppressor genes from the region on chromosome 18q are *SMAD2* and *SMAD4*, both encoding key signaling molecules within the transforming growth factor  $\beta$  (TGF $\beta$ ) pathway. Inactivation of *p53* located on chromosome 17p is found in approximately 75% of colorectal carcinomas but very rarely in benign lesions, suggesting that this is a late event associated with tumor progression (Baker *et al.* 1989; Rodrigues *et al.* 1990).

Other genetic events, besides the one mentioned above, seem to be needed for colorectal carcinogenesis. The fact that 50% of CRCs do not have detectable mutations in the *KRAS* genes suggests the involvement of other genes with equivalent function. In addition, the order of events rather than their accumulation is suggested to drive tumorigenesis (Kinzler and Vogelstein 1996). *KRAS* mutations are common among non-dysplastic lesions with limited potential to progress to cancer (Jen *et al.* 1994). Similarly, germline mutations in the *p53* gene do not predispose to CRC.

#### **1.4.2 Genomic instability**

As several genetic events are needed for tumor development the question is whether the normal mutation rate is sufficient for these changes to occur, or whether tumor cells exhibit an increased mutation rate as compared to normal cells. Identification of germline mutations in the mismatch repair (MMR) genes leading to hereditary non-polyposis colorectal cancer (HNPCC) provided an answer for at least 15% of all CRC cases. As a consequence of deficient MMR the mutation rate in these tumor cells is

100-1000-fold greater than in normal cells resulting in microsatellite instability (MSI). However, in 85% of colorectal cancers no MSI can be found. Instead, a high rate of instability at chromosomal level is detected (chromosomal instability - CIN). Whether CIN is the cause or the effect of tumor progression is still a matter of debate but the fact that tumor cells can exhibit either MSI or CIN implies that CIN may have similar role as MSI and that one type of instability is required for cancers to form.

#### *1.4.2.1 Chromosomal instability (CIN)*

Tumors from FAP patients as well as the majority of sporadic CRCs with inactivating mutations in the *APC* gene exhibit CIN. These tumors are typically located within the distal colon and they are characterized by multiple numerical and structural chromosomal abnormalities (Lindblom 2001). These aberrations are so complex that representative karyotypes, like the ones found in leukemia and lymphomas, can not be identified. CIN occurs early in the adenoma to carcinoma sequence as a dominant trait, however the mechanisms leading to CIN have not been completely elucidated (Lengauer *et al.* 1997). Alterations of genes involved in chromosome metabolism, spindle assembly, cell cycle regulation and mitotic checkpoint control generate CIN in yeast, however no clear connection between mutations in these genes and CRC has been found so far (Jallepalli and Lengauer 2001). A growing body of evidence suggests that the APC protein may play a role in the maintenance of chromosomal stability. APC is found to localize to kinetochores of metaphase chromosomes through the interaction between its C-terminus and end-binding protein EB1 (Fodde *et al.* 2001; Kaplan *et al.* 2001). APC deficient cells exhibit both numerical and structural chromosomal abnormalities resulting from defect in the proper attachment of the mitotic spindle to the dividing chromosomes and the regulation of centrosome duplication through the interaction of APC with tubulin and the centrosomes (Fodde *et al.* 2001). Recently, however, it has been demonstrated that aberrant WNT/ $\beta$ -catenin signaling caused by *APC* mutations leads to CIN through upregulation of the conductin protein, another member of WNT pathway (Hadjihannas *et al.* 2006). Interestingly, some studies have even found a connection between CIN and polyoma JC virus infection (Niv *et al.* 2005).

#### 1.4.2.2 *Microsatellite instability (MSI)*

Microsatellites are short tandemly repeated sequences of one to six nucleotides that are spread throughout the genome, occurring in both coding and noncoding regions. These repetitive elements are prone to slippage of polymerase machinery during replication resulting in deletions or insertions, depending on whether the template or the newly synthesized strand is affected. These errors are normally repaired by the MMR system whose major components are hMLH1, hMSH2, hMSH3, hMSH6, hPMS1 and hPMS2. Deficiency in the MMR system, like the one seen in HNPCC, substantially increases the mutation rate which can easily be seen as the expansion or reduction in microsatellite sequences, a phenomenon known as MSI (Bhattacharyya *et al.* 1994). In contrast to colorectal tumors exhibiting CIN, tumors exhibiting MSI are poorly differentiated, have nearly diploid karyotype and are located mostly in the proximal colon (Jass *et al.* 1998; Jass 2004).

As mentioned previously, caretakers regulate tumor growth indirectly. Thus, inactivation of one of MMR genes does not lead to tumor initiation on its own. As a consequence of defective MMR system a number of genes, particularly the ones with microsatellite or microsatellite-like sequences are affected. Frameshift mutations within the coding poly-A tract of the *TGF $\beta$ R2* gene, a member of the TGF $\beta$  signaling pathway, are found at high frequency in both sporadic and hereditary MSI tumors (Markowitz *et al.* 1995). Other genes with a role in tumorigenesis that are common target for inactivation due to MSI are *IGFIIR*, *PTEN*, *BAX*, *TCF-4*, *hMSH6*, *hMSH3* (Souza *et al.* 1996; Tashiro *et al.* 1997; Yamamoto *et al.* 1997; Duval *et al.* 1999; Schwartz *et al.* 1999).

In addition to HNPCC, approximately 15% of sporadic CRCs exhibit MSI. In the majority of these cancers hypermethylation of the *hMLH1* promoter associated with the loss of hMLH1 protein expression is found (Herman *et al.* 1998). Analysis of the methylation profile of these cancers revealed a high rate of methylation not associated with aging, which was labeled as a CpG island methylator phenotype (CIMP) (Toyota *et al.* 1999). Moreover, it has been shown that MMR deficiency in sporadic CRCs occurs almost exclusively as a consequence of CIMP-associated *hMLH1* methylation (Weisenberger *et al.* 2006).

### 1.4.3 Sporadic vs. familial CRC

Traditionally, CRC is classified as either sporadic or familial (hereditary), where sporadic cancers represent the majority of all cases. Only a few percent of the total burden of CRCs comprise a group of high-risk hereditary syndromes, such as FAP, HNPCC and MAP, in which disease causing genes have already been identified. Since the genetic contribution in CRC has been estimated to be around 35%, there is a major gap in our understanding of the disease. Additional CRC syndromes that contribute to familial clustering of the disease are expected to be found. However, aggregation of cancers in families can also occur due to chance or shared environment. On the other hand, it is not always possible to completely eliminate the effects of genetic factors in a given sporadic case. Small family size, reduced penetrance, lack of information on family history and poor diagnosis all contribute to misclassification of cases as sporadic. Moreover, late onset recessive cancer syndromes are very difficult to recognize and such patients would most probably be classified as sporadic cases, if gene defects causing recessive syndromes are not known. The first indication of recessive mode of CRC inheritance came after the finding that siblings have a significantly higher CRC risk than offspring of affected parents (Johns and Houlston 2001). Identification of inherited bi-allelic mutations in the *MUTYH* gene causing MAP syndrome served as a proof for the existence of recessive CRCs (Al-Tassan *et al.* 2002; Sampson *et al.* 2005). Estimation of a recessive inheritance that would account for 0.75% of all CRCs, based on the observation of an increased risk of right-sided colon cancer, speaks in favor of additional, yet to be found recessive syndromes (Hemminki and Chen 2005).

In 2002 Pharoah *et al.* proposed a polygenic model of breast cancer in which breast cancer susceptibility is mainly due to a large number of alleles acting additively or multiplicatively, each of them associated with a very small risk (Pharoah *et al.* 2002). It is now accepted that the same model can be applied to CRC, as well as other common cancers (Pharoah *et al.* 2002). Therefore, it is reasonable to expect that a certain fraction of CRCs considered to be sporadic will result from the inheritance of a number of low-penetrance variants suggesting that classification of cases as sporadic or familial is not as straightforward as previously thought.

#### 1.4.4 Familial adenomatous polyposis (FAP)

FAP was the first recognized form of inherited CRC. It is an autosomal, dominantly inherited disease found in not more than 1% of all CRC patients (Potter 1999). FAP is clinically characterized by the occurrence of hundreds to thousands of adenomatous polyps throughout the colon and rectum during the second and third decade of life. If left untreated, one or more adenomas will inevitably develop into carcinoma at a mean age of 40 years, implying a 100% penetrance. Tumors are primarily located within the distal colon and they exhibit chromosomal instability. In addition to CRC and adenomas, FAP patients frequently develop extracolonic manifestations such as congenital hypertrophy of the retinal pigment epithelium (CHRPE), osteomas, sebaceous and epidermoid cysts, and desmoid tumors. In some cases extracolonic malignancies are also found. Based on the patient's phenotype several FAP variants have been described. Gardner's syndrome is characterized by the presence of colonic adenomas, desmoid tumors, osteomas, epidermoid cysts and fibromas (Gardner and Richards 1953) while in Turcot syndrome a combination of colorectal and central nervous system tumors, usually childhood cerebellar medulloblastoma, is seen (Hamilton *et al.* 1995). Attenuated FAP (AFAP) is a mild clinical variant of FAP characterized by lower number of adenomas (<100), which tend to aggregate more often in the proximal colon, and it has a later age of CRC diagnosis (50-55 years). The lifetime risk for CRC is as high as in the case of FAP, however the spectrum of extracolonic manifestations in AFAP differs; desmoid tumors and CHRPE are rare or absent while lesions of the upper gastrointestinal tract are commonly found (Soravia *et al.* 1998).

FAP is caused by germline mutations in the adenomatous polyposis coli (*APC*) gene, located on chromosome 5q (Grodén *et al.* 1991; Nishisho *et al.* 1991). More than 600 mutations in the *APC* gene have been described so far, the majority of which results in the production of truncated protein with an abnormal function (<http://www.hgmd.cf.ac.uk/ac/index.php>). In addition, a good genotype-phenotype correlation is seen. The classical FAP phenotype is associated with mutations between codons 169 and 1393 while AFAP results from mutations in either the 5' end, the 3' end of the gene (after codon 1595) or mutations located in alternatively spliced exon 9 (Soravia *et al.* 1998; Moisisio *et al.* 2002). The Gardner variant of FAP is frequently caused by mutations occurring between codons 1403 and 1578 (Soravia *et al.* 1998; Wallis *et al.* 1999).

### 1.4.5 Hereditary nonpolyposis colorectal cancer (HNPCC)

HNPCC, or Lynch syndrome, is the most common form of hereditary CRC and it is found in about 2.5% of all CRC cases (Lynch *et al.* 2006). It is an autosomal, dominantly inherited disease with the predisposition to not only CRC but also to cancers of other sites, most commonly endometrium, stomach, small bowel and ovaries. In contrast to FAP, HNPCC patients develop a small number of adenomatous polyps which however rapidly develop into cancer, usually within two or three years. These tumors are predominantly located in the proximal colon and they tend to be poorly differentiated (Lynch and de la Chapelle 2003). The life-time risk for cancer is about 80% with the mean age at diagnosis of 44 years (Aarnio *et al.* 1999). In addition to colonic neoplasia patients with a rare variant of HNPCC known as Muir-Torre syndrome also develop sebaceous gland tumors, multiple keratoacanthomas, and basal cell carcinomas (Fusaro *et al.* 1996).

HNPCC is caused by germline mutations in one of the DNA mismatch repair (MMR) genes, most frequently involving *hMLH1* and *hMSH2* (Peltomaki and Vasen 2004). Mutations in the *hMSH6* gene have been found in about 5% of HNPCC families exhibiting an attenuated phenotype, where incomplete penetrance, late onset of cancer and a high risk for endometrial cancer among females are seen (Wagner *et al.* 2001; Jarvinen 2004). The role of other MMR genes, *hPMS1/2* and *hMLH3* in HNPCC is still not clear (Liu *et al.* 2001; Liu *et al.* 2003). More than 400 pathogenic mutations in the MMR genes have been reported in the international HNPCC database (<http://www.insight-group.org>). The majority of HNPCC causing mutations are nucleotide changes, however genomic rearrangements, mostly in the *hMSH2* gene, are also detected.

In 1991 The International Collaborative Group on HNPCC (ICG-HNPCC) proposed the first diagnostic criteria, so-called Amsterdam criteria I, with the role of providing uniformity in HNPCC classification (Vasen *et al.* 1991). These criteria were shown to be too stringent as they did not include extracolonic cancers and only large families with required number of cancer cases were considered. As a consequence a significant proportion of HNPCC families were missed. Thus, in 1999, a revised version, the Amsterdam criteria II, including also HNPCC-associated cancers were established (Table 1) (Vasen *et al.* 1999).

**Table 1.** The Amsterdam criteria II and the Bethesda criteria (revised)

---

Amsterdam criteria II

There should be at least three relatives with an HNPCC-associated cancer (CRC, cancers of endometrium, small bowel, ureter or renal pelvis) and:

1. One should be first degree relative of the other two
2. At least two successive generations should be affected
3. At least one cancer should be diagnosed before age 50
4. FAP should be excluded in the CRC case
5. Tumors should be verified by pathological examination

The Bethesda criteria (revised)

Tumors from individuals should be tested for MSI in the following situations:

1. CRC diagnosed in patients who is less than 50 years of age
  2. Presence of synchronous, metachronous CRC or other HNPCC associated tumor, regardless of age
  3. CRC with the MSI-H histology diagnosed in a patient who is less that 60 years of age
  4. CRC in a patient with one or more first-degree relatives with an HNPCC-related tumor, with one of the cancers being diagnosed under age of 50 years
  5. CRC in a patient with two or more first- or second-degree relatives with HNPCC-related tumors, regardless of age
- 

As MSI is a hallmark of HNPCC it is often used as a tool that can facilitate HNPCC identification. For this purpose a set of guidelines known as the Bethesda criteria, the primary role of which was to select patients for MSI screening, have been developed (Table 1) (Umar *et al.* 2004). The MSI test is performed on both tumor and matching normal DNA using a set of five microsatellite markers, two mononucleotide, BAT25 and BAT26, and three dinucleotide repeats, D2S123, D5S346 and D17S250. Tumors exhibiting instability at two or more markers are defined as MSI-high (MSI-H) whereas those with instability at one marker are of MSI-low (MSI-L) type. If no instability is found at any of the five markers tumors are considered to be MSI stable (MSS).

It is worth noting that a number of families fulfilling the Amsterdam criteria I but who do not show MSI are also identified. Whether these families can be classified as HNPCC families is still a matter of debate. However, it has been suggested that these

families are not of HNPCC type and they should be designated as familial CRC of unknown type (Lynch *et al.* 2006).

#### **1.4.6 MUTYH-associated polyposis (MAP)**

MAP is the first and so far the only identified recessively inherited CRC predisposing syndrome. It is caused by bi-allelic germline mutations in the *MUTYH* gene (human MutY homologue), a member of the BER system. MAP is also the only known disease associated with an inherited BER deficiency. The syndrome was identified after investigation of a family with three out of seven siblings exhibiting an AFAP-like phenotype but with no identifiable *APC* mutations (Al-Tassan *et al.* 2002). Mutation screening of the *APC* gene from tumors showed an increased rate of G:C>T:A transversions, suggesting a deficiency in the BER system. Subsequent analysis of the genes belonging to the BER system revealed that all three affected family members were compound heterozygotes for two missense mutations in the *MUTYH* gene, Y165C and G382D, while unaffected members were either heterozygote carriers of one of the mutations or they were homozygous normal. Population-based analysis of MAP suggested that the syndrome may be as frequent as FAP (Enholm *et al.* 2003; Farrington *et al.* 2005). Different founder mutations in *MUTYH* have been found in various ethnic populations. The Y165C and G382D mutations account for more than 80% of pathogenic mutations found in Caucasians whereas E466X and Y90X have been identified in patients of Indian and Pakistani origin, respectively (Chow *et al.* 2004).

MAP carcinogenesis proceeds through the pathway that differs from CIN and MSI but that shares characteristics of both. MAP tumors are near diploid, without MSI and with low level of loss of heterozygosity (LOH) (Lipton *et al.* 2003). A high frequency of somatic *APC* mutations have been found and all of the detected protein-truncating mutations were G>T transversions. Other common target for inactivation is the *KRAS* gene where a G>T transversion affecting codon 12 (G12C) was identified (Lipton *et al.* 2003).

Clinically, MAP is very difficult to distinguish from FAP and AFAP, although it presents later in life and is restricted to the gastrointestinal tract. Disease-causing mutations in the *MUTYH* gene have been found in 20-30% of patients with 15-100

colorectal adenomas and in 10% of patients with classical polyposis (Sampson *et al.* 2003; Sieber *et al.* 2003; Vandrovcova *et al.* 2004). In addition, *MUTYH* mutations have also been identified among CRC patients without adenomas (Farrington *et al.* 2005). Being a recessive disease, a significant number of patients appear to be sporadic and some studies show that in 50% of cases CRC was already found at presentation (Sampson *et al.* 2003; Sieber *et al.* 2003). MAP is shown to be a high risk recessive syndrome with bi-allelic mutations in *MUTYH* showing complete penetrance by the age of 60 years (Farrington *et al.* 2005). As for the heterozygote state alone, early reports including ours suggested a small increase in the CRC risk but no results reached the level of significance (Croitoru *et al.* 2004; Zhou *et al.* 2005). Recently, however, a 1.68-fold excess risk for the heterozygous carriers of 55 years of age or older has been observed (Farrington *et al.* 2005).

#### **1.4.7 Harmatomatous syndromes**

A small proportion of familial CRC arises in the settings of rare harmatomatous syndromes, where an intermediate risk of CRC has been observed.

Peutz-Jegher syndrome (PJS) is a disease that exhibits an autosomal dominant inheritance with reduced penetrance. In addition to harmatomatous polyps that are mostly located in the small intestine, typical mucocutaneous melanotic pigmentation is seen in most patients. An elevated risk of CRC, gastric, pancreatic, ovarian, breast testicular and endometrial cancer is seen (Boardman *et al.* 1998). Germline mutations in the *STK11* (serine/threonine-protein kinase 11) are found in 50-70% of PJS patients (Hemminki *et al.* 1998).

Familial juvenile polyposis syndrome (JPS) is the most common of all harmatomatous syndromes with an incidence of 1:100 000 to 1:160 000. It is a dominant disease characterized by the occurrence of harmatomatous polyps throughout the colon and rectum and with an elevated risk for CRC. In ~18% of JPS patient germline mutations in the *SMAD4* gene have been found, while an additional 20% of patients carry mutation in the *BMPRIA* gene (Howe *et al.* 2004). Both genes are members of the TGF $\beta$ -signaling pathway

Cowdens disease is a rare autosomal, dominant syndrome affecting 1 in a million individuals. Besides hamatomatous polyps, patients develop skin changes and are at risk of malignancies of the breast, thyroid and endometrium (Eng and Peacocke 1998). The majority of CD patients (80%) carry mutations in the *PTEN* gene.

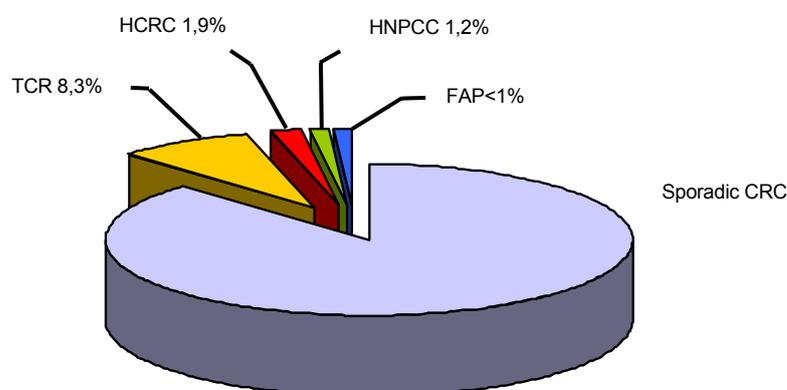
Bannayan-Ruvalcaba-Riley syndrome is also associated with mutations in the *PTEN* gene (Marsh *et al.* 1998). The clinical manifestations are broad and may include macrocephaly, mental retardation, multiple hemangiomas and lipomas, skeletal malformations and, in up to 45% patients, intestinal polyposis.

#### **1.4.8 Non-HNPCC/non-FAP familial CRC**

Based on segregation analysis it has been suggested that the etiology of approximately 15% of all CRCs might be explained by mutations in dominantly acting predisposing genes (Houlston *et al.* 1992). High-penetrant CRC syndromes with identified disease-causing genes, such as FAP and HNPCC, do not however account for all familial clustering of CRC, since they are found in not more than 5% of all cases. Among the remaining CRC families of unknown type the polyposis phenotype, like the one seen in FAP, has rarely been observed. A certain proportion of CRC families of unknown type does fulfill the Amsterdam criteria I for HNPCC classification but these families do not exhibit MSI and they do not have mutations in any of the MMR genes. A recent analysis of such groups of families has shown that they represent a distinct group with later age of onset, distal tumor localization, more adenomas and slower progression of adenomas to carcinomas, as compared to HNPCC families (Mueller-Koch *et al.* 2005). Analysis of the cancer incidence among MMR proficient CRC families fulfilling the Amsterdam criteria I also revealed that these families differ from HNPCC families, since an increase in the incidence of only CRC, although lower than in HNPCC, and no increase in the incidence of other HNPCC-associated cancers (endometrium, stomach, small bowel, bladder) was seen (Lindor *et al.* 2005). In addition, the mean age at diagnosis of CRC in MMR proficient families (60.7 years) was higher than in HNPCC families (Lindor *et al.* 2005). However, it was still significantly lower as compared to the general population strengthening the role of genetic factors in the etiology of the disease among these families. In order to make a clear distinction from HNPCC families, the term “familial colorectal cancer type X” is suggested for all Amsterdam criteria positive families without MMR deficiency (Lindor *et al.* 2005). This is

probably a heterogeneous group of families where a certain proportion of familial aggregation of CRC is expected to occur by chance or due to environmental factors. Nevertheless, some families are expected to carry mutations in as yet unidentified CRC predisposing genes.

In Sweden, around 10% of CRC patients are from families with recognized hereditary component but without identified mutations in already known CRC predisposing genes (Olsson and Lindblom 2003). Almost 2% of these families show dominant inheritance of the disease with three or more first-degree relatives affected with CRC in at least two generations (Figure 2). These families are termed hereditary colorectal cancer (HCRC) and they probably segregate dominantly acting high-risk predisposing genes. The lifetime cancer risk in HCRC families is similar to the risk observed in HNPCC, although with later onset (Lindgren *et al.* 2002). A group of low-risk families, termed two-close relatives (TCR) families, where two first-degree relatives are affected with CRC is found in additional 8.3% of Swedish CRC cases (Figure 2) (Olsson and Lindblom 2003). In this family group a higher number of adenomas is seen, as compared to both HNPCC and HCRC. However, an estimated CRC risk is only 20-30% suggesting the possible role of low to moderate penetrance alleles (Lindgren *et al.* 2002; Liljegren *et al.* 2003). In paper I families belonging to both HCRC and TCR group had been included in the genome wide linkage analysis. The lack of a single common predisposing locus provided evidence for genetic heterogeneity in the etiology of the disease among the analyzed families (Djureinovic *et al.* 2006).



**Figure 2.** The frequencies of different groups of CRC in Sweden with respect to genetic background (Olsson and Lindblom 2003). HCRC: Familial colorectal cancer; TCR: Two-close relatives; HNPCC: Hereditary nonpolyposis colorectal cancer; FAP: Familial adenomatous polyposis.

#### 1.4.8.1 *A hereditary mixed polyposis syndrome (HMPS)*

HMPS was described as a distinct CRC predisposing syndrome in Ashkenazi families. It is characterized by the development of a variety of different colonic tumors including atypical juvenile polyps, hyperplastic polyps, serrated adenomas and classical adenomas that eventually progress to carcinoma (Whitelaw *et al.* 1997). Affected individuals usually develop less than 15 polyps and the condition is inherited in a dominant manner. No extra-colonic features specific to FAP have been observed among HMPS patients. In addition, limited similarity to JPS is seen. CRC within HMPS seems to develop through a pathway in which serrated adenomas play an important role. Serrated adenomas are very heterogeneous group of lesions and no unique classification criteria have been established yet (Jass 2005). However, their most important features include morphological resemblance to hyperplastic polyps and cytological characteristics (dysplasia) of adenomas.

#### 1.4.8.2 *Hereditary breast and colorectal cancer (HBCC)*

Families with clustering of both breast and colorectal cancer have frequently been found. Many of these families may result from chance clustering since both cancer types are among the most common ones worldwide. However, some families have features suggestive of inherited disease where no mutations in any of the known breast or colorectal cancer predisposing genes have been found (Lipton *et al.* 2001).

In a search for new breast cancer predisposing genes the kinase-deficient variant in the *CHEK2* gene, *CHEK2* 1100delC, has been identified as a low-penetrance allele with an estimated twofold increased risk of developing breast cancer (Meijers-Heijboer *et al.* 2002). The same variant has been found in 18% of breast cancer families characterized by the presence of CRC cases, defining the hereditary breast and colorectal cancer phenotype – HBCC (Meijers-Heijboer *et al.* 2003). An increased frequency of *CHEK2* 1100delC has also been detected among HNPCC and HNPCC-like families with HBCC-like colorectal cancer suggesting that the variant is associated with an increase in the CRC risk, although lower than in the case of breast cancer. The data suggests that the *CHEK2* 1100delC is not a major predisposing factor for breast and colorectal cancers, but instead acts in synergy with as-yet-unidentified cancer susceptibility gene (Meijers-Heijboer *et al.* 2003). Several subsequent studies including ours (paper III) have failed to show a clear association between *CHEK2* 1100delC and increase in the

CRC risk (Kilpivaara *et al.* 2003; Lipton *et al.* 2003; de Jong *et al.* 2005). However, due to the low frequency of the variant, none of these studies was able to exclude a very low penetrance effect among CRC cases. Further studies are needed to establish the role of *CHEK2* 1100delC in colorectal cancer families with HBCC-like phenotype.

#### **1.4.9 Novel CRC loci**

##### *15q13-14*

Using linkage analysis in combination with LOH, a novel CRC locus was mapped to chromosome 15q14-22 in one Ashkenazi family affected with colorectal adenomas and carcinomas in which no *APC* mutations had been identified (Tomlinson *et al.* 1999). The locus was named *CRAC1* (colorectal adenoma and carcinoma). Interestingly, the same region of linkage was identified in three additional Ashkenazi families affected with HMPS (Jaeger *et al.* 2003). Moreover, haplotype analysis of all linked families showed that the same rare haplotype on chromosome 15q13-14 was found in all four of them, implying that the *CRAC1* gene is highly likely to be identical to the *HMPS* gene and that the shared haplotype probably originates from a common founder (Jaeger *et al.* 2003). The *HMPS/CRAC1* gene from this region has not been identified yet. Whether the *HMPS/CRAC1* locus plays a role in ethnic groups other than Ashkenazi remains to be determined.

##### *9q22.32-31.1*

In a recently performed sibling pair analysis of kindreds diagnosed with either CRC or advanced adenoma before age of 65 a novel region on chromosome 9q22.1-31.2 was identified (Wiesner *et al.* 2003). This region was among six identified regions of interest, however it provided the strongest evidence of linkage in the analysis of discordant sib pairs. Further support for the existence of a CRC susceptibility gene on chromosome 9q came after two independent analyses. In the study performed by our group evidence of linkage to the region on chromosome 9q22.32-31.1 with the multipoint LOD score of 2.4 was seen in an extended Swedish family affected with CRC and adenomas (Skoglund *et al.* 2006). The region of linkage identified in our family falls within the region identified by Wiesner *et al.*. The second confirmation of the locus on chromosome 9q was provided from the analysis of 57 CRC families from the United Kingdom where a suggestive linkage with the heterogeneity LOD (HLOD) score of 1.23 has been seen for the same region as the one found by our group,

9q22.32-31.1 (Kemp *et al.* 2006). No mutations in genes from this region have been identified so far.

### *3q21.1-26.2*

In a search for novel CRC predisposing genes we recently performed a genome-wide linkage analysis in 18 CRC families from Sweden (paper I). No common locus was identified providing evidence of genetic heterogeneity among tested families. Weak evidence of linkage was however observed to the regions on chromosomes 14q23.1–24.1, 11q13.2–13.4, 11q22.1–23.1 and 22q12.1 when analysis was performed under the assumption of locus heterogeneity. In order to further investigate these regions, a genome-wide scan was performed in additional 12 families and the data from both family groups were pooled. No confirmation to any of the suggested regions was seen. However, a novel locus on chromosome 3q21.2-26.2 was identified (Vandrovcova *et al.*, submitted). Chromosome 3 was the most consistent finding in the pooled analysis where suggestive linkage was obtained in both parametric and nonparametric analysis (HLOD score of 1.9 for  $\alpha=0.45$ ,  $\alpha$  value represents a proportion of linked families, and NPL score of 2.1). Eight families exhibiting linkage to this region were identified and subsequent finemapping performed in these families provided further evidence of linkage. One of the linked families was family 242 in which mutation screening of one candidate gene from this region, *SMARCA3*, was performed previously (paper II). However, no mutations were identified.

Interestingly, a genome-wide linkage analysis of 69 CRC pedigrees performed with a high-density single nucleotide polymorphisms (SNP) array identified the region on chromosome 3q21-24 as the most significant finding in the analysis where the affected status was restricted to cancer cases only (Kemp *et al.* 2006). The maximum HLOD score of 3.1 was obtained for  $\alpha=0.62$ . An NPL score of 3.4 was seen for the same region. The region identified in this study falls within the region identified by our group and therefore represents an independent confirmation of our finding.

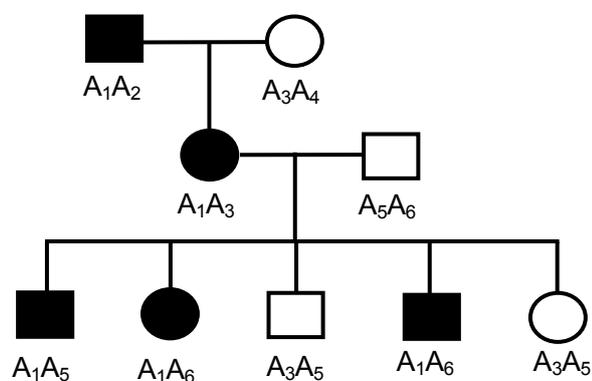
## **1.5 STRATEGIES FOR FINDING NOVEL CANCER GENES**

Ever since the elucidation of genetic basis of retinoblastoma a number of cancer predisposing genes have been identified. In most cases gene identification has been restricted to highly penetrant hereditary cancer syndromes that account for only a small

fraction of all cancer cases. As mentioned earlier, mutations in genes known to predispose to CRC have been found in only a proportion of patients with a family history of the disease. However, the genetic contribution in CRC has been estimated to be ~ 35% (Lichtenstein *et al.* 2000). Moreover, segregation analysis in CRC suggested the existence of additional dominantly acting genes (Houlston *et al.* 1992). These data suggest that a number of both high-risk and low-risk CRC predisposing genes await identification. Different approaches can be applied in a search for novel CRC predisposing genes as well as genes predisposing to other cancers. Their choice depends on several factors such as the genetic model of the disease in question and material available for the analysis. Some of the methods that can be used for gene identification are listed below.

### 1.5.1 Linkage analysis

In linkage analysis, a number of DNA markers (microsatellite markers or SNPs) of known position are tested in members of families segregating the trait and a chromosomal region that may harbor the gene responsible for the trait is determined. This is achieved by identification of DNA marker alleles that co-segregate with the trait more often than expected by random segregation (Figure 3). The closer two loci are on a chromosome the less likely they will be separated by recombination. The recombination fraction (theta -  $\theta$ ) is the probability for a recombination to occur. It ranges from 0, for loci that are completely linked, to 0.5 for unlinked loci that are far apart on the same chromosome or loci that are located on different chromosomes. Theta is used as a measure of genetic distance between two loci, with 1% of recombination corresponding to approximately 1 centimorgan (cM) of genetic distance.



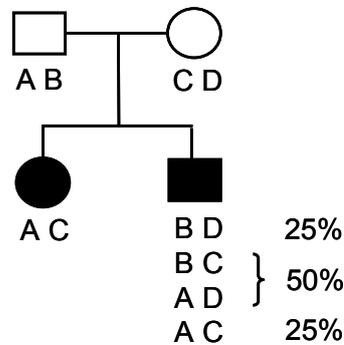
**Figure 3.** Segregation of marker alleles and the trait within a family. All affected family members have allele A<sub>1</sub> of the specific marker; therefore allele A<sub>1</sub> is co-segregating with the trait.

Classical linkage analysis is referred to as parametric or model-based analysis since a genetic model for the disease has to be provided. This includes the mode of inheritance, gene penetrance estimates and allele frequencies. The probability of linkage is given as the logarithm of the odds (LOD) score where the odds of linkage represents the ratio between two hypotheses; the alternative hypothesis that the loci are linked (recombination fraction is  $\theta$ ) and the null hypothesis that the loci are not linked (recombination fraction is 0). LOD scores are calculated for different  $\theta$  values and, if a set of families is used, they can be added up across families. Evidence of linkage is provided when the maximal LOD score exceeds a predefined threshold.

Parametric linkage analysis has been very successful in identifying a number of disease-causing genes including CRC predisposing genes *APC*, *hMLH1* and *hMSH2* (Bodmer *et al.* 1987; Lindblom *et al.* 1993; Peltomaki *et al.* 1993). It is still the method of choice for genetic mapping of simple Mendelian diseases and diseases showing near-Mendelian inheritance. One of the most important factors for successful linkage mapping is careful selection of families segregating the same phenotype. The *APC* gene was localized after selection of families with polyposis phenotype, thus excluding other types of CRC (Bodmer *et al.* 1987). Some diseases are genetically heterogeneous, that is they are caused by mutations in different genes. If families with different gene defects are analyzed together a linkage peak may not be detected. One way of avoiding the problem of genetic heterogeneity is analysis of single large families that are informative enough to be used on their own, the strategy that was used in mapping of HNPCC-causing genes *hMSH2* and *hMLH1* (Lindblom *et al.* 1993; Peltomaki *et al.* 1993).

Nowadays it is becoming more apparent that cancers, particularly the common ones, have to be looked at as a complex traits influenced by a number of different genes, environmental factors and their interaction. Because of difficulties in providing the correct genetic model, parametric linkage analysis has very little power, if any, in mapping of such traits. Instead, non-parametric (NPL) or model-free linkage analysis, in which no complete specification about the genetic model of the disease has to be given, is the method of choice. The NPL analysis was originally developed for sibling pairs but latter was also extended to general pedigrees. In many cases sib-pair method concentrates on affected sib-pairs and it is based on the idea that two affected siblings will share identical-by-descent (IBD) alleles near the disease gene more often than

expected under the hypothesis of random segregation (Figure 4). Alleles are said to be IBD if they originate from the same ancestor. The affected-relative-pair method includes relatives other than sib pairs and is based on the identity by state (IBS) relationship. Two genes are IBS if they are the same alleles regardless of their origin.



**Figure 4.** Affected sib-pairs with IBD sharing. According to Mendelian laws of inheritance the probability that sibs will share 1 IBD allele is 50% while probability of sharing 2 or 0 IBD alleles is 25%.

### 1.5.2 Allelic association

The proposed model of polygenic susceptibility to breast cancer suggests that a substantial proportion of cases are due to a large number of alleles acting additively or multiplicatively, each of them however conferring only a small genotypic risk (Pharoah *et al.* 2002). The same model is expected to be applicable to other types of cancers including CRC. Due to their small effect, low-penetrance alleles have very little chance to be detected by linkage analysis. Allelic association has been shown to be a powerful tool in identifying genes behind complex traits. Allelic association is present if the frequency of the occurrence of a specific marker allele together with a disease trait deviates from the expected occurrence based on random segregation. An association seen between a marker allele and a disease trait can result either from linkage disequilibrium with a closely located susceptibility gene or from a direct biological effect of the tested marker. Two types of association studies can be performed; population-based and family-based studies. Population-based association studies are designed as case-control studies in which genotype frequencies are determined in a set of unrelated affected individuals and compared to frequencies seen in matched controls. One of the major factors influencing the result of case-control studies is the sample size and it depends on frequency and estimated risk of tested alleles. Selection of adequate controls also plays an important role. If the tested population contains several genetically distinct population subsets (population stratification) a spurious associations

may be seen. The problem caused by stratification can be overcome by performing family-based association studies. Methods like the transmission disequilibrium test (TDT) compare the frequency of transmission of one allele over the other from heterozygous parents. No control samples are needed, instead the untransmitted allele serves as an internal control. Population-based association studies are however rarely used in cancer as it is often difficult to recruit parents of affected cases.

### **1.5.3 Somatic alterations in tumors**

Constitutional chromosomal changes are rare in solid tumors, however once identified these changes can pinpoint the region harboring a cancer susceptibility gene, as was done in the localization of the *APC* gene (Bodmer *et al.* 1987). In contrast, somatic chromosomal alterations are frequently detected in tumors. Identification of chromosomal regions that are consistently deleted in tumor cells compared to normal cells may suggest that tumor suppressor genes are located within these regions. Such findings were often used as a complementation to the findings from linkage analysis. The detection of frequently deleted chromosomal regions may also represent the starting point in the process of gene identification, this strategy was successful in identifying the gene causing the Peutz-Jegher syndrome (Hemminki *et al.* 1997). In addition to deleted regions, identification of frequently amplified regions may also be of importance.

Not just chromosomal alterations but also subtle changes detected at nucleotide level may assist in identification of disease-causing gene. Observation of an excess of G:C > T:A transversion in the *APC* gene detected in tumor DNA from patients affected with multiple adenomas and cancer prompted the authors to speculate whether the BER deficiency was associated with the disease. In a subsequent mutation screening of BER genes inactivating mutations in one member, *MUTYH* gene, were identified (Al-Tassan *et al.* 2002).

## 2 AIMS

The general aim of this thesis was to identify novel genetic factors with the role in colorectal cancer predisposition.

The specific aims were as follows:

1. To identify potential loci harboring novel CRC predisposing genes.
2. To determine whether the *SMARCA3* gene plays a role in the etiology of CRC as well as gastric cancer.
3. To determine the frequency of the *CHEK2* 1100delC variant among selected and unselected CRC patients from Sweden.
4. To determine the contribution of the *MUTYH* gene to the predisposition to CRC among Swedish familial and sporadic CRC patients.

## 3 MATERIAL AND METHODS

### 3.1 PATIENT MATERIAL

Families with a suggested predisposition to CRC used in papers I-IV were recruited from the cohort of patients undergoing counseling at the Cancer Family Clinic, Karolinska Hospital. Family history including clinicopathological details was obtained by interview and confirmed by medical records, pathological reports or death certificates. None of these families had classical or attenuated FAP, as determined by colonoscopic examination. HNPCC was excluded using the current clinical screening protocol consisting on MSI test, immunohistochemistry of MMR genes on selected tumors and mutation screening of the *hMLH1*, *hMSH2*, *hMSH6* or *hPMS2* genes performed in families fulfilling the Amsterdam criteria, in cases with MSI positive tumors or in families with MSI negative tumors but with a family member diagnosed with CRC before the age of 50 (Wahlberg *et al.* 1999; Liu *et al.* 2000). All family members were offered regular colonoscopy surveillance. Based on the family history of the disease, patients were classified into two groups: HCRC and TCR.

A cohort of unselected CRC cases was recruited from 13 hospitals from Stockholm-Gotland and Uppsala-Örebro health-care regions. From this cohort 665 cases were used in paper III to determine the prevalence of the *CHEK2* 1100delC variant. In addition, 95 patients with a family history of gastric cancer were selected from the cohort in order to test the prevalence of the identified *SMARCA3* variants (paper II).

A panel of 450 sporadic CRC cases recruited from the Department of Surgery at Uppsala University Hospital and the Department of Pathology at Linköping University Hospital were used in paper IV for testing the frequency of the two most common mutations in the *MUTYH* gene found in Caucasians.

Control samples used in paper II-IV were subjects randomly chosen from a cohort of anonymous blood donors from Stockholm region. The exact number of control samples used in different studies differs depending on the experimental design and PCR amplification status.

### **3.2 LINKAGE ANALYSIS**

The principles of linkage analysis are discussed in the Introduction. In paper I a genome-wide linkage analysis in 18 non-FAP/non-HNPCC colorectal cancer families from Sweden was performed using the ABI Prism Linkage Mapping Set version 2.5 (Applied Biosystems, Foster city, CA, USA) which consists of 400 fluorescently labeled microsatellite markers distributed across the genome. Genotyping was performed according to manufacturer's instruction. Multipoint linkage analyses were undertaken using parametric as well as nonparametric methods. In addition, haplotypes were generated for all families. Both linkage analyses and the generation of haplotypes were carried out using the SimWalk2 program version 2.83 (Sobel and Lange 1996). Data generated for the chromosome X were analyzed with the FASTLINK program (Cottingham *et al.* 1993). In the genome-wide linkage analysis two affected status criteria, stringent and less stringent, were used. Finemapping of the regions on chromosomes 11 and 14 was performed using an additional eleven markers on chromosome 11 and 19 markers on chromosome 14, selected for their position and heterozygosity.

### **3.3 DIRECT SEQUENCING**

Direct sequencing used today is a modification of an enzymatic sequencing method developed by Sanger *et al.* (Sanger *et al.* 1977). The reaction is based on the use of 2',3'-dideoxynucleotides (ddNTPs) which lack a hydroxyl group at the 3' position thus serving as a base-specific chain terminator. As four ddNTPs (A, G, C and T) are tagged with four different fluorescent dyes all four base-specific reactions can occur in the same tube. The DNA template used in the reaction is usually provided in the form of PCR product. Synthesis of a complementary strand is performed in the presence of labeled ddNTPs which are in competition with an excess of dNTP molecules and the reaction will proceed until one of the ddNTPs is incorporated instead of dNTP. Chain termination will occur randomly at one of the many possible choices resulting in a collection of labeled DNA fragments with a common 5' end but a variable 3' end. Fragments of different length are size fractionated either by gel electrophoresis or in capillaries and the emission spectra of each dye is recorded and registered as a chromatogram enabling direct reading of the sequences.

Mutation screening of the *SMARCA3* gene, presented in the paper II, was performed using direct sequencing. Primers used for amplification of all exons including exon/intron boundaries, together with the 5' and 3' untranslated regions and putative promoter sequence, were designed using online available Primer3 software package. Reactions were performed using the ABI Big Dye Terminator v3.1 kit and fragments were separated on an ABI 3730 XL capillary sequencer (Applied Biosystems, Forester City, California). In paper III, all fragments of the *MUTYH* gene exhibiting abnormal DHLPC profiles were reamplified and purified PCR products were sequenced using ABI Big Dye Terminator v3.0 kit. Sequencing reaction products were analyzed using ABI 377 automated sequencer (Applied Biosystems, Forester City, California).

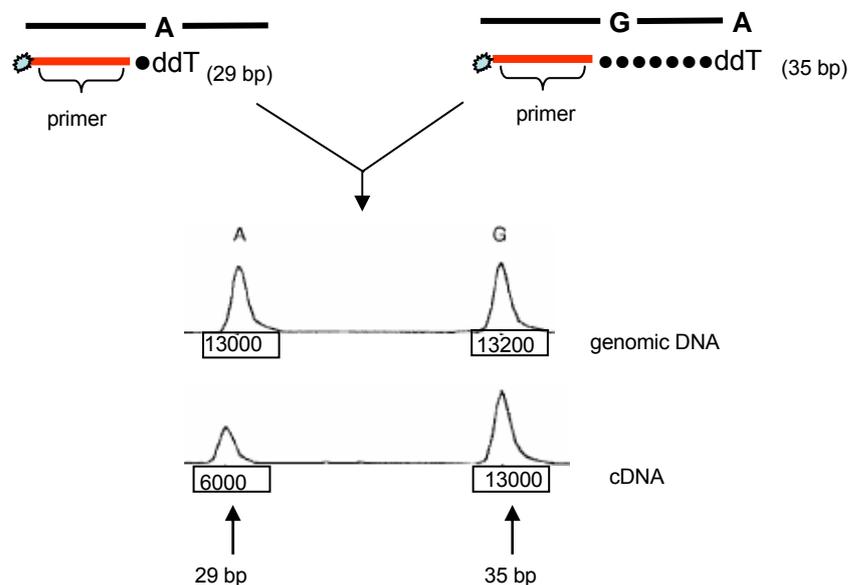
### **3.4 REVERSE-TRANSCRIPTASE PCR (RT-PCR)**

RT-PCR is used to study the gene expression within the tissue of interest, however it is also used as one of the mutation screening methods since it allows detection of large insertions or deletions (such as those involving the whole exons) and mutations affecting splicing, all of which are usually missed by standard mutation screening approaches. The first step is conversion of the total RNA or mRNA isolated from a certain tissue into single-stranded complementary DNA (cDNA) using reverse transcriptase. The cDNA synthesis can be initiated in at least two ways; either with hexamer primers constructed to bind randomly over the whole RNA molecule or with oligo (dT) primer binding to the poly A tail of the mRNA molecule. In the second step, cDNA is used as a template in the PCR reaction which is performed with exon-specific primers designed to amplify the gene of interest.

In paper II cDNA was prepared from total RNA isolated from EBV transformed lymphocytes using the random hexamer priming method. The coding region of the *SMARCA3* gene was amplified in two fragments with primers designed in such a way that at least one of them was covering two exons, in order to reduce genomic DNA contamination. Products of the reactions were detected on an agarose gel along with control samples.

### 3.5 SINGLE NUCLEOTIDE PRIMER EXTENSION (SNUPE) ANALYSIS

SNuPE analysis of the mRNA is a relatively simple method for testing the difference in the level of expression between two alleles of a gene of interest. SNPs are used as markers for allele discrimination, therefore the method depends on constitutional heterozygosity of the selected polymorphism. Fragments of genomic DNA and cDNA harboring the selected SNP are amplified in a PCR reaction and used as a template for subsequent primer extension (Figure 5). The extension reaction is performed with a fluorescently labeled primer designed to anneal close to the polymorphic site. In addition, a combination of three dNTPs with one ddNTP, selected to be complementary to one of the two possible bases at the polymorphic site, is used. This implies that primer extension of one allele will be terminated after reaching the polymorphic site while extension of the other allele will continue until a nucleotide complementary to ddNTP is encountered. Thus, the reaction results in the production of two fragments of different size that can easily be separated on denaturing gels. As these products are fluorescently labeled automated sequencers can be used for their detection and the calculated allelic ratios will indicate whether two alleles are differentially expressed.



**Figure 5.** Schematic view of SNUPE protocol. Difference in the expression of two alleles can be estimated based on the ratio of peak areas.

SNuPE analyses on RT-PCR products and PCR products from genomic DNA of the same individual, a member of family 242, was used to test the allelic difference in the expression of the *SMARCA3* gene. Both products were analyzed on an ABI 377 automated sequencer using 10% denaturing polyacrilamide gels and allelic ratio was

determined using the Genescan3.1<sup>®</sup> software program (Applied Biosystems, Forester City, California).

### **3.6 SOUTHERN BLOT**

Southern blot is used as a tool to localize a particular DNA sequence and, as one of a mutation screening methods, for detection of aberrations involving large parts of chromosomes. The target DNA is digested with one or more restriction endonucleases and resulting fragments are size-fractionated by agarose gel electrophoresis. The DNA is thereafter denatured and transferred from the gel to a nitrocellulose or a nylon membrane. Fragments of interest are identified by hybridization with a radiolabeled DNA or RNA probe and the signal is visualized by using a photographic film or a phosphoimager.

Southern blot was used in testing for large chromosomal aberrations involving the *SMARCA3* gene (paper II). DNA from one member of family 242 was digested with BglII, transferred onto nylon membrane and hybridised with two probes labelled with  $\alpha^{32}\text{P}$ -dCTP. The signal was detected using a phosphoimager.

### **3.7 ALLELE-SPECIFIC PCR AMPLIFICATION**

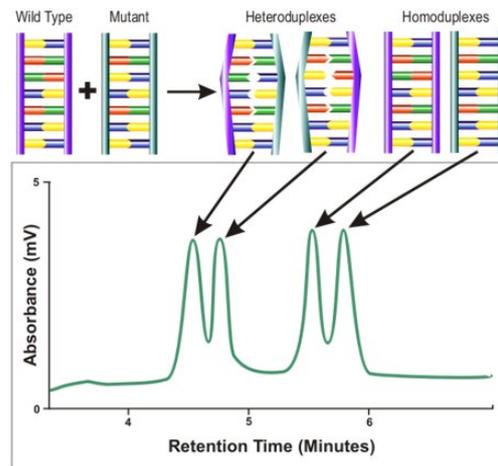
Allele-specific PCR is used for direct analysis of known mutations such as single base substitutions, small insertions or deletions. It is based on selective amplification of one allele over another, which is achieved by designing one of the primers to match the desired allele but to mismatch the other allele. The mismatch is usually located at the terminal 3'-nucleotide of a primer; however, in some instances when the melting temperature ( $T_m$ ) of the 3' end of a primer is high, additional mismatches are needed for improved discrimination of alleles. In addition, a third primer, common for both alleles, is also used. Every sample is tested in two reactions, one performed with a wild type-specific primer and the other performed with a mutation-specific primer. Products of reactions are simply visualized on an agarose gel.

Allele-specific PCR was employed for testing the prevalence of the *CHEK2* 1100delC variant in a set of selected and unselected CRC patients and controls (paper III). In addition to *CHEK2* primers, primers specific for the *SLC30A9* gene were used as an internal control of each reaction.

### **3.8 DENATURING HIGH PERFORMANCE LIQUID CHROMATOGRAPHY (DHPLC)**

DHPLC is an automated method for mutation screening of candidate genes. Single base substitutions and small insertions or deletions are detected by temperature modulated heteroduplex analysis on a reversed-phase column. Mutation detection is based on the different elution profiles between heteroduplex DNA molecules and perfectly matched homoduplexes under conditions of partial thermal denaturation. The first step in the procedure is PCR amplification of genomic DNA templates followed by final denaturation and renaturation which enables formation of heteroduplex DNA molecules. In the second step, DNA fragments are fractionated on a liquid chromatography column. Hydrophobic portion of triethylamine acetate (TEAA) interacts with hydrophobic beads in the column while negatively charged phosphate groups of partially denatured DNA fragments are attracted to a positively charged ammonium groups of TEAA. At increasing concentrations of Acetonitril (ACN) the DNA/TEAA attraction is reduced resulting in the elution of DNA fragments, first heteroduplex molecules are eluted followed by homoduplexes (Figure 6). As fragments pass through the UV detector the absorbance is measured and recorded as a chromatogram. DHPLC can detect the presence of all combinations of nucleotide substitutions with sensitivity close to 100%, if analyses are performed at multiple temperatures. However, the location and the exact type of change have to be determined by sequencing.

In paper IV mutation screening of the *MUTYH* gene was performed by DHPLC using Transgenomic Wave DNA Fragment Analysis System equipped with a DNASep column (Transgenomics, Crewe, United Kingdom). The denatured and reannealed PCR products were loaded onto the instrument directly without purification. The optimal conditions such as running column temperature and concentrations of TEAA and ACN were determined for each amplicon using the WAVEMAKER 3.4 software provided with the instrument.



**Figure 6.** Schematic representation of DHPLC ([www.transgenomics.com](http://www.transgenomics.com))

### 3.9 STATISTICAL ANALYSIS

The difference in the variant frequency between cases and controls was measured using the Fishers exact test in paper III and the Chi-square test in paper IV. P value <0.05 was considered to be statistically significant.

## 4 RESULTS AND DISCUSSION

### 4.1 PAPER I

#### **A genome-wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary nonpolyposis colorectal cancer**

Colorectal cancer is one of the most common types of malignancies worldwide with the estimated lifetime risk in the general population of 5%. A number of syndromes with a Mendelian or near-Mendelian inheritance of the disease, like FAP, HNPCC and MAP, have been described, however these syndromes are identified in only a small proportion of cases with a family history of the disease. In order to identify novel CRC predisposing genes a genome-wide linkage analysis was performed in 18 CRC families from Sweden using two affected status criteria, stringent and loose.

No common susceptibility locus was identified in either nonparametric or parametric analyses performed under the assumption of locus homogeneity using either of the two affected status criteria. Several loci were however identified when analyses were performed under the assumption of locus heterogeneity. Multipoint parametric linkage analysis performed with the stringent affected status criteria provided weak evidence of linkage to chromosome 22q12 with the HLOD score of 1.26 and  $\alpha=0.60$ . When using the less stringent criteria two chromosomes were suggested to be of interest in both parametric and nonparametric linkage analysis. The most significant result in this study was obtained for the region on chromosome 14q24 where the HLOD score of 2.61 for  $\alpha=0.25$  was obtained. The NPL score seen for this region was 2.88 ( $p=0.001$ ). In addition, two regions on chromosome 11 with evidence of linkage were identified. Marker D11S908 was suggested to be of interest in both parametric and nonparametric analysis (the HLOD score was 2.10 and the NPL score was 2.16) while parametric analysis provided evidence of linkage for D11S1314 (HLOD score of 1.96).

As chromosomes 11 and 14 showed the strongest evidence of linkage in this study additional markers within these regions were selected for finemapping. Subsequent analyses performed with new markers failed to reach the level obtained in the first screen. Although both the HLOD scores and the NPL scores were reduced across these regions they were still within the range of suggestive linkage. Six families were identified to be linked to chromosome 14 and haplotype analysis of these families

revealed an overlapping region between markers D14S1038 and D14S1069 (14q23.1-24.1). In addition, six families exhibited linkage to chromosome 11 where two overlapping regions were identified after haplotype analysis, one between D11S987 and D11S4207 (11q13.2-13.4) and the other between D11S4120 and D11S4090 (11q22.1-23.1).

In this study no evidence of linkage was seen to the recently identified regions on chromosome 15 and chromosome 9 (Jaeger *et al.* 2003; Wiesner *et al.* 2003; Skoglund *et al.* 2006). Moreover, weak evidence of linkage to chromosomes 11 and 14 seen in only a proportion of analyzed families supports the view that the genetic determinants of familial CRC are heterogeneous.

## **4.2 PAPER II**

### **Mutation screening of the *SMARCA3* gene in Swedish colorectal cancer patients**

The SWI/SNF multiprotein complex is involved in chromatin remodelling in an ATP dependent manner and is implicated in a variety of important cellular functions. Mutations in several of the SWI/SNF family genes (*RAD54B*, *hSNF/INI1*, *BRG1*) have been found in different tumors and tumor cell lines suggesting a potential tumor suppressor activity of this gene family (Hiramoto *et al.* 1999; Wong *et al.* 2000; Yuge *et al.* 2000; Schmitz *et al.* 2001). In addition, promoter hypermethylation of the *SMARCA3* gene, another member of the SWI/SNF family, has been reported in colon and gastric cancer (Moinova *et al.* 2002; Hamai *et al.* 2003; Hibi *et al.* 2003; Leung *et al.* 2003). In the genome-wide linkage analysis (paper I) a family presented with both colorectal and gastric cancer (family 242) exhibited suggestive linkage to the region on chromosome 3q harbouring *SMARCA3*. In order to investigate whether this gene plays a role in the predisposition to colorectal and gastric cancer mutation screening was performed in family 242 as well as 20 additional colorectal and/or gastric cancer families.

Altogether 15 germline sequence variants were identified in the *SMARCA3* gene of which seven variants were previously reported polymorphisms while eight variants represented novel sequence changes. None of the identified variants were clearly pathogenic. Heterozygote frequency of the eight novel variants was determined by

testing control individuals and for seven of these variants a slightly increased frequency was seen in patients. Their significance was further determined in 287 unrelated patients with a family history of colorectal cancer of whom 95 also had a family history of gastric cancer. A small difference in the frequency between patients and controls, although not statistically significant, was detected only for the c.2440C>T variant. Subsequent segregation analysis performed in three families with available additional material failed to show clear co-segregation of the variant and the disease. Two identified promoter variants, -454G>A and -291G>A, were found at low frequency among tested patients however neither of these variants was detected in controls. In family 242, besides germline mutation screening, RNA analysis and Southern blot analysis were carried out but failed to identify aberrations of *SMARCA3* associated with inherited predisposition to colorectal and gastric cancer.

Although a very low or modifying effect of variants like c.2440C>T, -454G>A and -291G>A could not be excluded, it seems that *SMARCA3* is not a common target for mutations, at least not in colorectal and gastric cancer. These data taken together with the frequent observation of gene silencing caused by promoter hypermethylation may indicate a more important role of *SMARCA3* in tumor progression rather than tumor initiation. The region of linkage on chromosome 3q22.1-26.31 identified in family 242 is however still of interest. The same region was identified as the most significant finding in the pooled analysis of 30 CRC families, of which 18 were from the first screen we have performed in paper I (Vandrovicova *et al.*). In addition, an independent confirmation of this region was obtained in the genome-wide linkage analysis of 69 CRC pedigrees performed with a high-density SNP array (Kemp *et al.* 2006).

### **4.3 PAPER III**

#### **The *CHEK2* 1100delC variant in Swedish colorectal cancer**

The *CHEK2* 1100delC variant has been suggested to act as a low penetrance breast cancer susceptibility allele in families without *BRCA1/2* mutations (Meijers-Heijboer *et al.* 2002). Moreover, a high frequency of this variant has been found among families with hereditary breast and colorectal cancer (HBCC families) suggesting that the variant could also confer a colorectal cancer risk (Meijers-Heijboer *et al.* 2003). Interestingly, the region on chromosome 22 harboring the *CHEK2* gene was suggested

to be of interest in our previously performed genome-wide linkage analysis (Djureinovic *et al.* 2006). To determine whether *CHEK2* 1100delC plays a role in CRC predisposition, the variant frequency was determined in selected familial cases, unselected cases, in 18 families previously analyzed in the linkage analysis and in controls.

No significant difference in the frequency of *CHEK2* 1100delC was found between CRC cases and controls. The variant frequency among selected familial cases, unselected cases and controls was 1.15%, 0.93% and 0.66%, respectively. One out of 45 familial cases (2.22%) with a family history of breast cancer was found to be a carrier of the variant, however, as the sample size was too small, the difference was not significant. None of the 18 families from linkage analysis was found to carry *CHEK2* 1100delC thus excluding the effect of the variant among these families. The significance of the region on chromosome 22 remains to be determined in future studies.

In addition to our study, several other studies failed to show increased frequency of *CHEK2* 1100delC among CRC cases as well as patients affected with multiple colorectal adenomas (Kilpivaara *et al.* 2003; Lipton *et al.* 2003; de Jong *et al.* 2005). The variant was also excluded as a major cause of double primary breast and colorectal cancer in females from southern Sweden (Isinger *et al.* 2006). However, due to the low frequency of *CHEK2* 1100delC, none of these studies including ours was able to exclude a very low penetrance effect of this variant in CRC predisposition.

#### **4.4 PAPER IV**

##### **Germline mutations in the *MYH* gene in Swedish familial and sporadic colorectal cancer**

The BER pathway is involved in the repair of mutations caused by oxidative DNA damage, like 8-oxo-guanine products (8-oxoG). 8-oxoG mispairs with adenine residues leading to an excess of GC>TA transversions. Recently bi-allelic germline mutations in the *MYH* gene, member of the BER system, have been identified in families with multiple colorectal adenomas and cancer without germline mutations in the *APC* gene (Al-Tassan *et al.* 2002; Sieber *et al.* 2003). In order to determine the role of the *MYH* gene in familial colorectal cancer associated with a fewer number of polyps germline

mutation screening of this gene was performed in 84 non-FAP/non-HNPCC colorectal cancer patients.

No obviously pathogenic monoallelic or biallelic germline *MYH* mutations were found among 84 unrelated patients with a family history of CRC, including the two most common pathogenic variants, Y165C and G382D, found in Caucasians (Al-Tassan *et al.* 2002; Sampson *et al.* 2003; Sieber *et al.* 2003). One silent and three missense variants, all previously described as polymorphisms, in addition to six intronic variant were identified. These data suggest a very limited role, if any, of the *MYH* gene in CRC predisposition among Swedish families with a low polyp number. It is however possible that families with *MYH* mutations were missed since the prevalence of MYH-associated colorectal cancers is very low and, due to recessive inheritance of the disease, some patients can be wrongly classified as sporadic cases.

In addition to patients with a family history of the disease, the prevalence of the two most common pathogenic variants, Y165C and G382D, were determined in 450 sporadic CRC cases from Sweden and 480 controls. Of the samples tested none were found to be homozygote or compound heterozygote carrier of the two variants. The heterozygous frequency of Y165C was 0.91% in cases and 0.43% in controls, while the heterozygous frequency of G382D was 0.45% and 0.21% in cases and controls, respectively. While genotypic distribution of both variants was found to be in Hardy-Weinberg equilibrium among control samples, their distribution was found to deviate from equilibrium among cases. Both Y165C and G382D were overrepresented in cases by approximately 2-fold as compared to controls, although the difference was not statistically significant. Thus, our data together with recently published studies suggest that even heterozygous *MYH* mutation carriers are predisposed to CRC (Croitoru *et al.* 2004; Farrington *et al.* 2005). In addition to two common pathogenic variants, three novel heterozygous sequence changes affecting the same amino acid position, R423Q, R423P and R423R, were detected among cases and not in controls. The significance of these variants however has to be determined in future studies.

## 5 CONCLUSIONS

The aim of this thesis was to identify novel genetic factors involved in CRC predisposition. Based on our findings the following conclusions can be made:

Genome-wide linkage analysis in 18 non-FAP/non-HNPCC families from Sweden failed to identify single common locus harboring a CRC predisposing gene. Three chromosomes were however suggested to be of interest when analysis was performed under the assumption of locus heterogeneity: chromosomes 22q, 11q and 14q. In addition, no evidence of linkage was detected to the recently identified regions on chromosomes 9q and 15q. Thus, our data provide further evidence of genetic heterogeneity in familial CRC of unknown type.

The lack of obviously pathogenic germline mutations in the *SMARCA3* gene suggests that the gene is not likely to be associated with an increased predisposition to CRC and gastric cancer. However, some variants in the *SMARCA3* gene could still play a role as modifying or a low-penetrance alleles acting in a multiplicative or additive manner.

We found no evidence for an association of the *CHEK2* 1100delC variant and the colorectal cancer risk. However, as the variant frequency in the Swedish population is very low, our study was not able to exclude a very low effect of *CHEK2* 1100delC.

Our data suggest that bi-allelic mutations in the *MUTYH* gene are not likely to account for familial CRC with low number of polyps. The two most common pathogenic *MUTYH* variants are detected in Swedish population indicating that mutations in the *MUTYH* gene could account for a small proportion of CRC cases in Sweden. In addition, evidence for a slightly increased CRC risk among heterozygote *MUTYH* mutation carriers was detected.

The analysis of the CRC of unknown type, performed by ours and other groups, clearly suggests that disease is genetically heterogeneous. Several novel loci have been recently identified and their significance, as well as the disease-causing genes residing within these regions, remains to be determined. A high-penetrance genes acting in a dominant manner may be identified, however their contribution to the overall CRC

incidence will probably be small. In contrast, a significant proportion of CRCs are expected to arise due to the action of a number of low-penetrance alleles acting additively or multiplicatively. Identification of such alleles is still difficult, but emerging technologies and human genome project mean that the identification of these variants will become easier.

## 6 ACKNOWLEDGEMENTS

This work has been carried out at the Department of Molecular Medicine and Surgery, Karolinska Institute. I would like to thank to all of you who have helped me in any way during my PhD years:

**Professor Annika Lindblom**, my supervisor, thank you for accepting me into your group and making this work possible, for being so enthusiastic and always having time for me and for all nice dinners, BBQs and picnics.

**Lennart Iselius**, my co-supervisor, for valuable discussions and comments.

**Professor Magnus Nordenskjöld** for providing a nice research environment and creating a positive atmosphere at CMM.

My dear lab-mates: **Barbro**, for always taking care of everything and making our lab-life easy. **Xiao Lei**, for being not just a great colleague but also a dear friend. Thanks for letting me bother you even now. **Paula**, what else can I say but thanks for everything you have done, for all the laughs and fun times we had, for being there for me no matter what, for listening and understanding and for all the emergency calls (and chats) in the time of crisis. Can't wait to see you! **Johanna S**, my "project companion", thanks for the nice time we had, both here and during our "colon trips", and for beautifying my manuscripts. Good luck with your thesis. **Jana**, thanks for accepting to share "the colon experience" with Johanna and me, I wish you all the best in the future. **Antonia**, for being a great student and for all the work you have done. **Barbara**, for being a nice friend and always willing to help. **Johanna R, Bu, Simone, Moraima, Annelie and Sara**, thanks for good company in the lab and for all the fun we had.

All the people at CMM: **professor Anna Wedel, professor Agneta Nordenskjöld, Michela, Virpi, Tiina, Svetlana, Micke, Biying, Keng-Ling, Fredrik, Anna S, Ana B, Hanh, Cilla, Clara, Sofia, Husameldin, Kazuko, Lina, Emma, Anya, Feng Quing, Katja, Zheng, Jacqueline, Elisabeth, Eva, Britt-Marie, Erik, Helena, Theo and Jan**, many thanks for the nice time over the years, lots of fun, delicious breakfasts and seminars.

**Dr. Günther Weber** thanks for taking time to answer all my questions and for interesting journal clubs.

**Dr. Fabio Sanchez**, for stimulating talks, general support and for interesting discussions.

**Anki, Christina, Anna-Lena, Sigrid and Catarina**, thank you for all the practical help but also for nice chats about everything.

**Yvonne Cowan, Kerstin Florell, Christina Bremer, Gunilla Risberg, Delphi Post and Majalena Granqvist** for all the help with administrative matters.

**Lennart Helleday** for always finding time to solve my computer problems.

**Gummi**, for becoming a dear friend, for always being so positive and ready to have fun. We are looking forward seeing you again.

**Dr. Sonja Pavlovic**, my mentor and friend, thank you for being so enthusiastic and caring, **Jelena U** and **Jelena P** (sorry, J), I will always remember the time we spent in the lab.

**Sandra, Dragan, Emma, Slavena, Ljubica, Nenad** for being great friends and bringing the "home spirit" into our life.

**Čika Bato and teta Vida**, thank you for all the Christmas and Eastern dinners. You always make us feel like at home.

**Dušica** and **Dragan**, our dear friends, thanks for constant on-line soul relief and for providing an all-inclusive pit-stop during our vacations.

**My "Swedish family" from Lidköping**, it meant a lot being close to you.

Mojoj dragoj **Stani** što još uvijek brine o meni.

My wonderful family: **mama i tata, Saša, Biljana** and **Dragan** (the best sambo one can possibly have), thank you for everything you've done (and still doing) for me, for letting me choose my own way, for all your support and encouragement, but must of all for your love.

## 7 REFERENCES

- Aarnio M, Sankila R, Pukkala E, et al. (1999). "Cancer risk in mutation carriers of DNA-mismatch-repair genes." **Int J Cancer** 81(2): 214-218.
- Al-Tassan N, Chmiel NH, Maynard J, et al. (2002). "Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors." **Nat Genet** 30(2): 227-232.
- Baker SJ, Fearon ER, Nigro JM, et al. (1989). "Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas." **Science** 244(4901): 217-221.
- Bhattacharyya NP, Skandalis A, Ganesh A, et al. (1994). "Mutator phenotypes in human colorectal carcinoma cell lines." **Proc Natl Acad Sci U S A** 91(14): 6319-6323.
- Boardman LA, Thibodeau SN, Schaid DJ, et al. (1998). "Increased risk for cancer in patients with the Peutz-Jeghers syndrome." **Ann Intern Med** 128(11): 896-899.
- Bodmer WF, Bailey CJ, Bodmer J, et al. (1987). "Localization of the gene for familial adenomatous polyposis on chromosome 5." **Nature** 328(6131): 614-616.
- Bos JL, Fearon ER, Hamilton SR, et al. (1987). "Prevalence of ras gene mutations in human colorectal cancers." **Nature** 327(6120): 293-297.
- Boyle P and Langman JS (2000). "ABC of colorectal cancer: Epidemiology." **Bmj** 321(7264): 805-808.
- Chan TL, Yuen ST, Kong CK, et al. (2006). "Heritable germline epimutation of MSH2 in a family with hereditary nonpolyposis colorectal cancer." **Nat Genet**.
- Chow E, Thirlwell C, Macrae F and Lipton L (2004). "Colorectal cancer and inherited mutations in base-excision repair." **Lancet Oncol** 5(10): 600-606.
- Cleaver JE (2005). "Cancer in xeroderma pigmentosum and related disorders of DNA repair." **Nat Rev Cancer** 5(7): 564-573.
- Cottingham RW, Jr., Idury RM and Schaffer AA (1993). "Faster sequential genetic linkage computations." **Am J Hum Genet** 53(1): 252-263.
- Croitoru ME, Cleary SP, Di Nicola N, et al. (2004). "Association between biallelic and monoallelic germline MYH gene mutations and colorectal cancer risk." **J Natl Cancer Inst** 96(21): 1631-1634.
- Cruz-Correa M, Cui H, Giardiello FM, et al. (2004). "Loss of imprinting of insulin growth factor II gene: a potential heritable biomarker for colon neoplasia predisposition." **Gastroenterology** 126(4): 964-970.
- Cui H, Horon IL, Ohlsson R, et al. (1998). "Loss of imprinting in normal tissue of colorectal cancer patients with microsatellite instability." **Nat Med** 4(11): 1276-1280.
- de Jong MM, Nolte IM, Te Meerman GJ, et al. (2005). "Colorectal cancer and the CHEK2 1100delC mutation." **Genes Chromosomes Cancer** 43(4): 377-382.
- de la Chapelle A (2004). "Genetic predisposition to colorectal cancer." **Nat Rev Cancer** 4(10): 769-780.

- Djureinovic T, Skoglund J, Vandrovцова J, et al. (2006). "A genome wide linkage analysis in Swedish families with hereditary non-familial adenomatous polyposis/non-hereditary non-polyposis colorectal cancer." **Gut** 55(3): 362-366.
- Duval A, Gayet J, Zhou XP, et al. (1999). "Frequent frameshift mutations of the TCF-4 gene in colorectal cancers with microsatellite instability." **Cancer Res** 59(17): 4213-4215.
- Eng C and Peacocke M (1998). "PTEN and inherited hamartoma-cancer syndromes." **Nat Genet** 19(3): 223.
- Enholm S, Hienonen T, Suomalainen A, et al. (2003). "Proportion and phenotype of MYH-associated colorectal neoplasia in a population-based series of Finnish colorectal cancer patients." **Am J Pathol** 163(3): 827-832.
- Falls JG, Pulford DJ, Wylie AA and Jirtle RL (1999). "Genomic imprinting: implications for human disease." **Am J Pathol** 154(3): 635-647.
- Fan T, Yan Q, Huang J, et al. (2003). "Lsh-deficient murine embryonal fibroblasts show reduced proliferation with signs of abnormal mitosis." **Cancer Res** 63(15): 4677-4683.
- Farrington SM, Tenesa A, Barnetson R, et al. (2005). "Germline susceptibility to colorectal cancer due to base-excision repair gene defects." **Am J Hum Genet** 77(1): 112-119.
- Fearon ER, Cho KR, Nigro JM, et al. (1990). "Identification of a chromosome 18q gene that is altered in colorectal cancers." **Science** 247(4938): 49-56.
- Fearon ER, Hamilton SR and Vogelstein B (1987). "Clonal analysis of human colorectal tumors." **Science** 238(4824): 193-197.
- Fearon ER and Vogelstein B (1990). "A genetic model for colorectal tumorigenesis." **Cell** 61(5): 759-767.
- Feinberg AP, Ohlsson R and Henikoff S (2006). "The epigenetic progenitor origin of human cancer." **Nat Rev Genet** 7(1): 21-33.
- Feinberg AP and Vogelstein B (1983). "Hypomethylation distinguishes genes of some human cancers from their normal counterparts." **Nature** 301(5895): 89-92.
- Feinberg AP and Vogelstein B (1983). "Hypomethylation of ras oncogenes in primary human cancers." **Biochem Biophys Res Commun** 111(1): 47-54.
- Fodde R, Kuipers J, Rosenberg C, et al. (2001). "Mutations in the APC tumour suppressor gene cause chromosomal instability." **Nat Cell Biol** 3(4): 433-438.
- Fodde R, Smits R and Clevers H (2001). "APC, signal transduction and genetic instability in colorectal cancer." **Nat Rev Cancer** 1(1): 55-67.
- Friend SH, Bernards R, Rogelj S, et al. (1986). "A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma." **Nature** 323(6089): 643-646.
- Fusaro RM, Lemon SJ and Lynch HT (1996). "Muir-Torre syndrome and defective DNA mismatch repair genes." **J Am Acad Dermatol** 35(3 Pt 1): 493-494.
- Gardner EJ and Richards RC (1953). "Multiple cutaneous and subcutaneous lesions occurring simultaneously with hereditary polyposis and osteomatosis." **Am J Hum Genet** 5(2): 139-147.

Gibbons RJ, McDowell TL, Raman S, et al. (2000). "Mutations in ATRX, encoding a SWI/SNF-like protein, cause diverse changes in the pattern of DNA methylation." **Nat Genet** 24(4): 368-371.

Groden J, Thliveris A, Samowitz W, et al. (1991). "Identification and characterization of the familial adenomatous polyposis coli gene." **Cell** 66(3): 589-600.

Hadjihannas MV, Bruckner M, Jerchow B, et al. (2006). "Aberrant Wnt/beta-catenin signaling can induce chromosomal instability in colon cancer." **Proc Natl Acad Sci U S A** 103(28): 10747-10752.

Hamai Y, Oue N, Mitani Y, et al. (2003). "DNA hypermethylation and histone hypoacetylation of the HLTF gene are associated with reduced expression in gastric carcinoma." **Cancer Sci** 94(8): 692-698.

Hamilton SR, Liu B, Parsons RE, et al. (1995). "The molecular basis of Turcot's syndrome." **N Engl J Med** 332(13): 839-847.

Hanahan D and Weinberg RA (2000). "The hallmarks of cancer." **Cell** 100(1): 57-70.

Hemminki A, Markie D, Tomlinson I, et al. (1998). "A serine/threonine kinase gene defective in Peutz-Jeghers syndrome." **Nature** 391(6663): 184-187.

Hemminki A, Tomlinson I, Markie D, et al. (1997). "Localization of a susceptibility locus for Peutz-Jeghers syndrome to 19p using comparative genomic hybridization and targeted linkage analysis." **Nat Genet** 15(1): 87-90.

Hemminki K and Chen B (2005). "Familial risks for colorectal cancer show evidence on recessive inheritance." **Int J Cancer** 115(5): 835-838.

Herman JG, Umar A, Polyak K, et al. (1998). "Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma." **Proc Natl Acad Sci U S A** 95(12): 6870-6875.

Hibi K, Nakayama H, Kanyama Y, et al. (2003). "Methylation pattern of HLTF gene in digestive tract cancers." **Int J Cancer** 104(4): 433-436.

Higuchi T and Jass JR (2004). "My approach to serrated polyps of the colorectum." **J Clin Pathol** 57(7): 682-686.

Hiramoto T, Nakanishi T, Sumiyoshi T, et al. (1999). "Mutations of a novel human RAD54 homologue, RAD54B, in primary cancer." **Oncogene** 18(22): 3422-3426.

Hoeijmakers JH (2001). "Genome maintenance mechanisms for preventing cancer." **Nature** 411(6835): 366-374.

Houlston RS, Collins A, Slack J and Morton NE (1992). "Dominant genes for colorectal cancer are not rare." **Ann Hum Genet** 56 ( Pt 2): 99-103.

Howe JR, Sayed MG, Ahmed AF, et al. (2004). "The prevalence of MADH4 and BMPR1A mutations in juvenile polyposis and absence of BMPR2, BMPR1B, and ACVR1 mutations." **J Med Genet** 41(7): 484-491.

Huebner RJ and Todaro GJ (1969). "Oncogenes of RNA tumor viruses as determinants of cancer." **Proc Natl Acad Sci U S A** 64(3): 1087-1094.

Isinger A, Bhat M, Borg A and Nilbert M (2006). "CHEK2 1100delC in patients with metachronous cancers of the breast and the colorectum." **BMC Cancer** 6: 64.

Jaeger EE, Woodford-Richens KL, Lockett M, et al. (2003). "An ancestral Ashkenazi haplotype at the HMPS/CRAC1 locus on 15q13-q14 is associated with hereditary mixed polyposis syndrome." **Am J Hum Genet** 72(5): 1261-1267.

Jallepalli PV and Lengauer C (2001). "Chromosome segregation and cancer: cutting through the mystery." **Nat Rev Cancer** 1(2): 109-117.

Jarvinen HJ (2004). "Hereditary cancer: guidelines in clinical practice. Colorectal cancer genetics." **Ann Oncol** 15 Suppl 4: iv127-131.

Jass JR (2004). "Role of the pathologist in the diagnosis of hereditary non-polyposis colorectal cancer." **Dis Markers** 20(4-5): 215-224.

Jass JR (2005). "Serrated adenoma of the colorectum and the DNA-methylator phenotype." **Nat Clin Pract Oncol** 2(8): 398-405.

Jass JR, Do KA, Simms LA, et al. (1998). "Morphology of sporadic colorectal cancer with DNA replication errors." **Gut** 42(5): 673-679.

Jen J, Powell SM, Papadopoulos N, et al. (1994). "Molecular determinants of dysplasia in colorectal lesions." **Cancer Res** 54(21): 5523-5526.

Johns LE and Houlston RS (2001). "A systematic review and meta-analysis of familial colorectal cancer risk." **Am J Gastroenterol** 96(10): 2992-3003.

Jones PA and Baylin SB (2002). "The fundamental role of epigenetic events in cancer." **Nat Rev Genet** 3(6): 415-428.

Kaplan KB, Burds AA, Swedlow JR, et al. (2001). "A role for the Adenomatous Polyposis Coli protein in chromosome segregation." **Nat Cell Biol** 3(4): 429-432.

Kemp Z, Carvajal-Carmona L, Spain S, et al. (2006). "Evidence for a colorectal cancer susceptibility locus on chromosome 3q21-q24 from a high-density SNP genome-wide linkage scan." **Hum Mol Genet**.

Kemp ZE, Carvajal-Carmona LG, Barclay E, et al. (2006). "Evidence of linkage to chromosome 9q22.33 in colorectal cancer kindreds from the United Kingdom." **Cancer Res** 66(10): 5003-5006.

Kilpivaara O, Laiho P, Aaltonen LA and Nevanlinna H (2003). "CHEK2 1100delC and colorectal cancer." **J Med Genet** 40(10): e110.

Kinzler KW and Vogelstein B (1996). "Lessons from hereditary colorectal cancer." **Cell** 87(2): 159-170.

Kinzler KW and Vogelstein B (1997). "Cancer-susceptibility genes. Gatekeepers and caretakers." **Nature** 386(6627): 761, 763.

Knudson AG, Jr. (1971). "Mutation and cancer: statistical study of retinoblastoma." **Proc Natl Acad Sci U S A** 68(4): 820-823.

Lengauer C, Kinzler KW and Vogelstein B (1997). "Genetic instability in colorectal cancers." **Nature** 386(6625): 623-627.

Leung WK, Yu J, Bai AH, et al. (2003). "Inactivation of helicase-like transcription factor by promoter hypermethylation in human gastric cancer." **Mol Carcinog** 37(2): 91-97.

Lichtenstein P, Holm NV, Verkasalo PK, et al. (2000). "Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland." **N Engl J Med** 343(2): 78-85.

Liljegren A, Lindblom A, Rotstein S, et al. (2003). "Prevalence and incidence of hyperplastic polyps and adenomas in familial colorectal cancer: correlation between the two types of colon polyps." **Gut** 52(8): 1140-1147.

Lindblom A (2001). "Different mechanisms in the tumorigenesis of proximal and distal colon cancers." **Curr Opin Oncol** 13(1): 63-69.

Lindblom A, Tannergard P, Werelius B and Nordenskjold M (1993). "Genetic mapping of a second locus predisposing to hereditary non-polyposis colon cancer." **Nat Genet** 5(3): 279-282.

Lindgren G, Liljegren A, Jaramillo E, et al. (2002). "Adenoma prevalence and cancer risk in familial non-polyposis colorectal cancer." **Gut** 50(2): 228-234.

Lindor NM, Rabe K, Petersen GM, et al. (2005). "Lower cancer incidence in Amsterdam-I criteria families without mismatch repair deficiency: familial colorectal cancer type X." **Jama** 293(16): 1979-1985.

Lipton L, Fleischmann C, Sieber OM, et al. (2003). "Contribution of the CHEK2 1100delC variant to risk of multiple colorectal adenoma and carcinoma." **Cancer Lett** 200(2): 149-152.

Lipton L, Halford SE, Johnson V, et al. (2003). "Carcinogenesis in MYH-associated polyposis follows a distinct genetic pathway." **Cancer Res** 63(22): 7595-7599.

Lipton L, Thomas HJ, Eeles RA, et al. (2001). "Apparent Mendelian inheritance of breast and colorectal cancer: chance, genetic heterogeneity or a new gene?" **Fam Cancer** 1(3-4): 189-195.

Liu HX, Zhou XL, Liu T, et al. (2003). "The role of hMLH3 in familial colorectal cancer." **Cancer Res** 63(8): 1894-1899.

Liu T, Wahlberg S, Burek E, et al. (2000). "Microsatellite instability as a predictor of a mutation in a DNA mismatch repair gene in familial colorectal cancer." **Genes Chromosomes Cancer** 27(1): 17-25.

Liu T, Yan H, Kuismanen S, et al. (2001). "The role of hPMS1 and hPMS2 in predisposing to colorectal cancer." **Cancer Res** 61(21): 7798-7802.

Lynch HT, Boland CR, Gong G, et al. (2006). "Phenotypic and genotypic heterogeneity in the Lynch syndrome: diagnostic, surveillance and management implications." **Eur J Hum Genet** 14(4): 390-402.

Lynch HT and de la Chapelle A (2003). "Hereditary colorectal cancer." **N Engl J Med** 348(10): 919-932.

Markowitz S, Wang J, Myeroff L, et al. (1995). "Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability." **Science** 268(5215): 1336-1338.

Marsh DJ, Coulon V, Lunetta KL, et al. (1998). "Mutation spectrum and genotype-phenotype analyses in Cowden disease and Bannayan-Zonana syndrome, two hamartoma syndromes with germline PTEN mutation." **Hum Mol Genet** 7(3): 507-515.

- Meijers-Heijboer H, van den Ouweland A, Klijn J, et al. (2002). "Low-penetrance susceptibility to breast cancer due to CHEK2(\*)1100delC in noncarriers of BRCA1 or BRCA2 mutations." **Nat Genet** 31(1): 55-59.
- Meijers-Heijboer H, Wijnen J, Vasen H, et al. (2003). "The CHEK2 1100delC mutation identifies families with a hereditary breast and colorectal cancer phenotype." **Am J Hum Genet** 72(5): 1308-1314.
- Moinova HR, Chen WD, Shen L, et al. (2002). "HLTF gene silencing in human colon cancer." **Proc Natl Acad Sci U S A** 99(7): 4562-4567.
- Moisio AL, Jarvinen H and Peltomaki P (2002). "Genetic and clinical characterisation of familial adenomatous polyposis: a population based study." **Gut** 50(6): 845-850.
- Molatore S and Ranzani GN (2004). "Genetics of colorectal polyps." **Tech Coloproctol** 8 Suppl 2: s240-242.
- Mueller-Koch Y, Vogelsang H, Kopp R, et al. (2005). "Hereditary non-polyposis colorectal cancer: clinical and molecular evidence for a new entity of hereditary colorectal cancer." **Gut** 54(12): 1733-1740.
- Muto T, Bussey HJ and Morson BC (1975). "The evolution of cancer of the colon and rectum." **Cancer** 36(6): 2251-2270.
- Nishisho I, Nakamura Y, Miyoshi Y, et al. (1991). "Mutations of chromosome 5q21 genes in FAP and colorectal cancer patients." **Science** 253(5020): 665-669.
- Niv Y, Goel A and Boland CR (2005). "JC virus and colorectal cancer: a possible trigger in the chromosomal instability pathways." **Curr Opin Gastroenterol** 21(1): 85-89.
- Olsson L and Lindblom A (2003). "Family history of colorectal cancer in a Sweden county." **Fam Cancer** 2(2): 87-93.
- Oshimo Y, Nakayama H, Ito R, et al. (2003). "Promoter methylation of cyclin D2 gene in gastric carcinoma." **Int J Oncol** 23(6): 1663-1670.
- Parkin DM, Bray F, Ferlay J and Pisani P (2005). "Global cancer statistics, 2002." **CA Cancer J Clin** 55(2): 74-108.
- Peltomaki P, Aaltonen LA, Sistonen P, et al. (1993). "Genetic mapping of a locus predisposing to human colorectal cancer." **Science** 260(5109): 810-812.
- Peltomaki P and Vasen H (2004). "Mutations associated with HNPCC predisposition -- Update of ICG-HNPCC/INSiGHT mutation database." **Dis Markers** 20(4-5): 269-276.
- Pharoah PD, Antoniou A, Bobrow M, et al. (2002). "Polygenic susceptibility to breast cancer and implications for prevention." **Nat Genet** 31(1): 33-36.
- Potter JD (1999). "Colorectal cancer: molecules and populations." **J Natl Cancer Inst** 91(11): 916-932.
- Powell SM, Zilz N, Beazer-Barclay Y, et al. (1992). "APC mutations occur early during colorectal tumorigenesis." **Nature** 359(6392): 235-237.
- Rodrigues NR, Rowan A, Smith ME, et al. (1990). "p53 mutations in colorectal cancer." **Proc Natl Acad Sci U S A** 87(19): 7555-7559.
- Sakai T, Toguchida J, Ohtani N, et al. (1991). "Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene." **Am J Hum Genet** 48(5): 880-888.

Sampson JR, Dolwani S, Jones S, et al. (2003). "Autosomal recessive colorectal adenomatous polyposis due to inherited mutations of MYH." **Lancet** 362(9377): 39-41.

Sampson JR, Jones S, Dolwani S and Cheadle JP (2005). "MutYH (MYH) and colorectal cancer." **Biochem Soc Trans** 33(Pt 4): 679-683.

Sanger F, Nicklen S and Coulson AR (1977). "DNA sequencing with chain-terminating inhibitors." **Proc Natl Acad Sci U S A** 74(12): 5463-5467.

Santarosa M and Ashworth A (2004). "Haploinsufficiency for tumour suppressor genes: when you don't need to go all the way." **Biochim Biophys Acta** 1654(2): 105-122.

Santoro M, Melillo RM, Carlomagno F, et al. (2002). "Molecular mechanisms of RET activation in human cancer." **Ann N Y Acad Sci** 963: 116-121.

Schmitz U, Mueller W, Weber M, et al. (2001). "INI1 mutations in meningiomas at a potential hotspot in exon 9." **Br J Cancer** 84(2): 199-201.

Schwartz S, Jr., Yamamoto H, Navarro M, et al. (1999). "Frameshift mutations at mononucleotide repeats in caspase-5 and other target genes in endometrial and gastrointestinal cancer of the microsatellite mutator phenotype." **Cancer Res** 59(12): 2995-3002.

Sieber OM, Lipton L, Crabtree M, et al. (2003). "Multiple colorectal adenomas, classic adenomatous polyposis, and germ-line mutations in MYH." **N Engl J Med** 348(9): 791-799.

Skoglund J, Djureinovic T, Zhou XL, et al. (2006). "Linkage analysis in a large Swedish family supports the presence of a susceptibility locus for adenoma and colorectal cancer on chromosome 9q22.32-31.1." **J Med Genet** 43(2): e7.

Sobel E and Lange K (1996). "Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics." **Am J Hum Genet** 58(6): 1323-1337.

Soravia C, Berk T, Madlensky L, et al. (1998). "Genotype-phenotype correlations in attenuated adenomatous polyposis coli." **Am J Hum Genet** 62(6): 1290-1301.

Souza RF, Appel R, Yin J, et al. (1996). "Microsatellite instability in the insulin-like growth factor II receptor gene in gastrointestinal tumours." **Nat Genet** 14(3): 255-257.

Suter CM, Martin DI and Ward RL (2004). "Germline epimutation of MLH1 in individuals with multiple cancers." **Nat Genet** 36(5): 497-501.

Tashiro H, Blazes MS, Wu R, et al. (1997). "Mutations in PTEN are frequent in endometrial carcinoma but rare in other common gynecological malignancies." **Cancer Res** 57(18): 3935-3940.

Tomlinson I, Rahman N, Frayling I, et al. (1999). "Inherited susceptibility to colorectal adenomas and carcinomas: evidence for a new predisposition gene on 15q14-q22." **Gastroenterology** 116(4): 789-795.

Toyota M, Ahuja N, Ohe-Toyota M, et al. (1999). "CpG island methylator phenotype in colorectal cancer." **Proc Natl Acad Sci U S A** 96(15): 8681-8686.

Umar A, Boland CR, Terdiman JP, et al. (2004). "Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability." **J Natl Cancer Inst** 96(4): 261-268.

- Vagefi PA and Longo WE (2005). "Colorectal cancer in patients with inflammatory bowel disease." **Clin Colorectal Cancer** 4(5): 313-319.
- Wagner A, Hendriks Y, Meijers-Heijboer EJ, et al. (2001). "Atypical HNPCC owing to MSH6 germline mutations: analysis of a large Dutch pedigree." **J Med Genet** 38(5): 318-322.
- Wahlberg S, Liu T, Lindblom P and Lindblom A (1999). "Various mutation screening techniques in the DNA mismatch repair genes hMSH2 and hMLH1." **Genet Test** 3(3): 259-264.
- Wallis YL, Morton DG, McKeown CM and Macdonald F (1999). "Molecular analysis of the APC gene in 205 families: extended genotype-phenotype correlations in FAP and evidence for the role of APC amino acid changes in colorectal cancer predisposition." **J Med Genet** 36(1): 14-20.
- Vandrovcova J, Stekrova J, Kebrdlova V and Kohoutova M (2004). "Molecular analysis of the APC and MYH genes in Czech families affected by FAP or multiple adenomas: 13 novel mutations." **Hum Mutat** 23(4): 397.
- Vasen HF, Mecklin JP, Khan PM and Lynch HT (1991). "The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC)." **Dis Colon Rectum** 34(5): 424-425.
- Vasen HF, Watson P, Mecklin JP and Lynch HT (1999). "New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC." **Gastroenterology** 116(6): 1453-1456.
- Weisenberger DJ, Siegmund KD, Campan M, et al. (2006). "CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer." **Nat Genet** 38(7): 787-793.
- Versteeg I, Sevenet N, Lange J, et al. (1998). "Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer." **Nature** 394(6689): 203-206.
- Whitelaw SC, Murday VA, Tomlinson IP, et al. (1997). "Clinical and molecular features of the hereditary mixed polyposis syndrome." **Gastroenterology** 112(2): 327-334.
- Wiesner GL, Daley D, Lewis S, et al. (2003). "A subset of familial colorectal neoplasia kindreds linked to chromosome 9q22.2-31.2." **Proc Natl Acad Sci U S A** 100(22): 12961-12965.
- Vogelstein B and Kinzler KW (1993). "The multistep nature of cancer." **Trends Genet** 9(4): 138-141.
- Vogelstein B and Kinzler KW (2004). "Cancer genes and the pathways they control." **Nat Med** 10(8): 789-799.
- Wong AK, Shanahan F, Chen Y, et al. (2000). "BRG1, a component of the SWI-SNF complex, is mutated in multiple human tumor cell lines." **Cancer Res** 60(21): 6171-6177.
- Yamamoto H, Sawai H and Perucho M (1997). "Frameshift somatic mutations in gastrointestinal cancer of the microsatellite mutator phenotype." **Cancer Res** 57(19): 4420-4426.
- Yuge M, Nagai H, Uchida T, et al. (2000). "hSNF5/INI1 gene mutations in lymphoid malignancy." **Cancer Genet Cytogenet** 122(1): 37-42.

Zhou XL, Djureinovic T, Werelius B, et al. (2005). "Germline mutations in the MYH gene in Swedish familial and sporadic colorectal cancer." **Genet Test** 9(2): 147-151.